Brisk guide to Mathematics

Jan Slovák

and

Martin Panák, Michal Bulant, Vladimir Ejov, Ray Booth

Brno, Adelaide, 2018

Authors:
Ray Booth
Michal Bulant
Vladimir Ezhov
Martin Panák
Jan Slovák

With further help of:
Aleš Návrat
Michal Veselý
...

Graphics and illustrations:
Petra Rychlá

2018 Masaryk University, Flinders University

Contents – practice

Contents – theory

Preface

The motivation for this textbook came from many years of lecturing Mathematics at the Faculty of Informatics at the Masaryk University in Brno. The programme requires introduction to genuine mathematical thinking and precision. The endeavor was undertaken by Jan Slovák and Martin Panák since 2004, with further collaborators joining later. Our goal was to cover seriously, but quickly, about as much of mathematical methods as usually seen in bigger courses in the classical Science and Technology programmes. At the same time, we did not want to give up the completeness and correctness of the mathematical exposition. We wanted to introduce and explain more demanding parts of Mathematics together with elementary explicit examples how to use the concepts and results in practice. But we did not want to decide how much of theory or practice the reader should enjoy and in which order.

All these requirements have lead us to the two column format of the textbook, where the theoretical explanation on one side and the practical procedures and exercises on the other side are split. This way, we want to encourage and help the readers to find their own way. Either to go through the examples and algorithms first, and then to come to explanations why the things work, or the other way round. We also hope to overcome the usual stress of the readers horrified by the amount of the stuff. With our text, they are not supposed to read through the book in a linear order. On the opposite, the readers should enjoy browsing through the text and finding their own thrilling paths through the new mathematical landscapes.

In both columns, we intend to present rather standard exposition of basic Mathematics, but focusing on the essence of the concepts and their relations. The exercises are addressing simple mathematical problems but we also try to show the exploitation of mathematical models in practice as much as possible.

We are aware that the text is written in a very compact and non-homogeneous way. A lot of details are left to readers, in particular in the more difficult paragraphs, while we try to provide a lot of simple intuitive explanation when introducing new concepts or formulating important theorems. Similarly, the examples display the variety from very simple ones to those requesting independent thinking.

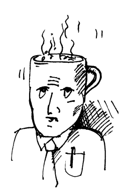
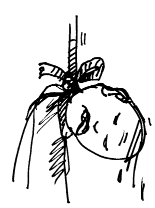
We would very much like to help the reader:

- to formulate precise definitions of basic concepts and to prove simple mathematical results;
- to percieve the meaning of roughly formulated properties, relations and outlooks for exploring mathematical tools;
- to understand the instructions and algorithms underlying mathematical models and to appreciate their usage.

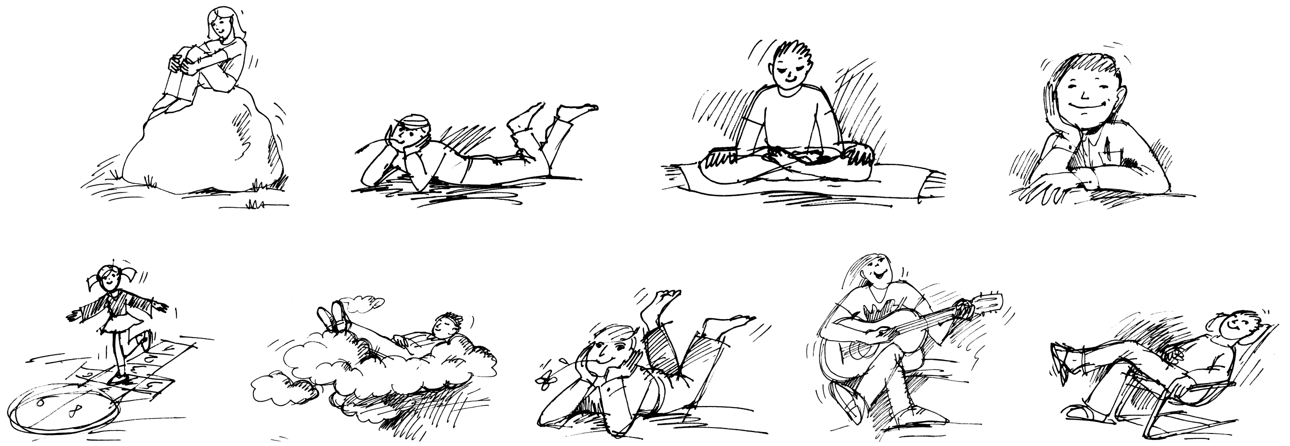
These goals are ambitious and there are no simple paths reaching them without failures on the way. This is one of the reasons why we come back to basic ideas and concepts several times with growing complexity and width of the discussions. Of course, this might also look chaotic but we very much hope that this approach gives a better chance to those who will persist in their efforts. We also hope, this textbook should be a perfect beginning and help for everybody who is ready to think and who is ready to return back to earlier parts again and again.

To make the task simpler and more enjoyable, we have added what we call "emotive icons". We hope they will spirit the dry mathematical text and indicate which parts should be read more carefully, or better left out in the first round.

The usage of the icons follows the feelings of the authors and we tried to use them in a systematic way. We hope the readers will assign the meaning to icons individually. Roughly speaking, we are using icons to indicate complexity, difficulty etc.:

Further icons indicate unpleasant technicality and need of patiance, or possible entertainment and pleasure:



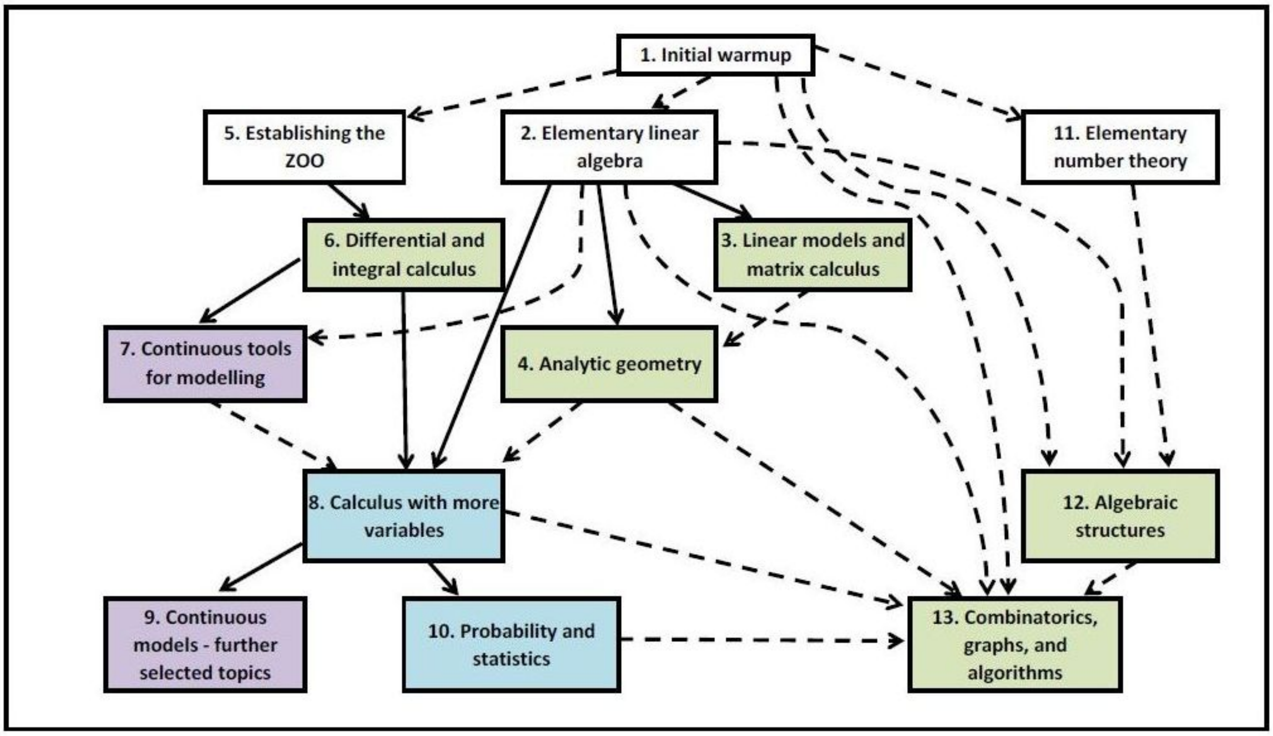
Similarly, we use various icons in the practical column:

The practical column with the solved problems and exercises should be readable nearly independently of the theory. Without the ambition to know the deeper reasons why the algorithms work, it should be possible to read mainly just this column. In order to help such readers, some definitions and descriptions in the theoretical text are marked in order to catch the eyes easily when reading the exercises. The exercises and theory are partly coordinated to allow jumping there and back, but the links are not tight. The numbering in the two columns is distinguished by using the different numberings of sections, i.e. those like **??** belong to the theoretical column, while **??** points to the practical column. The equations are numbered within subsections and their quotes include the subsection numbers if necessary.

In general, our approach stresses the fact that the methods of the so called discrete Mathematics seem to be more important for mathematical models nowadays. They seem also simpler to get percieved and grasped.

However, the continuous methods are strictly necessary too. First of all, the classical continuous mathematical analysis is essential for understanding of convergence and robustness of computations. It is hard to imagine how to deal with error estimates and computational complexity of numerical processes without it. Moreover, the continuous models are often the efficient and effectively computable approximations to discrete problems coming from practice.

The rough structure of the book and the dependencies between its chapters are depicted in the diagram below. The darker the color is, the more demanding is the particular chapter (or at least its essential parts). In particular, the chapters 7 and 9 include a lot of material which would perhaps not be covered in the regular course activities or required at exams in great detail. The solid arrows mean strong dependencies, while the dashed links indicate only partial dependencies. In particular, the textbook could support courses starting with any of the white boxes, i.e. aiming at standard linear algebra and geometry (chapters 2 through 4), discrete chapters of mathematics (11 through 13), and the rudiments of Calculus (5, 6, 8).

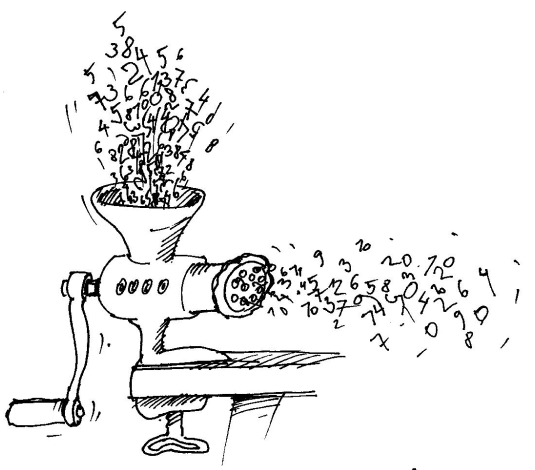


All topics covered in the book have been included (with more or less details) in our teaching of large four semester courses of Mathematics, complemented by numerical seminars, since 2005. In our teaching, the first semester covered chapters 1 and 2 and selected topics from chapters 3 and 4. The second semester fully included chapters 5 and 6 and selected topics from chapter 7. The third semester was split into two parts. The first one was covered by chapter 8, while the rest of the semester was devoted to chapter 10 (with only a few glimpses towards the more advanced topics from chapter 9). The last semester provided large parts of the content of chapters 11 through 13, although the entire graph theory was skipped (since it was tought elsewhere). Actually, the second semester could be offered in parallel with the first one, while the fourth semester could follow immediately after the first one. Indeed, some students were advised to go for the second and fourth semester simultaneously (those in the IT security programme).

CHAPTER 2

Elementary linear algebra

Can't you count with scalars yet?
– no worry, let us go straight to matrices...



A. Systems of linear equations and matrix manipulation

We approach vector spaces in a clever way. We begin with something we know – systems of linear equations and find that the vector spaces are hidden behind them.

In the previous chapter we warmed up by considering relatively simple problems which did not require any sophisticated tools. It was enough to use addition and multiplication of scalars. In this and subsequent chapters we shall add more sophisticated thoughts and tools.

First we restrict ourselves to concepts and operations consisting of a finite number of multiplications and additions to a finite number of scalars. This will take us three chapters and only then will we move on to infinitesimal concepts and tools. Typically we deal with finite collections of scalars of a given size. We speak about "linear objects" and "linear algebra". Although it might seem to be a very special tool, we shall see later that even more complicated objects are studied mostly using their "linear approximations".

In this chapter we will work with finite sequences of scalars. Such sequences arise in real-world problems whenever we deal with objects described by several parameters, which we shall call coordinates. Do not try much to imagine the space with more than three coordinates. You have to live with the fact that we are able to depict only one, two or three dimensions. However, we will deal with an arbitrary number of dimensions. For example, observing any parameter in a group 500 students (for instance, their study results), our data will have 500 elements and we would like to work with them. Our goal is to develop tools which will work well even if the number of elements is large.

Do not be afraid of terms like field or ring of scalars $\mathbb{K}$. Simply, imagine any specific domain of numbers. Rings of scalars are for instance integers $\mathbb{Z}$ and all residue classes $\mathbb{Z}_k$. Among fields we have seen only $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ and residue classes $\mathbb{Z}_k$ for k prime. $\mathbb{Z}_2$ is very specific among them, because the equation $x = -x$ does not imply $x = 0$ here, whereas in every other field it does.

1. Vectors and matrices

In the first two parts of this chapter, we will work with vectors and matrices in the simple context of finite sequences of scalars. We can imagine working with integers or residue classes as well as real or complex numbers. We hope to illustrate how easily a concise and formal reasoning can lead to strong results valid in a much broader context than just for real numbers.

2.A.1. A colourful example. A company of painters orders

810 litres of paint, to contain 270 litres each of red, green and blue coloured paint. The provider can satisfy this order by mixing the colours he usually sells (he has enough in his warehouse). He has

- reddish colour – it contains 50 % of red, 25 % of green and 25 % of blue colour;
- greenish colour – it contains 12,5 % of red, 75 % of green and 12,5 % of blue colour;
- bluish colour – it contains 20 % of red, 20 % of green and 60 % of blue colour.

How many litres of each of the colours at the warehouse have to be mixed in order to satisfy the order?

Solution. Denote by

- x – the number of litres of reddish colour to be used;
- y – the number of litres of bluish colour to be used;
- z – the number of litres) greenish colour to be used;

By mixing the colours we want a colour that contains 270 litres of red. Note that reddish contains 50 % red, greenish contains 12,5 % red and bluish 20 % red. Thus the following has to be satisfied:

$$0,5x \quad + \quad 0,125y \quad + \quad 0,2z \quad = \quad 270.$$

Similarly, we require (for blue and green colours respectively) that

$$\begin{array}{rcrcrcl} 0,25x & + & 0,75y & + & 0,2z & = & 270, \\ 0,25x & + & 0,125y & + & 0,6z & = & 270. \end{array}$$

From the first equation $x = 540 - 0,25y - 0,4z$. Substitute for x into the second and third equations to obtain two linear equations of two variables $2,75y + 0,4z = 540$ and $0,25y + 2z = 540$. From the second of these we express $z = 270 - 0,125y$ and substitute into the first one we obtain $2,7y = 432$, that is, $y = 160$. Therefore $z = 270 - 0,125 \cdot 160 = 250$ and hence $x = 540 - 0,25 \cdot 160 + 0,4 \cdot 250 = 400$.

An alternative approach is to deduce consequences from the given equations by a sequence of adding them or multiplying them by non-zero scalars. This is easily handled in the matrix notation (which we met when solving equations with two variables in the previous chapter already). The first row of the matrix consists of coefficients of the variables in the first equation, second of the coefficients in the second equation and third of the coefficients in the third. Therefore the

Later, we follow the general terminology where the notion of vectors is related to fields of scalars only.

2.1.1. Vectors over scalars. For now, a *vector* is for us an ordered n-tuple of scalars from $\mathbb{K}$, where the fixed $n \in \mathbb{N}$ is called *dimension*.

We can add and multiply scalars. We will be able to add vectors, but multiplying a vector will be possible only by a scalar. This corresponds to the idea we have already seen in the plane $\mathbb{R}^2$. There, addition is realized as vector composition (as composition of arrows having their direction and size and compared when emanating from the origin). Multiplication by scalar is realized as stretching the vectors.

A vector $u = (a_1, \dots, a_n)$ is multiplied by a scalar c by multiplying every element of the n-tuple u by c. Addition is defined coordinate-wise.

Basic vector operations

$$\begin{aligned} u + v &= (a_1, \dots, a_n) + (b_1, \dots, b_n) \\ &= (a_1 + b_1, \dots, a_n + b_n) \\ c \cdot u &= c \cdot (a_1, \dots, a_n) = (c \cdot a_1, \dots, c \cdot a_n). \\ cu &= c(a_1, \dots, a_n) = (ca_1, \dots, ca_n). \end{aligned}$$

For vector addition and multiplication by scalars we shall use the same symbols as for scalars, that is, respectively, plus and either dot or juxtaposition.

The vector notation convention. We shall not, unlike many

other textbooks, use any special notations for vectors and leave it to the reader to pay attention to the context. For scalars, we shall mostly use letters from the beginning of the alphabet, for the vector from the end of the alphabet. The middle part of the alphabet can be used for indices of variables or components and also for summation indices.

Later we shall require that the scalars are from some specific field, but for now we will work with the more relaxed properties of scalars as listed in **??**. In the general theory in the end of this chapter and later, we will work exclusively with fields of scalars.

For vector addition in $\mathbb{K}^n$, the properties (CG1)–(CG4) (see **??**) clearly hold with the zero element being (notice we define the addition coordinate-wise)

$$0 = (0, \dots, 0) \in \mathbb{K}^n.$$

We are purposely using the same symbol for both the zero vector element and the zero scalar element. Next, let us notice the following basic properties of vectors:

matrix of the system is

$$\begin{pmatrix} 0,5 & 0,125 & 0,2 \\ 0,25 & 0,75 & 0,2 \\ 0,25 & 0,125 & 0,6 \end{pmatrix},$$

The *extended matrix of the system* is obtained from the matrix of the system by inserting the column of the right-hand sides of the individual equations in the system:

$$\left(\begin{array}{ccc|c} 0,5 & 0,125 & 0,2 & 270 \\ 0,25 & 0,75 & 0,2 & 270 \\ 0,25 & 0,125 & 0,6 & 270 \end{array}\right)$$

By doing elementary row transformations sequentially (they all correspond to adding rows and multiplication by scalars with the equations, see 2.1.7) we can eliminate the variables in the equations, one by one:

$$\left(\begin{array}{ccc|c} 0,5 & 0,125 & 0,2 & 270 \\ 0,25 & 0,75 & 0,2 & 270 \\ 0,25 & 0,125 & 0,6 & 270 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0,25 & 0,4 & 540 \\ 1 & 3 & 0,8 & 1\,080 \\ 1 & 0,5 & 2,4 & 1\,080 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|c} 1 & 0,25 & 0,4 & 540 \\ 0 & 2,75 & 0,4 & 540 \\ 0 & 0,25 & 2 & 540 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0,25 & 0,4 & 540 \\ 0 & 11 & 1,6 & 2\,160 \\ 0 & 1 & 8 & 2\,160 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|c} 1 & 0,25 & 0,4 & 540 \\ 0 & 1 & 8 & 2\,160 \\ 0 & 11 & 1,6 & 2\,160 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0,25 & 0,4 & 540 \\ 0 & 1 & 8 & 2\,160 \\ 0 & 0 & -86,4 & -21\,600 \end{array}\right).$$

By back substitution, we compute successively

$$z = \frac{-21\,600}{-86,4} = 250,$$
$$y = 2\,160 - 8 \cdot 250 = 160,$$
$$x = 540 - 0,4 \cdot 250 - 0,25 \cdot 160 = 400.$$

Thus it is necessary to mix 400 litres of reddish, 160 litres of bluish and 250 litres of greenish colour. □

2.A.2. Solve the system of simultaneous linear equations

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 3x_3 & = & 2, \\ 2x_1 & - & 3x_2 & - & x_3 & = & -3, \\ -3x_1 & + & x_2 & + & 2x_3 & = & -3. \end{array}$$

Solution. We write the system of equations in the form of the extended matrix of the system

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

Every row of the matrix corresponds to one equation. As in the previous example, equivalent transformation of the equations correspond to the elementary row operations on the matrix and we use them to transform it into the row echelon form

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 2 \\ 2 & -3 & -1 & -3 \\ -3 & 1 & 2 & -3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & 3 & 2 \\ 0 & -7 & -7 & -7 \\ 0 & 7 & 11 & 3 \end{array}\right) \sim$$

Vector properties

For all vectors $v, w \in \mathbb{K}^n$ and scalars $a, b \in \mathbb{K}$ we have

(V1) $$a \cdot (v + w) = a \cdot v + a \cdot w$$

(V2) $$(a + b) \cdot v = a \cdot v + b \cdot v$$

(V3) $$a \cdot (b \cdot v) = (a \cdot b) \cdot v$$

(V4) $$1 \cdot v = v$$

5

The properties (V1)–(V4) of our vectors are easily checked for any specific ring of scalars $\mathbb{K}$, since we need just the corresponding properties of scalars as listed in **??** and **??**, applied to individual components of the vectors. In this way we shall work with, for instance, $\mathbb{R}^n$, $\mathbb{Q}^n$, $\mathbb{C}^n$, but also with $\mathbb{Z}^n$, $(\mathbb{Z}_k)^n$, $n = 1, 2, 3, \dots$.

2.1.2. Matrices over scalars. Matrices are slightly more complicated objects, useful when working with vectors.

Matrices of type m/n

A matrix of the type m/n over scalars $\mathbb{K}$ is a rectangular schema A with m rows and n columns

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

where $a_{ij} \in \mathbb{K}$ for all $1 \le i \le m,\ 1 \le j \le n$. For a matrix A with elements a_{ij} we also use the notation $A = (a_{ij})$.

The vector $(a_{i1}, a_{i2}, \dots, a_{in}) \in \mathbb{K}^n$ is called the (i-th) *row of the matrix* A, $i = 1, \dots, m$. The vector $(a_{1j}, a_{2j}, \dots, a_{mj}) \in \mathbb{K}^m$ is called the (j-th) *column of the matrix* A, $j = 1, \dots, n$.

Matrices of the type $1/n$ or $n/1$ are actually just vectors in $\mathbb{K}^n$.

All general matrices can be understood as vectors in

$\mathbb{K}^{mn}$, we just consider all the columns. In particular, matrix addition and matrix multiplication by scalars is defined:

$$A + B = (a_{ij} + b_{ij}), \quad a \cdot A = (a \cdot a_{ij})$$

where $A = (a_{ij})$, $B = (b_{ij})$, $a \in \mathbb{K}$.

The matrix $-A = (-a_{ij})$ is called the *additive inverse* to the matrix A and the matrix

$$0 = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix}$$

is called the *zero matrix*. By considering matrices as mn-dimensional vectors, we obtain the following:

Proposition. *The formulas for $A + B$, $a \cdot A$, $-A$, 0 define the operations of addition and multiplication by scalars for the set of all matrices of the type m/n, which satisfy axioms (V1)–(V4).*

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

First we subtracted from the second row twice the first row, and to the third row we added three times the first row. Then we added the second row to the third row and multiplied the second row by $-1/4$. Now we restore the system of equations

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

We see immediately that $x_3 = -1$. If we substitute $x_3 = -1$ into the equation $x_2 + x_3 = 1$, we obtain $x_2 = 2$. Then by substituting $x_3 = -1$, $x_2 = 2$ into the first equation, we obtain $x_1 = 1$. □

Systems of linear equations can be written in matrix notation. But is it an advantage, when we can solve the systems even without speaking about matrices? Yes it is, we can handle the equations more conceptually. We can easily decide how many solutions a system has. It is much more efficient in computer assisted computations. Thus we shall get familiar with various operations which can be done with matrices. As we have seen in previous examples, equivalent operations with linear equations correspond to elementary row (column) transformations. Further we have seen that transforming a matrix into a row echelon form, a process called Gaussian elimination, see 2.1.7), solves the system very easily. We demonstrate this on some examples, where we will see that a system can have infinitely many solutions or no solution at all.

2.A.3. Solve a system of linear equations

$$\begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 3x_3 & = & 0, \\ 3x_1 & + & 16x_2 & + & 7x_3 & = & 0, \\ 3x_1 & - & 5x_2 & + & 4x_3 & = & 0, \\ -7x_1 & + & 7x_2 & + & -10x_3 & = & 0. \end{array}$$

Solution. Because the right-hand side of all equations is zero (such a case is called a homogeneous system) we work with the matrix of the system only. We find the solution by transforming the matrix into the row echelon form using elementary row transformations. These correspond to changing the order of equations, multiplying an equation by a non-zero number and addition of multiples of equations. Furthermore, we can always go back and forth between the matrix notation and the original system notation with variables x_i. We obtain:

$$\begin{pmatrix} 2 & -1 & 3 \\ 3 & 16 & 7 \\ 3 & -5 & 4 \\ -7 & 7 & -10 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 \\ 0 & 35/2 & 5/2 \\ 0 & -7/2 & -1/2 \\ 0 & 7/2 & 1/2 \end{pmatrix}.$$

2.1.3. Matrices and equations. Many mathematical models are based on systems of linear equations. Matrices are useful for the description of such systems. In order to see this, let us introduce the notion of *scalar product* of two vectors, assigning to the vectors $(a_1, \dots, a_n)$ and $(x_1, \dots, x_n)$ their product

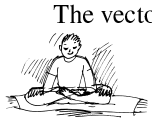
$$(a_1, \dots, a_n) \cdot (x_1, \dots, x_n) = a_1x_1 + \dots + a_nx_n.$$

This means, we multiply the corresponding coordinates of the vectors and sum the results.

Every system of m linear equations in n variables

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

can be seen as a constraint on values of m scalar products with one unknown vector $(x_1, \dots, x_n)$ (called the *vector of variables*, or *vector variable*) and the known vectors of coordinates $(a_{i1}, \dots, a_{in})$.



The vector of variables can be also seen as a column in a matrix of the type $n/1$, and similarly the values $b_1, \dots, b_n$ can be seen as a vector u, and that is again a single column of the matrix of the type $n/1$. Our system of equations can then be formally written as $A \cdot x = u$ as follows:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

where the left-hand side is interpreted as m scalar products of the individual rows of the matrix (giving rise to a column vector) with the vector variable x, whose values are prescribed by the equations. That means that the identity of the i-th coordinates corresponds to the original i-th equation

$$a_{i1}x_1 + \dots + a_{in}x_n = b_i$$

and the notation $A \cdot x = u$ gives the original system of equations.

2.1.4. Matrix product. In the plane, that is, for vectors of dimension two, we developed a matrix calculus. We noticed that it is effective to work with (see **??**). Now we generalize such a calculus and we develop all the tools we know already from the plane case to deal with higher dimensions n.

It is possible to define matrix multiplication only when the dimensions of the rows and columns allow it, that is, when the scalar product is defined for them as before:

From there we see that the second, third and fourth equations are multiples of the equation $7x_2 + x_3 = 0$. We continue:

$$\begin{pmatrix} 2 & -1 & 3 \\ 0 & 35/2 & 5/2 \\ 0 & -7/2 & -1/2 \\ 0 & 7/2 & 1/2 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 \\ 0 & 35/2 & 5/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} 2 & -1 & 3 \\ 0 & 7 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

Considered as equations, the last two are redundant, and we are left with just

$$\begin{array}{rcrcrcl} 2x_1 & - & x_2 & + & 3x_3 & = & 0, \\ & & 7x_2 & + & x_3 & = & 0 \end{array}$$

We substitute for the variable x_3 a parameter $t \in \mathbb{R}$ and express

$$x_2 = -\frac{1}{7}\,x_3 = -\frac{1}{7}\,t \quad \text{a} \quad x_1 = \frac{1}{2}\,(x_2 - 3x_3) = -\frac{11}{7}\,t.$$

If we now substitute $t = -7s$, we obtain the result in a simple form

$$(x_1,\, x_2,\, x_3) = (11s,\, s,\, -7s)\,, \qquad s \in \mathbb{R}.$$

The whole system has infinitely many solutions. □

2.A.4. Find all solutions of the system of linear equations

$$\begin{array}{rcrcrcrcl} 3x_1 & & & + & 3x_3 & - & 5x_4 & = & -8, \\ x_1 & - & x_2 & + & x_3 & - & x_4 & = & -2, \\ -2x_1 & - & x_2 & + & 4x_3 & - & 2x_4 & = & 0, \\ 2x_1 & + & x_2 & - & x_3 & - & x_4 & = & -3. \end{array}$$

Solution. The corresponding extended matrix of the system is

$$\left(\begin{array}{cccc|c} 3 & 0 & 3 & -5 & -8 \\ 1 & -1 & 1 & -1 & -2 \\ -2 & -1 & 4 & -2 & 0 \\ 2 & 1 & -1 & -1 & -3 \end{array}\right).$$

By changing the order of rows (equations) we obtain

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 2 & 1 & -1 & -1 & -3 \\ -2 & -1 & 4 & -2 & 0 \\ 3 & 0 & 3 & -5 & -8 \end{array}\right),$$

which we transform into the row echelon form:

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 2 & 1 & -1 & -1 & -3 \\ -2 & -1 & 4 & -2 & 0 \\ 3 & 0 & 3 & -5 & -8 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 0 & 3 & -3 & 1 & 1 \\ 0 & -3 & 6 & -4 & -4 \\ 0 & 3 & 0 & -2 & -2 \end{array}\right)$$

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

The system has thus infinitely many solutions, because we have three equations in four variables. These three equations have exactly one solution for any choice for the variable

MATRIX PRODUCT

For any matrix $A = (a_{ij})$ of the type m/n and any matrix $B = (b_{jk})$ of the type n/q over the ring of scalars $\mathbb{K}$ we define their product $C = A \cdot B = (c_{ik})$ as a matrix of the type m/q with the elements

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}, \text{ for arbitrary } 1 \le i \le m,\, 1 \le k \le q.$$

That is, the element c_{ik} of the product is exactly the scalar product of the i-th row of the matrix on the left and of the k-th column of the matrix on the right. For instance we have

$$\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 3 \\ 3 & 1 & 0 \end{pmatrix}.$$

2.1.5. Square matrices. If there is the same number of rows and columns in the matrix, we speak of a *square matrix*. The number of rows or columns is then called the *dimension of the matrix*. The matrix

$$E = (\delta_{ij}) = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}$$

is called the *unit matrix*, or alternatively, the *identity matrix*. The numbers δ_{ij} defined in such a way are also called the *Kronecker delta*. When we restrict ourselves to square matrices over $\mathbb{K}$ of fixed dimension n, the matrix product is defined for any two matrices. That is, there is the well defined multiplication operation there. Its properties are similar to that of scalars:

Proposition. *On the set of all square matrices of dimension n over an arbitrary ring of scalars $\mathbb{K}$, the multiplication operation is defined with the following properties of rings (see* **??***):*

(O1) Multiplication is associative.
(O3) The unit matrix $E = (\delta_{ij})$ is the unit element for multiplication.
(O4) Multiplication and addition is distributive.

In general, neither the property (O2) nor (OI) are true. Therefore, the square matrices for $n > 1$ do not form an integral domain, and consequently they cannot be a (commutative or non-commutative) field.

PROOF. Associativity of multiplication – (O1): Since scalars are associative, distributive and commutative, we can compute for any three matrices $A = (a_{ij})$ of type m/n, $B = (b_{jk})$ of type n/p and $C = (c_{kl})$ of type p/q:

$$A \cdot B = \Big(\sum_j a_{ij} \cdot b_{jk}\Big), \quad B \cdot C = \Big(\sum_k b_{jk} \cdot c_{kl}\Big),$$

$$(A \cdot B) \cdot C = \Big(\sum_k \big(\sum_j a_{ij} b_{jk}\big) c_{kl}\Big) = \Big(\sum_{j,k} a_{ij} b_{jk} c_{kl}\Big),$$

$$A \cdot (B \cdot C) = \Big(\sum_j a_{ij} \big(\sum_k b_{jk} c_{kl}\big)\Big) = \Big(\sum_{j,k} a_{ij} b_{jk} c_{kl}\Big).$$

$x_4 \in \mathbb{R}$. Thus for x_4 we substitute the parameter $t \in \mathbb{R}$ and go back from the matrix notation to the system of equations

$$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & + & x_3 & - & t & = & -2, \\ & & 3x_2 & - & 3x_3 & + & t & = & 1, \\ & & & & 3x_3 & - & 3t & = & -3. \end{array}$$

From the last equation we have $x_3 = t - 1$. Substituting for x_3 into the second equation gives

$$3x_2 - 3t + 3 + t = 1, \quad \text{that is,} \quad x_2 = \frac{1}{3}(2t - 2).$$

Finally, using the first equation, we have

$$x_1 - \frac{1}{3}(2t-2) + t - 1 - t = -2, \quad \text{tj.} \quad x_1 = \frac{1}{3}(2t - 5).$$

The set of solutions can be written (for $t = 3s$) in the form $\{(x_1, x_2, x_3, x_4) = \left(2s - \frac{5}{3}, 2s - \frac{2}{3}, 3s - 1, 3s\right), s \in \mathbb{R}\}$.

We return to the extended matrix of the system and transform it further by using the row transformations in order to have (still in the row echelon form) the first non-zero number of every row (the so-called *pivot*) equal to one and that all the other numbers in the column of the pivot are zero. We have

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

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

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

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

because first we have multiplied the second and the third row by $1/3$, then we have added the third row to the second and its (-1)-multiple to the first. Finally we have added the second row to the first. From the last matrix we easily obtain the result

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -5/3 \\ -2/3 \\ -1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2/3 \\ 2/3 \\ 1 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

Free variables are those whose columns do not contain any pivot (in our case there is no pivot in the fourth column, that is, the fourth variable is free and we use it as a parameter). □

Note that while computing, we relied on the fact that it does not matter in which order are we performing the sums and products, that is, we were relying on the properties of scalars.

We can easily see that multiplication by a unit matrix has the property of a unit element:

$$A \cdot E = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \\ a_{m1} & \cdots & a_{mm} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix} = A$$

and similarly from the left,

$$E \cdot A = A.$$

It remains to prove the distributivity of multiplication and addition. Again using the distributivity of scalars we can easily calculate for matrices $A = (a_{ij})$ of the type m/n, $B = (b_{jk})$ of the type n/p, $C = (c_{jk})$ of the type n/p, $D = (d_{kl})$ of the type p/q

$$\begin{aligned} A \cdot (B + C) &= \Big(\sum_j a_{ij}(b_{jk} + c_{jk})\Big) \\ &= \Big(\big(\sum_j a_{ij}b_{jk}\big) + \big(\sum_j a_{ij}c_{jk}\big)\Big) = A \cdot B + A \cdot C \\ (B + C) \cdot D &= \Big(\sum_k (b_{jk} + c_{jk})d_{kl}\Big) \\ &= \Big(\big(\sum_k b_{jk}d_{kl}\big) + \big(\sum_k c_{jk}d_{kl}\big)\Big) = B \cdot D + C \cdot D. \end{aligned}$$

As we have seen in **??**, two matrices of dimension two do not necessarily commute: for example

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This gives us immediately a counterexample to the validity of (O2) and (OI). For matrices of type $1/1$ both axioms clearly hold, because the scalars itself have them. For matrices of greater dimension the counterexamples can be obtained similarly. Simply place the counterexamples for dimension 2 in their left upper corner, and select the rest to be zero. (Verify this on your own!) □

In the proof we have actually worked with matrices of more general types, thus we have proved the properties in greater generality:

2.A.5. Determine the solutions of the system of equations

$$\begin{array}{rcrcrcrcr} 3x_1 & & & + & 3x_3 & - & 5x_4 & = & 8, \\ x_1 & - & x_2 & + & x_3 & - & x_4 & = & -2, \\ -2x_1 & - & x_2 & + & 4x_3 & - & 2x_4 & = & 0, \\ 2x_1 & + & x_2 & - & x_3 & - & x_4 & = & -3. \end{array}$$

Solution. Note that the system of equations in this exercise differs from the system of equations in the previous exercise only in the value 8 (instead of -8) on the right-hand side. If we do the same row transformations as in the previous exercise, we obtain

$$\left(\begin{array}{cccc|c} 3 & 0 & 3 & -5 & 8 \\ 1 & -1 & 1 & -1 & -2 \\ -2 & -1 & 4 & -2 & 0 \\ 2 & 1 & -1 & -1 & -3 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 2 & 1 & -1 & -1 & -3 \\ -2 & -1 & 4 & -2 & 0 \\ 3 & 0 & 3 & -5 & 8 \end{array}\right)$$

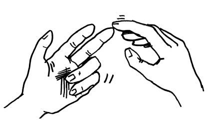
$$\sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 0 & 3 & -3 & 1 & 1 \\ 0 & 0 & 3 & -3 & -3 \\ 0 & 0 & 3 & -3 & 13 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & -1 & 1 & -1 & -2 \\ 0 & 3 & -3 & 1 & 1 \\ 0 & 0 & 3 & -3 & -3 \\ 0 & 0 & 0 & 0 & 16 \end{array}\right),$$

where the last operation was subtracting the third row from the fourth. From the fourth equation $0 = 16$ follows that the system has no solutions. Let us emphasize than whenever we obtain an equation of the form $0 = a$ for some $a \neq 0$ (that is, zero row on the left side and non-zero number after the vertical bar) when doing the row transformation, the system has no solutions. □

You can find more exercises for systems of systems of linear equations on the page 58

Now we are going to manipulate with matrices to get more familiar with their properties.

2.A.6. Matrix multiplication. Note that, in order to be able



to multiply two matrices, the necessary and sufficient condition is that the first matrix has the same number of columns as the number of rows of the second matrix. The number of rows of the resulting matrix is then given by the number of rows of the first matrix, the number of columns then equals the number of columns of the second matrix.

i) $\begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 1 \\ 5 & 4 \end{pmatrix}$,

ii) $\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 7 \end{pmatrix}$,

iii) $\begin{pmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 2 & 1 \\ 1 & 1 & -2 & -3 \\ 3 & 2 & 1 & 0 \end{pmatrix}$

$= \begin{pmatrix} 12 & 7 & 1 & -5 \\ 3 & 0 & 5 & 4 \end{pmatrix}$,

Associativity and distributivity

Matrix multiplication is associative and distributive, that is,

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C$$
$$A \cdot (B + C) = A \cdot B + A \cdot C,$$

whenever are all the given operations defined. The unit matrix is a unit element for multiplication (both from the right and from the left).

5

2.1.6. Inverse matrices. With scalars we can do the follow-

ing: from the equation $a \cdot x = b$ with a fixed invertible a we can express $x = a^{-1} \cdot b$ for any b. We would like to be able to do this for matrices too. So we need to solve the problem – how to tell that such a matrix exists, and if so, how to compute it?

We say that B is the *inverse* of A if

$$A \cdot B = B \cdot A = E.$$

Then we write $B = A^{-1}$. From the definition it is clear that both matrices must be square and of the same dimension n. A matrix which has an inverse is called an *invertible matrix* or a *regular square matrix*.

In the subsequent paragraphs we derive (among other things) that B is actually the inverse of A whenever just one of the above required equations holds. The other is then a consequence.

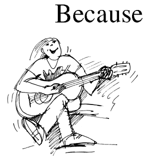
We easily check that if A^{-1} and B^{-1} exist, then there also is the inverse of the product $A \cdot B$

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}. \tag{1}$$

Indeed, because of the associativity of matrix multiplication proved a while ago, we have

$$(B^{-1} \cdot A^{-1}) \cdot (A \cdot B) = B^{-1} \cdot (A^{-1} \cdot A) \cdot B = E$$
$$(A \cdot B) \cdot (B^{-1} \cdot A^{-1}) = A \cdot (B \cdot B^{-1}) \cdot A^{-1} = E.$$

Because we can calculate with matrices similarly as with



scalars (they are just a little more complicated), the existence of an inverse matrix can really help us with the solution of systems of linear equations: if we express a system of n equations for n unknowns as a matrix product

$$A \cdot x = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \\ a_{m1} & \cdots & a_{mm} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = u$$

and when the inverse of the matrix A exists, then we can multiply from the left by A^{-1} to obtain

$$A^{-1} \cdot u = A^{-1} \cdot A \cdot x = E \cdot x = x,$$

that is, $A^{-1} \cdot u$ is the desired solution.

On the other hand, expanding the condition $A \cdot A^{-1} = E$ for unknown scalars in the matrix A^{-1} gives us n systems of linear equations for the same matrix on the left and different

iv) $\begin{pmatrix} 1 & 3 & 1 \\ -2 & 2 & -1 \\ 3 & 1 & -4 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \\ -3 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \\ 18 \end{pmatrix}$,

v) $\begin{pmatrix} 1 & 3 & -3 \end{pmatrix} \cdot \begin{pmatrix} 1 & -2 & 3 \\ 3 & 2 & 1 \\ 1 & -1 & -4 \end{pmatrix} = \begin{pmatrix} 7 & 7 & 18 \end{pmatrix}$,

vi) $\begin{pmatrix} 1 & 2 & -2 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \end{pmatrix}$.

Remark. Parts i) and ii) in the previous exercise show that multiplication of square matrices is not commutative in general. In part iii) we see that if we can multiply two rectangular matrices, then it is possible only in one of the orders. In parts iv) and v) note that $(A \cdot B)^T = B^T \cdot A^T$.

2.A.7. Let

$$A = \begin{pmatrix} 4 & 0 & -5 \\ 2 & 7 & 15 \\ 2 & 7 & 13 \end{pmatrix}, \quad B = \begin{pmatrix} 7 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & -19 & \sqrt{13} \end{pmatrix}.$$

Can the matrix A be transformed into B using only elementary row transformations (we say then that such matrices are row equivalent)?

Solution. Both matrices are row equivalent with the three-dimensional identity matrix. It is easy to see that row equivalence on the set of all matrices of given type is indeed an equivalence relation. Thus the matrices A and B are row equivalent. □

2.A.8. Find a matrix B for which the matrix $C = B \cdot A$ is in row echelon form, where

$$A = \begin{pmatrix} 3 & -1 & 3 & 2 \\ 5 & -3 & 2 & 3 \\ 1 & -3 & -5 & 0 \\ 7 & -5 & 1 & 4 \end{pmatrix}.$$

Solution. If we multiply the matrix A successively from the left by elementary matrices (consider what elementary row transformations does it correspond to)

$$E_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -5 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$E_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -7 & 0 & 0 & 1 \end{pmatrix},$$

$$E_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_6 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

vectors on the right. Thus we should think about methods for solutions of the systems of linear equations.

2.1.7. Equivalent operations with matrices. Let us gain some practical insight into the relation between systems of equations and their matrices. Clearly, searching for the inverse can be more complicated than finding the direct solution to the system of equations. But note that whenever we have to solve more systems of equations with the same matrix A but with different right sides u, then yielding A^{-1} can be really beneficial for us.

From the point of view of solving systems of equations $A \cdot x = u$, it is natural to consider the matrices A and vectors u equivalent whenever they give a system of equations with the same solution set. Let us think about possible operations which would simplify the matrix A such that obtaining the solution is easier.

We begin with simple manipulations of rows of equations which do not influence the solution, and similar modifications of the right-hand side vector. If we are able to change a square matrix into the unit matrix, then the right-hand side vector is a solution of the original system. If some of the rows of the system vanish during the course of manipulations (that is, they become zero), then we get some direct information about the solution. Our simple operations are:

Elementary row transformations

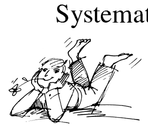
- interchanging two rows,
- multiplication of any given row by a non-zero scalar,
- adding another row to any given row.

These operations are called *elementary row transformations*. It is clear that the corresponding operations at the level of the equations in the system do not change the set of the solutions whenever our ring of coordinates is an integral domain.

Analogically, *elementary column transformations* of matrices are

- interchanging two columns
- multiplication of any given column by a non-zero scalar,
- adding another column to any given column.

These do not preserve the solution set, since they change the variables themselves.



Systematically we can use elementary row transformations for subsequent elimination of variables. This gives an algorithm which is usually called the *Gaussian elimination method*. Henceforth, we shall assume that our scalars come from a integral domain (e.g. integers are allowed, but not say $\mathbb{Z}_4$).

$$E_7 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -4 & 0 & 1 \end{pmatrix}, \quad E_8 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

we obtain

$$B = E_8E_7E_6E_5E_4E_3E_2E_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1/12 & -5/12 & 0 \\ 1 & -2/3 & 1/3 & 0 \\ 0 & -4/3 & -1/3 & 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & -3 & -5 & 0 \\ 0 & 1 & 9/4 & 1/4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

□

2.A.9. Complex numbers as matrices. Consider the set of

matrices $C = \left\{\begin{pmatrix} a & b \\ -b & a \end{pmatrix}, a, b \in \mathbb{R}\right\}$. Note that C is closed under addition and matrix multiplication, and further show that the mapping $f : C \to \mathbb{C}$, $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mapsto a + bi$ satisfies $f(M+N) = f(M)+f(N)$ and $f(M\cdot N) = f(M)\cdot f(N)$ (on the left-hand sides of the equations we have addition and multiplication of matrices, on the right-hand sides we have addition and multiplication of complex numbers). Thus the set C along with multiplication and addition can be seen as the field $\mathbb{C}$ of complex numbers. The mapping f is called an isomorphism (of fields). Thus for instance we have

$$\begin{pmatrix} 3 & 5 \\ -5 & 3 \end{pmatrix} \cdot \begin{pmatrix} 8 & -9 \\ 9 & 8 \end{pmatrix} = \begin{pmatrix} 69 & 13 \\ -13 & 69 \end{pmatrix},$$

which corresponds to $(3 + 5i) \cdot (8 - 9i) = 69 - 13i$.

2.A.10. Solve the equations for matrices

$$\begin{pmatrix} 1 & 3 \\ 3 & 8 \end{pmatrix} \cdot X_1 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad X_2 \cdot \begin{pmatrix} 1 & 3 \\ 3 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Solution. Clearly the unknowns X_1 and X_2 must be matrices of the type 2×2 (in order for the products to be defined and that the result is a matrix of the type 2×2). Set

$$X_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$$

and multiply out the matrices in the first given equation. We obtain

$$\begin{pmatrix} a_1 + 3c_1 & b_1 + 3d_1 \\ 3a_1 + 8c_1 & 3b_1 + 8d_1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

Gaussian elimination of variables

Proposition. *Any non-zero matrix over an arbitrary integral domain of scalars $\mathbb{K}$ can be transformed, using finitely many elementary row transformations, into row echelon form:*

- *For each j, if $a_{ik} = 0$ for all columns $k = 1, \dots, j$, then $a_{kj} = 0$ for all $k \geq i$,*
- *if $a_{(i-1)j}$ is the first non-zero element at the $(i-1)$-st row, then $a_{ij} = 0$.*

Proof. The matrix in row echelon form looks like

$$\begin{pmatrix} 0 & \dots & 0 & a_{1j} & \dots & \dots & \dots & a_{1m} \\ 0 & \dots & 0 & 0 & \dots & a_{2k} & \dots & a_{2m} \\ \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 & a_{lp} & \dots \\ \vdots \end{pmatrix}.$$

The matrix can (but does not have to) end with some zero rows. In order to transform an arbitrary matrix, we can use a simple algorithm, which will bring us, row by row, to the resulting echelon form:

Gaussian elimination algorithm

(1) By a possible interchange of rows we can obtain a matrix where the first row has a non-zero element in the first non-zero column. Let that column be column j. In other words, $a_{1j} \neq 0$, but $a_{iq} = 0$ for all i, and all q, $1 \leq q < j$.
(2) For each $i = 2, \dots$, multiply the first row by the element a_{ij}, multiply i-th row by the element a_{1j} and subtract, to obtain $a_{ij} = 0$ on the i-th row.
(3) By repeated application of the steps (1) and (2), always for the not-yet-echelon part of rows and columns in the matrix we reach, after a finite number of steps, the final form of the matrix.

This algorithm clearly stops after a finite number of steps and provides the proof of the proposition. □

The given algorithm is really the usual elimination of variables used in the systems of linear equations.

In a completely analogous manner we define the column echelon form of matrices and considering column elementary transformations instead the row ones, we obtain an algorithm for transforming matrices into the column echelon form.

Remark. Although we could formulate the Gaussian elim-

ination for general scalars from any ring, this does not make much sense in view of solving equations. Clearly having divisors of zero among the scalars, we might get zeros during the procedure and lose information this way. Think carefully about the differences between the choices $\mathbb{K} = \mathbb{Z}$, $\mathbb{K} = \mathbb{R}$ and possibly $\mathbb{Z}_2$ or $\mathbb{Z}_4$.

that is,

$$\begin{array}{ccccccccc} a_1 & & & + & 3c_1 & & & = & 1, \\ & & b_1 & & & + & 3d_1 & = & 2, \\ 3a_1 & & & + & 8c_1 & & & = & 3, \\ & & 3b_1 & & & + & 8d_1 & = & 4. \end{array}$$

By adding a (-3)-multiple of the first equation with the third equation we obtain $c_1 = 0$ and then $a_1 = 1$. Analogously, by adding a (-3)-multiple of the second equation to the fourth equation we obtain $d_1 = 2$ and then $b_1 = -4$. Thus we have

$$X_1 = \begin{pmatrix} 1 & -4 \\ 0 & 2 \end{pmatrix}.$$

We can find the values a_2, b_2, c_2, d_2 by a different approach. If A is a square matrix, we write A^{-1} to denote its inverse, so that $A \cdot A^{-1} = A^{-1} \cdot A = E$, the unit matrix) It is easy to check that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

which holds for any numbers $a, b, c, d \in \mathbb{R}$ provided $ad - bc \neq 0$. (This is easy to derive; it also directly follows from formula 1 in 2.2.10). We calculate

$$\begin{pmatrix} 1 & 3 \\ 3 & 8 \end{pmatrix}^{-1} = \begin{pmatrix} -8 & 3 \\ 3 & -1 \end{pmatrix}.$$

Multiplying the given equations by this matrix from the right gives

$$X_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} -8 & 3 \\ 3 & -1 \end{pmatrix},$$

and thus

$$X_2 = \begin{pmatrix} -2 & 1 \\ -12 & 5 \end{pmatrix}.$$

□

2.A.11. Solve the matrix equation

$$X \cdot \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & -6 \\ 2 & 1 \end{pmatrix}.$$

○

2.A.12. Computing the inverse matrix. Compute the inverse of the matrices

$$A = \begin{pmatrix} 4 & 3 & 2 \\ 5 & 6 & 3 \\ 3 & 5 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 3 & 3 & 4 \\ 2 & 2 & 3 \end{pmatrix}.$$

Then determine the matrix $\left(A^T \cdot B\right)^{-1}$.

Solution. We find the inverse by the following method: write next to each other the matrix A and the unit matrix. Then use elementary row transformations so that the sub-matrix A changes into the unit matrix. This will change the original unit sub-matrix to A^{-1}. We obtain

On the other hand, if we are dealing with fields of scalars, we can always arrive at a row echelon form where the non-zero entries on the "diagonal" are ones. This is done by applying the the appropriate scalar multiplication to each individual row. However, this is not possible in general – think for instance of the integers $\mathbb{Z}$.

2.1.8. Matrix of elementary row transformations. Let us now restrict ourselves to fields of scalars $\mathbb{K}$, that is, every non-zero scalar has an inverse.

Note that elementary row or column transformations correspond respectively to multiplication from the left or right by the following matrices:

(1) Interchanging the i-th and j-th row (column)

$$\begin{pmatrix} 1 & 0 & \dots & & & & \\ 0 & \ddots & & & & & \\ \vdots & & 0 & \dots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

.

(2) Multiplication of the i-th row (column) by the scalar a:

$$\begin{array}{cc} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \leftarrow i \end{array}$$

.

(3) To row i, add row j (column):

$$\begin{array}{cc} & \begin{pmatrix} 1 & 0 & & & & & \\ 0 & \ddots & & & & & \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & & 1 & & \ddots & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}. \\ i \rightarrow & \\ & \uparrow \\ & j \end{array}$$

This trivial observation is actually very important, since the product of invertible matrices is invertible (recall 1) and all elementary transformations over a field of scalars are invertible (the definition of the elementary transformation itself ensures that inverse transformations are of the same type and it is easy to determine the corresponding matrix).

Thus, the Gaussian elimination algorithm tells us, that for an arbitrary matrix A, we can obtain its equivalent row

$$\left(\begin{array}{ccc|ccc} 4 & 3 & 2 & 1 & 0 & 0 \\ 5 & 6 & 3 & 0 & 1 & 0 \\ 3 & 5 & 2 & 0 & 0 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & -1 \\ 5 & 6 & 3 & 0 & 1 & 0 \\ 3 & 5 & 2 & 0 & 0 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & -1 \\ 0 & 16 & 3 & -5 & 1 & 5 \\ 0 & 11 & 2 & -3 & 0 & 4 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & -1 \\ 0 & 5 & 1 & -2 & 1 & 1 \\ 0 & 11 & 2 & -3 & 0 & 4 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & -1 \\ 0 & 5 & 1 & -2 & 1 & 1 \\ 0 & 1 & 0 & 1 & -2 & 2 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -4 & 3 \\ 0 & 0 & 1 & -7 & 11 & -9 \\ 0 & 1 & 0 & 1 & -2 & 2 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -4 & 3 \\ 0 & 1 & 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -7 & 11 & -9 \end{array}\right).$$

In the first step we subtracted from the first row the third row, in the second step we added a (-5)-multiple of the first to the second row and added a (-3)-multiple of the first row to the third row, in the third step we subtracted from the second row the third row, in the fourth step we added a (-2)-multiple of the second row to the third row, in the fifth step we added a (-5)-multiple of the third row to the second row and added a 2-multiple of the third row to the first row, and in the last step we changed the second and the third row. We have obtained the result

$$A^{-1} = \begin{pmatrix} 3 & -4 & 3 \\ 1 & -2 & 2 \\ -7 & 11 & -9 \end{pmatrix}.$$

Note that when calculating the matrix A^{-1} we did not have to cope with fractions thanks to the suitably chosen row transformations. Although we could carry on similarly when doing the next exercise, that is, B^{-1}, we will rather do the more obvious row transformations. We have

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

echelon form $A' = P \cdot A$ by multiplying with a suitable invertible matrix $P = P_k \cdots P_1$ from the left (that is, sequential multiplication with k matrices of the elementary row transformations).

If we apply the same elimination procedure for the columns, we can transform any matrix B into its column echelon form B' by multiplying it from the right by a suitable invertible matrix $Q = Q_1 \cdots Q_\ell$. If we start with the matrix $B = A'$ in row echelon form, this procedure eliminates only the still non-zero elements out of the diagonal of the matrix and in the end we can transform the remaining elements to be units. Thus we have verified a very important result which we will use many times in the future:

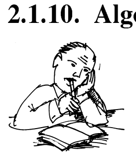
2.1.9. Theorem. *For every matrix A of the type m/n over a field of scalars $\mathbb{K}$, there exist square invertible matrices P and Q of dimensions m and n, respectively, such that the matrix $P \cdot A$ is in row echelon form and*

$$P \cdot A \cdot Q = \begin{pmatrix} 1 & \dots & 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & & & & & \\ 0 & \dots & 1 & 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & & \end{pmatrix}.$$

The number of the ones in the diagonal is independent of the particular choice of P and Q.

Proof. We already have proved everything but the last sentence. We shall see this last claim below in 2.1.11. □

2.1.10. Algorithm for computing inverse matrices. In the previous paragraphs we almost obtained the complete algorithm for computing the inverse matrix. Using the simple modification below, we find either that the inverse does not exist, or we compute the inverse. Keep in mind that we are still working over a field of scalars.



Equivalent row transformations of a square matrix A of dimension n leads to an invertible matrix P' such that $P' \cdot A$ is in row echelon form. If A has an inverse, then there exists also the inverse of $P' \cdot A$. But if the last row of $P' \cdot A$ is zero, then the last row of $P' \cdot A \cdot B$ is also zero for any matrix B of dimension n. Thus, the existence of a zero row in the result of (row) Gaussian elimination excludes the existence of A^{-1}.

Assume now that A^{-1} exists. As we have just seen, the row echelon form of A will have exclusively non-zero rows only, In particular, all diagonal elements of $P' \cdot A$ are non-zero. But now, we can employ row elimination by the elementary row transformation from the bottom-right corner backwards and also transform the diagonal elements to be units. In this way, we obtain the unit matrix E. Summarizing, we find another invertible matrix P'' such that for $P = P'' \cdot P'$ we have $P \cdot A = E$.

Now observe that we could clearly work with columns instead of row transformation and thus, under the assumption

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

that is,

$$B^{-1} = \begin{pmatrix} 1 & 2 & -3 \\ -1 & 1 & -1 \\ 0 & -2 & 3 \end{pmatrix}.$$

Using the identity

$$\left(A^T \cdot B\right)^{-1} = B^{-1} \cdot \left(A^T\right)^{-1} = B^{-1} \cdot \left(A^{-1}\right)^T$$

and the knowledge of the inverse matrices computed before, we obtain

$$\left(A^T \cdot B\right)^{-1} = \begin{pmatrix} 1 & 2 & -3 \\ -1 & 1 & -1 \\ 0 & -2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 & -7 \\ -4 & -2 & 11 \\ 3 & 2 & -9 \end{pmatrix}$$
$$= \begin{pmatrix} -14 & -9 & 42 \\ -10 & -5 & 27 \\ 17 & 10 & -49 \end{pmatrix}.$$

□

2.A.13. Compute the inverse of the matrix

$$A = \begin{pmatrix} 1 & 0 & -2 \\ 2 & -2 & 1 \\ 5 & -5 & 2 \end{pmatrix}.$$

○

2.A.14. Calculate A^5 and A^{-3}, if

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

○

2.A.15. Compute the inverse of the matrix

$$\begin{pmatrix} 8 & 3 & 0 & 0 & 0 \\ 5 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 & 5 \end{pmatrix}.$$

○

2.A.16. Determine whether there exists an inverse of the matrix

$$C = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

If yes, then compute C^{-1}.

of the existence of A^{-1}, we would find a matrix Q such that $A \cdot Q = E$. From this we see immediately that

$$P = P \cdot E = P \cdot (A \cdot Q) = (P \cdot A) \cdot Q = E \cdot Q = Q.$$

That is, we have found the inverse matrix

$$A^{-1} = P = Q$$

for the matrix A. Notice that at the point of finding the matrix P with the property $P \cdot A = E$, we do not have to do any further computation, since we have already obtained the inverse matrix.

In practice, we can work as follows:

Computing the inverse matrix

Write the unit matrix E to the right of the matrix A, producing an augmented matrix (A, E). Transform the augmented matrix using the elementary row transformations to row echelon form. This produces an augmented matrix (PA, PE), where P is invertible, and PA is in row echelon form. By the above, either $PA = E$, in which case A is invertible and $P = PE = A^{-1}$, or PA has a row of zeros, in which case we conclude that the inverse matrix for A does not exist.

2.1.11. Linear dependence and rank. In the previous

practical algorithms dealing with matrices we worked all the time with row and column additions and scalar multiplications, seeing them as vectors.

Such operations are called *linear combinations*. We shall return to such operations in an abstract sense later on in 2.3.1. But it will be useful to understand their core meaning right now. A linear combination of rows of a matrix $A = (a_{ij})$ of type m/n is understood as an expression of the form

$$c_1 u_{i_1} + \cdots + c_k u_{i_k},$$

where c_i are scalars, $u_j = (a_{j1}, \dots, a_{jn})$ are rows of the matrix A. Similarly, we can consider linear combinations of columns by replacing the above rows u_j by the columns $u_j = (a_{1j}, \dots, a_{mj})$.

If the zero row can be written as a linear combination of some given rows with at least one non-zero scalar coefficient, we say that these rows are *linearly dependent*. In the alternative case, that is, when the only possibility of obtaining the zero row is to select all the scalars c_j equal to zero, the rows are called *linearly independent*.

Analogously, we define linearly dependent and linearly independent columns.

The previous results about the Gaussian elimination can

be now interpreted as follows: the number of non-zero "steps" in the row (column) echelon form is always equal to the number of linearly independent rows (columns) of the matrix. Let E_h be the matrix from the theorem 2.1.9 with h ones on the diagonals and assume that by two different row transformation procedures into

2.A.17. Compute A^{-1}, if

(a) $A = \begin{pmatrix} 1 & i \\ -i & 3 \end{pmatrix}$, while i is the imaginary unit

(b) $A = \begin{pmatrix} 1 & -5 & -3 \\ -1 & 5 & 4 \\ -1 & 6 & 2 \end{pmatrix}$.

2.A.18. Find the inverse to the $n \times n$ matrix ($n > 1$)

$$A = \begin{pmatrix} 2-n & 1 & \cdots & 1 & 1 \\ 1 & 2-n & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \ddots & \ddots & 2-n & 1 \\ 1 & 1 & \cdots & 1 & 2-n \end{pmatrix}.$$

Solution. You can try for small n ($n = 2, 3, 4$), which is easy to compute with the known algorithm, and then guess the general form.

$$A^{-1} = \frac{1}{n-1} \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}.$$

□

We have already encountered systems of linear equations at the beginning of the chapter. Now we will deal with them in more detail. We use the inverse matrix to assist in computing the solution to the system of linear equations. Note that we do the same computation as before. To express the variables is the same as to bring the matrix of the system with equivalent transformation to the identity matrix and that is the same as to multiply the matrix of the system with the inverse matrix.

2.A.19. Participants of a trip. There were 45 participants of a two-day bus trip. On the first day, the fee for a watchtower visit was €30 for an adult, €16 for a child and €24 for a senior. The total fee for the first day was €1 116. On the second day, the fee for a bus with a palace and botanical garden tour was €40 for an adult, €24 for a child and €34 for a senior. The total fee for the second day was €1 542. How many adults, children and seniors were there among the participants?

Solution. Introduce the variables

x for the „number of adults";

y for the „number of children";

the echelon form we obtain two different $h' < h$. But then according to our algorithm there are invertible matrices P, P', Q, and Q' such that

$$E_h = P \cdot A \cdot Q, \; E_{h'} = P' \cdot A \cdot Q'.$$

In particular, $E_h = P \cdot P'^{-1} \cdot E_{h'} \cdot Q'^{-1} \cdot Q$ and so there are invertible matrices P'' and Q'' such that

$$P'' \cdot E_{h'} \cdot Q'' = E_h.$$

In the product $P'' \cdot E_{h'}$ there will be more zero rows in the bottom part of the echelon matrix than we see in E_h and we must be able to reach E_h using only elementary column transformations. This is clearly not possible, because the zero rows remain zero there.

Therefore the number of ones in the matrix $P \cdot A \cdot Q$ in theorem 2.1.9 is independent of the choice of our elimination procedure and it is always equal to the number of linearly independent rows in A, which must be the same as the number of linearly independent columns in A. This number is called the *rank of the matrix* and we denote it by $h(A)$. We have the following theorem:

Theorem. *Let A be a matrix of type m/n over a field of scalars $\mathbb{K}$. The matrix A has the same number $h(A)$ of linearly independent rows as linearly independent columns. In particular, the rank is always at most the minimum of the dimensions of the matrix A.*

The algorithm for computing the inverse matrix also says that a square matrix A of dimension m has an inverse if and only if its rank equals m.

2.1.12. Matrices as mappings. Similarly to the way we worked with matrices in the geometry of the plane (see **??**), we can interpret every matrix A of the type m/n as a mapping

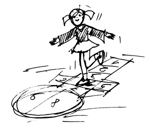
$$A : \mathbb{K}^n \to \mathbb{K}^m, \quad x \mapsto A \cdot x.$$

By the distributivity of matrix multiplication, it is clear how the linear combinations of vectors are mapped using such mappings:

$$A \cdot (a\,x + b\,y) = a\,(A \cdot x) + b\,(A \cdot y).$$

Straight from the definition we see, by the associativity of multiplication, that composition of mappings corresponds to matrix multiplication in given order. Thus invertible matrices of dimension n correspond to bijective mappings $A : \mathbb{K}^n \to \mathbb{K}^n$.

Remark. From this point of view, the theorem 2.1.9 is very interesting. We can see it as follows: the rank of the matrix determines how large is the image of the whole $\mathbb{K}^n$ under this mapping. In fact, if $A = P \cdot E_k \cdot Q$ where the matrix E_k has k ones as in 2.1.9, then the invertible Q first bijectively "shuffles" the n-dimensional vectors in $\mathbb{K}^n$, the matrix E_k then "copies" the first k coordinates and completes them with the remaining $m - k$ zeros.



z for the „number of seniors“;

There were 45 participants, therefore

$$x + y + z = 45.$$

The fees for the first and second days respectively imply that

$$\begin{aligned} 30x + 16y + 24z &= 1\,116,\\ 40x + 24y + 34z &= 1\,542. \end{aligned}$$

We write the system of three linear equations in the matrix notation as

$$\begin{pmatrix} 1 & 1 & 1\\ 30 & 16 & 24\\ 40 & 24 & 34 \end{pmatrix} \cdot \begin{pmatrix} x\\ y\\ z \end{pmatrix} = \begin{pmatrix} 45\\ 1\,116\\ 1\,542 \end{pmatrix}.$$

We compute

$$\begin{pmatrix} 1 & 1 & 1\\ 30 & 16 & 24\\ 40 & 24 & 34 \end{pmatrix}^{-1} = \frac{1}{6}\begin{pmatrix} 16 & 5 & -4\\ 30 & 3 & -3\\ -40 & -8 & 7 \end{pmatrix}.$$

Hence the solution is

$$\begin{pmatrix} x\\ y\\ z \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 16 & 5 & -4\\ 30 & 3 & -3\\ -40 & -8 & 7 \end{pmatrix} \cdot \begin{pmatrix} 45\\ 1\,116\\ 1\,542 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 132\\ 72\\ 66 \end{pmatrix} = \begin{pmatrix} 22\\ 12\\ 11 \end{pmatrix},$$

expressed in words, there were 22 adults, 12 children and 11 seniors. □

The latter approach is particularly efficient if we have to solve several systems with the same matrix on the left hand side but different values on the right hand side.

But what if the matrix of the system is not invertible? Then we cannot use the inverse matrix for solving the system. Such a system cannot have a single solution. As the reader may have noticed above, a system of linear equations either has no solution, has one solution or has infinitely many solutions, depending on one or more free parameters (for instance, it cannot have exactly two solutions). We should have also noticed when dealing with equations with two variables in the previous section, that the space of the solutions is either a vector space (in the case when the right-hand side of the system is zero, we speak of a *homogeneous system* of linear equations) or an affine space, see **??** (in the case when the right-hand side of at least one of the equations is non-zero, we speak of a *non-homogeneous system* of linear equations).

We can recognize all the possibilities from the rank of the matrices, i.e. the number of nonzero rows left in the row-echelon form.

This “k-dimensional” image then cannot be enlarged by multiplying with P. Multiplying by P can only bijectively reshuffle the coordinates.

2.1.13. Solving systems of linear equations. We shall return to the notions of dimension, linear independence and so on in the third part of this chapter.

But we should notice now what our results say about the solutions of the system of linear equations.

If we consider the matrix of the system of equations and add to it the column of the required results, we speak about the *extended matrix of the system*. The above Gaussian elimination approach corresponds to the sequential variable elimination in the equations and the deletion of the linearly dependent equations (these are simply consequences of other equations).

Thus we have derived complete information about the size of the set of solutions of the system of linear equations, based on the rank of the matrix of the system. If we are left with more non-zero rows in the row echelon form of the extended matrix than in the original matrix of the system, then there cannot be a solution (simply, we cannot obtain the given vector value with the corresponding linear mapping). If the rank of both matrices is the same, then the backwards elimination provides exactly as many free parameters as the difference between the number of variables n and the rank $h(A)$. In particular, there will be exactly one solution if and only if the matrix is invertible.

2. Determinants

In the fifth part of the first chapter, we introduced the scalar function $\det$ on square matrices of dimension 2 over the real numbers, called determinant, see **??**. We saw that the determinant assigned a non-zero number to a matrix if and only the matrix was invertible. We did not say it in exactly this way, but you can check for yourself in previous paragraphs starting with **??** and formula (**??**).

We saw also that determinants were useful in another way, see the paragraphs **??** and **??**. There we showed that the volume of the parallelepiped should be linearly dependent on every two of the vectors defining it. It was useful to require the change of the sign when changing the order of these vectors. Because determinants (and only determinants) have these properties, up to a constant scalar multiple, we concluded that it was determining the volume. Now we will see that we can proceed similarly for every finite dimension.

We work again with arbitrary scalars $\mathbb{K}$ and matrices over these scalars. Our results about determinants will thus hold for all commutative rings, notably also for integer matrices or matrices over any residue classes.

2.2.1. Definition of the determinant. Recall that the bijective mapping from a set X to itself is called a *permutation of the set X*, see **??**. If $X = \{1, 2, \dots, n\}$, the permutation can

2.A.20. Determine the rank of the matrix

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

Then determine the number of solutions of the system of linear equations

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & + & x_3 & - & 2x_4 & = & 4, \\ -3x_1 & - & 2x_2 & - & x_3 & - & x_4 & = & 5, \\ & + & 2x_2 & & & + & x_4 & = & 1, \\ x_1 & - & 4x_2 & + & x_3 & - & 2x_4 & = & 3 \end{array}$$

Determine also all solutions of the system

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & + & x_3 & - & 2x_4 & = & 0, \\ -3x_1 & - & 2x_2 & - & x_3 & - & x_4 & = & 0, \\ & + & 2x_2 & & & + & x_4 & = & 0, \\ x_1 & - & 4x_2 & + & x_3 & - & 2x_4 & = & 0 \end{array}$$

and of the system

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

Solution. Transforming the matrix to the row-echelon form, we check that the rank is four. (The rank cannot exceed the number of rows or columns). The first of the three given system is given by the extended matrix

$$\left(\begin{array}{rrrr|r} 1 & 1 & 1 & -2 & 4 \\ -3 & -2 & -1 & -1 & 5 \\ 0 & 2 & 0 & 1 & 1 \\ 1 & -4 & 1 & -2 & 3 \end{array}\right).$$

But the left-hand side is exactly A^T and thus we can get the column-echelon form the same way as before. In particular, the columns of the matrix are linearly indepent and the rank is maximal, i.e. four again. Therefore there exists a matrix $\left(A^T\right)^{-1}$ and the system has a unique solution

$$(x_1,\, x_2,\, x_3,\, x_4)^T = \left(A^T\right)^{-1} \cdot (4,5,1,3)^T .$$

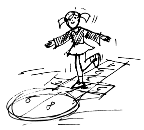
The second of the systems has the same left-hand side (given by the matrix A^T) as the first. Because the numbers on the right-hand side of the equations in the system do not influence the number of solutions and because every homogeneous system has a zero solution, the only solution of the second system is given by

$$(x_1,\, x_2,\, x_3,\, x_4) = (0,0,0,0) .$$

be written by putting the resulting ordering into a table:

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}.$$

The element $x \in X$ is called a fixed point of the permutation σ if $\sigma(x) = x$. If there exist exactly two distinct elements $x, y \in X$ such that $\sigma(x) = y$ while all other elements $z \in X$ are fixed points, then the permutation σ is called a *transposition*, and we denote it by (x, y). Of course, then $\sigma(y) = x$ holds for such a transformation.



For dimension 2, the formula for a determinant was simple – take all possible products of two elements, one from every column and every row of the matrix, give them a sign such that interchanging two columns leads to the change of the sign of the whole result, and sum all of them (that is, both):

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \ \det A = ad - bc.$$

Consider now square matrices $A = (a_{ij})$ of dimension n over $\mathbb{K}$. The formula for the determinant of the matrix A is also composed of all possible products from elements from individual rows and columns:

Definition of determinant

The determinant of the matrix A is a scalar $\det A = |A|$ defined by the relation

$$|A| = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

where Σ_n is the set of all possible permutations over $\{1, \dots, n\}$ and the symbol sgn for a permutation σ, called the parity of σ, will be described later. Each of the expressions

$$\operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

is called a *term in the determinant* $|A|$.

In dimensions 2 or 3 we can easily guess correct signs. The product of the elements on the diagonal should be with positive sign and we want anti-symmetry when interchanging two columns or rows.

Determinants in dimension 2 and 3

For $n = 2$ it is, as we have expected

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

Similarly for $n = 3$

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

This formula is often called the *Sarrus rule*.

The third system is given by the extended matrix

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

which is the matrix A (only the last column is given after the vertical bar). If we try to simplify the matrix into the row echelon form, we must obtain a row

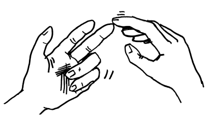
$$\left(\begin{array}{ccc|c} 0 & 0 & 0 & a \end{array}\right), \quad \text{where} \quad a \neq 0.$$

We know, that the column on the right-hand side is not a linear combination of the columns on the left-hand side (the rank of the matrix is 4). This system thus has no solution. □

For further examples see 2.H.7

B. Permutations and determinants

In order to be able to define the key object of the matrix



calculus, the determinant, we must deal with permutations (bijections of a finite set) and their parities.

We shall use the two-row notation for permutations (see 2.2.1). In the first row we list all elements of the given set, and every column then corresponds to a pair (preimage, image) in the given permutation. Because a permutation is a bijection, the second row is indeed a permutation (ordering) of the first row, in accordance with the definition from combinatorics.

2.B.1. Decompose the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 6 & 7 & 8 & 9 & 5 & 4 & 2 \end{pmatrix}$$

into a product of transpositions.

Solution. We first decompose the permutation into a product of independent cycles. Start with the first element 1 and look on the second row to see what the image of 1 is. It is 3. Now look on the column that starts with 3, and see that the image of 3 is 6, and so on. Continue until we again reach the starting element 1. We obtain the following sequence of elements, which map to each other under the given permutation:

$$1 \mapsto 3 \mapsto 6 \mapsto 9 \mapsto 2 \mapsto 1.$$

The mapping which maps elements in such a manner is called a cycle (see 2.2.3) which we denote by $(1, 3, 6, 9, 2)$.

Now choose any element not contained in the obtained cycle. With the same procedure as with 1, we obtain the cycle $(4, 7, 5, 8)$. From the method is clear that the result does not depend on the first obtained cycle. Each element from the set

2.2.2. Parity of permutation. How should we define the

sign of a permutation? We say that a pair of elements $a, b \in X = \{1, \dots, n\}$ forms an *inversion in the permutation* σ, if $a < b$ and $\sigma(a) > \sigma(b)$. A permutation σ is called *even* or *odd*, if it contains an even or odd number of inversions, respectively.

Thus, the *parity of the permutation* σ is $(-1)^{\text{number of inversions}}$ and we denote it by $\operatorname{sgn}(\sigma)$. This amounts to our definition of sign for computing determinant. But we should like to know how to calculate the parity. The following theorem reveals that the Sarrus rule really defines the determinant in dimension 3.

Theorem. *Over the set $X = \{1, 2, \dots, n\}$ there are exactly $n!$ distinct permutations. These can be ordered in a sequence such that every two consecutive permutations differ in exactly one transposition. Every transposition changes parity.*

For any chosen permutation σ there is such a sequence starting with σ.

Proof. For $n = 1$ or $n = 2$, the claim is trivial. We prove the theorem by induction on the size n of the set X.

Assume that the claim holds for all sets with $n - 1$ elements and consider a permutation $\sigma(1) = a_1, \dots, \sigma(n) = a_n$. According to the induction assumption, all the permutations that end with a_n can be obtained in a sequence, where every two consecutive permutations differ in one transposition. There are $(n - 1)!$ such permutations. In order to proceed further, we select the last of them, and use the transposition of $\sigma(n) = a_n$ with some element a_i which has not been at the last position yet. Once again, we form a sequence of all permutations that end with a_i. After doing this procedure n-times, we obtain $n(n - 1)! = n!$ distinct permutations – that is, all permutations on n elements. The resulting sequence satisfies the condition.

Note that the last sentence of the theorem does not seem to be useful in practice. But it is a very important part for proving the theorem by induction over the size of X.

It remains to prove the part of the theorem about parities. Consider the ordering

$$(a_1, \dots, a_i, a_{i+1}, \dots, a_n),$$

containing r inversions. Then in the ordering

$$(a_1, \dots, a_{i+1}, a_i, \dots, a_n)$$

there are either $r - 1$ or $r + 1$ inversions. Every transposition (a_i, a_j) is obtainable by doing $(j - i) + (j - i - 1) = 2(j - i) - 1$ transpositions of neighbouring elements. Therefore any transposition changes the parity. Also, we already know that all permutations can be obtained by applying transpositions. □

We found that applying a transposition changes the parity of a permutation and any ordering of numbers $\{1, 2, \dots, n\}$ can be obtained through transposing of neighbouring elements. Therefore we have proven

($\{1, 2, \dots, 9\}$) appears in one of the obtained cycles, we can thus write:

$$\sigma = (1, 3, 6, 9, 2) \circ (4, 7, 5, 8),$$

or

$$\sigma = (4, 7, 5, 8) \circ (1, 3, 6, 9, 2),$$

since independent cycles commute. For cycles the decomposition into transpositions is simple, we have

$$(1, 3, 6, 9, 2) = (1, 3) \circ (3, 6) \circ (6, 9) \circ (9, 2) =$$
$$(1, 3)(3, 6)(6, 9)(9, 2).$$

Thus we obtain:

$$\sigma = (1, 3)(3, 6)(6, 9)(9, 2)(4, 7)(7, 5)(5, 8).$$

□

Remark. The minimal number of transpositions in the decomposition of a permutation is obtained by carrying out exactly the procedure as above. That is, first decompose the permutation into the independent cycles, then the cycles canonically into the transpositions. Thus the found decomposition is the decomposition into the minimal number of transpositions.

Note also that the operation $\circ$ is a composition of mappings, thus it is necessary to carry out the composition "backwards", as we are used to in composition of mappings. Applying the given composition of transposition for instance on the element two we can successively write:

$$[(1, 3)(3, 6)(6, 9)(9, 2)](2) =$$
$$[(1, 3)(3, 6)(6, 9)]((9, 2)(2)) =$$
$$[(1, 3)(3, 6)(6, 9)](9) = [(1, 3)(3, 6)](6) = (1, 3)(3) = 1,$$

thus the mapping indeed maps the element 2 on the element 1 (it is actually just the cycle $(1, 3, 6, 9, 2)$ written in a different way). When writing a composition of permutations, we often omit the sign "$\circ$" and speak of the product of permutations.

When writing the cycle we write only the elements on which the cycle (that is, the mapping) nontrivially acts (that is, the element is mapped to some other element). Fixed-points of the cycle are not listed. Thus it is necessary to know on which set do we consider the given cycle (mostly it will be clear from the context). The cycle $c = (4, 7, 5, 8)$ from the previous example is thus a mapping (permutation), which, in the two-row notation, looks like this

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 7 & 8 & 6 & 5 & 4 & 9 \end{pmatrix}.$$

If the original permutation has some fixed-points they do not appear in the cycle decomposition.

Corollary. *On every finite set $X = \{1, \dots, n\}$ with n elements, $n > 1$, there are exactly $\frac{1}{2}n!$ even permutations, and $\frac{1}{2}n!$ odd permutations.*

If we compose two permutations, it means first doing all transpositions forming the first permutation and then all the transpositions forming the second one. Therefore for any two permutations $\sigma, \eta : X \to X$ we have

$$\operatorname{sgn}(\sigma \circ \eta) = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\eta)$$

and also

$$\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma).$$

2.2.3. Decomposing permutations into cycles. A good tool for practical work with permutations is the cycle decomposition, which is also a good exercise on the concept of equivalence.

CYCLES

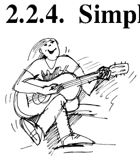
A permutation σ over the set $X = \{1, \dots, n\}$ is called a *cycle* of length k, if we can find elements $a_1, \dots, a_k \in X$, $2 \le k \le n$ such that $\sigma(a_i) = a_{i+1}$, $i = 1, \dots, k-1$, while $\sigma(a_k) = a_1$, and other elements in X are fixed-points of σ. Cycles of length two are transpositions.

Every permutation is a composition of cycles. Cycles of even length have parity -1, cycles of odd length have parity 1.

PROOF. The last claim has yet to be proved. Fix a permutation σ and define a relation R such that two elements $x, y \in X$ are R-related if and only if $\sigma^\ell(x) = y$ for some iteration $\ell \in \mathbb{Z}$ of the permutation σ (notice σ^{-1} means the inverse bijection to σ). Clearly, it is an equivalence relation (check it carefully!). Because X is a finite set, for some ℓ it must be that $\sigma^\ell(x) = x$. If we pick one equivalence class $\{x, \sigma(x), \dots, \sigma^{\ell-1}(x)\} \subset X$ and define other elements to be fixed-points, we obtain a cycle. Evidently, the original permutation X is then the composition of all these cycles for individual equivalence classes and it does not matter in which order we compose the cycles.

For determining the parity we just have to note that cycles of even length can be written as a composition of an odd number of transposition, therefore their parity is -1. Analogously, cycle of odd length can be obtained using an even number of transpositions and therefore it has parity 1. □

2.2.4. Simple properties of determinant. Knowing the properties of permutations and their parities from previous paragraphs allows us to derive quickly some basic properties of determinants.



For every matrix $A = (a_{ij})$ of the type m/n over scalars from $\mathbb{K}$ we define the *transpose* of A as the matrix $A^T = (a'_{ij})$ with elements $a'_{ij} = a_{ji}$. A^T is of the type n/m.

A square matrix A with the property $A = A^T$ is called *symmetric*. If $A = -A^T$, then A is called *antisymmetric*.

Note further that the notation $(1, 2, 3)$ gives the same cycle as for instance $(2, 3, 1)$ or $(3, 1, 2)$. But the notation $(1, 3, 2)$ is a different cycle.

2.B.2. Determine the parity of the following permutations:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 6 & 7 & 8 & 9 & 5 & 4 & 2 \end{pmatrix},$$

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

Solution. According to our definition (see 2.2.2) we compute the number of inversions of σ: we go sequentially through the second row in the two-row notation and for every number k there we count the number of numbers which are smaller than k and are located after k in the second row. It is not hard to see that the number of inversions in a given permutation is exactly the number of pairs "larger before smaller" in the second row. For σ we compute (stepping through the second row): after three there is one and two, thus we add 2; after one there is no smaller number and we add 0; after six there is five, four and two, thus we add 4, similarly for seven, eight and nine, for five we add 2, for four we add 1 and for two nothing. Thus we have 17 inversions in total and thus the permutation is odd.

But we can compute the parity of σ otherwise. The theorem 2.2.2 implies that the parity of a permutation is given by the parity of the number of transpositions in its decomposition (this number is, unlike the number of transposition in an arbitrary decomposition, always the same)

The previous exercise gives us
$\sigma = (1, 3)(3, 6)(6, 9)(9, 2)(4, 7)(7, 5)(5, 8)$. There are seven transpositions in the decomposition, thus the permutation is indeed odd.

Alternatively we can decompose τ into either a product of three transpositions (using the cycle decomposition):

$$\tau = (1, 2, 4)(3, 6) = (1, 2)(2, 4)(3, 6),$$

or we count the number of inversions in τ: $1+2+3+0+1 = 7$. Either way we find that τ is an odd permutation.

In general, as soon as the decomposition to cycles is ready, we may just count the lengths of the cycles, since each cycle including k elements is clearly built of $k-1$ transpositions and thus contributes $(-1)^{k-1}$ to the parity. □

For the following exercises, recall how to compute determinants of the type 2×2 ($a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$) and 3×3 (Sarrus rule), see 2.2.1.

Simple properties of determinants

Theorem. *Every square matrix $A = (a_{ij})$ satisfies the following conditions:*

(1) $|A^T| = |A|$.
(2) If one of the rows contains only zero elements from $\mathbb{K}$, then $|A| = 0$.
(3) If a matrix B was obtained from A by transposing two rows, then $|A| = -|B|$.
(4) If a matrix B was obtained from A by multiplying one row by a scalar $a \in \mathbb{K}$, then $|B| = a\,|A|$.
(5) If all elements of the k-th row in A are of the form $a_{kj} = c_{kj} + b_{kj}$ and all remaining rows in the matrices A, $B = (b_{ij})$, $C = (c_{ij})$ are identical, then $|A| = |B| + |C|$.
(6) A determinant $|A|$ does not change if we add to any row of A a linear combination of other rows.

Proof. (1) The terms of determinants $|A|$ and $|A^T|$ are in bijective correspondence, where the term $\operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ corresponds the following A^T term (notice it does not depend on the order of scalars)

$$\operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdot a_{\sigma(2)2} \cdots a_{\sigma(n)n} = \\ = \operatorname{sgn}(\sigma) a_{1\sigma^{-1}(1)} \cdot a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)},$$

and we have to ensure that this member has the correct sign. But the parities of σ and σ^{-1} are the same, and so this is really a term in the determinant $|A^T|$ and the first claim is proved.

(2) This comes straight from the definition of determinant, because all its terms contain exactly one member from every row. Thus, if one of the rows is zero, all terms of the determinant are also zero.

(3) The only change in the terms of $|B|$ compared to $|A|$ is the addition of one transposition in all permutations, therefore all the signs will be reversed.

(4) This follows straight from the definition, because terms of $|B|$ are just terms of $|A|$ multiplied by the scalar a.

(5) In every term of $|A|$, there is exactly one element from the k-th row of the matrix A. By the distributive law for multiplication and addition in $\mathbb{K}$, the claim follows directly from the definition of determinant.

(6) If there are two identical rows in A, then there are always two identical terms among all terms in the determinant, up to the sign. Therefore in this case $|A| = 0$. Thus, by (5), we can add any other row to the given row, without changing the value of the determinant. In view of the claims (4) and (5), we can in fact add a scalar multiple of any other row. □

2.B.3. Compute the determinant of the following matrices

$$\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & 2 & 3\\ 1 & -1 & 2\\ 3 & 2 & 2\end{pmatrix}, \quad \begin{pmatrix}1 & 1 & 1\\ 1 & 0 & 0\\ -2 & 0 & 1\end{pmatrix}.$$

○

Solution. The determinant of the first matrix is $1 \cdot 1 - 2 \cdot 2 = -3$.

As for the second matrix, according to the Sarrus rule we just have to enumerate the expression
$1{\cdot}(-1){\cdot}2+2{\cdot}2{\cdot}3+3{\cdot}1{\cdot}2-3{\cdot}(-1){\cdot}3-1{\cdot}2{\cdot}2-1{\cdot}2{\cdot}2 = 17$.

We can also bring the matrix into the row echelon form and then multiply the numbers on the diagonal but we have to remember that a multiplication of a row with a scalar changes the determinant of the matrix by the same multiple. Interchanging two rows changes the sign of the determinant of the matrix.

$$\begin{vmatrix}1 & 2 & 3\\ 1 & -1 & 2\\ 3 & 2 & 2\end{vmatrix} = \begin{vmatrix}1 & 2 & 3\\ 0 & -3 & -1\\ 0 & -4 & -7\end{vmatrix} = \frac{1}{-4} \cdot \frac{1}{3} \cdot \begin{vmatrix}1 & 2 & 3\\ 0 & 12 & 4\\ 0 & -12 & -21\end{vmatrix}$$
$$= -\frac{1}{12} \cdot \begin{vmatrix}1 & 2 & 3\\ 0 & 12 & 4\\ 0 & 0 & -17\end{vmatrix}$$

We finish with an upper triangular matrix. The determinant of such matrices is the product of the numbers on the main diagonal. So the result is $-\frac{1}{12}(1 \cdot 12 \cdot (-17)) = 17$.

We can see, that using the Sarrus rule is quicker.

For the third matrix we have
$1{\cdot}0{\cdot}1+1{\cdot}0{\cdot}1+1{\cdot}0{\cdot}(-2)-1{\cdot}0{\cdot}(-2)-1{\cdot}1{\cdot}1-1{\cdot}0{\cdot}0 = -1$.

□

It is important to realize, that Sarrus rule can be used for matrices 3×3 only. For higher dimension matrices you can either bring the matrix to the row echelon form (where you have to take in to account rules 2.2.4) or use the Laplace expansion (see 2.2.8).

2.B.4. Compute the determinant of the matrix

$$\begin{pmatrix}1 & 3 & 5 & 6\\ 1 & 2 & 2 & 2\\ 1 & 1 & 1 & 2\\ 0 & 1 & 2 & 1\end{pmatrix}.$$

Solution. We compute this in two ways. First, convert the matrix to row echelon form. We can use already known elementary transformations,

2.2.5. Computational corollaries. By the previous theorem, we can use elementary row transformations to bring any square matrix A into row echelon form, without changing the value of its determinant. We just have to be careful and add only linear combinations of other rows to a given one.

Let us look at the distribution of the elements in the products in individual terms of a determinant $|A|$ with dimension of A equal to $n > 1$. There is just one term with all of its elements on the diagonal. In all other terms, there must be elements both above and below the diagonal (if we place one element outside of the diagonal, we block two diagonal entries and we leave only $n - 2$ diagonal positions for the other $n - 1$ elements).

Thus, if the matrix A is in a row echelon form, then every term of $|A|$ is zero, except the term with exclusively diagonal entries. Thus we have proved the following algorithm:

Computing determinants using elimination

If A is in the row echelon form then

$$|A| = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}.$$

The previous theorem gives an effective method for computing determinants using the Gauss elimination method, see the paragraph 2.1.7.

Let us note a nice corollary of the first claim of the previous theorem about the equality of the determinants of the matrix and its transpose. It ensures that whenever we prove some claim about determinants formulated in terms of rows of the corresponding matrix, we immediately obtain an analogous claim in terms of the columns. For instance, we can immediately formulate all the claims (2)–(6) for linear combinations of columns.

As a useful (theoretical) illustration of this principle, we shall derive the following formula for direct calculation of solutions of systems of linear equations:

Cramer rule

Consider the system of n linear equations for n variables with matrix of the system $A = (a_{ij})$ and the column of values $b = (b_1, \dots, b_n)$. In matrix notation this means we are solving the equation $A \cdot x = b$.

If there exists the inverse A^{-1}, then the individual components of the unique solution $x = (x_1, \dots, x_n)$ are given as

$$x_i = |A_i||A|^{-1},$$

where the matrices A_i arise from the matrix A of the system by replacing the i-th column by the column b of values.

Proof. As we have already seen, the inverse of the matrix of the system exists if and only if the system has a unique solution. If we have such a solution x, we can express the

$$\begin{vmatrix} 1 & 3 & 5 & 6 \\ 1 & 2 & 2 & 2 \\ 1 & 1 & 1 & 2 \\ 0 & 1 & 2 & 1 \end{vmatrix} = - \begin{vmatrix} 1 & 1 & 1 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 3 & 5 & 6 \\ 0 & 1 & 2 & 1 \end{vmatrix} = - \begin{vmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 2 & 4 & 4 \\ 0 & 1 & 2 & 1 \end{vmatrix}$$

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

Note, that we have interchanged the rows twice in the course of computation.

The other way of computing the determinant is by cofactor expansion along the first column (the one with the greatest number (one) of zeroes). Successively we obtain

$$\begin{vmatrix} 1 & 3 & 5 & 6 \\ 1 & 2 & 2 & 2 \\ 1 & 1 & 1 & 2 \\ 0 & 1 & 2 & 1 \end{vmatrix} = 1 \cdot \begin{vmatrix} 2 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{vmatrix} - 1 \cdot \begin{vmatrix} 3 & 5 & 6 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{vmatrix} +$$

$$1 \cdot \begin{vmatrix} 3 & 5 & 6 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \end{vmatrix} \overset{\text{using the Sarrus rule}}{=} -2 - 2 + 6 = 2.$$

□

2.B.5. Compute the determinant of the matrix

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 3 & 0 & 3 \\ 4 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

Solution. We notice, that the last (fifth) row contains four zeros (as well as the second column). It is the most, we can find in a row or a column in the matrix, thus it will be advantageous to use Laplace theorem (2.3.10) and compute the determinant via expasion along the fifth row or second column. We present the expansion via fifth row:

$$\begin{vmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 3 & 0 & 3 \\ 4 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 & 5 \end{vmatrix} = 0 \cdot \begin{vmatrix} 0 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 4 & 4 \end{vmatrix} - 0 \cdot \begin{vmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 3 & 0 & 3 \\ 4 & 0 & 4 & 4 \end{vmatrix}$$

$$+0 \cdot \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 4 & 0 & 4 & 4 \end{vmatrix} - 0 \cdot \begin{vmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 3 \\ 4 & 0 & 0 & 4 \end{vmatrix} + 5 \cdot \begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{vmatrix}$$

$$= 5 \cdot \begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 \end{vmatrix} = 5 \cdot 2 \cdot \begin{vmatrix} 1 & 1 & 0 \\ 0 & 3 & 0 \\ 4 & 0 & 4 \end{vmatrix} = 120,$$

column b in the matrix A_i by the corresponding linear combination of the columns of the matrix A, that is the values $b_i = a_{i1}x_1 + \cdots + a_{in}x_n$. Then, by subtracting the x_k-multiples of all the other columns from this i-th column, we arrive at just the x_i-multiple of the original column of A. The number x_i can thus be brought in front of the determinant to obtain the equation $|A_i| = x_i\,|A|$, and thus $|A_i||A|^{-1} = x_i|A||A|^{-1} = x_i$, which is our claim. □

Notice also that the properties (3)–(5) from the previous theorem say that the determinant, (considered as a mapping which assigns a scalar to n vectors of dimension n), is an antisymmetric mapping linear in every argument, exactly as we required in analogy to the 2-dimensional case.

2.2.6. Further properties of the determinant. Later we will see that, exactly as in the dimension 2, the determinant of the matrix equals to the (oriented) volume of the parallelepiped determined by the columns of the matrix. We shall also see that considering the mapping $x \mapsto A \cdot x$ given by the square matrix A on $\mathbb{R}^n$ we can understand the determinant of this matrix as expressing the ratio between the volume of the parallelepipeds given by the vectors $x_1, \dots x_n$ and their images $A \cdot x_1, \dots, A \cdot x_n$.

Because the composition $x \mapsto A{\cdot}x \mapsto B{\cdot}(A{\cdot}x)$ of mappings corresponds to the matrix multiplication, the *Cauchy theorem* below is easy to understand:

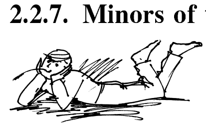
Cauchy theorem

Theorem. *Let $A = (a_{ij})$, $B = (b_{ij})$ be square matrices of dimension n over the ring of scalars $\mathbb{K}$. Then $|A \cdot B| = |A| \cdot |B|$.*

Notice, the claims (2), (3) and (6) from the theorem 2.2.4 are easily deduced from the Cauchy theorem and the representation of the elementary row transformations as multiplication by suitable matrices (cf. 2.1.8).

In the next paragraphs, we derive this theorem in a purely algebraic way, in particular because the previous argumentation based on geometrical intuition could hardly work for arbitrary scalars. The basic tool is the *determinant expansion* using one or more of the rows or columns. We will also need a little technical preparation. The reader who is not fond of too much abstraction can skip these paragraphs and note only the statement of the Laplace theorem and its corollaries.

2.2.7. Minors of the matrix. When investigating matrices and their properties we often work only with parts of the matrices. Therefore we need some new concepts.



where we have used the expansion along the second column in the second step and computed the determinant of the 3×3 matrix directly using the Sarrus rule.

Another option is to try to expand the determinant along several rows, exploiting vanishing of many sub-determinants there. For example, we may use the last two rows. Clearly there might be only two non-zero sub-determinants built from this row there. Thus the entire determinant must be (notice that choosing two lines and two columns always leads to the plus sign in the definition of the algebraic complement, see 2.3.10)

$$\begin{vmatrix} 4 & 4 \\ 0 & 5 \end{vmatrix} \cdot \begin{vmatrix} 0 & 1 & 0 \\ 2 & 0 & 2 \\ 0 & 3 & 0 \end{vmatrix} + \begin{vmatrix} 4 & 4 \\ 0 & 5 \end{vmatrix} \cdot \begin{vmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix} = \\ = 20 \cdot 0 + 20 \cdot 6 = 120$$

□

2.B.6. Find all the values of a such that

$$\begin{vmatrix} a & 1 & 1 & 1 \\ 0 & a & 1 & 1 \\ 0 & 1 & a & 1 \\ 0 & 0 & 0 & -a \end{vmatrix} = 1.$$

For complex a give either its algebraic or polar form.

Solution. We compute the determinant by expanding the first row of the matrix:

$$D = \begin{vmatrix} a & 1 & 1 & 1 \\ 0 & a & 1 & 1 \\ 0 & 1 & a & 1 \\ 0 & 0 & 0 & -a \end{vmatrix} = a \cdot \begin{vmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 0 & 0 & -a \end{vmatrix}.$$

Expand further using the last row:

$$D = a \cdot (-a) \begin{vmatrix} a & 1 \\ 1 & a \end{vmatrix} = -a^2(a^2 - 1).$$

We conclude that $a^4 - a^2 + 1 = 0$. Substituting $t = a^2$ we have $t^2 - t + 1$ with roots $t_1 = \frac{1+i\sqrt{3}}{2} = \cos(\pi/3) + i\sin(\pi/3)$, $t_2 = \frac{1-i\sqrt{3}}{2} = \cos(\pi/3) - i\sin(\pi/3) = \cos(-\pi/3) + i\sin(-\pi/3)$, from where we obtain four possible values for the parameter a: $a_1 = \cos(\pi/6) + i\sin(\pi/6) = \sqrt{3}/2 + i/2$, $a_2 = \cos(7\pi/6) + i\sin(7\pi/6) = -\sqrt{3}/2 - i/2$, $a_3 = \cos(-\pi/6) + i\sin(-\pi/6) = \sqrt{3}/2 - i/2$, $a_4 = \cos(5\pi/6) + i\sin(5\pi/6) = -\sqrt{3}/2 + i/2$.

Alternatively, we can multiply by $a^2 + 1$ to obtain

$$a^6 + 1 = (a^2 + 1)(a^4 - a^2 + 1) = 0.$$

The equation $a^6 = -1$ has six (complex) solutions given by $a = \cos\varphi + i\sin\varphi$ where $\varphi = \pi/6 + k\pi/3 = (2k+1)\pi/6$, $k = 0, 1, 2, 3, 4, 5$. Of these, we must discard the two choices $k = 1$, and $k = 4$, since these choices solve $a^2 + 1 = 0$ and

SUBMATRICES AND MINORS

Let $A = (a_{ij})$ be a matrix of the type m/n and let $1 \le i_1 < \ldots < i_k \le m$, $1 \le j_1 < \ldots < j_l \le n$ be fixed natural numbers. Then the matrix

$$M = \begin{pmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \ldots & a_{i_1 j_\ell} \\ \vdots & & & \vdots \\ a_{i_k j_1} & a_{i_k j_2} & \ldots & a_{i_k j_\ell} \end{pmatrix}$$

of the type k/ℓ is called a *submatrix of the matrix* A determined by the rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_\ell$. The remaining $(m-k)$ rows and $(n-\ell)$ columns determine a matrix M^* of the type $(m-k)/(n-\ell)$, which is called *complementary submatrix* to M in A. When $k = \ell$ we call the determinant $|M|$ the *subdeterminant* or *minor* of the order k of the matrix A. If $m = n$ and $k = \ell$, then M^* is also a square matrix and $|M^*|$ is called the *minor complement* to $|M|$, or *complementary minor* of the submatrix M in the matrix A. The scalar

$$(-1)^{i_1 + \cdots + i_k + j_1 + \cdots + j_l} \cdot |M^*|$$

is then called the *algebraic complement* of the minor $|M|$.

The submatrices formed by the first k rows and columns are called *leading principal submatrices*, and their determinants are called *leading principal minors* of the matrix A. If we choose k sequential rows and columns starting with the i-th row, we speak of *principal matrices* and *principal minors*.

Specially, when $k = \ell = 1$, $m = n$ we call the corresponding algebraic complementary minor the *algebraic complement* A_{ij} of the element a_{ij} of the matrix A.

2.2.8. Laplace determinant expansion. If the principal minor $|M|$ of the matrix A is of the order k, then, directly from the definition of the determinant, each of the individual $k!(n-k)!$ terms in the product of $|M|$ with its algebraic complement is a term of $|A|$.

In general, consider a square submatrix M, that is, a square matrix given by the rows $i_1 < i_2 < \cdots < i_k$ and columns $j_1 < \cdots < j_k$. Then using $(i_1 - 1) + \cdots + (i_k - k)$ exchanges of neighbouring rows and $(j_1 - 1) + \cdots + (j_k - k)$ exchanges of neighbouring columns in A we can transform this submatrix M into a principal submatrix and the complementary matrix gets transformed into its complementary matrix. The whole matrix A gets transformed into a matrix B satisfying (cf. 2.2.4 and the definition of the determinant) $|B| = (-1)^\alpha |A|$, where $\alpha = \sum_{h=1}^{k}(i_h - j_h) - 2(1 + \cdots + k)$. But $(-1)^\alpha = (-1)^\beta$ with $\beta = \sum_{h=1}^{k}(i_h + j_h)$. Therefore we have checked:

Proposition. *If A is a square matrix of dimension n and $|M|$ is its minor of the order $k < n$, then the product of any term of $|M|$ with any term of its algebraic complement is a term in $|A|$.*

not $a^4 - a^2 + 1 = 0$. We conclude that $a = \cos\varphi + i\sin\varphi$ where $\varphi = (2k+1)\pi/6$, $k = 0, 2, 3$, or 5. $\square$

2.B.7. Vandermonde determinant. Prove the formula for the Vandermonde determinant, that is, the determinant of the Vandermonde matrix:

$$V_n = \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2 & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i),$$

where $x_1, \dots, x_n \in \mathbb{R}$ and on the right-hand side of the equation there is the product of all terms $x_j - x_i$ where $j > i$.

Solution. We proceed by induction on n. From technical reasons we work with the transposed Vandermonde matrix (it has the same determinant). By subtracting the first row from all other rows and then expanding the first column we obtain

$$\begin{aligned} &V_n(x_1, x_2, \dots, x_n) \\ &= \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & \dots & x_2^{n-1} - x_1^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n - x_1 & x_n^2 - x_1^2 & \dots & x_n^{n-1} - x_1^{n-1} \end{vmatrix} \\ &= \begin{vmatrix} x_2 - x_1 & x_2^2 - x_1^2 & \dots & x_2^{n-1} - x_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_n - x_1 & x_n^2 - x_1^2 & \dots & x_n^{n-1} - x_1^{n-1} \end{vmatrix}. \end{aligned}$$

If we take out $x_{i+1} - x_1$ from the i-th row for $i \in \{1, 2, \dots, n-1\}$, we obtain

$$\begin{aligned} &V_n(x_1, x_2, \dots, x_n) = (x_2 - x_1) \cdots (x_n - x_1) \\ &\cdot \begin{vmatrix} 1 & x_2 + x_1 & \dots & \sum_{j=0}^{n-2} x_2^{n-j-2} x_1^j \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n + x_1 & \dots & \sum_{j=0}^{n-2} x_n^{n-j-2} x_1^j \end{vmatrix}. \end{aligned}$$

By subtracting from every column (starting with the last and ending with the second) x_1-multiple of the previous column, we obtain

$$\begin{vmatrix} 1 & x_2 + x_1 & \dots & \sum_{j=0}^{n-2} x_2^{n-j-2} x_1^j \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n + x_1 & \dots & \sum_{j=0}^{n-2} x_n^{n-j-2} x_1^j \end{vmatrix}$$

$$= \begin{vmatrix} 1 & x_2 & \dots & x_2^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{n-2} \end{vmatrix}.$$

This claim suggests that we could perhaps express the determinant of the matrix by using some products of smaller determinants. We see that $|A|$ contains exactly $n!$ distinct terms, exactly one for each permutation. These terms are mutually distinct as polynomials in the components of a general matrix A. If we can show that there are exactly that many mutually distinct expressions from the previous claim, we obtain the determinant $|A|$ as their sum.

It remains to show that the terms of the product $|M| \cdot |M^*|$ contain exactly $n!$ distinct members from $|A|$.

From the chosen k rows we can choose $\binom{n}{k}$ minors M and using the previous lemma each of the $k!(n-k)!$ terms in the products of $|M|$ with their algebraic complements is a term in $|A|$. But for distinct choices of M we can never obtain the same terms and the individual terms in $(-1)^{i_1+\dots+i_k+j_1+\dots+j_l} \cdot |M| \cdot |M^*|$ are also mutually distinct. Therefore we have exactly the required number $k!(n-k)!\binom{n}{k} = n!$ of terms.

Thus we have proved:

Laplace theorem

Theorem. *Let $A = (a_{ij})$ be a square matrix of dimension n over arbitrary ring of scalars with k rows fixed. Then $|A|$ is a sum of all $\binom{n}{k}$ products $(-1)^{i_1+\dots+i_k+j_1+\dots+j_l} \cdot |M| \cdot |M^*|$ of minors of the order k chosen among the fixed rows with their algebraic complements.*

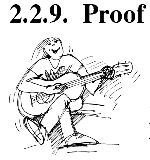
The Laplace theorem transforms the computation of $|A|$ into the computation of determinants of lower dimension. This method of computation is called the *Laplace expansion* along the chosen rows (or columns). For instance, the expansion along the i-th row or the j-th column is:

$$|A| = \sum_{j=1}^{n} a_{ij} A_{ij}$$

where A_{ij} denotes the algebraic complement of the element a_{ij} (that is, minor of order one).

In practical computations, it is often efficient to combine the Laplace expansion with a direct method of Gaussian elimination.

2.2.9. Proof of the Cauchy theorem. The theorem is based on a clever but elementary application of the Laplace theorem. We just use the Laplace expansion twice on a particular arrangement of a well chosen matrix.



Consider first the following matrix H of dimension $2n$ (we are using the so-called block symbolics, that is, we write the matrix as if composed of the (sub)matrices A, B, and so on).

Therefore

$$V_n(x_1, x_2, \dots, x_n)$$
$$= (x_2 - x_1) \cdots (x_n - x_1)\; V_{n-1}(x_2, \dots, x_n).$$

Because it is clear that

$$V_2(x_{n-1}, x_n) = x_n - x_{n-1},$$

it follows by induction that

$$V_n(x_1, x_2, \dots, x_n) = \prod_{1 \le i < j \le n} (x_j - x_i).$$

Note that the determinant is non-zero whenever the numbers $x_1, \dots, x_n$ are mutually distinct. □

Remark. Another (more beautiful?) proof of the formula can be found in **??**.

2.B.8. Find whether or not the matrix

$$\begin{pmatrix} 3 & 2 & -1 & 2 \\ 4 & 1 & 2 & -4 \\ -2 & 2 & 4 & 1 \\ 2 & 3 & -4 & 8 \end{pmatrix}$$

is invertible.

Solution. The matrix is invertible (that is, there is an inverse matrix) whenever we can transform it by elementary row transformations into the unit matrix. That is equivalent for instance to the property that it has non-zero determinant. That we can compute using the Laplace Theorem (2.3.10) by expanding for instance the first row:

$$\begin{vmatrix} 3 & 2 & -1 & 2 \\ 4 & 1 & 2 & -4 \\ -2 & 2 & 4 & 1 \\ 2 & 3 & -4 & 8 \end{vmatrix} = 3 \cdot \begin{vmatrix} 1 & 2 & -4 \\ 2 & 4 & 1 \\ 3 & -4 & 8 \end{vmatrix}$$
$$-2 \cdot \begin{vmatrix} 4 & 2 & -4 \\ -2 & 4 & 1 \\ 2 & -4 & 8 \end{vmatrix} + (-1) \cdot \begin{vmatrix} 4 & 1 & -4 \\ -2 & 2 & 1 \\ 2 & 3 & 8 \end{vmatrix}$$
$$- 2 \cdot \begin{vmatrix} 4 & 1 & 2 \\ -2 & 2 & 4 \\ 2 & 3 & -4 \end{vmatrix}$$
$$= 3 \cdot 90 - 2 \cdot 180 + (-1) \cdot 110 - 2 \cdot (-100) = 0,$$

that is, the given matrix is not invertible. □

2.B.9. Solve the system from 2.A.2 using the Cramer rule (see 2.2.5).

$$H = \begin{pmatrix} A & 0 \\ -E & B \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} & 0 & \dots & 0 \\ -1 & & 0 & b_{11} & \dots & b_{1n} \\ & \ddots & & \vdots & & \vdots \\ 0 & & -1 & b_{n1} & \dots & b_{nn} \end{pmatrix}$$

The Laplace expansion along the first n rows gives $|H| = |A| \cdot |B|$.

Now in sequence, we add linear combinations of the first n columns to the last n columns in order to obtain a matrix with zeros in the bottom right corner. We obtain

$$K = \begin{pmatrix} a_{11} & \dots & a_{1n} & c_{11} & \dots & c_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} & c_{n1} & \dots & c_{nn} \\ -1 & & 0 & 0 & \dots & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & & -1 & 0 & \dots & 0 \end{pmatrix}.$$

The elements of the submatrix on the top right part must satisfy

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj},$$

that is, they are exactly the components of the product $A \cdot B$ and $|K| = |H|$. The expansion of the last n columns gives us $|K| = (-1)^n(-1)^{1+\dots+2n}|A \cdot B| = (-1)^{2n\cdot(n+1)} \cdot |A \cdot B| = |A \cdot B|$. This proves the Cauchy theorem.

2.2.10. Determinant and the inverse matrix. Assume first that there is an inverse matrix of the matrix A, that is, $A \cdot A^{-1} = E$. Since the unit matrix always satisfies $|E| = 1$, it follows that for every invertible matrix its determinant is an invertible scalar and by the Cauchy theorem we have $|A^{-1}| = |A|^{-1}$.

But we can say more, combining the Laplace and Cauchy theorems.

Inverse matrix determinant formula

For any square matrix $A = (a_{ij})$ of dimension n we define a matrix $A^* = (a^*_{ij})$, where $a^*_{ij} = A_{ji}$ are algebraic complements of the elements a_{ji} in A. The matrix A^* is called the *algebraically adjoint matrix* of the matrix A.

Theorem. *For every square matrix A over a ring of scalars $\mathbb{K}$ we have that*

(1) $$AA^* = A^*A = |A| \cdot E.$$

In particular,

(1) A^{-1} exists as a matrix over the ring of scalars $\mathbb{K}$ if and only if $|A|^{-1}$ exists in $\mathbb{K}$.
(2) If A^{-1} exists, then $A^{-1} = |A|^{-1} \cdot A^*$.

5

Solution. We just plug in the values to the rule:

$$x_1 = \frac{\begin{vmatrix} 2 & 2 & 3 \\ -3 & -3 & -1 \\ -3 & 1 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 3 \\ 2 & -3 & 1 \\ -3 & 1 & 2 \end{vmatrix}} = 1, \; x_2 = \frac{\begin{vmatrix} 1 & 2 & 3 \\ 2 & -3 & -1 \\ -3 & -3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 3 \\ 2 & -3 & 1 \\ -3 & 1 & 2 \end{vmatrix}} = 2$$

$$x_3 = \frac{\begin{vmatrix} 1 & 2 & 2 \\ 2 & -3 & -3 \\ -3 & 1 & -3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 3 \\ 2 & -3 & 1 \\ -3 & 1 & 2 \end{vmatrix}} = -1.$$

□

2.B.10. Find the algebraically adjoint matrix and the inverse of the matrix

$$A = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 3 & 0 & 4 \\ 5 & 0 & 6 & 0 \\ 0 & 7 & 0 & 8 \end{pmatrix}.$$

Solution. The adjoint matrix is

$$A^* = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix}^T,$$

where A_{ij} is the algebraic complement of the element a_{ij} of the matrix A, that is, the product of the number $(-1)^{i+j}$ and the determinant of the matrix given by A without the i-th row and j-th column. We have

$$A_{11} = \begin{vmatrix} 3 & 0 & 4 \\ 0 & 6 & 0 \\ 7 & 0 & 8 \end{vmatrix} = -24, \quad A_{12} = -\begin{vmatrix} 0 & 0 & 4 \\ 5 & 6 & 0 \\ 0 & 0 & 8 \end{vmatrix} = 0,$$

$$A_{13} = \begin{vmatrix} 0 & 3 & 4 \\ 5 & 0 & 0 \\ 0 & 7 & 8 \end{vmatrix} = 20, \quad A_{14} = -\begin{vmatrix} 0 & 3 & 0 \\ 5 & 0 & 6 \\ 0 & 7 & 0 \end{vmatrix} = 0,$$

$$A_{21} = -\begin{vmatrix} 0 & 2 & 0 \\ 0 & 6 & 0 \\ 7 & 0 & 8 \end{vmatrix} = 0, \quad A_{22} = \begin{vmatrix} 1 & 2 & 0 \\ 5 & 6 & 0 \\ 0 & 0 & 8 \end{vmatrix} = -32,$$

$$A_{23} = -\begin{vmatrix} 1 & 0 & 0 \\ 5 & 0 & 0 \\ 0 & 7 & 8 \end{vmatrix} = 0, \quad A_{24} = \begin{vmatrix} 1 & 0 & 2 \\ 5 & 0 & 6 \\ 0 & 7 & 0 \end{vmatrix} = -28,$$

$$A_{31} = \begin{vmatrix} 0 & 2 & 0 \\ 3 & 0 & 4 \\ 7 & 0 & 8 \end{vmatrix} = 8, \quad A_{32} = -\begin{vmatrix} 1 & 2 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 8 \end{vmatrix} = -0,$$

$$A_{33} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 7 & 8 \end{vmatrix} = -4, \quad A_{34} = -\begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 0 & 7 & 0 \end{vmatrix} = -0,$$

Proof. As already mentioned, the Cauchy theorem shows that the existence of A^{-1} implies the invertibility of $|A| \in \mathbb{K}$.

For an arbitrary square matrix A we can directly compute $A \cdot A^* = (c_{ij})$, where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} a^*_{kj} = \sum_{k=1}^{n} a_{ik} A_{jk}.$$

If $i = j$, it is exactly the Laplace expansion of $|A|$ along the i-th row. If $i \neq j$, it is the expansion of the determinant of the matrix where the i-th and j-th row is the same, therefore $c_{ij} = 0$. This implies that $A \cdot A^* = |A| \cdot E$, and we have proven the equality (1).

Let us further assume that $|A|$ is an invertible scalar. If we repeat the previous computation for $A^* \cdot A$, we obtain $|A|^{-1} A^* \cdot A = E$. Therefore our computation really gives the inverse matrix of A, as claimed in the theorem. □

Notice that for fields of scalars we have already proved that the right inverse of. a matrix is automatically the left inverse and thus the inverse, too. Here we have obtained the same result for all rings of scalars, together with a strong and effective existence condition. On the other hand the exact formula for the inverse has become rather theoretical with little practical value.

As a direct corollary of this theorem we can once again prove the Cramer rule for solving the systems of linear equations, see 2.2.5. Really, for the solution of the system $A \cdot x = b$ we just need to read in the equation

$$x = A^{-1} \cdot b = |A|^{-1} A^* \cdot b$$

the individual components of the expression $A^* \cdot b$ as the Laplace expansions of the determinant of the matrix A_i which arose through the exchange of the i-th column of A for the column b.

3. Vector spaces and linear mappings

2.3.1. Abstract vector spaces. Let us go back for a while to the systems of m linear equations of n variables from 2.1.3 and further, let us assume that the system is the homogeneous system $A \cdot x = 0$, that is

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

By the distributivity of the matrix multiplication it is clear that the sum of two solutions $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ satisfies

$$A \cdot (x + y) = A \cdot x + A \cdot y = 0$$

and thus is also a solution. Similarly, a scalar multiple $a \cdot x$ is also a solution. The set of all solutions of a fixed system of equations is therefore closed under vector addition and scalar multiplication. These are the basic properties of vectors of dimension n in $\mathbb{K}^n$, see 2.1.1. Now we have the vectors in the

$$A_{41} = -\begin{vmatrix} 0 & 2 & 0 \\ 3 & 0 & 4 \\ 0 & 6 & 0 \end{vmatrix} = 0, \quad A_{42} = \begin{vmatrix} 1 & 2 & 0 \\ 0 & 0 & 4 \\ 5 & 6 & 0 \end{vmatrix} = -16,$$

$$A_{43} = -\begin{vmatrix} 1 & 0 & 0 \\ 0 & 3 & 4 \\ 5 & 0 & 0 \end{vmatrix} = 0, \quad A_{44} = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 5 & 0 & 6 \end{vmatrix} = -12.$$

By substitution we obtain

$$A^* = \begin{pmatrix} -24 & 0 & 20 & 0 \\ 0 & -32 & 0 & 28 \\ 8 & 0 & -4 & 0 \\ 0 & 16 & 0 & -12 \end{pmatrix}^T = \begin{pmatrix} -24 & 0 & 8 & 0 \\ 0 & -32 & 0 & 16 \\ 20 & 0 & -4 & 0 \\ 0 & 28 & 0 & -12 \end{pmatrix}.$$

We compute the inverse matrix A^{-1} from the relation $A^{-1} = |A|^{-1} \cdot A^*$. The determinant of the matrix A is (expanding the first row) equal to

$$|A| = \begin{vmatrix} 1 & 0 & 2 & 0 \\ 0 & 3 & 0 & 4 \\ 5 & 0 & 6 & 0 \\ 0 & 7 & 0 & 8 \end{vmatrix} = \begin{vmatrix} 3 & 0 & 4 \\ 0 & 6 & 0 \\ 7 & 0 & 8 \end{vmatrix} + 2\begin{vmatrix} 0 & 3 & 4 \\ 5 & 0 & 0 \\ 0 & 7 & 8 \end{vmatrix} = 16.$$

By substitution, we obtain

$$A^{-1} = \begin{pmatrix} -3/2 & 0 & 1/2 & 0 \\ 0 & -2 & 0 & 1 \\ 5/4 & 0 & -1/4 & 0 \\ 0 & 7/4 & 0 & -3/4 \end{pmatrix}.$$

□

C. Vector spaces

Typical properties of vector spaces (met already in the plane or three dimensional space) can be observed in many other situations. We illustrate this by examples.

2.C.1. Vector space – yes or no? Decide whether following sets form a vector space over the field of real numbers:

i) The set of solutions of the system

$$\begin{aligned} x_1 + x_2 + \cdots + x_{98} + x_{99} + x_{100} &= 100x_1, \\ x_1 + x_2 + \cdots + x_{98} + x_{99} &= 99x_1, \\ x_1 + x_2 + \cdots + x_{98} &= 98x_1, \\ &\vdots \\ x_1 + x_2 &= 2x_1. \end{aligned}$$

solution space with n coordinates. The "dimension" of this space is given by the difference of the number of variables and the rank of the matrix A. Thus we can easily deal with the solution of a system of 1000 equations in 1000 variables and need only one or two free parameters. Thus the whole solution space will behave as a plane or a line, as we have already seen in **??** at the page **??**, although the vectors themselves are given by so many components.

We go further. Already in paragraph **??** we have encountered an interesting example of a space of all solutions of a homogeneous linear difference equation of first order. All solutions have been obtained from a single one by scalar multiplication and are also closed under addition and scalar multiples. These "vectors" of solutions are infinite sequences of numbers, although we intuitively expect that the "dimension" of the whole space of solutions should be one. We shall understand such phenomena with the help of a more general definition of vector space and its dimension.

Vector space definition

A vector space V over a field of scalars $\mathbb{K}$ is a set where we define the operations

- addition, which satisfies the axioms (CG1)–(CG4) from the paragraph **??** on the page **??**,
- scalar multiplication, for which the axioms (V1)–(V4) from the paragraph 2.1.1 on the page 5 hold.

Recall our simple notational convention: scalars are usually denoted by letters from the beginning of the alphabet, that is, $a, b, c, \dots$, while for vectors we shall use letters from the end, that is, u, v, w, x, y, z. Usually, x, y, z will denote n-tuples of scalars. For completeness, the letters from the centre of the alphabet, for instance i, j, k, ℓ, will mostly denote indices.

In order to gain some practice in the formal approach, we check some simple properties of vectors. These are trivial for n-tuples for scalars, but not so evident for general vectors in our new abstract sense.

2.3.2. Proposition. *Let V be a vector space over a field of scalars $\mathbb{K}$. Suppose $a, b, a_i \in \mathbb{K}$, and $u, v, u_j \in V$. Then*

(1) $a \cdot u = 0$ if and only if $a = 0$ or $u = 0$,
(2) $(-1) \cdot u = -u$,
(3) $a \cdot (u - v) = a \cdot u - a \cdot v$,
(4) $(a - b) \cdot u = a \cdot u - b \cdot u$,
(5) $\left(\sum_{i=1}^n a_i\right) \cdot \left(\sum_{j=1}^m u_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i \cdot u_j$.

Proof. We can expand

$$(a + 0) \cdot u \overset{(V2)}{=} a \cdot u + 0 \cdot u = a \cdot u$$

which, according to the axiom (CG4), implies $0 \cdot u = 0$. Now

$$u + (-1) \cdot u \overset{(V2)}{=} (1 + (-1)) \cdot u = 0 \cdot u = 0$$

and thus $-u = (-1) \cdot u$. Further,

$$a \cdot (u + (-1) \cdot v) \overset{(V2,V3)}{=} a \cdot u + (-a) \cdot v = a \cdot u - a \cdot v,$$

ii) The set of solutions of the equation

$$x_1 + x_2 + \cdots + x_{100} = 0$$

iii) The set of solutions of the equation

$$x_1 + 2x_2 + 3x_3 + \cdots + 100x_{100} = 1.$$

iv) The set of all real (or complex) sequences. (Real or complex sequence is a mapping $f : \mathbb{N} \to \mathbb{R}$ or $f : \mathbb{N} \to \mathbb{C}$. The image of number n is then called n-th member of the sequence, we usually denote it by lower index, say a_n.)

v) The set of solutions of a homogeneous difference equation.

vi) The set of solutions of a non-homogeneous difference equation.

vii) $\{f : \mathbb{R} \to \mathbb{R} | f(1) = f(2) = c, c \in \mathbb{R}\}$

Solution. We check the properties of a vector space, see 2.3.1. Actually all we have to do is to check whether the given sets are closed to linear combinations of it's elements. Then all the axioms of a vector space are satisfied.

i) Yes. They all are real multiples of the vector $\underbrace{(1, 1, 1 \ldots, 1)}_{\text{100 ones}}$. A sum of two multiples of the same vector is again a multiple of the vector. The reverse vector is again a multiple of the vcetor and all other axioms are trivially satisfied. By the way, the solution space is thus a vector space of dimension 1, see also 2.3.7.

ii) Yes. It is a space of dimension 99 (corresponds to the number of free parameters of the solution). In general the set of all solutions of any system of homogeneous linear equations forms a vector space.

iii) No. For instance, taking twice the solution $x_1 = 1, x_i = 0, i = 2, \ldots 100$ we do not obtain a solution. But the set of solutions forms an affine space (see **??**).

iv) Yes. The set of all real or complex sequences clearly forms a real (complex) vector space. Adding the sequences and scalar multiplication is defined term-wise, where it is clearly the vector space of all real (complex) numbers.

v) Yes. In order to show that the set of sequences which satisfy given difference homogeneous equation it is enough to show that it is closed under addition and real number multiplication (as the set of all real sequences is a vector space, as we know). Consider two sequences $(x_j)_{j=0}^{\infty}$

which proves (3). It follows that

$$(a - b) \cdot u \overset{(V2,V3)}{=} a \cdot u + (-b) \cdot u = a \cdot u - b \cdot u$$

which proves (4). Property (5) follows using induction with (V2) and (V1).

It remains to prove (1): $a \cdot 0 = a \cdot (u - u) = a \cdot u - a \cdot u = 0$, which along with the first derived proposition in this proof proves one implication. For the other implication, we use an axiom for the field of scalars, and axiom (V4) for vector spaces: if $p \cdot u = 0$ and $p \neq 0$, then $u = 1 \cdot u = (p^{-1} \cdot p) \cdot u = p^{-1} \cdot 0 = 0$. □

2.3.3. Linear (in)dependence. In paragraph 2.1.11 we worked with linear combinations of rows of a matrix. With vectors we work analogously:

Linear combination and independence

An expression of the form $a_1 \, v_1 + \cdots + a_k \, v_k$ is called a *linear combination* of vectors $v_1, \ldots, v_k \in V$.

A finite sequence of vectors $v_1, \ldots, v_k$ is called *linearly independent*, if the only zero linear combination is the one with all coefficients zero. That is, for scalars $a_1, \ldots, a_k \in \mathbb{K}$,

$$a_1 \, v_1 + \cdots + a_k \, v_k = 0 \implies a_1 = a_2 = \cdots = a_k = 0.$$

It is clear that for an independent sequence of vectors, all vectors are mutually distinct and nonzero.

The set of vectors $M \subset V$ in a vector space V over $\mathbb{K}$ is called *linearly independent*, if every finite k-tuple of vectors $v_1, \ldots, v_k \in M$ is linearly independent.

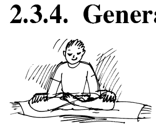
The set of vectors M is *linearly dependent*, if it is not linearly independent.

A nonempty subset M of vectors in a vector space over a field of scalars $\mathbb{K}$ is dependent if and only if one of its vectors can be expressed as a finite linear combination using other vectors in M. This follows directly from the definition.

At least one of the coefficients in the corresponding linear combination must be nonzero, and since we are over a field of scalars, we can multiply whole combination by the inverse of this nonzero coefficient and thus express its corresponding vector as a linear combination of the others.

Every subset of a linearly independent set M is clearly also linearly independent (we require the same conditions on a smaller set of vectors). Similarly, we can see that $M \subset V$ is linearly independent if and only if every finite subset of M is linearly independent.

2.3.4. Generators and subspaces. A subset $M \subset V$ is called a *vector subspace* if it forms, together with the restricted operations of addition and scalar multiplication, a vector space. That is, we require



$$\forall a, b \in \mathbb{K}, \; \forall v, w \in M, \; a \cdot v + b \cdot w \in M.$$

and $(y_j)_{j=0}^{\infty}$ satisfying the given equation, that is,

$$\begin{aligned} a_n x_{n+k} + a_{n-1}x_{n+k-1} + \cdots + a_0 x_k &= 0 \\ a_n y_{n+k} + a_{n-1}y_{n+k-1} + \cdots + a_0 y_k &= 0. \end{aligned}$$

By adding these equations, we obtain

$$\begin{aligned} a_n(x_{n+k} + y_{n+k}) + a_{n-1}(x_{n+k-1} + y_{n+k-1}) \\ + \cdots + a_0(x_k + y_k) = 0, \end{aligned}$$

therefore also the sequence $(x_j + y_j)_{j=0}^{\infty}$ satisfies the given equation. Analogously, if the sequence $(x_j)_{j=0}^{\infty}$ satisfies the given equation, then also $(ux_j)_{j=0}^{\infty}$, where $u \in \mathbb{R}$.

vi) No. The sum of two solutions of a non-homogeneous equation

$$a_n x_{n+k} + a_{n-1}x_{n+k-1} + \cdots + a_0 x_k = c$$
$$a_n y_{n+k} + a_{n-1}y_{n+k-1} + \cdots + a_0 y_k = c,\ c \in \mathbb{R} - \{0\}$$

satisfies the equation

$$\begin{aligned} a_n(x_{n+k} + y_{n+k}) + a_{n-1}(x_{n+k-1} + y_{n+k-1}) \\ + \cdots + a_0(x_k + y_k) = 2c, \end{aligned}$$

that is, it does not satisfy the original non-homogeneous equation. But the set of solutions forms an affine space, see **??**.

vii) It is a vector space if and only if $c = 0$. If we take two functions f and g from the given set, then $(f+g)(1) = (f+g)(2) = f(1) + g(1) = 2c$. Thus if $f+g$ is to be a member of the given set, it must be that $(f+g)(1) = c$, therefore $2c = c$, hence $c = 0$.

□

2.C.2. Find out, whether the set

$$U_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3;\ |\,x_1\,| = |\,x_2\,| = |\,x_3\,|\}$$

is a subspace of a vector space $\mathbb{R}^3$ and the set

$$U_2 = \{ax^2 + c;\ a, c \in \mathbb{R}\}$$

a subspace of the space of polynomials of degree at most 2.

Solution. The set U_1 is not a vector (sub)space. We can see that, for instance,

$$(1,1,1) + (-1,1,1) = (0,2,2) \notin U_1.$$

The set U_2 is a subspace (there is a clear identification with $\mathbb{R}^2$), because

$$\left(a_1x^2 + c_1\right) + \left(a_2x^2 + c_2\right) = (a_1 + a_2)\,x^2 + (c_1 + c_2),$$

We investigate a couple of cases: The space of m-tuples of scalars $\mathbb{R}^m$ with coordinate-wise addition and multiplication is a vector space over $\mathbb{R}$, but also a vector space over $\mathbb{Q}$. For instance for $m = 2$, the vectors $(1,0), (0,1) \in \mathbb{R}^2$ are linearly independent, because from

$$a \cdot (1,0) + b \cdot (0,1) = (0,0)$$

follows $a = b = 0$. Further, the vectors $(1,0), (\sqrt{2}, 0) \in \mathbb{R}^2$ are linearly dependent over $\mathbb{R}$, because $\sqrt{2} \cdot (1,0) = (\sqrt{2}, 0)$, but over $\mathbb{Q}$ they are linearly independent! Over $\mathbb{R}$ these two vectors “generate” a one-dimensional subspace, while over $\mathbb{Q}$ the subspace is “larger”.

Polynomials of degree at most m form a vector space $\mathbb{R}_m[x]$. We can consider the polynomials as mappings $f : \mathbb{R} \to \mathbb{R}$ and define the addition and scalar multiplication like this: $(f+g)(x) = f(x) + g(x)$, $(a \cdot f)(x) = a \cdot f(x)$. Polynomials of all degrees also form a vector space $\mathbb{R}[x]$ and $\mathbb{R}_m[x] \subset \mathbb{R}_n[x]$ is a vector subspace for any $m \le n \le \infty$. Further examples of subspaces is given by all even polynomials or all odd polynomials, that is, polynomials satisfying $f(-x) = \pm f(x)$.

In complete analogy with polynomials, we can define a vector space structure on a set of all mappings $\mathbb{R} \to \mathbb{R}$. or of all mappings $M \to V$ of an arbitrary fixed set M into the vector space V.

Because the condition in the definition of subspace consists only of universal quantifiers, the intersection of subspaces is still a subspace. We can see this also directly: Let W_i, $i \in I$, be vector subspaces in V, $a, b \in \mathbb{K}$, $u, v \in \cap_{i \in I} W_i$. Then $a \cdot u + b \cdot v \in W_i$ for all $i \in I$. Hence $a \cdot u + b \cdot v \in \cap_{i \in I} W_i$.

It can be noted that the intersection of all subspaces $W \subset V$ that contain some given set of vectors $M \subset V$ is a subspace. It is called span M.

We say that a set M *generates* the subspace $\operatorname{span} M$, or that the elements of M are *generators* of the subspace $\operatorname{span} M$.

We formulate a few simple claims about subspace generation:

Proposition. *For every nonempty set $M \subset V$, we have*

(1) $\operatorname{span} M = \{a_1 \cdot u_1 + \cdots + a_k \cdot u_k;\ k \in \mathbb{N}, a_i \in \mathbb{K}, u_j \in M, j = 1, \ldots, k\}$*;*
(2) $M = \operatorname{span} M$ *if and only if* M *is a vector subspace;*
(3) if $N \subset M$ *then* $\operatorname{span} N \subset \operatorname{span} M$ *is a vector subspace; the subspace* $\operatorname{span} \emptyset$ *generated by the empty subspace is the trivial subspace* $\{0\} \subset V$.

Proof. (1) The set of all linear combinations

$$a_1u_1 + \cdots + a_ku_k$$

on the right-hand side of (1) is clearly a vector subspace and of course it contains M. On the other hand, each of the linear combinations must be in $\operatorname{span} M$ and thus the first claim is proved.

$$k \cdot \left(ax^2 + c\right) = (ka)\, x^2 + kc$$

for all numbers $a_1, c_1, a_2, c_2, a, c, k \in \mathbb{R}$. □

D. Linear (in)dependence

2.D.1. Determine whether or not the vectors $(1, 2, 3, 1)$, $(1, 0, -1, 1)$, $(2, 1, -1, 3)$ and $(0, 0, 3, 2)$ are linearly independent.

Solution. Because

$$\begin{vmatrix} 1 & 2 & 3 & 1 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & -1 & 3 \\ 0 & 0 & 3 & 2 \end{vmatrix} = 10 \neq 0,$$

the given vectors are linearly independent. □

2.D.2. Given arbitrary linearly independent vectors u, v, w, z in a vector space V, decide whether or not in V the vectors

$$u - 2v, \quad 3u + w - z, \quad u - 4v + w + 2z, \quad 4v + 8w + 4z$$

are linearly independent.

Solution. Considered vectors are linearly independent if and only if the vectors $(1, -2, 0, 0)$, $(3, 0, 1, -1)$, $(1, -4, 1, 2)$, $(0, 4, 8, 4)$ are linearly independent in $\mathbb{R}^4$. We have

$$\begin{vmatrix} 1 & -2 & 0 & 0 \\ 3 & 0 & 1 & -1 \\ 1 & -4 & 1 & 2 \\ 0 & 4 & 8 & 4 \end{vmatrix} = -36 \neq 0,$$

thus the vectors are linearly independent. □

2.D.3. The vectors

$$(1, 2, 1), \quad (-1, 1, 0), \quad (0, 1, 1)$$

are linearly independent, and therefore together form a basis of $\mathbb{R}^3$ (for basis it is important to give an order of the vectors). Every three-dimensional vector is therefore some linear combination of them. What linear combination corresponds to the vector $(1, 1, 1)$, or equivalently, what are the coordinates of the vector $(1, 1, 1)$ in the basis formed by the given vectors?

Solution. We seek a, b, $c \in \mathbb{R}$ such that $a(1, 2, 1) + b(-1, 1, 0) + c(0, 1, 1) = (1, 1, 1)$. The equation must hold in every coordinate, so we have a system of three linear equations in three variables:

$$\begin{aligned} a - b &= 1 \\ 2a + b + c &= 1 \\ a + c &= 1, \end{aligned}$$

Claim (2) follows immediately from claim (1) and from the definition of vector space. Analogously, (1) implies the third claim.

Finally, the smallest possible vector subspace is $\{0\}$. Notice that the empty set is contained in every subspace and each of them contains the vector 0. This proves the last claim. □

Basis and dimension

A subset $M \subset V$ is called a *basis of the vector space V* if $\operatorname{span} M = V$ and M is linearly independent.

A vector space with a finite basis is called *finitely dimensional*. The number of elements of the basis is called the *dimension of V*.

If V does not have a finite basis, we say that V is *infinitely dimensional*. We write $\dim V = k$, $k \in \mathbb{N}$ or $k = \infty$.

In order to be satisfied with such a definition of dimension, we must know that different bases of the same space will always have the same number of elements. We shall show this below. But we note immediately, that the trivial subspace is generated by the empty set, which is an "empty" basis. Thus it has dimension zero.

2.3.5. Back to systems of linear equations. It is a good time now to recall the properties of systems of linear equation in terms of abstract vector spaces and their bases. As we have already noted in the introduction to this section (cf. 2.3.1), the set of all solutions of the *homogeneous system*

$$A \cdot x = 0$$

is a vector space. If A is a matrix with m rows and n columns, and the rank of the matrix is k, then using the row echelon transformation (see 2.1.7)to solve the system, we find that the dimension of the space of all solutions is exactly $n - k$.

Indeed, the left hand side of the equation can be understood as the linear combination of the columns of A with coefficients given by x and the rank k of the matrix provides the number of linearly independent columns in A, thus the dimension of the subspace of all possible linear combinations of the given form. Therefore, after transforming the system into row echelon form, exactly $m - k$ zero rows remain. In the next step, we are left with exactly $n - k$ free parameters. By setting one of them to have value one, while all others are zero, we obtain exactly $n - k$ linearly independent solutions. Then all solutions are given by all the linear combinations of these $n - k$ solutions. Every such $(n - k)$-tuple of solutions is called a *fundamental system of solutions* of the given homogeneous system of equations. We have proved:

Proposition. *The set of all solutions of the homogeneous system of equations*

$$A \cdot x = 0$$

for n variables with the matrix A of rank k is a vector subspace in $\mathbb{K}^n$ of dimension $n - k$. Every basis of this space forms a fundamental system of solutions of the given homogeneous system.

whose solution gives us $a = \frac{1}{2}$, $b = -\frac{1}{2}$, $c = \frac{1}{2}$, thus we have

$$(1,1,1) = \frac{1}{2}\cdot(1,2,1) - \frac{1}{2}\cdot(-1,1,0) + \frac{1}{2}\cdot(0,1,1),$$

that is, the coordinates of the vector $(1,1,1)$ in the basis $((1,2,1),(-1,1,0),(0,1,1))$ are $(\frac{1}{2},-\frac{1}{2},\frac{1}{2})$. □

2.D.4. Determine all constants $a \in \mathbb{R}$ such that the polynomials ax^2+x+2, $-2x^2+ax+3$ and x^2+2x+a are linearly dependent (in the vector space $P_3[x]$ of polynomials of one variable of degree at most three over real numbers).

Solution. In the basis 1, x, x^2 the coefficients of the given vectors (polynomials) are $(a,1,2)$, $(-2,a,3)$, $(1,2,a)$. Polynomials are linearly independent if and only if the matrix whose columns are given by the coordinates of the vectors has a rank lower than the number of the vectors. In this case the rank must be two or less. In the case of a square matrix, a rank less than the number of rows means that the determinant is zero. The condition for a thus reads

$$\begin{vmatrix} a & -2 & 1 \\ 1 & a & 2 \\ 2 & 3 & a \end{vmatrix} = 0,$$

that is, a is a root of the polynomial $a^3 - 6a - 5 = (a+1)(a^2-a-5)$, thus there are 3 such constants $a_1 = -1$, $a_2 = \frac{1+\sqrt{21}}{2}$, $a_3 = \frac{1-\sqrt{21}}{2}$. □

2.D.5. Consider the complex numbers $\mathbb{C}$ as a real vector space. Determine the coordinates of the number $2+i$ in the basis given by the roots of the polynomial x^2+x+1.

Solution. Because roots of the given polynomial are $-\frac{1}{2} + i\frac{\sqrt{3}}{2}$ and $-\frac{1}{2} - i\frac{\sqrt{3}}{2}$, we have to determine the coordinates (a,b) of the vector $2+i$ in the basis $(-\frac{1}{2}+i\frac{\sqrt{3}}{2}, -\frac{1}{2}-i\frac{\sqrt{3}}{2})$. These real numbers a, b are uniquely determined by the condition

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

By equating separately the real and the imaginary parts of the equation, we obtain a system of two linear equations in two variables:

$$\begin{aligned} -\frac{1}{2}a - \frac{1}{2}b &= 2 \\ \frac{\sqrt{3}}{2}a - \frac{\sqrt{3}}{2}b &= 1. \end{aligned}$$

The solution gives us $a = -2 + \frac{\sqrt{3}}{3}$, $b = -2 - \frac{\sqrt{3}}{3}$, therefore the coordinates are $(-2+\frac{1}{\sqrt{3}}, -2-\frac{1}{\sqrt{3}})$. □

Next, consider the general system of equations

$$A \cdot x = b.$$

Notice that the columns of the matrix A are actually images of the vectors of the standard basis in $\mathbb{K}^n$ under the mapping assigning the vector $A\cdot x$ to each vector x. If there should be a solution, b must be in the image under this mapping and thus it must be a linear combination of the columns in A.

If we extend the matrix A by the column b, the number of linearly independent columns and thus also rows might increase (but does not have to). If this number increases, then b is not in the image and the system of equations does not have a solution. If on the other hand the number of linearly independent rows does not change after adding the column b to the matrix A, it means that b must be a linear combination of the columns of A. Coefficients of such combinations are then exactly the solutions of our system.

Consider now two fixed solutions x and y of our system and some solution z of the homogeneous system with the same matrix. Then clearly

$$\begin{aligned} A\cdot(x-y) &= b - b = 0 \\ A\cdot(x+z) &= 0 + b = b. \end{aligned}$$

Thus we can summarise in the form of the so called *Kronecker-Capelli theorem*[1]:

KRONECKER-CAPELLI THEOREM

Theorem. *The solution of a non-homogeneous system of linear equations $A\cdot x = b$ exists if and only if adding the column b to the matrix A does not increase the number of linearly independent rows. In such a case the space of all solution is given by all sums of one fixed particular solution of the system and all solutions of the homogeneous system that has the same matrix.*

2.3.6. Sums of subspaces. Since we now have some intuition about generators and the subspaces generated by them, we should understand the possibilities of how some subspaces can generate the whole space V.

[1] A common formulation of this fact is "system has a solution if and only if the rank of its matrix equals the rank of its extended matrix". Leopold Kronecker was a very influential German Mathematician, who dealt with algebraic equations in general and in particular pushed forward Number Theory in the middle of 19th century. Alfredo Capelli, an Italian, worked on algebraic identities. This theorem is equally often called by different names, e.g. *Rouché-Frobenius theorem* or *Rouché-Capelli theorem* etc. This is a very common feature in Mathematics.

2.D.6. Remark. As a perceptive reader may have spotted, the problem statement is not unambiguous – we are not given the order of the roots of the polynomial, thus we do not have the order of the basis vectors. The result is thus given up to the permutation of the coordinates.

We add a remark about rationalising the denominator, that is, removing the square roots from the denominator. The authors do not have a distinctive attitude whether this should always be done or not (Does $\frac{\sqrt{3}}{3}$ look better than $\frac{1}{\sqrt{3}}$?). In some cases the rationalising is undesirable: from the fraction $\frac{6}{\sqrt{35}}$ we can immediately spot that its value is a little greater than 1 (because $\sqrt{35}$ is just a little smaller than 6), while for the rationalised fraction $\frac{6\sqrt{35}}{35}$ we cannot spot anything. But in general the convention is to normalize.

2.D.7. Consider complex numbers $\mathbb{C}$ as a real vector space. Determine the coordinates of the number $2+i$ in the basis given by the roots of the polynomial $x^2 - x + 1$. ○

2.D.8. For what values of the parameters $a, b, c \in \mathbb{R}$ are the vectors $(1, 1, a, 1)$, $(1, b, 1, 1)$, $(c, 1, 1, 1)$ linearly dependent? ○

2.D.9. Let a vector space V be given along with a basis formed by the vectors u, v, w, z. Determine whether or not the vectors

$$u - 3v + z, \quad v - 5w - z, \quad 3w - 7z, \quad u - w + z$$

are linearly independent. ○

2.D.10. Complete the vectors $1 - x^2 + x^3$, $1 + x^2 + x^3$, $1 - x - x^3$ to a basis of the space of polynomials of degree at most 3. ○

2.D.11. Do the matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 4 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} -5 & 0 \\ 3 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & -2 \\ 0 & 3 \end{pmatrix}$$

form a basis of the vector space of square two-dimensional matrices?

Solution. The four given matrices are as vectors in the space of 2×2 matrices linearly independent. It follows from the fact that the matrix

$$\begin{pmatrix} 1 & 1 & -5 & 1 \\ 0 & 4 & 0 & -2 \\ 1 & 0 & 3 & 0 \\ -2 & -1 & 0 & 3 \end{pmatrix}$$

Sum of subspaces

Let V_i, $i \in I$ be subspaces of V. Then the subspace generated by their union, that is, $\operatorname{span} \cup_{i \in I} V_i$, is called the *sum of subspaces* V_i. We denote it as $W = \sum_{i \in I} V_i$. Notably, for a finite number of subspaces $V_1, \dots, V_k \subset V$ we write

$$W = V_1 + \cdots + V_k = \operatorname{span} V_1 \cup V_2 \cup \cdots \cup V_k.$$

We see that every element in the considered sum W can be expressed as a linear combination of vectors from the subspaces V_i. Because vector addition is commutative, we can aggregate summands that belong to the same subspace and for a finite sum of k subspaces we obtain

$$V_1 + V_2 + \cdots + V_k = \{v_1 + \cdots + v_k;\ v_i \in V_i, i = 1, \dots, k\}.$$

The sum $W = V_1 + \cdots + V_k \subset V$ is called the *direct sum* of subspaces if the intersection of any two is trivial, that is, $V_i \cap V_j = \{0\}$ for all $i \neq j$. We show that in such a case, every vector $w \in W$ can be written in a unique way as the sum

$$w = v_1 + \cdots + v_k,$$

where $v_i \in V_i$. Indeed, if we could simultaneously write w as $w = v_1' + \cdots + v_k'$, then

$$0 = w - w = (v_1 - v_1') + \cdots + (v_k - v_k').$$

If $v_i - v_i'$ is the first nonzero term of the right-hand side, then this vector from V_i can be expressed using vectors from the other subspaces. This is a contradiction to the assumption that V_i has zero intersection with all the other subspaces. The only possibility is then that all the vectors on the right-hand side are zero and thus the expression of w is unique.

For direct sums of subspaces we write

$$W = V_1 \oplus \cdots \oplus V_k = \oplus_{i=1}^{k} V_i.$$

2.3.7. Basis. Now we have everything prepared for understanding minimal sets of generators as we understood them in the plane $\mathbb{R}^2$ and to prove the promised indepence of the number of basis elements on any choices.

A basis of a k-dimensional space will usually be denoted as a k-tuple $\underline{v} = (v_1 \dots, v_k)$ of basis vectors. This is just a matter of convention: with finitely dimensional vector spaces we shall always consider the bases along with a given order of the elements, even if we have not defined it that way (strictly speaking).

Clearly, if $(v_1, \dots, v_n)$ is a basis of V, then the whole space V is the direct sum of the one-dimensional subspaces

$$V = \operatorname{span}\{v_1\} \oplus \cdots \oplus \operatorname{span}\{v_n\}.$$

An immediate corollary of the derived uniqueness of decomposition of any vector w in V into the components in the direct sum gives a unique decomposition

$$w = x_1 v_1 + \cdots + x_n v_n.$$

This allows us, after choosing a basis, to see the abstract vectors again as n-tuples of scalars. We shall return to this idea is invertible (which is by the way equivalent to any of the following claims: its rank equals its dimension; it can be transformed into the unit matrix by elementary row transformations; it has the inverse matrix; it has a non-zero determinant (equal to 116); it stands for a system of homogeneous linear equations with only zero solution; every non-homogeneous linear system with left-hand side given by this matrix has a unique solution; the range of a linear mapping given by this matrix is a vector space of dimension 4 – this mapping is injective). □

2.D.12. In the vector space $\mathbb{R}^4$ we are given three-dimensional subspaces

$$U = \operatorname{span}\{u_1, u_2, u_3\}, \qquad V = \operatorname{span}\{v_1, v_2, v_3\},$$

while

$$u_1 = \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, \; u_2 = \begin{pmatrix}1\\1\\0\\1\end{pmatrix}, \; u_3 = \begin{pmatrix}1\\0\\1\\1\end{pmatrix}, \quad v_1 = \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix},$$

$v_2 = (1, -1, 1, -1)^T$, $v_3 = (1, -1, -1, 1)^T$. Determine the dimension and find a basis of the subspace $U \cap V$.

Solution. The subspace $U \cap V$ contains exactly the vectors that can be obtained as a linear combinations of vectors u_i and also as a linear combination of vectors v_i. Thus we search for numbers $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$ such that the following holds:

$$x_1\begin{pmatrix}1\\1\\1\\0\end{pmatrix} + x_2\begin{pmatrix}1\\1\\0\\1\end{pmatrix} + x_3\begin{pmatrix}1\\0\\1\\1\end{pmatrix} = y_1\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} + y_2\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix} + y_3\begin{pmatrix}1\\-1\\-1\\1\end{pmatrix},$$

that is, we are looking for a solution of a system

$$\begin{array}{ccccccccccc} x_1 & + & x_2 & + & x_3 & = & y_1 & + & y_2 & + & y_3, \\ x_1 & + & x_2 & & & = & y_1 & - & y_2 & - & y_3, \\ x_1 & & & + & x_3 & = & -y_1 & + & y_2 & - & y_3, \\ & & x_2 & + & x_3 & = & -y_1 & - & y_2 & + & y_3. \end{array}$$

Using matrix notation of this homogeneous system (and preserving the order of the variables) we have

$$\begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 0 & -1 & 1 & 1 \\ 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & 1 & 1 & 1 & -1 \end{pmatrix}$$

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

in paragraph 2.3.11, when we finish the discussion of the existence of bases and sums of subspaces in the general case.

2.3.8. Theorem. *From any finite set of generators of a vector space V we can choose a basis. Every basis of a finitely dimensional space V has the same number of elements.*

Proof. The first claim is easily proved using induction on the number of generators k.

Only the zero subspace does not need a generator and thus we are able to choose an empty basis. On the other hand, we are not able to choose the zero vector (the generators would then be linearly dependent) and there is nothing else in the subspace.

In order to have our inductive step more natural, we deal with the case $k = 1$ first. We have $V = \operatorname{span}\{v\}$ and $v \neq 0$, because $\{v\}$ is a linearly independent set of vectors. Then $\{v\}$ is also a basis of the vector space V and any other vector is a multiple of v, so all bases of V must contain exactly one vector, which can be chosen from any set of generators.

Assume that the claim holds for $k = n$ and consider $V = \operatorname{span}\{v_1, \dots, v_{n+1}\}$. If $v_1, \dots, v_{n+1}$ are linearly independent, then they form a basis. If they are linearly dependent, there exists i such that

$$v_i = a_1v_1 + \cdots + a_{i-1}v_{i-1} + a_{i+1}v_{i+1} + \cdots + a_{n+1}v_{n+1}.$$

Then $V = \operatorname{span}\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{n+1}\}$ and we can choose a basis, using the inductive assumption.

In remains to show that bases always have the same number of elements. Consider a basis $(v_1, \dots, v_n)$ of the space V and for an arbitrary nonzero vector u, consider

$$u = a_1v_1 + \cdots + a_nv_n \in V$$

with $a_i \neq 0$ for some i. Then

$$v_i = \frac{1}{a_i}\big(u - (a_1v_1 + \cdots + a_{i-1}v_{i-1} + a_{i+1}v_{i+1} + \cdots + a_nv_n)\big)$$

and therefore also $\operatorname{span}\{u, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n\} = V$.

We show that this is again a basis. For if adding u to the linearly independent vectors $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n$ leads to a set of linearly dependent vectors, then

$$V = \operatorname{span}\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n\},$$

which implies a basis of $n-1$ vectors, which is not possible.

Thus we have proved that for any nonzero vector $u \in V$ there exists i, $1 \leq i \leq n$, such that $(u, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ is again a basis of V.

Similarly, instead of one vector u, we can consider a linearly independent set $u_1, \dots, u_k$. We will sequentially add $u_1, u_2, \dots$, always exchanging for some v_i using our previous approach. We have to ensure that there always is such v_i to be replaced (that is, that the vectors u_i will not consequently replace each other).

Assume thus that we have already placed $u_1, \dots, u_\ell$ instead of some v_j's. Then the vector $u_{\ell+1}$ can be expressed as a linear combination of the latter vectors u_i and the remaining v_j's. If any of the coefficients at $u_1, \dots, u_\ell$ were nonzero,

$$\sim \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 0 & 1 & 1 & 1 & 1 & -1 \\ 0 & 0 & -1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 3 & 1 & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 0 & 1 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 & -2 & -2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & -2 & -2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & -2 & -2 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

We obtain a solution

$$x_1 = -2t,\ x_2 = -2s,\ x_3 = 2s + 2t,\ y_1 = -s - t,\ y_2 = s,$$
$$y_3 = t, \quad t, s \in \mathbb{R}.$$

We obtain a general vector of the intersection by substituting

$$\begin{pmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 \\ x_1 + x_3 \\ x_2 + x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -2t - 2s \\ 2s \\ 2t \end{pmatrix}.$$

We see that

$$\dim U \cap V = 2, \quad U \cap V = \operatorname{span}\left\{ \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

□

2.D.13. Let there be in $\mathbb{R}^3$ two vector spaces U and V generated by the vectors

$$(1, 1, -3)\,, (1, 2, 2) \quad \text{and} \quad (1, 1, -1)\,, (1, 2, 1)\,, (1, 3, 3)\,,$$

respectively. Determine the intersection of these two subspaces.

Solution. According to the definition of intersection, the vectors in the intersection are in both, the span of the vectors $(1, 1, -3)\,, (1, 2, 2)$, as well as in the span of the vectors $(1, 1, -1)\,, (1, 2, 1)\,, (1, 3, 3)$. It helps to consider first the geometry. Firstly, U is spanned by two linearly independent vectors. So U is a plane in $\mathbb{R}^3$. Next, V is spanned by three vectors. But these are linearly dependent since

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ -1 & 1 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{vmatrix} = 0.$$

So V is also a plane.

then it would mean that the vectors $u_1, \dots, u_{\ell+1}$ were linearly dependent, which is a contradiction.

Summarizing, for every $k \le n$ we can arrive after k steps at a basis in which k vectors from the original basis were exchanged for the new u_i's. If $k > n$, then in the n-th step we would obtain a basis consisting only of new vectors u_i, which means that the original set could not be linearly independent. In particular, it is not possible for two bases to have a different number of elements. □

In fact, we have proved a stronger claim, the *Steinitz exchange lemma*, which says that for every finite basis $\underline{v}$ and every system of linearly independent vectors u_i in V we can find a subset of the basis vectors v_i which will complete the set of u_i's into a new basis.

2.3.9. Corollaries of the Steinitz lemma. Because of the possibility of freely choosing and replacing basis vectors we can immediately derive nice (and intuitively expectable) properties of bases of vector spaces:

Proposition. *(1) Every two bases of a finite dimensional vector space have the same number of elements, that is, our definition of dimension is basis-independent.*
(2) If V has a finite basis, then every linearly independent set can be extended to a basis.
(3) A basis of a finite dimensional vector space is a maximal linearly independent set of vectors.
(4) The bases of a vector space are the minimal sets of generators.

A little more complicated, but now easy to deal with, is the situation of dimensions of subspaces and their sums:

Corollary. *Let $W, W_1, W_2 \subset V$ be subspaces of a space V of finite dimension. Then*
(1) $\dim W \le \dim V$,
(2) $V = W$ if and only if $\dim V = \dim W$,
(3) $\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2)$.

Proof. It remains to prove only the last claim. This is evident if the dimension of one of the spaces is zero. Assume $\dim W_1 = r \ge 1$, $\dim W_2 = s \ge 1$ and let $(w_1 \dots, w_t)$ be a basis of $W_1 \cap W_2$ (or empty set, if the intersection is trivial).

According to the Steinitz exchange lemma this basis of the intersection can be extended to a basis $(w_1, \dots, w_t, u_{t+1} \dots, u_r)$ for W_1 and to a basis $(w_1 \dots, w_t, v_{t+1}, \dots, v_s)$ for W_2. Vectors

$$w_1, \dots, w_t, u_{t+1}, \dots, u_r, v_{t+1} \dots, v_s$$

clearly generate $W_1 + W_2$. We show that they are linearly independent. Let

$$a_1 w_1 + \dots + a_t w_t + b_{t+1} u_{t+1} + \dots$$
$$\dots + b_r u_r + c_{t+1} v_{t+1} + \dots + c_s v_s = 0.$$

If the vector (x_1, x_2, x_3) lies in U, then $(x_1, x_2, x_3) = \lambda(1, 1, -3) + \mu(1, 2, 2)$ for some scalars λ, μ. Similarly (x_1, x_2, x_3) lies in V, so $(x_1, x_2, x_3) = \alpha(1, 1, -1) + \beta(1, 2, 1) + \gamma(1, 3, 3)$ for scalars α, β, γ. When written in full, this is a set of six equations in eight unknowns. Solving these is possible but can be quite cumbersome. Some simplification is obtained as follows:

The first three equations, which describe U are

$$x_1 = \lambda + \mu$$

$$x_2 = \lambda + 2\mu$$

$$x_3 = -3\lambda + 2\mu$$

If we solve these three equations for the two "unknowns" λ and μ, (which in any case we do not want), or alternatively if we eliminate λ and μ, from these equations, we obtain the single equation $8x_1 - 5x_2 + x_3 = 0$ to replace the first three.

The second set of three equations, which describe V are

$$x_1 = \alpha + \beta + \gamma$$

$$x_2 = \alpha + 2\beta + \gamma$$

$$x_3 = -\alpha + \beta + 3\gamma$$

If we solve these three equations for the three "unknowns" α β and γ, (which in any case we do not want), or alternatively if we eliminate α β and γ, from these equations, we obtain the single equation $3x_1 - 2x_2 + x_3 = 0$ to describe V. Introducing the parameter t, it is straightforward write the solution as the line $(x_1, x_2, x_3) = t(3, 5, 1)$.

□

2.D.14. Determine the vector subspace (of the space $\mathbb{R}^4$) generated by the vectors $u_1 = (-1, 3, -2, 1)$, $u_2 = (2, -1, -1, 2)$, $u_3 = (-4, 7, -3, 0)$, $u_4 = (1, 5, -5, 4)$, by choosing a maximal set of linearly independent vectors u_i (that is, by choosing a basis).

Solution. Write the vectors u_i into the columns of a matrix and transform it using elementary row transformations. This way we obtain

Then necessarily

$$- (c_{t+1} \cdot v_{t+1} + \cdots + c_s \cdot v_s) =$$
$$= a_1 \cdot w_1 + \cdots + a_t \cdot w_t + b_{t+1} \cdot u_{t+1} + \cdots + b_r \cdot u_r$$

must belong to $W_2 \cap W_1$. This implies that

$$b_{t+1} = \cdots = b_r = 0,$$

since this is the way we have defined our bases. Then also

$$a_1 \cdot w_1 + \cdots + a_t \cdot w_t + c_{t+1} \cdot v_{t+1} + \cdots + c_s \cdot v_s = 0$$

and because the corresponding vectors form a basis W_2, all the coefficients are zero.

The claim (3) now follows by directly counting the generators. □

2.3.10. Examples. (1) $\mathbb{K}^n$ has (as a vector space over $\mathbb{K}$) dimension n. The n-tuple of vectors

$$((1, 0, \dots, 0), (0, 1, \dots, 0) \dots, (0, \dots, 0, 1))$$

is clearly a basis, we call it the *standard basis of* $\mathbb{K}^n$.

Note that in the case of a finite field of scalars, say $\mathbb{Z}_k$, the whole space $\mathbb{K}^n$ has only a finite number k^n of elements.

(2) $\mathbb{C}$ as a vector space over $\mathbb{R}$ has dimension 2. A basis is for instance the pair of numbers 1 and i, or any other two complex numbers which are not a real multiple of each other, eg. $1 + i$ and $1 - i$.

(3) $\mathbb{K}_m[x]$, that is, the space of all polynomials of degree at most m, has dimension $m + 1$. A basis is for instance the sequence $1, x, x^2, \dots, x^m$.

The vector space of all polynomials $\mathbb{K}[x]$ has dimension ∞, but we can still find a basis (although infinite in size): $1, x, x^2, \dots$.

(4) The vector space $\mathbb{R}$ over $\mathbb{Q}$ has dimension ∞. It does not have a countable basis.

(5) The vector space of all mappings $f : \mathbb{R} \to \mathbb{R}$ has also dimension ∞. It does not have any countable basis.

2.3.11. Vector coordinates. If we fix a basis $(v_1, \dots, v_n)$ of a finite dimensional space V, then every vector $w \in V$ can be expressed as a linear combination $v = a_1 v_1 + \cdots + a_n v_n$ in a unique way. Indeed, assume that we can do it in two ways:

$$w = a_1 v_1 + \cdots + a_n v_n = b_1 v_1 + \cdots + b_n v_n.$$

Then

$$0 = (a_1 - b_1) \cdot v_1 + \cdots + (a_n - b_n) \cdot v_n$$

and thus $a_i = b_i$ for all $i = 1, \dots, n$, because the vectors v_i are linearly independent. We have reached the concept of coordinates:

Definition. The coefficients of the unique linear combination expressing the given vector $w \in V$ in the chosen basis $\underline{v} = (v_1, \dots, v_n)$ are called the *coordinates of the vector* w in this basis.

$$\begin{pmatrix} -1 & 2 & -4 & 1 \\ 3 & -1 & 7 & 5 \\ -2 & -1 & -3 & -5 \\ 1 & 2 & 0 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 & 4 \\ -1 & 2 & -4 & 1 \\ 3 & -1 & 7 & 5 \\ -2 & -1 & -3 & -5 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 4 & -4 & 5 \\ 0 & -7 & 7 & -7 \\ 0 & 3 & -3 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 5/4 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 5/4 \\ 0 & 0 & 0 & -1/4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

From this it follows that the vectors u_1, u_2, u_4, are linearly independent.

Furthermore, (see the third column)

$$2 \cdot (-1, 3, -2, 1) - (2, -1, -1, 2) = (-4, 7, -3, 0).$$

so that $u_3 = 2u_1 - u_2$, and hence u_1, u_2, and u_4, form a basis for the subspace. □

2.D.15. Find a basis of the subspace

$$U = \operatorname{span}\left\{ \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} -2 & -1 \\ 0 & 1 \\ 2 & 3 \end{pmatrix} \right\}$$

of the vector space of real matrices 3×2. Extend this basis to a basis of the whole space.

Solution. Recall that a basis of a subspace is a set of linearly independent vectors which generate given subspace. By writing the entries of the matrices in a row, we can consider the matrices as vectors in R^6. In this way, the four given matrices can be identified with the rows of the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 \\ -1 & 0 & 1 & 2 & 3 & 4 \\ -2 & -1 & 0 & 1 & 2 & 3 \end{pmatrix}.$$

It is easy to show that this matrix has rank 2, and hence that the subspace U is generated just by the first two matrices, which consequently form a basis for U. In fact, it follows easily that

$$\begin{pmatrix} -1 & 0 \\ 1 & 2 \\ 3 & 4 \end{pmatrix} = -1 \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{pmatrix}$$
$$\begin{pmatrix} -2 & -1 \\ 0 & 1 \\ 2 & 3 \end{pmatrix} = -2 \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{pmatrix}.$$

There are many options for extending this basis to be a basis for the whole space. One option is to choose the first two of

Whenever we speak about coordinates $(a_1, \dots, a_n)$ of a vector w, which we express as a sequence, we must have a fixed ordering of the basis vectors $\underline{v} = (v_1, \dots, v_n)$. Although we have defined the basis as a minimal set of generators, in reality we work with them as with sequences (that is, with ordered sets).

Assigning coordinates to vectors

A mapping assigning the vector $v = a_1 v_1 + \dots + a_n v_n$ to its coordinates in the basis $\underline{v}$ will be denoted by the same symbol $\underline{v} : V \to \mathbb{K}^n$. It has the following properties:

(1) $\underline{v}(u + w) = \underline{v}(u) + \underline{v}(w);\ \forall u, w \in V$,
(2) $\underline{v}(a \cdot u) = a \cdot \underline{v}(u);\ \forall a \in \mathbb{K}, \forall u \in V$.

Note that the operations on the two sides of these equations are not identical. Quite the opposite; they are operations on different vector spaces! Sometimes it is useful to understand vectors as mappings from fixed independent generators to coordinates. In this way, we may think about the basis M of infinite dimensional vector spaces V. Even though the set M will be infinite, there can be only a finite number of non-zero values for any mapping representing a vector. The vector space of all polynomials $\mathbb{K}_\infty[x]$, with the basis $M = \{1, x, x^2, \dots\}$ is a good example.

2.3.12. Linear mappings. The above properties of the assignments of coordinates are typical for what we have called linear mappings in the geometry of the plane $\mathbb{R}^2$.

For any vector space (of finite or infinite dimension) we define "linearity" of a mapping between spaces in a similar way to the case of the plane $\mathbb{R}^2$:

Linear mappings

Let V and W be vector spaces over the same field of scalars $\mathbb{K}$. The mapping $f : V \to W$ is called a *linear mapping*, or *homomorphism*, if the following holds:

(1) $f(u + v) = f(u) + f(v),\ \forall u, v \in V$
(2) $f(a \cdot u) = a \cdot f(u),\ \forall a \in \mathbb{K},\ \forall u \in V$.

We have seen such mappings already in the case of matrix multiplication:

$$f : \mathbb{K}^n \to \mathbb{K}^m, \quad x \mapsto A \cdot x$$

with a fixed matrix A of the type m/n over $\mathbb{K}$.

The *image* of a linear mapping, $\operatorname{Im} f = f(V) \subset W$, is always a vector subspace, since for any set of vectors u_i, the linear combination of images $f(u_i)$ is the image of the linear combination of the vectors u_i with the same coefficients.

Analogously, the set of all vectors $\operatorname{Ker} f = f^{-1}(\{0\}) \subset V$ is a subspace, since the linear combination of zero images will always be a zero vector. The subspace $\operatorname{Ker} f$ is called the *kernel of the linear mapping f*.

A linear mapping which is a bijection is called an *isomorphism*.

the given matrices together with the last four (actually, any four would do) of the six linearly independent matrices

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

. Linear independence of these six matrices is established by computing

$$\begin{vmatrix}1&2&3&4&5&6\\0&1&2&3&4&5\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{vmatrix}=1\neq 0.$$

Clearly the dimension is 6, so spanning is automatic, and hence we have a basis. □

E. Linear mappings

How can we describe simple mappings analytically? For example,how can we describe a rotation, an axial symmetry, a mirror symmetry, a projection of a three-dimensional space onto a two-dimensional one in the plane or in the space? How can we describe the scaling of a diagram? What do they have in common? These all are linear mappings. This means that they preserve a certain structure of the space or a subspace. What structure? The structure of a vector space. Every point in the plane is described by two coordinates, every point in the 3-dimensional space is described by three coordinates. If we fix the origin, then it makes sense to say that a point is in some direction twice that far from the origin as some other point. We also know where arrive at if we translate or shift by some amount in a given direction and then by some other amount in another direction. These properties can be formalized – we speak of vectors in the plane or in space, and we consider their multiplication and addition. Linear mappings have the property that the image of a sum of vectors is a sum of the images of the vectors. The image of a multiple of a vector is the same multiple as the image of the vector. These properties are shared among the mappings stated at the beginning of this paragraph. Such a mapping is then uniquely determined by its behaviour on the vectors of a basis. (In the plane, a basis consists of two vectors not on the same line. In space a basis consists of three vectors not all in the same plane).

How can we write down some linear mapping f on a vector space V? For simplicity, we start with the plane $\mathbb{R}^2$. Assume that the image of the point (vector) $(1,0)$ is (a,b) and the image of the point (vector) $(0,1)$ is (c,d). This

Analogously to the abstract definition of vector spaces, it is again necessary to prove seemingly trivial claims that follow from the axioms:

Proposition. *Let $f : V \to W$ be a linear mapping between two vector spaces over the same field of scalars $\mathbb{K}$. The following is true for all vectors $u, u_1, \dots, u_k \in V$ and scalars $a_1, \dots, a_k \in \mathbb{K}$*

(1) $f(0) = 0$,
(2) $f(-u) = -f(u)$,
(3) $f(a_1 \cdot u_1 + \dots + a_k \cdot u_k) = a_1 \cdot f(u_1) + \dots + a_k \cdot f(u_k)$,
(4) for every vector subspace $V_1 \subset V$, its image $f(V_1)$ is a vector subspace in W,
(5) for every vector subspace $W_1 \subset W$, the set $f^{-1}(W_1) = \{v \in V;\ f(v) \in W_1\}$ is a vector subspace in V.

Proof. We rely on the axioms, definitions and already proved results (in case you are not sure what has been used, look it up!):

$$f(0) = f(u-u) = f((1-1)\cdot u) = 0 \cdot f(u) = 0,$$
$$f(-u) = f((-1)\cdot u) = (-1)\cdot f(u) = -f(u).$$

Property (3) is derived easily from the definition for two summands, using induction on the number of summands.

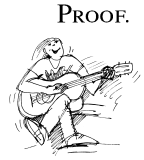
Next, (3) implies $\operatorname{span} f(V_1) = f(V_1)$, thus it is a vector subspace. On the other hand, if $f(u) \in W_1$ and $f(v) \in W_1$ then for any scalars we arrive at $f(a\cdot u + b\cdot v) = a\cdot f(u) + b\cdot f(v) \in W_1$. □

2.3.13. Proposition (Simple corollaries)**.** *(1) The composition $g \circ f : V \to Z$ of two linear mappings $f : V \to W$ and $g : W \to Z$ is again a linear mapping.*
(2) The linear mapping $f : V \to W$ is an isomorphism if and only if $\operatorname{Im} f = W$ and $\operatorname{Ker} f = \{0\} \subset V$. The inverse mapping of an isomorphism is again an isomorphism.
(3) For any two subspaces $V_1, V_2 \subset V$ and linear mapping $f : V \to W$,

$$f(V_1 + V_2) = f(V_1) + f(V_2),$$
$$f(V_1 \cap V_2) \subset f(V_1) \cap f(V_2).$$

(4) The "coordinate assignment" mapping $\underline{u} : V \to \mathbb{K}^n$ given by an arbitrarily chosen basis $\underline{u} = (u_1, \dots, u_n)$ of a vector space V is an isomorphism.
(5) Two finitely dimensional vector spaces are isomorphic if and only if they have the same dimension.
(6) The composition of two isomorphisms is an isomorphism.

Proof. Proving the first claim is a very easy exercise left to the reader. In order to verify (2), notice that f is surjective if and only if $\operatorname{Im} f = W$. If $\operatorname{Ker} f = \{0\}$ then $f(u) = f(v)$ ensures $f(u-v) = 0$, that is, $u = v$. In this case f is injective. Finally, if f is a linear bijection, then the vector w is the preimage of a linear combination $au + bv$, that is $w = f^{-1}(au + bv)$, if and only if



uniquely determines the image of an arbitrary point with coordinates (u, v): $f((u, v)) = f(u(1, 0)+v(0, 1)) = uf(1, 0)+ vf(1, 0) = (ua, ub) + (vc, vd) = (au + cv, bu + dv)$. This can be written down more efficiently as follows:

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au + cv \\ bu + dv \end{pmatrix}$$

A linear mapping is thus a mapping uniquely determined (in a fixed basis) by a matrix. Furthermore, when we have another linear mapping g given by the matrix $\begin{pmatrix} e & f \\ g & h \end{pmatrix}$, then we can easily compute (an interested reader can fill in the details by himself) that their composition $g \circ f$ is given by the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + fc & be + df \\ ag + ch & bg + dh \end{pmatrix}.$$

This leads us to the definition of matrix multiplication in exactly this way. That is, an application of a mapping on a vector is given by the matrix multiplication of the matrix of the mapping with the given vector, and that the mapping of a composition is given by the product of the corresponding matrices. This works analogously in the spaces of higher dimension. Further, this again shows what has already been proven in (2.1.5), namely, that matrix multiplication is associative but not commutative, just as with mapping composition. That is another motivation to study vector spaces.

Recall that already in the first chapter we worked with the matrices of some linear mappings in the plane $\mathbb{R}^2$, notably with the rotation around a point and with axial symmetry (see **??** and **??**).

We try now to write down matrices of linear mappings from $\mathbb{R}^3$ to $\mathbb{R}^3$. What does the matrix of a rotation in three dimensions look like? We begin with some special (easier for description) rotations about coordinate axes:

2.E.1. Matrix of rotation about coordinate axes in $\mathbb{R}^3$. We write down the matrices of rotations by the angle φ, about the (oriented) axes x, y and z in $\mathbb{R}^3$.

Solution. When rotating a particular point about the given axis (say x), the corresponding coordinate (x) does not change. The remaining two coordinates are then given by the rotation in the plane which we already know (a matrix of the type 2×2).

Thus we obtain the following matrices – rotation about the axis z:

$$\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$f(w) = au + bv = f(a \cdot f^{-1}(u) + b \cdot f^{-1}(v)).$$

Thus we also get $w = af^{-1}(u) + bf^{-1}(v)$ and therefore the inversion of a linear bijection is again a linear bijection.

The third property is obvious from the definition, but try finding an example showing that the inequality in the second equation can indeed by sharp.

The remaining claims all follow immediately from the definition. $\square$

2.3.14. Coordinates again. Consider any two vector spaces

V and W over $\mathbb{K}$ with $\dim V = n$, $\dim W = m$ and consider some linear mapping $f : V \to W$. For every choice of basis $\underline{u} = (u_1, \dots, u_n)$ on V, $\underline{v} = (v_1, \dots, v_n)$ on W there are the following linear mappings as shown in the diagram:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\underline{u}"', "\simeq"] & W \arrow[d, "\simeq"', "\underline{v}"] \\ \mathbb{K}^n \arrow[r, dotted, "f_{\underline{u},\underline{v}}"] & \mathbb{K}^m \end{tikzcd}$$

The bottom arrow $f_{\underline{u},\underline{v}}$ is defined by the remaining three, i.e. the composition of linear mappings

$$f_{\underline{u},\underline{v}} = \underline{v} \circ f \circ \underline{u}^{-1}.$$

MATRIX OF A LINEAR MAPPING

Every linear mapping is uniquely determined by its values on an arbitrary set of generators, in particular, on the vectors of a basis $\underline{u}$. Denote by

$$\begin{aligned} f(u_1) &= a_{11} \cdot v_1 + a_{21} \cdot v_2 + \dots + a_{m1} v_m \\ f(u_2) &= a_{12} \cdot v_1 + a_{22} \cdot v_2 + \dots + a_{m2} v_m \\ &\vdots \\ f(u_n) &= a_{1n} \cdot v_1 + a_{2n} \cdot v_2 + \dots + a_{mn} v_m, \end{aligned}$$

that is, scalars a_{ij} form a matrix A, where the columns are coordinates of the values $f(u_j)$ of the mapping f on the basis vectors expressed in the basis $\underline{v}$ on the target space W.

A matrix $A = (a_{ij})$ is called the *matrix of the mapping* f in the bases $\underline{u}$, $\underline{v}$.

For a general vector $u = x_1 u_1 + \dots + x_n u_n \in V$ we calculate (recall that vector addition is commutative and distributive with respect to scalar multiplication)

$$\begin{aligned} f(u) &= x_1 f(u_1) + \dots + x_n f(u_n) \\ &= x_1(a_{11}v_1 + \dots + a_{m1}v_m) + \dots + x_n(a_{1n}v_1 + \dots) \\ &= (x_1 a_{11} + \dots + x_n a_{1n}) v_1 + \dots + (x_1 a_{m1} + \dots) v_m. \end{aligned}$$

Using matrix multiplication we can now very easily and clearly write down the values of the mapping $f_{\underline{u},\underline{v}}(w)$ defined uniquely by the previous diagram. Recall that vectors in $\mathbb{K}^\ell$ are understood as columns, that is, matrices of the type $\ell/1$

$$f_{\underline{u},\underline{v}}(\underline{u}(w)) = \underline{v}(f(w)) = A \cdot \underline{u}(w).$$

rotation about the axis x:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

rotation about the axis y:

$$\begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix}$$

Note the sign of φ in the matrix for rotation about y. We want, as with any other rotation, the rotation about the y axis to be in the positive sense — that is, when we look in the opposite direction of the direction of the y axis, the world turns anticlockwise. The signs in the matrices depend on the orientation of our coordinate system. Usually, in the 3-dimensional space the "dextrorotary coordinate system" is chosen: if we place our hand on the x axis such that the fingers point in the direction of the axis and such that we can rotate the x axis in the xy plane so that x coincides with the y axis and they point in the same direction, then the thumb should point in the direction of the z axis. In such a system, this is a rotation in the negative sense in the plane xz (that is, the axis z turns in the direction towards x). Think about the positive and negative sense of rotations by all three axes. The sign is also consistent with the cycle x to y to z to x to y etc.... or 1 to 2 to 3 to 1 to..... etc. □

Knowledge of matrices allows us to write the matrix of rotation about any oriented axis. Let us start with a specific example:

2.E.2. Find the matrix of the rotation in the positive sense by the angle $\pi/3$ about the line passing through the origin with the oriented directional vector $(1,1,0)$ under the standard basis $\mathbb{R}^3$.

Solution. The given rotation is easily obtained by composing these three mappings:

- rotation through the angle $\pi/4$ in the negative sense about the axis z (the axis of the rotation goes over on the x axis);
- rotation through the angle $\pi/3$ in the positive sense about the x axis;
- rotation through the angle $\pi/4$ in the positive sense about the z axis (the x axis goes over on the axis of the rotation).

The matrix of the resulting rotation is the product of the matrices corresponding to the given three mappings, while the order of the matrices is given by the order of application of

On the other hand, if we have fixed bases on V and W, then every choice of a matrix A of the type m/n gives a unique linear mapping $\mathbb{K}^n \to \mathbb{K}^m$ and thus also a mapping $f : V \to W$. We have found the bijective correspondence between matrices of the fixed types (determined by dimensions of V and W) and linear mappings $V \to W$.

2.3.15. Coordinate transition matrix. If we choose $V = W$ to be the same space, but with two different bases $\underline{u}$, $\underline{v}$, and consider the identity mapping for f, then the approach from the previous paragraph expresses the vectors of the basis $\underline{u}$ in coordinates with respect to the basis $\underline{v}$. Let the resulting matrix be T.

Thus, we are just applying the concept of the matrix of a linear mapping to the special case of the identity mapping id_V.

$$\begin{tikzcd} V \arrow[r, "\mathrm{id}_V"] \arrow[d, "\underline{u}"', "\simeq"] & V \arrow[d, "\simeq"', "\underline{v}"] \\ \mathbb{K}^n \arrow[r, dotted, "T=(\mathrm{id}_V)_{\underline{u},\underline{v}}"] & \mathbb{K}^n \end{tikzcd}$$

The resulting matrix T is called the *coordinate transition matrix* for changing the basis from $\underline{u}$ to the basis $\underline{v}$.

The fact that the matrix T of the identity mapping yields exactly the transformation of coordinates between the two bases is easily seen.

Consider the expression of u with the basis $\underline{u}$

$$u = x_1u_1 + \cdots + x_nu_n,$$

and replace the vectors u_i by their expressions as linear combinations of the vectors v_i in the basis $\underline{v}$. Collecting the terms properly, we obtain the coordinate expression $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$ of the same vector u in the basis $\underline{v}$. It is enough just to reorder the summands and express the individual scalars at the vectors of the basis. But this is exactly what we do when forming the matrix for the identity mapping, thus $\bar{x} = T \cdot x$.

We have arrived at the following instruction for building the coordinate transition matrix:

Calculating the matrix for changing the basis

Proposition. *The matrix T for the transition from the basis $\underline{u}$ to the basis $\underline{v}$ is obtained by taking the coordinates of the vectors of the basis $\underline{u}$ expressed in the basis $\underline{v}$ and writing them as the columns of the matrix T. The new coordinates $\bar{x}$ in terms of the new basis $\underline{v}$ are then $\bar{x} = T \cdot x$, where x is the coordinate vector in the original basis $\underline{u}$.*

Because the inverse mapping to the identity mapping is again the identity mapping, the coordinate transition matrix is always invertible and its inverse T^{-1} is the coordinate transition matrix in the opposite direction, that is from the basis $\underline{v}$ to the basis $\underline{u}$ (just have a look at the diagram above and invert all the arrows).

the mappings – the first mapping applied is in the product the rightmost one. Thus we obtain the desired matrix

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{\sqrt{6}}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{\sqrt{6}}{4} \\ -\frac{\sqrt{6}}{4} & \frac{\sqrt{6}}{4} & \frac{1}{2} \end{pmatrix}$$

Note that the resulting rotation could be also obtained for instance by taking the composition of the three following mappings:

- rotation through the angle $\pi/4$ in the positive sense about the axis z (the axis of rotation goes over on the axis y);
- rotation through the angle $\pi/3$ in the positive sense about the axis y;
- rotation through the angle $\pi/4$ in the negative sense about the axis z (the axis y goes over to the axis of rotation).

Analogously we obtain

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{\sqrt{6}}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{\sqrt{6}}{4} \\ -\frac{\sqrt{6}}{4} & \frac{\sqrt{6}}{4} & \frac{1}{2} \end{pmatrix}$$

□

2.E.3. Matrix of general rotation in $\mathbb{R}^3$. Derive the matrix of a general rotation in $\mathbb{R}^3$.

Solution. We can do the same things as in the previous example with general values. Consider an arbitrary unit vector (x, y, z). Rotation in the positive sense by the angle φ about this vector can be written down as a composition of the following rotations whose matrices we already know:

i) rotation $\mathcal{R}_1$ in the negative sense about the z axis through the angle with cosine equal to $x/\sqrt{x^2+y^2} = x/\sqrt{1-z^2}$, that is, with sine $y/\sqrt{1-z^2}$, under which the line with the directional vector (x, y, z) goes over on the line with the directional vector $(0, y, z)$. The matrix of this rotation is

$$R_1 = \begin{pmatrix} x/\sqrt{1-z^2} & y/\sqrt{1-z^2} & 0 \\ -y/\sqrt{1-z^2} & x/\sqrt{1-z^2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

2.3.16. More coordinates. Next, we are interested in the matrix of a composition of the linear mappings. Thus, consider another vector space Z over $\mathbb{K}$ of dimension k with basis $\underline{w}$, linear mapping $g : W \to Z$ and denote the corresponding matrix by $g_{\underline{v},\underline{w}}$.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\underline{u}"', "\simeq"] & W \arrow[r, "g"] \arrow[d, "\simeq"', "\underline{v}"] & Z \arrow[d, "\simeq"', "\underline{w}"] \\ \mathbb{K}^n \arrow[r, dotted, "f_{\underline{u},\underline{v}}"] & \mathbb{K}^m \arrow[r, dotted, "g_{\underline{v},\underline{w}}"] & \mathbb{K}^k \end{tikzcd}$$

The composition $g \circ f$ on the upper row corresponds to the matrix of the mapping $\mathbb{K}^n \to \mathbb{K}^k$ on the bottom and we calculate directly (we write A for the matrix of f and B for the matrix of g in the chosen bases):

$$\begin{aligned} g_{\underline{v},\underline{w}} \circ f_{\underline{u},\underline{v}}(x) &= \underline{w} \circ g \circ \underline{v}^{-1} \circ \underline{v} \circ f \circ \underline{u}^{-1} \\ &= B \cdot (A \cdot x) = (B \cdot A) \cdot x = (g \circ f)_{\underline{u},\underline{w}}(x) \end{aligned}$$

for every $x \in \mathbb{K}^n$. By the associativity of matrix multiplications, the composition of mappings corresponds to multiplication of the corresponding matrices. Note that the isomorphisms correspond exactly to invertible matrices and that the matrix of the inverse mapping is the inverse matrix.

The same approach shows how the matrix of a linear mapping changes, if we change the coordinates on both the domain and the codomain:

$$\begin{tikzcd} V \arrow[r, "\mathrm{id}_V"] \arrow[d, "\underline{u}'"', "\simeq"] & V \arrow[r, "f"] \arrow[d, "\underline{u}"', "\simeq"] & W \arrow[r, "\mathrm{id}_W"] \arrow[d, "\simeq"', "\underline{v}"] & W \arrow[d, "\simeq"', "\underline{v}'"] \\ \mathbb{K}^n \arrow[r, dotted, "T"] & \mathbb{K}^n \arrow[r, dotted, "f_{\underline{u},\underline{v}}"] & \mathbb{K}^m \arrow[r, dotted, "S^{-1}"] & \mathbb{K}^m \end{tikzcd}$$

where T is the coordinate transition matrix from $\underline{u}'$ to $\underline{u}$ and S is the coordinate change matrix from $\underline{v}'$ to $\underline{v}$. If A is the original matrix of the mapping, then the matrix of the new mapping is given by $A' = S^{-1} AT$.

In the special case of a linear mapping $f : V \to V$, that is the domain and the codomain are the same space V, we express f usually in terms of a single basis $\underline{u}$ of the space V. Then the change from the old basis to the new basis $\underline{u}'$ with the coordinate transition matrix T leads to the new matrix $A' = T^{-1}AT$.

2.3.17. Linear forms. A simple but very important case of linear mappings on an arbitrary vector space V over the scalars $\mathbb{K}$ appears with the codomain being the scalars themselves, i.e. mappings $f : V \to \mathbb{K}$. We call them *linear forms*.

If we are given the coordinates on V, the assignments of a single i-th coordinate to the vectors is an example of a linear form. More precisely, for every choice of basis $\underline{v} = (v_1, \dots, v_n)$, there are the linear forms $v_i^* : V \to \mathbb{K}$ such that $v_i^*(v_j) = \delta_{ij}$, that is, $v_i^*(v_j) = 1$ when $i = j$, and $v_i^*(v_j) = 0$ when $i \neq j$.

The vector space of all linear forms on V is denoted by V^* and we call it the *dual space* of the vector space V. Let us now assume that the vector space V has finite dimension

ii) rotation $\mathcal{R}_2$ in the positive sense about the y axis through the angle with cosine $\sqrt{1-z^2}$, that is, with sine z, under which the line with the directional vector $(0, y, z)$ goes over on the line with the directional vector $(1, 0, 0)$. The matrix of this rotation is

$$R_2 = \begin{pmatrix} \sqrt{1-z^2} & 0 & z \\ 0 & 1 & 0 \\ -z & 0 & \sqrt{1-z^2} \end{pmatrix},$$

iii) rotation $\mathcal{R}_3$ in the positive sense about the x axis through the angle φ with the matrix

$$R_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi) & \cos(\varphi) \end{pmatrix},$$

iv) rotation $\mathcal{R}_2^{-1}$ with the matrix R_2^{-1},

v) rotation $\mathcal{R}_1^{-1}$ with the matrix R_1^{-1}.

The matrix of the composition of these mappings, that is, the matrix we are looking for, is given by the product of the rotations in the reverse order:

$$R_1^{-1} \cdot R_2^{-1} \cdot R_3 \cdot R_2 \cdot R_1 =$$

$$= \begin{pmatrix} 1-t+tx^2 & txy-zs & txz+ys \\ yxt+zs & 1-t+ty^2 & tyz-xs \\ zxt-ys & tzy+xs & 1-t+tz^2 \end{pmatrix},$$

where $t = 1 - \cos\varphi$ and $s = \sin\varphi$.

□

2.E.4. We are given a linear mapping $\mathbb{R}^3 \to \mathbb{R}^3$ in the standard basis as the following matrix:

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

Write down the matrix of this mapping in the basis

$(f_1, f_2, f_3) = ((1, 1, 0), (-1, 1, 1), (2, 0, 1))$.

Solution. The transition matrix T for changing the basis from the basis $\underline{f} = (f_1, f_2, f_3)$ to the standard basis, that is, to the basis given by the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, can be obtained, according to the Claim 2.25, by writing down the coordinates of the vectors f_1, f_2, f_3 in the standard basis as the columns of the matrix T. Thus we have

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

n. The basis of V^*, $\underline{v}^* = (v_1^*, \dots, v_n^*)$, composed of assignments of individual coordinates as above, is called the *dual basis* to $\underline{v}$. Clearly this is a basis of the space V^*, because these forms are evidently linearly independent (prove it!) and if $\alpha \in V^*$ is an arbitrary form, then for every vector $u = x_1 v_1 + \cdots + x_n v_n$

$$\begin{aligned} \alpha(u) &= x_1\alpha(v_1) + \cdots + x_n\alpha(v_n) \\ &= \alpha(v_1)v_1^*(u) + \cdots + \alpha(v_n)v_n^*(u) \end{aligned}$$

and thus the linear form α is a linear combination of the forms v_i^*.

Taking into account the standard basis $\{1\}$ on the one-dimensional space of scalars $\mathbb{K}$, any choice of a basis $\underline{v}$ on V identifies the linear forms α with matrices of the type $1/n$, that is, with rows y. The components of these rows are coordinates of the general linear forms α in the dual basis $\underline{v}^*$. Expressing such a form on a vector is then given by multiplying the corresponding row vector y with the column of the coordinates x of the vector $u \in V$ in the basis $\underline{v}$:

$$\alpha(u) = y \cdot x = y_1 x_1 + \cdots + y_n x_n.$$

Thus we can see that for every finitely dimensional space V, the dual space V^* is isomorphic to the space V. The choice of the dual basis provides such an isomorphism.

In this context we meet again the scalar product of a row of n scalars with a column of n scalars. We have worked with it already in the paragraph 2.1.3 on the page 7.

The situation is different for infinitely dimensional

spaces. For instance the simplest example of the space of all polynomials $\mathbb{K}[x]$ in one variable is a vector space with a countable basis with elements $v_i = x^i$. As before, we can define linearly independent forms v_i^*. Every formal infinite sum $\sum_{i=0}^{\infty} a_i v_i^*$ is now a well-defined linear form on $\mathbb{K}[x]$, because it will be evaluated only for a finite linear combination of the basis polynomials x^i, $i = 0, 1, 2, \dots$.

The countable set of all v_i^* is thus not a basis. Actually, it can be proved that this dual space cannot have a countable basis.

2.3.18. The length of vectors and scalar product. When

dealing with the geometry of the plane $\mathbb{R}^2$ in the first chapter we also needed the concept of the length of vectors and their angles, see **??**. of vectors and their angles. For defining these concepts we used the scalar product of two vectors $v = (x, y)$ and $v' = (x', y')$ in the form $u \cdot v = xx' + yy'$. Indeed, the expression for the length of $v = (x, y)$ is given by

$$\|v\| = \sqrt{x^2 + y^2} = \sqrt{v \cdot v},$$

while the (oriented) angle φ of two vectors $v = (x, y)$ and $v' = (x', y')$ is in the planar geometry given by the formula

$$\cos\varphi = \frac{xx' + yy'}{\|v\|\|v'\|}.$$

The transition matrix for changing the basis from the standard basis to the basis $\underline{f}$ is then given by

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

The matrix of the mapping in the basis $\underline{f}$ is then given by

$$T^{-1}AT = \begin{pmatrix} \frac{1}{4} & 2 & -\frac{3}{4} \\ \frac{5}{4} & 0 & \frac{7}{4} \\ \frac{3}{4} & -2 & \frac{9}{4} \end{pmatrix}.$$

□

2.E.5. Consider the vector space of polynomials of one variable of degree at most 2 with real coefficients. In this space, consider the basis 1, x, x^2. Write down the matrix of the derivative mapping in this basis and also in the basis $1+x^2$, x, $x+x^2$.

Solution. $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 3 \\ 0 & -1 & -1 \end{pmatrix}$. □

2.E.6. In the standard basis in $\mathbb{R}^3$, determine the matrix of the rotation through the angle $90°$ in the positive sense about the line (t,t,t), $t \in \mathbb{R}$, oriented in the direction of the vector $(1,1,1)$. Further, find the matrix of this rotation in the basis $\underline{g} = ((1,1,0),(1,0,-1),(0,1,1))$.

Solution. We can easily determine the matrix of the given rotation in a suitable basis, that is, in a basis given by the directional vector of the line and by two mutually perpendicular vectors in the plane $x+y+z=0$, that is, in the plane of vectors perpendicular to the vector $(1,1,1)$. We note that the matrix of the rotation in the positive sense through $90°$ in an orthonormal basis in $\mathbb{R}^2$ is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. In the orthogonal basis with vectors of length k, l respectively, it is $\begin{pmatrix} 0 & -k/l \\ l/k & 0 \end{pmatrix}$. If we choose perpendicular vectors $(1,-1,0)$ and $(1,1,-2)$ in the plane $x+y+z=0$ with lengths $\sqrt{2}$ and $\sqrt{6}$, then in the basis $\underline{f} = ((1,1,1),(1,-1,0),(1,1,-2))$ the rotation we are looking for has matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\sqrt{3} \\ 0 & 1/\sqrt{3} & 0 \end{pmatrix}$. In order to obtain the matrix of the rotation in the standard basis, it is enough to change the basis. The transition matrix T for changing the basis from the basis $\underline{f}$ to the standard basis is obtained by writing the coordinates (under the standard basis) of the vectors of the basis $\underline{f}$ as the columns of the matrix

Note that this scalar product is linear in each of its arguments, and we denote it by $u \cdot v$ or by $\langle v, v' \rangle$. The scalar product defined in such a way is symmetric in its arguments and of course $\|v\| = 0$ if and only if $v = 0$. We also see immediately that two vectors in the Euclidean plane are perpendicular whenever their scalar product is zero.

Now we shall mimic this approach for higher dimensions. First, observe that the angle between two vectors is always a two-dimensional concept (we want the angle to be the same in the two-dimensional space containing the two vectors u and v). In the subsequent paragraphs, we shall consider only finitely dimensional vector spaces over real scalars $\mathbb{R}$.

Scalar product and orthogonality

A scalar product on a vector space V over real numbers is a mapping $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ which is symmetric in its arguments, linear in each of them, and such that $\langle v, v \rangle \geq 0$ and $\|v\|^2 = \langle v, v \rangle = 0$ if and only if $v = 0$.

The number $\|v\| = \sqrt{\langle v, v \rangle}$ is called the length of the vector v.

Vectors v and $w \in V$ are called *orthogonal* or *perpendicular* whenever $\langle v, w \rangle = 0$. We also write $v \perp w$. The vector v is called *normalised* whenever $\|v\| = 1$.

The basis of the space V composed exclusively of mutually orthogonal vectors is called an *orthogonal basis*. If the vectors in such a basis are all normalised, we call the basis *orthonormal*.

A scalar product is very often denoted by the common dot, that is, $\langle u, v \rangle = u \cdot v$. Thus, it is then necessary to recognize from the context whether the dot means a product of two vectors (the result is a scalar) or something different (e.g. we often denote the product of matrices and product of scalars in the same way).

Because the scalar product is linear in each of its arguments, it is completely determined by its values on pairs of basis vectors. Indeed, choose a basis $\underline{u} = (u_1, \dots, u_n)$ of the space V and denote

$$s_{ij} = \langle u_i, u_j \rangle.$$

Then from the symmetry of the scalar product we know $s_{ij} = s_{ji}$ and from the linearity of the product in each of its arguments we get

$$\Big\langle \sum_i x_i u_i, \sum_j y_j u_j \Big\rangle = \sum_{i,j} x_i y_j \langle u_i, u_j \rangle = \sum_{i,j} s_{ij} x_i y_j.$$

If the basis is orthonormal, the matrix S is the unit matrix. This proves the following useful claim:

T: $T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{pmatrix}$. Finally, for the desired matrix R, we have

$$R = T \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\sqrt{3} \\ 0 & 1/\sqrt{3} & 0 \end{pmatrix} \cdot T^{-1}$$
$$= \begin{pmatrix} 1/3 & 1/3 - \sqrt{3}/3 & 1/3 + \sqrt{3}/3 \\ 1/3 + \sqrt{3}/3 & 1/3 & 1/3 - \sqrt{3}/3 \\ 1/3 - \sqrt{3}/3 & 1/3 + \sqrt{3}/3 & 1/3 \end{pmatrix}$$

This result can be checked by substituting into the matrix of general rotation (2.E.3). By normalizing the vector $(1, 1, 1)$ we obtain the vector $(x, y, z) = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, $\cos(\varphi) = 0, \sin(\varphi) = 1$. □

2.E.7. Matrix of general rotation revisited. We derive

the matrix of (general) rotation from (2.E.3) through the angle φ in the positive sense about the unit vector (x, y, z) in a different way, analogically to the previous exercise. In the basis $\underline{f} = ((x, y, z), (-y, x, 0), (zx, zy, z^2 - 1))$, that is, in the orthogonal basis composed of the directional vector of the axis of rotation and of two mutually perpendicular vectors with sizes $\sqrt{1 - z^2}$ lying in a plane perpendicular to the axis of rotation, the matrix corresponding to the rotation is $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi) & \cos(\varphi) \end{pmatrix}$. The matrix for changing the basis from $\underline{f}$ to the standard basis is then $T = \begin{pmatrix} x & -y & zx \\ y & x & zy \\ z & 0 & z^2 - 1 \end{pmatrix}$ with the inverse matrix

$$T^{-1} = \begin{pmatrix} x & y & z \\ -\frac{y}{1-z^2} & \frac{x}{1-z^2} & 0 \\ \frac{zx}{1-z^2} & \frac{zy}{1-z^2} & -1 \end{pmatrix}.$$

Finally, for the matrix R of the rotation we obtain

$$R = T \cdot A \cdot T^{-1}$$
$$= \begin{pmatrix} 1 - t + tx^2 & txy - zs & txz + ys \\ yxt + zs & 1 - t + ty^2 & tyz - xs \\ zxt - ys & tzy + xs & 1 - t + tz^2 \end{pmatrix},$$

where again $t = 1 - \cos\varphi$ and $s = \sin\varphi$, and we get the same matrix as before.

When multiplying and simplifying, we must repeatedly use the assumption $x^2 + y^2 + z^2 = 1$.

Scalar product in coordinates

Proposition. *For every orthonormal basis, the scalar product is given by the coordinate expression*
$$\langle x, y\rangle = y^T \cdot x.$$
For each basis of the space V there is the symmetric matrix S such that the coordinate expression of the scalar product is
$$\langle x, y\rangle = y^T \cdot S \cdot x.$$

Notice, that with symmetric matrix S it is just a matter of convention in which order we insert the vectors: the formula
$$x^T \cdot S \cdot y = (x^T \cdot S \cdot y)^T = y^T \cdot S^T \cdot x$$
produces the same value. However, we shall later consider the second argument as a linear form, thus it seems to be more convenient to use the expression $y^T \cdot S \cdot x$.

2.3.19. Orthogonal complements and projections. For ev-

ery fixed subspace $W \subset V$ in a space with scalar product, we define its *orthogonal complement* as

$$W^\perp = \{u \in V;\ u \perp v \text{ for all } v \in W\}.$$

It follows directly from the definition that $W^\perp$ is a vector subspace. If $W \subset V$ has a basis $(u_1, \dots, u_k)$ then the description for $W^\perp$ is given as k homogeneous equations for n variables. Thus $W^\perp$ will have dimension at least $n - k$. Also $u \in W \cap W^\perp$ means that $\langle u, u\rangle = 0$, and thus also $u = 0$ by the definition of scalar product. Clearly then, V is the direct sum
$$V = W \oplus W^\perp.$$

A linear mapping $f : V \to V$ on any vector space is called a *projection*, if we have
$$f \circ f = f.$$
In such a case, we can write, for every vector $v \in V$,
$$v = f(v) + (v - f(v)) \in \mathrm{Im}(f) + \mathrm{Ker}(f) = V$$
and if $v \in \mathrm{Im}(f)$ and $f(v) = 0$, then also $v = 0$. Thus the above sum of the subspaces is direct. We say that f is a projection to the subspace $W = \mathrm{Im}(f)$ along the subspace $U = \mathrm{Ker}(f)$. In words, the projection can be described naturally as follows: we decompose the given vector into a component in W and a component in U, and forget the second one.

If V has a scalar product, we say that the projection is orthogonal if the kernel is orthogonal to the image.

Every subspace $W \neq V$ thus defines an *orthogonal projection* to W. It is a projection to W along $W^\perp$, given by the unique decomposition of every vector u into components $u_W \in W$ and $u_{W^\perp} \in W^\perp$, that is, linear mapping which maps $u_W + u_{W^\perp}$ to u_W.

Through a more detailed analysis of properties of various types of linear mapping we now obtain a deeper understanding of tools we are given by vector spaces for linear modeling of processes and systems.

2.E.8. Consider complex numbers as a real vector space and choose 1 and i for its basis. Determine in this basis the matrix of the following linear mappings:

a) conjugation,
b) multiplication by the number $(2+i)$.

Determine the matrix of these mappings in the basis $\underline{f} = ((1-i),(1+i))$.

Solution. In order to determine the matrix of a linear mapping in some basis, it is enough to determine the images of the basis vectors.

a) For conjugation we have $1 \mapsto 1$, $i \mapsto -i$, written in the coordinates $(1,0) \mapsto (1,0)$ and $(0,1) \mapsto (0,-1)$. By writing the images into the columns we obtain the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, In the basis $\underline{f}$ the conjugation interchanges basis vectors, that is, $(1,0) \mapsto (0,1)$ and $(0,1) \mapsto (1,0)$ and the matrix of conjugation under this basis is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

b) For the basis $(1,i)$ we obtain $1 \mapsto 2+i$, $i \mapsto 2i-1$, that is, $(1,0) \mapsto (2,1)$, $(0,1) \mapsto (2,-1)$. Thus the matrix of multiplication by the number $2+i$ under the basis $(1,i)$ is: $\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$.

We determine the matrix in the basis $\underline{f}$. Multiplication by $(2+i)$ gives us: $(1-i) \mapsto (1-i)(2+i) = 3-i$, $(1+i) \mapsto (1+3i)$. Coordinates $(a,b)_{\underline{f}}$ of the vector $3-i$ in the basis $\underline{f}$ are given, as we know, by the equation $a\cdot(1-i)+b\cdot(1+i) = 3+i$, that is, $(3+i)_{\underline{f}} = (2,1)$. Analogously $(1+3i)_{\underline{f}} = (-1,2)$. Altogether, we obtain the matrix $\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$.

Think about the following: why is the matrix of multiplication by $2+i$ the same in both bases? Would the two matrices in these bases be the same for multiplication by any complex number? □

2.E.9. Determine the matrix A which, under the standard basis of the space $\mathbb{R}^3$, gives the orthogonal projection on the vector subspace generated by the vectors $u_1 = (-1,1,0)$ and $u_2 = (-1,0,1)$.

Solution. Note first that the given subspace is a plane containing the origin with normal vector $u_3 = (1,1,1)$. The ordered

2.3.20. Existence of orthonormal bases. It is easy to see

that on every finite dimensional real vector space there exist scalar products. Just choose any basis. Define lengths so that each basis vector is of unit length. Immediately we have a scalar product. Call it orthonormal. In this basis the scalar products of vectors are computed as in the formula in the Theorem 2.3.18.

More often we are given a scalar product on a vector space V, and we want to find an appropriate orthonormal basis for it. We present an algorithm using suitable orthogonal projections in order to transform any basis into an orthogonal one. It is called the *Gramm-Schmidt orthogonalization process*.

The point of this procedure is to transform a given sequence of independent generators $v_1,\dots,v_k$ of a finite dimensional space V into an orthogonal set of independent generators of V.

Gramm-Schmidt orthogonalization

Proposition. *Let $(u_1,\dots,u_k)$ be a linearly independent k-tuple of vectors of a space V with scalar product. Then there exists an orthogonal system of vectors $(v_1,\dots,v_k)$ such that $v_i \in \operatorname{span}\{u_1,\dots,u_i\}$, and $\operatorname{span}\{u_1,\dots,u_i\} = \operatorname{span}\{v_1,\dots,v_i\}$, for all $i = 1,\dots,k$. We obtain it by the following procedure:*

- *The independence of the vectors u_i ensures that $u_1 \neq 0$; we choose $v_1 = u_1$.*
- *If we have already constructed the vectors $v_1,\dots,v_\ell$ with the required properties and if $\ell < k$, we choose $v_{\ell+1} = u_{\ell+1} + a_1 v_1 + \dots + a_\ell v_\ell$, where $a_i = -\frac{\langle u_{\ell+1}, v_i\rangle}{\|v_i\|^2}$.*

Proof. We begin with the first (nonzero) vector v_1 and calculate the orthogonal projection v_2 to

$$\operatorname{span}\{v_1\}^\perp \subset \operatorname{span}\{v_1, v_2\}.$$

The result is nonzero if and only if v_2 is independent of v_1. All other steps are similar:

In step ℓ, $\ell > 1$ we seek the vector $v_{\ell+1} = u_{\ell+1} + a_1 v_1 + \dots + a_\ell v_\ell$ satisfying $\langle v_{\ell+1}, v_i\rangle = 0$ for all $i = 1,\dots,\ell$. This implies

$$0 = \langle u_{\ell+1} + a_1 v_1 + \dots + a_\ell v_\ell, v_i\rangle = \langle u_{\ell+1}, v_i\rangle + a_i \langle v_i, v_i\rangle$$

and we can see that the vectors with the desired properties are determined uniquely up to a scalar multiple. □

Whenever we have an orthogonal basis of a vector space V, we just have to normalise the vectors in order to obtain an orthonormal basis. Thus, starting the Gramm-Schmidt orthogonalization with any basis of V, we have proven:

Corollary. *On every finite dimensional real vector space with scalar product there exists an orthonormal basis.*

In an orthonormal basis, the coordinates and orthogonal projections are very easy to calculate. Indeed, suppose we triple $(1,1,1)$ is clearly a solution to the system

$$\begin{array}{rcrcrcl} -x_1 & + & x_2 & & & = & 0, \\ -x_1 & & & + & x_3 & = & 0, \end{array}$$

that is, the vector u_3 is perpendicular to the vectors u_1, u_2.

Under the given projection the vectors u_1 and u_2 must map to themselves and the vector u_3 on the zero vector. In the basis composed of u_1, u_2, u_3 (in this order) is thus the matrix of this projection

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

Using the the transition matrix for changing the basis

$$T = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

from the basis (u_1, u_2, u_3) to the standard basis, and from the standard basis to the basis (u_1, u_2, u_3) we obtain

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

□

F. Inner products and linear maps

2.F.1. Write down the matrix of the mapping of orthogonal projection on the plane passing through the origin and perpendicular to the vector $(1,1,1)$.

Solution. The image of an arbitrary point (vector) $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ under the considered mapping can be obtained by subtracting from the given vector its orthogonal projection onto the direction normal to the considered plane, that is, onto the direction $(1,1,1)$. This projection $\mathbf{p}$ is given by (see 1) as

$$\frac{\langle \mathbf{x}, (1,1,1)\rangle}{|(1,1,1)|^2}$$
$$= (\frac{x_1+x_2+x_3}{3}, \frac{x_1+x_2+x_3}{3}, \frac{x_1+x_2+x_3}{3}).$$

The resulting mapping is thus

$$\mathbf{x} - \mathbf{p}$$
$$= (\frac{2x_1}{3} - \frac{x_2+x_3}{3}, \frac{2x_2}{3} - \frac{x_1+x_3}{3}, \frac{2x_3}{3} - \frac{x_1+x_2}{3}) =$$
$$= \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

have an orthonormal basis $(e_1, \dots, e_n)$ for a space V. Then every vector $v = x_1 e_1 + \dots + x_n e_n$ satisfies

$$\langle e_i, v\rangle = \langle e_i, x_1 e_1 + \dots + x_n e_n\rangle = x_i$$

and so we can always express

$$(1) \qquad v = \langle e_1, v\rangle e_1 + \dots + \langle e_n, v\rangle e_n.$$

If we are given a subspace $W \subset V$ and its orthonormal basis $(e_1, \dots, e_k)$, then we can extend it to an orthonormal basis $(e_1, \dots, e_n)$ for V. Orthogonal projection of a general vector $v \in V$ to W is then given by the expression

$$v \mapsto \langle e_1, v\rangle e_1 + \dots + \langle e_n, v\rangle e_k.$$

In particular, we need only consider an orthonormal basis of the subspace W in order to write the orthogonal projection to W explicitly.

Note that in general the projection f to the subspace W along U and the projection g to U along W is constrained by the equality $g = \mathrm{id}_V - f$. Thus, when dealing with orthogonal projections to a given subspace W, it is always more efficient to calculate the orthonormal basis of that space W or $W^\perp$ whose dimension is smaller.

Note also that the existence of an orthonormal basis guarantees that for every real space V of dimension n with a scalar product, there exists a linear mapping which is an isomorphism between V and the space $\mathbb{R}^n$ with the standard scalar product. Similarly it has been shown already in Theorem 2.3.18, where we found that the desired isomorphism is exactly the coordinate assignment. In words – in every orthonormal basis the scalar product is computed by the same formula as the standard scalar product in $\mathbb{R}^n$.

We shall return to the questions of the length of a vector and to projections in the following chapter in a more general context.

2.3.21. Angle between two vectors. As we have already noted, the angle between two linearly independent vectors in the space must be the same as when we consider them in the two-dimensional subspace they generate. Basically, this is the reason why the notion of angle is independent of the dimension of the original space. If we choose an orthogonal basis such that its first two vectors generate the same subspace as the two given vectors u and v (whose angle we are measuring), we can simply take the definition from the planar geometry. Independently of the choice of coordinates we can formulate the definition as follows:

Angle between two vectors

The angle φ between two vectors v and w in a vector space with a scalar product is given by the relation

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

The angle defined in this way does not depend on the order of the vectors v, w and it is chosen in the interval $0 \le \varphi \le \pi$.

We shall return to scalar products and angles between vectors in further chapters.

We have (correctly) obtained the same matrix as in the exercise 2.E.9. □

2.F.2. In $\mathbb{R}^3$ write down the matrix of the mirror symmetry with respect to the plane containing the origin and $(1,1,1)$ being its normal vector.

Solution. As in 2.F.1 we get the image of an arbitrary vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ with the help of the orthogonal projection onto the direction $(1,1,1)$. Unlike in the previous example, we need to subtract the projection twice (see image). Thus we get the matrix:

$$x - 2p =$$
$$\left(\frac{x_1}{3} - \frac{2(x_2+x_3)}{3}, \frac{x_2}{3} - \frac{2(x_1+x_3)}{3}, \frac{x_3}{3} - \frac{2(x_1+x_2)}{3}\right)$$
$$= \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Second solution. The normed normal vector of the mirror plane is $n = \frac{1}{\sqrt{3}}(1,1,1)$. We can express the mirror image of v under the mirror symmetry Z as follows: $Z(v) = v - 2\langle v, n\rangle n = v - 2n \cdot (n^T \cdot v) = v - 2(n \cdot n^T) \cdot v = ((E - 2n \cdot n^T)v$ (where we have used $\langle v, n\rangle = v \cdot n^T$ for the standard scalar product and the associativity of the matrix multiplication). We get the same matrix:

$$E - 2n \cdot n^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix}.$$

□

2.F.3. Consider $\mathbb{R}^3$, with the standard coordinate system. In the plane $z = 0$ there is a mirror and at the point $[4,3,5]$ there is a candle. The observer at the point $[1,2,3]$ is not aware of the mirror, but sees in it the reflection of the candle. Where does he think the candle is?

Solution. Independently of our position, we see the mirror image of the scene in the mirror (that is why it is called a mirror image). The mirror image is given by reflecting the scene (space) by the plane of the mirror, the plane $z = 0$. The reflection with respect to this plane changes the sign of the z-coordinate. That is we can see the candle at the point $[4,3,-5]$. □

By using the inner product we can determine the (angular) deflection of the vectors:

2.3.22. Multilinear forms. The scalar product was given as a mapping from the product of two copies of a vector space V into the space of scalars, which was linear in each of its arguments. Similarly, we will work with mappings from the product of k copies of a vector space V into the scalars, which are linear in each of its k arguments. We speak of k-*linear* forms.

Most often we will meet *bilinear forms*, that is, the case $\alpha : V \times V \to \mathbb{K}$, where for any four vectors u, v, w, z and scalars a, b, c and d we have

$$\alpha(au + bv, cw + dz) = ac\,\alpha(u,w) + ad\,\alpha(u,z) + bc\,\alpha(v,w) + bd\,\alpha(v,z).$$

If additionally we always have

$$\alpha(u,w) = \alpha(w,u),$$

then we speak of a *symmetric bilinear form*. If interchanging the arguments leads to a change of sign, we speak of an *antisymmetric bilinear form*.

Already in planar geometry we have defined the determinant as a bilinear antisymmetric form α, that is, $\alpha(u,w) = -\alpha(w,u)$. In general, due to the theorem 2.2.4, we know that the determinant with dimension n can be seen as an n-linear antisymmetric form.

As with linear mappings it is clear that every k-linear form is completely determined by its values on all k-tuples of basis elements in a fixed basis. In analogy to linear mappings we can see these values as k-dimensional analogues to matrices. We show this by an example with $k = 2$, where it will correspond to matrices as we have defined them.

Matrix of a bilinear form

If we choose a basis $\underline{u}$ on V and define for a given bilinear form α scalars $a_{ij} = \alpha(u_i, u_j)$ then we obtain for vectors v, w with coordinates x and y (as columns of coordinates)

$$\alpha(v,w) = \sum_{i,j=1}^{n} a_{ij} x_i y_j = x^T \cdot A \cdot y,$$

where A is a matrix $A = (a_{ij})$.

Directly from the definition of the matrix of a bilinear form we see that the form is symmetric or antisymmetric if and only if the corresponding matrix has this property.

Every bilinear form α on a vector space V defines a mapping $V \to V^*$, $v \mapsto \alpha(\ ,v)$. That is, by placing a fixed vector in the first argument we obtain a linear form which is the image of this vector. If we choose a fixed basis on a finitely dimensional space V and a dual basis V^*, then we have the mapping

$$x \mapsto (y \mapsto x^T \cdot A \cdot y).$$

All this is a matter of convention. Also we may fix the second vector and get a linear form again.

2.F.4. Determine the deflection of the roots of the polynomial $x^2 - i$ considered as vectors in the complex plane.

Solution. The roots of the given polynomial are square roots of i. The arguments of the square roots of any complex numbers differ according to the de Moivre theorem by π. Their deflection is thus always π. $\square$

2.F.5. Determine the cosine of the deflection of the lines p, q in $\mathbb{R}^3$ given by the equations

$$\begin{aligned} p \quad &: \quad -2x + y + z = 1 \\ & \qquad x + 3y - 4z = 5 \\ q \quad &: \quad x - y = -2 \\ & \qquad z = 6 \end{aligned}$$

$\bigcirc$

2.F.6. Using the Gram-Schmidt orthogonalisation, obtain the orthogonal basis of the subspace

$$U = \left\{(x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4;\ x_1 + x_2 + x_3 + x_4 = 0\right\}$$

of the space $\mathbb{R}^4$.

Solution. The set of solutions of the given homogeneous linear equation is clearly a vector space with the basis

$$u_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

shall be denoted Denote by v_1, v_2, v_3, vectors of the orthogonal basis obtained using the Gram-Schmidt orthogonalisation process.

First set $v_1 = u_1$. Then let

$$v_2 = u_2 - \frac{u_2^T \cdot v_1}{||v_1||^2}\, v_1 = u_2 - \frac{1}{2}\, v_1 = \left(-\frac{1}{2}, -\frac{1}{2}, 1, 0\right)^T,$$

that is, choose a multiple $v_2 = (-1, -1, 2, 0)^T$. Then let

$$v_3 = u_3 - \frac{u_3^T \cdot v_1}{||v_1||^2}\, v_1 - \frac{u_3^T \cdot v_2}{||v_2||^2}\, v_2 = u_3 - \frac{1}{2}\, v_1 - \frac{1}{6}\, v_2 =$$
$$= \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 1\right)^T.$$

Altogether we have

$$v_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ -1 \\ 2 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 3 \end{pmatrix}.$$

Due to the simplicity of the exercise we can immediately give an orthogonal basis of the vectors

$$(1, -1, 0, 0)^T, \quad (0, 0, 1, -1)^T, \quad (1, 1, -1, -1)^T$$

4. Properties of linear mappings

In order to exploit vector spaces and linear mappings in modelling real processes and systems in other sciences, we need a more detailed analysis of properties of diverse types of linear mappings.

2.4.1. We begin with four examples in the lowest dimension of interest.

With the standard basis of the plane $\mathbb{R}^2$ and with the standard scalar product we consider the following matrices of mappings $f : \mathbb{R}^2 \to \mathbb{R}^2$:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, C = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, D = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The matrix A describes the orthogonal projection along the subspace

$$W = \{(0, a);\ a \in \mathbb{R}\} \subset \mathbb{R}^2$$

to the subspace

$$V = \{(a, 0);\ a \in \mathbb{R}\} \subset \mathbb{R}^2,$$

that is, the projection to the x-axis along the y-axis. Evidently for this $f : \mathbb{R}^2 \to \mathbb{R}^2$ we have $f \circ f = f$ and thus the restriction $f|_V$ of the given mapping on its codomain is the identity mapping. The kernel of f is exactly the subspace W.

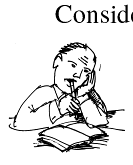
The matrix B has the property $B^2 = 0$, therefore the same holds for the corresponding mapping f. We can envision this as the differentiation of polynomials $\mathbb{R}_1[x]$ of degree at most one in the basis $(1, x)$ (we shall come to differentiation in chapter five, see **??**).

The matrix C gives a mapping f, which rescales the first vector of the basis a-times, and the second one b-times. Therefore the whole plane divides into two subspaces, which are preserved under the mapping and where it is only a *homothety*, that is, scaling by a scalar multiple (the first case was a special case with $a = 1$, $b = 0$). For instance the choice $a = 1$, $b = -1$ corresponds to axial symmetry (mirror symmetry) under the x-axis, which is the same as complex conjugation $x+iy \mapsto x-iy$ on the two-dimensional real space $\mathbb{R}^2 \simeq \mathbb{C}$ in basis $(1, i)$. This is a linear mapping of the two-dimensional real vector space $\mathbb{C}$, but not of the one-dimensional complex space $\mathbb{C}$.

The matrix D is the matrix of rotation by 90 degrees (the angle $\pi/2$) centered at the origin in the standard basis. We can see at first glance that none of the one-dimensional subspaces is preserved under this mapping.

Such a rotation is a bijection of the plane onto itself, therefore we can surely find distinct bases in the domain and codomain, where its matrix will be the unit matrix E. We simply take any basis of the domain and its image in the codomain. But we are not able to do this with the same basis for both the domain and the codomain.

Consider the matrix D as a matrix of the mapping $g : \mathbb{C}^2 \to \mathbb{C}^2$ with the standard basis of the complex vector space $\mathbb{C}^2$. Then we can find vectors $u = (i, 1)$, $v = (-i, 1)$, for which we have



or

$$(-1,1,1,-1)^T, \quad (1,-1,1,-1)^T, \quad (-1,-1,1,1)^T.$$

$\square$

2.F.7. Write down a basis of the real vector space of the matrices 3×3 over $\mathbb{R}$ with zero trace. (The trace of a matrix is the sum of the elements on the diagonal). Write the coordinates of the matrix

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

in this basis.

2.F.8. Find the orthogonal complement $U^\perp$ of the subspace

$$U = \{(x_1, x_2, x_3, x_4);\ x_1 = x_3, x_2 = x_3 + 6x_4\} \subset \mathbb{R}^4.$$

Solution. The orthogonal complement $U^\perp$ consists of just those vectors that are perpendicular to every solution of the system

$$\begin{array}{ccccccc} x_1 & & - & x_3 & & = & 0, \\ & x_2 & - & x_3 & - \ 6x_4 & = & 0. \end{array}$$

A vector is a solution of this system if and only if it is perpendicular to both vectors $(1, 0, -1, 0)$, $(0, 1, -1, -6)$. Thus we have

$$U^\perp = \{a \cdot (1, 0, -1, 0) + b \cdot (0, 1, -1, -6);\ a, b \in \mathbb{R}\}.$$

$\square$

2.F.9. Find an orthonormal basis of the subspace $V \subset \mathbb{R}$, where $V = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 + 2x_2 + x_3 = 0\}$.

Solution. The fourth coordinate does not appear in the restriction for the subspace, thus it seems reasonable to select $(0, 0, 0, 1)$ as one of the vectors of the orthonormal basis and reduce the problem into the subspace $\mathbb{R}^3$. If we set the second coordinate equal to zero, then in the investigated space there are vectors with reverse first and third coordinate, notably, the unit vector $(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, 0)$. This vector is perpendicular to any vector which has first coordinate equal to the third coordinate. In order to get into the investigated subspace, we choose the second coordinate equal to the negative of the sum of the first and the third coordinate, and then normalise. Thus we choose the vector $(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, 0)$ and we are finished. $\square$

$$g(u) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \cdot u,$$

$$g(v) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} -1 \\ -i \end{pmatrix} = -i \cdot v.$$

That means that in the basis (u, v) on $\mathbb{C}^2$, the mapping g has the matrix

$$K = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Notice that by extending the scalars to $\mathbb{C}$, we arrive at an analogy to the matrix C with diagonal elements $a = \cos(\frac{1}{2}\pi) + i\sin(\frac{1}{2}\pi)$ and its complex conjugate $\bar{a}$. In other words, the argument of the number a in polar form provides the angle of the rotation.

This is easy to understand, if we denote the real and imaginary part of the vector u as follows

$$u = x_u + iy_u = \operatorname{Re} u + i \operatorname{Im} u = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + i \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The vector v is the complex conjugate of u. We are interested in the restriction of the mapping g to the real vector subspace $V = \mathbb{R}^2 \cap \operatorname{span}\{u, v\} = \mathbb{C}^2$. Evidently,

$$V = \operatorname{span}\{u + \bar{u}, i(u - \bar{u})\} = \operatorname{span}\{x_u, -y_u\}$$

is the whole plane $\mathbb{R}^2$. The restriction of g to this plane is exactly the original mapping given by the matrix A. From the definition of multiplication by the complex unit, it is a rotation through the angle $\frac{1}{2}\pi$ in the positive sense with respect to the chosen basis $x_u, -y_u$. Work it by yourself with a direct calculation. Note also why exchanging the order of the vectors u and v leads to the same result, although in a different real basis!

2.4.2. Eigenvalues and eigenvectors of mappings. A key to the description of mappings in the previous examples was the answer to the question "what are the vectors satisfying the equation $f(u) = a \cdot u$ for some suitable scalars a?".

We consider this question for any linear mapping $f : V \to V$ on a vector space of dimension n over scalars $\mathbb{K}$. If we imagine such an equality written in coordinates, i.e. using the matrix of the mapping A in some bases, we obtain a system of linear equations

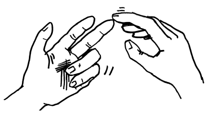
$$A \cdot x - a \cdot x = (A - a \cdot E) \cdot x = 0$$

with an unknown parameter a. We know already that such a system of equations has only the solution $x = 0$ if the matrix $A - aE$ is invertible. Thus we want to find such values $a \in \mathbb{K}$ for which $A - aE$ is not invertible, and for that, the necessary and sufficient condition reads (see Theorem 2.2.10)

$$\det(A - a \cdot E) = 0. \tag{1}$$

G. Eigenvalues and eigenvectors

2.G.1. Find the eigenvalues and the associated subspaces of eigenvectors of the matrix



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

Solution. First we find the characteristic polynomial of the matrix:

$$\begin{vmatrix} -1-\lambda & 1 & 0 \\ -1 & 3-\lambda & 0 \\ 2 & -2 & 2-\lambda \end{vmatrix} = \lambda^3 - 4\lambda^3 + 2\lambda + 4.$$

This polynomial has roots 2, $1+\sqrt{3}$, $1-\sqrt{3}$, which are then the eigenvalues of the matrix. Their algebraic multiplicity is one (they are simple roots of the polynomial), thus each has associated only one (up to a non-zero multiple) eigenvector. Otherwise stated, the geometric multiplicity of the eigenvalue is one, see 2.4.8).

We determine the eigenvector associated with the eigenvalue 2. It is a solution of the homogeneous linear system with the matrix $A - 2E$:

$$\begin{aligned} -3x_1 + x_2 &= 0 \\ -1x_1 + x_2 &= 0 \\ 2x_1 - 2x_2 &= 0. \end{aligned}$$

The system has solution $x_1 = x_2 = 0$, $x_3 \in \mathbb{R}$ arbitrary. So the eigenvector associated with the value 2 is then the vector $(0,0,1)$ (or any multiple of it).

Similarly we determine the remaining two eigenvectors – as solutions of the system $[A - (1+\sqrt{3})E]\mathbf{x} = 0$. The solution of the system

$$\begin{aligned} (-2-\sqrt{3})x_1 + x_2 &= 0 \\ -1x_1 + (2-\sqrt{3})x_2 &= 0 \\ 2x_1 - 2x_2 + (1-\sqrt{3})x_3 &= 0 \end{aligned}$$

is the space $\{(2-\sqrt{3},1,2)\,t,\ t \in \mathbb{R}\}$.

That is the space of eigenvectors associated with the eigenvalue $1+\sqrt{3}$.

Similarly we obtain that the space of eigenvectors associated with the eigenvalue $1-\sqrt{3}$ is $\{(2+\sqrt{3},1,-2)\,t,\ t \in \mathbb{R}\}$. □

2.G.2. Determine the eigenvalues and eigenvectors of the matrix

If we consider $\lambda = a$ as a variable in the previous scalar equation, we are actually looking for the roots of a polynomial of degree n. As we have seen in the case of the matrix D, the roots may exist in an extension of our field of scalars, if they are not in $\mathbb{K}$.

Eigenvalues and eigenvectors

Scalars $\lambda \in \mathbb{K}$ satisfying the equation $f(u) = \lambda \cdot u$ for some nonzero vector $u \in V$ are called the *eigenvalues of mapping f*. The corresponding nonzero vectors u are called the *eigenvectors of the mapping f*.

If u, v are eigenvectors associated with the same eigenvalue λ, then for every linear combination of u and v,

$$f(au + bv) = af(u) + bf(v) = \lambda(au + bv).$$

Therefore the eigenvectors associated with the same eigenvalue λ, together with the zero vector, form a nontrivial vector subspace $V_\lambda \subset V$. We call it the *eigenspace associated with λ*. For instance, if $\lambda = 0$ is an eigenvalue, the kernel $\operatorname{Ker} f$ is the eigenspace V_0.

We have seen how to compute the eigenvalues in coordinates. The independence of the eigenvalues from the choice of coordinates is clear from their definition. But let us look explicitely what happens if we change the basis. As a direct corollary of the transformation properties from the paragraph 2.3.16 and the Cauchy theorem 2.2.6 for calculation of the determinant of product, the matrix A' in the new coordinates will be $A' = P^{-1}AP$ with an invertible matrix P. Thus

$$\begin{aligned} |P^{-1}AP - \lambda E| &= |P^{-1}AP - P^{-1}\lambda EP| \\ &= |P^{-1}(A - \lambda E)P| \\ &= |P^{-1}||(A - \lambda E)||P| \\ &= |A - \lambda E|, \end{aligned}$$

because the scalar multiplication is commutative and we know that $|P^{-1}| = |P|^{-1}$.

For these reasons we use the same terminology for matrices and mappings:

Characteristic polynomials

For a matrix A of dimension n over $\mathbb{K}$ we call the polynomial $|A - \lambda E| \in \mathbb{K}_n[\lambda]$ *the characteristic polynomial of the matrix A*.

Roots of this polynomial are the *eigenvalues of the matrix A*. If A is the matrix of the mapping $f : V \to V$ in a certain basis, then $|A - \lambda E|$ is also called the *characteristic polynomial of the mapping f*.

Because the characteristic polynomial of a linear mapping $f : V \to V$ is independent of the choice of the basis of V, the coefficients of individual powers of the variable λ are scalars expressing some properties of f. In particular, they too cannot depend on the choice of the basis. Suppose $\dim V = n$ and $A = (a_{ij})$ is the matrix of the mapping in

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

Describe the geometric interpretation of this mapping and write down its matrix in the basis:

$$\begin{aligned} e_1 &= (1, -1, 1) \\ e_2 &= (1, 2, 0) \\ e_3 &= (0, 1, 1) \end{aligned}$$

Solution. The characteristic polynomial of the matrix A is

$$\begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 1 & 2 & 1-\lambda \end{vmatrix} = -\lambda^3+4\lambda^2-2\lambda = -\lambda(\lambda^2-4\lambda+2).$$

The roots of this polynomial are the eigenvalues, thus the eigenvalues are 0, $2+\sqrt{2}$, $2-\sqrt{2}$. Thus eigenvalues are 0, $2+\sqrt{2}$, $2-\sqrt{2}$. We compute the eigenvectors associated with the particular eigenvalues:

- 0: We solve the system
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$
Its solutions form a one-dimensional vector space of eigenvectors: $\operatorname{span}\{(1, -1, 1)\}$.
- $2+\sqrt{2}$: We solve the system
$$\begin{pmatrix} -(1+\sqrt{2}) & 1 & 0 \\ 1 & -\sqrt{2} & 1 \\ 1 & 2 & -(1+\sqrt{2}) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$
The solutions form a one-dimensional space $\operatorname{span}\{(1, 1+\sqrt{2}, 1+\sqrt{2})\}$.
- $2-\sqrt{2}$: We solve the system
$$\begin{pmatrix} (\sqrt{2}-1) & 1 & 0 \\ 1 & \sqrt{2} & 1 \\ 1 & 2 & (\sqrt{2}-1) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$
Its solutions form a space of eigenvectors $\operatorname{span}\{(1, 1-\sqrt{2}, 1-\sqrt{2})\}$.

Hence the given matrix has eigenvalues 0, $2+\sqrt{2}$ and $2-\sqrt{2}$, with the associated one-dimensional spaces of eigenvectors $\operatorname{span}\{(1, -1, 1)\}$, $\operatorname{span}\{(1, 1+\sqrt{2}, 1+\sqrt{2})\}$ and $\operatorname{span}\{(1, 1-\sqrt{2}, 1-\sqrt{2})\}$ respectively.

The mapping can thus be interpreted as a projection along the vector $(1, -1, 1)$ into the plane given by the vectors $(1, 1+\sqrt{2}, 1+\sqrt{2})$ and $(1, 1-\sqrt{2}, 1-\sqrt{2})$ composed with the linear mapping given by "stretching" by the factor corresponding to the eigenvalues in the directions of the associated eigenvectors.

some basis. Then

$$|A - \lambda \cdot E| = (-1)^n \lambda^n + (-1)^{n-1}(a_{11} + \cdots + a_{nn})\lambda^{n-1} + \cdots + |A|\lambda^0.$$

The coefficient at the highest power says whether the dimension of the space V is even or odd.

The most interesting coefficient is the sum of the diagonal elements of the matrix. We have just proved that it does not depend on the choice of the basis and we call it the *trace of the matrix A* and denote it by $\operatorname{Tr} A$. The *trace of the mapping f* is defined as a trace of the matrix in an arbitrary basis.

In fact, this is not so surprising once we notice that the trace is actually the linear approximation of the determinant in the neighbourhood of the unit matrix in the direction A. We shall deal with such concepts in Chapter 8 only. But since the determinant is a polynomial, we may see easily that the only terms in $\det(E + tA)$ which are linear in the real parameter t are just the trace.

We discuss a few important properties of eigenspaces now.

2.4.3. Theorem. *Eigenvectors of linear mappings $f : V \to V$ associated to different eigenvalues are linearly independent.*

Proof. Let $a_1, \dots, a_k$ be distinct eigenvalues of the mapping f and $u_1, \dots, u_k$ eigenvectors with these eigenvalues. The proof is by induction on the number of linearly independent vectors among the chosen ones. Assume that $u_1, \dots, u_\ell$ are linearly independent and $u_{l+1} = \sum_i c_i u_i$ is their linear combination. We can choose $\ell = 1$, because the eigenvectors are nonzero. But then $f(u_{\ell+1}) = a_{l+1} \cdot u_{l+1} = \sum_{i=1}^{l} a_{l+1} \cdot c_i \cdot u_i$, that is,

$$f(u_{l+1}) = \sum_{i=1}^{l} a_{l+1} \cdot c_i \cdot u_i = \sum_{i=1}^{l} c_i \cdot f(u_i) = \sum_{i=1}^{l} c_i \cdot a_i \cdot u_i.$$

By subtracting the second and the fourth expression in the equalities we obtain $0 = \sum_{i=1}^{l} (a_{l+1} - a_i) \cdot c_i \cdot u_i$. All the differences between the eigenvalues are nonzero and at least one coefficient c_i is nonzero. This is a contradiction with the assumed linear independence $u_1, \dots, u_\ell$, therefore also the vector u_{l+1} must be linearly independent of the others. □

The latter theorem can be seen as a decomposition of a linear mapping f into a sum of much simpler mappings. If there are $n = \dim V$ distinct eigenvalues λ_i, we obtain the entire V as a direct sum of one-dimensional eigenspaces V_{λ_i}. Each of them then describes a projection on this invariant one-dimensional subspace, where the mapping is given just as multiplication by the eigenvalue λ_i. Furthermore, this decomposition can be easily calculated:

Now we express it in the given basis. For this we need the matrix T for changing the basis from the standard basis to the new basis. This can be obtained by writing the coordinates of the vectors of the original basis under the new basis into the columns of the matrix T. But we shall do it in a different way – we obtain first the matrix for changing the basis from the new one to the original one, that is, the matrix T^{-1}. We just write the coordinates of the vectors of the new basis into the columns:

$$T^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Then

$$T = {T^{-1}}^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ -2 & 1 & 3 \end{pmatrix},$$

and for the matrix B of a mapping under new basis we have (see 2.3.16)

$$B = TAT^{-1} = \begin{pmatrix} 0 & 5 & 2 \\ 0 & -2 & -1 \\ 0 & 14 & 6 \end{pmatrix}.$$

□

You can find more exercises on computing with eigenvalues and eigenvectors on the page 67.

In the case of a 3×3 matrix, you can use this special formula to find its characteristic polynomial:

2.G.3. For any $n \times n$ matrix A its characteristic polynomial $|\,A - \lambda\, E\,|$ is of degree n, that is, it is of the form

$$|\,A - \lambda\, E\,| = c_n\, \lambda^n + c_{n-1}\, \lambda^{n-1} + \cdots + c_1\, \lambda + c_0, \quad c_n \neq 0,$$

while we have

$$c_n = (-1)^n, \quad c_{n-1} = (-1)^{n-1} \operatorname{tr} A, \quad c_0 = |\,A\,|.$$

If the matrix A is three-dimensional, we obtain

$$|\,A - \lambda\, E\,| = -\lambda^3 + (\operatorname{tr} A)\, \lambda^2 + c_1\, \lambda + |\,A\,|.$$

By choosing $\lambda = 1$ we obtain

$$|\,A - E\,| = -1 + \operatorname{tr} A + c_1 + |\,A\,|.$$

From there we obtain

$$|\,A - \lambda\, E\,| = \\ -\lambda^3 + (\operatorname{tr} A)\, \lambda^2 + (|\,A - E\,| + 1 - \operatorname{tr} A - |\,A\,|)\, \lambda + |\,A\,|.$$

Use this expression for determining the characteristic polynomial and the eigenvalues of the matrix

$$A = \begin{pmatrix} 32 & -67 & 47 \\ 7 & -14 & 13 \\ -7 & 15 & -6 \end{pmatrix}.$$

○

Basis of eigenvectors

Corollary. *If there exist n mutually distinct roots λ_i of the characteristic polynomial of the mapping $f : V \to V$ on the n-dimensional space V, then there exists a decomposition of V into a direct sum of eigenspaces each of dimension 1. This means that there exists a basis for V consisting only of eigenvectors and in this basis the matrix for f is the diagonal matrix with the eigenvalues on the diagonal. This basis is uniquely determined up to the order of the elements and scale of the vectors.*

The corresponding basis (expressed in the coordinates in an arbitrary basis of V) is obtained by solving n systems of homogeneous linear equations of n variables with matrices $(A - \lambda_i \cdot E)$, where A is the matrix of f in a chosen basis.

2.4.4. Invariant subspaces. We have seen that every eigenvector v of the mapping $f : V \to V$ generates a subspace $\operatorname{span}\{v\} \subset V$, which is preserved by the mapping f.

More generally, we say that a vector subspace $W \subset V$ is an *invariant subspace* for a linear mapping f, if $f(W) \subset W$.

If V is a finite dimensional vector space and we choose some basis $(u_1, \dots, u_k)$ of a subspace W, we can always extend it to be a basis $(u_1, \dots, u_k, u_{k+1}, \dots, u_n)$ for the whole space V. For every such basis, the mapping will have a matrix A of the form

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix} \tag{1}$$

where B is a square matrix of dimension k, D is a square matrix of dimension $n - k$ and C is a matrix of the type $n/(n-k)$. On the other hand, if for some basis $(u_1, \dots, u_n)$ the matrix of the mapping f is of the form (1), then $W = \operatorname{span}\{u_1, \dots, u_k\}$ is invariant under the mapping f.

By the same arguments, the mapping with the matrix A as in (1) leaves the subspace $\operatorname{span}\{u_{k+1}, \dots, u_n\}$ invariant, if and only if the submatrix C is zero.

From this point of view the eigenspaces of the mapping are special cases of invariant subspaces. Our next task is to find some conditions under which there are invariant complements of invariant subspaces.

2.4.5. We illustrate some typical properties of mappings on the spaces $\mathbb{R}^3$ and $\mathbb{R}^2$ in terms of eigenvalues and eigenvectors.

(1) Consider the mapping given in the standard basis by the matrix A

$$f : \mathbb{R}^3 \to \mathbb{R}^3, \ A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

2.G.4. Find the orthonormal complement of the vectorspace spaned by the vectors $(2,1,3)$, $(3,16,7)$, $(3,5,4)$, $(-7,7,-10)$.

Solution. In fact the task consists of solving the system 2.A.3, which we have done already. $\square$

2.G.5. Pauli matrices. In physics, the state of a particle with spin $\frac{1}{2}$ is described with Pauli matrices. They are the 2×2 matrices over complex numbers:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

For square matrices we define their *commutator* (denoted by square brackets) as $[\sigma_1, \sigma_2] := \sigma_1\sigma_2 - \sigma_2\sigma_1$

Show that $[\sigma_1, \sigma_2] = 2i\sigma_3$ and similarly $[\sigma_1, \sigma_3] = 2i\sigma_2$ and $[\sigma_2, \sigma_3] = 2i\sigma_1$. Furthermore, show that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ and that the eigenvalues of the matrices σ_1, σ_2, σ_3 are ± 1.

Show that for matrices describing the state of the particle with spin 1, namely

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

, the commuting relations are the same as in the case of Pauli matrices.

Equivalently it can be shown that under the notation $1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I := i\sigma_3, J := i\sigma_2, K := i\sigma_1$ forms the vector space with basis $(1, I, J, K)$ of an algebra of quaternions (the algebra is a vector space with binary bilinear operation of multiplication, in this case the multiplication is given by matrix multiplication). In order for the vector space to be an algebra of quaternions it is necessary and sufficient to show the following properties: $I^2 = J^2 = K^2 = -1$ and $IJ = -JI = K$, $JK = -KJ = I$ and $KI = -IK = J$.

2.G.6. Can the matrix

$$B = \begin{pmatrix} 5 & 6 \\ 6 & 5 \end{pmatrix}$$

be expressed in the form of the product $B = P^{-1} \cdot D \cdot P$ for some diagonal matrix D and invertible matrix P? If possible, give an example of such matrices D, P, and find out how many such pairs there are.

Solution. The matrix B has two distinct eigenvalues, and thus such an expression exists. For instance it holds that

$$\begin{pmatrix} 5 & 6 \\ 6 & 5 \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{pmatrix} \cdot \begin{pmatrix} 11 & 0 \\ 0 & -1 \end{pmatrix} \cdot \tfrac{1}{2}\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix}.$$

We compute

$$|A - \lambda E| = \begin{vmatrix} -\lambda & 0 & 1 \\ 0 & 1-\lambda & 0 \\ 1 & 0 & -\lambda \end{vmatrix} = -\lambda^3 + \lambda^2 + \lambda - 1,$$

with roots $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = -1$. The eigenvectors with eigenvalue $\lambda = 1$ can be computed:

$$\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

with the basis of the space of solutions, that is, of all eigenvectors with this eigenvalue

$$u_1 = (0,1,0), \quad u_2 = (1,0,1).$$

Similarly for $\lambda = -1$ we obtain the third independent eigenvector

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow u_3 = (-1,0,1).$$

Under the basis u_1, u_2, u_3 (note that u_3 must be linearly independent of the remaining two because of the previous theorem and u_1, u_2 were obtained as two independent solutions) f has the diagonal matrix

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

The whole space $\mathbb{R}^3$ is a direct sum of eigenspaces, $\mathbb{R}^3 = V_1 \oplus V_2$, with $\dim V_1 = 2$, and $\dim V_2 = 1$. This decomposition is uniquely determined and says much about the geometric properties of the mapping f. The eigenspace V_1 is furthermore a direct sum of one-dimensional eigenspaces, which can be selected in other ways (thus such a decomposition has no further geometrical meaning).

(2) Consider the linear mapping $f : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ defined by polynomial differentiation, that is, $f(1) = 0$, $f(x) = 1$, $f(x^2) = 2x$. The mapping f thus has in the usual basis $(1, x, x^2)$ the matrix

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

The characteristic polynomial is $|A - \lambda \cdot E| = -\lambda^3$, thus it has only one eigenvalue, $\lambda = 0$. We compute the eigenvectors:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The space of the eigenvectors is thus one-dimensional, generated by the constant polynomial 1.

The striking property of this mapping is that is no basis for which the matrix would be diagonal. There is the "chain" of vectors mapping four independent generators as follows: $\frac{1}{2}x^2 \mapsto x \mapsto 1 \mapsto 0$ builds a sequence of subspaces without invariant complements.

There exist exactly two diagonal matrices D:

$$\begin{pmatrix} 11 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 0 \\ 0 & 11 \end{pmatrix},$$

but the columns of the matrix P^{-1} can be substituted with their arbitrary non-zero scalar multiples, thus there are infinitely many pairs D, P. □

As we have already seen in 2.G.2, based on the eigenvalues and eigenvectors of the given 3×3 matrix, we can often interpret geometrically the mapping it induces in $\mathbb{R}^3$. In particular, we notice that can do so in the following situations: If the matrix has 0 as eigenvalue and 1 as an eigenvalue with geometric multiplicity 2, then it is a projection in the direction of the eigenvector associated with the eigenvalue 0 on the plane given by the eigenspace of the eigenvalue 1. If the eigenvector associated with 0 is perpendicular to that plane, then the mapping is an orthogonal projection.

If the matrix has eigenvalue -1 with the eigenvector perpendicular to the plane of the eigenvectors associated with the eigenvalue 1, then it is a mirror symmetry through the plane of the eigenvectors associated with 1.

If the matrix has eigenvalue 1 with an eigenvector perpendicular to plane of the eigenvectors associated with the eigenvalue -1, then it is an axial symmetry (in space) through the axis given by the eigenvector associated with 1.

2.G.7. Determine what linear mapping $\mathbb{R}^3 \to \mathbb{R}^3$ is given by the matrix

$$\begin{pmatrix} -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{4}{3} & -\frac{7}{3} & -\frac{8}{3} \\ -1 & 1 & 1 \end{pmatrix}$$

Solution. The matrix has a double eigenvalue -1, its associated eigenspace is $\operatorname{span}\{(2,0,1), (1,1,0)\}$. Further, the matrix has 0 as the eigenvalue, with eigenvector $(1,4,-3)$. The mapping given by this matrix under the standard basis is then an axial symmetry through the line given by the last vector composed with the projection on the plane perpendicular to the last vector, that is, given by the equation $x+4y-3z=0$. □

2.G.8. The theorem 2.4.7 gives us tools for recognising a matrix of a rotation in $\mathbb{R}^3$. It is orthogonal (rows orthogonal to each other equivalently the same for the columns). It has three distinct eigenvalues with absolute value 1. One of them is the number 1 (its associated eigenvector is the axis of the rotation). The argument of the remaining two, which are necessarily complex conjugates, gives the angle of the rotation

2.4.6. Orthogonal mappings. We consider the special case

of the mapping $f : V \to W$ between spaces with scalar products, which preserve lengths for all vectors $u \in V$.

Orthogonal mappings

A linear mapping $f : V \to W$ between spaces with scalar product is called an *orthogonal mapping*, if for all $u \in V$

$$\langle f(u), f(u)\rangle = \langle u, u\rangle.$$

The linearity of f and the symmetry of the scalar product imply that for all pairs of vectors the following equality holds:

$$\langle f(u+v), f(u+v)\rangle = \langle f(u), f(u)\rangle + \langle f(v), f(v)\rangle + 2\langle f(u), f(v)\rangle.$$

Therefore all orthogonal mappings satisfy also the seemingly stronger condition for all vectors $u, v \in V$:

$$\langle f(u), f(v)\rangle = \langle u, v\rangle,$$

i.e. the mapping f leaves the scalar product invariant if and only if it leaves invariant the length of the vectors. (We should have noticed that this is true for all fields of scalars, where $1+1 \neq 0$, but it does hold true for $\mathbb{Z}_2$.)

In the initial discussion about the geometry in the plane we proved in the Theorem **??** that a linear mapping $\mathbb{R}^2 \to \mathbb{R}^2$ preserves lengths of the vectors if and only if its matrix in the standard basis (which is orthonormal with respect to the standard scalar product) satisfies $A^T \cdot A = E$, that is, $A^{-1} = A^T$.

In general, orthogonal mappings $f : V \to W$ must be always injective, because the condition $\langle f(u), f(u)\rangle = 0$ implies $\langle u, u\rangle = 0$ and thus $u = 0$. In such a case, the dimension of the range is always at least as large as the dimension of the domain of f. But then both dimensions are equal and $f : V \to \operatorname{Im} f$ is a bijection. If $\operatorname{Im} f \neq W$, we extend the orthonormal basis of the image of f to an orthonormal basis of the range space and the matrix of the mapping then contains a square regular submatrix A along with zero rows so that it has the required number of rows. Without loss of generality we can assume that $W = V$.

Our condition for the matrix of an orthogonal mapping in any orthonormal basis requires that for all vectors x and y in the space $\mathbb{K}^n$:

$$(A \cdot x)^T \cdot (A \cdot y) = x^T \cdot (A^T \cdot A) \cdot y = x^T \cdot y.$$

Special choice of the standard basis vectors for x and y yields directly $A^T \cdot A = E$, that is, the same result as for dimension two. Thus we have proved the following theorem:

Matrix of orthogonal mappings

Theorem. *Let V be a real vector space with scalar product and let $f : V \to V$ be a linear mapping. Then f is orthogonal if and only if in some orthogonal basis (and then consequently in all of them) its matrix A satisfies $A^T = A^{-1}$.*

in the positive sense in the plane given by the basis $u_\lambda + \overline{u_\lambda}$, $i(u_\lambda - \overline{u_\lambda})$.

2.G.9. Determine what linear mapping is given by the matrix

$$\begin{pmatrix} \frac{3}{5} & \frac{16}{25} & \frac{-12}{25} \\ \frac{-16}{25} & \frac{93}{125} & \frac{24}{125} \\ \frac{12}{25} & \frac{24}{125} & \frac{107}{125} \end{pmatrix}.$$

Solution. First we notice, that the matrix is orthogonal (rows are mutually orhogonal, and equivalently the same with columns). The matrix has the following eigenvalues and corresponding eigenvectors: 1, $v_1 = (0, 1, \frac{4}{3})$; $\frac{3}{5} + \frac{4}{5}i, v_2 = (1, \frac{4}{5}i, -\frac{3}{5}i)$; $\frac{3}{5} - \frac{4}{5}i, v_3 = (1, -\frac{4}{5}i, \frac{3}{5}i)$. All three eigenvalues have absolute value one, which together with the observation of orthogonality tells us that the matrix is a matrix of rotation. Its axis is given by the eigenvector corresponding to the eigenvalue 1, that is the vector $(0, 1, \frac{4}{3})$. The plane of rotation is the real plane in $\mathbb{R}^3$, which is given by the intersection of two dimensional complex space in $\mathbb{C}^3$ generated by the remaining eigenvectors with $\mathbb{R}^3$. It is the plane $\operatorname{span}\{(1,0,0),(0,-4,3)\}$ (the first generator is the (real multiple of) $v_2 + v_3$, the other one is the (real multiple of) $i(v_2 - v_3)$, see 2.4.7). We can determine the rotation angle in this plane, It is a rotation by the angle $\arccos(\frac{3}{5}) \doteq 0,295\pi$, which is the argument of the eigenvalue $\frac{3}{5} + \frac{4}{5}i$ (or minus that number, if we would choose the other eigenvalue).

It remains to determine the direction of the rotation. First, recall that the meaning of the direction of the rotation changes when we change the orientation of the axis (it has no meaning to speak of the direction of the rotation if we do not have an orientation of the axis). Using the ideas from the proof of the theorem 2.4.7, we see that the given matrix acts by rotating by $\arccos(\frac{3}{5}))$ in the positive sense in the plane given by the basis $((1,0,0),(0,-\frac{4}{5},\frac{3}{5}))$. The first vector of the basis is the imaginary part of the eigenvector associated with the eigenvalue $\frac{3}{5} + \frac{4}{5}i$, the second is then the (common) real part of the eigenvectors associated with the complex eigenvalues. The order of the vectors in the basis is important (by changing their order the meaning of the direction changes). The axis of rotation is perpendicular to the plane. If we orient using the right-hand rule (the perpendicular direction is obtained by taking the product of the vectors in the basis) then the direction of the rotation agrees with the direction of rotation in the plane with the given basis. In our case we obtain by the vector product $(0,1,-1) \times (1,1,-1) = (0,-1,-1)$. It is

Proof. Indeed, if f preserves lengths, it must have the claimed property in every orthonormal basis. On the other hand, the previous calculations show that this property for the matrix in one such basis ensures length preservation. □

Square matrices which satisfy the equality $A^T = A^{-1}$ are called *orthogonal matrices*.

The shape of the coordinate transition matrices between

orthonormal bases is a direct corollary of the above theorem. Each such matrix must provide a mapping $\mathbb{K}^n \to \mathbb{K}^n$ which preserves lengths and thus satisfies the condition $S^{-1} = S^T$. When changing from one orthonormal basis to another one, the matrix of any linear mapping changes according to the relation

$$A' = S^T A\, S.$$

2.4.7. Decomposition of an orthogonal mapping. We take a more detailed look at eigenvectors and eigenvalues of orthogonal mappings on a real vector space V with scalar product.

Consider a fixed orthogonal mapping $f : V \to V$ with the matrix A in some orthonormal basis. We continue as with the matrix D of rotation in 2.4.1.

We think first about invariant subspaces of orthogonal mappings and their orthogonal complements. Namely, given any subspace $W \subset V$ invariant with respect to an orthogonal mapping $f : V \to V$, then for all $v \in W^\perp$ and $w \in W$ we immediately see

$$\langle f(v), w\rangle = \langle f(v), f \circ f^{-1}(w)\rangle = \langle v, f^{-1}(w)\rangle = 0$$

since $f^{-1}(w) \in W$, too. But this means that also $f(W^\perp) \subset W^\perp$ and we have proved a simple but very important proposition:

Proposition. *The orthogonal complement of a subspace invariant with respect to an orthogonal mapping is also invariant.*

If all eigenvalues of an orthogonal mapping are real, this

claim ensures that there always exists a basis of V composed of eigenvectors. Indeed, the restriction of f to the orthogonal complement of an invariant subspace is again an orthogonal mapping, therefore we can add one eigenvector to the basis after another, until we obtain the whole decomposition of V. However, mostly the eigenvalues of orthogonal mappings are not real. We need to deviate into complex vector spaces. We formulate the result right away:

thus a rotation through $\arccos(\frac{3}{5})$ in the positive sense about the vector $(0, -1, -1)$, that is, a rotation through $\arccos(\frac{3}{5})$ in the negative sense about the vector $(0, 1, 1)$. $\square$

2.G.10. Determine what linear mapping is given by the matrix

$$\begin{pmatrix} \frac{-1}{5} & \frac{3}{5} & \frac{-1}{5} \\ \frac{-8}{5} & \frac{9}{5} & \frac{2}{5} \\ \frac{8}{5} & \frac{-4}{5} & \frac{3}{5} \end{pmatrix}.$$

Solution. By already known method we find out that the matrix has the following eigenvalues and corresponding eigenvectors: $1, (1, 2, 0)$; $\frac{3}{5} + \frac{4}{5}i, 1, (1, 1 + i, -1 - i)$; $\frac{3}{5} - \frac{4}{5}i, (1, 1 - i, -1 + i)$. Though all three eigenvectors have absolute value 1, they are not orthogonal to each other, thus the matrix is not orthogonal. Consequently it is not a matrix of rotation. Nevertheless, it is a linear mapping which is "close" to a rotation. It is a rotation in the plane given by two complex eigenvectors (but this plane is not orthogonal to the vector $(1, 2, 0)$, but it is preserved by the map). It remains to determine the direction of the rotation. First, we should recall that the meaning of the direction of the rotation changes when we change the orientation of the axis (it has no meaning to speak of the direction of the rotation if we do not have an orientation of the axis).

Using the same ideas as in the previous example, we see that the given matrix acts by rotating by $\arccos(\frac{3}{5}))$ in the positive sense in the plane given by the basis $((1, 1, -1), (0, 1,).$ The first vector of the basis is the imaginary part of the eigenvector associated with the eigenvalue $\frac{3}{5} + \frac{4}{5}i$, the second is then the (common) real part of the eigenvectors associated with the complex eigenvalues. The order of the vectors in the basis is important (by changing their order the meaning of the direction changes). The "axis" of rotation is not perpendicular to the plane, but we can orient the vectors lying in the whole half-plane using the right-hand rule (the perpendicular direction is obtained by taking the product of the vectors in the basis) then the direction of the rotation agrees with the direction of rotation in the plane with the given basis. In our case we obtain by the vector product $(0, 1, -1) \times (1, 1, -1) = (0, -1, -1)$. It is thus a rotation through $\arccos(\frac{3}{5})$ in the positive sense about the vector $(0, -1, -1)$, that is, a rotation through $\arccos(\frac{3}{5})$ in the negative sense about the vector $(0, 1, 1)$. $\square$

Orthogonal mapping decomposition

Theorem. *Let $f : V \to V$ be an orthogonal mapping on a real vector space V with scalar product. Then all the (in general complex) roots of the characteristic polynomial f have length one. There exists the decomposition of V into one-dimensional eigenspaces corresponding to the real eigenvalues $\lambda = \pm 1$ and two-dimensional subspaces $P_{\lambda, \bar{\lambda}}$ with $\lambda \in \mathbb{C} \backslash \mathbb{R}$, where f acts by the rotation by the angle equal to the argument of the complex number λ in the positive sense. All these subspaces are mutually orthogonal.*

Proof. Without loss of generality we can work with the space $V = \mathbb{R}^m$ with the standard scalar product. The mapping is thus given by an orthogonal matrix A which can be equally well seen as the matrix of a (complex) linear mapping on the complex space $\mathbb{C}^m$ (which just happens to have all of its coefficients real).

There exist exactly m (complex) roots of the characteristic polynomial of A, counting their algebraic multiplicities (see the fundamental theorem of algebra, **??**). Furthermore, because the characteristic polynomial of the mapping has only real coefficients, the roots are either real or there are a pair of roots which are complex conjugates λ and $\bar{\lambda}$. The associated eigenvectors in $\mathbb{C}^m$ for such pairs of complex conjugates are actually solutions of two systems of linear homogeneous equations which are also complex conjugate to each other – the corresponding matrices of the systems have real components, except for the eigenvalues λ. Therefore the solutions of this systems are also complex conjugates (check this!).

Next, we exploit the fact that for every invariant subspace its orthogonal complement is also invariant. First we find the eigenspaces $V_{\pm 1}$ associated with the real eigenvalues, and restrict the mapping to the orthogonal complement of their sum. Without loss of generality we can thus assume that our orthogonal mapping has no real eigenvalues and that $\dim V = 2n > 0$.

Now choose an eigenvalue λ and let u_λ be the eigenvector in $\mathbb{C}^{2n}$ associated to the eigenvalue $\lambda = \alpha + i\beta$, $\beta \neq 0$. Analogously to the case of rotation in the plane discussed in paragraph 2.4.1 in terms of the matrix D, we are interested in the real part of the sum of two one-dimensional (complex) subspaces $W = \operatorname{span}\{u_\lambda\} \oplus \operatorname{span}\{\bar{u}_\lambda\}$, where $\bar{u}_\lambda$ is the eigenvector associated to the conjugated eigenvalue $\bar{\lambda}$.

Now we want the intersection of the 2-dimensional complex subspace W with the real subspace $\mathbb{R}^{2n} \subset \mathbb{C}^{2n}$, which is clearly generated (over $\mathbb{R}$) by the vectors $u_\lambda + \bar{u}_\lambda$ and $i(u_\lambda - \bar{u}_\lambda)$. We call this real 2-dimensional subspace $P_{\lambda, \bar{\lambda}} \subset \mathbb{R}^{2n}$ and notice, this subspace is generated by the basis given by the real and imaginary part of u_λ

$$x_\lambda = \operatorname{Re} u_\lambda, \quad -y_\lambda = -\operatorname{Im} u_\lambda.$$

Because $A \cdot (u_\lambda + \bar{u}_\lambda) = \lambda u_\lambda + \bar{\lambda}\bar{u}_\lambda$ and similarly with the second basis vector, it is clearly an invariant subspace with

2.G.11. Without any computation, write down the spectrum of the linear mapping $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by $(x_1, x_2, x_3) \mapsto (x_1 + x_3, x_2, x_1 + x_3)$. ○

2.G.12. Find the dimension of the eigenspaces of the eigenvalues λ_i of the matrix

$$\begin{pmatrix} 4 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 5 & 2 & 3 & 0 \\ 0 & 4 & 0 & 3 \end{pmatrix}.$$

○

respect to multiplication by the matrix A and we obtain

$$A \cdot x_\lambda = \alpha x_\lambda + \beta y_\lambda, \; A \cdot y_\lambda = -\alpha y_\lambda + \beta x_\lambda.$$

Because our mapping preserves lengths, the absolute value of the eigenvalue λ must equal one. But that means that the restriction of our mapping to $P_{\lambda,\bar{\lambda}}$ is the rotation by the argument of the eigenvalue λ. Note that the choice of the eigenvalue $\bar{\lambda}$ instead of λ leads to the same subspace with the same rotation, we would just have expressed it in the basis x_λ, y_λ, that is, the same rotation will in these coordinates go by the same angle, but with the opposite sign, as expected.

The proof of the whole theorem is completed by restricting the mapping to the orthogonal complement and finding another 2-dimensional subspace, until we get the required decomposition. □

We return to the ideas in this proof once again in chapter three, where we study complex extensions of the Euclidean vector spaces, see **??**.

Remark. The previous theorem is very powerful in dimension three. Here at least one eigenvalue must be real ± 1, since three is odd. But then the associated eigenspace is an axis of the rotation of the three-dimensional space through the angle given by the argument of the other eigenvalues. Try to think how to detect in which direction the space is rotated. Note also that the eigenvalue -1 means an additional reflection through the plane perpendicular to the axis of the rotation.

We shall return to the discussion of such properties of matrices and linear mappings in more details at the end of the next chapter, after illustrating the power of the matrix calculus in several practical applications. We close this section with a general quite widely used definition:

Spectrum of linear mapping

2.4.8. Definition. The *spectrum of a linear mapping* $f : V \to V$, or the spectrum of a square matrix A, is a sequence of roots of the characteristic polynomial f or A, along with their multiplicities, respectively. The *algebraic multiplicity* of an eigenvalue means the multiplicity of the root of the characteristic polynomial, while the *geometric multiplicity* of the eigenvalue is the dimension of the associated subspace of eigenvectors.

The *spectral diameter* of a linear mapping (or matrix) is the greatest of the absolute values of the eigenvalues.

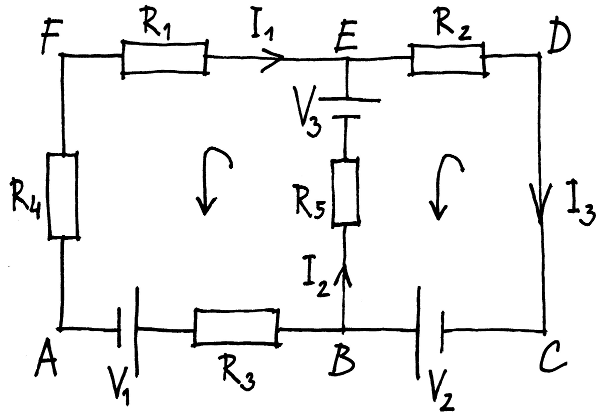
In this terminology, our results about orthogonal mappings can be formulated as follows: the spectrum of an orthogonal mapping is always a subset of the unit circle in the complex plane. Thus only the values ± 1 may appear in the real part of the spectrum and their algebraic and geometric multiplicities are always the same. Complex values of the spectrum then correspond to rotations in suitable two-dimensional subspaces which are mutually perpendicular.

H. Additional exercises for the whole chapter

2.H.1. Kirchhoff's Circuit Laws. We consider an application of Linear Algebra to analysis of electric circuits, using Ohm's law and Kirchhoff's voltage and current laws.

Consider an electric circuit as in the figure and write down the values of the currents there if you know the values $V_1 = 20, \quad V_2 = 120, \quad V_3 = 50, \quad R_1 = 10, \quad R_2 = 30, \quad R_3 = 4, \quad R_4 = 5, \quad R_5 = 10,$

Notice that the quantities I_i denote the electric currents, while R_j are resistances, and V_k are voltages.



Solution. There are two closed loops, namely $ABEF$ and $EBCD$ and two branching vertices B and E of degree no less than 3. On every segment of the circuit, bounded by branching points, the electric current is constant. Set it to be I_1 on the segment $EFAB$, I_2 on EB, and I_3 on $BCDE$.

Applying Kirchhoff's current law to branching points B and E we obtain: $I_1 + I_2 = I_3$ and $I_3 - I_1 = I_2$, which are, of course the same equations. In case there are many branching vertices, we write all Kirchhhoff's Current Law equations to the system, having at least one of those equations redundant.

Choose the counter clockwise orientations of the loops $ABEF$ and $EBCD$. Applying Kirchhoff Voltage Law and Ohm's Law to the loop $ABEF$ we obtain the equation:

$$V_1 + I_1R_3 - I_2R_5 + V_3 + I_1R_1 + I_1R_4 = 0.$$

Similarly, the loop $EBCD$ implies

$$-V_2 + I_3R_2 - V_3 + R_5I_2 = 0.$$

By combining all equations, we obtain the system

$$\begin{array}{rcrcrcr} I_1 & + & I_2 & - & I_3 & = & 0, \\ (R_3+R_1+R_4)I_1 & - & R_5I_2 & + & & = & -V_1 - V_3, \\ & & R_5I_2 & + & R_2I_3 & = & V_2 + V_3. \end{array}$$

Substituing the prescribed values we obtain the linear system

$$\begin{array}{rcrcrcr} I_1 & + & I_2 & - & I_3 & = & 0, \\ 19I_1 & - & 10I_2 & + & & = & -70, \\ & & 10I_2 & + & 30I_3 & = & 170. \end{array}$$

This has solutions $I_1 = -\frac{80}{53} \approx -1.509, \quad I_2 = \frac{219}{53} \approx 4.132, \quad I_3 = \frac{139}{53} \approx 2.623.$ □

2.H.2. The general case. In general, the method for electrical circuit analysis can be formulated along the following steps:

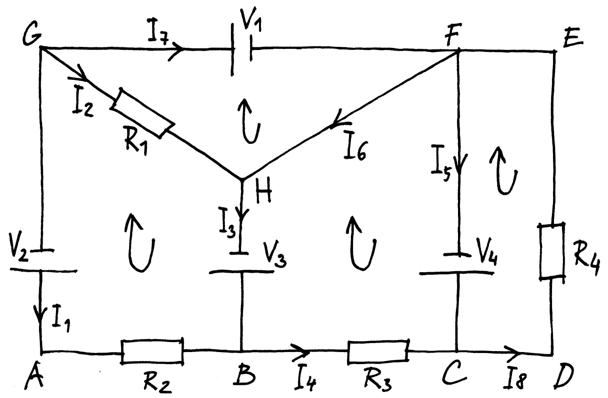
i) Identify all branching vertices of the circuit, i.e vertices of degree no less than 3;
ii) Identify all closed loops of the circuit;
iii) Introduce variables I_k, denoting oriented currents on each segment of the circuit between two branching vertices;

iv) Write down Kirchhoff's current conservation law for each branching vertex. The total incoming current equals the total outgoing current;

v) Choose an orientation on every closed loop of the circuit and write down Kirchhoff's voltage conservation law according to the chosen orientation. If you find an electric charge of voltage V_j and you go from the short bar to the long bar, the contribution of this charge is V_j. It is $-V_j$ if you go from the long bar to the short one. If you go in the positive direction of a current I and find a resistor with resistance R_j, the contribution is $-R_jI$, and it is R_jI if the orientation of the loop is opposite to the direction of the current I. The total voltage change along each closed loop must be zero.

vi) Compose the system of linear equations collecting all equations, representing Kirchhoff's current and voltage laws and solve it with respect to the variables, representing currents. Notice that some equations may be redundant, however, the solution should be unique.

To illustrate this general approach, consider the circuit example in the diagram.



Solution.

i) The set of branching vertices is $\{B, C, F, G, H\}$.

ii) The set of closed loops is $\{ABHG, FHBC, GHF, CDEF\}$.

iii) Let I_1 be the current on the segment GAB, I_2 on the segment GH, I_3 on the segment HB, I_4 on the segment BC, I_5 on the segment FC, I_6 on the segment FH, I_7 on GF, and I_8 on $CDEF$.

iv) Write Kirchhoff's current conservation laws for the branching vertices:
   - vertex B: $I_1 + I_3 = I_4$
   - vertex C: $I_4 + I_5 = I_8$
   - vertex F: $I_8 = I_5 + I_6 - I_7$
   - vertex G: $-I_7 = I_1 + I_2$
   - vertex H: $I_2 + I_6 = I_3$

v) Write Kirchhoff's voltage conservation for each of the closed loops traversed counter-clockwise:
   - loop $ABHG$: $-R_1I_2 + V_3 + R_2I_1 - V_2 = 0$
   - loop $FHBC$: $V_4 + R_3I_4 - V_3 = 0$
   - loop GHF: $R_1I_2 - V_1 = 0$
   - loop $CDEF$: $R_4I_8 - V_4 = 0$

   Set the parameters: $R_1 = 4, \quad R_2 = 7, \quad R_3 = 9, \quad R_4 = 12, \quad V_1 = 10, \quad V_2 = 20, \quad ,V_3 = 60, \quad ,V_4 = 120$, to obtain the system

$$I_1 + I_3 - I_4 = 0$$
$$I_4 + I_5 - I_87 = 0$$
$$I_5 + I_6 - I_7 - I_8 = 0$$
$$I_1 + I_2 + I_7 = 0$$

$$\begin{aligned} I_2 - I_3 + I_6 &= 0 \\ 7I_1 - 4I_2 &= -40 \\ 9I_4 &= -60 \\ 4I_2 &= 10 \\ 12I_8 &= 120 \end{aligned}$$

with the solution set $I_1 = \frac{-30}{7}, \quad I_2 = \frac{5}{2}, \quad I_3 = \frac{-50}{21}, \quad I_4 = \frac{-20}{3}, \quad I_5 = \frac{50}{3}, \quad I_6 = \frac{-205}{42}, \quad I_7 = \frac{25}{14}, \quad I_8 = 10.$

□

2.H.3. Solve the system of equations

$$\begin{array}{rcrcrcrcrcr} x_1 & + & x_2 & + & x_3 & + & x_4 & - & 2x_5 & = & 3, \\ & & 2x_2 & + & 2x_3 & + & 2x_4 & - & 4x_5 & = & 5, \\ -x_1 & - & x_2 & - & x_3 & + & x_4 & + & 2x_5 & = & 0, \\ -2x_1 & + & 3x_2 & + & 3x_3 & & & - & 6x_5 & = & 2. \end{array}$$

Solution. The extended matrix of the system is

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -2 & 3 \\ 0 & 2 & 2 & 2 & -4 & 5 \\ -1 & -1 & -1 & 1 & 2 & 0 \\ -2 & 3 & 3 & 0 & -6 & 2 \end{array} \right).$$

Adding the first row to the third, adding its 2-multiple to the fourth, and adding the $(-5/2)$-multiple of the second to the fourth we obtain

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -2 & 3 \\ 0 & 2 & 2 & 2 & -4 & 5 \\ 0 & 0 & 0 & 2 & 0 & 3 \\ 0 & 5 & 5 & 2 & -10 & 8 \end{array} \right) \sim \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -2 & 3 \\ 0 & 2 & 2 & 2 & -4 & 5 \\ 0 & 0 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & -3 & 0 & -9/2 \end{array} \right).$$

The last row is clearly a multiple of the previous, and thus we can omit it. The pivots are located in the first, second and fourth. Thus the free variables are x_3 and x_5 which we substitute by the real parameters t and s. Thus we consider the system

$$\begin{array}{rcrcrcrcrcr} x_1 & + & x_2 & + & t & + & x_4 & - & 2s & = & 3, \\ & & 2x_2 & + & 2t & + & 2x_4 & - & 4s & = & 5, \\ & & & & & & 2x_4 & & & = & 3. \end{array}$$

We see that $x_4 = 3/2$. The second equation gives

$$2x_2 + 2t + 3 - 4s = 5, \quad \text{that is,} \quad x_2 = 1 - t + 2s.$$

From the first we have

$$x_1 + 1 - t + 2s + t + 3/2 - 2s = 3, \quad \text{tj.} \quad x_1 = 1/2.$$

Altogether,

$$(x_1, x_2, x_3, x_4, x_5) = (1/2, 1{-}t{+}2s, t, 3/2, s), \quad t, s \in \mathbb{R}.$$

Alternatively, we can consider the extended matrix and transform it using the row transformations into the row echelon form. We arrange it so that the first non-zero number in every row is 1, and the remaining numbers in the column containing this 1 are 0. We omit the fourth equation, which is a combination of the first three. Sequentially, multiplying the second and the third row by the number $1/2$, subtracting the third row from the second and from the first and by subtracting the second row from the first we obtain

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -2 & 3 \\ 0 & 2 & 2 & 2 & -4 & 5 \\ 0 & 0 & 0 & 2 & 0 & 3 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -2 & 3 \\ 0 & 1 & 1 & 1 & -2 & 5/2 \\ 0 & 0 & 0 & 1 & 0 & 3/2 \end{array}\right) \sim$$

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

If we choose again $x_3 = t$, $x_5 = s$ ($t, s \in \mathbb{R}$), we obtain the general solution (2.H.3) as above. □

2.H.4. Find the solution of the system of linear equations given by the extended matrix

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 & 4 \\ 0 & 5 & -4 & 3 & 1 \\ 5 & 3 & 3 & -3 & 5 \end{array}\right).$$

Solution. We transform the given extended matrix into the row echelon form. We first copy the first three rows and into the last row we write the sum of the (2)-multiple of the first and of the (-3)-multiple of the last row. By this we obtain

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 & 4 \\ 0 & 5 & -4 & 3 & 1 \\ 5 & 3 & 3 & -3 & 5 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 5 & -4 & 3 & 1 \\ 0 & 6 & 1 & 14 & 0 \end{array}\right).$$

Copying the first two rows and adding a 5-multiple of the second row to the 3-multiple of the third and its 2-multiple to the fourth gives

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 5 & -4 & 3 & 1 \\ 0 & 6 & 1 & 14 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & -17 & -1 & 33 \\ 0 & 0 & -1 & 10 & 12 \end{array}\right).$$

Copying the first, second and fourth row, and adding the fourth to the third, yields

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & -17 & -1 & 33 \\ 0 & 0 & -1 & 10 & 12 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & -18 & 9 & 45 \\ 0 & 0 & -1 & 10 & 12 \end{array}\right).$$

With three more row transformations, we arrive at

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & -18 & 9 & 45 \\ 0 & 0 & -1 & 10 & 12 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & 2 & -1 & -5 \\ 0 & 0 & 1 & -10 & -12 \end{array}\right) \sim$$

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & 1 & -10 & -12 \\ 0 & 0 & 2 & -1 & -5 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & 1 & -10 & -12 \\ 0 & 0 & 0 & 19 & 19 \end{array}\right).$$

The system has exactly 1 solution. We determine it by backwards elimination

$$\left(\begin{array}{cccc|c} 3 & 3 & 2 & 1 & 3 \\ 0 & -3 & -1 & -2 & 6 \\ 0 & 0 & 1 & -10 & -12 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cccc|c} 3 & 3 & 2 & 0 & 2 \\ 0 & -3 & -1 & 0 & 8 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \sim$$

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

The solution is

$$x_1 = 4, \quad x_2 = -2, \quad x_3 = -2, \quad x_4 = 1.$$

□

2.H.5. Find all the solutions of the homogeneous system

$$x + y = 2z + v, \quad z + 4u + v = 0, \quad -3u = 0, \quad z = -v$$

of four linear equations with 5 variables x, y, z, u, v.

Solution. We rewrite the system into a matrix such that in the first column there are coefficients of x, in the second there are coefficients of y, and so on. We put all the variables in equations to the left side. By this, we obtain the matrix

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

We add $(4/3)$-multiple of the third row to the second and subtract then the second row from the fourth to obtain

$$\begin{pmatrix} 1 & 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We multiply the third row by the number $-1/3$ and add the 2-multiple of the second row to the first, which gives

$$\begin{pmatrix} 1 & 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

From the last matrix, we get immediately (reading from bottom to top) $u = 0,\ z + v = 0,\ x + y + v = 0$. Letting ?? and $v = s$ and $y = t$, the complete solution is

$$(x,\, y,\, z,\, u,\, v) = (-t - s,\, t,\, -s,\, 0,\, s)\,, \quad t, s \in \mathbb{R}.$$

which can be rewritten as

$$\begin{pmatrix} x \\ y \\ z \\ u \\ v \end{pmatrix} = t \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \quad t, s \in \mathbb{R},$$

Notice that the second and the fifth column of the matrix together form a basis for the solutions. These are the columns which do not contain a leading 1 in any of its entries. □

2.H.6. Determine the number of solutions for the systems

(a)

$$\begin{aligned} 12x_1 + \sqrt{5}x_2 + 11x_3 &= -9, \\ x_1 \qquad\qquad - 5x_3 &= -9, \\ x_1 \qquad\qquad + 2x_3 &= -7; \end{aligned}$$

(b)

$$\begin{aligned} 4x_1 + 2x_2 - 12x_3 &= 0, \\ 5x_1 + 2x_2 - x_3 &= 0, \\ -2x_1 - x_2 + 6x_3 &= 4; \end{aligned}$$

(c)

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 12x_3 & = & 0, \\ 5x_1 & + & 2x_2 & - & x_3 & = & 1, \\ -2x_1 & - & x_2 & + & 6x_3 & = & 0. \end{array}$$

Solution. The vectors $(1, 0, -5)$, $(1, 0, 2)$ are clearly linearly independent, (they are not multiples of each other) and the vector $(12, \sqrt{5}, 11)$ cannot be their linear combination (its second coordinate is non-zero). Therefore the matrix whose rows are these three linearly independent vectors (from the left side) is invertible. Thus the system for case (a) has exactly one solution.

For cases (b) and (c), it is enough to note that

$$(4, 2, -12) = -2(-2, -1, 6).$$

In case (b) adding the first equation to the third multiplied by two gives $0 = 8$, hence there is no solution for the system. In case (c) the third equation is a multiple of the first, so the system has infinitely many distinct solutions. □

2.H.7. Find a linear system, whose set of solutions is exactly

$$\{(t + 1,\ 2t,\ 3t,\ 4t);\ t \in \mathbb{R}\}.$$

Solution. Such a system is for instance

$$2x_1 - x_2 = 2, \quad 2x_2 - x_4 = 0, \quad 4x_3 - 3x_4 = 0.$$

These solutions are satisfied for every $t \in \mathbb{R}$. The vectors

$$(2, -1, 0, 0), \quad (0, 2, 0, -1), \quad (0, 0, 4, -3)$$

giving the left-hand sides of the equations are linearly independent (the set of solutions contains a single parameter). □

2.H.8. Solve the system of homogeneous linear equations given by the matrix

$$\begin{pmatrix} 0 & \sqrt{2} & \sqrt{3} & \sqrt{6} & 0 \\ 2 & 2 & \sqrt{3} & -2 & -\sqrt{5} \\ 0 & 2 & \sqrt{5} & 2\sqrt{3} & -\sqrt{3} \\ 3 & 3 & \sqrt{3} & -3 & 0 \end{pmatrix}.$$

○

2.H.9. Determine all solutions of the system

$$\begin{array}{rcrcrcrcr} & & x_2 & & & + & x_4 & = & 1, \\ 3x_1 & - & 2x_2 & - & 3x_3 & + & 4x_4 & = & -2, \\ x_1 & + & x_2 & - & x_3 & + & x_4 & = & 2, \\ x_1 & & & - & x_3 & & & = & 1. \end{array}$$

○

2.H.10. Solve

$$\begin{array}{rcrcrcrcr} 3x & - & 5y & + & 2u & + & 4z & = & 2, \\ 5x & + & 7y & - & 4u & - & 6z & = & 3, \\ 7x & - & 4y & + & & + & 3z & = & 4, \\ x & + & 6y & - & 2u & - & 5z & = & 2 \end{array}$$

○

2.H.11. Determine whether or not the system of linear equations

$$\begin{array}{rcrcrcr} 3x_1 & + & 3x_2 & + & x_3 & = & 1, \\ 2x_1 & + & 3x_2 & - & x_3 & = & 8, \\ 2x_1 & - & 3x_2 & + & x_3 & = & 4, \\ 3x_1 & - & 2x_2 & + & x_3 & = & 6 \end{array}$$

of three variables x_1, x_2, x_3 has a solution. ○

2.H.12. Determine the number of solutions of the system of 5 linear equations

$$A^T \cdot x = (1, 2, 3, 4, 5)^T,$$

where

$$x = (x_1, x_2, x_3)^T \quad \text{and} \quad A = \begin{pmatrix} 3 & 1 & 7 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 2 & 1 & 4 & 3 & 0 \end{pmatrix}.$$

Repeat the question for the system

$$A^T \cdot x = (1, 1, 1, 1, 1)^T$$

○

2.H.13. Depending on the parameter $a \in \mathbb{R}$, determine the solution of the system of linear equations

$$\begin{array}{ccccccc} ax_1 & + & 4x_2 & +2 & x_3 & = & 0, \\ 2x_1 & + & 3x_2 & - & x_3 & = & 0. \end{array}$$

○

2.H.14. Depending on the parameter $a \in \mathbb{R}$, determine the number of solutions of the system

$$\begin{pmatrix} 4 & 1 & 4 & a \\ 2 & 3 & 6 & 8 \\ 3 & 2 & 5 & 4 \\ 6 & -1 & 2 & -8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 3 \\ -3 \end{pmatrix}.$$

○

2.H.15. Decide whether or not there is a system of homogeneous linear equations of three variables whose set of solutions is exactly

(a) $\{(0, 0, 0)\}$;
(b) $\{(0, 1, 0), (0, 0, 0), (1, 1, 0)\}$;
(c) $\{(x, 1, 0);\ x \in \mathbb{R}\}$;
(d) $\{(x, y, 2y);\ x, y \in \mathbb{R}\}$.

○

2.H.16. Solve the system of linear equations, depending on the real parameters a, b.

$$\begin{aligned} x + 2y + bz &= a \\ x - y + 2z &= 1 \\ 3x - y &= 1. \end{aligned}$$

○

2.H.17. Using the inverse matrix, compute the solution of the system

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 2, \\ x_1 + x_2 - x_3 - x_4 &= 3, \\ x_1 - x_2 + x_3 - x_4 &= 3, \\ x_1 - x_2 - x_3 + x_4 &= 5. \end{aligned}$$

○

2.H.18. For what values of parameters $a, b \in \mathbb{R}$ has the system of linear equations

$$\begin{array}{ccccccc} x_1 & - & ax_2 & - & 2x_3 & = & b, \\ x_1 & + & (1-a)x_2 & & & = & b-3, \\ x_1 & + & (1-a)x_2 & + & ax_3 & = & 2b-1 \end{array}$$

(a) exactly one solution;

(b) no solution;

(c) at least 2 solutions? (i.e. infinitely many solutions)

Solution. We rewrite it, as usual, in the extended matrix, and transform:

$$\begin{pmatrix} 1 & -a & -2 & b \\ 1 & 1-a & 0 & b-3 \\ 1 & 1-a & a & 2b-1 \end{pmatrix} \sim \begin{pmatrix} 1 & -a & -2 & b \\ 0 & 1 & 2 & -3 \\ 0 & 1 & a+2 & b-1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & -a & -2 & b \\ 0 & 1 & 2 & -3 \\ 0 & 0 & a & b+2 \end{pmatrix}.$$

At the first step we subtract the first row from the second and the third; and at the second step we subtract the second from the third. We see that the system has a unique solution (determined by backward elimination) if and only if $a \neq 0$. If $a = 0$ and $b = -2$, we have a zero row in the extended matrix. Choosing $x_3 \in \mathbb{R}$ as a parameter then gives infinitely many distinct solutions. For $a = 0$ and $b \neq -2$ the last equation $a = b + 2$ cannot be satisfied and the system has no solution.

Note that for $a = 0$, $b = -2$ the solutions are

$$(x_1, x_2, x_3) = (-2 + 2t, -3 - 2t, t), \quad t \in \mathbb{R}$$

and for $a \neq 0$ the unique solution is the triple

$$\left(\frac{-3a^2 - ab - 4a + 2b + 4}{a}, -\frac{2b + 3a + 4}{a}, \frac{b+2}{a} \right).$$

□

2.H.19. Let

$$A = \begin{pmatrix} 4 & 5 & 1 \\ 3 & 4 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Find real numbers b_1, b_2, b_3 such that the system of linear equations $A \cdot x = b$ has:

(a) infinitely many solutions;

(b) unique solution;

(c) no solution;

(d) exactly four solutions.

Solution. It is enough to choose $b_1 = b_2 + b_3$ in case a) and $b_1 \neq b_2 + b_3$ in case c). Since all possibilities for $b_1,\ b_2,\ b_3$ are catered for, variant d) cannot occur. Variant b) cannot occur, since the matrix A is not invertible. □

2.H.20. Factor the following permutations into a product of transpositions:

i) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}$,

ii) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 4 & 1 & 2 & 5 & 8 & 3 & 7 \end{pmatrix}$,

iii) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 6 & 1 & 10 & 2 & 5 & 9 & 8 & 3 & 7 \end{pmatrix}$.

2.H.21. Determine the parity of the given permutations:

i) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 5 & 6 & 4 & 1 & 2 & 3 \end{pmatrix}$,

ii) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 7 & 1 & 2 & 3 & 8 & 4 & 5 \end{pmatrix}$,

iii) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 7 & 1 & 10 & 2 & 5 & 4 & 9 & 3 & 6 \end{pmatrix}$.

2.H.22. Find the algebraically adjoint matrix F^* for

$$F = \begin{pmatrix} \alpha & \beta & 0 \\ \gamma & \delta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \alpha, \beta, \gamma, \delta \in \mathbb{R}.$$

2.H.23. Calculate the algebraically adjoint matrix for the matrices

$$\text{(a)} \begin{pmatrix} 3 & -2 & 0 & -1 \\ 0 & 2 & 2 & 1 \\ 1 & -2 & -3 & -2 \\ 0 & 1 & 2 & 1 \end{pmatrix}, \qquad \text{(b)} \begin{pmatrix} 1+i & 2i \\ 3-2i & 6 \end{pmatrix},$$

where i denotes the imaginary unit.

2.H.24. Is the set $V = \{(1, x);\, x \in \mathbb{R}\}$ with operations

$\oplus : V \times V \to V, \quad (1, y) \oplus (1, z) = (1, z + y) \quad$ for all $z, y \in \mathbb{R}$

$\odot : \mathbb{R} \times V \to V, \quad z \odot (1, y) = (1, y \cdot z) \quad$ for all $z, y \in \mathbb{R}$

a vector space?

2.H.25. Express the vector $(5, 1, 11)$ as a linear combination of the vectors $(3, 2, 2)$, $(2, 3, 1)$, $(1, 1, 3)$, that is, find numbers $p, q, r \in \mathbb{R}$, for which

$$(5, 1, 11) = p\,(3, 2, 2) + q\,(2, 3, 1) + r\,(1, 1, 3)\,.$$

2.H.26. In $\mathbb{R}^3$, determine the matrix of rotation through the angle $120°$ in the positive sense about the vector $(1, 0, 1)$

2.H.27. In the vector space $\mathbb{R}^3$, determine the matrix of the orthogonal projection onto the plane $x + y - 2z = 0$.

2.H.28. In the vector space $\mathbb{R}^3$, determine the matrix of the orthogonal projection on the plane $2x - y + 2z = 0$.

2.H.29. Determine whether the subspaces $U = \langle (2, 1, 2, 2) \rangle$ and $V = \langle (-1, 0, -1, 2)\,, (-1, 0, 1, 0)\,, (0, 0, 1, -1) \rangle$ of the space $\mathbb{R}^4$ are orthogonal. If they are, is $\mathbb{R}^4 = U \oplus V$, that is, is $U^\perp = V$?

2.H.30. Let p be a given line:

$$p : [1, 1] + (4, 1)t,\; t \in \mathbb{R}$$

Determine the parametric expression of all lines q that pass through the origin and have deflection $60°$ with the line p.

2.H.31. Depending on the parameter $t \in \mathbb{R}$, determine the dimension of the subspace U of the vector space $\mathbb{R}^3$, if U is generated by the vectors

(a) $u_1 = (1, 1, 1), \quad u_2 = (1, t, 1), \quad u_3 = (2, 2, t)$;
(b) $u_1 = (t, t, t), \quad u_2 = (-4t, -4t, 4t), \quad u_3 = (-2, -2, -2)$.

2.H.32. Construct an orthogonal basis of the subspace

$$\langle\,(1,1,1,1),\ (1,1,1,-1),\ (-1,1,1,1)\,\rangle$$

of the space $\mathbb{R}^4$.

2.H.33. In the space $\mathbb{R}^4$, find an orthogonal basis of the subspace of all linear combinations of the vectors $(1,0,1,0)$, $(0,1,0,-7)$, $\quad (4,-2,4,14)$.

Find an orthogonal basis of the subspace generated by the vectors $(1,2,2,-1)$, $\quad (1,1,-5,3)$, $\quad (3,2,8,-7)$.

2.H.34. For what values of the parameters $a, b \in \mathbb{R}$ are the vectors

$$(1,1,2,0,0), \quad (1,-1,0,1,a), \quad (1,b,2,3,-2)$$

in the space $\mathbb{R}^5$ pairwise orthogonal?

2.H.35. In the space $\mathbb{R}^5$, consider the subspace generated by the vectors $(1,1,-1,-1,0)$, $(1,-1,-1,0,-1)$, $(1,1,0,1,1)$, $(-1,0,-1,1,1)$. Find a basis for its orthogonal complement.

2.H.36. Describe the orthogonal complement of the subspace V of the space $\mathbb{R}^4$, if V is generated by the vectors $(-1,2,0,1)$, $(3,1,-2,4)$, $(-4,1,2,-4)$, $(2,3,-2,5)$.

2.H.37. In the space $\mathbb{R}^5$, determine the orthogonal complement $W^\perp$ of the subspace W, if

(a) $W = \{(r+s+t, -r+t, r+s, -t, s+t);\ r,s,t \in \mathbb{R}\}$;

(b) W is the set of the solutions of the system of equations $x_1 - x_3 = 0$, $x_1 - x_2 + x_3 - x_4 + x_5 = 0$.

2.H.38. In the space $\mathbb{R}^4$, let

$$(1,-2,2,1), \quad (1,3,2,1)$$

be given vectors. Extend these two vectors into an orthogonal basis of the whole $\mathbb{R}^4$. (You can do this in any way you wish, for instance by using the Gram-Schmidt orthogonalization process.)

2.H.39. Define an inner product on the vector space of the matrices from the previous exercise. Compute the norm of the matrix from the previous exercise, induced by the product you have defined. ○

2.H.40. Find a basis for the vector space of all antisymmetric real square matrices of the type 4×4. Consider the standard inner product in this basis and using this inner product, express the size of the matrix

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

. ○

2.H.41. Find the eigenvalues and the associated eigenspaces of eigenvectors of the matrix:

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

Solution. The characteristic polynomial of the matrix is $\lambda^3 - 6\lambda^2 + 12\lambda - 8$, which is $(\lambda-2)^3$. The number 2 is thus an eigenvalue with algebraic multiplicity three. Its geometric multiplicity is either one, two or three. We determine the vectors associated to this eigenvalue as the solutions of the system

$$(A-2E)\mathbf{x} = \begin{array}{rrcl} -x_1 & +x_2 & = & 0, \\ -x_1 & +x_2 & = & 0, \\ 2x_1 & -2x_2 & = & 0. \end{array}$$

Its solutions form the two-dimensional space $\langle(1,-1,0),(0,0,1)\rangle$. Thus the eigenvalue 2 has algebraic multiplicity 3 and geometric multiplicity 2.

□

2.H.42. Determine the eigenvalues of the matrix

$$\begin{pmatrix} -13 & 5 & 4 & 2 \\ 0 & -1 & 0 & 0 \\ -30 & 12 & 9 & 5 \\ -12 & 6 & 4 & 1 \end{pmatrix}.$$

○

2.H.43. Given that the numbers 1, -1 are eigenvalues of the matrix

$$A = \begin{pmatrix} -11 & 5 & 4 & 1 \\ -3 & 0 & 1 & 0 \\ -21 & 11 & 8 & 2 \\ -9 & 5 & 3 & 1 \end{pmatrix},$$

find all solutions of the characteristic equation $|\,A - \lambda\,E\,| = 0$. Hint: if you denote all the roots of the polynomial $|\,A - \lambda\,E\,|$ by $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, then

$$|\,A\,| = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \cdot \lambda_4, \quad \textit{and} \quad \operatorname{tr} A = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4.$$

○

2.H.44. Find a four-dimensional matrix with eigenvalues $\lambda_1 = 6$ and $\lambda_2 = 7$ such that the multiplicity of λ_2 as a root of the characteristic polynomial is three, and that

(a) the dimension of the subspace of eigenvectors of λ_2 is 3;

(b) the dimension of the subspace of eigenvectors of λ_2 is 2;

(c) the dimension of the subspace of eigenvectors of λ_2 is 1;

○

2.H.45. Find the eigenvalues and the eigenvectors of the matrix:

$$\begin{pmatrix} -1 & -\frac{5}{6} & \frac{5}{3} \\ 0 & -\frac{2}{3} & -\frac{2}{3} \\ 0 & \frac{1}{6} & -\frac{4}{3} \end{pmatrix}.$$

2.H.46. Determine the characteristic polynomial $|\,A - \lambda\,E\,|$, eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}.$$

○

respectively.

Solutions to the exercises

2.A.11. There is only one such matrix X, and it is

$$\begin{pmatrix} 18 & -32 \\ 5 & -8 \end{pmatrix}.$$

2.A.13. $A^{-1} = \begin{pmatrix} 1 & 10 & -4 \\ 1 & 12 & -5 \\ 0 & 5 & -2 \end{pmatrix}.$

2.A.14.

$$A^5 = \begin{pmatrix} 122 & -121 & 121 \\ -121 & 122 & -121 \\ 0 & 0 & 1 \end{pmatrix}, \quad A^{-3} = \frac{1}{27} \begin{pmatrix} 14 & 13 & -13 \\ 13 & 14 & 13 \\ 0 & 0 & 27 \end{pmatrix}.$$

2.A.15. $\begin{pmatrix} 2 & -3 & 0 & 0 & 0 \\ -5 & 8 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -5 & 2 \\ 0 & 0 & 0 & 3 & -1 \end{pmatrix}.$

2.A.16. $C^{-1} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$

2.A.17. In the first case we have

$$A^{-1} = \frac{1}{2} \cdot \begin{pmatrix} 3 & -i \\ i & 1 \end{pmatrix};$$

in the second

$$A^{-1} = \begin{pmatrix} 14 & 8 & 5 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

2.D.7. $(2 + \frac{1}{\sqrt{3}}, 2 - \frac{1}{\sqrt{3}})$.

2.D.8. The vectors are dependent whenever at least one of the conditions

$$a = b = 1, \qquad a = c = 1, \qquad b = c = 1$$

is satisfied.

2.D.9. Vectors are linearly independent.

2.D.10. It suffices to add for instance the polynomial x.

2.F.5. $\cos = \frac{\sqrt{2}}{\sqrt{3}}$.

2.G.3. Je $|\, A - \lambda\, E \,| = -\lambda^3 + 12\lambda^2 - 47\lambda + 60,$. $\lambda_1 = 3$, $\lambda_2 = 4$, $\lambda_3 = 5$.

2.G.11. The solution is the sequence $0, 1, 2$.

2.G.12. The dimension is 1 for $\lambda_1 = 4$ and 2 for $\lambda_2 = 3$.

2.H.8. The solutions are all scalar multiples of the vector

$$\left(1 + \sqrt{3},\, -\sqrt{3},\, 0,\, 1,\, 0\right).$$

2.H.9. $x_1 = 1 + t, \quad x_2 = \frac{3}{2}, \quad x_3 = t, \quad x_4 = -\frac{1}{2}, \qquad t \in \mathbb{R}.$

2.H.10. The system has no solution.

2.H.11. The system has a solution, because

$$3 \cdot \begin{pmatrix} 3 \\ 2 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 3 \\ -3 \\ -2 \end{pmatrix} - 5 \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 8 \\ 4 \\ 6 \end{pmatrix}.$$

2.H.12. The system of linear equations

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

has no solution, while the system

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

has a unique solution $x_1 = -1$, $x_2 = 1$, $x_3 = 2$.

2.H.13. The set of all solutions is given by

$$\{(-10t,\ (a+4)t,\ (3a-8)t)\,;\ t \in \mathbb{R}\}.$$

2.H.14. For $a = 0$, the system has no solution. For $a \neq 0$ the system has infinitely many solutions.

2.H.15. The correct answers are „yes“, „no“, „no“ and „yes“ respectively.

2.H.16. i) If $b \neq -7$, then $x = z = (2+a)/(b+7)$, $y = (3a-b-1)/(b+7)$. ii) If $b = -7$ and $a \neq -2$, then there is no solution. iii) If $a = -2$ and $b = -7$ then the solution is $x = z = t$, $y = 3t - 1$, for any t.

2.H.17.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}^{-1} \quad \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

We can then easily obtain

$$x_1 = \frac{13}{4}, \quad x_2 = -\frac{3}{4}, \quad x_3 = -\frac{3}{4}, \quad x_4 = \frac{1}{4}.$$

2.H.20. i) $(1,7)(2,6)(5,3)$, ii) $(1,6)(6,8)(8,7)(7,3)(2,4)$, iii) $(1,4)(4,10)(10,7)(7,9)(9,3)(2,6)(6,5)$

2.H.21. i) 17 inversions, odd, ii) 12 inversions, even iii) 25 inversions, odd

2.H.22. From the knowledge of the inverse matrix F^{-1} we obtain

$$F^* = (\alpha\delta - \beta\gamma)\, F^{-1} = \begin{pmatrix} \delta & -\beta & 0 \\ -\gamma & \alpha & 0 \\ 0 & 0 & \alpha\delta - \beta\gamma \end{pmatrix},$$

for any α, β, γ, $\delta \in \mathbb{R}$.

2.H.23. The matrices are

$$\text{(a)} \begin{pmatrix} 1 & 1 & -2 & -4 \\ 0 & 1 & 0 & -1 \\ -1 & -1 & 3 & 6 \\ 2 & 1 & -6 & -10 \end{pmatrix}, \qquad \text{(b)} \begin{pmatrix} 6 & -2i \\ -3+2i & 1+i \end{pmatrix}.$$

2.H.24. It is easy to check that it is a vector space. The first coordinate does not affect the results of the operations – it is just the vector space $(\mathbb{R}, +, \cdot)$ written in a different way.

2.H.25. There is a unique solution

$$p = 2, \qquad q = -2, \qquad r = 3.$$

2.H.26.

$$\begin{pmatrix} 1/4 & -\sqrt{6}/4 & 3/4 \\ \sqrt{6}/4 & -1/2 & -\sqrt{6}/4 \\ 3/4 & \sqrt{6}/4 & 1/4 \end{pmatrix}$$

2.H.27.

$$\begin{pmatrix} 5/6 & -1/6 & 1/3 \\ -1/6 & 5/6 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

2.H.28.
$$\begin{pmatrix} 5/9 & 2/9 & -4/9 \\ 2/9 & 8/9 & 2/9 \\ -4/9 & 2/9 & 5/9 \end{pmatrix}$$

2.H.29. The vector that determines the subspace U is perpendicular to each of the three vectors that generate V. The subspaces are thus orthogonal. But it is not true that $\mathbb{R}^4 = U \oplus V$. The subspace V is only two-dimensional, because
$$(-1, 0, -1, 2) = (-1, 0, 1, 0) - 2\,(0, 0, 1, -1)\,.$$

2.H.30.
$$q_1 : (2 - \frac{\sqrt{3}}{2}, 2\sqrt{3} + \frac{1}{2})t, \quad q_2 : (2 + \frac{\sqrt{3}}{2}, -2\sqrt{3} + \frac{1}{2})t.$$

2.H.31. In the first case we have $\dim U = 2$ for $t \in \{1, 2\}$, otherwise we have $\dim U = 3$. In the second case we have $\dim U = 2$ for $t \neq 0$ and $\dim U = 1$ for $t = 0$.

2.H.32. Using the Gram-Schmidt orthogonalization process we can obtain the result
$$((1, 1, 1, 1),\ (1, 1, 1, -3),\ (-2, 1, 1, 0))\,.$$

2.H.33. We have for instance the orthogonal bases
$$((1, 0, 1, 0), (0, 1, 0, -7))$$
for the first part, and
$$((1, 2, 2, -1), (2, 3, -3, 2), (2, -1, -1, -2)).$$
for the second part.

2.H.34. The solution is $a = 9/2$, $b = -5$, because
$$1 + b + 4 + 0 + 0 = 0, \quad 1 - b + 0 + 3 - 2a = 0.$$

2.H.35. The basis must contain a single vector. It is
$$(3, -7, 1, -5, 9).$$
(or any non-zero scalar multiple thereof.

2.H.36. The orthogonal complement $V^\perp$ is the set of all scalar multiples of the vector $(4, 2, 7, 0)$.

2.H.37.
(a) $W^\perp = \langle\, (1, 0, -1, 1, 0),\ (1, 3, 2, 1, -3)\,\rangle\,;$
(b) $W^\perp = \langle\, (1, 0, -1, 0, 0),\ (1, -1, 1, -1, 1)\,\rangle.$

2.H.38. There are infinitely many possible extensions, of course. A very simple one is
$$(1, -2, 2, 1), \quad (1, 3, 2, 1), \quad (1, 0, 0, -1), \quad (1, 0, -1, 1).$$

2.H.39. For instance, one can use the inner product that follows from the isomorphism of the space of all real 3×3 matrices with the space $\mathbb{R}^9$. If we use the product from $\mathbb{R}^9$, we obtain an inner product that assigns to two matrices the sum of products of two corresponding elements. For the given matrix we obtain
$$\left\| \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 1 & -2 & -3 \end{pmatrix} \right\| = \left\langle \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 1 & -2 & -3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 1 & -2 & -3 \end{pmatrix} \right\rangle = \sqrt{1^2 + 2^2 + 0^2 + 0^2 + 2^2 + 0^2 + 1^2 + (-2)^2 + (-3)^2} = \sqrt{23}.$$

2.H.40.

2.H.42. The matrix has only one eigenvalue, namely -1, since the characteristic polynomial is $(\lambda + 1)^4$.

2.H.43. The root -1 of the polynomial $|\,A - \lambda\,E\,|$ has multiplicity three.

2.H.44. Possible examples are,
$$\text{(a)} \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}; \quad \text{(b)} \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix};$$
$$\text{(c)} \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 7 \end{pmatrix}.$$

2.H.45. There is a triple eigenvalue -1. The corresponding eigenspace is $\langle(1,0,0),(0,2,1)\rangle$.

2.H.46. The characteristic polynomial is $-(\lambda-2)^2(\lambda-9)$, that is, the eigenvalues are 2 and 9 with associated eigenvectors

$$(1,2,0)\,,(-3,0,1) \quad \text{a} \quad (1,1,1)$$