INTRODUCTION TO LINEAR ALGEBRA

LINEAR SPACE, MATRICES, DETERMINANTS

SYSTEM OF LINEAR EQUATION

LINEAR TRANSFORMATIONS

ELEMENTS OF 3D GEOMETRY

Pavel Ptk

Contents

Preface

This textbook consists of extended lecture notes of the lecture I held at The Czech Technical University of Prague. It comprises basic linear algebra needed in engineering sciences (for a more detailed account, see the contents). The theory is supplemented with illustrating examples. Problems for individual study are formulated at the end of each chapter. Two appendices – mathematical induction and elementary theory of polynomials and rational functions – are provided. (The only prerequisity for reading this textbook is elementary algebra and geometry in the extent of the secondary school curricula.)

Bibliography is included at the end of the textbook. There is also my acknowledgement of the help extended to me by my colleagues and students.

Prague, July 1997

Prof. RNDr. Pavel Ptk, DrSc.

Department of Mathematics
&
Center for Machine Perception

Faculty of Electrical Engineering
Czech Technical University, Prague

Acknowledgement: The author acknowledges the support of the professional association "ELEKTRA" which enabled publication of this textbook.

Chapter 1

Linear space

1.1. The definition of linear space. Examples

The notion of linear space came into existence in the effort to obtain a *unifying principle* for certain algebraic phenomena. After realizing that constructions of a "linear nature" repeat themselves in several situations, it became appropriate to adopt an axiomatic approach. It is the axiomatic approach that gives rise to the notion of linear space. The advantage of this modern and largely used attitude is that one obtains results simultaneously applicable in apparently totally different areas. The effort needed for proving these more general results is in fact the same as the effort needed for proving them in a particular model. The alleged flaw seems to be a higher degree of abstraction. A beginner may indeed experience a certain inconvenience, particularly in the case when the secondary school in which he studied neglected the preparation of "mathematical thinking". But this can be overcome. Sooner or later the hard-working student will appreciate the axiomatic approach and begin to like it. There is no doubt that he (she) will profit from his (her) skill acquired in the axiomatic approach when confronted with the algebraic problems of engineering sciences.

We will define the notion of linear space first and then we will illustrate this notion by several examples. Before, let us make the following conventions. The letter R will denote the set of all real numbers understood with their usual algebraic structure. The letter N will stand for the set of all natural numbers. The symbols $\cup$, $\cap$, $-$ will denote the standard set theory operations of union, intersection and complementation. The symbol $P \times Q$ will denote the cartesian product of the sets P and Q. Finally, the logical quantifiers $\forall$ ("for every") and $\exists$ ("there exists") as well as the logical implication $\Rightarrow$ will be occasionally used to simplify the formulation of the results. The ends of proofs will be denoted by $\square$.

1.1.1. Definition. Let L be a nonvoid set. Let L be equipped with the operations $\sharp\colon L \times L \to L$ and $\circ\colon R \times L \to L$ such that the following 8 conditions are satisfied:

Axiom 1. $\overrightarrow{x} \sharp \overrightarrow{y} = \overrightarrow{y} \sharp \overrightarrow{x}$ (the commutativity of $\sharp$), $\forall \overrightarrow{x}, \overrightarrow{y} \in L$,

Axiom 2. $\overrightarrow{x} \sharp (\overrightarrow{y} \sharp \overrightarrow{z}) = (\overrightarrow{x} \sharp \overrightarrow{y}) \sharp \overrightarrow{z}$ (the associativity of $\sharp$), $\forall \overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z} \in L$,

Axiom 3. $\exists \overrightarrow{o} \in L$ such that $\overrightarrow{x} \sharp \overrightarrow{o} = \overrightarrow{x}$ (the existence of the zero vector), $\forall \overrightarrow{x} \in L$,

Axiom 4. $\forall \overrightarrow{x} \exists \overrightarrow{y}$ such that $\overrightarrow{x} \sharp \overrightarrow{y} = \overrightarrow{o}$ (the existence of the $\sharp$-inverse),

Axiom 5. $1 \circ \overrightarrow{x} = \overrightarrow{x}$ (the neutrality of 1 with respect to $\circ$), $\forall \overrightarrow{x} \in L$,

Axiom 6. $(\lambda\mu) \circ \overrightarrow{x} = \lambda \circ (\mu \circ \overrightarrow{x})$ (the interplay of $\circ$ with multiplication in R), $\forall \overrightarrow{x} \in L, \forall \lambda, \mu \in R$, Axiom 7. $(\lambda+\mu)\circ\overrightarrow{x}=\lambda\circ\overrightarrow{x}\ \sharp\ \mu\circ\overrightarrow{x}$ (the distributivity of $\circ$ with addition in R), $\forall\overrightarrow{x}\in L, \forall\lambda,\mu\in R$,

Axiom 8. $\lambda\circ(\overrightarrow{x}\ \sharp\ \overrightarrow{y})=\lambda\circ\overrightarrow{x}\ \sharp\ \lambda\circ\overrightarrow{y}$ (distributivity of $\sharp$ and $\circ$), $\forall\overrightarrow{x},\overrightarrow{y}\in L, \forall\lambda\in R$.

The triple $(L,\sharp,\circ)$ is then called *a linear space.*

Examples:

1. **The linear space R^n**

   Fix a natural number n and let $L=R^n$ (i.e., R^n is the set of all ordered n-tuples of real numbers). Thus, if $\overrightarrow{a}\in R^n$, then $\overrightarrow{a}=(a_1,a_2,\dots,a_n)$. For instance, $(1,1/2,-\pi)\in R^3$, $(0,0)\in R^2$, $(1,2,3,4)\in R^4$, etc.

   Let us define the operations $\sharp,\circ$ as follows: If

   $$\overrightarrow{a}=(a_1,a_2,\dots,a_n)\in R^n \quad\text{and}\quad \overrightarrow{b}=(b_1,b_2,\dots,b_n)\in R^n,$$

   then

   $$\overrightarrow{a}\sharp\overrightarrow{b}=(a_1+b_1,a_2+b_2,\dots,a_n+b_n), \qquad \lambda\circ\overrightarrow{a}=(\lambda a_1,\lambda a_2,\dots,\lambda a_n).$$

   One can easily check that $(R^n,\sharp,\circ)$ is a linear space. We will show it for R^3. Observe how the algebraic structure of R (the commutativity of addition, the associativity and distributivity of operations on R, etc.) is used in verifying the axioms of linear space.

   Ax. 1: $\overrightarrow{x}\ \sharp\ \overrightarrow{y}=(x_1,x_2,x_3)\ \sharp\ (y_1,y_2,y_3)=(x_1+y_1,x_2+y_2,x_3+y_3)=(y_1+x_1,y_2+x_2,y_3+x_3)=\overrightarrow{y}\ \sharp\ \overrightarrow{x}$,

   Ax. 2: $\overrightarrow{x}\ \sharp\ (\overrightarrow{y}\ \sharp\ \overrightarrow{z})=(x_1,x_2,x_3)\ \sharp\ [(y_1,y_2,y_3)\ \sharp\ (z_1,z_2,z_3)]=(x_1,x_2,x_3)\ \sharp\ (y_1+z_1,y_2+z_2,y_3+z_3)=(x_1+(y_1+z_1),x_2+(y_2+z_2),x_3+(y_3+z_3))=((x_1+y_1)+z_1,(x_2+y_2)+z_2,(x_3+y_3)+z_3)=(x_1+y_1,x_2+y_2,x_3+y_3)\ \sharp\ (z_1,z_2,z_3)=(\overrightarrow{x}\ \sharp\ \overrightarrow{y})\ \sharp\ \overrightarrow{z}$,

   Ax. 3: $\overrightarrow{o}=(0,0,0)$,

   Ax. 4: If $\overrightarrow{x}=(x_1,x_2,x_3)$, then we take $\overrightarrow{y}=(-x_1,-x_2,-x_3)$ for the inverse,

   Ax. 5: $1\circ\overrightarrow{x}=1\circ(x_1,x_2,x_3)=1x_1,1x_2,1x_3)=(x_1,x_2,x_3)=\overrightarrow{x}$,

   Ax. 6: $(\lambda\mu)\circ\overrightarrow{x}=(\lambda\mu)\circ(x_1,x_2,x_3)=((\lambda\mu)x_1,(\lambda\mu)x_2,(\lambda\mu)x_3)=(\lambda(\mu x_1),\lambda(\mu x_2),\lambda(\mu x_3))=\lambda\circ(\mu x_1,\mu x_2,\mu x_3)=\lambda\circ(\mu\circ\overrightarrow{x})$,

   Ax. 7: $(\lambda+\mu)\circ\overrightarrow{x}=(\lambda+\mu)\circ(x_1,x_2,x_3)=((\lambda+\mu)x_1,(\lambda+\mu)x_2,(\lambda+\mu)x_3)=(\lambda x_1+\mu x_1,\lambda x_2+\mu x_2,\lambda x_3+\mu x_3)=(\lambda x_1,\lambda x_2,\lambda x_3)\ \sharp\ (\mu x_1,\mu x_2,\mu x_3)=\lambda\circ\overrightarrow{x}\ \sharp\ \mu\circ\overrightarrow{x}$,

   Ax. 8: $\lambda\circ(\overrightarrow{x}\ \sharp\ \overrightarrow{y})=\lambda\circ[(x_1,x_2,x_3)\ \sharp\ (y_1,y_2,y_3)]=\lambda\circ(x_1+y_1,x_2+y_2,x_3+y_3)=$
   $=(\lambda(x_1+y_1),\lambda(x_2+y_2),\lambda(x_3+y_3))=(\lambda x_1+\lambda y_1,\lambda x_2+\lambda y_2,\lambda x_3+\lambda y_3)=$
   $=(\lambda x_1,\lambda x_2,\lambda x_3)\ \sharp\ (\lambda y_1,\lambda y_2,\lambda y_3)=\lambda\circ\overrightarrow{x}\ \sharp\ \lambda\circ\overrightarrow{y}$.

   It should be noted that the set $R(=R^1)$ of real numbers can be viewed as a linear space.

2. **The linear space V_3 of geometric vectors.**

   Let V be the set of "free" vectors in the three-dimensional Euclidean space E_3. If we introduce the operations $\sharp,\circ$ the standard way (see the figures below) then $(V_3,\sharp,\circ)$ will become a linear space. We will make this space a topic of special attention later (see Chap. 4).

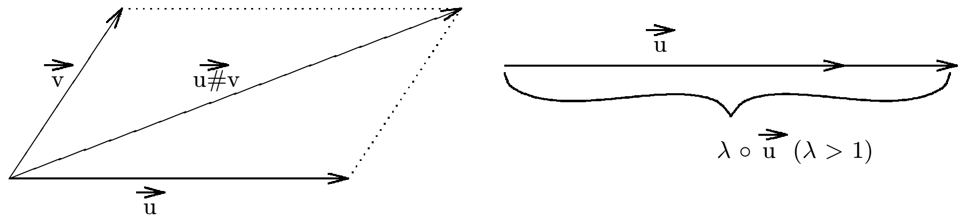


Fig. 1

3. **The linear space $\mathscr{F}(\mathscr{S})$ of all real functions defined on a given set S.**
If $f, g \in \mathscr{F}(S)$, let us define the operations $\sharp, \circ$ "pointwise":

$$(f \sharp g)(x) = f(x) + g(x), \quad \forall x \in S$$
$$(\lambda \circ f)(x) = \lambda \cdot f(x), \quad \forall x \in S.$$

It is easily seen that the requirements of Axioms 1–8 are then fulfilled. Thus, $\mathscr{F}(\mathscr{S})$ is a linear space.

4. **The linear space of all matrices of a given size**

1.1.2. Definition. Let us fix two natural numbers $m, n \in N$. By *a matrix* of the size (m, n) we define a rectangular array of real numbers that consists of m rows and n columns.

Example: $\mathbf{A} = \begin{Vmatrix} 1, & 2 \\ 3, & 0 \\ 1, & -1 \end{Vmatrix}$ is a 3×2 matrix ($\mathbf{A}$ is a "3 by 2" matrix),

$$\mathbf{B} = \begin{Vmatrix} 1, & 2, & -1, & 4 \\ \frac{1}{2}, & \frac{2}{7}, & 5, & 0 \\ 0, & 1, & 2, & \pi \\ 1, & 1, & -1, & 0 \end{Vmatrix} \text{ is a } 4 \times 4 \text{ matrix,}$$

$\mathbf{C} = \|1, 2, 3, 4, 5\|$ is a 1×5 matrix.

Matrix $\mathbf{M} = \|m_{ij}\|$, where m_{ij} denotes the intersection of the i-th row with the j-th column (the numbers m_{ij} in the matrix array are called *the components* of $\mathbf{M}$ or "entries" of $\mathbf{M}$). In the example above, by writing $\mathbf{A} = \|a_{ij}\|$ we mean $a_{11} = 1$, $a_{12} = 2$, $a_{21} = 3$, etc.

Let us fix the pair m, n of natural numbers and denote by $\mathscr{M}$ the set of all matrices of size $m \times n$. The set $\mathscr{M}^{mn}$ can be understood as a linear space if we introduce the operations "componentwise":

If $\mathbf{M} = \|m_{ij}\| \in \mathscr{M}^{mn}$, $\mathbf{K} = \|k_{ij}\| \in \mathscr{M}^{mn}$, then $\mathbf{M} \sharp \mathbf{K} = \|m_{ij} + k_{ij}\|$, $\lambda \circ \mathbf{M} = \|\lambda m_{ij}\|$.
It is easy to verify that $\mathscr{M}^{mn}$ with the above operations $\sharp, \circ$ forms a linear space. (The reader is supposed to check all Axioms 1–8 of linear space on his (her) own. It is advisable to make a habit of checking "easy" or "trivial" statements in detail, it helps very much in learning the subject.)

1.1.3. Remark. Observe that the space R^n is in fact "identical" with the space $\mathscr{M}$ (resp. with the space $\mathscr{M}$).

When may $(L, \sharp, \circ)$ not be a linear space?

(a) Let $L = R$ and define $\overrightarrow{a} \sharp \overrightarrow{b} = 5 \quad \forall \overrightarrow{a}, \overrightarrow{b} \in L,\ \lambda \circ \overrightarrow{a} = 2 \quad \forall \lambda \in R,\ \forall \overrightarrow{a} \in L$. Then $(R, \sharp, \circ)$ does *not* satisfy the axioms 3, 4, 5, 7, 8. Thus, $(R, \sharp, \circ)$ is not a linear space.

(b) Let $L = R$, and define $\overrightarrow{a} \sharp \overrightarrow{b} = \overrightarrow{a} \quad \forall \overrightarrow{a}, \overrightarrow{b} \in L,\ \lambda \circ \overrightarrow{a} = \lambda \overrightarrow{a} \quad \forall \lambda \in R$, $\forall \overrightarrow{a} \in L$. Then $(L, \sharp, \circ)$ fails to be a linear space. Which axioms are not fulfilled ? (Answer: 1, 2, 3, 4, 7, 8)

(c) Let $L = R$, and define $\overrightarrow{a} \sharp \overrightarrow{b} = 0 \quad \forall \overrightarrow{a}, \overrightarrow{b} \in L, \quad \lambda \circ \overrightarrow{a} = 0 \quad \forall \lambda \in R,\ \overrightarrow{a} \in L$. Then $(L, \sharp, \circ)$ fails to be a linear space. Which axioms are not fulfilled ? (Answer: 3, 4, 5)

Sometimes a linear space may be given rather nonstandard operations.

Example: Put $L = R^+ = (0, +\infty)$, and define

$$\begin{aligned} \overrightarrow{x} \sharp \overrightarrow{y} &= xy \qquad && \forall \overrightarrow{x} \overrightarrow{y} \in L, \\ \lambda \circ \overrightarrow{x} &= x^\lambda && \forall \lambda \in R,\ \overrightarrow{x} \in L. \end{aligned}$$

Then $(L, \sharp, \circ)$ is a linear space.

Solution:

Ax.1: $\overrightarrow{x} \sharp \overrightarrow{y} = \overrightarrow{y} \sharp \overrightarrow{x}$; this means $xy = yx$ O.K.

Ax.2: $\overrightarrow{x} \sharp (\overrightarrow{y} \sharp \overrightarrow{z}) = (\overrightarrow{x} \sharp \overrightarrow{y}) \sharp \overrightarrow{z}$; this means $x(yz) = (xy)z$ O.K.

Ax.3: $\exists \overrightarrow{o}, \overrightarrow{x} \sharp \overrightarrow{o} = \overrightarrow{x}$; this means $\overrightarrow{o} = 1$ O.K.

Ax.4: $\forall \overrightarrow{x}\ \exists \overrightarrow{y}, \overrightarrow{x} \sharp \overrightarrow{y} = \overrightarrow{o}$; this means $\overrightarrow{x} = x,\ \overrightarrow{y} = \frac{1}{x},\ \overrightarrow{x} \sharp \overrightarrow{y} = x \cdot \frac{1}{x} = 1 = \overrightarrow{o}$ O.K.

Ax.5: $1 \circ \overrightarrow{x} = \overrightarrow{x}$; this means $x = x^1$ O.K.

Ax.6: $(\lambda\mu) \circ \overrightarrow{x} = \lambda \circ (\mu \circ \overrightarrow{x})$; this means $x^{\lambda\mu} = (x^\mu)^\lambda$ O.K.

Ax.7: $(\lambda + \mu) \circ \overrightarrow{x} = \lambda \overrightarrow{x} \sharp \mu \circ \overrightarrow{x}$; this means $x^{\lambda+\mu} = x^\lambda \cdot x^\mu$ O.K.

Ax.8: $\lambda \circ (\overrightarrow{x} \sharp \overrightarrow{y}) = \lambda \circ \overrightarrow{x} \sharp \lambda \circ \overrightarrow{y}$; this means $(xy)^\lambda = x^\lambda \cdot y^\lambda$ O.K.

□

We will be interested in the "calculus" of a linear space. Before, let us make a few conventions. It is customary to call the elements of L **vectors**. For this reason, a linear space L is sometimes called a vector space. The real numbers $\lambda \in R$ used in the "outer" operation $\circ$ are called scalars. Some more conventions: $\overrightarrow{o}$ is called *the zero vector* (the null vector), *the inverse* vector to $\overrightarrow{x}$, i. e. the vector $\overrightarrow{y}$ guaranteed for the given $\overrightarrow{x}$ by Ax. 4 - will be denoted by $(-\overrightarrow{x})$, the vector $\overrightarrow{x} \sharp (-\overrightarrow{y})$ will sometimes be called *the difference* of $\overrightarrow{x}$ and $\overrightarrow{y}$ and denoted by $\overrightarrow{x} - \overrightarrow{y}$. In what follows, the letter L will be reserved for linear spaces.

The following two propositions establish that the behavior of linear spaces cannot be too wild.

1.1.4. Proposition. There exists exactly one zero vector, $\overrightarrow{o}$, in L. Furthermore, for any $\overrightarrow{x}$ of L, there exists exactly one inverse in L – the vector $(-\overrightarrow{x})$.

PROOF: We will prove Prop. 1.1.4 by way of contradiction. Suppose that the zero vector in L is not unique. Thus, suppose that $\overrightarrow{o}, \overrightarrow{o'}$ are two different vectors with the property required in Ax. 3. Then $\overrightarrow{o} \sharp \overrightarrow{o'} = \overrightarrow{o}$ (Ax.3) and, also $\overrightarrow{o} \sharp \overrightarrow{o'} = \overrightarrow{o'}$ (Ax.3 and 1). It follows that $\overrightarrow{o} = \overrightarrow{o'}$ but this contradicts our assumption.

Let $(-\overrightarrow{x}), (-\overrightarrow{x})'$ be two "$\sharp$-inverses" to $\overrightarrow{x}$. Then $\overrightarrow{x} \sharp (-\overrightarrow{x}) = \overrightarrow{o}$ (Ax.4)."Adding" $(-\overrightarrow{x})'$ to both sides of the last equation, we obtain $(-\overrightarrow{x})' \sharp [\overrightarrow{x} \sharp; (-\overrightarrow{x})] = (-\overrightarrow{x})' \sharp \overrightarrow{o} = = (-\overrightarrow{x})'$. According to Ax.2, Ax.1 and Ax.4, $(-\overrightarrow{x})' \sharp [\overrightarrow{x} \sharp (-\overrightarrow{x})] = [(-\overrightarrow{x})' \sharp \overrightarrow{x}] \sharp (-\overrightarrow{x}) = = [\overrightarrow{x} \sharp (-\overrightarrow{x})'] \sharp (-\overrightarrow{x}) = \overrightarrow{o} \sharp (-\overrightarrow{x}) = (-\overrightarrow{x})$.

□

1.1.5. Proposition. The following equalities are valid in L:

(i) $0\circ\overrightarrow{x} = \overrightarrow{o} \quad \forall x \in L$,

(ii) $\lambda\circ\overrightarrow{o} = \overrightarrow{o} \quad \forall \lambda \in R$,

(iii) $(-1)\circ\overrightarrow{x} = (-\overrightarrow{x}) \ \forall x \in L$.

PROOF: Making use of Ax.6, we see that $0\circ\overrightarrow{x} = (0+0)\circ\overrightarrow{x} = (0\circ\overrightarrow{x}) \sharp (0\circ\overrightarrow{x})$. By adding the vector $-(0\circ\overrightarrow{x})$ to both sides of the equality $0\circ\overrightarrow{x} = 0\circ\overrightarrow{x} \sharp 0\circ\overrightarrow{x}$, we consecutively obtain

$$0\circ\overrightarrow{x} \sharp -(0\circ\overrightarrow{x}) = [0\circ\overrightarrow{x} \sharp 0\circ\overrightarrow{x}] \sharp -(0\circ\overrightarrow{x}) = (\text{Ax. } 2) =$$

$$= 0\circ\overrightarrow{x} \sharp [0\circ\overrightarrow{x} \sharp -(0\circ\overrightarrow{x})] = (\text{Ax. } 4) = 0\circ\overrightarrow{x} \sharp \overrightarrow{o} = (\text{Ax. } 3) = 0\circ\overrightarrow{x}.$$

But $0\circ\overrightarrow{x} \sharp -(0\circ\overrightarrow{x}) = \overrightarrow{o}$ and therefore $\overrightarrow{o} = 0\circ\overrightarrow{x}$. Similarly, $\lambda\circ\overrightarrow{o} = (\text{Ax. } 3) = \lambda\circ(\overrightarrow{o} \sharp \overrightarrow{o}) = (\text{Ax. } 6) = \lambda\circ\overrightarrow{o} \sharp \lambda\circ\overrightarrow{o}$, etc. Furthermore,

$$(1+(-1))\circ\overrightarrow{x} = 0\circ\overrightarrow{x} = \overrightarrow{o} \Rightarrow 1\circ\overrightarrow{x}\sharp(-1)\circ\overrightarrow{x} = \overrightarrow{o} \Rightarrow (\text{Ax. } 5)\overrightarrow{x}\sharp(-1)\circ\overrightarrow{x} = \overrightarrow{o}$$

$$\Rightarrow (\text{Ax. } 4)\ (-1)\circ\overrightarrow{x} = (-\overrightarrow{x}).$$

The above proposition will often be used in the calculus of linear space. In particular, Prop. 1.1.5 (iii) will be applied in many places (later on without explicitly mentioning it).

□

1.1.6. Example. Show that the equality $\overrightarrow{x} \sharp 2\circ\overrightarrow{y} = \overrightarrow{o}$ is equivalent to the equality $\overrightarrow{y} = \left(-\frac{1}{2}\right)\circ\overrightarrow{x}$ in any linear space.

SOLUTION: $\overrightarrow{x} \sharp 2\circ\overrightarrow{y} = \overrightarrow{o} \Leftrightarrow \frac{1}{2}\circ(\overrightarrow{x} \sharp 2\circ\overrightarrow{y}) = \frac{1}{2}\circ\overrightarrow{o} \Leftrightarrow \frac{1}{2}\circ\overrightarrow{x} \sharp \left[\frac{1}{2}\circ(2\circ\overrightarrow{y})\right] = \overrightarrow{o} \Leftrightarrow \frac{1}{2}\circ\overrightarrow{x} \sharp (1)\circ\overrightarrow{y} = \overrightarrow{o} \Leftrightarrow \frac{1}{2}\circ\overrightarrow{x} \sharp \overrightarrow{y} = \overrightarrow{o} \Leftrightarrow -\left(\frac{1}{2}\circ\overrightarrow{x}\right) \sharp \left(\frac{1}{2}\circ\overrightarrow{x} \sharp \overrightarrow{y}\right) = -\left(\frac{1}{2}\circ\overrightarrow{x}\right) \Leftrightarrow \left[-(\frac{1}{2}\circ\overrightarrow{x}) \sharp \frac{1}{2}\circ\overrightarrow{x}\right] \sharp \overrightarrow{y} = -(\frac{1}{2}\circ\overrightarrow{x}) \Leftrightarrow \overrightarrow{o} \sharp \overrightarrow{y} = -(\frac{1}{2}\circ\overrightarrow{x}) \Leftrightarrow$ Ax. 1, Ax. 3, Prop. 1.1.5 (iii) $\overrightarrow{y} = (-1)\circ(\frac{1}{2}\circ\overrightarrow{x}) \Leftrightarrow$ Ax. 6 $\overrightarrow{y} = (-\frac{1}{2})\circ\overrightarrow{x}$.

The following considerations model the situation known from the space of geometric vectors.

1.1.7. Definition. Let $\overrightarrow{x_1}, \overrightarrow{x_2}, \ldots, \overrightarrow{x_k}$ be vectors of L, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be scalars. The vector $\overrightarrow{y}$ defined by the equality

$$\overrightarrow{y} = \lambda_1\circ\overrightarrow{x}_1 \sharp \lambda_2\circ x_2 \sharp \ \ldots \ \sharp \lambda_k x_k$$

is called *the linear combination* of the vectors $\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_k}$ with the coefficients $\lambda_1, \lambda_2, \dots, \lambda_k$.

1.1.8. Example. In the space R^2 we have: $3\circ(1,2) \sharp (-2)\circ(5,3) = (3,6) \sharp (-10,-6) = (-7,0)$. Thus, the vector $(-7,0)$ is the linear combination of $(1,2),(5,3)$ with the coefficients 3, -2.

Is the vector (5,3) a linear combination of the vectors (0,0), (2,1), (4,2)? If it is so, there are scalars $\lambda_1, \lambda_2, \lambda_3$ such that $(5,3) = \lambda_1 \circ (0,0) \sharp \lambda_2 \circ (2,1) \sharp \lambda_3 \circ (4,2)$. This means that $(5,3) = (2\lambda_2 + 4\lambda_3,\ \lambda_2 + 2\lambda_3)$, and therefore

$$\begin{aligned} 5 &= 2\lambda_2 + 4\lambda_3 \\ 3 &= \ \ \lambda_2 + 2\lambda_3. \end{aligned}$$

Eliminating λ_2 or λ_3, we have 1=0. This is a contradiction. Thus, (5,3) is not a linear combination of (0,0), (2,1), (4,1).

1.1.9. Example. Let us consider the space $\mathscr{M}$. Is the vector $\begin{Vmatrix} 0, & -1 \\ 2, & 3 \end{Vmatrix}$ a linear combination of the vectors $\begin{Vmatrix} 1, & 2 \\ 1, & 1 \end{Vmatrix}, \begin{Vmatrix} 2, & 3 \\ 4, & 5 \end{Vmatrix}$?

[Answer: Yes, the coefficients are $\lambda_1 = -2, \lambda_2 = 1$].

1.1.10. Example. Let us consider the space $\mathscr{F}(\mathscr{R})$. Let us show that the vector $\sin^2 x$ is a linear combination of the vectors 1, $\cos 2x$. (Indeed, $\sin^2 x = \frac{1}{2}\circ 1 \sharp \left(-\frac{1}{2}\right)\circ \cos 2t$.)

1.1.11. Example. Let us remain in the space $\mathscr{F}(\mathscr{R})$. Let us show that $\overrightarrow{a}$ is a linear combination of $\overrightarrow{b}$ and $\overrightarrow{c}$, where

$$\begin{aligned} \overrightarrow{a} &= 2 + 3x + 5x^2 \\ \overrightarrow{b} &= 3 + 7x + 8x^2 \\ \overrightarrow{c} &= 1 - 6x + x^2. \end{aligned}$$

SOLUTION: We ask whether there are coefficients α_1, α_2 such that

$$2 + 3x + 5x^2 = \alpha_1 \circ (3 + 7x + 8x^2) \sharp \alpha_2 \circ (1 - 6x + x^2).$$

If it is so, we get

$$2 + 3x + 5x^2 = (3\alpha_1 + \alpha_2) + (7\alpha_1 - 6\alpha_2)x + (8\alpha_1 + \alpha_2)x^2 \Longrightarrow$$

$$\left.\begin{array}{rcl} 2 &=& 3\alpha_1 + \alpha_2 \\ 3 &=& 7\alpha_1 - 6\alpha_2 \\ 5 &=& 8\alpha_1 + \alpha_2 \end{array}\right\} \Longrightarrow \left.\begin{array}{rcl} 2 &=& 3\alpha_1 + \alpha_2 \\ 5 &=& 25\alpha_2 \\ -1 &=& -5\alpha_2 \end{array}\right\} \Longrightarrow \alpha_2 = \frac{1}{5},\ \alpha_1 = \frac{3}{5}$$

Thus, $\overrightarrow{a} = \frac{3}{5}\circ \overrightarrow{b} \sharp \frac{1}{5}\circ \overrightarrow{c}$.

The following definition introduces the subsets of a linear space that are linear spaces in their own right. It allows us to construct interesting new spaces.

1.1.12. Definition. Let L be a linear space and let M be a nonvoid subset of L. Then M is called *a linear subspace* (in short, a subspace) of L if the following two conditions are satisfied:

(i) for any pair $\overrightarrow{x} \in M$, $\overrightarrow{y} \in M$ we have $\overrightarrow{x} \sharp \overrightarrow{y} \in M$,

(ii) for any $\lambda \in R$ and $\overrightarrow{x} \in M$ we have $\lambda \circ \overrightarrow{x} \in M$.

Let us exhibit some illustrating examples.

1.1.13. Example. Let us consider the space R^2.

(i) Let $M = \{(1,2)\}$, $M \subset R^2$. Is M a linear subspace of R^2? We see that $(1,2) \sharp (1,2) = (2,4) \notin M$. Thus, M is not a linear subspace of R^2. Also, $7 \circ (1,2) = (7,14) \notin M$.

(ii) Let $M = \{(0,0\}$. Then M is a linear subspace of R^2. (In fact, this is the only possibility in a linear space when a singleton may be a subspace. Why?)

(iii) Let $M = \{(a,b) \in R^2 |\, b = 2a,\ a,b \in R\}$, $M \subset R^2$. Then M is a linear subspace of R^2.

(iv) Let $M = \{(a,b) \in R^2 | a \in R, b = 0\}$. Obviously, M is a subspace of R^2, too.

(v) Let $M = \{(a,b) \in R^2 |\ a \text{ is a rational number}\}$. Then M is *not* a subspace of R^2. (Indeed, $(1,1) \in M$, but $\pi \circ (1,1) = (\pi, \pi) \notin M$; the operation $\sharp$ is "closed" in M.)

1.1.14. Example. Consider the space $\mathscr{M}$.

(i) Let $M = \{\mathbf{A} \in \mathscr{M} | = = \}$. Thus, $A \in \mathscr{M} \iff$ the diagonal of $\mathbf{A}$ consists of zeros. Thus M is a subspace of $\mathscr{M}$.

(ii) Let $M = \{\mathbf{A} \in \mathscr{M} | \geq \}$. Then $\begin{Vmatrix} 1, & 2 \\ 3, & 4 \end{Vmatrix} \in M$, but $(-1) \circ \begin{Vmatrix} 1, & 2 \\ 3, & 4 \end{Vmatrix} = \begin{Vmatrix} -1, & -2 \\ -3, & -4 \end{Vmatrix} \notin M$. Thus M is not a subspace of M.

1.1.15. Example. Consider the space $\mathcal{F}(R)$.

(i) Let M be the set of all polynomials. Then M is a subspace of $\mathcal{F}(R)$. *We shall denote this important space by* $\mathcal{P}$ (the linear space of all polynomials).

(ii) Let $M = \{p(x) \in \mathcal{P} |$ the degree of $p(x) \leq 5\}$. Then M is a subspace of $\mathcal{F}(R)$. Obviously, M is also a subspace of $\mathcal{P}$.

(iii) Let $M = \{p(x) \in \mathcal{P} |$ the degree of $p(x)$ is $5\}$. Then M is *not* a subspace of $\mathcal{F}(R)$. Indeed, $x^5 \sharp (-x^5) = 0 \notin M$.

(iv) Let $M = \{p(x) \in \mathcal{P} |\ p(x)$ is divisible by the polynomial $(x-1)\}$. Then M is a subspace of $\mathcal{P}$.

(v) Let $M = \{f(x) \in \mathcal{F}(R) |\ f(1) = 0\}$. Then M is a subspace of $\mathcal{F}(R)$.

(vi) Let $M = \{f(x) \in \mathcal{F}(R) |\ f(1) \geq 0\}$. Then M is not a subspace of $\mathcal{F}(R)$.

Let us again consider Def 1.1.6. We immediately see that M is a subspace of $L \iff$ all the linear combinations of vectors from M belong to M. In short,

$$M \text{ is a subspace of } L \iff \forall\, \lambda_1, \lambda_2, \dots, \lambda_n \in R \quad \text{and} \quad \forall\, \overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_n \in M$$

we have $\lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_2 \circ \overrightarrow{x}_n \in M$.

If we are in need of a formal proof of the above assertion, we would proceed by induction. If $n = 1$ then the statement is obvious. Suppose that the statement is true for some k. Is it then true for $k+1$? Indeed, consider a vector $\overrightarrow{y} = \lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_k \circ \overrightarrow{x}_k \sharp \lambda_{k+1} \circ \overrightarrow{x}_{k+1}$. Since by the inductive assumption $\lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_k \circ \overrightarrow{x}_k \in M$, and $\lambda_{k+1} \circ \overrightarrow{x}_{k+1} \in M$ in view of M being a subspace of L, we see that $\overrightarrow{y}$ is a sum of two vectors from M. Thus, $\overrightarrow{y} \in M$.

1.1.16. Proposition. If M is a linear subspace of L, then M with the operations $\sharp\, , \circ$ inherited from L is a linear space in its own right.

Proof: Consider $(M, \sharp, \circ)$, where $\sharp, \circ$ are the operations in L. Obviously Ax. 1, 2, 5, 6, 7, 8 are all right. Let us verify Ax. 3 and Ax. 4. Since $M \neq 0$, there is a vector $\overrightarrow{x}$ with $\overrightarrow{x} \in M$. Since $0 \circ \overrightarrow{x} \in M$, we have $0 \circ \overrightarrow{x} = \overrightarrow{o} \in M$. Thus, $\overrightarrow{o} \in M$. (An alternative proof: $M \neq \emptyset \Rightarrow \overrightarrow{x} \in M \Rightarrow (-1) \circ \overrightarrow{x} \in M \Rightarrow (-1) \circ \overrightarrow{x} = (-\overrightarrow{x}) \Rightarrow \overrightarrow{x} \sharp (-\overrightarrow{x}) \in M \Rightarrow \overrightarrow{x} \sharp (-\overrightarrow{x}) = \overrightarrow{o} \in M$). Ax. 4 can be verified analogously (consider the reasoning used in the “alternative proof” of the previous statement).

□

1.1.17. Remark. Every linear space L contains the smallest linear subspace - the subspace $\{\overrightarrow{o}\}$ - and it also contains the largest linear subspace - the space L itself.

We are approaching a crucial notion of linear algebra.

1.1.18. Definition. Let $x_1, x_2, \dots, x_k$ be vectors in L. If $k \geq 2$, then the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ are said to be *linearly dependent* (in shorthand LD), if at least one of the vectors $x_1, x_2, \dots, x_k$ can be written as a linear combination of the others. If $k = 1$, then $\overrightarrow{x}_1$ is said to be linearly dependent when $\overrightarrow{x}_1 = \overrightarrow{o}$. The vectors $x_1, x_2, \dots, x_k$ are said to be *linearly independent* (in shorthand LI) if they are *not* linearly dependent.

Observe that two vectors are linearly dependent if and only if one of them is a scalar multiple of the other. The following statement will conveniently allow us to decide whether a family of vectors is LD or LI.

1.1.19. Proposition. The vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ are linearly independent (LI) exactly when the following implication holds true: If the equality $\lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_k \circ \overrightarrow{x}_k = \overrightarrow{o}$ is satisfied, then $\lambda_1 = 0, \lambda_2 = 0, \dots, \lambda_k = 0$. (In other words, the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ are LI if the only linear combination of the vectors that equals the null vector is the trivial combination.)

Proof: It is obvious that if $k = 1$ the proposition holds true. Suppose that $k \geq 2$. Let us denote by “LI” the property that the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ satisfy Prop. 1.1.18. Upon this convention, we have to show that LI $\iff$ “LI”. Let us first derive LI $\Longrightarrow$ “LI”. Hoping to reach a contradiction, assume that it is not the case. In other words, assume non "LI" and show non LI. This would establish the required implication LI $\Longrightarrow$ "LI". As we remember from the course of elementary logic, (LI $\Longrightarrow$ "LI") $\Longleftrightarrow$ (non "LI" $\Longrightarrow$ non LI).

After setting up the strategy of the proof, let us embark on showing LI $\Longrightarrow$ "LI". Intending to show non "LI" $\Longrightarrow$ non LI , suppose that there is a non-trivial linear combination of $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ which makes $\overrightarrow{o}$. Thus, suppose that $\lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_k \circ \overrightarrow{x}_k = \overrightarrow{o}$ for some λ_i ($1 \leq i < k$) distinct from 0. Without any loss of generality we may assume that $\lambda_i = \lambda_1$ (otherwise we would simply rename the coefficients and vectors). We then can write $\lambda_1 \circ \overrightarrow{x}_1 = (-\lambda_2) \circ \overrightarrow{x}_2 \sharp (-\lambda_3) \circ \overrightarrow{x}_3 \sharp \dots \sharp (-\lambda_k) \circ \overrightarrow{x}_k$ (prove it in detail making use of the axioms of a LS!). It follows that (supply the details again!)

$$\overrightarrow{x}_1 = \left(-\frac{\lambda_2}{\lambda_1}\right) \circ \overrightarrow{x}_2 \sharp \left(-\frac{\lambda_3}{\lambda_1}\right) \circ \overrightarrow{x}_3 \sharp \dots \sharp \left(-\frac{\lambda_k}{\lambda_1}\right) \circ \overrightarrow{x}_k.$$

This establishes that $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ are LI.

Secondly, let us show "LI" $\Longrightarrow$ LI. Again, we re-phrase the implication in the form non LI $\Longrightarrow$ non "LI". Assume therefore that among the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_n$ there exists at least one that is a linear combination of the others. We may assume that this vector is $\overrightarrow{x}_1$. Therefore we can write $\overrightarrow{x}_1 = \beta_2 \circ \overrightarrow{x}_2 \sharp \beta_3 \circ \overrightarrow{x}_3 \sharp \dots \sharp \beta_k \circ \overrightarrow{x}_k$. Adding $(-\overrightarrow{x})$ to both sides, we obtain

$$\overrightarrow{o} = (-\overrightarrow{x}_1) \sharp \beta_2 \overrightarrow{x}_2 \sharp \dots \sharp \beta_k \circ \overrightarrow{x}_k.$$

Applying Prop. 1.3, we see that $\overrightarrow{o} = (-1) \circ \overrightarrow{x}_1 \sharp \beta_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \beta_k \overrightarrow{x}_k$. This proves that the vectors are LD, i. e. they are not LI.

□

1.1.20. Example. Consider the space R^3 . Show that the vectors

$$\overrightarrow{x}_1 = (1, 2, 3), \qquad \overrightarrow{x}_2 = (1, -1, 0), \qquad \overrightarrow{x}_3 = (1, 1, 1)$$

are LI in R^3 .

SOLUTION: We shall make use of Prop. 1.1.19. Applying the standard Gaussian elimination, we successively obtain

$$\lambda_1 \circ (1,2,3) \sharp \lambda_2 \circ (1,-1,0) \sharp \lambda_3 \circ (1,1,1) = \overrightarrow{o} = (0,0,0) \Longrightarrow$$

$$\begin{array}{rcl} \lambda_1 + \lambda_2 + \lambda_3 = 0 & & \lambda_1 + \lambda_2 + \lambda_3 = 0 & & \lambda_1 + \lambda_2 + \lambda_3 = 0 & & \lambda_3 = 0 \\ 2\lambda_1 - \lambda_2 + \lambda_3 = 0 & \Longrightarrow & -3\lambda_2 - \lambda_3 = 0 & \Longrightarrow & -3\lambda_2 - \lambda_3 = 0 & \Longrightarrow & \lambda_2 = 0 \\ 3\lambda_1 \qquad + \lambda_3 = 0 & & -3\lambda_2 - 2\lambda_3 = 0 & & -\lambda_3 = 0 & & \lambda_1 = 0 \end{array}$$

From now on, let us agree to solve the systems of linear equations by Gaussian elimination as above. This technique is most efficient. Working first with the first equation, we eliminate the unknown x_1 from the remaining equations. Leaving then the first equation unchanged and working with the second equation, we eliminate the unknown x_2 from the remaining equations, etc.

1.1.21. Example. Decide whether the vectors $\overrightarrow{x}_1 = (2,1,3), \overrightarrow{x}_2 = (-1,1,0), \overrightarrow{x}_3 = (1,0,1)$, $\overrightarrow{x}_4 = (2,2,4)$ are LI or LD.

SOLUTION:

$$\lambda_1 \circ (2,1,3) \sharp \lambda_2 \circ (-1,1,0) \sharp \lambda_3 \circ (1,0,1) \sharp \lambda_4 \circ (2,2,4) = (0,0,0) \Longrightarrow$$

$$\Rightarrow \begin{array}{r} 2\lambda_1 - \lambda_2 + \lambda_3 + 2\lambda_4 = 0 \\ \lambda_1 + \lambda_2 \qquad + 2\lambda_4 = 0 \\ 3\lambda_1 \qquad + \lambda_3 + 4\lambda_4 = 0 \end{array} \Rightarrow \begin{array}{r} 2\lambda_1 - \lambda_2 + \lambda_3 + 2\lambda_4 = 0 \\ -3\lambda_2 + \lambda_3 - 2\lambda_4 = 0 \\ -3\lambda_2 + \lambda_3 - 2\lambda_4 = 0 \end{array} \Rightarrow \begin{array}{r} 2\lambda_1 - \lambda_2 + \lambda_3 + 2\lambda_4 = 0 \\ -3\lambda_2 + \lambda_3 - 2\lambda_4 = 0 \\ 0 = 0 \end{array}$$

It follows that we may put, e. g., $\lambda_4 = 1, \lambda_3 = -1$, and compute from the bottom to the top, obtaining $\lambda_2 = -1, \lambda_1 = -1$. We conclude that there is a nontrivial linear combination of $\overrightarrow{x}_1, \overrightarrow{x}_2, \overrightarrow{x}_3, \overrightarrow{x}_4$ giving $\overrightarrow{o}$. Thus, the vectors are LD.

1.1.22. Remark. Observe that the polynomials $1, x, x^2, \dots, x^n$, for a given $n \in N$, are always LI when viewed as vectors of $\mathcal{F}(R)$ or $\mathcal{P}$.

1.1.23. Example. Consider the space $\mathcal{F}(R)$. Show that the functions $f_1 = -\frac{1}{2}\sin^2 x, f_2 = 3\cos^2 x, f_3 = 2$ are LD.

SOLUTION: $(-2)\circ f_1 \sharp (\frac{1}{3})\circ f_2 \sharp (-\frac{1}{2})\circ f_3 = \overrightarrow{o}$.

1.1.24. Proposition. If the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$ are LI in L then they are mutually distinct and $\overrightarrow{x}_i \neq \overrightarrow{o}$ for any i $(i \leq k)$.

PROOF: If $\overrightarrow{x}_1 = \overrightarrow{x}_2$, then $1\circ \overrightarrow{x}_1 \sharp (-1)\circ \overrightarrow{x}_2 \sharp 0\circ \overrightarrow{x}_3 \sharp \dots \sharp 0\circ \overrightarrow{x}_k = \overrightarrow{o}$. Thus, $x_1, x_2, \dots x_k$ are LD - a contradiction. Further if $\overrightarrow{x} = \overrightarrow{o}$, then $1\circ \overrightarrow{x}_1 \sharp 0\circ \overrightarrow{x}_2 \sharp \dots \sharp 0\circ \overrightarrow{x}_k = \overrightarrow{o}$ - a contradiction again.

□

1.1.25. Proposition. If we “add” to a given vector of a LI family a linear combination of the other vectors from that family and if we leave the other vectors unchanged, we obtain a LI family.

PROOF: Without any loss of generality, let us assume that we “add” to the first vector, $\overrightarrow{x}_1$, of the family. So we start with an LI family, $\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k$, and we are to show that the vectors $\overrightarrow{x}_1 \sharp \mu_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \mu_k \circ \overrightarrow{x}_k, \overrightarrow{x}_2, \overrightarrow{x}_3, \dots \overrightarrow{x}_k$ are LI. Using Prop. 1.1.18, we consecutively get $\lambda_1 \circ (\overrightarrow{x}_1 \sharp \mu_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \mu_k \circ \overrightarrow{x}_k) \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \dots \sharp \lambda_k \circ \overrightarrow{x}_k = \overrightarrow{o} \Longrightarrow \lambda_1 \circ \overrightarrow{x}_1 \sharp (\lambda_1\mu_2 + \lambda_2)\circ \overrightarrow{x}_2 \sharp \dots \sharp (\lambda_1\mu_k + \lambda_k)\circ \overrightarrow{x}_k = \overrightarrow{o} \Longrightarrow \lambda_1 = 0,\ \lambda_1\mu_2 + \lambda_2 = 0, \dots, \lambda_1\mu_k + \lambda_k = 0 \Longrightarrow \lambda_1 = 0, \lambda_2 = 0, \dots \lambda_k = 0$.

□

1.1.26. Proposition. If the vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k$ are LI and if $\lambda \neq 0$, then the vectors $\lambda \circ \overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k$ are LI.

PROOF: Obvious (use Prop. 1.1.18).

The following definition will turn out to be related to the "internal" construction of a linear space.

1.1.27. Definition. Let L be a linear space. Let M be a subset of L, $M \neq \emptyset$. Then the set of all linear combinations of vectors from M (i. e., the set of all vectors $\overrightarrow{x} = \lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \ldots \sharp \lambda_k \circ \overrightarrow{x}_k$, where $k \in N, \overrightarrow{x}_i \in M (i \leq k), \lambda_i \in R (i \leq k)$) is called a *linear span* of M. (We denote the span of M by $\operatorname{Span} M$.)

1.1.28. Proposition. $\operatorname{Span} M$ is a linear subspace of L. Moreover, $\operatorname{Span} M$ is *the smallest* linear subspace of L among those linear subspaces of L that contain M.

PROOF: We are to show that $\operatorname{Span} M$ is closed (in itself!) under the formation of the operations $\sharp$ and $\circ$. If $\overrightarrow{x}, \overrightarrow{y} \in \operatorname{Span} M$, then $\overrightarrow{x} = \lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \ldots \sharp \lambda_k \circ \overrightarrow{x}_k$, where each $\overrightarrow{x}_i (i \leq k)$ belongs to M and $\overrightarrow{y} = \mu_1 \circ \overrightarrow{y}_1 \sharp \mu_2 \circ \overrightarrow{y}_2 \sharp \ldots \sharp \mu_k \circ \overrightarrow{y}_k$, where each $\overrightarrow{y}_i (i \leq k)$ belongs to M. Therefore $\overrightarrow{x} \sharp \overrightarrow{y}$ can again be expressed as a linear combination of vectors from M! (Indeed, $\overrightarrow{x} \sharp \overrightarrow{y} = \lambda_1 \circ \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \ldots \sharp \lambda_k \circ \overrightarrow{x}_k \sharp \mu_1 \circ \overrightarrow{y}_1 \sharp \mu_2 \circ \overrightarrow{y}_2 \sharp \ldots \sharp \mu_k \circ \overrightarrow{y}_k$ where each vector occurring on the right-hand side belongs to M.) Similarly, if $\overrightarrow{x} \in \operatorname{Span} M$ then $\lambda \circ \overrightarrow{x} \in \operatorname{Span} M$. Finally if L' is a linear subspace of $L, M \subset L'$, then all linear combinations of vectors from M must remain in L'. Thus, $\operatorname{Span} M \subset L'$.

□

1.1.29. Remark. Observe that, for any subset M of L, $\operatorname{Span}(\operatorname{Span} M) = \operatorname{Span} M$. Indeed, for any linear subspace K of L we have $\operatorname{Span} K = K$.

1.1.30. Proposition. If $\overrightarrow{y}_1 = \overrightarrow{x}_1 \sharp \lambda_2 \circ \overrightarrow{x}_2 \sharp \lambda_3 \circ \overrightarrow{x}_3 \sharp \ldots \sharp \lambda_k \circ \overrightarrow{x}_k$ for some scalars $\lambda_2, \lambda_3, \ldots, \lambda_k$, then $\operatorname{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \overrightarrow{x}_3, \ldots, \overrightarrow{x}_k\} = \operatorname{Span}\{\overrightarrow{y}_1, \overrightarrow{x}_2, \overrightarrow{x}_3, \ldots, \overrightarrow{x}_k\}$. If $\overrightarrow{y}_1 = \lambda \circ \overrightarrow{x}_1$, where $\lambda \neq 0$, then $\operatorname{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k\} = \operatorname{Span}\{\overrightarrow{y}_1, \overrightarrow{x}_2 \ldots, \overrightarrow{x}_k\}$.

PROOF: We must show that if a vector is expressed as a linear combination of $\overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k$, then it is also expressed as a linear combination of $\overrightarrow{y}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k$, and vice versa. If $\overrightarrow{x} = \mu_1 \circ x_1 \sharp \ldots \sharp \mu_k \circ x_k$ and if we write $\overrightarrow{x}_1 = \overrightarrow{y}_1 \sharp (-\lambda_2) \circ \overrightarrow{x}_2 \sharp \ldots \sharp (-\lambda_k) \circ \overrightarrow{x}_k$, then $\overrightarrow{x} = \mu_1 \circ (\overrightarrow{y}_1 \sharp (-\lambda_2) \circ \overrightarrow{x}_2 \sharp \ldots \sharp (-\lambda_k) \circ \overrightarrow{x}_k) \sharp \ldots \sharp \mu_k \circ \overrightarrow{x}_k = \mu_1 \circ \overrightarrow{y}_1 \sharp (-\mu_1 \lambda_2 + \mu_2) \circ x_2 \sharp \ldots \sharp (\mu_1 \lambda_k + \mu_k) \circ x_k$. This proves that $\operatorname{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \ldots \overrightarrow{x}_k\} \subset \operatorname{Span}\{\overrightarrow{y}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k\}$. Analogously, $\operatorname{Span}\{\overrightarrow{y}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_k\} \subset \operatorname{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_n\}$. The rest is easy.

The notion of a LI and LD family can also be defined for infinite families. Let us recall this definition before the fundamental Def. 1.1.32 that follows.

1.1.31. Definition. A (possibly infinite) subset M of L is called (linearly) independent (LI) if every choice of finitely many vectors of M gives a LI family in L. If M is not LI, then it is called linearly dependent (LD).

We are coming to another main notion of linear algebra.

1.1.32. Definition. Let L be a linear space. A set B, $B \subset L$, is called *a basis* of L if B is linearly independent and $\operatorname{Span} B = L$.

1.1.33. Example.

1. The set $B = \{(1,0),(0,1)\}$ is a basis of R^2. More generally, the set $B = \{(1,0,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,\dots,1,0),(0,0,\dots,0,1)\}$ is a basis of R^n. This basis is called *the standard basis of* R^n.
2. The set
$$B = \left\{ \begin{Vmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{Vmatrix}, \dots, \begin{Vmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{Vmatrix} \right\}$$
is a basis of $\mathscr{M}$. This basis has $3 \cdot 2 = 6$ elements. More generally, in the space $\mathscr{M}$ there is an analogous basis consisting of $m \cdot n$ elements. This basis is called *the standard basis of* $\mathscr{M}$.
3. The space $\mathcal{P}$ of all polynomials does *not* have a *finite* basis. Indeed, if $p_1, p_2, \dots, p_k$ are polynomials and if $m = \max\{\text{ dg } p_i (i \le k)\}$, then the polynomial x^{m+1} does not belong to $\operatorname{Span}\{p_1, p_2, \dots, p_n\}$. On the other hand, the set $\{1, x, x^2, \dots, x^n, \dots\}$ constitutes a basis of $\mathcal{P}$.

The standard bases are "very nice" bases. There are also less nice bases. For instance the set $B = \{(1,1),(1,-2)\}$ is a basis of R^2 (check it on your own). The situation may be even more complicated.

1.1.34. Example.

a) Find a basis in R^3 of $L = \operatorname{Span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4\}$, where $\vec{a}_1 = (2,0,-1)$, $\vec{a}_2 = (1,-1,2)$, $\vec{a}_3 = (3,-1,1)$, $\vec{a}_4 = (4,-2,3)$.

Solution: Let us use the symbol $\sim$ to denote the coincidence of the corresponding Spans. Thus,
$$\begin{array}{l} a_1 = (2,0,-1) \\ a_2 = (1,-1,2) \\ a_3 = (3,-1,1) \\ a_4 = (4,-2,3) \end{array} \sim \begin{array}{l} b_1 = (1,-1,2) = \vec{a_2} \\ b_2 = (0,2,-5) = \vec{a}_1 - 2\vec{a}_2 \\ b_3 = (3,-1,1) = \vec{a}_3 \\ b_4 = (4,-2,3) = \vec{a}_4. \end{array}$$
Indeed, by Prop. 1.1.30 we have $\operatorname{Span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4\} = \operatorname{Span}\{\vec{b}_1, \vec{b}_2, \vec{b}_3, \vec{b}_4\}$. Furthermore,
$$\begin{array}{l} (1,-1,2) \\ (0,2,-5) \\ (3,-1,1) \\ (4,-2,3) \end{array} \sim \begin{array}{l} (1,-1,2) \\ (0,2,-5) \\ (0,2,-5) \\ (0,2,-5) \end{array} \sim \begin{array}{l} (1,-1,2) \\ (0,2,-5) \end{array}$$
By Prop. 1.1.30 again, we infer that
$$L = \operatorname{Span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4\} = \operatorname{Span}\{(1,-1,2),(0,2,-5)\}.$$
It follows that the set $B = \{(1,-1,2),(0,2,-5)\}$ is a basis of L.

b) Suppose that we are given the following matrices:

$$A_1 = \left\|\begin{matrix} 1, & 1 \\ -1, & 2 \end{matrix}\right\|, A_2 = \left\|\begin{matrix} 2, & 1 \\ 0, & 1 \end{matrix}\right\|, A_3 = \left\|\begin{matrix} 1, & 0 \\ 1, & 1 \end{matrix}\right\|, A_4 = \left\|\begin{matrix} 4, & 2 \\ 0, & 4 \end{matrix}\right\|, A_5 = \left\|\begin{matrix} 3, & 1 \\ 1, & 2 \end{matrix}\right\|$$

Find a basis in $\mathscr{M}$ of $L = \mathrm{Span}\{A_1, A_2, A_3, A_4, A_5\}$.

SOLUTION:

$$\left\|\begin{matrix} 1, & 1 \\ -1, & 2 \end{matrix}\right\| \left\|\begin{matrix} 2, & 1 \\ 0, & 1 \end{matrix}\right\| \left\|\begin{matrix} 1, & 0 \\ 1, & 1 \end{matrix}\right\| \left\|\begin{matrix} 4, & 2 \\ 0, & 4 \end{matrix}\right\| \left\|\begin{matrix} 3, & 1 \\ 1, & 2 \end{matrix}\right\| \sim$$

$$\sim \left\|\begin{matrix} 1, & 1 \\ -1, & 2 \end{matrix}\right\| \left\|\begin{matrix} 0, & -1 \\ 2, & -3 \end{matrix}\right\| \left\|\begin{matrix} 0, & -1 \\ 2, & -1 \end{matrix}\right\| \left\|\begin{matrix} 0, & -2 \\ 4, & -4 \end{matrix}\right\| \left\|\begin{matrix} 0, & -2 \\ 4, & -4 \end{matrix}\right\| \sim$$

$$\sim \left\|\begin{matrix} 1, & 1 \\ -1, & 2 \end{matrix}\right\| \left\|\begin{matrix} 0, & -1 \\ 2, & -3 \end{matrix}\right\| \left\|\begin{matrix} 0, & 0 \\ 0, & 2 \end{matrix}\right\| \left\|\begin{matrix} 0, & 0 \\ 0, & 2 \end{matrix}\right\| \left\|\begin{matrix} 0, & 0 \\ 0, & 2 \end{matrix}\right\| \Rightarrow B = \left\{ \left\|\begin{matrix} 1, & 1 \\ -1, & 2 \end{matrix}\right\| \left\|\begin{matrix} 0, & -1 \\ 2, & -3 \end{matrix}\right\| \left\|\begin{matrix} 0, & 0 \\ 0, & 1 \end{matrix}\right\| \right\}$$

is a basis of L.

c) Let $L = \mathrm{Span}\{\overrightarrow{p}_1, \overrightarrow{p}_2, \overrightarrow{p}_3, \overrightarrow{p}_4, \overrightarrow{p}_5\}$ in $\mathcal{P}$, where

$$\begin{aligned} \overrightarrow{p}_1 &= x^3 + x^2 + 3x + 2 \\ \overrightarrow{p}_2 &= 2x^2 + 4x + 4 \\ \overrightarrow{p}_3 &= -x^3 + x^2 + x + 2 \\ \overrightarrow{p}_4 &= x^3 + x + 1 \\ \overrightarrow{p}_5 &= x^2 + 2x + 1 \end{aligned}$$

Find a basis of L.

SOLUTION:

$$\begin{array}{rcrcrcr} x^3 & + & x^2 & + & 3x & + & 2 \\ & & 2x^2 & + & 4x & + & 4 \\ -x^3 & + & x^2 & + & x & + & 2 \\ x^3 & & & + & x & + & 1 \\ & & x^2 & + & 2x & + & 1 \end{array} \sim \begin{array}{r} x^3 + x^2 + 3x + 2 \\ 2x^2 + 4x + 4 \\ 2x^2 + 4x + 4 \\ -x^2 - 2x - 1 \\ x^2 + 2x + 1 \end{array} \sim \begin{array}{r} x^3 + x^2 + 3x + 2 \\ 2x^2 + 4x + 4 \\ 0 \\ +2 \\ +2 \end{array} \sim$$

$$\sim \begin{array}{r} x^3 + x^2 + 3x + 2 \\ 2x^2 + 4x + 4 \\ 2 \\ +2 \end{array} \sim \begin{array}{r} x^3 + x^2 + 3x + 2 \\ x^2 + 2x + 2 \\ 1 \end{array}$$

We see that the set $B = \{x^3 + x^2 + 3x + 2, x^2 + 2x + 2, 1\}$ constitutes a basis of L.

A question naturally arises whether every linear space L has a basis. It is true provided L is nontrivial (i. e., $L \neq \{\overrightarrow{o}\}$).

1.1.35. Theorem. Every nontrivial linear space L has a basis.

PROOF: Let B be a linearly independent family of vectors in L and let B be *maximal* (i. e., let B be such that if we add a vector, $\overrightarrow{v}$, to B then the family $B \cup \{\overrightarrow{v}\}$ fails to be linearly independent). Such a maximal family of vectors exists in any nontrivial space L. (The last statement can be rigorously proved in the theory of sets – one however has to make friends with the axiom of choice. In the courses for "practical" men we usually allow ourselves to understand the existence of maximal family B as intuitively plausible.) The properties of B are then easily seen. Since B is LI, we only have to show that $\operatorname{Span} B = L$. If this is not the case, we can find a vector $\overrightarrow{v}$ such that $\overrightarrow{v} \in L$ and $\overrightarrow{v} \notin \operatorname{Span} B$. We prove now that $B \cup \{\overrightarrow{v}\}$ is still LI. Suppose that $B \cup \{\overrightarrow{v}\}$ is LD. Then there is a nontrivial linear combination $\alpha_1 \circ \overrightarrow{v}_1 \sharp \alpha_2 \circ \overrightarrow{v}_2 \sharp \ldots \sharp \alpha_k \circ \overrightarrow{v}_k \sharp \alpha_{k+1} \circ \overrightarrow{v} = \overrightarrow{o}$. If $\alpha_{k+1} = 0$, then at least one of α_i $(i \leq k)$ is nonzero - a contradiction. If $\alpha_{k+1} \neq 0$ then $\overrightarrow{v} = \left(-\dfrac{\alpha_1}{\alpha_{k+1}}\right) \circ \overrightarrow{v}_1 \sharp \ldots \sharp \left(-\dfrac{\alpha_k}{\alpha_{k+1}}\right) \circ \overrightarrow{v}_k \Longrightarrow$ $\overrightarrow{v} \in \operatorname{Span} B$. This is a contradiction.

□

Finite dimensional linear spaces.

We will mostly deal with spaces that have a finite basis. It seems worth studying them in more detail. The entry point is the following remarkable result.

1.1.36. Proposition. (The Steinitz interchange lemma). Let L be a linear space. Let $s \in N$ and let $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \ldots, \overrightarrow{b}_s\}$ be a family of LI vectors in L. Let $r \in N$ and let $M = \{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_r\}$ be a family of vectors in L such that $\operatorname{Span} M = L$. Then

(i) $r \geq s$,

(ii) There is a subset P of M such that P consists of $r - s$ vectors and such that $\operatorname{Span}(P \cup \{\overrightarrow{b}_1, \overrightarrow{b}_2, \ldots, \overrightarrow{b}_s\}) = L$.

PROOF: We have to prove that as many as s vectors in M can be replaced by the vectors $\overrightarrow{b}_1, \overrightarrow{b}_2, \ldots, \overrightarrow{b}_s$ without reducing the span. We will proceed by induction over the number s. Suppose first that $s = 1$. Then $\overrightarrow{b}_1 \neq \overrightarrow{o}$. Since $L = \operatorname{Span} M = \operatorname{Span}\{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_r\}$, we have $\overrightarrow{b}_1 \in \operatorname{Span}\{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_r\}$. Thus, $\overrightarrow{b}_1 = \lambda_1 \circ \overrightarrow{a}_1 \sharp \lambda_2 \circ \overrightarrow{a}_2 \sharp \ldots \sharp \lambda_r \circ \overrightarrow{a}_r$ for some scalars $\lambda_1, \lambda_2, \ldots, \lambda_r$. Then we have $r \geq 1 = s$. Further, at least one λ_i $(i \leq r)$ has to be nonzero. Let us assume that $\lambda_1 \neq 0$. It follows that

$$\overrightarrow{a}_1 = \left(\frac{1}{\lambda_1}\right) \circ \overrightarrow{b_1} \sharp \left(-\frac{\lambda_2}{\lambda_1}\right) \circ \overrightarrow{a}_2 \sharp \ldots \sharp \left(-\frac{\lambda_r}{\lambda_1}\right) \overrightarrow{a}_r \Longrightarrow L = \operatorname{Span}\{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_r\} =$$

$$= \operatorname{Span}\left\{\left(\frac{1}{\lambda_1}\right) \circ \overrightarrow{b_1} \sharp \left(-\frac{\lambda_2}{\lambda_1}\right) \circ \overrightarrow{a}_2 \sharp \ldots \sharp \left(-\frac{\lambda_r}{\lambda_1} \overrightarrow{a}_r\right), \overrightarrow{a}_2, \overrightarrow{a}_3, \ldots, \overrightarrow{a}_r\right\} =$$

=(Prop. 1.1.29) $= \operatorname{Span}\{\overrightarrow{b}_1, \overrightarrow{a}_2, \overrightarrow{a}_3, \ldots, \overrightarrow{a}_r\}$.

Let us proceed to the second step of induction. Assuming that the statement is true for some $s \in N$, we want to show that the statement is true for $\lambda + 1$. Suppose that we are given $s + 1$ vectors, some $\overrightarrow{b}_1, \overrightarrow{b}_2, \ldots, \overrightarrow{b}_s, \overrightarrow{b}_{s+1}$, and suppose that these vectors are linearly independent. We have to "substitute" the vectors into $M = \{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_r\}$ without changing the span. According to the inductive assumption, the result is true for s. Thus, we may suppose that we have interchanged $\overrightarrow{a}_1$ for $\overrightarrow{b}_1$, $\overrightarrow{a}_2$ for $\overrightarrow{b}_2, \ldots, \overrightarrow{a}_s$ for $\overrightarrow{b}_s$, and

remained with the equality $L = \mathrm{Span}\{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s, \overrightarrow{a}_{s+1}, \dots, \overrightarrow{a}_r\}$. Thus, $\overrightarrow{b}_{s+1} = \mu_1 \circ \overrightarrow{b}_1 \sharp \mu_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp \mu_s \circ \overrightarrow{b}_s \sharp \mu_{s+1} \circ \overrightarrow{a}_{s+1} \sharp \dots, \sharp \mu_r \circ \overrightarrow{a}_r$. It is not possible that all $\mu_{s+1}, \mu_{s+2}, \dots, \mu_r$ be zero – this would mean that $\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s, \overrightarrow{b}_{s+1}$ are LD which is absurd. Without any loss of generality we assume that $\mu_{s+1} \neq 0$. Then $\overrightarrow{a}_{s+1} = \left(\frac{1}{\mu_{s+1}}\right) \circ \overrightarrow{b}_{s+1} \sharp \left(-\frac{\mu_1}{\mu_{s+1}}\right) \circ \overrightarrow{b}_1 \sharp \dots \sharp \left(-\frac{\mu_s}{\mu_{s+1}}\right) \circ \overrightarrow{b}_s \sharp \left(-\frac{\mu_{s+2}}{\mu_{s+1}}\right) \circ \overrightarrow{a}_{s+2} \sharp \dots \sharp \left(-\frac{\mu_r}{\mu_{s+1}}\right) \circ \overrightarrow{a}_r$. Now, using Prop. 1.1.29, $\mathrm{Span}\{\overrightarrow{b}_1. \overrightarrow{b}_2, \dots, \overrightarrow{b}_s, \overrightarrow{b}_{s+1}, \overrightarrow{a}_{s+2}, \dots, \overrightarrow{a}_r\} =$
$= \mathrm{Span}\{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s, \overrightarrow{a}_{s+1}, \overrightarrow{a}_{s+2}, \dots, \overrightarrow{a}_r\} = L$.

□

The above result has some important consequences.

1.1.37. Theorem. If B_1, B_2 are bases of a finite dimensional space L, then $cardB_1 = cardB_2$ i. e., both bases B_1, B_2 have an equal number of elements.

PROOF: Let $B_1 = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_n\}$ and $B_2 = \{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_m\}$. Since $\mathrm{Span}\, B_2 = L$ and since B_1 is a LI family we have $m \geq n$ (Prop. 1.1.35) If we exchange the rôles of B_1, B_2, we obtain $n \leq m$.

□

1.1.38. Definition. Suppose that $B = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_n\}$ is a basis of L. The number n (= the number of vectors in B) is called *the dimension of* L. We write $\dim L = n$. (If $L = \{\overrightarrow{o}\}$, we agree to let $\dim L = 0$.) If L does not have any finite basis, we say that L is infinite dimensional ($\dim L = \infty$).

1.1.39. Remark. The definition of dimension is correct in view of Prop. 1.1.36! Indeed, it does not matter which basis we take to determine $\dim L$.

1.1.40. Proposition. Suppose that $p \in N$ and $p > \dim L$. Suppose further that $S = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_p\}$ is a subset of L. Then S is LD. A corollary: If L is finite dimensional, then $\dim L$ is the number of vectors in a maximal family of LI vectors in L.

PROOF: If $\dim L = n$ and $n \geq 1$, then there is a basis of L that possesses n elements. If S is LI then, by Prop. 1.1.36, $n = \dim L \geq p$. This is absurd.

□

1.1.41. Proposition. If M is a linear subspace of L, then $\dim M \leq \dim L$.

PROOF: If $M = \{\overrightarrow{o}\}$ then Prop. 1.20 is trivial. If $M \neq \overrightarrow{o}$, then M has a basis $B = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s\}$ consisting of LI vectors. According to Prop. 1.1.36, if $\dim L = n$ then $s \leq n$.

□

Important observation. Every LI family in L can be completed to a basis of L. (Indeed, if $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s\}$ is LI in L and if $A = \{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_n\}$ is an arbitrary basis of L, then we can replace some vectors in A by $\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s$ without changing the span. The resulting family $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_s, \overrightarrow{a}_{s+1}, \overrightarrow{a}_{s+2}, \dots, \overrightarrow{a}_n\}$ has to be a basis of L (prove in detail!).

1.1.42. Theorem. Let M be a linear subspace of L and let $\dim M = \dim L$. Then $M = L$.

PROOF: By Prop. 1.1.36 we can replace *the whole basis* of L with a basis of M.

1.1.43. Definition. Let $B = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_n\}$ be a basis of L. Take a vector $\overrightarrow{x} \in L$. The real numbers $x_1, x_2, \dots, x_n$ such that $\overrightarrow{x} = x_1 \circ \overrightarrow{b}_1 \sharp x_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp x_n \circ \overrightarrow{b}_n$ are called *the coordinates of* $\overrightarrow{x}$ *with respect to* B. (We write $\overrightarrow{x} = (x_1, x_2, \dots, x_n)_B$.)

1.1.44. Proposition. For any basis B of L and any vector $\overrightarrow{x} \in L$, the coordinates of $\overrightarrow{x}$ with respect to B are determined uniquely.

PROOF: Suppose that $B = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_n\}$ is a basis of L. Suppose that $\overrightarrow{x} \in L$. Hoping to reach a contradiction, assume that $\overrightarrow{x}$ allows for two expressions in coordinates with respect to B. Thus, assume that $\overrightarrow{x} = x_1 \circ \overrightarrow{b}_1 \sharp x_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp x_n \circ \overrightarrow{b}_n = y_1 \circ \overrightarrow{b}_1 \sharp y_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp y_n \circ \overrightarrow{b}_n$, where the coordinates do not coincide. Then $(x_1 - y_1) \circ \overrightarrow{b}_1 \sharp (x_2 - y_2) \circ \overrightarrow{b}_2 \sharp \dots \sharp (x_n - y_n) \circ \overrightarrow{b}_n = \overrightarrow{o}$. Since B is LI, this combination has to be trivial. We infer that $x_1 - y_1 = 0, x_2 - y_2 = 0, \dots, x_n - y_n = 0$ and therefore $x_1 = y_1, x_2 = y_2, \dots, x_n = y_n$. This contradicts the assumption.

□

1.1.45. Remark. Obviously, a vector may have different coordinates with respect to different bases. For instance, $(3, 2) = 3 \circ (1, 0) \sharp 2 \circ (0, 1)$, but $(3, 2) = 1 \circ (1, 1) \sharp (-1) \circ (-2, -1)$. Thus, (3,2) has the coordinates (3, 2) with respect to the standard basis $\{(1, 0), (0, 1)\}$, whereas (3,2) has the coordinates (1, -1) with respect to the basis $\{(1, 1), (-2, -1)\}$.

1.1.46. Example. Show that the vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ constitute a basis, B, of R^3. Find the coordinates of $\overrightarrow{v}$ with respect to B. We set $\overrightarrow{v}_1 = (2, 1, -3), \overrightarrow{v}_2 = (3, 2, -5), \overrightarrow{v}_3 = (1, -1, 1), \overrightarrow{v} = (7, 6, -14)$.

SOLUTION: It suffices to show that $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ are LI. Since $\dim R^3$ is known to be 3, we use Prop. 1.1.25 and Prop. 1.1.26 to obtain

$$\begin{array}{llll} (2,1,-3) & (1,-1,1) & (1,-1,1) & (1,-1,\ \ 1) \\ (3,2,-5) & \approx (2,1,-3) & \approx (0,3,-5) & \approx (0,\ \ 3,-5) \\ (1,-1,1) & (3,2,-5) & (0,5,-8) & (0,\ \ 0,\ \ 1). \end{array}$$

Since the last vectors are LI, the original three vectors have to be LI, too. Thus, B is a basis.

Further, write $\overrightarrow{v} = v_1 \circ (2, 1, -3) \sharp v_2 \circ (3, 2, -5) \sharp v_3 \circ (1, -1, 1)$. Then

$$(7, 6, -14) = (2v_1 + 3v_2 + v_3, v_1 + 2v_2 - v_3, -3v_1 - 5v_2 + v_3).$$

Expressing the corresponding equations in the matrix form, we have

$$\left|\begin{array}{rrr|r} 2 & 3 & 1 & 7 \\ 1 & 2 & -1 & 6 \\ -3 & -5 & 1 & -14 \end{array}\right| \approx \left|\begin{array}{rrr|r} 2 & 3 & 1 & 7 \\ 0 & -1 & 3 & -5 \\ 0 & -1 & 5 & -7 \end{array}\right| \approx \left|\begin{array}{rrr|r} 2 & 3 & 1 & 7 \\ 0 & -1 & 3 & -5 \\ 0 & 0 & 2 & -2 \end{array}\right|$$

(the symbol $\approx$ means here the equivalence as far as the solution is concerned). It follows that $v_3 = -1,\ v_2 = 2$ and $v_1 = 1$. Thus, $\overrightarrow{v} = (1, 2, -1)_B$.

1.1.47. Remark. Do not overlook that in the notion of coordinates we understand the basis as *an ordered basis!* Thus, $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \overrightarrow{b}_3\}$ and $\{\overrightarrow{b}_3, \overrightarrow{b}_1, \overrightarrow{b}_2\}$ are different bases in this sense.

1.1.48. Example. We have shown that the dimension of the space R^n is n. We may consider the standard basis $\{(1,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,\dots,1,0),(0,0,\dots,0,1)\}$. The space $\mathscr{M}$ has the dimension $m \cdot n$ (consider the standard basis). The space $\mathcal{F}(R)$ of all functions on R is infinite dimensional. Indeed, even its subspace $\mathcal{P}$ of all polynomials is infinite dimensional (see Example 1.1.33, 3).

1.1.49. Example. Find a basis of the linear space $\mathscr{M}$ of all symmetric 3×3 matrices. (A matrix $A \in \mathscr{M}$ is called *symmetric* if $a_{ij} = a_{ji}$ for any pair i, j. Observe that the set $\mathscr{M}$ forms a linear subspace of $\mathscr{M}$.)

SOLUTION: Consider a typical matrix of $\mathscr{M}$, some $\mathbf{A} = \left\|\begin{matrix} a, & b, & c \\ b, & d, & e \\ c, & e, & f \end{matrix}\right\|$. We can write

$$\mathbf{A} = \left\|\begin{matrix} a, & b, & c \\ b, & d, & e \\ c, & e, & f \end{matrix}\right\| = a \circ \left\|\begin{matrix} 1, & 0, & 0 \\ 0, & 0, & 0 \\ 0, & 0, & 0 \end{matrix}\right\| \sharp\, b \circ \left\|\begin{matrix} 0, & 1, & 0 \\ 1, & 0, & 0 \\ 0, & 0, & 0 \end{matrix}\right\| \sharp\, c \circ \left\|\begin{matrix} 0, & 0, & 1 \\ 0, & 0, & 0 \\ 1, & 0, & 0 \end{matrix}\right\| \sharp$$

$$\sharp\, d \circ \left\|\begin{matrix} 0, & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 0 \end{matrix}\right\| \sharp\, e \circ \left\|\begin{matrix} 0, & 0, & 0 \\ 0, & 0, & 1 \\ 0, & 1, & 0 \end{matrix}\right\| \sharp\, f \circ \left\|\begin{matrix} 0, & 0, & 0 \\ 0, & 0, & 0 \\ 0, & 0, & 1 \end{matrix}\right\| .$$

The matrices on the left-hand side are obviously LI and therefore constitute a basis of $\mathscr{M}$. Thus, $\dim \mathscr{M} = $. (Generalize: $\dim \mathscr{M} = + + + \dots + = \frac{(+)}{}$.)

1.1.50. Example. Let

$$L_1 = \operatorname{Span}\{(1,-1,2),(1,0,1),(3,-1,0)\}$$
$$L_2 = \operatorname{Span}\{(2,-1,3),(1,-2,3),(3,-1,4)\}$$

(i) Check that $L_1 \cap L_2$ is a linear subspace of L (this fact is true in general, see Ex. 16 at the end of this chapter).

(ii) Find a basis of $L_1 \cap L_2$.

SOLUTION: It can be easily shown that if both L_1, L_2 satisfy the properties of Def. 1.1.12, than so does the set intersection $L_1 \cap L_2$. Thus, $L_1 \cap L_2$ is a linear subspace of R^3. Suppose that $\overrightarrow{x} \in L_1 \cap L_2$. Then

$$\overrightarrow{x} = a\circ(1,-1,2) \sharp\, b\circ(1,0,1) \sharp\, c\circ(3,-1,0) \qquad \text{and, also,}$$
$$\overrightarrow{x} = d\circ(2,-1,3) \sharp\, e\circ(1,-2,3) \sharp\, f\circ(3,-1,4).$$

It follows that $\overrightarrow{x} \in L_1 \cap L_2$ is true exactly if there are two triples of coefficients, some a, b, c and d, e, f, such that

$$a\circ(1,-1,2) \sharp\, b\circ(1,0,1) \sharp\, c\circ(3,-1,0) = d\circ(2,-1,3) \sharp\, e\circ(1,-2,3) \sharp\, f\circ(3,-1,4).$$

After a little manipulation, we obtain

$$\left.\begin{array}{rcrcrcrcrcrcrcl} a & + & b & + & 3c & - & 2d & - & e & - & 3f & = & 0 \\ -a & & & - & c & + & d & + & 2e & + & f & = & 0 \\ 2a & + & b & & & - & 3d & - & 3e & - & 4f & = & 0 \end{array}\right\} \Longrightarrow \left.\begin{array}{r} a+b+3c-2d-e-3f=0 \\ b+2c-d+e-2f=0 \\ -b-6c+d-e+2f=0 \end{array}\right\} \Longrightarrow$$

$$\Longrightarrow \begin{array}{rcrcrcrcrcrcl} a & + & b & + & 3c & - & 2d & - & e & - & 3f & = & 0 \\ & & b & + & 2c & - & d & + & e & - & 2f & = & 0 \\ & & & - & 4c & & & & & & & = & 0. \end{array}$$

It follows that $c=0$ but the most important thing is that for every triple d,e,f there is a triple a,b,c such that the collection a,b,c,d,e,f solves the last system of equations. It follows that $L_1 \cap L_2 = L_2$. Let us find a basis of L_2. This is easy:

$$\begin{array}{c} (2,-1,3) \\ (1,-2,3) \\ (3,-1,4) \end{array} \sim \begin{array}{c} (2,-1,3) \\ (0,3,-3) \\ (0,-1,1) \end{array} \sim \begin{array}{c} (2,-1,3) \\ (0,1,-1). \end{array}$$

Thus, the set $B=\{(2,-1,3),(0,1,-1)\}$ is a basis of $L_1 \cap L_2$ $(=L_2)$ and $\dim L_1 \cap L_2 = 2$. (Observe that $L_1 = R^3$.)

1.1.51. Example. : Put $\mathcal{P}^3 = \{p \in P \mid \deg p \leq 3\}$ (i. e., $\mathcal{P}^3$ is the linear space of all polynomials whose degrees are not greater than 3). Let $Q = \{p \in \mathcal{P}^3 \mid p$ is divisible by the polynomial $x-2\}$. Check that Q is a linear subspace of $\mathcal{P}^3$ (and, therefore, Q is a linear subspace of $\mathcal{P}$). Find a basis of Q and determine $\dim Q$.

Solution: Let us first proceed directly. The subset Q is a linear subspace of P^3 since if two polynomials p_1, p_2 are divisible by $x-2$, then the polynomial $p_1 \sharp p_2$ as well as the polynomial $\lambda \circ p_1$ is also divisible by $x-2$. Indeed, the divisibility by a given polynomial means that the remainder after division is zero. This property is obviously stable under addition and scalar multiplication of polynomials.

Suppose that $p = ax^3+bx^2+cx+d$ is a polynomial. Let us divide it by $x-2$. We obtain

$$(ax^3+bx^2+cx+d):(x-2) = ax^2+(b+2a)x+(4a+2b+c)+\frac{8a+4b+2c+d}{x-2}$$

$$\begin{array}{ll} -(ax^3-2ax^2) & \\ \hline (2a+b)x^2+cx & \\ -((2a+b)x^2-2(2a+b)x) & \\ \hline (4a+2b+c)x & +d \\ -((4a+2b+c)x & -2(4a+2b+c)) \\ \hline & (8a+4b+2c+d) \end{array}$$

We see that a polynomial $p = ax^3 + bx^2 + cx + d$ is divisible by $x-2$ if and only if $8a+4b+2c+d=0$. This implies that a polynomial, p, belongs to Q if and only if there are coefficients a,b,c such that $p = ax^3+bx^2+cx-8a-4b-2c$. Thus, $p = a\circ(x^3-8) \sharp b\circ(x^2-4) \sharp c\circ(x-2)$. We conclude that the set $B = \{x^3-8, x^2-4, x-2\}$ is a basis of Q. It follows that $\dim Q = 3$.

We can also try and use a think-first-then-do approach. Since $\dim \mathcal{P}^3 = 4$ and since Q is a strictly smaller subspace of $\mathcal{P}^3$, we see that $\dim Q \leq 3$ (Th. 1.1.42). But the polynomials $x-2, x(x-2)$ and $x^2(x-2)$ certainly belong to Q. Since they are obviously LI, we see that they must constitute a basis of Q. Thus, $\dim Q = 3$.

The following result will be used later. The result shows that the linear independence of general vectors can sometimes be translated into coordinate n-tuples.

1.1.52. Theorem. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be LI vectors in L. Let us form new vectors $\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m$ as linear combinations of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$:

$$\begin{aligned}
\vec{z}_1 &= a_{11}\circ\vec{v}_1 \sharp a_{12}\circ\vec{v}_2 \sharp \dots \sharp a_{1n}\circ\vec{v}_n,\\
\vec{z}_2 &= a_{21}\circ\vec{v}_1 \sharp a_{22}\circ\vec{v}_2 \sharp \dots \sharp a_{2n}\circ\vec{v}_n,\\
&\vdots\\
\vec{z}_m &= a_{m1}\circ\vec{v}_1 \sharp a_{m2}\circ\vec{v}_2 \sharp \dots \sharp a_{mn}\circ\vec{v}_n.
\end{aligned}$$

Write $\vec{a_1} = (a_{11}, a_{12}, \dots, a_{1n}), \vec{a_2} = (a_{21}, a_{22}, \dots, a_{2n}), \dots, \vec{a_m} = (a_{m1}, a_{m2}, \dots, a_{mn})$. Thus, the vectors $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ belong to R^n. The following result is then true: The vectors $\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m$ are LI (in L) if and only if the vectors $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ are LI (in R^n).

A corollary:

a) $\dim \operatorname{Span}\{\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m\} = \dim \operatorname{Span}\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$.

b) If $m = n$ and $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$ is a basis of L, then $\{\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m\}$ is a basis of L if and only if $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$ is a basis of R^n.

PROOF: We will first prove that if the vectors $\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m$ are LI in L. then so are the vectors $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$ in R^n.

Suppose that $\vec{z}_1, \vec{z}_2, \dots, \vec{z}_m$ are LI. Hoping to reach a contradiction, suppose that $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$ are LD. This means that there is a nontrivial linear combination of the vectors $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$ which equals $\vec{o}$. In other words, $\lambda_1\circ\vec{a}_1 \sharp \lambda_2\circ\vec{a}_2 \sharp \dots \sharp \lambda_m\circ\vec{a}_m = \vec{o}$ in R^n for some nontrivial coefficients $\lambda_1, \lambda_2, \dots, \lambda_m$ (with a certain abuse of notation, we have used the symbols $\sharp, \circ$ for the operations in R^n). The last equation gives us the following equalities:

$$\begin{aligned}
\lambda_1 a_{11} + \lambda_2 a_{21} + \dots + \lambda_m a_{m1} &= 0\\
\lambda_1 a_{12} + \lambda_2 a_{22} + \dots + \lambda_m a_{m2} &= 0\\
&\vdots\\
\lambda_1 a_{1n} + \lambda_2 a_{2n} + \dots + \lambda_m a_{mn} &= 0
\end{aligned}$$

We will show that $\lambda_1\circ\vec{z}_1 \sharp \lambda_2\circ\vec{z}_2 \sharp \dots \sharp \lambda_m\circ\vec{z}_m = \vec{o}$ in L. Indeed, $\lambda_1\circ\vec{z}_1 \sharp \lambda_2\circ\vec{z}_2 \sharp \dots \sharp \lambda_m\circ\vec{z}_m = \lambda_1\circ(a_{11}\circ\vec{v}_1 \sharp a_{12}\circ\vec{v}_2 \sharp \dots \sharp a_{1n}\circ\vec{v}_n) \sharp \lambda_2\circ(a_{21}\circ\vec{v}_1 \sharp a_{22}\circ$

$\overrightarrow{v}_2 \sharp \ldots \sharp a_{2n} \circ \overrightarrow{v}_n) \sharp \ldots \sharp \lambda_m \circ (a_{m1} \circ \overrightarrow{v}_1 \sharp a_{m2} \circ \overrightarrow{v}_2 \sharp \ldots \sharp a_{mn} \circ \overrightarrow{v}_n) = (\lambda_1 a_{11} + \lambda_2 a_{21} + \cdots + \lambda_m a_{m1}) \circ \overrightarrow{v}_1 \sharp (\lambda_1 a_{12} + \lambda_2 a_{22} + \cdots + \lambda_m a_{m2}) \circ \overrightarrow{v}_2 \sharp \ldots \sharp (\lambda_1 a_{1n} + \lambda_2 a_{2n} + \cdots + \lambda_m a_{mn}) \circ \overrightarrow{v}_n = 0 \circ \overrightarrow{v}_1 \sharp 0 \circ \overrightarrow{v}_2 \sharp \ldots \sharp 0 \circ \overrightarrow{v}_n = \overrightarrow{o}$. This is a contradiction. (Observe that for this implication we did not need the assumption on $\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots \overrightarrow{v}_n$ being LI.)

What remains to be proved is the reverse implication: $\{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots, \overrightarrow{a}_m\}$ LI $\Rightarrow \{\overrightarrow{z}_1, \overrightarrow{z}_2, \ldots, \overrightarrow{z}_m\}$ LI. Suppose that $\{\overrightarrow{z}_1, \overrightarrow{z}_2, \ldots, \overrightarrow{z}_m\}$ are LD. Then $\lambda_1 \circ \overrightarrow{z}_1 \sharp \lambda_2 \circ \overrightarrow{z}_2 \sharp \ldots \sharp \lambda_m \circ \overrightarrow{z}_m = \overrightarrow{o}$ (in L) for a nontrivial linear combination. Substituting for $\{\overrightarrow{z}_1, \overrightarrow{z}_2, \ldots, \overrightarrow{z}_m\}$, we get $\lambda_1 \circ (a_{11} \circ \overrightarrow{v}_1 \sharp a_{12} \circ \overrightarrow{v}_2 \sharp \ldots \sharp a_{1n} \circ \overrightarrow{v}_n) \sharp \lambda_2 \circ (a_{21} \circ \overrightarrow{v}_1 \sharp a_{22} \circ \overrightarrow{v}_2 \sharp \ldots \sharp a_{2n} \circ \overrightarrow{v}_n) \sharp \ldots \sharp \lambda_m \circ (a_{m1} \circ \overrightarrow{v}_1 \sharp a_{m2} \circ \overrightarrow{v}_2 \sharp \ldots \sharp a_{mn} \circ \overrightarrow{v}_n) = \overrightarrow{o}$. It follows that $(\lambda_1 a_{11} + \lambda_2 a_{21} + \cdots + \lambda_m a_{m1}) \circ \overrightarrow{v}_1 \sharp (\lambda_1 a_{12} + \lambda_2 a_{22} + \cdots + \lambda_m a_{m2}) \circ \overrightarrow{v}_2 \sharp \ldots \sharp (\lambda_1 a_{1n} + \lambda_2 a_{2n} + \cdots + \lambda_m a_{mn}) \circ \overrightarrow{v}_n = \overrightarrow{o}$. Since the vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots, \overrightarrow{v}_n$ are LI, we obtain

$$\begin{aligned}
\lambda_1 a_{11} + \lambda_2 a_{21} + \cdots + \lambda_m a_{m1} &= 0 \\
\lambda_1 a_{12} + \lambda_2 a_{22} + \cdots + \lambda_m a_{m2} &= 0 \\
&\vdots \\
\lambda_1 a_{1n} + \lambda_2 a_{2n} + \cdots + \lambda_m a_{mn} &= 0
\end{aligned}$$

But the latter equations are equivalent to the equation $\lambda_1 \circ \overrightarrow{a}_1 \sharp \lambda_2 \circ \overrightarrow{a}_2 \sharp \ldots \sharp \lambda_m \circ \overrightarrow{a}_m = \overrightarrow{o}$ (in R^n). We have proved the second implication and the proof is complete.

As for the corollary, part a) follows easily since $\dim \mathrm{Span}\{\overrightarrow{z}_1, \overrightarrow{z}_2, \ldots \overrightarrow{z}_m\}$ (resp. $\dim \mathrm{Span}\{\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots \overrightarrow{a}_m\}$) is exactly the number of vectors in a maximal LI subset of $\overrightarrow{z}_1, \overrightarrow{z}_2, \ldots, \overrightarrow{z}_m$ (resp. $\overrightarrow{a}_1, \overrightarrow{a}_2, \ldots \overrightarrow{a}_m$). Part b) is a direct consequence of part a).

1.1.53. Example. Prove that if the vectors $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}$ are LI, then so are the vectors $\overrightarrow{a} \sharp \overrightarrow{b} - \overrightarrow{c}$, $\overrightarrow{a} - \overrightarrow{b} \sharp \overrightarrow{c}$, $-\overrightarrow{a} \sharp \overrightarrow{b} \sharp \overrightarrow{c}$.

Solution: Consider the "coordinate" vectors $(1, 1, -1), (1, -1, 1), (-1, 1, 1)$ in R^3. According to Prop. 1.1.52, all we have to verify is the linear independence of these vectors. This can be done easily. Indeed,

$$\begin{matrix} (1,1,-1) \\ (1,-1,1) \\ (-1,1,1) \end{matrix} \sim \begin{matrix} (1, & 1, -1) \\ (0, -2, & 2) \\ (0, & 2, & 0) \end{matrix} \sim \begin{matrix} (1, & 1, -1) \\ (0, -1, & 1) \\ (0, & 0, & 1). \end{matrix}$$

Problems

(A problem indicated with a $*$ is supposed to be a challenge for an interested student.)

1. Show that if we replace Axioms 3 and 4 in the definition of linear space by the axiom "for any $\overrightarrow{x} \in L$, $0 \circ \overrightarrow{x} = \overrightarrow{o}$", we obtain the same theory.
2. Show that the following properties are fulfilled in any linear space L:

(i) $-(-\overrightarrow{x}) = \overrightarrow{x}$,
(ii) if $\lambda \circ \overrightarrow{u} = \lambda \circ \overrightarrow{v}$, then either $\lambda = 0$ or $\overrightarrow{u} = \overrightarrow{v}$,
(iii) if $\lambda_1 \circ \overrightarrow{u} = \lambda_2 \circ \overrightarrow{u}$, then either $\lambda_1 = \lambda_2$ or $\overrightarrow{u} = \overrightarrow{o}$.

3. Check that the set C of all complex numbers can be understood as a linear space (we may take for the operations $\sharp, \circ$ the usual operations $+, \cdot$ in complex numbers). Verify all Axioms 1–8.

4. Let $L = \{3\}$. Can we define operations $\sharp, \circ$ on L so that L becomes a linear space?
[Yes]

5. * Let S denote the set of all sequences of real numbers. Let, for any $\overrightarrow{a} = (a_1, a_2, \dots, a_n, \dots)$, $\overrightarrow{b} = (b_1, b_2, \dots, b_n, \dots)$,

$$\overrightarrow{a} \sharp \overrightarrow{b} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n, \dots)$$
$$\lambda \circ \overrightarrow{a} = (\lambda \cdot a_1, \lambda \cdot a_2, \dots, \lambda \cdot a_n, \dots).$$

(i) Show that with these "coordinate-wise" operations the set S is converted to a linear space.
(ii) Show that the subset S_{conv} of S which consists of all convergent sequences is a linear subspace of S .
(iii) Let $S_{\text{fin}} = \{(a_1, a_2, \dots, a_n, \dots) \in S_{\text{conv}} | a_n \neq 0$ for at most finitely many $n \in N\}$. Show that S_{fin} is a linear subspace of S_{conv} . Does S_{fin} have a finite basis?
[No]

6. * Show that the set $\mathscr{R}$ of all rational functions forms a linear space when $\mathscr{R}$ is naturally endowed with the operations $\sharp$ (the ordinary addition) and $\circ$ (the ordinary multiplication). Show that the rational functions of the type $\dfrac{1}{x+a}, \dfrac{1}{x^2+px+q}, \dfrac{x}{x^2+rx+s}$, where a, p, q, r, s are real numbers and $p^2 - 4q < 0$, $r^2 - 4s < 0$, constitute a basis of $\mathscr{R}$ (see Appendix 2).

7. Find out whether the following vectors $\overrightarrow{x}_1, \overrightarrow{x}_2, \overrightarrow{x}_3$ are LI or LD in R^3 :

$$\overrightarrow{x}_1 = (3, 1, 5), \quad \overrightarrow{x}_2 = (1, 2, -2), \quad \overrightarrow{x}_3 = (2, 3, -1)$$

[LI]

8. Consider the vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ defined as follows:

$$\overrightarrow{v}_1 = (1, \alpha, 1), \quad \overrightarrow{v}_2 = (0, 1, \alpha), \quad \overrightarrow{v}_3 = (\alpha, 1, 0),$$

where α is a real parameter. Find all $\alpha \in R$ for which the vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ are LD.
$[0, \sqrt{2}, -\sqrt{2}]$

9. (i) Find out whether the vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are LI or LD in $\mathscr{M}$:

$$\mathbf{A} = \begin{Vmatrix} 1, & 10 \\ 6, & 3 \end{Vmatrix}, \quad \mathbf{B} = \begin{Vmatrix} 2, & 1 \\ 1, & -1 \end{Vmatrix}, \quad \mathbf{C} = \begin{Vmatrix} 3, & 7 \\ 3, & 2 \end{Vmatrix}, \quad \mathbf{D} = \begin{Vmatrix} 1, & -3 \\ 1, & -4 \end{Vmatrix}.$$

[LD]

(ii) Consider the vectors $\mathbf{E}, \mathbf{F}, \mathbf{G}$ defined as follows:

$$\mathbf{E} = \left\| \begin{matrix} 1, & 3 \\ -4, & 0 \end{matrix} \right\|, \qquad \mathbf{F} = \left\| \begin{matrix} 1, & 2 \\ 3, & 0 \end{matrix} \right\|, \qquad \mathbf{G} = \left\| \begin{matrix} 2, & -1 \\ \alpha, & 0 \end{matrix} \right\|,$$

where α is a real parameter. Find all $\alpha \in R$ for which the vectors $\mathbf{E}, \mathbf{F}, \mathbf{G}$ are LD in $\mathscr{M}$.

[$\alpha = 41$]

10. (i) Show that the following vectors f_1, f_2, f_3, f_4 are LD in the space $\mathcal{F}$ of all functions:

$$f_1 = 2t^3 + 6, \qquad f_2 = t^3 + t^2 + t + 1, \qquad f_3 = 2t^2, \qquad f_4 = 2t^3 + 2t^2 - t + 8.$$

(ii) Show that $g \in \mathrm{Span}\{g_1, g_2, g_3\}$, where

$$g = 5t - 7t^2$$
$$g_1 = 1 + t - 2t^2, \qquad g_2 = 7 - 8t + 7t^2, \qquad g_3 = 3 - 2t + t^2.$$

11. Let $\overrightarrow{x}_1, \overrightarrow{x}_2, \overrightarrow{x}_3, \overrightarrow{x}$ be the following vectors in R^3:

$$\overrightarrow{x}_1 = (-1, 2, 1), \quad \overrightarrow{x}_2 = (1, 1, 2), \quad \overrightarrow{x}_3 = (-5, 8, 3), \quad \overrightarrow{x} = (0, 1, \alpha),$$

where α is a real parameter. Find an $\alpha \in R$ such that $\overrightarrow{x} \in \mathrm{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \overrightarrow{x}_3\}$.

[$\alpha = 1$]

12. Show that $\mathrm{Span}\{\overrightarrow{a}, \overrightarrow{b}\} = \mathrm{Span}\{\overrightarrow{c}, \overrightarrow{d}\}$, where $\overrightarrow{a} = (1, 1, -1)$, $\overrightarrow{b} = (2, 0, 1)$, $\overrightarrow{c} = (3, 1, 0)$, $\overrightarrow{d} = (1, -1, 2)$.

13. * Let L be a linear space.

(i) Prove that if $\{\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z}\} \subset L$ such that $\overrightarrow{x} \sharp \overrightarrow{y} \sharp \overrightarrow{z} = \overrightarrow{o}$, then $\mathrm{Span}\{\overrightarrow{x}, \overrightarrow{y}\} = \mathrm{Span}\{\overrightarrow{y}, \overrightarrow{z}\}$.

(ii) Prove that if $\{\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k\} \subset L$, $\overrightarrow{y} \in \mathrm{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k, \overrightarrow{x}\}$ and $\overrightarrow{y} \notin \mathrm{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k\}$, then $\overrightarrow{x} \in \mathrm{Span}\{\overrightarrow{x}_1, \overrightarrow{x}_2, \dots, \overrightarrow{x}_k, \overrightarrow{y}\}$.

14. Show that the following vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ constitute a basis of R^3. Find the coordinates of the vector $\overrightarrow{v}$ in this basis.

$$\overrightarrow{v}_1 = (3, 1, -3), \quad \overrightarrow{v}_2 = (-5, 2, -1), \quad \overrightarrow{v}_3 = (7, 3, 4); \qquad \overrightarrow{v} = (11, -8, -4)$$

[Coordinates: $(1, -3, -1)$]

15. Show that $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ form a basis of $\mathscr{M}$.

$$\mathbf{A} = \left\| \begin{matrix} 3, & -1 \\ 0, & 0 \end{matrix} \right\|, \quad \mathbf{B} = \left\| \begin{matrix} 1, & 1 \\ 0, & 0 \end{matrix} \right\|, \quad \mathbf{C} = \left\| \begin{matrix} 3, & 0 \\ 5, & -3 \end{matrix} \right\|, \quad \mathbf{D} = \left\| \begin{matrix} 4, & -1 \\ 3, & -1 \end{matrix} \right\|$$

16. * Exhibit two bases $\mathcal{B}_1 = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \overrightarrow{b}_3, \overrightarrow{b}_4\}$, $\mathcal{B}_2 = \{\overrightarrow{c}_1, \overrightarrow{c}_2, \overrightarrow{c}_3, \overrightarrow{c}_4\}$ of R^4 such that $\overrightarrow{b}_1 = (1, 0, 0, 0)$, $\overrightarrow{b}_2 = (1, 1, 0, 0)$, $\overrightarrow{c}_3 = (1, 1, 1, 0)$, $\overrightarrow{c}_4 = (1, 1, 1, 1)$ and $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$.

17. Let L_1, L_2 be linear subspaces of L.

(i) Show that $L_1 \cap L_2$ is a linear subspace of L.

(ii) Show by example that $L_1 \cup L_2$ does *not* have to be a linear subspace of L. (Take, for instance, $L_1 = \{(0,a)|a \in R\}$ and $L_2 = \{(b,0)|b \in R\}$ as subspaces of $L = R^2$.)

18. In R^4 find a basis of $L_1 \cap L_2$, where $L_1 = \text{Span}\{\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}\}$ and $L_2 = \text{Span}\{\overrightarrow{d}, \overrightarrow{e}, \overrightarrow{f}\}$:

$$\begin{aligned} \overrightarrow{a} &= (1,1,0,2) & \overrightarrow{d} &= (2,-1,3,1) \\ \overrightarrow{b} &= (-1,1,2,3) & \overrightarrow{e} &= (0,1,1,2) \\ \overrightarrow{c} &= (-2,0,2,1) & \overrightarrow{f} &= (2,2,2,1) \end{aligned}$$

[$L_1 \cap L_2 = \text{Span}\{(-1,5,6,13)\}$, thus $\{(-1,5,6,13)\}$ is a basis of $L_1 \cap L_2$]

19. In R^4 find a basis of $L_1 \cap L_2$, where $L_1 = \text{Span}\{\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}\}$ and $L_2 = \text{Span}\{\overrightarrow{d}, \overrightarrow{e}, \overrightarrow{f}\}$:

$$\begin{aligned} \overrightarrow{a} &= (1,2,0,3) & \overrightarrow{d} &= (2,0,-2,1) \\ \overrightarrow{b} &= (1,1,-1,2) & \overrightarrow{e} &= (3,1,0,2) \\ \overrightarrow{c} &= (0,1,1,1) & \overrightarrow{f} &= (4,2,2,3) \end{aligned}$$

[$L_1 \cap L_2 = \{(0,0,0,0)\}$ - it does not have a basis]

20. Show that $B = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ constitutes a basis of $\mathscr{M}$. Find the coordinates of $\mathbf{E}$ with respect to B.

$$\mathbf{A} = \begin{Vmatrix} 3, & 0 \\ -1, & 0 \end{Vmatrix}, \quad \mathbf{B} = \begin{Vmatrix} 1, & 0 \\ 1, & 0 \end{Vmatrix}, \quad \mathbf{C} = \begin{Vmatrix} 3, & 5 \\ 0, & -3 \end{Vmatrix}, \quad \mathbf{D} = \begin{Vmatrix} 4, & 3 \\ -1, & -1 \end{Vmatrix} \quad \mathbf{E} = \begin{Vmatrix} 3, & 2 \\ 1, & -2 \end{Vmatrix}$$

[$\mathbf{E}_B = (1,1,1,-1)$]

21. * Call a matrix $\mathbf{A} \in \mathscr{M}$ *magical* if it enjoys the following property: The sum along any row of $\mathbf{A}$ equals the sum down any column, and this equals the sum over the diagonal and the antidiagonal (the antidiagonal consists of the elements $a_{i,4-i},\ i = 1,2,3$). Verify that the set $\mathscr{M}_{\text{mag}}$ of all magical matrices forms a linear subspace of $\mathscr{M}$. Find a basis of $\mathscr{M}_{\text{mag}}$ and determine $\dim \mathscr{M}_{\text{mag}}$.

22. Let $\mathcal{P}^4 = \{p \in \mathcal{P} | \deg p \leq 4\}$ and set $\mathcal{R} = \{p \in \mathcal{P}^4 | p \text{ is divisible by } x^2 + x + 1\}$. Verify that $\mathcal{R}$ is a linear subspace of $\mathcal{P}^4$. Find a basis of $\mathcal{R}$ and determine $\dim \mathcal{R}$.

23. Let $\mathcal{P}^4 = \{p \in \mathcal{P} | \deg p \leq 4\}$. Let $\mathcal{Q} = \{p \in \mathcal{P}^4 | p(x) = x^4 \cdot p(\frac{1}{x}) \text{ for any } x \in R,\ x \neq 0\}$. Show that $\mathcal{Q}$ is a linear subspace of $\mathcal{P}^4$. Find a basis of $\mathcal{Q}$.

[Answer: A basis of $\mathcal{Q}$ is e.g. $\{x^4 + 1, x^3 + x, x^2\}$, $\dim \mathcal{Q} = 3$.]

24. Show that the vectors $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}, \overrightarrow{d}$ constitute a basis of R^4 if and only if the vectors $\overrightarrow{a}, \overrightarrow{a} \sharp \overrightarrow{b}, \overrightarrow{a} \sharp \overrightarrow{b} \sharp \overrightarrow{c}, \overrightarrow{a} \sharp \overrightarrow{b} \sharp \overrightarrow{c} \sharp \overrightarrow{d}$ constitute a basis of R^4.

Chapter 2

Matrices, determinants, systems of linear equations

This chapter could be subtitled "linear algebra for a practical man". Indeed, matrices, determinants and systems of linear equations arise in most real-life problems. As we know, a complete treatment of practical problems is usually based on a solid piece of theory. This chapter is no exception – we will find many situations in which the linear algebra theory developed in the first chapter proves its practical value.

2.1. The multiplication of matrices

Let us first formulate an alternative definition of a matrix. This definition is sometimes more appropriate for formal considerations.

2.1.1. Definition. A mapping $\mathbf{A}\colon \{1,2,3,\dots,m\} \times \{1,2,3,\dots,n\} \to R$ is called a (real) matrix of size $m \times n$. Alternately, $\mathbf{A}$ is called an $m \times n$ matrix (an "m by n matrix").

It is obvious that this definition agrees with the "naive" definition formulated before. Indeed, if we write $\mathbf{A} = \|a_{ij}\|$, we can understand matrix $\mathbf{A}$ as the mapping $(i,j) \to a_{ij}$.

Let us make the following convention:

By *the i-th row* of $\mathbf{A}$ we mean the set of all components a_{ij}, where $j \in \{1,2,\dots,n\}$. Analogously, by *the j-th column* of $\mathbf{A}$ we mean the set of all components a_{ij}, where $i \in \{1,2,\dots,m\}$.

2.1.2. Definition. Let $\mathbf{A} = \|a_{ij}\|$ be a matrix of size $m \times n$. The matrix $\mathbf{A}^T = \|a_{ji}\|$ is called *the transpose of* $\mathbf{A}$. Thus, $\mathbf{A}^T$ is an $n \times m$ matrix.

A square matrix is a matrix of size $n \times n$ for some $n \in N$. A square matrix $\mathbf{A}$ is called *symmetric* if $\mathbf{A} = \mathbf{A}^T$.

2.1.3. Example. If $\mathbf{A} = \left\| \begin{matrix} 1, & 2, & 0 \\ -1, & 3 & 2 \end{matrix} \right\|$, then $\mathbf{A}^T = \left\| \begin{matrix} 1, & -1 \\ 2, & 3 \\ 0, & 2 \end{matrix} \right\|$. A matrix $\mathbf{B} = \left\| \begin{matrix} 1, & 0, & -2 \\ 0, & 2, & 7 \\ \alpha, & \beta, & 5 \end{matrix} \right\|$ is symmetric exactly when $\alpha = -2$ and $\beta = 7$.

Recall that the set $\mathscr{M}$ of all matrices of size $m \times n$ forms a linear space – the operations $\sharp$ and $\circ$ are defined "component-wise". Thus, we can "add" matrices and "multiply" matrices with a scalar. Sometimes we relax the correct notation $\mathbf{A}\sharp\mathbf{B}$ and $\lambda \circ \mathbf{A}$ to the less formal $\mathbf{A} + \mathbf{B}$ and $\lambda\mathbf{A}$ when there is no danger of misunderstanding.

The following definition introduces another operation on matrices that is very important in applications. The definition may not be very obvious at first sight – a typical symptom of something that aspires to being involved in deeper results.

2.1.4. Definition. Let $\mathbf{A} = \|a_{ij}\| \in \mathscr{M}$ and $\mathbf{B} = \|b_{ij}\| \in \mathscr{M}$ be matrices. Then the matrix $\mathbf{C} = \|c_{ij}\| \in \mathscr{M}$, where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik} \cdot b_{kj},$$

is called *the product of* $\mathbf{A}$ *and* $\mathbf{B}$. We write $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ or simply $\mathbf{C} = \mathbf{AB}$.

2.1.5. Examples.

a) Let $\mathbf{A} = \begin{Vmatrix} 1, & 2 \\ -3, & 1 \end{Vmatrix}$ and $\mathbf{B} = \begin{Vmatrix} -1, & 0, & 3 \\ 2, & 6, & 9 \end{Vmatrix}$. Compute $\mathbf{AB}$.

Solution:

$$\begin{aligned} \mathbf{AB} &= \begin{Vmatrix} 1, & 2 \\ -3, & 1 \end{Vmatrix} \begin{Vmatrix} -1, & 0, & 3 \\ 2, & 6, & 9 \end{Vmatrix} \\ &= \begin{Vmatrix} 1 \cdot (-1) + 2 \cdot 2, & 1 \cdot 0 + 2 \cdot 6, & 1 \cdot 3 + 2 \cdot 9 \\ (-3) \cdot (-1) + 1 \cdot 2, & (-3) \cdot 0 + 1 \cdot 6, & (-3) \cdot 3 + 1 \cdot 9 \end{Vmatrix} = \begin{Vmatrix} 3, & 12, & 21 \\ 5, & 6, & 0 \end{Vmatrix} \end{aligned}$$

b) If $\mathbf{A} = \begin{Vmatrix} 2, & 1 \\ -3, & 4 \end{Vmatrix}$ and $\mathbf{B} = \begin{Vmatrix} 2 \\ 5 \end{Vmatrix}$, then $\mathbf{AB} = \begin{Vmatrix} 9 \\ 14 \end{Vmatrix}$.

Let us explicitly note the following consequences of the definition of the matrix product.

(i) The matrix product $\mathbf{AB}$ has for the (i,j)-th component "the product of the i-th row of $\mathbf{A}$ with the j-th column of $\mathbf{B}$".

(ii) In order for the product $\mathbf{AB}$ to be meaningful, the number of columns of $\mathbf{A}$ has to coincide with the number of rows of $\mathbf{B}$ (otherwise the product is *not* defined!). For instance, the product $\begin{Vmatrix} 1, & 2 \\ 5, & 3 \end{Vmatrix} \cdot \begin{Vmatrix} 1, & 2 \\ 2, & -1 \\ 3, & 5 \end{Vmatrix}$ is not defined.

(iii) Matrix multiplication is generally *not commutative.* For instance, if $\mathbf{A} = \begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix}$ and $\mathbf{B} = \begin{Vmatrix} 1, & 1 \\ 0, & 0 \end{Vmatrix}$, then

$$\mathbf{AB} = \begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix} \begin{Vmatrix} 1, & 1 \\ 0, & 0 \end{Vmatrix} = \begin{Vmatrix} 1, & 1 \\ 1, & 1 \end{Vmatrix}, \quad \text{whereas} \quad \mathbf{BA} = \begin{Vmatrix} 1, & 1 \\ 0, & 0 \end{Vmatrix} \begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix} = \begin{Vmatrix} 2, & 1 \\ 0, & 0 \end{Vmatrix}.$$

The following proposition reveals basic properties of matrix multiplication.

2.1.6. Proposition. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be matrices and let $\lambda \in R$ be a scalar. If the following matrix products exist, then

(i) $\lambda \circ (\mathbf{AB}) = (\lambda \circ \mathbf{A})\mathbf{B} = \mathbf{A}(\lambda \circ \mathbf{B})$,

(ii) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (the associativity of matrix multiplication),

(iii) $(\mathbf{A} \sharp \mathbf{B})\mathbf{C} = \mathbf{AC} \sharp \mathbf{BC}$ (the "left" distributivity),

(iv) $\mathbf{C}(\mathbf{A} \sharp \mathbf{B}) = \mathbf{CA} \sharp \mathbf{CB}$ (the "right" distributivity),

(v) $\mathbf{EA} = \mathbf{AE} = \mathbf{A}$, when $\mathbf{E}$ is the unit matrix,

$$\mathbf{E} = \begin{Vmatrix} 1, & 0, & 0, & \dots, & 0 \\ 0, & 1, & 0, & \dots, & 0 \\ \vdots & & & & \\ 0, & 0, & \dots, & 1, & 0 \\ 0, & 0, & \dots, & 0, & 1 \end{Vmatrix},$$

(vi) $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$.

PROOF: Let us only check property (vi), the other properties can easily be verified from the definition of matrix multiplication. Let $\mathbf{A} = \|a_{ij}\| \in \mathscr{M}$ and $\mathbf{B} = \|b_{ij}\| \in \mathscr{M}$. Then $\mathbf{A}^T = \|\alpha_{ij}\| \in \mathscr{M}$, where $\alpha_{ij} = a_{ji}$, and $\mathbf{B}^T = \|\beta_{ij}\| \in \mathscr{M}$, where $\beta_{ij} = b_{ji}$. We have $\mathbf{B}^T\mathbf{A}^T = \|\gamma_{ij}\|$, where $\gamma_{ij} = \sum_{k=1}^{p} \beta_{ik}\alpha_{kj} = \sum_{k=1}^{p} a_{jk}b_{ki}$. On the other hand, if $\mathbf{AB} = \mathbf{C}$ and $\mathbf{C} = \|c_{ij}\| \in \mathscr{M}$, then $c_{ij} = \sum_{k=1}^{p} a_{ik} \cdot b_{kj} = \gamma_{ji}$. This proves that $\mathbf{B}^T\mathbf{A}^T = (\mathbf{AB})^T$. □

Obviously, a matrix $\mathbf{A} \in \mathscr{M}$ can also be viewed as a collection of m vectors of R^n (the rows of $\mathbf{A}$) or, as a collection of n vectors of R^m (the columns of $\mathbf{A}$). Thus, individual properties of a matrix can sometimes be treated with the help of the linear space theory of R^n or R^m. This will be the case in the considerations which follow.

The rank of a matrix

2.1.7. Definition. Let us consider a matrix $\mathbf{A} \in \mathscr{M}$. Let us view the rows of $\mathbf{A}$ as vectors of $\mathbf{R}^n$. Let us denote by $\mathscr{R}$ the linear span of these vectors in $\mathbf{R}^n$. Then $\dim \mathscr{R}$ is called *the rank of* $\mathbf{A}$. (We denote the rank of $\mathbf{A}$ by $r(\mathbf{A})$.)

Let us remark that an alternative definition of $r(\mathbf{A})$ would be the number of a maximal linearly independent family of rows of $\mathbf{A}$ (see Prop. 1.1.40). Thus, $r(\mathbf{A}) \leq n$.

2.1.8. Examples.

a) If $\mathbf{A} = \begin{Vmatrix} 1, & 1 \\ 1, & -1 \end{Vmatrix}$, then $r(\mathbf{A}) = 2$

b) If $\mathbf{A} = \begin{Vmatrix} 1 \\ 3 \\ 5 \end{Vmatrix}$, then $r(\mathbf{A}) = 1$.

c) If $\mathbf{A} = \begin{Vmatrix} 1, & 2, & 3 \\ 2, & 4, & 6 \\ 0, & 0, & 0 \end{Vmatrix}$, then $r(\mathbf{A}) = 1$.

d) If $\mathbf{A} = \begin{Vmatrix} 1, & 0, & -1, & 1 \\ 2, & 1, & 1, & 0 \\ 4, & 1, & -1, & 2 \end{Vmatrix}$, then $r(A) = 2$ (it is easily seen that the third row is a linear combination of the other two rows).

Obviously, the definition of $r(\mathbf{A})$ implicitly contains a kind of discrimination: Why should we prefer rows to columns? Fortunately, nature itself restores the justice here – if we defined the rank of $\mathbf{A}$ "via columns", we would obtain the same number. The following theorem takes care of this remarkable fact.

2.1.9. Theorem Let $\mathbf{A}$ be a matrix. Then

$$r(\mathbf{A}) = r(\mathbf{A}^T).$$

PROOF: We will only prove that $r(\mathbf{A}) \geq r(\mathbf{A}^T)$ for any $\mathbf{A}$. This is easily seen to be equivalent to the required equality $r(\mathbf{A}) = r(\mathbf{A}^T)$. Indeed, if we put $\mathbf{B} = \mathbf{A}^T$, we have $r(\mathbf{B}) \geq r(\mathbf{B}^T)$, and therefore $r(\mathbf{A}^T) \geq r((\mathbf{A}^T)^T) = r(\mathbf{A})$. Thus, we also have $r(\mathbf{A}) \leq r(\mathbf{A}^T)$, leading to $r(\mathbf{A}) = r(\mathbf{A}^T)$.

Suppose that $\mathbf{A} \in \mathscr{M}$. We will divide the proof into two steps. Let us first assume that $r(\mathbf{A}) = m$. This means that all rows of $\mathbf{A}$ are LI. Since the column vectors of $\mathbf{A}$ belong to R^m, we infer (Prop. 1.1.40) that $\mathbf{A}$ cannot have more than m LI columns. It follows that $r(\mathbf{A}^T) \leq m$. Thus, in this case, $r(\mathbf{A}) \geq r(\mathbf{A}^T)$.

Let us now take up the general case. Write

$$\mathbf{A} = \begin{Vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn} \end{Vmatrix}.$$

Suppose that $r(\mathbf{A}) = p,\ p < m$. This means that there are p LI rows in $\mathbf{A}$ such that the other rows can be expressed as linear combinations of these p rows. We may assume that these p rows are the i_1-th, i_2-th, $\ldots, i_p$-th rows. Put

$$\mathbf{A}' = \begin{Vmatrix} a_{i_1 1}, & a_{i_1 2}, & \ldots, & a_{i_1 n} \\ a_{i_2 1}, & a_{i_2 2}, & \ldots, & a_{i_2 n} \\ \vdots & & & \\ a_{i_p 1}, & a_{i_p 2}, & \ldots, & a_{i_p n} \end{Vmatrix}$$

Then $r(\mathbf{A}') = p$, and we know from the previous investigation that $r(\mathbf{A}') \geq r(\mathbf{A}'^T)$. All we have to prove is $r(\mathbf{A}'^T) = r(\mathbf{A}^T)$. The proof needs a bit of thinking. We want to show that $r(\mathbf{A}^T)$ is not greater than $r(\mathbf{A}'^T)$, the inequality $r(\mathbf{A}^T) \geq r(\mathbf{A}'^T)$ being trivial. We know that every row $(a_{q1}, a_{q2}, \ldots, a_{qn})$, where $q \notin \{i_1, i_2, \ldots, i_p\}$, is a linear combination of the rows of $\mathbf{A}'$. It follows that if a linear combination of columns of $\mathbf{A}'$ makes the zero vector in

R^p, then so does the same linear combination of the corresponding extended columns of $\mathbf{A}$ in R^m (model it on examples and consult Problem 14 formulated in the problem block after this chapter). This in turn gives $r(\mathbf{A}^T) \leq r(\mathbf{A}'^T)$. The proof is complete. □

Let us summarize what we have shown on the $r(\mathbf{A})$ so far.

2.1.10. Theorem The number of a maximal family of LI rows in $\mathbf{A}$ coincides with the number of a maximal family of LI columns in $\mathbf{A}$. Thus, $r(\mathbf{A})$ can be computed equally well from the "row space" as from the "column space".

Having re-phrased the notion of $r(\mathbf{A})$ into the language of linear space, we are in the position to apply results of the previous chapter.

2.1.11. Proposition. Let $\mathbf{A}$ be a matrix.

(i) If $\mathbf{B}$ has arisen from $\mathbf{A}$ by exchanging two rows (resp. two columns), then $r(\mathbf{B}) = r(\mathbf{A})$.

(ii) If $\mathbf{B}$ has arisen from $\mathbf{A}$ by multiplying a row (resp. a column) of $\mathbf{A}$ by a nonzero number, then $r(\mathbf{B}) = r(\mathbf{A})$.

(iii) If $\mathbf{B}$ has arisen from $\mathbf{A}$ by adding to a given row of $\mathbf{A}$ a linear combination of remaining rows (resp., by adding to a given column of $\mathbf{A}$ a linear combination of remaining columns), then $r(\mathbf{B}) = r(\mathbf{A})$.

(iv) If $\mathbf{B}$ has arisen from $\mathbf{A}$ by deleting null rows (resp., null columns), then $r(\mathbf{B}) = r(\mathbf{A})$.

PROOF: It follows directly from Prop. 1.1.26 and 1.1.40.

In applying the above proposition, we usually proceed recursively, knowing that the adequate "Gaussian elimination" does not change the rank.

2.1.12. Example. Compute the rank of $\mathbf{A}$, where

$$\mathbf{A} = \begin{Vmatrix} 1, & 2, & 0, & -1 \\ 2, & 1, & 3, & 5 \\ -4, & 1, & -9, & -17 \end{Vmatrix}.$$

SOLUTION: Using the symbol $\sim$ to denote the coincidence of ranks, we proceed with elimination to obtain

$$\begin{Vmatrix} 1, & 2, & 0, & -1 \\ 2, & 1, & 3, & 5 \\ -4, & 1, & -9, & -17 \end{Vmatrix} \sim \begin{Vmatrix} 1, & 2, & 0, & -1 \\ 0, & -3, & 3, & 7 \\ 0, & 9, & -9, & -21 \end{Vmatrix} \sim \begin{Vmatrix} 1, & 2, & 0, & -1 \\ 0, & -3, & 3, & 7 \\ 0, & 0, & 0, & 0 \end{Vmatrix}.$$

It follows that $r(\mathbf{A}) = 2$. As an illustration of Th. 2.2.3, compute $r(\mathbf{A}^T)$. We have

$$\mathbf{A}^T = \begin{Vmatrix} 1, & 2, & -4 \\ 2, & 1, & 1 \\ 0, & 3, & -9 \\ -1, & 5, & -17 \end{Vmatrix} \sim \begin{Vmatrix} 1, & 2, & -4 \\ 0, & -3, & 9 \\ 0, & 3, & -9 \\ 0, & 7, & -21 \end{Vmatrix} \sim \begin{Vmatrix} 1, & 2, & -4 \\ 0, & -1, & 3 \\ 0, & 0, & 0 \\ 0, & 0, & 0 \end{Vmatrix} \Rightarrow r(\mathbf{A}^T) = 2.$$

2.1.13. Example. Discuss the rank of matrix $\mathbf{A}$ with respect to the parameter α:

$$\mathbf{A} = \left\| \begin{matrix} -1, & 3, & -6, & 4 \\ 1, & \alpha, & 5, & -3 \\ -1, & -2, & -1, & -1 \\ 3, & 2, & 3, & -1 \end{matrix} \right\| .$$

SOLUTION: We first observe that if we interchange rows or columns in $\mathbf{A}$, the rank does not change (Th. 2.2.3).

Thus by interchanging rows 2 and 4 and columns 2 and 4, we see that

$$\left\| \begin{matrix} -1, & 3, & -6, & 4 \\ 1, & \alpha, & 5, & -3 \\ -1, & -2, & -1, & -1 \\ 3, & 2, & 3, & -1 \end{matrix} \right\| \sim \left\| \begin{matrix} -1, & 3, & -6, & 4 \\ 3, & 2, & 3, & -1 \\ -1, & -2, & -1, & -1 \\ 1, & \alpha, & 5, & -3 \end{matrix} \right\| \sim \left\| \begin{matrix} -1, & 4, & -6, & 3 \\ 3, & -1, & 3, & 2 \\ -1, & -1, & -1, & -2 \\ 1, & -3, & 5, & \alpha \end{matrix} \right\| \sim$$

$$\sim \left\| \begin{matrix} -1, & 4, & -6, & 3 \\ 0, & 11, & -15, & 11 \\ 0, & -5, & 5, & -5 \\ 0, & 1, & -1, & 3+\alpha \end{matrix} \right\| \sim \left\| \begin{matrix} -1, & 4, & -6, & 3 \\ 0, & -1, & 1, & -1 \\ 0, & 11, & -15, & 11 \\ 0, & 1, & -1, & 3+\alpha \end{matrix} \right\| \sim \left\| \begin{matrix} -1, & 4, & -6, & 3 \\ 0, & -1, & 1, & -1 \\ 0, & 0, & -4, & 0 \\ 0, & 0, & 0, & 2+\alpha \end{matrix} \right\| .$$

We conclude that if $\alpha = -2$, then $r(\mathbf{A}) = 3$, and if $\alpha \neq -2$, then $r(\mathbf{A}) = 4$.

In dealing with square matrices, it is worth distinguishing the ones whose rank is the highest possible.

2.1.14. Definition. Let $\mathbf{A}$ be a square matrix of size $n \times n$. Then $\mathbf{A}$ is called *regular* if $r(\mathbf{A}) = n$. If $\mathbf{A}$ is not regular, it is called *singular*.

2.1.15. Example. Find all parameters $\alpha \in R$ such that $\mathbf{A}$ is singular:

$$\mathbf{A} = \left\| \begin{matrix} 4, & 0, & 1 \\ 2, & \alpha, & 0 \\ 0, & 2, & 1 \end{matrix} \right\| .$$

SOLUTION:

$$\left\| \begin{matrix} 4, & 0, & 1 \\ 2, & \alpha, & 0 \\ 0, & 2, & 1 \end{matrix} \right\| \sim \left\| \begin{matrix} 4, & 1, & 0 \\ 2, & 0, & \alpha \\ 0, & 1, & 2 \end{matrix} \right\| \sim \left\| \begin{matrix} 4, & 1, & 0 \\ 0, & 1, & 2 \\ 2, & 0, & \alpha \end{matrix} \right\| \sim \left\| \begin{matrix} 4, & 1, & 0 \\ 0, & 1, & 2 \\ 0, & 1, & -2\alpha \end{matrix} \right\| .$$

It follows that $\mathbf{A}$ is singular exactly when $\alpha = -1$.

The following important result, which will be proved later, states that matrix multiplication cannot increase the rank of matrices. Moreover, multiplication with a regular matrix preserves the rank.

2.1.16. Theorem. Suppose that the product $\mathbf{AB}$ is defined.
Then $r(\mathbf{AB}) \leq \min\{r(\mathbf{A}), r(\mathbf{B})\}$. Moreover, if $\mathbf{A}$ (resp. $\mathbf{B}$) is regular, then $r(\mathbf{AB}) = r(\mathbf{B})$ (resp. $r(\mathbf{AB}) = r(\mathbf{A})$).

It should be observed that the rank may actually decrease in matrix multiplication. For instance, if $\mathbf{A} = \begin{Vmatrix} 0 & 0 \\ 1 & 0 \end{Vmatrix}$, then $r(\mathbf{A}) = 1$, but $r(\mathbf{AA}) = r\begin{Vmatrix} 0 & 0 \\ 0 & 0 \end{Vmatrix} = 0$. (As an amusing excercise, show how the "resp." version of the "moreover" statement follows from the original version. A hint: Use $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.)

The inverse matrix

In applications we often need to deal with "inverses" taken with respect to matrix multiplication. In this paragraph we develop a method for computing inverses.

2.1.17. Definition. Let $\mathbf{A} \in \mathscr{M}$ (i.e. let $\mathbf{A}$ be a square matrix). Let $\mathbf{E} \in \mathscr{M}$ denote the unit matrix (i. e., $\mathbf{E} = \|e_{ij}\|$, where $e_{ij} = 0$ if $i \neq j$ and $e_{ij} = 1$ if $i = j$). A matrix $\mathbf{B} \in \mathscr{M}$ such that $\mathbf{BA} = \mathbf{E}$ is called *an inverse of the matrix* $\mathbf{A}$.

Obviously, the definition of inverse is meaningful for regular matrices only (indeed, since $\mathbf{E}$ is regular, $\mathbf{A}$ also has to be regular in view of Th. 2.2.10). It turns out (see the next proposition), that for regular matrices the above definition determines "inverses" uniquely. Thus, we may talk about *the* inverse matrix of $\mathbf{A}$, denoting the inverse matrix of $\mathbf{A}$ by $\mathbf{A}^{-1}$.

2.1.18. Proposition. If $\mathbf{B}, \mathbf{B}'$ are two inverses to $\mathbf{A}$, then $\mathbf{B} = \mathbf{B}'$. Moreover, if $\mathbf{B}$ is the inverse of $\mathbf{A}$, then $\mathbf{AB} = \mathbf{BA} = \mathbf{E}$ and $\mathbf{B}$ is regular.

PROOF: If $\mathbf{BA} = \mathbf{B}'\mathbf{A} = \mathbf{E}$, then $(\mathbf{B} - \mathbf{B}')\mathbf{A} = \mathbf{O}$, where $\mathbf{O}$ is the null matrix (the matrix $\mathbf{O}$ has all components equal to 0). Since $\mathbf{A}$ is regular, it follows (Th. 2.2.10) that $r(\mathbf{B} - \mathbf{B}') = r(\mathbf{O}) = 0$. Thus, $\mathbf{B} - \mathbf{B}' = \mathbf{O}$ and therefore $\mathbf{B} = \mathbf{B}'$.

Suppose that $\mathbf{B}$ is the inverse of $\mathbf{A}$. Consider the matrix $(\mathbf{AB})^2$. We can write $(\mathbf{AB})^2 = (\mathbf{AB})(\mathbf{AB}) = \mathbf{A}(\mathbf{BA})\mathbf{B} = \mathbf{AB}$. This means that $(\mathbf{AB})^2 = \mathbf{AB}$. Thus, $(\mathbf{AB})(\mathbf{AB} - \mathbf{E}) = \mathbf{O}$. Since $\mathbf{B}$ is also regular (Th. 2.2.10), we see that $\mathbf{AB}$ is regular (Th. 2.2.10) and therefore $(\mathbf{AB} - \mathbf{E}) = 0$. Thus, $\mathbf{AB} = \mathbf{E}$. □

In what follows we will show that *every regular matrix does have an inverse.* Prior to that, we will examine the Gaussian elimination procedure from the point of view of matrix multiplication. We will establish the following remarkable "meta-theorem": Everything that can be obtained by Gaussian elimination can also be obtained by matrix multiplication! In other words, the Gaussian elimination procedure can be simulated by matrix multiplication. This simple but enormously useful result will also be applicable in other places.

2.1.19. Observation. Let $\mathbf{A}$ be a square matrix. Suppose that we want to multiply the k-th row of $\mathbf{A}$ by a scalar λ. This may alternately be performed by the matrix multiplication $\mathbf{UA}$, where the matrix $\mathbf{U} = \|u_{ij}\|$ equals the unit matrix $\mathbf{E}$ on all components but u_{kk} for which we set $u_{kk} = \lambda$.

2.1.20. Example. The following example illustrates multiplication of the second row of $\mathbf{A}$ by the scalar 5.

$$\mathbf{A} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} \qquad \mathbf{U} = \begin{Vmatrix} 1, & 0, & 0, & 0 \\ 0, & 5, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & 0, & 0, & 0 \\ 0, & 5, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix} \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 5, & -5, & 35, & 30 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix}$$

2.1.21. Observation. Let $\mathbf{A}$ be a square matrix. Suppose that we want to interchange the r-th row of $\mathbf{A}$ with the s-th row of $\mathbf{A}$. This may alternately be performed by the matrix multiplication $\mathbf{UA}$, where the matrix $\mathbf{U} = \|u_{ij}\|$ equals the unit matrix $\mathbf{E}$ on all components with the exception of $u_{rr},\ u_{rs},\ u_{sr},\ u_{ss}$ for which $u_{rr} = 0,\ u_{rs} = 1,\ u_{sr} = 1,\ u_{ss} = 0$.

2.1.22. Example. The following example illustrates the interchange of the second row of $\mathbf{A}$ with the third row of $\mathbf{A}$.

$$\mathbf{A} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} \qquad \mathbf{U} = \begin{Vmatrix} 1, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix} \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 5, & 2, & 3, & 2 \\ 1, & -1, & 7, & 6 \\ 0, & 1, & 0, & 3 \end{Vmatrix}$$

2.1.23. Observation. Let $\mathbf{A}$ be a square matrix. Suppose that we want to add a linear combination of rows of $\mathbf{A}$ to the k-th row of $\mathbf{A}$. This may alternately be performed by the matrix multiplication $\mathbf{UA}$, where the matrix $\mathbf{U} = \|u_{ij}\|$ equals the unit matrix $\mathbf{E}$ on all components but the k-th row for which, for the combination with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \lambda_{k+1}, \ldots, \lambda_n$, the components are $u_{kj} = \lambda_j$ for $j \neq k$, and $u_{kk} = 1$.

2.1.24. Example. The following example illustrates the adding of a linear combination of rows to a given row: We want to add the combination "$2\cdot$ (the first row) $+(-1)\cdot$ (the third row) $+3\cdot$ (the fourth row)" to the second row of $\mathbf{A}$.

$$\mathbf{A} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} \qquad \mathbf{U} = \begin{Vmatrix} 1, & 0, & 0, & 0 \\ 2, & 1, & -1, & 3 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & 0, & 0, & 0 \\ 2, & 1, & -1, & 3 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & 1 \end{Vmatrix} \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 1, & -1, & 7, & 6 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ -2, & 4, & 10, & 21 \\ 5, & 2, & 3, & 2 \\ 0, & 1, & 0, & 3 \end{Vmatrix}$$

The previous considerations suggest the way to compute $\mathbf{A}^{-1}$ for a given regular matrix $\mathbf{A}$. If we manipulate the matrix $\mathbf{A}$ with Gaussian elimination so that we end up with a unit matrix, $\mathbf{E}$, and if we consecutively "copy" the same steps of the Gaussian elimination we used on the matrix $\mathbf{E}$, the matrix $\mathbf{A}$ will become $\mathbf{E}$ and the matrix $\mathbf{E}$ will become $\mathbf{A}^{-1}$. Symbolically,

$$\begin{aligned}\mathbf{A} &\Longrightarrow \mathbf{E}\\ \mathbf{E} &\Longrightarrow \mathbf{A}^{-1}.\end{aligned}$$

Before formulating and proving this in the next theorem, let us consider two examples.

2.1.25. Example. Compute $\mathbf{A}^{-1}$ for $\mathbf{A} = \begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix}$.

SOLUTION: We write $\mathbf{A}$ and $\mathbf{E}$ and carry on the elimination commented on above. We consecutively obtain

$$\begin{array}{cc}
\begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix} & \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix}, \\[2ex]
\begin{Vmatrix} 1, & 2 \\ 0, & 3 \end{Vmatrix} & \begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix} \\[2ex]
\begin{Vmatrix} 3, & 0 \\ 0, & 3 \end{Vmatrix} & \begin{Vmatrix} 1, & -2 \\ 1, & 1 \end{Vmatrix}, \\[2ex]
\begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix} & \dfrac{1}{3} \circ \begin{Vmatrix} 1, & -2 \\ 1, & 1 \end{Vmatrix} = \mathbf{A}^{-1}.
\end{array}$$

Indeed, $\begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix} \cdot \left(\frac{1}{3} \circ \begin{Vmatrix} 1, & -2 \\ 1, & 1 \end{Vmatrix} \right) = \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix}$.

2.1.26. Example. Compute $\mathbf{A}^{-1}$ for $\mathbf{A} = \begin{Vmatrix} 0, & 1, & 2 \\ 1, & -1, & 1 \\ 2, & 0, & 1 \end{Vmatrix}$.

SOLUTION:

$$\begin{Vmatrix} 0, & 1, & 2 \\ 1, & -1, & 1 \\ 2, & 0, & 1 \end{Vmatrix} \qquad \begin{Vmatrix} 1, & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & -1, & 1 \\ 0, & 1, & 2 \\ 2, & 0, & 1 \end{Vmatrix} \qquad \begin{Vmatrix} 0, & 1, & 0 \\ 1, & 0, & 0 \\ 0, & 0, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & -1, & 1 \\ 0, & 1, & 2 \\ 0, & 2, & -1 \end{Vmatrix} \qquad \begin{Vmatrix} 0, & 1, & 0 \\ 1, & 0, & 0 \\ 0, & -2, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & -1, & 1 \\ 0, & 1, & 2 \\ 0, & 0, & -5 \end{Vmatrix} \qquad \begin{Vmatrix} 0, & 1, & 0 \\ 1, & 0, & 0 \\ -2, & -2, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 5, & -5, & 0 \\ 0, & 5, & 0 \\ 0, & 0, & -5 \end{Vmatrix} \qquad \begin{Vmatrix} -2, & 3, & 1 \\ 1, & -4, & 2 \\ -2, & -2, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 5, & 0, & 0 \\ 0, & 5, & 0 \\ 0, & 0, & -5 \end{Vmatrix} \qquad \begin{Vmatrix} -1, & -1, & 3 \\ 1, & -4, & 2 \\ -2, & -2, & 1 \end{Vmatrix}$$

$$\begin{Vmatrix} 1, & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{Vmatrix} \qquad \frac{1}{5}\circ \begin{Vmatrix} -1, & -1, & 3 \\ 1, & -4, & 2 \\ 2, & 2, & -1 \end{Vmatrix} = \mathbf{A}^{-1}$$

The general case will be taken care of analogously.

2.1.27. Theorem. Let $\mathbf{A}$ be a regular matrix. Then $\mathbf{A}^{-1}$ exists.

PROOF: Let us only sketch the proof with the details being obvious. Suppose that $\mathbf{A} \in \mathscr{M}$ is a regular matrix. Let us show that there is a (finite) family of matrices, some matrices $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \dots, \mathbf{U}_{p-1}, \mathbf{U}_p$ such that $\mathbf{U}_p\mathbf{U}_{p-1}\dots\mathbf{U}_3\mathbf{U}_2\mathbf{U}_1\mathbf{A} = \mathbf{E}$. This is sufficient – we will have $\mathbf{A}^{-1} = \mathbf{U}_p\mathbf{U}_{p-1}\dots\mathbf{U}_3\mathbf{U}_2\mathbf{U}_1$.

Let us formally divide the proof into a few steps.

1. If $a_{11} = 0$, then we take for $\mathbf{U}_1$ a matrix that interchanges the rows of $\mathbf{A}$ so that $\mathbf{UA}$ has a nonzero element at the place $1,1$. This is possible since the matrix $\mathbf{A}$ cannot have zeros in the entire first column ($\mathbf{A}$ is regular). If $a_{11} \neq 0$, nothing happens in this step (i.e., $\mathbf{U}_1 = \mathbf{E}$).
2. We find matrices $\mathbf{U}_2, \mathbf{U}_3, \dots, \mathbf{U}_n$ such that the matrix $\mathbf{U}_n\mathbf{U}_{n-1}\dots\mathbf{U}_2\mathbf{U}_1\mathbf{A}$ has zeros in the first column at all places except for $1,1$. Thus, we find matrices that perform the usual elimination in the first column (such matrices exist as we have established in the course of observations).
3. We will proceed in an analogous way so that we obtain a matrix that possesses zeros everywhere below the diagonal. An important feature of this matrix is that all its components on the diagonal must be nonzero! (Indeed, if it is not the case, the matrix is

singular (prove in detail!). Since this matrix has been obtained from the original one only by applying operations that do not change the rank, the original matrix would have to be singular, too. This contradicts our assumption.)

4. At this stage we have obtained a matrix, $\tilde{\mathbf{A}} = \|\tilde{a}_{ij}\|$, such that $\tilde{\mathbf{A}} = \mathbf{U}_r\mathbf{U}_{r-1}\dots \mathbf{U}_3\mathbf{U}_2\mathbf{U}_1\mathbf{A}$ for some matrices $\mathbf{U}_i$ $(i \le r)$. The matrices $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n$ have simulated the Gaussian elimination procedure. Moreover, we know that $\tilde{a}_{ij} = 0$ whenever $i > j$ and $\tilde{a}_{i,i} \neq 0$ for any $i \in \{1, 2, \dots, n\}$. In particular, $\tilde{a}_{n,n} \neq 0$ and we can carry out the elimination "from the bottom to the top". This will be performed by matrices $\mathbf{U}_{r+1}, \mathbf{U}_{r+2}, \dots, \mathbf{U}_p$ with the effect that $\mathbf{U}_p\mathbf{U}_{p-1}\dots\mathbf{U}_{r+1}\tilde{\mathbf{A}} = \mathbf{E}$. Putting this together with the previous steps, we have $\mathbf{U}_p\mathbf{U}_{p-1}\dots\mathbf{U}_3\mathbf{U}_2\mathbf{U}_1\mathbf{A} = \mathbf{E}$. This completes the proof.

□

It is natural to call the matrices used in Observations 2.3.3, 2.3.5 and 2.3.7 *the matrices of (Gaussian) elimination*. We have thoroughly applied them in the above theorem. As another application, let us sketch the proof of Th. 2.2.10. We want to show that $r(\mathbf{AB}) \le \min\{r(\mathbf{A}), r(\mathbf{B})\}$. Let us first realize two facts that we formulate as statements.

Statement 1: If $\mathbf{U}$ is a matrix of elimination type, then $\mathbf{U}^{-1}$ is also a matrix of elimination type.

The proof of this statement is easy and amounts to direct verification. We leave it to the reader as an exercise. (For instance, if

$$\mathbf{U} = \begin{Vmatrix} 1, & 0, & 0 \\ \lambda_1, & 1, & \lambda_2 \\ 0, & 0, & 1 \end{Vmatrix}, \text{ then } \mathbf{U}^{-1} = \begin{Vmatrix} 1, & 0, & 0 \\ -\lambda_1, & 1, & -\lambda_2 \\ 0, & 0, & 1 \end{Vmatrix}, \text{ etc.}).$$

Statement 2: Suppose that $\mathbf{A} \in \mathscr{M}$ and suppose that $r(\mathbf{A}) = p$. Then we can find a matrix $\mathbf{C} \in \mathscr{M}, \mathbf{C} = \|\|$ such that

(i) $c_{ij} = 0$ for $i > p$ (i.e., all the row vectors of $\mathbf{C}$ that are strictly below the p-th vector consist of zeros),

(ii) there are matrices $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_k$ of elimination type such that $\mathbf{A} = \mathbf{U}_k\mathbf{U}_{k-1}\dots \mathbf{U}_2\mathbf{U}_1\mathbf{C}$.

PROOF: Performing the Gaussian elimination on $\mathbf{A}$, we obtain $\mathbf{C} = \tilde{\mathbf{U}}_k\tilde{\mathbf{U}}_{k-1}\dots\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_1\mathbf{A}$ for some matrices $\tilde{\mathbf{U}}_i$ $(i \le k)$ of elimination type. Then $\mathbf{A} = \tilde{\mathbf{U}}_1^{-1}\tilde{\mathbf{U}}_2^{-1}\dots\tilde{\mathbf{U}}_k^{-1}\mathbf{C}$, and it suffices to put $\mathbf{U}_j = \tilde{\mathbf{U}}_{k-j}^{-1}$ $(j \le k)$.

The proof of Th. 2.2.10 can now be easily provided. Consider the matrix product $\mathbf{AB}$. Suppose that $\mathbf{A} \in \mathscr{M}$ and $r(\mathbf{A}) = p$. Then $\mathbf{AB} = \mathbf{U}_k\mathbf{U}_{k-1}\dots\mathbf{U}_1\mathbf{CB}$, where $\mathbf{C}$ has zeros below the p-th row and $\mathbf{U}_i$ $(i \le k)$ are matrices of elimination type. Thus, $r(\mathbf{CB}) \le r(\mathbf{B})$. Since the matrices $\mathbf{U}_i$ $(i \le k)$ do not change the rank (Prop. 2.2.5), we obtain $r(\mathbf{U}_k\mathbf{U}_{k-1}\dots\mathbf{U}_1\mathbf{CB}) = r(\mathbf{CB}) \le r(\mathbf{B})$. This gives us the inequality $r(\mathbf{AB}) \le r(\mathbf{B})$. The rest easily follows from the duality $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ and Th. 2.2.3. If $\mathbf{A}$ is regular, then we apply only the matrices of elimination type, i.e. we put $\mathbf{B} = \mathbf{E}$ in Statement 2. This means that $r(\mathbf{AB}) = r(\mathbf{B})$ (Prop. 2.2.5).

2.2. Determinants

Let us consider a square matrix, $\mathbf{A} \in \mathscr{M}$, taken arbitrarily but kept fixed in the following considerations. Thus,

$$\mathbf{A} = \begin{Vmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & & & \\ a_{n1}, & a_{n2}, & \dots, & a_{nn} \end{Vmatrix}.$$

It is often desirable to find out whether $\mathbf{A}$ is regular or singular. There is an evaluation of matrices that *determines it.* This evaluation is called a determinant and will be the topic of our interest in this paragraph.

In order to define the determinant of our matrix $\mathbf{A}$, let us select the n-tuples $(a_{1k_1}, a_{2k_2}, \dots, a_{nk_n})$ of the components of $\mathbf{A}$ that contain *exactly one element of every row of* $\mathbf{A}$ *and exactly one element of every column of* $\mathbf{A}$. In other words, let us consider just those n-tuples for which the indices $k_1, k_2, \dots, k_n$ are mutually distinct. Let us call these n-tuples *distinguished n-tuples.* Let us denote the set of all distinguished n-tuples by $\mathscr{T}(\mathbf{A})$. (Observe that for $n = 8$ the choice of n-tuples corresponds to the positions of eight rooks on the chess board so that the rooks do not tackle each other. This interpretation is sometimes helpful).

2.2.1. Proposition. The cardinality of $\mathscr{T}(\mathbf{A})$ is $n!$ (i.e. $\mathscr{T}(\mathbf{A})$ possesses exactly n-factorial elements).

PROOF: Choosing from the first row we have exactly n possibilities. Choosing than from the second row, we have exactly $(n-1)$ possibilities, etc. Thus, the total number of ways to pick a distinguished n-tuple of $\mathbf{A}$ is $n \cdot (n-1) \cdot (n-2) \dots 2 \cdot 1 = n!$. □

2.2.2. Definition. Let $t = (a_{1k_1}, a_{2k_2}, \dots, a_{nk_n})$ be a distinguished n-tuple of $\mathbf{A}$. Let v be the number of those pairs (a_{pq}, a_{rs}) in t such that $p > r$ and $q < s$. Then the number $(-1)^v$ is called *the sign of* t ($\operatorname{sign} t$). By *an oriented product* of t, we mean the product of all elements of t multiplied with $\operatorname{sign} t$. (Thus, $\operatorname{sign} t$ is $+1$ if v is even and $\operatorname{sign} t = -1$ if v is odd.)

In computing $\operatorname{sign} t$ for a given distinguished n-tuple t, it is handy to realize the following mnemonic. If (a_{pq}, a_{rs}) is a pair of t for which the equalities $p > r$ and $q < s$ are true, then the line segment drawn in $\mathbf{A}$ connecting a_{pq} with a_{rs} is directed left-downwards. Let us call this pair *a leftdown.* Thus, if we want to find $\operatorname{sign} t$, we compute all the leftdowns of t and attach 1 or -1 if the number of leftdowns is found to be even or odd.

After these prerequisites we are in a position to formulate the definition of determinant.

2.2.3. Definition. The determinant of $\mathbf{A}$ (in symbols, $\det \mathbf{A}$) is the sum of all oriented products when we go through all distinguished n-tuples of $\mathbf{A}$.

2.2.4. Examples.

1. If $n = 1$, then $\mathbf{A} = a_{11}$. In this case $\det \mathbf{A} = a_{11}$.
2. If $n = 2$, then $\mathbf{A} = \begin{Vmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \end{Vmatrix}$. In this case $\det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}$.
3. If $n = 3$, then $\mathbf{A} = \begin{Vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{Vmatrix}$. In this case we have to go through all distinguished triples of $\mathbf{A}$. Since every distinguished triple has to hit the first row and, moreover, since it has to do it *exactly ones*, we can select all distinguished triples by checking the elements belonging to the first row. Observing this, we organize the distinguished triples corresponding to a_{11}, a_{12} and a_{13}. We see that the distinguished triples corresponding to a_{11} are $a_{11}a_{22}a_{33}$ and $a_{11}a_{23}a_{32}$, the distinguished triples corresponding to a_{12} are $a_{12}a_{21}a_{33}$ and $a_{12}a_{23}a_{31}$, and the distinguished triples corresponding to a_{13} are $a_{13}a_{22}a_{31}$ and $a_{13}a_{21}a_{32}$. After computing the signs, which we can do by applying our mnemonic, we conclude that

$$\det \mathbf{A} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{32} - a_{13}a_{22}a_{31} + a_{13}a_{21}a_{32}.$$

   This agrees with the well-known Sarruss' rule dealt with at the secondary school.

In principle, even for $n > 3$, we could compute the determinants from definition. However, this would not be practical. Also, we will need a better understanding of determinants for theoretical reasons.

2.2.5. Proposition. $\det \mathbf{A} = \det \mathbf{A}^T$.

PROOF: This equality can be verified directly – every distinguished n-tuple of $\mathbf{A}$ is a distinguished n-tuple of $\mathbf{A}^T$, and vice versa. This one-to-one correspondance immediately gives $\det \mathbf{A} = \det \mathbf{A}^T$. □

The above result in turn establishes a kind of duality - the results on rows that concern determinants also allow a literary translation into the corresponding column formulation. In the propositions which follow we will make use of this circumstance without explicitly commenting on it.

2.2.6. Proposition. Let us interchange two rows (resp. two columns) of $\mathbf{A}$ and let us denote the resulting matrix by $\mathbf{B}$. Then $\det \mathbf{B} = -\det \mathbf{A}$.

PROOF: Let us first assume that we have interchanged the rows that are next to each other, say the i-th row with the $i+1$-st row. Then all distinguished n-tuples of $\mathbf{A}$ obviously become distinguished n-tuples of the resulting matrix $\mathbf{B}$ and moreover, all n-tuples of $\mathbf{B}$ will be exhausted this way. The important fact is the change of sign - if the elements in the i-th and $i+1$-st row made a leftdown in $\mathbf{A}$, then they would not make it after the interchange. This shows $\det \mathbf{B} = -\det \mathbf{A}$ in view of the definition of determinants.

Let us now take up a general case. Then we apply the following simple combinatorial fact: an arbitrary interchange of rows can be reached in the course of interchanging the rows that are next to each other and moreover, to complete the interchange we need *an odd number of steps!*

Indeed, for the interchange of the i-th row with the j-th row we need exactly $2(j-1)-1$ such steps. It follows that $\det \mathbf{B} = (-1)^{2(j-i)-1} \det \mathbf{A}$, which means $\det \mathbf{B} = -\det \mathbf{A}$. □

To illustrate the meaning of the next result, let us first consider an example.

2.2.7. Example. Without computing the determinant, convince yourself that

$$\det \begin{Vmatrix} 1, & -2, & 3 \\ 1, & 2, & -1 \\ 1, & 0, & -1 \end{Vmatrix} + \det \begin{Vmatrix} 1, & -2, & 3 \\ 2, & 1, & -2 \\ 1, & 0, & -1 \end{Vmatrix} = \det \begin{Vmatrix} 1, & -2, & 3 \\ 3, & 3, & -3 \\ 1, & 0, & -1 \end{Vmatrix}$$

SOLUTION: It suffices to observe that all the matrices involved agree on all but the second row, where the matrix on the right-hand side has the sum of the components of the matrices on the left-hand side. The equality then immediately follows from the definition of determinant (provide the details considering the oriented products).

2.2.8. Proposition. Suppose that for a natural number $p \in N$ square matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_p, \mathbf{B}$ have the same components on all but the i-th row. Suppose further that every component of $\mathbf{B}$ lying in the i-th row is the sum of all corresponding components of all matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$. Then $\det \mathbf{B} = \det \mathbf{A}_1 + \det \mathbf{A}_2 + \cdots + \det \mathbf{A}_p$.

PROOF: It follows from the definition of determinant — for every oriented product $\pm b_{1k_1} b_{2k_2} \ldots b_{nk_n}$ we have

$$\pm b_{1k_1} b_{2k_2} \ldots b_{nk_n} = \pm b_{1k_1} b_{2k_2} \ldots b_{i-1,k_{i-1}} (a^1_{ik_i} + a^2_{ik_i} + \cdots + a^p_{ik_i}) b_{i+1,k_{i+1}} \ldots b_{nk_n},$$

where we set $\mathbf{A}_p = \|a^p_{i,j}\|$.

The previous result is worth remembering - it will be heavily applied in the important Prop. 2.4.11 that follows. (Observe in passing that there is no general additivity property for determinants - the alleged equality $\det \mathbf{A}_1 + \det \mathbf{A}_2 = \det(\mathbf{A}_1 + \mathbf{A}_2)$ is almost never valid.)

2.2.9. Proposition. Let us multiply a row (resp. a column) of $\mathbf{A}$ with a scalar λ ($\lambda \in R$). Let us denote by $\mathbf{B}$ the resulting matrix. Then $\det \mathbf{B} = \lambda \cdot \det \mathbf{A}$. A consequence: $\det(\lambda \circ \mathbf{A}) = \lambda^n \cdot \det \mathbf{A}$.

PROOF: Trivial. □

2.2.10. Proposition. Suppose that the i-th row (resp. the i-th column) of $\mathbf{A}$ coincides with the j-th row (resp. with the j-th column) of $\mathbf{A}$ ($i \neq j$). Then $\det \mathbf{A} = 0$.

PROOF: We employ the following neat idea. Suppose $\det \mathbf{A} = a$. Observing that $\mathbf{A}$ does not change by interchanging the i-th and j-th row and applying Prop. 2.4.6, we have $a = \det \mathbf{A} = -\det \mathbf{A} = -a$. It follows that $a = -a$, and therefore $a = 0$. Thus, $\det \mathbf{A} = 0$. □

2.2.11. Proposition. Let us add to a given row (resp. column) of $\mathbf{A}$ a linear combination of the other rows (resp. the other columns). Let us leave the other rows (resp. columns) unchanged and let us denote the resulting matrix by $\mathbf{B}$. Then $\det \mathbf{B} = \det \mathbf{A}$.

PROOF: Assume that we add a linear combination we consider to the first row (this is a harmless assumption since we might easily modify a general situation to this case). Assume that the coefficients of the combination are scalars $\lambda_2, \lambda_3, \ldots, \lambda_n$. We will now apply Prop. 2.4.8. Indeed, for our matrix $\mathbf{B}$ it is not difficult to find matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ such that the following properties are fulfilled:

(i) $\mathbf{A}_1 = \mathbf{A}$ and for any $p, p > 1$, the matrix $\mathbf{A}_p$ has for its first row a λ_p-multiple of the p-th row,

(ii) the matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n, \mathbf{B}$ satisfy the assumptions of Prop. 2.4.8.

The rest is easy. We obtain $\det \mathbf{B} = \det \mathbf{A}_1 + \det \mathbf{A}_2 + \cdots + \det \mathbf{A}_n$. Since $\det \mathbf{A}_1 = \det \mathbf{A}$ and since $\det \mathbf{A}_p = 0$ for any $p > 1$ (Prop. 2.4.9, 2.4.10), we obtain $\det \mathbf{B} = \det \mathbf{A}$. □

The result derived above provides a method for efficiently computing determinants. Before we show it by examples, let us observe that the determinant of matrices which have all components under the diagonal equal to zero compute easily. Indeed, if $\mathbf{A}$ is an "upper triangular" matrix as described above then $\det \mathbf{A} = a_{11} a_{22} \ldots a_{nn}$. This fact is easy to prove - the product written on the right-hand side is the only (oriented) product of $\mathbf{A}$ that may be nonzero (give a precise argument for this!).

2.2.12. Example. Compute $\det \mathbf{A}$, where

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

SOLUTION: Using Prop. 2.4.11, we have

$$\det \mathbf{A} = \det \begin{Vmatrix} 1, & 1, & 2 \\ -1, & 0, & 3 \\ 2, & 1, & 4 \end{Vmatrix} = \det \begin{Vmatrix} 1, & 1, & 2 \\ 0, & 1, & 5 \\ 0, & -1, & 0 \end{Vmatrix} = \det \begin{Vmatrix} 1, & 1, & 2 \\ 0, & 1, & 5 \\ 0, & 0, & 5 \end{Vmatrix} = 1 \cdot 1 \cdot 5 = 5$$

2.2.13. Example. Compute $\det \mathbf{A}$, where

$$\mathbf{A} = \begin{Vmatrix} 3, & -3, & -2, & -5 \\ 2, & 5, & 4, & 6 \\ 5, & 5, & 8, & 7 \\ 1, & 1, & 3, & 1 \end{Vmatrix}.$$

Solution:

$$\det \mathbf{A} = \det \begin{Vmatrix} 3, & -3, & -2, & -5 \\ 2, & 5, & 4, & 6 \\ 5, & 5, & 8, & 7 \\ 1, & 1, & 3, & 1 \end{Vmatrix} = -\det \begin{Vmatrix} 1, & 1, & 3, & 1 \\ 2, & 5, & 4, & 6 \\ 5, & 5, & 8, & 7 \\ 3, & -3, & -2, & -5 \end{Vmatrix} =$$

$$= -\det \begin{Vmatrix} 1, & 1, & 3, & 1 \\ 0, & 3, & -2, & 4 \\ 0, & 0, & -7, & 2 \\ 0, & -6, & -11, & -8 \end{Vmatrix} = -\det \begin{Vmatrix} 1, & 1, & 3, & 1 \\ 0, & 3, & -\ 2, & 4 \\ 0, & 0, & -\ 7, & 2 \\ 0, & 0, & -15, & 0 \end{Vmatrix} =$$

$$= \frac{-1}{15 \cdot (-7)} \det \begin{Vmatrix} 1, & 1, & 3, & 1 \\ 0, & 3, & -2. & 4 \\ 0, & 0, & -105, & 30 \\ 0, & 0, & 105, & 0 \end{Vmatrix} = \frac{-1}{15 \cdot (-7)} \det \begin{Vmatrix} 1, & 1, & 3, & 1 \\ 0, & 3, & -2. & 4 \\ 0, & 0, & -105, & 30 \\ 0, & 0, & 0, & 30 \end{Vmatrix} =$$

$$= \frac{-1}{15 \cdot (-7)} \cdot (1 \cdot 3 \cdot (-105) \cdot 30) = -90.$$

The above example should remind a careless student that the “elimination method” for computing determinants requires a higher degree of concentration than the elimination method for computing the rank. When we multiply a row (column) we must not forget that the determinant is adequately multiplied (reconsider Prop. 2.4.10). Thus, in order to maintain the equality of determinants, we have to divide by the scalar with which we multiplied (this is where the fraction $\dfrac{-1}{15 \cdot (-7)}$ came from).

2.2.14. Example. Suppose we know that the numbers 357, 714 and 476 are divisible by 17. Show that the number $\det \begin{Vmatrix} 3, & 5, & 7 \\ 7, & 1, & 4 \\ 4, & 7, & 6 \end{Vmatrix}$ is divisible by 17.

Solution:

$$d = \det \begin{Vmatrix} 3, & 5, & 7 \\ 7, & 1, & 4 \\ 4, & 7, & 6 \end{Vmatrix} = \frac{1}{100} \cdot \det \begin{Vmatrix} 300, & 5, & 7 \\ 700, & 1, & 4 \\ 400, & 7, & 6 \end{Vmatrix} = \frac{1}{100} \cdot \det \begin{Vmatrix} 357, & 5, & 7 \\ 714, & 1, & 4 \\ 476, & 7, & 6 \end{Vmatrix} =$$

$$= \frac{1}{100} \cdot 17e$$

(d, e are natural numbers). Thus, $100d = 17e$ and therefore d is divisible by 17.

The above examples indicate how one utilizes the “elimination” properties of determinants for the purpose of computation. An even more important application is the following theoretical result which reveals an elegant characterization of regular matrices.

2.2.15. Theorem. Let $\mathbf{A}$ be a square matrix. Then $\mathbf{A}$ is regular if and only if $\det \mathbf{A} \neq 0$.

Proof: We have to verify the validity of two implications. Let us first take up the implication $\det \mathbf{A} \neq 0 \Rightarrow \mathbf{A}$ is regular.

Suppose that $\mathbf{A} \in \mathscr{M}$ and $\det \mathbf{A} \neq 0$. By the consecutive application of "elimination" operations on the rows of $\mathbf{A}$, we can convert the matrix $\mathbf{A}$ to an upper triangular matrix (i.e. to a matrix which has all components under the diagonal equal to 0). Let us denote this matrix by $\mathbf{B}$. Thus,

$$\mathbf{B} = \begin{Vmatrix} b_1, & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & b_2, & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & b_3, & \cdot & \cdot & \cdot \\ \vdots & & & & & \\ 0, & 0, & 0, & \dots, & 0, & b_n \end{Vmatrix}.$$

It is obvious (Prop. 2.4.6, Prop. 2.4.9 and Prop. 2.4.11) that the assumption of $\det \mathbf{A} \neq 0$ implies $\det \mathbf{B} \neq 0$. Thus, $\det \mathbf{B} = b_1 \cdot b_2 \dots b_n \neq 0$ and therefore $b_i \neq 0$ for any i $(i \leq n)$. This implies that $\mathbf{B}$ is regular. But $\mathbf{B}$ has arisen from $\mathbf{A}$ in the course of "elimination" and therefore $\mathbf{A}$ is regular (Prop. 2.2.5). This completes the proof of the implication $\det \mathbf{A} \neq 0 \Rightarrow \mathbf{A}$ is regular.

The reverse implication derives analogously. We only sketch the argument.

Suppose that $\mathbf{A} \in \mathscr{M}$ and $\mathbf{A}$ is regular. By applying the elimination procedure on the rows of $\mathbf{A}$, we convert $\mathbf{A}$ into $\mathbf{B}$, where

$$\mathbf{B} = \begin{Vmatrix} b_1, & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & b_2, & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & b_3, & \cdot & \cdot & \cdot \\ \vdots & & & & & \\ 0, & 0, & 0, & \dots, & 0, & b_n \end{Vmatrix}.$$

Since $\mathbf{A}$ is regular then so is $\mathbf{B}$ (Prop. 2.2.5). It follows that any b_i $(i \leq n)$ must differ from 0. Indeed, if this is not the case, then there must be *the greatest index* i $(i \leq n)$ for which $b_i = 0$. In this case it is easy to see that the i-th row is a linear combination of the rows lying below the i-th row. This would be a contradiction. Thus, we conclude that $b_i \neq 0$ for any i $(i \leq n)$. It follows that $\det \mathbf{B} = b_1 \cdot b_2 \dots b_n \neq 0$, and therefore (Prop. 2.4.6, Prop. 2.4.9 and Prop. 2.4.11) $\det \mathbf{A} \neq 0$. This completes the proof. □

The previous theorem can be restated the following way.

2.2.16. Corollary . If $\mathbf{A} \in \mathscr{M}$, then $\det \mathbf{A} = 0 \Leftrightarrow r(\mathbf{A}) < n \Leftrightarrow$ the row vectors (resp. the column vectors) are LD in $\mathbf{R}^n$.

2.2.17. Example. Discuss the rank of $\mathbf{A}$ with respect to the parameter α:

$$\mathbf{A} = \begin{Vmatrix} \alpha, & 1, & 1 \\ 1, & \alpha, & 1 \\ 1, & 1, & \alpha \end{Vmatrix}.$$

SOLUTION: Computing $\det \mathbf{A}$, we obtain

$$\det \mathbf{A} = \alpha(\alpha^2 - 1) - 1(\alpha - 1) + 1(1 - \alpha) = (\alpha - 1)^2(\alpha + 2).$$

Thus,

(i) if $\alpha \neq 1, \alpha \neq -2 \Rightarrow \mathbf{A}$ is regular, i.e. $r(\mathbf{A}) = 3$,

(ii) if $\alpha = 1$, then we have $r(\mathbf{A}) = 1$

(iii) if $\alpha = -2$, then we have $r(\mathbf{A}) = 2$.

Probably the most practical method for computing determinants is the following recurrent technique which will be the subject of our interest now. This technique also has considerable theoretical bearing. The following notion will play a central rôle.

2.2.18. Definition. Let $\mathbf{A} = \|a_{ij}\| \in \mathscr{M}$. Let us assume that we have deleted both the i-th row and the j-th column from $\mathbf{A}$. Then we have obtained a matrix of the size $(n-1, n-1)$. Let us denote the determinant of this smaller matrix by D_{ij} and set $A_{ij} = (-1)^{i+j} D_{ij}$. Then A_{ij} is called *the* i, j*-th cofactor of* $\mathbf{A}$.

The notion of cofactor allows us to reduce the computing of $\det \mathbf{A}$, for $\mathbf{A} \in \mathscr{M}$, to computing n determinants of matrices belonging to $\mathscr{M}^{-,-}$. The following important result says how this can be achieved.

2.2.19. Theorem(the expansion of determinant along a row):Suppose that $\mathbf{A} \in \mathscr{M}$. Let i be a fixed row index. Then $\det \mathbf{A} = \sum_{j=1}^{n} a_{ij} A_{ij}$.

Before proving this theorem, let us first explicitly formulate its dual setup and illustrate it by examples.

2.2.20. Theorem(the expansion of determinant down a column): Suppose that $\mathbf{A} \in \mathscr{M}$. Let j be a fixed column index. Then $\det \mathbf{A} = \sum_{i=1}^{n} a_{ij} A_{ij}$.

It is obvious that Th. 2.4.20 follows from Th. 2.4.19 (one makes use of Prop. 2.4.5).

2.2.21. Example. Using the theorem on expansion of determinants, compute $\det \mathbf{A}$, where

$$\mathbf{A} = \begin{Vmatrix} 6, & 1, & -8 \\ 9, & 2, & 7 \\ -1, & 3, & 0 \end{Vmatrix}.$$

SOLUTION: Let us first expand the determinant along the second row. According to Th. 2.4.19, we obtain

$$\det \mathbf{A} = 9 \cdot (-1)^{2+1} \cdot \det \begin{Vmatrix} 1, & -8 \\ 3, & 0 \end{Vmatrix} + 2 \cdot (-1)^{2+2} \cdot \det \begin{Vmatrix} 6, & -8 \\ -1, & 0 \end{Vmatrix} + 7 \cdot (-1)^{2+3} \cdot \det \begin{Vmatrix} 6, & 1 \\ -1, & 3 \end{Vmatrix} =$$
$$= -9 \cdot 24 + 2 \cdot (-8) - 7 \cdot 19 = -216 - 16 - 133 = -365$$

Let us now compute it alternately by expanding the determinant down the third column. This seems to be handy, for the third column contains 0. According to Th. 2.4.20, we obtain

$$\det \mathbf{A} = (-8) \cdot (-1)^{1+3} \cdot \det \begin{Vmatrix} 9, & 2 \\ -1, & 3 \end{Vmatrix} + 7 \cdot (-1)^{2+3} \cdot \det \begin{Vmatrix} 6, & 1 \\ -1, & 3 \end{Vmatrix} =$$
$$= (-8) \cdot 29 - 7 \cdot 19 = -232 - 133 = -365.$$

Let us return to the proof of Th. 2.4.19. Let us keep the row index i fixed. Then every distinguished n-tuple contains exactly one a_{ij} $(j \leq n)$. Let us denote by Z_{ij} the sum of all oriented products of all distinguished n-tuples that contain a_{ij}. Obviously, $\det \mathbf{A} = \sum_{j=1}^{n} Z_{ij}$. We will show that $Z_{ij} = a_{ij} A_{ij}$, and this will complete the proof.

Consider the element a_{ij}. In order to avoid complications that might occur in computing leftdowns of distinguished n-tuples which contain a_{ij}, let us first move the component a_{ij} to the place $1,1$ (the component $1,1$ does not make any leftdowns!). Making $i-1$ steps of row interchanges and $j-1$ steps of column interchanges, we obtain a new matrix, some $A' = \|a'_{ij}\|$, so that $a_{ij} = a'_{11}$. By Prop. 2.4.6, $\det \mathbf{A} = (-1)^{i-1+j-1} \cdot \det \mathbf{A}' = (-1)^{i+j} \cdot \det \mathbf{A}'$. Moreover, the sum of all oriented products of the distinguished n-tuples of $\mathbf{A}$ that contain a_{ij}, which is nothing but Z_{ij}, can then be expressed as follows:

$Z_{ij} = (-1)^{i+j} \cdot$ (the sum of all oriented products of the distinguished n-tuples of $\mathbf{A}'$ that contain a'_{11}) $= (-1)^{i+j} \cdot \det \mathbf{B}$, where $\mathbf{B}$ is the matrix which arises when we delete the first row and the first column of $\mathbf{A}'$. The definition of A_{ij} now gives $Z_{ij} = a_{ij} A_{ij}$. Thus,

$$\det \mathbf{A} = \sum_{j=1}^{n} Z_{ij} = \sum_{j=1}^{n} a_{ij} A_{ij}.$$

□

The previous theorem, which is sometimes formulated as an alternative definition of determinants, can be adopted as the main tool for computing determinants of matrices of all sizes. Besides, the theorem has several important theoretical consequences. Let us see first how it can be applied to the problem of constructing $\mathbf{A}^{-1}$. The construction is sometimes called *the determinant method for computing* $\mathbf{A}^{-1}$.

2.2.22. Theorem. Let $\mathbf{A}$ be a regular square matrix, $\mathbf{A} \in \mathscr{M}$. Then

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{Vmatrix} A_{11}, & A_{21}, & \dots, & A_{n1} \\ A_{12}, & A_{22}, & \dots, & A_{n2} \\ \vdots & & & \\ A_{1n}, & A_{2n}, & \dots, & A_{nn} \end{Vmatrix},$$

where A_{ij} are the respective cofactors.

PROOF: Put $\tilde{\mathbf{A}} = \|\tilde{a}_{ij}\|$, where $\tilde{a}_{ij} = \dfrac{A_{ji}}{\det \mathbf{A}}$. We must show that $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$. Thus, we must show that $\mathbf{A}\tilde{\mathbf{A}} = \mathbf{E}$, where $\mathbf{E}$ is the unit matrix. Let us set $\mathbf{C} = \mathbf{A}\tilde{\mathbf{A}}$ and write $\mathbf{C} = \|c_{ij}\|$. We must show that c_{ij} is "the Cronecker delta" (i.e. $c_{ij} = 1$ if $i = j$ and $c_{ij} = 0$ otherwise). We see that

$$c_{ij} = \sum_{k=1}^{n} a_{ik} \tilde{a}_{kj} = \frac{1}{\det \mathbf{A}} \left(\sum_{k=1}^{n} a_{ik} A_{jk} \right).$$

If $i = j$, then $c_{ii} = \dfrac{1}{\det \mathbf{A}} \left(\sum_{k=1}^{n} a_{ik} A_{ik} \right)$ and since $\sum_{k=1}^{n} a_{ik} A_{ik} = \det \mathbf{A}$ (Th. 2.4.19 - the expansion along the i-th row), we obtain $c_{ii} = 1$ $(i \leq n)$.

If $i \neq j$, then the formula $\sum_{k=1}^{n} a_{ik} A_{jk}$ can be viewed (Th. 2.4.19) as the expansion of the determinant of a matrix in which the i-th and j-th row coincide! This is the point –

Prop. 2.4.10 then gives $\sum_{k=1}^{n} a_{ik}A_{jk} = 0$. It follows that $c_{ij} = 0$ for $i \neq j$, and this completes the proof. $\square$

In computing $\mathbf{A}^{-1}$ the above described way, we must keep in mind that the formula for $\mathbf{A}^{-1}$ involves *the transposes of* A_{ij} $(i, j \leq n)$.

2.2.23. Example. Compute A^{-1}, where

$$\mathbf{A} = \begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix} \qquad \text{(compare with Ex. 2.3.9).}$$

SOLUTION: Obviously, $\det \mathbf{A} = \frac{1}{3}, A_{11} = 1, A_{21} = -2, A_{12} = 1$ and $A_{22} = 1$. Thus, $\mathbf{A}^{-1} = \frac{1}{3} \begin{Vmatrix} 1, & -2 \\ 1, & 1 \end{Vmatrix}$.

2.2.24. Example. Compute $\mathbf{A}^{1}$, where $\mathbf{A} = \begin{Vmatrix} 0, & 1, & 2 \\ 1, & -1, & 1 \\ 2, & 0, & 1 \end{Vmatrix}$ (compare with Ex. 2.3.10).

SOLUTION: $\det \mathbf{A} = 0 \cdot (-1)^{1+1} \cdot \det \begin{Vmatrix} -1, & 1 \\ 0, & 1 \end{Vmatrix} + 1 \cdot (-1)^{1+2} \cdot \det \begin{Vmatrix} 1, & 1 \\ 2, & 1 \end{Vmatrix} + 2 \cdot (-1)^{1+3} \cdot \det \begin{Vmatrix} 1, & -1 \\ 2, & 0 \end{Vmatrix} = 0 + 1 + 4 = 5$. Further,

$$\begin{aligned}
A_{11} &= (-1)^{1+1} \cdot \det \begin{Vmatrix} -1, & 1 \\ 0, & 1 \end{Vmatrix} = -1, \\
A_{21} &= (-1)^{1+2} \cdot \det \begin{Vmatrix} 1, & 2 \\ 0, & 1 \end{Vmatrix} = -1, \\
A_{31} &= (-1)^{1+3} \cdot \det \begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix} = 3, \\
A_{12} &= (-1)^{2+1} \cdot \det \begin{Vmatrix} 1, & 1 \\ 2, & 1 \end{Vmatrix} = 1, \\
&\vdots \\
A_{23} &= (-1)^{2+3} \cdot \det \begin{Vmatrix} 0, & 1 \\ 2, & 0 \end{Vmatrix} = 2, \\
A_{33} &= (-1)^{3+3} \cdot \det \begin{Vmatrix} 0, & 1 \\ 1, & -1 \end{Vmatrix} = -1.
\end{aligned}$$

Thus, $\mathbf{A}^{-1} = \dfrac{1}{5} \begin{Vmatrix} -1, & -1, & 3 \\ 1, & -4, & 2 \\ 2, & 2, & -1 \end{Vmatrix}$.

Another important consequence of Th. 2.4.19 is the famous Cramer's rule. It is based on the following observation which is interesting in its own right. It says, phrased in a simplified form, that if we can construct inverse matrices, we can solve systems of linear equations whose matrices are regular. The systems will be addressed in the next section. Here we only outline the idea which has to be acquired before Cramer's rule is proved.

Suppose that we want to solve a system of n equations with n unknowns, say a system such as follows:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

This system can be rewritten in the matrix form: $\mathbf{A}\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} = \begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}$, where $\mathbf{A} = \|a_{ij}\|$ $(i, j \leq n)$ and where the left-hand side is understood as a matrix product. Since $\mathbf{A}$ is supposed to be regular, $\mathbf{A}^{-1}$ exists. Multiplying with $\mathbf{A}^{-1}$ from the left, we obtain $\mathbf{A}^{-1}\left(\mathbf{A}\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}\right) = \mathbf{A}^{-1}\begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}$. But matrix multiplication is associative (Prop. 2.1.6 (ii)) and therefore we are allowed to shift the brackets:

$$\mathbf{A}^{-1}\left(\mathbf{A}\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}\right) = (\mathbf{A}^{-1}\mathbf{A})\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}.$$

However, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$ and therefore $\mathbf{E}\begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix} = \mathbf{A}^{-1}\begin{Vmatrix} b_1 \\ \vdots \\ b_n \end{Vmatrix}$. This finally reads

$$\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} = \mathbf{A}^{-1}\begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}.$$

2.2.25. Example. Solve the system

$$\begin{array}{rcrcrcr} 3x_1 & - & x_2 & + & 4x_3 & = & 1 \\ x_1 & + & 2x_2 & + & x_3 & = & 3 \\ 5x_1 & - & x_2 & - & 2x_3 & = & 11 \end{array}$$

SOLUTION: The matrix form of our system is

$$\begin{Vmatrix} 3, & -1, & 4 \\ 1, & 2, & 1 \\ 5, & -1, & -2 \end{Vmatrix} \begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} 1 \\ 3 \\ 11 \end{Vmatrix}.$$

Let us set $\mathbf{A} = \begin{Vmatrix} 3, & -1, & 4 \\ 1, & 2, & 1 \\ 5, & -1, & -2 \end{Vmatrix}$. Then $\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \mathbf{A}^{-1} \begin{Vmatrix} 1 \\ 3 \\ 11 \end{Vmatrix}$. Let us compute $\mathbf{A}^{-1}$. Using the determinant method, or the elimination method if we prefer, we obtain

$$\mathbf{A}^{-1} = \frac{1}{-60} \begin{Vmatrix} -3, & -6, & -9 \\ 7, & -26, & 1 \\ -11, & -2, & 7 \end{Vmatrix} \qquad (\det \mathbf{A} = -60).$$

It follows that

$$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \frac{1}{-60} \begin{Vmatrix} -3, & -6, & -9 \\ 7, & -26, & 1 \\ -11, & -2, & 7 \end{Vmatrix} \begin{Vmatrix} 1 \\ 3 \\ 11 \end{Vmatrix} = \frac{1}{-60} \begin{Vmatrix} -3 & - & 18 & - & 99 \\ 7 & - & 78 & + & 11 \\ -11 & - & 6 & + & 77 \end{Vmatrix} =$$

$$= \frac{1}{-60} \begin{Vmatrix} -120 \\ -\ 60 \\ 60 \end{Vmatrix} = \begin{Vmatrix} 2 \\ 1 \\ -1 \end{Vmatrix}.$$

Thus, $x_1 = 2, x_2 = 1$ and $x_3 = -1$.

2.2.26. Theorem(Cramer's rule): Let us consider the following system of n linear equations with n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

Let us put $\mathbf{A} = \|a_{ij}\|$ $(i, j \leq n)$ and let us assume that $\mathbf{A}$ is regular. Let us denote by $\mathbf{C}_i$ $(i \leq n)$ the matrix which we obtain when we replace the i-th column of $\mathbf{A}$ with the right-hand side column $\begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}$. Then the (only) solution $\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}$ of the system is given by the following formula:

$$x_i = \frac{\det \mathbf{C}_i}{\det \mathbf{A}} \quad (i = 1, 2, \ldots, n).$$

Proof: Copying the reasoning employed in the considerations preceding this theorem, we first rewrite the system in the matrix form, $\mathbf{A}\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} = \begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}$, and then separate the unknowns, $\begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} = \mathbf{A}^{-1}\begin{Vmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{Vmatrix}$. Since $\mathbf{A}^{-1} = \|\tilde{a}_{ij}\|$, where $\tilde{a}_{ij} = \dfrac{A_{ji}}{\det A}$ (Th. 2.4.22), we see that the matrix product reads, for a fixed i, $i \le n$,

$$x_i = \sum_{j=1}^{n} \tilde{a}_{ij} b_j = \sum_{j=1}^{n} \frac{A_{ji}}{\det A} \cdot b_j = \frac{1}{\det A}\left(\sum_{j=1}^{n} b_j A_{ji}\right).$$

But the last formula $\sum_{j=1}^{n} b_j A_{ji}$ is the expansion of $\det \mathbf{C}_i$ down the i-th column. Thus,

$$x_i = \frac{\det \mathbf{C}_i}{\det \mathbf{A}}.$$

□

2.2.27. Example. Solve the system

$$\begin{array}{rcrcrcr} 3x_1 & - & x_2 & + & 4x_3 & = & 1 \\ x_1 & + & 2x_2 & + & x_3 & = & 3 \\ 5x_1 & - & x_2 & - & 2x_3 & = & 11 \end{array} \qquad \text{(compare with Example 2.4.25)}$$

Solution: Observing that $\det \mathbf{A} \neq 0$, we apply Cramer's rule and obtain

$$x_1 = \frac{\det \mathbf{C}_1}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 1, & -1, & 4 \\ 3, & 2, & 1 \\ 11, & -1, & -2 \end{Vmatrix}}{-60} = \frac{-120}{-60} = 2,$$

$$x_2 = \frac{\det \mathbf{C}_2}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 3, & 1, & 4 \\ 1, & 3, & 1 \\ 5, & 11, & -2 \end{Vmatrix}}{-60} = \frac{-60}{-60} = 1,$$

$$x_3 = \frac{\det \mathbf{C}_3}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 3, & -1, & 1 \\ 1, & 2, & 3 \\ 5, & -1, & 11 \end{Vmatrix}}{-60} = \frac{60}{-60} = -1.$$

It should be noted that Cramer's rule applies even in solving systems with many more unknowns, where this technique is generally less handy, provided one is interested in only a few unknowns. This situation often occurs in the technical and economics sciences.

Let us finish this paragraph with a deeper theorem on determinants.

2.2.28. Theorem(Laplace's theorem): Suppose that $\mathbf{A}, \mathbf{B} \in \mathscr{M}$ ($\in \mathscr{N}$). Then

$$\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}.$$

A corollary:

(i) $\det \mathbf{A}^{-1} = \dfrac{1}{\det \mathbf{A}}$,

(ii) if $\mathbf{A}, \mathbf{B}$ are both regular, then so is $\mathbf{AB}$.

PROOF: If $\mathbf{A}$ is singular, then $\det \mathbf{A} = 0$ (Th. 2.4.15). Further, $\mathbf{AB}$ is also singular (Th. 2.2.10). It follows that $\det(\mathbf{AB}) = 0$ (Th. 2.4.15) and therefore $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$.

Suppose that $\mathbf{A}$ is regular. Then $\mathbf{A} = \mathbf{U}_k \mathbf{U}_{k-1} \dots \mathbf{U}_2 \cdot \mathbf{U}_1$, where $\mathbf{U}_i$ $(i \le k)$ are suitable matrices of elimination. Using now a simple inductive argument, all we have to show is the validity of the equality $\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}$ whenever $\mathbf{A}$ is a matrix of an elimination type. This can easily be verified by a direct computation. We will review all the cases possible (illustrate it on examples!).

1. Suppose that $\mathbf{A} = \mathbf{U}$, where $\mathbf{U}$ models the interchange of two rows of $\mathbf{B}$. Then a straightforward computation gives $\det \mathbf{U} = -1$. Thus, according to Prop. 2.4.6, $\det(\mathbf{UB}) = -\det \mathbf{B} = \det \mathbf{U} \cdot \det \mathbf{B}$.
2. Suppose that $\mathbf{A} = \mathbf{U}$, where $\mathbf{U}$ models the multiplication of a row of $\mathbf{B}$ by a scalar λ $(\lambda \in \mathbf{R})$. Then a straightforward computation gives $\det \mathbf{U} = \lambda$. Thus, according to Prop. 2.4.9, $\det(\mathbf{UR}) = \lambda \cdot \det \mathbf{B} = \det \mathbf{U} \cdot \det \mathbf{B}$.
3. Suppose that $\mathbf{A} = \mathbf{U}$, where $\mathbf{U}$ models the addition of a linear combination of rows to a given row. Then a straightforward computation gives $\det \mathbf{U} = 1$. Thus, according to Prop. 2.4.11, $\det(\mathbf{UB}) = \det \mathbf{B} = \det \mathbf{U} \cdot \det \mathbf{B}$.

□

2.3. Systems of linear equations

In this section we will study systems of linear equations. They arise quite frequently in the mathematical formulations of engineering problems. We will see that the results of previous investigations will be used in many places. In particular, the theory of linear space developed in the first chapter will be heavily applied.

Let us consider the following system of equations (m equations with n unknowns):

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

We will refer to $(\mathscr{S})$ when considering this system. The matrix

$$\mathbf{A} = \left\| \begin{array}{llll} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn} \end{array} \right\|$$

will be referred to as *the matrix of the system* $(\mathscr{S})$. The matrix of the system therefore does *not* depend upon the right-hand side. If we join to $\mathbf{A}$ the column of the right-hand side, obtaining thus a new matrix $\mathbf{A}'$,

$$\mathbf{A} = \left\| \begin{array}{lllll} a_{11}, & a_{12}, & \ldots, & a_{1n}, & b_1 \\ a_{21}, & a_{22}, & \ldots, & a_{2n}, & b_2 \\ \vdots & & & & \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn}, & b_m \end{array} \right\|,$$

we refer to $\mathbf{A}'$ as *the augmented matrix of* $(\mathscr{S})$.

Of course, a system may or may not have a solution. It is obviously desirable to characterize the situations in which a system does have a solution. The following result brings such a characterization. The relative transparency of the result is achieved by making use of matrix-theoretic language.

2.3.1. Theorem(Frobenius's theorem): The system $(\mathscr{S})$ possesses *at least* one solution if and only if the rank of the matrix of $(\mathscr{S})$ equals the rank of the augmented matrix of $(\mathscr{S})$. In other words, $(\mathscr{S})$ possesses a solution $\Leftrightarrow r(\mathbf{A}) = r(\mathbf{A}')$. A corollary: If $\mathbf{A} \in \mathscr{M}$ and if $r(\mathbf{A}) = m$, then $(\mathscr{S})$ has a solution.

PROOF: Assume first that $(\mathscr{S})$ possesses a solution, some $x_1, x_2, \ldots, x_n$. This means that

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

Alternately, the vector of the right-hand side is a linear combination of the column vectors of $\mathbf{A}$. It follows, in view of Th. 2.2.3, that $r(\mathbf{A}) = r(\mathbf{A}')$.

Conversely, assume that $r(\mathbf{A}) = r(\mathbf{A}')$. Assume further that $r(\mathbf{A}) = r(\mathbf{A}') = p$ $(p \leq m)$. Then the largest collection of LI columns of $\mathbf{A}$ consists of p vectors. Without loss of generality, let us suppose that this collection consists of the first p columns. Since $r(\mathbf{A}) = r(\mathbf{A}')$, the column vectors

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{mp} \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

must be LD in R^m. Thus, the above vectors allow for a nontrivial linear combination that equals the zero vector. Thus,

$$\lambda_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + \lambda_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + \lambda_p \begin{pmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{mp} \end{pmatrix} + \lambda_{p+1} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where $\lambda_j \neq 0$ for some j $(j \leq p+1)$. If $\lambda_{p+1} = 0$, we obtain a contradiction with the LI of the first p vectors. In effect, $\lambda_{p+1} \neq 0$, and we infer that

$$\left(-\frac{\lambda_1}{\lambda_{p+1}}\right) \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + \left(-\frac{\lambda_2}{\lambda_{p+1}}\right) \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + \left(-\frac{\lambda_p}{\lambda_{p+1}}\right) \begin{pmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{mp} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

We see that if we put $x_1 = -\dfrac{\lambda_1}{\lambda_{p+1}}, x_2 = -\dfrac{\lambda_2}{\lambda_{p+1}}, \ldots, x_p = -\dfrac{\lambda_p}{\lambda_{p+1}}, x_{p+1} = x_{p+2} = \cdots = x_n = 0$, we have produced a solution of $(\mathscr{S})$. $\square$

2.3.2. Example. Find the parameter $\alpha (\alpha \in R)$ for which the following system has a solution:

$$\begin{array}{rcrcrcrcr} x & + & 10y & - & 13z & - & 7u & = & -9 \\ & & 7y & - & 9z & - & 3u & = & \alpha \\ -x & + & 4y & - & 5z & + & u & = & -1 \end{array}$$

SOLUTION: We can find the appropriate $\alpha (\alpha \in R)$ without Frobenius's theorem – we simply apply the elimination procedure and see when we can complete it without a contradiction. Let us use this approach first.

$$\left\| \begin{array}{rrrr|r} 1, & 10, & -13, & -7 & -9 \\ 0, & 7, & -9, & -3 & \alpha \\ -1, & 4, & -5, & 1 & -1 \end{array} \right\| \sim \left\| \begin{array}{rrrr|r} 1, & 10, & -13, & -7 & -9 \\ 0, & 7, & -9, & -3 & \alpha \\ 0, & 14, & -18, & -6 & -10 \end{array} \right\| \sim \left\| \begin{array}{rrrr|c} 1, & 10, & -13, & -7 & -9 \\ 0, & 7, & -9, & -3 & \alpha \\ 0, & 0, & 0, & 0 & -10-2\alpha \end{array} \right\|.$$

From the last system we immediately see that we avoid the contradiction only if $\alpha = -5$. This condition is also easily seen to be sufficient for the existence of a solution.

We may however proceed a more sophisticated way making explicit use of Frobenius's theorem. Observing first that $r(\mathbf{A}) = r \left\| \begin{array}{rrrr} 1, & 10, & -13, & -7 \\ 0, & 7, & -9, & -3 \\ -1, & 4, & -5, & 1 \end{array} \right\| = 2$, we look for the parameter α $(\alpha \in R)$ such that $r(\mathbf{A}') = r \left\| \begin{array}{rrrr|r} 1, & 10, & -13, & 7 & -9 \\ 0, & 7, & -9, & -3 & \alpha \\ -1, & 4, & -5, & 1 & -1 \end{array} \right\| = 2$. Since $r(\mathbf{A}') = r(\mathbf{A}'^T)$, it is obvious that not more than two columns of $\mathbf{A}'$ are LI. Thus, if we add the column of the right-hand side to *any pair* of columns of $\mathbf{A}$, we must obtain a matrix the rank of which is 2. Thus, taking e. g. the first two columns, we infer that the parameter $\alpha (\alpha \in R)$ we seek has to satisfy the condition $r \left\| \begin{array}{rrr} 1, & 10, & -9 \\ 0, & 7, & \alpha \\ -1, & 4, & -1 \end{array} \right\| = 2$. This is true exactly when $\det \left\| \begin{array}{rrr} 1, & 10, & -9 \\ 0, & 7, & \alpha \\ -1, & 4, & -1 \end{array} \right\| = 0$.

Expanding the determinant down the first column, we obtain $1(-7-4\alpha)-1(10\alpha+63)=0$, which gives us $-14\alpha-70=0$. Thus, we conclude that $\alpha=-5$ is the only choice of parameter that admits a solution. $\square$

Let us now take up the problem of finding all solutions of a given system (provided the system has any solution at all, of course). It turns out that we can thoroughly use the results on linear spaces, especially the notion of basis. The first contact with the problem is however quite discouraging when seen from the standpoint of linear algebra. Let us consider the system $(\mathscr{S})$ and let us view any solution $(x_1, x_2, \dots, x_n)$ as a vector in the linear space R^n. Let us denote by $\mathscr{R}$ the set of all solutions of $(\mathscr{S})$. Thus, $\mathscr{R} \subset \mathscr{R}$. The following question arises naturally: When is $\mathscr{R}$ a linear subspace of R^n? To our regret, it is not always the case. In fact, *the set $\mathscr{R}$ of all solutions is a linear subspace of R^n exactly when the right-hand side consists of zeros.*

Of course, the systems with general right-hand sides attract our interest most (such systems often arise from engineering problems). The set of solutions of a general system $(\mathscr{S})$ is not a linear space. Let us try to replace the right-hand side of $(\mathscr{S})$ by zeros. This new system obviously has a different set of solutions than the original system but, a piece of good news, the set of solutions of the new system will be a linear subspace of R^n. The feeling is that the set of all solutions of the new system may help in describing the set of all solutions of the original system. It is indeed so. Let us now analyze the situation in rigorous terms.

2.3.3. Definition. A system of linear equations is called *homogeneous* if the right-hand side of the system consists of zeros (i.e., $b_1 = b_2 = \cdots = b_n = 0$). A system is called *nonhomogeneous* if it is not homogeneous.

Consider now the system $(\mathscr{S})$ set up at the beginning of this paragraph. A system is called the homogeneous system associated with $(\mathscr{S}$ (or simply the system associated with $(\mathscr{S})$) if the system is obtained from $(\mathscr{S})$ by replacing all $b_i (i \leq m)$ by 0.

The following simple but very useful observation gives the justification for the notion of a homogeneous associated system. (Since we will only deal with vectors in R^n, we will again relax the correct notation $\overrightarrow{x} \sharp \overrightarrow{y}, \lambda \circ \overrightarrow{x}$ to a simplified form $\overrightarrow{x} + \overrightarrow{y}, \lambda \overrightarrow{x}$.)

2.3.4. Proposition. Consider the (generally nonhomogeneous) system $(\mathscr{S})$. If $\overrightarrow{x} = (x_1, x_2, \dots, x_n)$ and $\overrightarrow{y} = (y_1, y_2, \dots, y_n)$ are two solutions of $(\mathscr{S})$, then $\overrightarrow{z} = \overrightarrow{x} - \overrightarrow{y}$ is a solution of the homogeneous system associated with $(\mathscr{S})$. Further, if $\overrightarrow{x} = (x_1, x_2, \dots, x_n)$ is a solution of $(\mathscr{S})$ and $\overrightarrow{z} = (z_1, z_2, \dots, z_n)$ is a solution of the homogeneous system associated with $(\mathscr{S})$, then $\overrightarrow{x} + \overrightarrow{z}$ is a solution of $(\mathscr{S})$.

PROOF: A routine verification.

2.3.5. Proposition. Let $\overrightarrow{x} = (x_1, x_2, \dots, x_n)$ be a solution of $(\mathscr{S})$. Let us keep $\overrightarrow{x}$ fixed. Then for any solution $\overrightarrow{y} = (y_1, y_2, \dots, y_n)$ of $(\mathscr{S})$ we can find a solution $\overrightarrow{z} = (z_1, z_2, \dots, z_n)$ of the homogeneous system associated with $(\mathscr{S})$ such that $\overrightarrow{y} = \overrightarrow{x} + \overrightarrow{z}$.

PROOF: We put $\overrightarrow{z} = \overrightarrow{y} - \overrightarrow{x}$ and use Prop. 2.5.4.

The previous result transforms the problem of describing solutions in general to the problem of describing solutions in the case of homogeneous systems. For the homogeneous systems the following important result can be proved. Before we formulate it, let us make a convention. Let $(\mathscr{H})$ and $(\mathscr{K})$ be two homogeneous systems and let $\mathscr{R}$ and $\mathscr{R}$ be the respective spaces of their solutions. We say that $(\mathscr{H})$ *is equivalent* to $(\mathscr{K})$ if $\dim\mathscr{R} = \dim\mathscr{R}$.

2.3.6. Theorem(on the set of solutions of a homogeneous system): Consider the homogeneous system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \qquad (\mathscr{A})$$

Let $\mathbf{A} = \|a_{ij}\|, i \leq n, j \leq m$ ($\mathbf{A}$ is the matrix of the system). Then the set $\mathscr{R}$ of all solutions of this system forms a linear subspace of R^n and, moreover, $\dim\mathscr{R} = \; - (\mathbf{A})$.

Proof: In order to verify that $\mathscr{R}$ $(\mathscr{R} \subset \mathscr{R})$ is a linear subspace of R^n, we must show that

(i) if both $\overrightarrow{x}$ and $\overrightarrow{y}$ are solutions of the above system $(\mathscr{A})$, then so is $\overrightarrow{x} + \overrightarrow{y}$,

(ii) if $\overrightarrow{x}$ is a solution of the above system $(\mathscr{A})$, then so is $\lambda\overrightarrow{x}$ for any scalar $\lambda(\lambda \in R)$.

But the verification of properties (i) and (ii) is easy. Thus, the set $\mathscr{R}$ of all solutions of the system $(\mathscr{A})$ is a linear subspace of R^n.

Let us take up the dimension of $\mathscr{R}$. Suppose that $r(\mathbf{A}) = p, p \leq n$. Let us first show that in this case there is a system $(\mathscr{B})$ equivalent to $(\mathscr{A})$ such that the matrix of $(\mathscr{B})$, $\mathbf{B}$, looks as follows:

$$\mathbf{B} = \left\| \begin{matrix} b_{11}, & 0, & 0, & \dots, & 0, & 0, & b_{1,p+1}, & b_{1,p+2}, & \dots, & b_{1,n} \\ 0, & b_{22}, & 0, & \dots, & 0, & 0, & b_{2,p+1}, & b_{2,p+2}, & \dots, & b_{2,n} \\ \vdots \\ 0, & 0, & 0, & \dots, & b_{p-1,p-1}, & 0, & b_{p-1,p+1}, & b_{p-1,p+2}, & \dots, & b_{p-1,n} \\ 0, & 0, & 0, & \dots, & 0, & b_{pp}, & b_{p,p+1}, & b_{p,p+2}, & \dots, & b_{p,n} \\ 0, & 0, & 0, & \dots, & 0, & 0, & 0, & 0, & \dots, & 0 \\ 0, & 0, & 0, & \dots, & 0, & 0, & 0, & 0, & \dots, & 0 \end{matrix} \right\|$$

Formally, writing $\mathbf{B} = \|b_{ij}\|$ for the matrix of $(\mathscr{B})$, we want to show that there is a $(\mathscr{B})$ equivalent to $(\mathscr{A})$ such that

$$\begin{aligned} b_{ii} &\neq 0 && \text{whenever } i \leq p, \\ b_{ij} &= 0 && \text{whenever } i > j, \\ b_{ij} &= 0 && \text{whenever } i < j \text{ and } j \leq p, \\ b_{ij} &= 0 && \text{whenever } i > p. \end{aligned}$$

Let us see why such a system $(\mathscr{B})$ exists. Since $r(\mathbf{A}) = p$, we may apply the Gaussian elimination procedure to obtain a system, $(\mathscr{C})$, the matrix of which, $\mathbf{C} = \|c_{ij}\|$, has only the first p rows nontrivial and all components under the diagonal equal to 0 (i.e., $c_{ij} = 0$ whenever $i > p$ and $c_{ij} = 0$ whenever $i > j$). Obviously, $(\mathscr{C})$ is equivalent to $(\mathscr{A})$.

We now may (and shall) assume that $c_{ii} \neq 0$ for any $i, i \leq p$. Indeed, if it is not the case, we may exchange suitable columns of $\mathbf{C}$ to obtain a matrix $\mathbf{D}$ with $d_{ii} \neq 0$. If we denote by $(\mathscr{D})$ the system corresponding to $\mathbf{D}$, we easily see that $(\mathscr{D})$ is equivalent to $(\mathscr{C})$ – we have only renamed some of the unknowns Thus, $(\mathscr{D})$ is equivalent to $(\mathscr{A})$ and we may take $(\mathscr{D})$ for $(\mathscr{C})$.

We can now proceed analogously to the "elimination construction" of inverse matrices. Since $c_{ii} \neq 0$ for any $i \leq p$, we may go on with the elimination "from the bottom to the top". This would change $\mathbf{C}$ into $\mathbf{B}$ such that $b_{ij} = 0$ whenever $i < j$ and $j \leq p$. If $(\mathscr{B})$ is taken to be the system corresponding to $\mathbf{B}$, we see that we have established the first part of the proof of Th. 2.5.6.

Let us now complete the proof. We have found ourselves in a considerably simplified situation than at the beginning – we only have to prove Th. 2.5.6 for the system $(\mathscr{B})$. If we again denote by $(\mathscr{R})$ the space of solutions of $(\mathscr{B})$, we must show that $\dim \mathscr{R} = -(\mathbf{B})$. Thus, we must find a basis of $\mathscr{R}$ which has $n - r(\mathbf{B})$ elements. In other words, putting $s = n - r(\mathbf{B})$, we must exhibit s linearly independent solutions of $(\mathscr{B})$, some solutions $\overrightarrow{x}^1, \overrightarrow{x}^2, \ldots, \overrightarrow{x}^s$, such that $\mathscr{R} = \operatorname{Span}\{\overrightarrow{x}, \overrightarrow{x}, \ldots, \overrightarrow{x}\}$. We define $\overrightarrow{x}^i (i \leq n)$ by setting up their "tail" coordinates. We put

$$\begin{aligned}
\overrightarrow{x}^1 &= (x_1^1, x_2^1, \ldots, x_p^1, 1, 0, 0, \ldots, 0)\\
\overrightarrow{x}^2 &= (x_1^2, x_2^2, \ldots, x_p^2, 0, 1, 0, \ldots, 0)\\
\overrightarrow{x}^3 &= (x_1^3, x_2^3, \ldots, x_p^3, 0, 0, 1, \ldots, 0)\\
&\vdots\\
\overrightarrow{x}^s &= (x_1^s, x_2^s, \ldots, x_p^s, 0, 0, 0, \ldots, 1).
\end{aligned}$$

It is evident that such solutions are determined uniquely and that the vectors $\overrightarrow{x}^1, \overrightarrow{x}^2, \ldots, \overrightarrow{x}^s$ are LI in R^n. It remains to be proven that $\mathscr{R} = \operatorname{Span}\{\overrightarrow{x}, \overrightarrow{x}, \ldots, \overrightarrow{x}\}$. In other words, it remains to be proven that every solution $\overrightarrow{y} \in \mathscr{R}$ of the system $(\mathscr{B})$ can be expressed as a linear combination of the vectors $\overrightarrow{x}^1, \overrightarrow{x}^2, \ldots, \overrightarrow{x}^s$. We can now argue as follows. We can obviously find a linear combination of vectors $\overrightarrow{x}^1, \overrightarrow{x}^2, \ldots, \overrightarrow{x}^s$ which agrees with $\overrightarrow{y}$ on the last s coordinates. Indeed, if $\overrightarrow{y} = (\overrightarrow{y_1}, \overrightarrow{y_2}, \ldots, \overrightarrow{y}_p, \overrightarrow{y}_{p+1}, \ldots, \overrightarrow{y}_n)$, then this linear combination is $y_{p+1}\overrightarrow{x}^1 + y_{p+2}\overrightarrow{x}^2 + \cdots + y_n\overrightarrow{x}^s$. But since this combination must again be a solution of $(\mathscr{B})$ and since the first $n - s$ coordinates of any solution are determined *uniquely* by the last s coordinates, we see that $\overrightarrow{y} = y_{p+1}\overrightarrow{x}^1 + y_{p+2}\overrightarrow{x}^2 + \cdots + y_n\overrightarrow{x}^s$. The proof is complete. □

2.3.7. Example. Solve the system

$$\begin{array}{ccccccccccc}
2x_1 & - & 3x_2 & + & x_3 & + & x_4 & + & 3x_5 & = & 0\\
4x_1 & - & x_2 & + & 3x_3 & + & 3x_4 & + & 5x_5 & = & 0\\
2x_1 & + & 7x_2 & + & 3x_3 & + & 3x_4 & + & x_5 & = & 0\\
3x_1 & - & 2x_2 & + & 2x_3 & + & 2x_4 & + & 4x_5 & = & 0
\end{array}$$

SOLUTION: We will first follow the strategy of the proof of Th. 2.5.6. This procedure is algorithmic in its nature. It can be recommended when a personal computer is used or when the fractions which are likely to occur can easily be treated. A more "human" approach, which one uses as a rule, will be demonstrated afterwords.

Let us first embark on the algorithmic solution. We consecutively obtain

$$\left\|\begin{array}{rrrrr} 2, & -3, & 1, & 1, & 3 \\ 4, & -1, & 3, & 3, & 5 \\ 2, & 7, & 3, & 3, & 1 \\ 3, & -2, & 2, & 2, & 4 \end{array}\right\| \sim \left\|\begin{array}{rrrrr} 2, & -3, & 1, & 1, & 3 \\ 0, & 5, & 1, & 1, & -1 \\ 0, & 10, & 2, & 2, & -2 \\ 0, & 5, & 1, & 1, & -1 \end{array}\right\| \sim$$

$$\left\|\begin{array}{rrrrr} 2, & -3, & 1, & 1, & 3 \\ 0, & 5, & 1, & 1, & -1 \\ 0, & 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0 \end{array}\right\| \sim \left\|\begin{array}{rrrrr} 10, & 0, & 8, & 8, & 12 \\ 0, & 5, & 1, & 1, & -1 \end{array}\right\| \sim$$

$$\sim \left\|\begin{array}{rrrrr} 5, & 0, & 4, & 4, & 6 \\ 0, & 5, & 1, & 1, & -1 \end{array}\right\|.$$

We now set up for the three solutions that will constitute a basis for the linear space of all solutions:

$$\begin{aligned} \overrightarrow{x}^1 &= (.,.,1,0,0) \\ \overrightarrow{x}^2 &= (.,.,0,1,0) \\ \overrightarrow{x}^3 &= (.,.,0,0,1). \end{aligned}$$

Computing the first two coordinates, which are determined uniquely, we obtain

$$\begin{aligned} \overrightarrow{x}^1 &= (-\frac{4}{5}, -\frac{1}{5}, 1, 0, 0) \\ \overrightarrow{x}^2 &= (-\frac{4}{5}, -\frac{1}{5}, 0, 1, 0) \\ \overrightarrow{x}^3 &= (-\frac{6}{5}, \frac{1}{5}, 0, 0, 1). \end{aligned}$$

It follows that if $\mathscr{R}$ is the space of solutions, then $\mathscr{R} = \mathrm{Span}\{\overrightarrow{x}, \overrightarrow{x}, \overrightarrow{x}\}$. Alternately, the general solution of this system reads

$$\begin{aligned} \overrightarrow{x} &= (x_1, x_2, x_3, x_4, x_5) \\ &= \alpha(-\frac{4}{5}, -\frac{1}{5}, 1, 0, 0) + \beta(-\frac{4}{5}, -\frac{1}{5}, 0, 1, 0) + \gamma(-\frac{6}{5}, \frac{1}{5}, 0, 0, 1), \end{aligned}$$

$\alpha, \beta, \gamma \in R$. Obviously, if we find the fractions cumbersome, we may (and usually shall) come to a "fractionless" form:

$$\overrightarrow{x} = \alpha'(4, 1, -5, 0, 0) + \beta'(4, 1, 0, -5, 0) + \gamma'(6, -1, 0, 0, -5),\ \alpha', \beta', \gamma' \in R.$$

In reality we proceed as follows. After simplifying the matrix to the form

$$\left\|\begin{array}{rrrrr} 2, & -3, & 1, & 1, & 3 \\ 0, & 5, & 1, & 1, & -1 \\ 0, & 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0 \end{array}\right\|$$

which we obtain by the same way as before, we realize that $\dim \mathscr{R} = n - r(\mathbf{A}) = 5 - 2 = 3$. Thus, we just need 3 linearly independent solutions of the system to describe *any solution* of the system. These 3 linearly independent solutions can easily be guessed when proceeding "from the bottom to the top" in the simplified matrix. For instance, the vector $\overrightarrow{u} = (.,0,0,1,1)$ satisfying the second equation can be completed to a solution by setting $u_1 = -2$ (we consider the first equation). Thus, $\overrightarrow{u} = (-2,0,0,1,1)$. We proceed similarly for the other two solutions – we only have to guarantee the linear independance. We get the independance by a right choice of the last four coordinates. For instance, the vector $\overrightarrow{v} = (.,0,-1,1,0)$ will be LI with $\overrightarrow{u}$ for any choice of first coordinate. Computing the first coordinate from the first equation, we obtain $\overrightarrow{v} = (0,0,-1,1,0)$. Finally, if we set $\overrightarrow{w} = (.,1,-5,0,0)$ we see that $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}$ will always be LI. Thus, setting $w_1 = 4$, which is forced by the first equation, we obtain $\overrightarrow{w} = (4,1,-5,0,0)$. We have constructed 3 linearly independent solutions $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}$ which we needed. Thus, a general solution reads

$$\overrightarrow{x} = (x_1,x_2,x_3,x_4,x_5) = a(-2,0,0,1,1) + b(0,0,-1,1,0) + c(4,1,-5,0,0),$$

$$a,b,c \in R.$$

Let us now combine Prop. 2.5.5 and Th. 2.5.6 to formulate the main theorem on solutions of systems.

2.3.8. Theorem(on the set of solution of a nonhomogeneous system): Consider the following system of linear equations (m equations in n unknowns):

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

Let $\mathbf{A} = \|a_{ij}\|$ be the matrix of the system and let $r(\mathbf{A}) = p$. Let $\overrightarrow{x} = (x_1,x_2,\ldots,x_n)$ be a solution of ($\mathscr{S}$) ("a particular solution") and let $\{\overrightarrow{u}_1, \overrightarrow{u}_2, \ldots, \overrightarrow{u}_{n-p}\}$ be a basis of the linear space of solutions of the homogeneous system associated with ($\mathscr{S}$). Then any solution of ($\mathscr{S}$) can be expressed in the form

$$\overrightarrow{x} + \alpha_1 \overrightarrow{u}_1 + \alpha_2 \overrightarrow{u}_2 + \cdots + \alpha_{n-p} \overrightarrow{u}_{n-p}$$

for some coefficients $\alpha_1, \alpha_2, \ldots, \alpha_{n-p}$.

PROOF: The proof immediately follows from Prop. 2.5.5 and Th. 2.5.6.

2.3.9. Example. Solve the system

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

SOLUTION: Let us first manipulate the augmented matrix of the system:

$$\left\| \begin{array}{rrrr|r} 1, & -2, & 1, & -1 & 1 \\ 1, & 1, & -1, & 1 & 2 \\ 2, & -1, & 1, & -1 & 1 \end{array} \right\| \sim \left\| \begin{array}{rrrr|r} 1, & -2, & 1, & -1 & 1 \\ 0, & 3, & -2, & 2 & 1 \\ 0, & 3, & -1, & 1 & -1 \end{array} \right\| \sim$$

$$\sim \quad \left\| \begin{array}{rrrr|r} 1, & -2, & 1, & -1 & 1 \\ 0, & 3, & -2, & 2 & 1 \\ 0, & 0, & 1, & -1 & -2 \end{array} \right\| \qquad \left(\text{assoc. hom.} \quad \begin{array}{rcr} \cancel{1} & \longrightarrow & 0 \\ \cancel{1} & \longrightarrow & 0 \\ -\cancel{2} & \longrightarrow & 0 \end{array} \right]$$

We see that $n - r(\mathbf{A}) = 4 - 3 = 1$. According to the previous theorem, we need produce one solution of this system plus one nontrivial solution of the associated homogeneous system. We easily guess these solutions by proceeding "from the bottom to the top":

The particular solution:

$$\begin{aligned} \overrightarrow{x} &= (.\,,\ .\,, 0, 2), \\ \overrightarrow{x} &= (.\,, -1, 0, 2), \\ \overrightarrow{x} &= (1, -1, 0, 2). \end{aligned}$$

The solution of the associated system (we replace the right-hand side by zeros):

$$\begin{aligned} \overrightarrow{u} &= (.\,, .\,, 1, 1), \\ \overrightarrow{u} &= (.\,, 0, 1, 1), \\ \overrightarrow{u} &= (0, 0, 1, 1). \end{aligned}$$

The solution then reads

$$\overrightarrow{x} = (x_1, x_2, x_3, x_4) = (1, -1, 0, 2) + \alpha(0, 0, 1, 1), \ \alpha \in R.$$

Expressed in coordinates,

$$\begin{aligned} x_1 &= 1 \\ x_2 &= -1 \\ x_3 &= \alpha \\ x_4 &= 2 + \alpha, \qquad\qquad \alpha \in R. \end{aligned}$$

2.3.10. Remark. We may remember the method of solving systems we were taught in high school – the method of introducing parameters. The system of Example 2.5.9 could be re-solved as follows. After simplifying the system by Gaussian elimination the way we did above and after obtaining the equivalent "triangular" system,

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

we substitute parameters. In this case we would need only one parameter. Thus, we set $x_4 = \alpha$ and obtain

$$\begin{aligned} x_3 &= -2 + \alpha \\ x_2 &= \frac{1}{3}(1 + 2x_3 - 2x_4) \\ &= \frac{1}{3} + \frac{1}{3}[2(-2+\alpha) - 2\alpha] = \frac{1}{3} - \frac{4}{3} \\ &= -1 \\ x_1 &= 1 + 2x_2 - x_3 + x_4 = 1 + 2\cdot(-1) - (-2+\alpha) + \alpha \\ &= 1 \end{aligned}$$

When we express the latter result in the vector form, we obtain

$$\overrightarrow{x} = (x_1, x_2, x_3, x_4) = (1, -1, -2, 0) + \alpha(0, 0, 1, 1),\ \ \alpha \in R.$$

The above procedure of introducing parameters can also be applied in general. A suspicion may arise: Do we not prefer our way just because it is "more scientific"? Not at all. There are at least four reasons why we advocate the approach based on Th. 2.5.8. First, it is usually handier and shorter. Secondly, the geometric interpretation of the solution suggests itself – the set of all solutions can be viewed as a linear variety in E^n such that the particular solution plays the rôle of origin and the basis of the associated homogeneous system determines the "size" of the variety (see also Chap. 4). Thirdly, the idea of describing all solutions by finding a particular solution and a basis for the homogeneous space repeats itself in a more complex form in many other areas of engineering mathematics, notably in differential equations and the Laplace transform. It is enormously important that this idea be properly acquired in the study of systems of linear equations. And fourthly, the method of introducing parameters is more appropriate for the student who wants to have this method the ultimate of his algebraic knowledge ... .

Let us finish this chapter by answering the following natural question: When does a system of n linear equations with n unknowns possess *exactly one solution*?

2.3.11. Theorem Let us consider the system of n linear equations with n unknowns. Let $\mathbf{A} \in M^{nn}$ be the matrix of this system. Then the system has exactly one solution if and only if $\det \mathbf{A} \neq 0$.

PROOF: By Th. 2.4.15, $\det \mathbf{A} \neq 0$ if and only if $\mathbf{A}$ is regular. We therefore want to show that the system has exactly one solution if and only if $\mathbf{A}$ is regular. We will now establish this equivalence.

Suppose first that $\mathbf{A}$ is regular. Then the solution is given by Cramer's rule and is therefore unique. (Here is another proof. Since $r(\mathbf{A}) = n$, we see that $r(\mathbf{A}') = n$, where $\mathbf{A}'$ stands for the augmented matrix; $r(\mathbf{A}') = n$ because $\mathbf{A}'$ has only n rows. By Frobenius's theorem, the system has a solution. Since $n - r(\mathbf{A}) = 0$, we see by Th. 2.5.8 that the system cannot have more than one solution.)

Suppose that the system has exactly one solution. Then $n - r(\mathbf{A}) = 0$ (otherwise there is an infinite set of solutions by Th. 2.5.8). Thus, $r(\mathbf{A}) = n$ and therefore $\mathbf{A}$ is regular.

Miscellaneous examples

2.3.12. Example. Solve the following (homogeneous) system of linear equations:

$$\begin{array}{rcrcrcrcl} 2x & + & 3y & - & z & - & u & = & 0 \\ x & - & 2y & + & 3z & + & 2u & = & 0 \\ 7x & & & + & 7z & + & 4u & = & 0 \\ 5x & + & 11y & - & 6z & - & 5u & = & 0 \end{array}$$

Solution: By the verb "solve" we mean "find all solutions", of course. We will apply the matrix form of Gaussian elimination, obtaining

$$\left\|\begin{array}{rrrr|r} 2, & 3, & -1, & -1 & 0 \\ 1, & -2, & 3, & 2 & 0 \\ 7, & 0, & 7, & 4 & 0 \\ 5, & 11, & -6, & -5 & 0 \end{array}\right\| \sim \left\|\begin{array}{rrrr|r} 1, & -2, & 3, & 2 & 0 \\ 2, & 3, & -1, & -1 & 0 \\ 7, & 0, & 7, & 4 & 0 \\ 5, & 11, & -6, & -5 & 0 \end{array}\right\| \sim$$

$$\sim \left\|\begin{array}{rrrr|r} 1, & -2, & 3, & 2 & 0 \\ 0, & 7, & -7, & -5 & 0 \\ 0, & 14, & -14, & -10 & 0 \\ 0, & 21, & -21, & -15 & 0 \end{array}\right\| \sim \left\|\begin{array}{rrrr|r} 1, & -2, & 3, & 2 & 0 \\ 0, & 7, & -7, & -5 & 0 \\ 0, & 0, & 0, & 0 & 0 \\ 0, & 0, & 0, & 0 & 0 \end{array}\right\|$$

Since the rank of the matrix of the system is 2, $r(\mathbf{A}) = 2$, and since the number of unknowns is 4, we see that the dimension of the space of all solutions is 2 (Th. 2.5.6). So all we need to obtain are two LI solutions of the system. These can easily be found from the system that is the output of the previous elimination. This system is

$$\begin{aligned} x - 2y + 3z + 2u &= 0 \\ 7y - 7z - 5u &= 0. \end{aligned}$$

Thus, starting with the second equation, we can easily guess the two solutions we need. For instance, we may put

$$\begin{aligned} \overrightarrow{v} &= (x, y, z, u) = (-1, 1, 1, 0) \\ \overrightarrow{w} &= (x, y, z, u) = (-1, 0, 5, -7) \end{aligned}$$

Summary: Every solution $\overrightarrow{r} = (x, y, z, u)$ of the system can be expressed as a linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$. Thus, $\overrightarrow{r} = \alpha(-1, 1, 1, 0) + \beta(-1, 0, 5, -7)$, $\alpha, \beta \in R$.

2.3.13. Example. Given the matrix $\mathbf{A} = \left\|\begin{array}{rr} 1, & 2 \\ -1, & 1 \end{array}\right\|$, convince yourself that the set $S = \{\mathbf{B} \in M^{22} | \mathbf{AB} = \mathbf{BA}\}$ is a linear subspace of M^{22}. Find a basis of S. Is the set $\{\mathbf{A}, \mathbf{A}^2\}$ also a basis of S?

Solution: Obviously, $S \neq \emptyset$ $(\mathbf{E} \in S)$. We have to show that the set S is closed under the formation of the "inner" operation and the "outer" multiplication with scalars. In other words, we have to show that if $\mathbf{B}_1, \mathbf{B}_2 \in S$ then $\mathbf{B}_1 + \mathbf{B}_2 \in S$ and $\lambda\mathbf{B}_1 \in S$ $(\lambda \in R)$. Let us verify these conditions. If $\mathbf{B}_1, \mathbf{B}_2 \in S$, then $\mathbf{AB}_1 = \mathbf{B}_1\mathbf{A},\ \mathbf{AB}_2 = \mathbf{B}_2\mathbf{A}$. Adding these equalities, we obtain $\mathbf{AB}_1 + \mathbf{AB}_2 = \mathbf{B}_1\mathbf{A} + \mathbf{B}_2\mathbf{A}$. Since matrix multiplication is distributive (Prop. 2.1.6 iii, iv), we see that $\mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2) = (\mathbf{B}_1 + \mathbf{B}_2)\mathbf{A}$. Thus, $\mathbf{B}_1 + \mathbf{B}_2 \in S$. We proceed analogously with the scalar multiple - if $\mathbf{B}_1 \in S$ then $\mathbf{AB}_1 = \mathbf{B}_1\mathbf{A}$ and therefore $\mathbf{A}(\lambda\mathbf{B}_1) = (\lambda\mathbf{B}_1)\mathbf{A}$ (Prop. 2.1.6 i). We have verified that S is a linear subspace of M^{22}.

We now want to find a basis of S. Let, for a general matrix $\mathbf{B} \in M^{22},\ \mathbf{B} = \begin{Vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{Vmatrix}$. Since $(\mathbf{B} \in S \Leftrightarrow \mathbf{AB} = \mathbf{BA})$, we infer that $\mathbf{B} \in S$ if and only if

$$\begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix} \begin{Vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{Vmatrix} = \begin{Vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{Vmatrix} \begin{Vmatrix} 1, & 2 \\ -1, & 1 \end{Vmatrix}.$$

Multiplying the matrix, we further obtain

$$\begin{aligned} b_{11} + 2b_{21} &= b_{11} - b_{12} \\ b_{12} + 2b_{22} &= 2b_{11} + b_{12} \\ -b_{11} + b_{21} &= b_{21} - b_{22} \\ -b_{12} + b_{22} &= 2b_{21} + b_{22}. \end{aligned}$$

We will first rearrange this system,

$$\begin{array}{rrrrrrrr} & & b_{12} & + & 2b_{21} & & & = & 0 \\ b_{11} & & & & & - & b_{22} & = & 0 \\ -b_{11} & & & & & + & b_{22} & = & 0 \\ & & -b_{12} & - & 2b_{21} & & & = & 0 \end{array},$$

finding out that the system reduces to

$$\begin{array}{rrrrrrrr} b_{11} & & & & & - & b_{22} & = & 0 \\ & & b_{12} & + & 2b_{21} & & & = & 0 \end{array}.$$

We therefore need only two LI solutions, which we can guess:

$$\begin{aligned} \mathbf{B}_1 &= \begin{Vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{Vmatrix} = \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix} \\ \mathbf{B}_2 &= \begin{Vmatrix} b_{11}, & b_{12} \\ b_{21}, & b_{22} \end{Vmatrix} = \begin{Vmatrix} 0, & -2 \\ 1, & 0 \end{Vmatrix}. \end{aligned}$$

Thus, the set $\left\{ \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix}, \begin{Vmatrix} 0, & -2 \\ 1, & 0 \end{Vmatrix} \right\}$ constitutes a basis of S. As a consequence, $\dim S = 2$. (The last question has the answer "yes", because $\mathbf{A}, \mathbf{A}^2$ are LI elements of S.)

2.3.14. Example. Solve the following system as a system consisting of one equation for five variables:

$$x_2 - x_4 = 1.$$

SOLUTION: Obviously, $r(\mathbf{A}) = 1$. According to Th. 2.5.8, we need one (quite arbitrary) solution of the system plus four LI solutions of the associated homogeneous system. These can easily be guessed. For instance, a "particular" solution can be $\overrightarrow{v} = (0,1,0,0,0)$ and the required solutions of the homogeneous system can be $\overrightarrow{v}_1 = (1,0,0,0,0), \overrightarrow{v}_2 = (0,0,1,0,0), \overrightarrow{v}_3 = (0,0,0,0,1), \overrightarrow{v}_4 = (0,1,0,1,0)$. Thus, all solutions of this system read $(0,1,0,0,0) + \alpha(1,0,0,0,0) + \beta(0,0,1,0,0) + \gamma(0,0,0,0,1) + \delta(0,1,0,1,0)$.

2.3.15. Example. Solve the following system

$$\begin{array}{rcrcrcr} x & + & 7y & - & 3z & = & 16 \\ 3x & - & 4y & + & 2z & = & 3 \\ 13x & - & 34y & + & 16z & = & 10 \end{array}$$

How does the solution comply with Frobenius' theorem?

SOLUTION:

$$\left\| \begin{array}{rrr|r} 1, & 7, & -3 & 16 \\ 3, & -4, & 2 & 3 \\ 13, & -34, & 16 & 10 \end{array} \right\| \sim \left\| \begin{array}{rrr|r} 1, & 7, & -3 & 16 \\ 0, & -25, & 11 & -45 \\ 0, & -125, & 55 & -198 \end{array} \right\| \sim \left\| \begin{array}{rrr|r} 1, & 7, & -3 & 16 \\ 0, & -25, & 11 & -45 \\ 0, & 0, & 0 & 27 \end{array} \right\|$$

This system does not have any solution - we would get a contradiction at the third equation. In accord with Frobenius' theorem, $r(\mathbf{A}) = 2 < r(\mathbf{A}') = 3$.

2.3.16. Example. Solve the system

$$\begin{array}{rcrcrcrcrcr} x_1 & + & x_2 & + & x_3 & + & x_4 & + & x_5 & = & 7 \\ 3x_1 & + & 2x_2 & + & x_3 & + & x_4 & - & 3x_5 & = & -2 \\ & & x_2 & + & 2x_3 & + & 2x_4 & + & 6x_5 & = & 23 \\ 5x_1 & + & 4x_2 & + & 3x_3 & + & 3x_4 & - & x_5 & = & 12 \end{array}$$

SOLUTION:

$$\left\| \begin{array}{rrrrr|r} 1, & 1, & 1, & 1, & 1 & 7 \\ 3, & 2, & 1, & 1, & -3 & -2 \\ 0, & 1, & 2, & 2, & 6 & 23 \\ 5, & 4, & 3, & 3, & -1 & 12 \end{array} \right\| \sim \left\| \begin{array}{rrrrr|r} 1, & 1, & 1, & 1, & 1 & 7 \\ 0, & -1, & -2, & -2, & -6 & -23 \\ 0, & 1, & 2, & 2, & 6 & 23 \\ 0, & -1, & -2, & -2, & -6 & -23 \end{array} \right\| \sim \left\| \begin{array}{rrrrr|r} 1, & 1, & 1, & 1, & 1 & 7 \\ 0, & 1, & 2, & 2, & 6 & 23 \\ 0, & 0, & 0, & 0, & 0 & 0 \\ 0, & 0, & 0, & 0, & 0 & 0 \end{array} \right\|$$

The solution then reads, after guessing the particular solution and 3 linearly independent solutions of the associated homogeneous system,

$$(x_1, x_2, x_3, x_4, x_5) = (4, -1, 0, 0, 4) + \alpha(5, -6, 0, 0, 1) + \beta(1, -2, 0, 1, 0) + \gamma(1, -2, 1, 0, 0),$$

$$\alpha, \beta, \gamma \in R.$$

Let us pause for a minute at this example. Let us suppose that a teacher has taken this example for the written part of an examination, knowing the correct answer displayed above. Let us suppose that a student came up with the following answer:

$$(x_1, x_2, x_3, x_4, x_5) = (-16, 23, 0, 0, 0) + p(6, -8, 0, 1, 1) + q(2, -4, 1, 1, 0) + r(6, -8, 1, 0, 1),$$

$$p, q, r \in R.$$

How can the teacher verify whether the answer is right? Consulting Th. 2.5.8 again, all that the teacher has to check is

(i) $(-16, 23, 0, 0, 0)$ is a solution (or, alternately, $(4, -1, 0, 0, 4) - (-16, 23, 0, 0, 0)$ is a solution of the associated homogeneous system which might be easier to find out),

(ii) $\mathrm{Span}\{(5, -6, 0, 0, 1), (1, -2, 0, 1, 0), (1, -2, 1, 0, 0)\} =$
$= \mathrm{Span}\{(6, -8, 0, 1, 1), (2, -4, 1, 1, 0), (6, -8, 1, 0, 1)\}$.

Question (i) can be answered easily by direct substitution. Question (ii) can be approached as follows. Since the vectors of the student are immediately seen to be LI, for the "tails" $(0, 1, 1), (1, 1, 0), (1, 0, 1)$ are LI, we have to show that the vectors of the student span a linear space which is not larger than the linear space spanned by the original vectors. The elimination technique helps us again (to be precise, we use Prop. 1.1.30) – if we write the second triple of vectors under the first one and work with rows, the second triple has to disappear in the course of elimination (think about a rigorous argument for this!):

$$\begin{array}{lllll} (5, & -6, & 0, & 0, & 1) \\ (1, & -2, & 0, & 1, & 0) \\ (1, & -2, & 1, & 0, & 0) \\ & & \dots & & \\ (6, & -8, & 0, & 1, & 1) \\ (2, & -4, & 1, & 1, & 0) \\ (6, & -8, & 1, & 0, & 1) \end{array} \sim \begin{array}{lllll} (1, & -2, & 0, & 1, & 0) \\ (1, & -2, & 1, & 0, & 0) \\ (5, & -6, & 0, & 0, & 1) \\ & & \dots & & \\ (6, & -8, & 0, & 1, & 1) \\ (2, & -4, & 1, & 1, & 0) \\ (6, & -8, & 1, & 0, & 1) \end{array} \sim \begin{array}{lllll} (1, & -2, & 0, & 1, & 0) \\ (0, & 0, & 1, & -1, & 0) \\ (0, & 4, & 0, & -5, & 1) \\ & & \dots & & \\ (0, & 4, & 0, & -5, & 1) \\ (0, & 0, & 1, & -1, & 0) \\ (0, & 4, & 1, & -6, & 1) \end{array} \sim$$

$$\begin{array}{lllll} (1, & -2, & 0, & 1, & 0) \\ (0, & 4, & 0, & -5, & 1) \\ (0, & 0, & 1, & -1, & 0) \\ & & \dots & & \\ (0, & 4, & 0, & -5, & 1) \\ (0, & 0, & 1, & -1, & 0) \\ (0, & 4, & 1, & -6, & 1) \end{array} \sim \begin{array}{lllll} (1, & -2, & 0, & 1, & 0) \\ (0, & 4, & 0, & -5, & 1) \\ (0, & 0, & 1, & -1, & 0) \\ & & \dots & & \\ (0, & 0, & 0, & 0, & 0) \\ (0, & 0, & 1, & -1, & 0) \\ (0, & 0, & 1, & -1, & 0) \end{array} \sim \begin{array}{lllll} (1, & -2, & 0, & 1, & 0) \\ (0, & 4, & 0, & -5, & 1) \\ (0, & 0, & 1, & -1, & 0) \\ & & \dots & & \\ (0, & 0, & 0, & 0, & 0) \\ (0, & 0, & 0, & 0, & 0) \\ (0, & 0, & 0, & 0, & 0) \end{array}$$

Thus, the student's solution has been found to be correct to the satisfaction of both.

2.3.17. Example. Solve the following system

$$\begin{array}{ccccccccccc} x_1 & - & x_2 & + & 2x_3 & + & 3x_4 & - & x_5 & = & 2 \\ x_1 & & & + & x_3 & + & 5x_4 & - & 3x_5 & = & 3 \\ x_1 & & & + & x_3 & + & 5x_4 & & & = & 6 \end{array}$$

Observe that although the set of solutions is infinite, the coordinate x_5 of any solution is fixed.

SOLUTION:

$$\left\|\begin{array}{rrrrr|r} 1, & -1, & 2, & 3, & -1 & 2 \\ 1, & 0, & 1, & 5, & -3 & 3 \\ 1, & 0, & 1, & 5, & 0 & 6 \end{array}\right\| \sim \left\|\begin{array}{rrrrr|r} 1, & -1, & 2, & 3, & -1 & 2 \\ 0, & 1, & -1, & 2, & -2 & 1 \\ 0, & 1, & -1, & 2, & 1 & 4 \end{array}\right\| \sim \left\|\begin{array}{rrrrr|r} 1, & -1, & 2, & 3, & -1 & 2 \\ 0, & 1, & -1, & 2, & -2 & 1 \\ 0, & 0, & 0, & 0, & 3 & 3 \end{array}\right\|$$

We see that $x_5 = 1$. Since the number of unknowns is 5 and since $r(\mathbf{A}) = 3$, we can produce the general form of solution (as always, we make use of Th. 2.5.8):

$$(x_1, x_2, x_3, x_4, x_5) = (6,3,0,0,1) + \alpha(-5,0,2,1,0) + \beta(-1,1,1,0,0),\ \ \alpha, \beta \in R.$$

As a by-product, we see that $x_5 = 1$ for any solution of the system.

2.3.18. Example. Discuss the solution of the following system with respect to the parameters a, b:

$$\begin{array}{rcrcrcr} 2x & - & y & + & 3z & = & 1 \\ x & + & 2y & - & z & = & -b \\ x & + & ay & - & 6z & = & -10 \end{array}$$

Solution: The system may in principle have exactly one solution, no solution or an infinite set of solutions. By Th. 2.5.11, the system has exactly one solution if and only if $\det \mathbf{A} \neq 0$. Let us first examine this case. Computing the determinant of the matrix by expanding along the first row, we obtain

$$\det \begin{Vmatrix} 2, & -1, & 3 \\ 1, & 2, & -1 \\ 1, & a, & -6 \end{Vmatrix} = 2(-12+a) - 1 \cdot (-1)(-6+1) + 3(a-2) = -35 + 5a$$

It follows that $\det \mathbf{A} \neq 0 \Leftrightarrow a \neq 7$. We therefore see that the system has exactly one solution if and only if $a \neq 7$ (the fact of having exactly one solution does *not* depend on the right-hand side!). This solution can then be computed by Cramer's rule:

$$x = \frac{\det \mathbf{C}_1}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 1, & -1, & 3 \\ -b, & 2, & -1 \\ -10, & a, & -6 \end{Vmatrix}}{-35+5a} = \frac{38 + a + 6b - 3ab}{-35+5a}$$

$$y = \frac{\det \mathbf{C}_2}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 2, & 1, & 3 \\ 1, & -b, & -1 \\ 1, & -10, & -6 \end{Vmatrix}}{-35+5a} = \frac{-45+15b}{-35+5a} = \frac{-9+3b}{-7+a}$$

$$z = \frac{\det \mathbf{C}_3}{\det \mathbf{A}} = \frac{\det \begin{Vmatrix} 2, & -1, & 1 \\ 1, & 2, & -b \\ 1, & a, & -10 \end{Vmatrix}}{\det \mathbf{A}} = \frac{-52 + a + b + 2ab}{-35+5a}.$$

Suppose further that $a = 7$. In this case the system may have either no solution or an infinite set of solutions which can be described (Th. 2.5.8). Let us consider the system for $a = 7$. Putting it in the matrix form and applying elimination, we consecutively obtain

$$\left\Vert \begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 1, & 2, & -1 & -b \\ 1, & 7, & -6 & -10 \end{array} \right\Vert \sim \left\Vert \begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 0, & -5, & 5 & 2b+1 \\ 0, & -15, & 15 & 21 \end{array} \right\Vert \sim \left\Vert \begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 0, & -5, & 5 & 2b+1 \\ 0, & 0, & 0 & -6b+18 \end{array} \right\Vert$$

We see that if $a = 7$ and $b \neq 3$, then the system has no solution at all. If $a = 7$ and $b = 3$, then we have the following system to consider:

$$\left\| \begin{array}{lll|r} 2, & -1, & 3 & 1 \\ 0, & -5, & 5 & 7 \end{array} \right\|.$$

The solution is easy to describe:

$$(x, y, z) = \left(-\frac{8}{5}, 0, \frac{7}{5}\right) + \alpha(-1, 1, 1), \quad \alpha \in R.$$

We are ready to formulate the summary of this problem. Before, let us observe that we could also have had $b = 3$ (for $a = 7$) using determinants only. Indeed, Th. 2.5.1 (Frobenius's theorem) could have been applied in the following amusing way. Consider again the system arrived at for $a = 7$, $\left\| \begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 1, & 2, & -1 & -b \\ 1, & 7, & -6 & -10 \end{array} \right\|$. We want to pick the parameter $b (b \in \mathbf{R})$ for which the system has a solution. Since this is the case exactly when $r(\mathbf{A}) = r(\mathbf{A}')$, where $\mathbf{A}'$ is the augmented matrix, and since $r(\mathbf{A}) = 2$, we see that $b (b \in R)$ has to be such that the rank of the following matrix $\mathbf{B}$, $\mathbf{B} = \left\| \begin{array}{rrr} 2, & -1, & 1 \\ 1, & 2, & -b \\ 1, & 7, & -10 \end{array} \right\|$, equals 2. This is equivalent to $\det \mathbf{B} = 0$. Let us calculate $\det \mathbf{B}$. We obtain, by expanding along the first row,

$$\det \mathbf{B} = 2(-20 + 7b) - 1 \cdot (-1)(-10 + b) + 1 \cdot (7 - 2) = -45 + 15b.$$

This shows that $b = 3$ is the critical value. (Observe also that we could have substituted the right-hand side vector for any arbitrary column in $\mathbf{A}$, not only for the third one as we did in the case of our matrix $\mathbf{B}$.)

Summary: If $a \neq 7$, then the system has exactly one solution,

$$x = \frac{38 + a + 6b - 3ab}{-35 + 5a}, \quad y = \frac{-9 + 3b}{-7 + a}, \quad z = \frac{-52 + a + b + 2ab}{-35 + 5a}.$$

If $p = 7$ and $b \neq 3$, then the system does not have any solution.

If $p = 7$ and $b = 3$, then the system has an infinite set of solutions which read

$$(x, y, z) = \left(-\frac{8}{5}, 0, \frac{7}{5}\right) + \alpha(-1, 1, 1), \ \alpha \in R.$$

2.3.19. Example. Solve the matrix equation

$$\mathbf{AX} = \mathbf{C},$$

where

$$\mathbf{A} = \left\| \begin{array}{rrr} 1, & 1, & -1 \\ 1, & -4, & 2 \\ 1, & -1, & 1 \end{array} \right\|, \qquad \mathbf{C} = \left\| \begin{array}{rrr} 1, & 1, & 0 \\ 0, & -1, & 2 \\ 0, & 0, & 3 \end{array} \right\|.$$

($\mathbf{X}$ is an unknown matrix.) Use $\mathbf{A}^{-1}$.

SOLUTION: If $\mathbf{A}$ is a regular matrix, then $\mathbf{A}^{-1}$ exists and we have the following series of implications:

$$\mathbf{AX}=\mathbf{C}\Longrightarrow \mathbf{A}^{-1}(\mathbf{AX})=\mathbf{A}^{-1}\mathbf{C}\Longrightarrow \text{ the matrix product is associative }\Longrightarrow$$

$$(\mathbf{A}^{-1}\mathbf{A})\mathbf{X}=\mathbf{A}^{-1}\mathbf{C}\Longrightarrow \mathbf{EX}=\mathbf{A}^{-1}\mathbf{C}\Longrightarrow \mathbf{X}=\mathbf{A}^{-1}\mathbf{C}.$$

We therefore need $\mathbf{A}^{-1}$. After choosing a method for computing $\mathbf{A}^{-1}$ (both are about equally technical), we obtain

$$\mathbf{A}^{-1}=\frac{1}{4}\left\|\begin{matrix}2, & 0, & 2\\ -1, & -2, & 3\\ -3, & -2, & 5\end{matrix}\right\|.$$

As a result,

$$\mathbf{X}=\mathbf{A}^{-1}\mathbf{C}=\left\|\begin{matrix}2, & 0, & 2\\ -1, & -2, & 3\\ -3, & -2, & 5\end{matrix}\right\|\left\|\begin{matrix}1, & 1, & 0\\ 0, & -1, & 2\\ 0, & 0, & 3\end{matrix}\right\|=\frac{1}{4}\left\|\begin{matrix}2, & 2, & 6\\ -1, & 1, & 5\\ -3, & 1, & 11\end{matrix}\right\|.$$

Problems

(A problem indicated with an * is supposed to be a challange for an interested student.)

1. Find the transpose, $\mathbf{A}^T$, of the following matrix $\mathbf{A}$. What size is $\mathbf{A}^T$?

$$\mathbf{A}=\begin{bmatrix}3 & 1 & 2 & -3\\ 4 & 2 & 1 & 4\end{bmatrix}.$$

$$\left(\text{Answer: }\begin{bmatrix}3 & 4\\ 1 & 2\\ 2 & 1\\ -3 & 4\end{bmatrix},\ (4,2).\right)$$

2. Supply the components of $\mathbf{A}$ so that a symmetric matrix is obtained.

$$\mathbf{A}=\left\|\begin{matrix}3 & 0 & .\\ . & 1 & -4\\ 2 & . & 5\end{matrix}\right\|.$$

$$\left(\text{Answer: }\left\|\begin{matrix}3 & 0 & 2\\ 0 & 1 & -4\\ 2 & -4 & 5\end{matrix}\right\|.\right)$$

3. Solve the following matrix equation:

$$3 \circ \mathbf{A} \sharp 2 \circ \mathbf{X} = \mathbf{B}, \quad \mathbf{A} = \begin{Vmatrix} 5, & 2 \\ -8, & 1 \end{Vmatrix}, \ \mathbf{B} = \begin{Vmatrix} 3, & 4 \\ 2, & 7 \end{Vmatrix}.$$

$$\left(\text{Answer: } \mathbf{X} = \begin{Vmatrix} -6, & -1 \\ 13, & 2 \end{Vmatrix}.\right)$$

4. Find the following matrix products whenever the setup is meaningful:

$$\text{a)} \ \begin{Vmatrix} 1, & 2, & 1 \end{Vmatrix} \begin{Vmatrix} 3 \\ 2 \\ 1 \end{Vmatrix}, \quad \text{b)} \ \begin{Vmatrix} 1, & 2, & 0, & 3 \end{Vmatrix} \begin{Vmatrix} 2, & 0 \\ 0, & 3 \\ -4, & 1 \\ 1, & -4 \end{Vmatrix}, \quad \text{c)} \ \begin{Vmatrix} 1, & -1 \\ 1, & -2 \\ 1, & -3 \end{Vmatrix} \begin{Vmatrix} 0, & 1, & -1, & 0 \\ 0, & 1, & -2, & -1 \end{Vmatrix}$$

$$\text{d)} \ \begin{Vmatrix} 1, & 1, & -1 \\ 2, & 0, & 1 \\ 3, & 2, & 4 \end{Vmatrix} \begin{Vmatrix} 1, & 2, & 3 \\ -1, & 0, & 1 \end{Vmatrix} \qquad \text{e)} \ \begin{Vmatrix} 5, & 1, & -2 \\ 7, & 1, & -3 \\ 3, & 2, & 0 \end{Vmatrix} \begin{Vmatrix} 3, & 1, & 2 \\ 2, & 2, & 1 \\ 1, & 3, & 3 \end{Vmatrix}.$$

$$\left(\text{Answer: a) } \|8\|, \text{ b) } \|5, -6\|, \text{ c) } \begin{Vmatrix} 0, & 0, & 1, & 1 \\ 0, & -1, & 3, & 2 \\ 0, & -2, & 5, & 3 \end{Vmatrix}, \text{ d) meaningless, e) } \begin{Vmatrix} 15, & 1, & 5 \\ 20, & 0, & 6 \\ 13, & 7, & 8 \end{Vmatrix}.\right)$$

5. Using mathematical induction, prove the following identities:

$$\begin{Vmatrix} 1, & 1 \\ 0, & 1 \end{Vmatrix}^n = \begin{Vmatrix} 1, & n \\ 0, & 1 \end{Vmatrix}, \quad \begin{Vmatrix} \cos\varphi, & -\sin\varphi \\ \sin\varphi, & \cos\varphi \end{Vmatrix}^n = \begin{Vmatrix} \cos n\varphi, & -\sin n\varphi \\ \sin n\varphi, & \cos n\varphi \end{Vmatrix}.$$

6. Let $\mathbf{A} = \begin{Vmatrix} 7, & 5 \\ 6, & 4 \end{Vmatrix}$ and $\mathbf{B} = \begin{Vmatrix} 1, & 2 \\ 4, & 3 \end{Vmatrix}$. Let $S = \{\mathbf{C} \in \mathscr{M} | \mathbf{AC} = \mathbf{BC}\}$. Verify that S is a linear subspace of $\mathscr{M}$. Find a basis of S. What is $\dim S$?

$$\left(\text{Answer: } \left\{ \begin{Vmatrix} 1, & 0 \\ -2, & 0 \end{Vmatrix}, \begin{Vmatrix} 0, & 1 \\ 0, & -2 \end{Vmatrix} \right\}, \dim S = 2.\right)$$

7. Let $\mathbf{A} = \begin{Vmatrix} 1, & 2 \\ 3, & 4 \end{Vmatrix}$. Let $S = \{\mathbf{B} \in \mathscr{M} | \mathbf{AB} = \mathbf{BA}\}$ (i.e., S is the set of all matrices that commute with $\mathbf{A}$). Verify that S is a linear subspace of $\mathscr{M}$. Find a basis of S. What is $\dim S$? Exhibit a basis of S that contains the element $\mathbf{A}^2$.

$$\left(\text{Answer: } \left\{ \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix}, \begin{Vmatrix} 0, & 2, \\ 3, & 3 \end{Vmatrix} \right\}, \dim S = 2; \text{e.g. } \{\mathbf{A}, \mathbf{A}^2\}.\right)$$

8. Let $\mathbf{A} = \begin{Vmatrix} 3, & 1, & 0 \\ 0, & 0, & 3 \end{Vmatrix}$. Let $S = \{\mathbf{B} \in \mathscr{M} | \mathbf{AB} = \mathbf{BA}\}$. Verify that S is a linear subspace of $\mathscr{M}$. Find a basis of S. What is $\dim S$?

$$\left(\text{Answer: } \left\{ \begin{Vmatrix} 1, & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{Vmatrix}, \begin{Vmatrix} 0, & 1, & 0 \\ 0, & 0, & 1 \\ 0, & 0, & 0 \end{Vmatrix}, \begin{Vmatrix} 0, & 0, & 1 \\ 0, & 0, & 0 \\ 0, & 0, & 0 \end{Vmatrix} \right\}, \dim S = 3.\right)$$

9. Use both the elimination method and the determinant method and find the inverse matrices to the following matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$.

$$\mathbf{A} = \| -3 \|, \qquad \mathbf{B} = \begin{Vmatrix} 1, & 2 \\ 2, & 2 \end{Vmatrix},$$

$$\mathbf{C} = \begin{Vmatrix} 3, & 2, & 4 \\ 2, & 0, & 4 \\ 1, & 1, & 1 \end{Vmatrix}, \qquad \mathbf{D} = \begin{Vmatrix} 1, & -2, & 3 \\ 1, & -2, & 1 \\ 5, & 2, & -5 \end{Vmatrix}.$$

$\Bigg($Answer: $\mathbf{A}^{-1} = \| -\frac{1}{3} \|$, $\mathbf{B}^{-1} = \begin{Vmatrix} -1, & 1 \\ 1, & \frac{1}{2} \end{Vmatrix}$, $\mathbf{C}^{-1}$ does not exist, $\mathbf{D}^{-1} = \frac{1}{12} \begin{Vmatrix} 4, & -2, & 2 \\ 5, & -10, & 1 \\ 6, & -6, & 0 \end{Vmatrix}$. $\Bigg)$

10. Use both the elimination method and the determinant method and find the inverse matrices to the following matrices $\mathbf{A}$ and $\mathbf{B}$.

$$\mathbf{A} = \begin{Vmatrix} 1, & 0, & -1, & 2 \\ 1, & 1, & -1, & 1 \\ -2, & 0, & 3, & -6 \\ -4, & -1, & 6, & -10 \end{Vmatrix}, \qquad \mathbf{B} = \begin{Vmatrix} 1, & 1, & 1, & 1 \\ 1, & 1, & -1, & -1 \\ 1, & -1, & 1, & -1 \\ 1, & -1, & -1, & 1 \end{Vmatrix}$$

$\left(\text{Answer: } \mathbf{A}^{-1} = \begin{Vmatrix} 3, & 0, & 1, & 0 \\ -2, & 2, & -2, & 1 \\ 0, & 2, & -3, & 2 \\ -1, & 1, & -2, & 1 \end{Vmatrix}, \mathbf{B}^{-1} = \frac{1}{4} \begin{Vmatrix} 1, & 1, & 1, & 1 \\ 1, & 1, & -1, & -1 \\ 1, & -1, & 1, & -1 \\ 1, & -1, & -1, & 1 \end{Vmatrix}\right)$

11. Consider the following matrices $\mathbf{A}$ and $\mathbf{B}$ (α is a real parameter). Find for which α ($\alpha \in R$) the matrices $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist and then compute $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$.

$$\mathbf{A} = \begin{Vmatrix} -2\alpha, & -6\alpha - 5, & 2\alpha + 1 \\ \alpha, & 3\alpha + 1, & -\alpha \\ -\alpha - 1, & -3\alpha - 4, & \alpha + 1 \end{Vmatrix} \qquad \mathbf{B} = \begin{Vmatrix} 1, & a, & a \\ 1, & 1, & 2 \\ 1 + a, & 2, & 3 \end{Vmatrix}$$

$\left(\text{Answer: } \mathbf{A}^{-1} \text{ exists for every } \alpha \in R; \mathbf{A}^{-1} = \begin{Vmatrix} 1, & 1, & -1 \\ 0, & 1 + a, & a \\ 1, & 5 + 3a, & 3a \end{Vmatrix}, \mathbf{B}^{-1} \text{ exists } \Leftrightarrow\right.$ $\left. \alpha \neq \pm 1; \mathbf{B}^{-1} = \frac{1}{a^2 - 1} \begin{Vmatrix} -1, & -a, & a \\ 2a - 1, & 3 - a - a^2, & a - 2 \\ 1 - a, & -2 + a + a^2, & 1 - a \end{Vmatrix}\right)$

12. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are two regular matrices of the same size. Show that
    (i) $(\lambda \circ \mathbf{A})^{-1} = \frac{1}{\lambda} \circ \mathbf{A}^{-1}$ for any λ, $\lambda \neq 0$,
    (ii) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
    (iii) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$,
    (iv) $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$ ($\mathbf{A}^T$ denotes the transpose matrix)

13. Find the rank of the following matrices:

a) $\left\| \begin{matrix} 1, & -1 \\ 2, & 1 \end{matrix} \right\|$, b) $\left\| \begin{matrix} 1, & 1, & -1 \\ 2, & 0, & 3 \\ 3, & 1, & 2 \end{matrix} \right\|$, c) $\left\| \begin{matrix} 1, & 1, & 1, & 2 \\ 1, & 2, & 3, & 4 \\ 1, & 2, & 1, & 2 \end{matrix} \right\|$,

d) $\left\| \begin{matrix} 1, & 3, & 2, & 0 \\ 0, & -1, & -3, & 2 \\ 2, & 1, & 0, & -1 \\ 5, & 0, & 2, & -7 \end{matrix} \right\|$, e) $\left\| \begin{matrix} 41, & 133, & 36, & 29, & 22 \\ 7, & 6, & 3, & 4, & 1 \\ 6, & 103, & 21, & 9, & 17 \\ 14, & 12, & 6, & 8, & 2 \end{matrix} \right\|$, f) $\left\| \begin{matrix} 0, & 0, & 0, & 1, & 0 \\ 1, & 1, & 1, & 1, & 1 \\ 0, & 1, & 0, & 0, & 0 \\ 0, & 0, & 1, & 0, & 0 \\ 1, & 2, & 3, & 4, & 5 \\ 1, & 3, & 4, & 5, & 1 \end{matrix} \right\|$

(Answer: a) 2, b) 2, c) 3, d) 3, e) 2, f) 5.)

14. Consider the following matrix $\mathbf{A}$:

$$\mathbf{A} = \left\| \begin{matrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & & & \\ a_{p1}, & a_{p2}, & \dots, & a_{pn} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} \end{matrix} \right\|.$$

Suppose that every row of $\mathbf{A}$ is a linear combination of the first p rows. Denote by $\overrightarrow{c}_i$ ($\overrightarrow{c}_i \in R^m$) the i-th column of $\mathbf{A}$ and by $\overrightarrow{d}_i$ ($\overrightarrow{d}_i \in R^p$) the restriction of $\overrightarrow{c}_i$ to the first p coordinates. Then if

$$\overrightarrow{d}_j = \sum_{r=1}^{s} \alpha_{i_p} \cdot \overrightarrow{d}_{i_p}$$

for some indices $i_1, i_2, \dots, i_r$ then

$$\overrightarrow{c}_j = \sum_{r=1}^{s} \alpha_{i_p} \cdot \overrightarrow{c}_{i_p}.$$

(Compare with Th. 2.2.3.)

15. For the following matrix $\mathbf{A}$, check that $r(\mathbf{A}) = r(\mathbf{A}^T)$.

$$\mathbf{A} = \left\| \begin{matrix} 2, & 1, & 3, & -1, & 1 \\ 3, & -1, & 2, & 0, & 3 \\ 1, & 3, & 4, & -2, & -1 \\ 4, & -3, & 1, & 1, & 5 \end{matrix} \right\|.$$

$\Big($Answer: $r(\mathbf{A}) = 2 = r(\mathbf{A}^T)$.$\Big)$

16. For the following matrices $\mathbf{A}$ and $\mathbf{B}$, compute $r(\mathbf{AB})$. Check that $\mathbf{A}$ is regular and therefore $r(\mathbf{AB}) = r(\mathbf{B})$.

$$\mathbf{A} = \left\| \begin{matrix} 3, & 2, & 1, & 2 \\ 4, & 0, & -1, & 0 \\ 2, & 1, & 0, & -1 \\ 3, & 1, & 0, & 4 \end{matrix} \right\|, \qquad \mathbf{B} = \left\| \begin{matrix} 2, & 0, & 2 \\ 3, & 1, & 4 \\ 2, & -1, & 1 \\ 1, & 0, & 1 \end{matrix} \right\|.$$

(Answer: $r(\mathbf{AB}) = 2$.)

17. For the following matrices $\mathbf{A}$ and $\mathbf{B}$, compute $r(\mathbf{AB})$. Check that in this case $r(\mathbf{AB}) < r(\mathbf{A})$ and $r(\mathbf{AB}) < r(\mathbf{B})$.

$$\mathbf{A} = \begin{Vmatrix} 5, & 2, & -2, & 3 \\ 6, & 4, & -3, & 5 \\ 9, & 2, & -3, & 4 \\ 7, & 6, & -4, & 7 \end{Vmatrix}, \qquad \mathbf{B} = \begin{Vmatrix} 2, & 2, & 2, & 2 \\ -1, & -5, & 3, & 11 \\ 16, & 24, & 8, & -8 \\ 8, & 16, & 0, & -16 \end{Vmatrix}.$$

(Answer: $r(\mathbf{AB}) = 0$.)

18. Discuss the rank of the following matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ with respect to the parameter α $(\alpha \in R)$:

$$\text{a) } \mathbf{A} = \begin{Vmatrix} 4, & 1, & 9 \\ 5, & 3, & \alpha \\ 0, & 1, & 1 \end{Vmatrix}, \text{ b) } \mathbf{B} = \begin{Vmatrix} 2, & \alpha, & -1, & 1 \\ 5, & -1, & \alpha, & 2 \\ 1, & 10, & -6, & 1 \end{Vmatrix}, \text{ c) } \mathbf{C} = \begin{Vmatrix} 0, & 3, & 2, & -3 \\ 5, & \alpha, & -4, & 12 \\ 2, & -1, & 0, & 3 \\ 1, & 1, & -2, & 3 \end{Vmatrix}.$$

(Answer: a) $\alpha = 13 \Rightarrow r(\mathbf{A}) = 2$, b) $\alpha = 3 \Rightarrow r(\mathbf{B}) = 2$, c) $\alpha = -4 \Rightarrow r(\mathbf{C}) = 3$
$\alpha \neq 13 \Rightarrow r(\mathbf{A}) = 3$, $\alpha \neq 3 \Rightarrow r(\mathbf{B}) = 3$, $\alpha \neq -4 \Rightarrow r(\mathbf{C}) = 4$.)

19. * Prove that for any matrix, $r(\mathbf{A}) = r(\mathbf{AA}^T)$.

20. Using only the definition of determinant, compute the determinants of the following matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$:

$$\mathbf{A} = \|2\|, \mathbf{B} = \begin{Vmatrix} 2, & -1 \\ 1, & 1 \end{Vmatrix}, \mathbf{C} = \begin{Vmatrix} 3, & 2, & 4 \\ 4, & 3, & 9 \\ 5, & 4, & 16 \end{Vmatrix}, \mathbf{D} = \begin{Vmatrix} 4, & 2, & 3, & 5 \\ 0, & 2, & 3, & 1 \\ 1, & 0, & -1, & 1 \\ 2, & -3, & 1, & 0 \end{Vmatrix}.$$

(Answer: $\det \mathbf{A} = 2$, $\det \mathbf{B} = 3$, $\det \mathbf{C} = 2$, $\det \mathbf{D} = 4$.)

21. Compute $\det \mathbf{A}$. First use the elimination method, then by expansion along the first row, and then by expansion down the second column.

$$\mathbf{A} = \begin{Vmatrix} 1, & 6, & -8 \\ 2, & 9, & 7 \\ 5, & ,8, & 7 \end{Vmatrix}.$$

(Answer: $\det \mathbf{A} = 365$.)

22. Compute $\det \mathbf{A}$. First use the elimination method then by expansion along the second row, and then by expansion down the third column.

$$\mathbf{A} = \begin{Vmatrix} 3, & -3, & -2, & -5 \\ 2, & 5, & 4, & 6 \\ 5, & 5, & 8, & 7 \\ 4, & 4, & 5, & 6 \end{Vmatrix}.$$

(Answer: $\det \mathbf{A} = 90$.)

23. Compute the determinants of the following matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$.

$$\mathbf{A} = \left\| \begin{matrix} 4, & 2, & 1 \\ 2, & 2, & 1 \\ 5, & 1, & 2 \end{matrix} \right\|, \qquad \mathbf{B} = \left\| \begin{matrix} 2, & 3, & -4, & -1 \\ 0, & -2, & 1, & 1 \\ -1, & -7, & 2, & 8 \\ -3, & -5, & 4, & 2 \end{matrix} \right\|,$$

$$\mathbf{C} = \left\| \begin{matrix} -4, & 0, & 2, & -1, & 0 \\ 1, & 3, & -3, & -1, & 4 \\ 2, & 0, & 1, & 3, & 0 \\ -2, & 1, & -3, & -1, & 5 \\ 1, & -5, & 1, & 0, & 5 \end{matrix} \right\|, \qquad \mathbf{D} = \left\| \begin{matrix} 5, & 1, & 4, & 2, & 7, & 3 \\ -1, & 0, & -2, & 0, & -3, & 0 \\ 1, & 0, & 4, & 0, & 9, & 0 \\ 8, & 1, & 5, & 3, & 7, & 6 \\ 1, & 0, & 8, & 0, & 27, & 0 \\ 9, & 1, & 5, & 4, & 3, & 10 \end{matrix} \right\|.$$

(Answer: $\det \mathbf{A} = 6$, $\det \mathbf{B} = -40$, $\det \mathbf{C} = -997$, $\det \mathbf{D} = 12$.)

24. Compute the determinants of the following matrices $\mathbf{A}$ and $\mathbf{B}$. Then compute the determinants of $\mathbf{A}^{-2}$ and $\mathbf{B}^{-3}$. (We write $\mathbf{A}^{-2} = (\mathbf{A}^{-1})^2$ and $\mathbf{B}^{-3} = (\mathbf{B}^{-1})^3$. Use Th. 2.4.28 in computing $\det(\mathbf{A}^{-2})$ and $\det(\mathbf{B}^{-3})$.)

$$\mathbf{A} = \left\| \begin{matrix} 3, & 2, & -3 \\ 1, & 6, & -3 \\ 0, & 1, & -1 \end{matrix} \right\| \qquad \mathbf{B} = \left\| \begin{matrix} 1, & 0, & -2, & 3 \\ 0, & 1, & -3, & 4 \\ 3, & 2, & 0, & -5 \\ 4, & 3, & -5, & 0 \end{matrix} \right\|$$

(Answer: $\det \mathbf{A} = -10$, $\det(\mathbf{A}^{-2}) = 10^{-2}$, $\det \mathbf{B} = -24$, $\det(\mathbf{B}^{-3}) = -24^3$)

25. Consider the following matrix $\mathbf{A}$. Show that $\det \mathbf{A} = 0$, and then find a linear combination of the row vectors which is equal to the zero vector in R^4.

$$\mathbf{A} = \left\| \begin{matrix} 5, & 0, & 2, & -7 \\ 0, & -1, & -3, & 2 \\ 2, & 1, & 0, & -1 \\ 1, & 3, & 2, & 0 \end{matrix} \right\|$$

26. Using determinants, find all parameters λ $(\lambda \in R)$ for which the following matrices are regular:

$$\text{a)} \left\| \begin{matrix} 3, & 2, & 10 \\ 2, & -1, & \lambda \\ 5, & 10, & 30 \end{matrix} \right\|, \qquad \text{b)} \left\| \begin{matrix} 4, & 3, & -2, & 1 \\ 2, & -1, & 3, & 4 \\ \lambda, & 2, & 1, & 1 \\ -1, & -3, & 2, & 1 \end{matrix} \right\|.$$

(Answer: a) $\lambda \neq 2$, b) $\lambda \neq 0$.)

27. Let $\mathbf{A} = \left\| \begin{matrix} 1, & 0, & \alpha \\ 0, & 1, & \alpha \\ 0, & 0, & -1 \end{matrix} \right\|$ and $\mathbf{B} = \left\| \begin{matrix} 9, & 0, & 0 \\ 0, & 0, & 0 \\ -\alpha, & 0, & 0 \end{matrix} \right\|$ (α is a real parameter). Let $\mathbf{C} = \mathbf{B} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^7$. Discuss the rank of $\mathbf{C}$ with respect to the parameter α. Use first the elimination method on $\mathbf{C}$, then, in the second solution, consider $\det \mathbf{C}$. (Hint: Consider $\mathbf{A}^2$ first.)

$\Big($Answer: $\mathbf{C} = \begin{Vmatrix} 16, & 0, & 4\alpha \\ 0, & 7, & 4\alpha \\ -\alpha, & 0, & -1 \end{Vmatrix}$, if $\alpha = 2$ or -2, then $r(\mathbf{C}) = 2$, if $\alpha \in R = \{-2, 2\}$, then $r(\mathbf{C}) = 3.\Big)$

28. Find the parameter $\alpha (\alpha \in R)$ such that the following vectors $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}$ are LD. Use first the notion of LD only, then use determinants.

$$\overrightarrow{a} = (5, 4, 3), \ \overrightarrow{b} = (3, 3, 2), \ \overrightarrow{c} = (8, \alpha, 3).$$

(Answer: $\alpha = 1$.)

29. Using determinants, find all parameters $\alpha (\alpha \in R)$ for which the following vectors $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}$ do not constitute a basis of R^3 :

$$\overrightarrow{a} = (\alpha, 1, 1), \quad \overrightarrow{b} = (1, \alpha, 1), \quad \overrightarrow{c} = (1, 1, \alpha).$$

(Answer: $\alpha = 1$ or $\alpha = -2$.)

30. Find the parameter $\alpha (\alpha \in R)$ such that the vector $\overrightarrow{d}$ belongs to $\mathrm{Span}\{\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}\}$. Use first the notion of Span only, then use determinants.

$$\overrightarrow{a} = (1, 0, 2), \ \overrightarrow{b} = (0, 1, 1), \ \overrightarrow{c} = (4, 1, 9), \ \overrightarrow{d} = (5, 3, \alpha).$$

(Answer: $\alpha = 13$.)

31. Suppose you know that the numbers $228, 323$ and 437 are divisible by 19 . Show that the number

$$\det \begin{Vmatrix} 2, & 3, & 4 \\ 2, & 2, & 3 \\ 8, & 3, & 7 \end{Vmatrix}$$

is also divisible by 19 . Do not compute the determinant.

32. Let $\mathbf{E}$ be the unit matrix in $\mathscr{M}$ and let $\mathbf{A} = \begin{Vmatrix} \alpha + 1, & -\alpha \\ \alpha + 1, & -\alpha \end{Vmatrix}$, where α $(\alpha \in R)$ is a parameter. Solve the equation

$$\det \left(\alpha \mathbf{E} + \sum_{i=1}^{10} \mathbf{A}^i \right) = 24.$$

(Hint: Compute $\mathbf{A}^2$ first.)

(Answer: $\alpha_1 = 2, \alpha_2 = -12$.)

33. * Let $d \colon \mathscr{M} \to \mathscr{R}$ be a mapping which satisfies the following two conditions:

(i) $d(\mathbf{E}) = 1$ ($\mathbf{E}$ is the unit matrix),

(ii) $d(\mathbf{AB}) = d(\mathbf{A}) \cdot d(\mathbf{B})$.

Prove that $d(\mathbf{A}) = \det \mathbf{A}$. (This result holds true for a general $\mathscr{M}$.)

34. Compute the determinants of the following matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ (the matrices belong to $\mathscr{M}$):

$$\mathbf{A} = \begin{Vmatrix} 0 & 0 & & \dots & 0 & a_{1,n} \\ 0 & 0 & & \dots & a_{2,n-1} & 0 \\ \vdots & & & & & \\ 0 & a_{n-1,2} & 0 & \dots & & 0 \\ a_{n,1} & 0 & 0 & \dots & & 0 \end{Vmatrix}, \qquad \mathbf{B} = \begin{Vmatrix} 1 & 2 & 3 & \dots & n \\ x & 2 & 3 & \dots & n \\ 1 & x & 3 & \dots & n \\ \vdots & & & & \\ 1 & 2 & 3 & \dots & n \end{Vmatrix}$$

$$\mathbf{C} = \begin{Vmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1-x & 1 & \dots & 1 \\ 1 & 1 & 2-x & \dots & 1 \\ \vdots & & & & \\ 1 & 1 & 1 & \dots & n-x \end{Vmatrix}.$$

(Answer (the symbol $\prod$ denotes the product):

$$\det \mathbf{A} = (-1)^{\frac{(n-1)(n+4)}{2}} \cdot \prod_{j=1}^{n} a_{j,n-j+1}, \qquad \det \mathbf{B} = (-1)^{n-1} \cdot n \prod_{j=1}^{n-1} (x-j),$$

$$\det \mathbf{C} = (-1)^n \cdot \prod_{j=0}^{n-1} (x-j).)$$

35. Compute the determinants of the following matrices $\mathbf{A}$ and $\mathbf{B}$ (the matrices belong to $\mathscr{M}$). A hint: Derive the recursive formula from n to $n-1$.

$$\mathbf{A} = \begin{Vmatrix} 1+x^2 & x & 0 & 0 & \dots & 0 \\ x & 1+x^2 & x & 0 & \dots & 0 \\ 0 & x & 1+x^2 & x & \dots & \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & x & 1+x^2 \end{Vmatrix},$$

$$\mathbf{B} = \begin{Vmatrix} x & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & x & -1 & 0 & \dots & 0 & 0 \\ 0 & 0 & x & -1 & \dots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \dots & x & -1 \\ a_0 & a_1 & a_2 & a_3 & \dots & a_{n-1} & a_n \end{Vmatrix}.$$

(Answer: $\det \mathbf{A} = \sum_{k=0}^{n} x^{2k}, \det \mathbf{B} = \sum_{k=0}^{n} a_k x^k$.)

36. * Prove the formula for the Vandermond determinant. Suppose that $\mathbf{A}$ is the following matrix $(\mathbf{A} \in \mathscr{M})$:

$$\mathbf{A} = \begin{Vmatrix} 1, & 1, & 1, & \dots, & 1 \\ x_1, & x_2, & x_3, & \dots, & x_n \\ x_1^2, & x_2^2, & x_3^2, & \dots, & x_n^2 \\ \vdots & & & & \\ x_1^{n-1}, & x_2^{n-1}, & x_3^{n-1}, & \dots, & x_n^{n-1} \end{Vmatrix}.$$

Then $\det \mathbf{A} = \prod\limits_{i<j\le n} (x_j - x_i)\,.$

37. Solve the following homogeneous systems:

a) $\begin{array}{rcrcl} 2x & + & 3y & = & 0 \\ 3x & - & y & = & 0 \end{array}$

b) $\begin{array}{rcrcrcl} 5x_1 & - & 2x_2 & - & 3x_3 & = & 0 \\ 9x_1 & + & 3x_2 & - & x_3 & = & 0 \end{array}$

c) $\begin{array}{rcrcrcl} 2x & + & y & - & 4z & = & 0 \\ 3x & + & 5y & - & 7z & = & 0 \\ 4x & - & 5y & - & 6z & = & 0 \end{array}$

d) $\begin{array}{rcrcrcl} 2x_1 & + & x_2 & + & 3x_3 & = & 0 \\ 2x_1 & + & 3x_2 & + & x_3 & = & 0 \\ 3x_1 & + & 2x_2 & + & x_3 & = & 0 \end{array}$

e) $\begin{array}{rcrcrcrcl} 2x & + & 3y & - & z & + & 5u & = & 0 \\ 3x & - & y & + & 2z & - & 7u & = & 0 \\ 4x & + & y & - & 3z & + & 6u & = & 0 \\ x & - & 2y & + & 4z & - & 7u & = & 0 \end{array}$

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

g) $\begin{array}{rcrcrcrcl} x & + & 2y & - & 3z & + & 4u & = & 0 \\ 2x & - & 5y & + & 7z & - & u & = & 0 \\ 4x & - & y & + & z & + & 7u & = & 0 \\ 3x & + & 15y & - & 22z & + & 21u & = & 0 \\ 3x & - & 3y & + & 4z & + & 3u & = & 0 \end{array}$

h) $x - 2y + z + u - 0$ i) $x_1 - x_2 + x_4 + x_6 = 0$

(Answer: a) $\overrightarrow{x} = (0,0)$, b) $\overrightarrow{x} = \alpha(1,-2,3), \alpha \in R$, c) $\overrightarrow{x} = \alpha(13,2,7), \alpha \in R$, d) $\overrightarrow{x} = (0,0,0)$, e) $\overrightarrow{x} = (0,0,0,0)$, f) $\overrightarrow{x} = \alpha(5,-7,5,6), \alpha \in R$, g) $\overrightarrow{x} = \alpha(1,13,9,0) + \beta(2,1,0,-1)$, h) $\overrightarrow{x} = \alpha(2,1,0,0) + \beta(0,1,2,0) + \gamma(0,0,-1,1), \alpha,\beta,\gamma \in R$, i) $\overrightarrow{x} = \alpha(0,0,1,0,0,0) + \beta(0,0,0,0,1,0) + \gamma(1,-1,0,0,0,0) + \delta(0,0,0,1,0,-1) + \varepsilon(0,1,0,0,0,1), \alpha,\beta,\gamma,\delta,\varepsilon \in R$.)

38. Using Frobenius's theorem, show that the following systems do not have any solutions. Determinants may help.

a) $\begin{array}{rcrcr} 2x & - & 6y & = & -1 \\ x & - & 3y & = & 4 \end{array}$

b) $\begin{array}{rcrcrcr} x & + & y & + & z & = & 5 \\ x & - & y & + & z & = & 2 \\ x & & & + & z & = & 3 \end{array}$

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

d) $\begin{array}{rcrcrcrcr} 3x & - & 2y & + & 5z & - & 6u & = & 1 \\ 7x & + & y & - & 3z & - & 4u & = & 1 \\ 6x & + & 5y & - & 13z & + & 3u & = & 1 \end{array}$

39. Solve the following nonhomogeneous systems:

a) $\begin{array}{rcrcr} 3x & + & y & = & -1 \\ 2x & + & y & = & 2 \end{array}$

b) $\begin{array}{rcrcr} -2x_1 & + & x_2 & = & 2 \\ 4x_1 & - & 2x_2 & = & -4 \end{array}$

c) $\begin{array}{rcrcr} 3x_1 & - & 2x_2 & = & 4 \\ -6x_1 & + & 4x_2 & = & 1 \end{array}$

d) $\begin{array}{rcrcrcr} x & - & 4y & + & z & = & 1 \\ 3x & - & 9y & - & 4z & = & 11 \\ 5x & + & 2y & + & z & = & 111 \end{array}$

e)
$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 5x_3 & = & 20 \\ 3x_1 & - & x_2 & + & 2x_3 & = & 7 \\ x_1 & - & 5x_2 & - & 8x_3 & = & -33 \end{array}$$

f)
$$\begin{array}{rcrcrcrcr} 6x_1 & - & 9x_2 & + & 7x_3 & + & 10x_4 & = & 3 \\ 2x_1 & - & 3x_2 & - & 3x_3 & - & 4x_4 & = & 1 \\ 2x_1 & - & 3x_2 & + & 13x_3 & + & 18x_4 & = & 1 \end{array}$$

g)
$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & - & 3x_3 & + & x_4 & = & -5 \\ 2x_1 & + & 3x_2 & - & x_3 & + & 2x_4 & = & 0 \\ 7x_1 & - & x_2 & + & 4x_3 & - & 3x_4 & = & 15 \\ x_1 & + & x_2 & - & 2x_3 & - & x_4 & = & -3 \end{array}$$

h)
$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & + & 3x_3 & - & x_4 & = & 7 \\ 4x_1 & - & 7x_2 & - & 12x_3 & + & 11x_4 & = & 4 \\ x_1 & - & 8x_2 & - & 13x_3 & + & 9x_4 & = & -9 \\ 2x_1 & - & x_2 & - & 2x_3 & + & 3x_4 & = & 6 \\ 3x_1 & + & x_2 & + & x_3 & + & 2x_4 & = & 13 \end{array}$$

i) $x_1 + x_2 - x_4 + x_5 = 1$

j)
$$\begin{array}{rcrcrcrcrcrcr} 2x_1 & - & x_2 & + & 3x_3 & + & x_4 & + & 2x_5 & + & 3x_6 & = & 1 \\ 2x_1 & - & 3x_2 & - & x_3 & + & 3x_4 & - & 2x_5 & + & x_6 & = & 1 \\ 2x_1 & & & + & 5x_3 & + & 2x_4 & + & 5x_5 & + & 3x_6 & = & 4 \end{array}$$

(Answer: a) $\overrightarrow{x} = (-3, 8)$, b) $\overrightarrow{x} = (-1, 0) + \alpha(1, 2), \alpha \in R$, c) no solution, d) $\overrightarrow{x} = (20, 5, 1)$, e) $\overrightarrow{x} = (1, 2, 3) + \alpha(9, 13, -7), \alpha \in R$, f) $\overrightarrow{x} = (0, 0, -11, 8) + \alpha(1, 0, 22, -16) + \beta(0, 1, -33, 24), \alpha, \beta \in R$, g) $\overrightarrow{x} = (1, 0, 2, 0)$, h) $\overrightarrow{x} = (3, 1, 1, 1) + \alpha(1, -1, 0, -1) + \beta(1, -8, 5, 0), \alpha, \beta \in R$, i) $\overrightarrow{x} = (1, 0, 0, 0, 0) + t(1, -1, 0, 0, 0) + u(0, 0, 1, 0, 0) + v(0, 1, 0, 1, 0) + w(0, 0, 0, 1, 1)$, $t, u, v, w \in R$, j) $\overrightarrow{x} = (-\frac{11}{2}, -6, 3, 0, 0, 0) + a(0, -3, 1, 0, 0, 1) + b(4, 5, -2, 0, 1, 0) + c(-5, -4, 0, 2, 0, 0), a, b, c \in R$.)

40. Using the inverse matrix technique, solve the following systems. Solve them also by Cramer's rule.

a)
$$\begin{array}{rcrcr} 8x & - & 3y & = & -12 \\ 3x & + & 2y & = & 33 \end{array}$$

b)
$$\begin{array}{rcrcrcr} 5x & + & 5y & + & z & = & 2 \\ 3x & - & 4y & - & 3z & = & 1 \\ -2x & + & y & + & z & = & -1 \end{array}$$

c)
$$\begin{array}{rcrcrcr} -3x & - & y & + & 6z & = & -5 \\ 4x & & & + & 7z & = & 19 \\ -x & - & y & + & 11z & = & 6 \end{array}$$

d)
$$\begin{array}{rcrcrcrcr} 4x & + & 3y & - & 5z & & & = & 5 \\ 3x & + & 2y & & & - & 5u & = & 12 \\ x & & & - & 2z & + & 3u & = & -4 \\ x & + & y & - & 5z & + & 7u & = & -9 \end{array}$$

(Answer: a) $\overrightarrow{x} = (3, 12)$, b) $\overrightarrow{x} - (1, -1, 2)$, c) $\overrightarrow{x} = (3, 2, 1)$, d) $\overrightarrow{x} = (1, 2, 1, 1)$.)

41. Solve the following matrix equations. Use inverse matrices.

a) $\mathbf{AXB} = \mathbf{C}$, where $\mathbf{A} = \begin{Vmatrix} 2, & 3 \\ 4, & 2 \end{Vmatrix}$, $\mathbf{B} = \begin{Vmatrix} 1, & 0 \\ 2, & 1 \end{Vmatrix}$, $\mathbf{C} = \begin{Vmatrix} 7, & 5 \\ 6, & 1 \end{Vmatrix}$.

b) $\mathbf{XA} = \mathbf{B}$, where $\mathbf{A} = \begin{Vmatrix} 2, & 7, & 3 \\ 3, & 9, & 4 \\ 1, & 5, & 3 \end{Vmatrix}$, $\mathbf{B} = \begin{Vmatrix} 1, & 0, & 2 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{Vmatrix}$.

c) $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} = \begin{Vmatrix} 1, & 2, & 3, & 4 \\ 0, & 1, & 2, & 3 \\ 0, & 0, & 1, & 2 \\ 0, & 0, & 0, & 1 \end{Vmatrix}$, $\mathbf{B} = \begin{Vmatrix} 7, & 9, & 6, & 3 \\ 5, & 6, & 3, & 1 \\ 3, & 3, & 1, & 0 \\ 1, & 1, & 0, & 0 \end{Vmatrix}$.

[Answer: a) $\mathbf{X} = \frac{1}{8} \begin{Vmatrix} 18, & -7 \\ -20, & 18 \end{Vmatrix}$, b) $\mathbf{X} = \frac{1}{3} \begin{Vmatrix} -19, & 12, & 5 \\ 5, & -3, & -1 \\ -6, & 3, & 3 \end{Vmatrix}$, c) $\mathbf{X} = \begin{Vmatrix} 0, & 0, & 1, & 1 \\ 0, & 1, & 1, & 1 \\ 1, & 1, & 1, & 0 \\ 1, & 1, & 0, & 0 \end{Vmatrix}$.]

42. Discuss the solution of the following systems with respect to the parameter λ:

a) $\begin{array}{rcrcl} \lambda x & + & y & = & 1 \\ x & + & \lambda y & = & \lambda \end{array}$

b) $\begin{array}{rcrcrcr} x & - & 7y & - & 5z & = & 0 \\ -2x & + & \lambda y & + & z & = & -3 \\ -x & + & 3y & + & \lambda z & = & -1 \end{array}$

c) $\begin{array}{rcrcrcl} x & + & y & + & \lambda z & = & \lambda \\ x & + & \lambda y & + & z & = & 1 \\ \lambda x & + & y & + & z & = & 1 \end{array}$

d) $\begin{array}{rcrcrcr} \lambda x_1 & + & x_2 & + & x_3 & = & 1 \\ x_1 & + & \lambda x_2 & + & x_3 & = & -1 \\ x_1 & + & x_2 & + & \lambda x_3 & = & (\lambda-1)^2 \end{array}$

(Answer: a) $\lambda = 1 \Rightarrow (x, y) = (1, 0) + \alpha(-1, 1),\ \alpha \in R$
$\lambda = -1 \Rightarrow (x, y) = (-1, 0) + \alpha(1, 1),\ \alpha \in R$
$\lambda \neq 1, \lambda \neq -1 \Rightarrow (x, y) = (0, 1)$

b) $\lambda = 17 \Rightarrow$ no solution
$\lambda = 2 \Rightarrow (x, y, z) = (1, -2, 3) + \alpha(1, 3, -4),\ \alpha \in R$
$\lambda \neq 17, \lambda \neq 2 \Rightarrow (x, y, z) = \left(\frac{26}{17-\lambda}, \frac{3}{17-\lambda}, \frac{1}{17-\lambda}\right)$

c) $\lambda = 1 \Rightarrow (x, y, z) = (1, 0, 0) + \alpha(1, 0, -1) + \beta(0, 1, -1),\ \alpha, \beta \in R$
$\lambda = -2 \Rightarrow (x, y, z) = (0, 0, 1) + \alpha(1, 1, 1),\ \alpha \in R$
$\lambda \neq 1, \lambda \neq -2 \Rightarrow (x, y, z) = (0, 0, 1)$

d) $\lambda = 1 \Rightarrow$ no solution
$\lambda = -2 \Rightarrow$ no solution
$\lambda \neq 1, \lambda \neq -2 \Rightarrow (x, y, z) = \left(\frac{-\lambda^2+3\lambda+1}{(\lambda-1)(\lambda+2)}, \frac{-\lambda^2+\lambda-3}{(\lambda-1)(\lambda+2)}, \frac{\lambda^2-1}{\lambda+2}\right)$)

43. Discuss the solution of the following systems with respect to the parameters $a, b (a, b \in R)$:

a) $\begin{array}{rcrcrcr} -2x & + & y & + & z & = & 7 \\ ax & + & 2y & - & z & = & 2 \\ -3x & + & y & + & 2z & = & -b \end{array}$

b) $\begin{array}{rcrcrcl} ax & + & y & + & z & = & 1 \\ x & + & ay & + & z & = & a \\ x & + & y & + & az & = & b \end{array}$

(Answer: a)$a \neq -1 \Rightarrow$ exactly one solution (it can be computed by Cramer's rule)
$a = -1, b \neq -11 \Rightarrow$ no solution
$a = -1, b = -11 \Rightarrow (x, y, z) = (-3, 0, 1) + \alpha(1, 1, 1), \alpha \in R.$
b)$a \neq 1, a \neq -2 \Rightarrow$ exactly one solution (it can be computed by Cramer's rule)
$a = 1, b \neq 1 \Rightarrow$ no solution
$a = 1, b = 1 \Rightarrow (x, y, z) = (1, 0, 0) + \alpha(1, -1, 0) + \beta(0, 1, -1), \alpha \in R, \beta \in R$
$a = -2, b \neq 1 \Rightarrow$ no solution
$a = -2, b = 1 \Rightarrow (x, y, z) = (0, 1, 0) + \alpha(1, 1, 1), \alpha \in R.$

44. Discuss the solution of the following systems with respect to the parameter λ $(\lambda \in R)$:

a)
$$\begin{array}{rcrcrcrcr} x & + & \lambda y & - & u & + & z & = & -3 \\ -7x & + & \lambda y & + & 3u & + & z & = & -3 \\ -2x & + & y & + & 3u & - & 4z & = & 4 \\ 3x & + & y & & & - & 3z & = & 1 \end{array}$$

b)
$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 6 \\ 3x_1 & + & 7x_2 & - & 4x_3 & = & 16 \\ x_1 & + & \lambda x_2 & - & 8x_3 & = & 4 \\ x_1 & + & 5x_2 & - & 8x_3 & = & 4 \end{array}$$

(Answer: a)$\lambda \neq 0 \Rightarrow$ exactly one solution (it can be computed by Cramer's rule)
$\lambda = 0 \Rightarrow (x, y, z, u) = (0, -8, 0, -3) + \alpha(1, 0, 2, 1), \alpha \in R.$
b)$\lambda = 5 \Rightarrow (x_1, x_2, x_3) = (2, 2, 1) + \alpha(-9, 5, 2),\ \alpha \in R, \lambda \neq 5 \Rightarrow$ no solution.)

Chapter 3

Linear transformations

3.1. Introduction. Basic examples

In this chapter we will study mappings that preserve the linear structure of the respective linear spaces. We will call these mappings *linear transformations*. They naturally arise in a variety of engineering problems, in particular in problems from control theory. Since we sometimes explicitly need to distinguish the operations of the "domain" and the "range" space, we occasionally return to the full notation $(L, \sharp, \circ), (L', \sharp', \circ')$, etc., when we deal with linear spaces. We will however refer to the underlying sets only, L, L', etc., if there is no danger of misunderstanding. Also, we will relax the precise notation of operations $\sharp, \circ$, etc., to the simplified form of using $+$ and omitting $\circ$ when we cannot confuse the reader.

3.1.1. Definition. Let $(L, \sharp, \circ)$ and $(L', \sharp', \circ')$ be linear spaces. A mapping $l\colon L \to L'$ is called *a linear transformation* (or simply a *linear mapping*) if the following two conditions are fulfilled:

(i) $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(\overrightarrow{x}) \sharp' l(\overrightarrow{y})$ for any $\overrightarrow{x}, \overrightarrow{y} \in L$,

(ii) $l(\lambda \circ \overrightarrow{x}) = \lambda \circ' l(\overrightarrow{x})$ for any $\overrightarrow{x} \in L$ and any $\lambda \in R$.

According to the above definition, l is linear if the following commutativity law holds true: The mapping l sends a sum (resp., a scalar multiple) of the *inputs* in L to the sum (resp., the scalar multiple) of *outputs* in L'. This property is often called *a superposition principle* in the technical literature.

3.1.2. Examples.

1. Let $L = L' = R$. Define $l\colon R \to R$ by putting $l(\overrightarrow{x}) = l(x) = a \cdot x$, where a is a real constant. Then l is a linear transformation. Indeed, $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(x+y) = a(x+y) = ax + ay = l(\overrightarrow{x}) \sharp' l(\overrightarrow{y})$, $l(\lambda \circ \overrightarrow{x}) = l(\lambda x) = a(\lambda x) = \lambda(ax) = \lambda \circ' l(\overrightarrow{x})$.
2. Let $L = L' = R$. Define $l\colon R \to R$ by putting $l(\overrightarrow{x}) = l(x) = ax + b$, where a, b are real constants, $b \neq 0$. Then l is *not* a linear transformation. (For instance, if $\overrightarrow{x} = 0$ and $\overrightarrow{y} = 1$ then $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(0+1) = l(1) = a + b$ whereas $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y}) = l(0) \sharp' l(1) = b + (a+b) = a + 2b$.)
3. Let $L = (R^+, \sharp, \circ)$ be endowed with the following operation: $x \sharp y = x \cdot y$, $\lambda \circ x = x^\lambda$ (see the illustrating example on p. 4). Let $L' = R$. Let $l\colon R^+ \to R$ be defined as follows: $l(x) = \ln x$. Then l is a linear transformation.
4. Let $L = R^2$ and $L' = R$. Define $l\colon R^2 \to R$ by putting $l(\overrightarrow{x}) = l(x_1, x_2) = x_1^2 + x_2^2$. Then l is not a linear mapping. (For instance, $l(2 \circ (1,1)) = l(2,2) = 4 + 4 = 8$, whereas $2 \circ' l(1,1) = 2(1+1) = 4$.)

5. Let $L = R^3$ and $L' = R^2$. Define $l\colon R^3 \to R^2$ by putting $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 3x_1 + x_2 - x_3)$. Then l is a linear mapping. Let us verify it in detail.

   (i)
$$\begin{aligned} l(\overrightarrow{x} \sharp \overrightarrow{y}) &= l((x_1, x_2, x_3)) \sharp (y_1, y_2, y_3)) = l(x_1 + y_1, x_2 + y_2, x_3 + y_3) \\ &= ((x_1 + y_1) - (x_2 + y_2) + 2(x_3 + y_3), 3(x_1 + y_1) + x_2 + y_2 - (x_3 + y_3)), \\ l(\overrightarrow{x} \sharp' l(\overrightarrow{y}) &= l(x_1, x_2, x_3)) \sharp' l(y_1, y_2, y_3) \\ &= (x_1 - x_2 + 2x_3, 3x_1 + x_2 - x_3) \sharp' (y_1 - y_2 + 2y_3, 3y_1 + y_2 - y_3) \\ &= ((x_1 - x_2 + 2x_3) + (y_1 - y_2 + 2y_3), (3x_1 + x_2 - x_3) + (3y_1 + y_2 - y_3)). \end{aligned}$$

   We see that $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(\overrightarrow{x}) \sharp' l(\overrightarrow{y})$ for any $\overrightarrow{x}, \overrightarrow{y} \in L$.

   (ii)
$$\begin{aligned} l(\lambda \circ \overrightarrow{x}) &= l(\lambda \circ (x_1, x_2, x_3)) = l(\lambda x_1, \lambda x_2, \lambda x_3) \\ &= (\lambda x_1 - \lambda x_2 + 2\lambda x_3, 3\lambda x_1 + \lambda x_2 - \lambda x_3), \\ \lambda \circ' l(\overrightarrow{x}) &= \lambda \circ' l(x_1, x_2, x_3) = \lambda \circ' (x_1 - x_2 + 2x_3, 3x_1 + x_2 - x_3) \\ &= (\lambda(x_1 - x_2 + 2x_3), \lambda(3x_1 + x_2 - x_3)). \end{aligned}$$

   We see that $l(\lambda \circ \overrightarrow{x}) = \lambda \circ' l(\overrightarrow{x})$ for any $\overrightarrow{x} \in L$ and any $\lambda \in R$.

6. Let $L = R$ and $L' = R^2$. Define $l\colon L \to L'$ by putting $l(\overrightarrow{x}) = l(x) = (x, 0)$. Then l is a linear transformation.

7. Let $L = \mathscr{M}$ and $L' = R^2$. Suppose that $\begin{Vmatrix} a_1 \\ a_2 \\ a_3 \end{Vmatrix}$ is a constant column matrix. Define $l\colon \mathscr{M} \to \mathscr{R}$ by putting
$$l(\overrightarrow{x}) = l\left(\begin{Vmatrix} x_{11}, & x_{12}, & x_{13} \\ x_{21}, & x_{22}, & x_{23} \end{Vmatrix}\right) = \begin{Vmatrix} x_{11}, & x_{12}, & x_{13} \\ x_{21}, & x_{22}, & x_{23} \end{Vmatrix} \begin{Vmatrix} a_1 \\ a_2 \\ a_3 \end{Vmatrix}.$$
   Then l is a linear transformation.

8. Let L, L' be arbitrary linear spaces. Define $l\colon L \to L'$ by putting $l(\overrightarrow{x}) = \overrightarrow{o}'$ for any $\overrightarrow{x} \in L$. Then l is a linear transformation.

9. Suppose that $\mathbf{A} \in \mathscr{M}$. Define $l\colon \mathscr{M} \to \mathscr{M}$ by putting, for a matrix $\mathbf{X} \in \mathscr{M}$, $(\mathbf{X}) = \mathbf{XA}$. Then l is a linear transformation.

10. Let $C\langle a, b\rangle$ denote the linear space of all continuous functions on $\langle a, b\rangle$. Define l: $C\langle a, b\rangle \to R$ by putting for a function $f(x) \in C\langle a, b\rangle$, $l(f) = \int_a^b f(x)dx$. Then l is a linear mapping.

11. Let $\mathscr{D}(\mathscr{R})$ denote the linear space of all infinitely differentiable functions on R. Then for every $n \in N$, the mapping $l\colon \mathscr{D}(\mathscr{R}) \to \mathscr{D}(\mathscr{R})$ defined by $l(f(x)) = f^{(n)}(x)$ ($f^{(n)}(x)$ means the n-th derivative of $f(x)$) is a linear mapping.

Let us collect basic properties of linear transformations.

3.1.3. Proposition. If $l\colon (L, \sharp, \circ) \to (L', \sharp', \circ')$ is a linear transformation, then $l(\overrightarrow{o}) = \overrightarrow{o}'$.

PROOF: We have $l(\overrightarrow{o}) = l(\overrightarrow{o} \sharp \overrightarrow{o}) = l(\overrightarrow{o}) \sharp' l(\overrightarrow{o}) = 2 \circ' l(\overrightarrow{o})$. Thus, if we write $l(\overrightarrow{o}) = \overrightarrow{a}$, then $\overrightarrow{a} = 2 \circ' \overrightarrow{a}$. Thus, by the calculus of linear space, $\overrightarrow{a} = \overrightarrow{o}'$, and therefore $l(\overrightarrow{o}) = \overrightarrow{o}'$.

3.1.4. Proposition. Suppose that $l\colon (L,\ \sharp\ ,\circ) \to (L',\ \sharp\ ',\circ')$ is a linear transformation. If $\lambda_1 \circ \overrightarrow{a_1} \ \sharp\ \lambda_2 \circ \overrightarrow{a_2} \ \sharp\ \ldots\ \sharp\ \lambda_k \circ \overrightarrow{a_k}$ is a linear combination of vectors in L, then

$$l(\lambda_1 \circ \overrightarrow{a_1} \ \sharp\ \lambda_2 \circ \overrightarrow{a_2} \ \sharp\ \ldots\ \sharp\ \lambda_k \circ \overrightarrow{a_k}) = \lambda_1 \circ' l(\overrightarrow{a_1}) \ \sharp\,'\lambda_2 \circ' l(\overrightarrow{a_2}) \ \sharp\,' \ldots\ \sharp\,'\lambda_k \circ' l(\overrightarrow{a_k}).$$

Proof: The proof of the last statement is obvious – conditions (i) and (ii) of Definition 3.1 can easily be extended to arbitrary linear combinations. A formal proof, which can be done by mathematical induction over the number of vectors in the combination, would proceed as follows. If $k = 1$, then the statement is obvious – the property $l(\lambda_1 \circ \overrightarrow{a_1}) = \lambda_1 \circ' l(\overrightarrow{a_1})$ is guaranteed by Def. 3.1 (ii). Suppose now that l preserves all linear combinations of vectors whose number does not exceed $k-1$. Consider now $l(\lambda_1 \circ \overrightarrow{a}_1 \ \sharp\ \lambda_2 \circ \overrightarrow{a}_2 \ \sharp\ \ldots\ \sharp\ \lambda_k \circ \overrightarrow{a}_k)$. We have

$$\begin{aligned}
&l(\lambda_1 \circ \overrightarrow{a_1} \ \sharp\ \lambda_2 \circ \overrightarrow{a_2} \ \sharp\ \ldots\ \sharp\ \lambda_k \circ \overrightarrow{a_k})\\
&\qquad = l\,[(\lambda_1 \circ \overrightarrow{a_1} \ \sharp\ \lambda_2 \circ \overrightarrow{a_2} \ \sharp\ \ldots\ \sharp\ \lambda_{k-1} \circ \overrightarrow{a}_{k-1}) \ \sharp\ \lambda_k \circ \overrightarrow{a}_k] \quad \text{(Def 3.1 (i))}\\
&\qquad = l(\lambda_1 \circ \overrightarrow{a_1} \ \sharp\ \lambda_2 \circ \overrightarrow{a_2} \ \sharp\ \ldots\ \sharp\ \lambda_{k-1} \circ \overrightarrow{a}_{k-1}) \ \sharp\ l(\lambda_k \circ \overrightarrow{a}_k)\\
&\qquad \text{(by the inductive assumption and Def. 3.1 (ii))}\\
&\qquad = \lambda_1 \circ' l(\overrightarrow{a_1}) \ \sharp\,'\lambda_2 \circ' l(\overrightarrow{a_2}) \ \sharp\,' \ldots\ \sharp\,'\lambda_{k-1} \circ' l(\overrightarrow{a}_{k-1} \ \sharp\,'\lambda_k \circ' l(\overrightarrow{a}_k).
\end{aligned}$$

□

Let us now analyze linear transformations in more detail. Let us first recall two standard set-theory notions.

3.1.5. Definition. Let $l\colon L \to L'$ be a linear transformation. Then l is called *injective* provided the following condition is fulfilled: If $\overrightarrow{x} \neq \overrightarrow{y}$ $(\overrightarrow{x}, \overrightarrow{y} \in L)$, then $l(\overrightarrow{x}) \neq l(\overrightarrow{y})$. Dually, l is called *surjective* provided the following condition is fulfilled: If $\overrightarrow{y}' \in L'$, then there exists a vector $\overrightarrow{x}$, $\overrightarrow{x} \in L$, such that $l(\overrightarrow{x}) = \overrightarrow{y}'$. The mapping l is called *a linear isomorphism* (or simple an isomorphism) if l is both injective and surjective.

In applications we often need to know how far a mapping is from being injective (resp. surjective). For linear transformations this can be found out relatively easily. The following notions are helpful.

3.1.6. Definition. Let $l\colon L \to L'$ be a linear transformation. Let $\mathrm{Ker}(l) = \{\overrightarrow{x} \in L \mid l(\overrightarrow{x}) = \overrightarrow{o}'\}$ and call $\mathrm{Ker}(l)$ *the kernel of* l. Obviously, $\mathrm{Ker}(l) \subset L$. Further, let $\mathrm{Im}(l) = \{l(\overrightarrow{x}) \mid \overrightarrow{x} \in L\}$ and call $\mathrm{Im}(l)$ *the image* of L under the transformation l. Obviously, $\mathrm{Im}(l) \subset L'$.

3.1.7. Proposition. Let $l\colon L \to L'$ be a linear transformation. Then

(i) $\mathrm{Ker}(l)$ is a linear subspace of L,

(ii) $\mathrm{Im}(l)$ is a linear subspace of L'.

Proof: (i) Suppose that $\overrightarrow{x}, \overrightarrow{y} \in \mathrm{Ker}(l)$. We have to show that $\overrightarrow{x} \ \sharp\ \overrightarrow{y} \in \mathrm{Ker}(l)$ and $\lambda \circ \overrightarrow{x} \in \mathrm{Ker}(l) (\lambda \in R)$. Computing $l(\overrightarrow{x} \ \sharp\ \overrightarrow{y})$ and $l(\lambda \circ \overrightarrow{x})$, we obtain

$$\begin{aligned}
l(\overrightarrow{x} \ \sharp\ \overrightarrow{y}) &= l(\overrightarrow{x}) \ \sharp\,' l(\overrightarrow{y}) = \overrightarrow{o}' \ \sharp\,' \overrightarrow{o}' = o',\\
l(\lambda \circ \overrightarrow{x}) &= \lambda \circ' l(\overrightarrow{x}) = \lambda \circ' \overrightarrow{o}' = o'.
\end{aligned}$$

Thus, $\vec{x} \sharp \vec{y} \in \mathrm{Ker}(l)$ and $\lambda \circ \vec{x} \in \mathrm{Ker}(l)$. We see that $\mathrm{Ker}(l)$ is a linear subspace of L.

(ii) Suppose that $\vec{u}, \vec{v} \in \mathrm{Im}(l)$. We have to show that $\vec{u} \sharp' \vec{v} \in \mathrm{Im}(l)$ and $\lambda \circ' \vec{u} \in \mathrm{Im}(l)$. According to the definition of $\mathrm{Im}(l)$, there are vectors $\vec{x}, \vec{y} \in L$ such that $\vec{u} = l(\vec{x})$ and $\vec{v} = l(\vec{y})$. Computing $\vec{u} \sharp' \vec{v}$ and $\lambda \circ' \vec{u}$, we obtain

$$\begin{aligned} \vec{u} \sharp' \vec{v} &= l(\vec{x}) \sharp' l(\vec{y}) = l(\vec{x} \sharp \vec{y}), \\ \lambda \circ' \vec{u} &= \lambda \circ' l(\vec{x}) = l(\lambda \circ \vec{x}). \end{aligned}$$

We see that both vectors $\vec{u} \sharp' \vec{v}$ and $\lambda \circ' \vec{u}$ have a "preimage" in L. Thus, $\vec{u} \sharp' \vec{v} \in \mathrm{Im}(l)$ and $\lambda \circ' \vec{u} \in \mathrm{Im}(l)$, which proves that $\mathrm{Im}(l)$ is a linear subspace of L'. □

3.1.8. Definition. Let $l\colon L \to L'$ be a linear transformation. Then $\dim \mathrm{Ker}(l)$ is called *the defect* of l. We denote the defect of l by $d(l)$. Dually, $\dim \mathrm{Im}(l)$ is called *the rank of* l. We denote the rank of l by $r(l)$.

3.1.9. Proposition. Let $l\colon L \to L'$ be a linear transformation.

(i) The transformation l is injective if and only if $d(l) = 0$ (i.e., $\mathrm{Ker}\, l = \{\vec{o}\}$).

(ii) Suppose that L' is finite dimensional. Then the transformation l is surjective if and only if $r(l) = \dim L'$.

Proof: (i) We have to prove that l is injective exactly when the zero vector of L, $\vec{o}$, is *the only vector* that is mapped into $\vec{o}\,'$. Let us verify the two implications required.

Suppose that l is injective. Since $l(\vec{o}) = \vec{o}\,'$ for any linear transformation, we immediately see that $l(\vec{x}) \neq \vec{o}\,'$ provided $\vec{x} \neq \vec{o}$. Thus, $\mathrm{Ker}\, l = \{\vec{o}\}$.

Suppose on the contrary that $\mathrm{Ker}(l) = \{\vec{o}\}$. We want to show that l is injective. By the definition, we want to show that $l(\vec{x}) \neq l(\vec{y})$ provided $\vec{x} \neq \vec{y}$ ($\vec{x}, \vec{y} \in L$). We will prove it by contradiction. Suppose that $\vec{x} \neq \vec{y}$ and $l(\vec{x}) = l(\vec{y})$. Then $l(\vec{x}) - l(\vec{y}) = \vec{o'}$ and since l is linear, we have $l(\vec{x} - \vec{y}) = \vec{o}\,'$. By our assumption $\mathrm{Ker}(l) = \{\vec{o}\}$ and therefore $\vec{x} - \vec{y} = \vec{o}$. This implies $\vec{x} = \vec{y}$ which is a contradiction. This shows that if $\mathrm{Ker}(l) = \{\vec{o}\}$, then l is injective.

(ii) This is obvious (see Th. 1.1.42).

The subspace $\mathrm{Ker}(l)$ can usually be determined directly from the definition. As far as $\mathrm{Im}(l)$ is concerned, the following result can often be applied.

3.1.10. Proposition. Let $l\colon L \to L'$ be a linear transformation. Let $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ be a basis of L. Then $\mathrm{Im}(l) = \mathrm{Span}\{l(\vec{b}_1), l(\vec{b}_2), \ldots, l(\vec{b}_n)\}$. If, moreover, $\mathrm{Ker}(l) = \vec{o}$, then $\{l(\vec{b}_1), l(\vec{b}_2), \ldots, l(\vec{b}_n)\}$ is a basis of $\mathrm{Im}(l)$.

Proof: Let $\vec{y} \in \mathrm{Im}(l)$. Then there is a vector $\vec{x} \in L$ such that $l(\vec{x}) = \vec{y}$. Since B is a basis of L, we can write $\vec{x} = \beta_1 \circ \vec{b}_1 \sharp \beta_2 \circ \vec{b}_2 \sharp \ldots \sharp \beta_n \circ \vec{b}_n$. Thus,

$$\begin{aligned} \vec{y} = l(\vec{x}) &= l(\beta_1 \circ \vec{b}_1 \sharp \beta_2 \circ \vec{b}_2 \sharp \ldots \sharp \beta_n \circ \vec{b}_n) \\ &= \beta_1 \circ' l(\vec{b}_1) \sharp' \beta_2 \circ' l(\vec{b}_2) \sharp' \ldots \sharp' \beta_n \circ' l(\vec{b}_n), \end{aligned}$$

which gives $\vec{y} \in \mathrm{Span}\{l(\vec{b}_1), l(\vec{b}_2), \ldots, l(\vec{b}_n)\}$.

Suppose now that $\mathrm{Ker}(l) = \overrightarrow{o}$. We have to show that the collection $\{l(\overrightarrow{b}_1), l(\overrightarrow{b}_2), \dots, l(\overrightarrow{b}_n)\}$ is LI. Let us assume that $\alpha_1 \circ' l(\overrightarrow{b}_1) \sharp' \alpha_2 \circ' l(\overrightarrow{b}_2) \sharp' \dots \sharp' \alpha_n \circ' l(\overrightarrow{b}_n) = \overrightarrow{o}'$. We want to show that $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$. Since l is linear, we can rewrite the above equality as follows:

$$l(\alpha_1 \circ \overrightarrow{b}_1 \sharp \alpha_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp \alpha_n \circ \overrightarrow{b}_n) = \overrightarrow{o}'.$$

But $\mathrm{Ker}(l) = \overrightarrow{o}$, and therefore $\alpha_1 \circ \overrightarrow{b}_1 \sharp \alpha_2 \circ \overrightarrow{b}_2 \sharp \dots \sharp \alpha_n \circ \overrightarrow{b}_n = \overrightarrow{o}$. We have assumed that $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \dots, \overrightarrow{b}_n\}$ is LI. Thus, $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$. □

3.1.11. Example. Let $l\colon R^3 \to R^3$ be the mapping defined by the following formula:

$$l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 4x_1 + 5x_2 + 6x_3, x_1 + x_2 + x_3).$$

Find $\mathrm{Ker}(l)$ and $\mathrm{Im}(l)$ (resp., find $d(l)$ and $r(l)$).

Solution: By the definition of $\mathrm{Ker}(l)$, we see that $\overrightarrow{x} \in \mathrm{Ker}(l) \Leftrightarrow l(\overrightarrow{x}) = \overrightarrow{o}$. Writing $\overrightarrow{x} = (x_1, x_2, x_3)$, we have $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 4x_1 + 5x_2 + 6x_3, x_1 + x_2 + x_3)$. Thus, $\overrightarrow{x} = (x_1, x_2, x_3) \in \mathrm{Ker}(l) \Leftrightarrow (x_1 + 2x_2 + 3x_3, 4x_1 + 5x_2 + 6x_3, x_1 + x_2 + x_3) = (0, 0, 0)$. Using the matrix form and the elimination procedure, we consecutively obtain

$$\left\|\begin{array}{ccc|c} 1, & 2, & 3 & 0 \\ 4, & 5, & 6 & 0 \\ 1, & 1, & 1 & 0 \end{array}\right\| \sim \left\|\begin{array}{ccc|c} 1, & 2, & 3 & 0 \\ 0, & -3, & -6 & 0 \\ 0, & 1, & 2 & 0 \end{array}\right\| \sim \left\|\begin{array}{ccc|c} 1, & 2, & 3 & 0 \\ 0, & 1, & 2 & 0 \\ 0, & 0, & 0 & 0 \end{array}\right\|$$

It follows that $(x_1, x_2, x_3) = \alpha(1, -2, 1)$. Thus, $\mathrm{Ker}(l) = \mathrm{Span}\{(1, -2, 1)\}$. Using the previous result, we know that

$$\mathrm{Im}(l) = \mathrm{Span}\{l(\overrightarrow{b}_1), l(\overrightarrow{b}_2), \dots, l(\overrightarrow{b}_n)\},$$

where $\{\overrightarrow{b}_1, \overrightarrow{b}_2, \overrightarrow{b}_3\}$ is a basis of L. Taking e. g. the standard basis of R^3, $\overrightarrow{b}_1 = (1, 0, 0)$, $\overrightarrow{b}_2 = (0, 1, 0)$, $\overrightarrow{b}_3 = (0, 0, 1)$, we have

$$\begin{aligned} l(1, 0, 0) &= (1, 4, 1) \\ l(0, 1, 0) &= (2, 5, 1) \\ l(0, 0, 1) &= (3, 6, 1). \end{aligned}$$

Consider the space $\mathrm{Span}\{(1, 4, 1), (2, 5, 1), (3, 6, 1)\}$. Let us find a basis of this space (this leads us back to problems studied in Chapter 1). Using elimination and the symbol $\sim$ to denote the coincidence of Spans, we obtain

$$\begin{array}{ccc} (1, & 4, & 1) \\ (2, & 5, & 1) \\ (3, & 6, & 1) \end{array} \sim \begin{array}{ccc} (1, & 4, & 1) \\ (0, & -3, & -1) \\ (0, & -6, & -2) \end{array} \sim \begin{array}{ccc} (1, & 4, & 1) \\ (0, & 3, & 1) \end{array}.$$

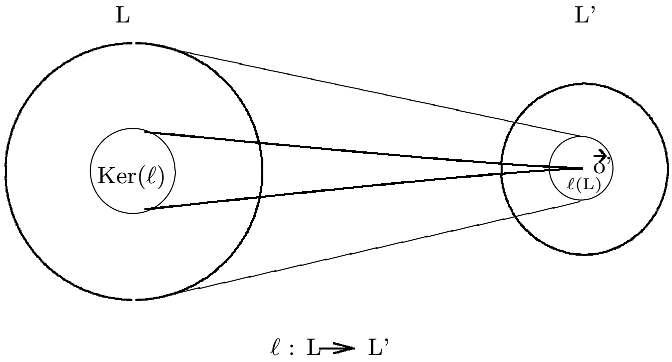
Thus, $\mathrm{Im}(l) = \mathrm{Span}\{(1, 4, 1), (0, 3, 1)\}$. As a consequence, $d(l) = 1$ and $r(l) = 2$.

The following result reveals a remarkable interplay between $d(l)$ and $r(l)$. One should not overlook the corollary to this theorem, Theorem 3.1.14, which brings an elegant proof to Th. 2.5.6 on the set of solutions of a homogeneous system.

3.1.12. Theorem Let $l\colon L \to L'$ be a linear transformation and let the space L be finite dimensional. Then

$$\dim L = d(l) + r(l).$$

Proof: The situation is illustrated by the figure below.



Suppose first that $d(l) = 0$. Then $\mathrm{Ker}(l) = \overrightarrow{o}$ and the result follows from the second part of Prop. 3.1.10. Suppose therefore that $d(l) > 0$. Put $d(l) = p$. Let $\{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_p\}$ be a basis of $\mathrm{Ker}(l)$. This basis can be completed to a basis of the entire space L (see the "important observation" after Prop. 1.1.41). Let $\dim L = n$ and let $\{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_p, \overrightarrow{a}_{p+1}, \dots, \overrightarrow{a}_n\}$ be a basis of L. We will show that the set $\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\}$ is a basis of $\mathrm{Im}(l)$. This will prove the theorem.

We have to show that $\mathrm{Span}\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\} = \mathrm{Im}(l)$ and, also, we have to show that the set $\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\}$ is LI. In order to show that $\mathrm{Span}\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\} = \mathrm{Im}(l)$, let us assume that $\overrightarrow{z} \in \mathrm{Im}(l)$. Then there is a vector $\overrightarrow{x} \in L$ such that $l(\overrightarrow{x}) = \overrightarrow{z}$. Let us express the vector $\overrightarrow{x}$ in the basis $\{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_n\}$:

$$\overrightarrow{x} = x_1 \circ \overrightarrow{a}_1 \sharp\, x_2 \circ \overrightarrow{a}_2 \sharp \,\dots\, \sharp\, x_n \circ \overrightarrow{a}_n.$$

Then

$$\begin{aligned}
\overrightarrow{z} = l(\overrightarrow{x}) &= l(x_1 \circ \overrightarrow{a}_1 \sharp\, x_2 \circ \overrightarrow{a}_2 \sharp \,\dots\, \sharp\, x_n \circ \overrightarrow{a}_n) \\
&= x_1 \circ' l(\overrightarrow{a}_1) \sharp' x_2 \circ' l(\overrightarrow{a}_2) \sharp' \dots \sharp' x_n \circ' l(\overrightarrow{a}_n) \\
&= x_{p+1} \circ' l(\overrightarrow{a}_{p+1}) \sharp' x_{p+2} \circ' l(\overrightarrow{a}_{p+2}) \sharp' \dots \sharp' x_n \circ' l(\overrightarrow{a}_n)
\end{aligned}$$

(the vectors $\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_p$ belong to $\mathrm{Ker}(l)$ and therefore $l(\overrightarrow{a}_i) = \overrightarrow{o}'$ for any i, $i \le p$). Thus, $\overrightarrow{z} \in \mathrm{Span}\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\}$.

We now want to show that the set $\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \dots, l(\overrightarrow{a}_n)\}$ is LI. Suppose that $\alpha_{p+1} \circ' l(\overrightarrow{a}_{p+1}) \sharp' \alpha_{p+2} \circ' l(\overrightarrow{a}_{p+2}) \sharp' \dots \sharp' \alpha_n \circ' l(\overrightarrow{a}_n) = \overrightarrow{o}'$. This implies that

$$l(\alpha_{p+1} \circ \overrightarrow{a}_{p+1} \sharp\, \alpha_{p+2} \circ \overrightarrow{a}_{p+2} \sharp \,\dots\, \sharp\, \alpha_n \circ \overrightarrow{a}_n) = \overrightarrow{o}'$$

and therefore the vector $\alpha_{p+1} \circ \overrightarrow{a}_{p+1} \sharp\, \alpha_{p+2} \circ \overrightarrow{a}_{p+2} \sharp \,\dots\, \sharp\, \alpha_n \circ \overrightarrow{a}_n$ belongs to $\mathrm{Ker}(l)$. Thus,

$$\alpha_{p+1} \circ \overrightarrow{a}_{p+1} \sharp\, \alpha_{p+2} \circ \overrightarrow{a}_{p+2} \sharp \,\dots\, \sharp\, \alpha_n \circ \overrightarrow{a}_n = \alpha_1 \circ \overrightarrow{a}_1 \sharp\, \alpha_2 \circ \overrightarrow{a}_2 \sharp \,\dots\, \sharp\, \alpha_p \circ \overrightarrow{a}_p$$

for some coefficients $\alpha_1, \alpha_2, \dots, \alpha_p$. This gives the equality $\alpha_1 \circ \overrightarrow{a}_1 \sharp\, \alpha_2 \circ \overrightarrow{a}_2 \sharp \,\dots\, \sharp\, \alpha_p \circ \overrightarrow{a}_p \sharp\, (-\alpha_{p+1}) \circ \overrightarrow{a}_{p+1} \sharp \,\dots\, \sharp\, (-\alpha_n) \circ \overrightarrow{a}_n = \overrightarrow{o}$. Since the vectors $\{\overrightarrow{a}_1, \overrightarrow{a}_2, \dots, \overrightarrow{a}_n\}$

are LI, we see that $\alpha_1 = \alpha_2 = \cdots = \alpha_{p+1} = \cdots = \alpha_n = 0$. We have verified that if $\alpha_{p+1} \circ' l(\overrightarrow{a}_{p+1}) \sharp' \alpha_{p+2} \circ' l(\overrightarrow{a}_{p+2}) \sharp' \ldots \sharp' \alpha_n \circ' l(\overrightarrow{a}_n) = \overrightarrow{o}'$, then $\alpha_{p+1} = \cdots = \alpha_n = 0$. This means that the set $\{l(\overrightarrow{a}_{p+1}), l(\overrightarrow{a}_{p+2}), \ldots, l(\overrightarrow{a}_n)\}$ is LI which we were to prove. $\square$

3.1.13. Example. Consider the linear mapping $l\colon R^4 \to R^3$ defined by the following formula:

$$l(x_1, x_2, x_3, x_4) = (3x_1 - 2x_2 + 5x_3 + 4x_4, 6x_1 - 4x_2 + 4x_3 + 3x_4, 9x_1 - 6x_2 + 3x_3 + 2x_4).$$

Find $d(l)$ and $r(l)$.

Solution: Let us find $d(l)$. We will compute $\mathrm{Ker}(l)$ the way we did in Ex. 3.1.11. We obtain

$$\left\|\begin{array}{cccc|c} 3, & -2, & 5, & 4 & 0 \\ 6, & -4, & 4, & 3 & 0 \\ 9, & -6, & 3, & 2 & 0 \end{array}\right\| \sim \left\|\begin{array}{cccc|c} 3, & -2, & 5, & 4 & 0 \\ 0, & 0, & -6, & -5 & 0 \\ 0, & 0, & -12, & -10 & 0 \end{array}\right\| \sim \left\|\begin{array}{cccc|c} 3, & -2, & 5, & 4 & 0 \\ 0, & 0, & 6, & 5 & 0 \\ 0, & 0, & 0, & 0 & 0 \end{array}\right\|$$

We see that $d(l) = 2$ and, as a consequence of Th. 3.1.12, $r(l) = 2$.

Let us now take up the corollary advertised before. We will present another proof to the fundamental Th. 2.5.6.

3.1.14. Theorem (homogeneous systems revisited): Consider the homogeneous system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

Let $\mathbf{A} = \|a_{ij}\|, i \le m, j \le n$ ($\mathbf{A}$ is the matrix of the system). Then the set $\mathscr{R}$ of all solutions of this system forms a linear subspace of R^n and, moreover, $\dim \mathscr{R} = -(\mathbf{A})$.

Proof: We will show how the "moreover" part of the above theorem follows from Th. 3.1.12. It is a nice piece of linear algebra. Let us first observe that the matrix $\mathbf{A}$ represents a linear mapping $l\colon R^n \to R^m$ such that, for any $\overrightarrow{x} = (x_1, x_2, \ldots, x_n) \in R^n, l(\overrightarrow{x}) = \mathbf{A}\left\|\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right\|$.

In this representation, $\mathscr{R} = \mathrm{Ker}()$. According to Th. 3.1.12, $n = \dim \mathscr{R} + ()$. Thus, it remains to be shown that $r(l) = r(\mathbf{A})$. But this follows immediately from Prop. 3.1.10 and Th. 2.2.3. Indeed, if B is the standard basis of $R^n, B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \ldots, \overrightarrow{b_n}\} = \{(1, 0, 0, \ldots, 0), (0, 1, 0, \ldots), \ldots, (0, 0, 0, \ldots, 1)\}$, then

$$\begin{aligned} l(\overrightarrow{b_1}) &= (a_{11}, a_{21}, \ldots, a_{m1}) \\ l(\overrightarrow{b_2}) &= (a_{12}, a_{22}, \ldots, a_{m2}) \\ &\vdots \\ l(\overrightarrow{b_n}) &= (a_{1n}, a_{2n}, \ldots, a_{mn}) \end{aligned}$$

By Prop. 3.1.10, $r(l) = \dim \mathrm{Span}\{l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \ldots, l(\overrightarrow{b_n})\}$. But $\dim \mathrm{Span}\{l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \ldots, l(\overrightarrow{b_n})\} = r(\mathbf{A}^T)$. By Th. 2.2.3, $r(\mathbf{A}^T) = r(\mathbf{A})$, and therefore $r(l) = r(\mathbf{A})$. The proof of Th. 3.1.14 is complete. □

3.2. Linear transformations given by determining the images of bases

Suppose that $l\colon L \to L'$ is a linear transformation. Suppose that $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \ldots, \overrightarrow{b_n}\}$ is a basis of L. Since l preserves linear combinations (Prop. 3.1.4), it is obvious that if we know $l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \ldots, l(\overrightarrow{b_n})$, then we know $l(\overrightarrow{x})$ for any $\overrightarrow{x} \in L$. Indeed, if $\overrightarrow{x} = x_1 \circ \overrightarrow{b_1} \sharp x_2 \circ \overrightarrow{b_2} \sharp \ldots \sharp x_n \circ \overrightarrow{b_n}$, then $l(\overrightarrow{x}) = x_1 \circ l(\overrightarrow{b_1}) \sharp x_2 \circ l(\overrightarrow{b_2}) \sharp \ldots \sharp x_n \circ l(\overrightarrow{b_n})$. This simple but very useful observation allows us to define linear transformations "in bases".

3.2.1. Example. The linear transformation $l\colon R^3 \to \mathscr{M}^{22}$ is defined as follows:

$$l(1,2,3) = \begin{Vmatrix} 1, & 1 \\ 0, & -1 \end{Vmatrix}$$

$$l(1,2,0) = \begin{Vmatrix} 1, & 2 \\ 1, & 1 \end{Vmatrix}$$

$$l(1,0,0) = \begin{Vmatrix} 1, & -1 \\ 2, & 3 \end{Vmatrix}$$

Convince yourself that the definition is correct and then find $l(-2,-2,3)$.

SOLUTION: Since l is supposed to be linear, the definition is correct provided the collection $\{(1,2,3),(1,2,0),(1,0,0)\}$ constitutes a basis of R^3. The latter fact is easy to verify (for instance, $\det \begin{Vmatrix} 1 & 2 & 3 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{Vmatrix} = -6 \neq 0$). Let us now compute $l(-2,-2,3)$. We first have to express the vector to be transformed in the given basis $\{(1,2,3),(1,2,0),(1,0,0)\}$. In order to do so,let us write $(-2,-2,3) = \alpha_1(1,2,3) + \alpha_2(1,2,0) + \alpha_3(1,0,0)$. Thus,

$$\begin{aligned} -2 &= \alpha_1 + \alpha_2 + \alpha_3 \\ -2 &= 2\alpha_1 + 2\alpha_2 \\ 3 &= 3\alpha_1 \end{aligned}$$

and therefore $\alpha_1 = 1, \alpha_2 = -2$ and $\alpha_3 = -1$. Using linearity, we consecutively obtain

$$\begin{aligned} l(-2,-2,3) &= l(1 \circ (1,2,3) + (-2) \circ (1,2,0) + (-1) \circ (1,0,0)) \\ &= 1 \circ' l(1,2,3) - 2 \circ' l(1,2,0) - 1 \circ' l(1,0,0) \\ &= 1 \circ' \begin{Vmatrix} 1, & 1 \\ 0, & -1 \end{Vmatrix} - 2 \circ' \begin{Vmatrix} 1, 2 \\ 1, 1 \end{Vmatrix} - 1 \circ' \begin{Vmatrix} 1, & -1 \\ 2, & 3 \end{Vmatrix} \\ &= \begin{Vmatrix} -2, & -2 \\ -4, & -6 \end{Vmatrix}. \end{aligned}$$

Thus, $l(-2,-2,3) = \left\|\begin{matrix} -2, & -2 \\ -4, & -6 \end{matrix}\right\|$.

Since a linear transformation is fully determined by images of vectors of a basis, we in principle can compute $\mathrm{Ker}(l), \mathrm{Im}(l)$ for a transformation given "in a basis", too. Obviously, we have nothing at our disposal but linearity and the images of a given basis so we have to proceed accordingly.

3.2.2. Example. The linear transformation $l\colon R^4 \to R^3$ is defined as follows:

$$\begin{aligned} l(1,0,0,0) &= (1,1,-1) \\ l(1,-1,0,0) &= (1,0,1) \\ l(1,1,1,0) &= (2,1,0) \\ l(0,0,1,1) &= (4,3,-2) \end{aligned}$$

Find $\mathrm{Ker}(l)$ and $\mathrm{Im}(l)$.

SOLUTION: Strictly speaking, we would first have to check that the four vectors whose images are given constitute a basis of R^4. Let us assume that the problem is set up correctly (a suspicious reader may verify "the soundness" on his (her) own). Let us take up $\mathrm{Ker}(1)$. We want to characterize those vectors $\overrightarrow{x} \in R^4$ such that $\overrightarrow{x} \in \mathrm{Ker}(l)$. To do so, let us take an arbitrary vector $\overrightarrow{x} \in R^4$ and express it with respect to the given basis. Thus, $\overrightarrow{x} = x_1(1,0,0,0)+x_2(1,-1,0,0)+x_3(1,1,1,0)+x_4(0,0,1,1)$. (We will use the casual notation being sure that the reader cannot be confused.) We ask when $l(\overrightarrow{x}) = \overrightarrow{o}'$. Since

$$l(\overrightarrow{x}) = x_1 l(1,0,0,0) + x_2 l(1,-1,0,0) + x_3 l(1,1,1,0) + x_4 l(0,0,1,1),$$

we obtain

$$l(\overrightarrow{x}) = x_1(1,1,-1) + x_2(1,0,1) + x_3(2,1,0) + x_4(4,3,-2).$$

It follows that the equation $l(\overrightarrow{x}) = \overrightarrow{o}'$ reads

$$x_1(1,1,-1) + x_2(1,0,1) + x_3(2,1,0) + x_4(4,3,-2) = (0,0,0).$$

This is equivalent to the following system:

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

Let us solve this system. Following the usual method we obtain

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

It is easily seen that $(x_1,x_2,x_3,x_4) = \mathrm{Span}\{(-1,-1,1,0)(-3,-1,0,1)\}$. We are almost finished with this part of the problem, only we must not forget that (x_1,x_2,x_3,x_4) are *the coordinates of* $\overrightarrow{x}$ *with respect to the given basis*! Thus, if we want to express our vector $\overrightarrow{x}$ in the standard coordinates, we have to get back to its original setup. We have

$$\overrightarrow{x} = x_1(1,0,0,0) + x_2(1,-1,0,0) + x_3(1,1,1,0) + x_4(0,0,1,1).$$

Since
$(x_1, x_2, x_3, x_4) = \mathrm{Span}\{(-1,-1,1,0)\dot{(}-3,-1,0,1)\} = \{\alpha(-1,-1,1,0)+\beta(-3,-1,0,1) | \alpha \in R\}$,
we obtain ($\alpha \in R, \beta \in R$)

$$\begin{aligned}\overrightarrow{x} &= (-\alpha - 3\beta)(1,0,0,0) + (-\alpha - \beta)(1,-1,0,0) + \alpha(1,1,1,0) + \beta(0,0,1,1)\\ &= \alpha(-1,2,1,0) + \beta(-4,-1,1,1).\end{aligned}$$

Thus,

$$\mathrm{Ker}(l) = \mathrm{Span}\{(-1,2,1,0), (-4,-1,1,1)\} \text{ and } d(l) = 2.$$

Let us find $\mathrm{Im}(l)$. According to Prop. 3.1.10, $\mathrm{Im}(l) = \mathrm{Span}\{l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \ldots, l(\overrightarrow{b_n})\}$, where $\{\overrightarrow{b_1}, \overrightarrow{b_2}, \ldots, \overrightarrow{b_n}\}$ is an arbitrary basis of L. In our case, $\mathrm{Im}(l) = \mathrm{Span}\{(1,1,-1), (1,0,1), (2,1,0), (4,3,-2)\}$. Recalling the stability properties of Span under the elimination technique (Chap. 1), we have

$$\begin{matrix}(1, & 1, & -1)\\ (1, & 0, & 1)\\ (2, & 1, & 0)\\ (4, & 3, & -2)\end{matrix} \quad\sim\quad \begin{matrix}(1, & 1, & -1)\\ (0, & -1, & 2)\\ (0, & -1, & 2)\\ (0, & -1, & 2)\end{matrix} \quad\sim\quad \begin{matrix}(1, & 1, & -1)\\ (0, & -1, & 2)\\ (0, & 0, & 0)\\ (0, & 0, & 0)\end{matrix}$$

We see that

$$\mathrm{Im}(l) = \mathrm{Span}\{(1,1,-1), (0,-1,2)\}.$$

Thus, $h(l) = 2$.

3.3. The matrix of linear transformation with respect to bases

Let us start with the following definition which will turn out to be of central importance in the study of linear transformations. As peculiar as it may seem at first sight, it provides a useful technical tool for considering engineering problems. The approach is often referred to as matrix representation of linear transformations. (It is recommended that you read the definition first formally, then consult the comments written below the definition, then compute the examples that follow and then carefully read the definition with the comments again.)

3.3.1. Definition. Let $l\colon L \to L$ be a linear transformation. Let $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \ldots, \overrightarrow{b_n}\}$ be a basis of L and let $B' = \{\overrightarrow{b_1}', \overrightarrow{b_2}', \ldots, \overrightarrow{b_m}'\}$ be a basis of L'. Let

$$\begin{array}{lcl} l(\overrightarrow{b_1}) & = & a_{11} \circ' \overrightarrow{b_1}' \;\sharp'\; a_{21} \circ' \overrightarrow{b_2}' \;\sharp'\; \ldots \;\sharp'\; a_{m1} \circ' \overrightarrow{b_m}',\\ l(\overrightarrow{b_2}) & = & a_{12} \circ' \overrightarrow{b_1}' \;\sharp'\; a_{22} \circ' \overrightarrow{b_2}' \;\sharp'\; \ldots \;\sharp'\; a_{m2} \circ' \overrightarrow{b_m}',\\ \quad\vdots & & \\ l(\overrightarrow{b_n}) & = & a_{1n} \circ' \overrightarrow{b_1}' \;\sharp'\; a_{2n} \circ' \overrightarrow{b_2}' \;\sharp'\; \ldots \;\sharp'\; a_{mn} \circ' \overrightarrow{b_m}'.\end{array}$$

Then the matrix $\mathbf{A} = \|a_{ij}\| \in \mathscr{M}^{mn}$ is called *the matrix of the transformation l with respect to the bases B and B'*. We sometimes write $\mathbf{A} = \mathbf{A}(l, B, B')$ when we want to refer to the bases B, B' in question.

The definition above calls for some comments. First, it is worthwhile observing that the matrix $\mathbf{A} = \mathbf{A}(l, B, B')$ is correctly defined (the coefficients a_{ij} are determined uniquely!). Second, the components a_{ij} heavily depend upon the choice of the bases B and B'. And thirdly, it is important to realize that if we fix the bases B and B', every matrix $\mathbf{A} \in \mathscr{M}$ is a matrix of a uniquely defined linear transformation $l\colon L \to L'$. Thus, having B and B' fixed, *there is a one-to-one correspondence between linear transformations from L to L' and the matrices in $\mathscr{M}$.* We will discuss this correspondence later. (One should not brood too much about why the definition is taken "transpose-wise". The reason is essentially cosmetic and will be clear later from the formulation of Prop. 3.3.5 and Th. 3.3.7.)

The following examples illustrate the notion of matrix of l (we will again use the simplified notation instead of correctly writing $\circ, \sharp$ and $\circ', \sharp'$).

3.3.2. Example. Let $l\colon R \to R$ be the linear transformation which is determined as follows: If $\overrightarrow{x} \in R, \overrightarrow{x} = x$, then $l(\overrightarrow{x}) = 5x$. Let $B = \{7\}$ and $B' = \{\frac{1}{3}\}$. Find the matrix of l with respect to B and B'.

SOLUTION: In this case $B = \{\overrightarrow{b_1}\} = \{7\}$ and $B' = \{\overrightarrow{b_1'}\} = \frac{1}{3}$. According to Definition 3.3.1, we have to find $l(\overrightarrow{b_1})$ and express it with respect to B'. Formally, we write $l(\overrightarrow{b_1}) = a_{11} \circ' \overrightarrow{b_1}'$, and $\mathbf{A} = \|a_{11}\|$ will be the matrix of l with respect to B and B'. Thus, $l(\overrightarrow{b_1}) = l(7) = 5 \cdot 7 = 35 = a_{11} \cdot \frac{1}{3}$. It follows that $a_{11} = 105$. We see that $\mathbf{A} = \|105\| \in \mathscr{M}$ is the matrix we were to find.

3.3.3. Example. Let a linear transformation $l\colon R^3 \to R^2$ be defined by the following formula:

$$l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 + x_2, x_2 - x_3).$$

Let $B = \{(1,0,0), (0,-1,1), (1,1,1)\}$ be a basis of R^3 and let $B' = \{(1,1), (1,2)\}$ be a basis of R^2. Find the matrix of l with respect to B and B'.

SOLUTION: According to the definition of the matrix of l with respect to B and B', we have to find the images under l of the vectors of B and then express these images with respect to the basis B'. Writing $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \overrightarrow{b_3}\}$ and $B' = \{\overrightarrow{b_1}', \overrightarrow{b_2}'\}$, we obtain

$$\begin{aligned}
l(\overrightarrow{b_1}) &= l(1, \quad 0, 0) = (\ \ 1, \quad 0) = a_{11} \overrightarrow{b_1}' + a_{21} \overrightarrow{b_2}' = a_{11}(1,1) + a_{21}(1,2) \\
l(\overrightarrow{b_2}) &= l(0, -1, 1) = (-1, -2) = a_{12} \overrightarrow{b_1}' + a_{22} \overrightarrow{b_2}' = a_{12}(1,1) + a_{22}(1,2) \\
l(\overrightarrow{b_3}) &= l(1, \quad 1, 1) = (\ \ 2, \quad 0) = a_{13} \overrightarrow{b_1}' + a_{23} \overrightarrow{b_2}' = a_{13}(1,1) + a_{23}(1,2).
\end{aligned}$$

It is now usually handy to use the matrix formalism to compute the coefficients a_{ij} ($i \leq 2$, $j \leq 3$). Our equations can be equivalently expressed as the following matrix equation:

$$\begin{Vmatrix} 1, & -1, & 2 \\ 0, & -2, & 0 \end{Vmatrix} = \begin{Vmatrix} 1, & 1 \\ 1, & 2 \end{Vmatrix} \begin{Vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \end{Vmatrix}.$$

This gives us

$$\begin{Vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \end{Vmatrix} = \begin{Vmatrix} 1, & 1 \\ 1, & 2 \end{Vmatrix}^{-1} \begin{Vmatrix} 1, & -1, & 2 \\ 0, & -2, & 0 \end{Vmatrix} = \begin{Vmatrix} 2, & -1 \\ -1, & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 1, & -1, & 2 \\ 0, & -2, & 0 \end{Vmatrix} = \begin{Vmatrix} 2, & 0, & 4 \\ -1, & -1, & -2 \end{Vmatrix}.$$

Thus, $\mathbf{A}(l,B,B') = \begin{Vmatrix} 2, & 0, & 4 \\ -1, & -1, & -2 \end{Vmatrix}$.

3.3.4. Example. Let the linear mapping $l\colon R^2 \to R^3$ be defined as follows:

$$l(1,1) = (0,1,-1) \qquad\qquad l(1,-1) = (2,1,-1)$$

Find the matrix of this mapping with respect to the bases $B = \{(0,2),(3,1)\}$ and $B' = \{(1,0,-1),(1,1,2),(0,1,1)\}$.

SOLUTION: We first have to find the images $l(0,2)$ and $l(3,1)$. According to the definition of l – the transformation l is determined "in the basis $\{(1,1),(1,-1)\}$" – we set

$$\begin{aligned}(0,2) &= \alpha_{11}(1,1) + \alpha_{12}(1,-1)\\ (3,1) &= \alpha_{21}(1,1) + \alpha_{22}(1,-1),\end{aligned}$$

which means the matrix equation $\begin{Vmatrix} 0, & 2 \\ 3, & 1 \end{Vmatrix} = \begin{Vmatrix} \alpha_{11}, & \alpha_{12} \\ \alpha_{21}, & \alpha_{22} \end{Vmatrix} \begin{Vmatrix} 1, & 1 \\ 1, & -1 \end{Vmatrix}$. Thus,

$$\begin{Vmatrix} \alpha_{11}, & \alpha_{12} \\ \alpha_{21}, & \alpha_{22} \end{Vmatrix} = \begin{Vmatrix} 0, & 2 \\ 3, & 1 \end{Vmatrix} \begin{Vmatrix} 1, & 1 \\ 1, & -1 \end{Vmatrix}^{-1} = \begin{Vmatrix} 0, & 2 \\ 3, & 1 \end{Vmatrix} \cdot \left(-\frac{1}{2}\right) \begin{Vmatrix} -1, & -1 \\ -1, & 1 \end{Vmatrix} = \begin{Vmatrix} 1, & -1 \\ 2, & 1 \end{Vmatrix}.$$

We have obtained

$$\begin{aligned}(0,2) &= 1(1,1) - 1(1,-1),\\ (3,1) &= 2(1,1) + 1(1,-1).\end{aligned}$$

Using linearity of l, we conclude that

$$\begin{aligned}l(0,2) &= 1\cdot l(1,1) - 1\cdot l(1,-1) = 1(0,1,-1) - 1(2,1,-1) = (-2,0,0)\\ l(3,1) &= 2\cdot l(1,1) + 1\cdot l(1,-1) = 2(0,1,-1) + 1(2,1,-1) = (2,3,-3).\end{aligned}$$

We will now take up the second step – we express the images of vectors of B in the basis B'. We write

$$\begin{aligned}l(0,2) = (-2,0,0) &= a_{11}(1,0,-1) + a_{21}(1,1,2) + a_{31}(0,1,1)\\ l(3,1) = (2,3,-3) &= a_{12}(1,0,-1) + a_{22}(1,1,2) + a_{32}(0,1,1).\end{aligned}$$

The above equations are again best solved in the matrix formulation. In order to obtain immediately the matrix of the transformation l, the appropriate matrix reformulation of the equations reads

$$\begin{Vmatrix} -2, & 2 \\ 0, & 3 \\ 0, & -3 \end{Vmatrix} = \begin{Vmatrix} 1, & 1, & 0 \\ 0, & 1, & 1 \\ -1, & 2, & 1 \end{Vmatrix} \begin{Vmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \\ a_{31}, & a_{32} \end{Vmatrix}.$$

Multiplying with the inverse in the standard way, we obtain

$$\begin{Vmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \\ a_{31}, & a_{32} \end{Vmatrix} = \begin{Vmatrix} 1, & 1, & 0 \\ 0, & 1, & 1 \\ -1, & 2, & 1 \end{Vmatrix}^{-1} \begin{Vmatrix} -2, & 2 \\ 0, & 3 \\ 0, & -3 \end{Vmatrix}.$$

Thus,

$$\begin{Vmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \\ a_{31}, & a_{32} \end{Vmatrix} = \left(-\frac{1}{2}\right) \begin{Vmatrix} -1, & -1, & 1 \\ -1, & 1, & -1 \\ 1, & -3, & 1 \end{Vmatrix} \begin{Vmatrix} -2, & 2 \\ 0, & 3 \\ 0, & -3 \end{Vmatrix} = \begin{Vmatrix} -1, & 4 \\ -1, & -2 \\ 1, & 5 \end{Vmatrix}.$$

Answer: The matrix $\mathbf{A}$ of l with respect to B and B' is

$$\mathbf{A} = \begin{Vmatrix} -1, & 4 \\ -1, & -2 \\ 1, & 5 \end{Vmatrix}.$$

The following result shows how the matrix of transformation may be used in computing the *geometric* notion of rank. Recall that the rank of l, $r(l)$, $\dim \mathrm{Im}(l)$ by definition.

3.3.5. Theorem Let $l\colon L \to L'$ be a linear transformation, let $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \dots, \overrightarrow{b_n}\}$ and $B' = \{\overrightarrow{b_1}', \overrightarrow{b_2}', \dots, \overrightarrow{b_n}'\}$ be bases of L and L'. Let $\mathbf{A}$ be the matrix of l with respect to B and B'. Then $r(l) = r(\mathbf{A})$. Corollaries:

1. The rank of the matrix of a linear transformation does not depend upon the choice of the bases B and B' ($r(A)$ is a so called invariant).

2. The transformation $l\colon L \to L'$ is a linear isomorphism if and only if $\mathbf{A}$ is regular.

PROOF: Let $\mathbf{A} = \|a_{ij}\|, i \le m, j \le n$, be the matrix of l with respect to B and B'. By Theorem 1.1.52, Cor. a), $\dim \mathrm{Span}\{l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \dots, l(\overrightarrow{b_n})\} = \dim \mathrm{Span}\{(a_{11}, a_{21}, \dots, a_{m1}), (a_{12}, a_{22}, \dots, a_{m2}), \dots, (a_{1n}, a_{2n}, \dots, a_{mn})\}$. But $\dim \mathrm{Span}\{l(\overrightarrow{b_1}), l(\overrightarrow{b_2}), \dots, l(\overrightarrow{b_n})\} = r(l)$ (Prop. 3.1.10) and $\dim \mathrm{Span}\{(a_{11}, a_{21}, \dots, a_{m1}), (a_{12}, a_{22}, \dots, a_{m2}), \dots, (a_{1n}, a_{2n}, \dots, a_{mn})\} = r(\mathbf{A}^T)$. And $r(\mathbf{A}^T) = r(\mathbf{A})$ (Th. 2.2.3) and the proof is complete. □

The following technical result shows how the matrix of linear mapping acts in the "input–output" process presented by a linear transformation.

3.3.6. Proposition. Let $l\colon L \to L'$ be a linear transformation. Let $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \dots, \overrightarrow{b_n}\}$ be a basis of L and let $B' = \{\overrightarrow{b_1}', \overrightarrow{b_2}', \dots, \overrightarrow{b_m}'\}$ be a basis of L'. Let $\mathbf{A} = \mathbf{A}(l, B, B') = \|a_{ij}\| \in \mathscr{M}$ be the matrix of l with respect to B and B'. Let $\overrightarrow{x} \in L$ and $\overrightarrow{x} = x_1 \circ \overrightarrow{b_1} \sharp x_2 \circ \overrightarrow{b_2} \sharp \dots \sharp x_n \overrightarrow{b_n}$. Let $l(\overrightarrow{x}) = \overrightarrow{y} = y_1 \circ' \overrightarrow{b_1}' \sharp' y_2 \circ' \overrightarrow{b_2}' \sharp' \dots \sharp' y_m \circ' \overrightarrow{b_m}'$. Then

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j \qquad (i \le m).$$

In other words, if $(x_1, x_2, \dots, x_n)$ are coordinates of $\overrightarrow{x}$ with respect to B and if $(y_1, y_2, \dots, y_n)$ are coordinates of $l(\overrightarrow{x})$ with respect to B', then $\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Proof: Computing $l(\overrightarrow{x})$, we consecutively obtain

$$\begin{aligned}
l(\overrightarrow{x}) &= l(x_1 \circ \overrightarrow{b_1} \sharp x_2 \circ \overrightarrow{b_2} \sharp \dots \sharp x_n \circ \overrightarrow{b_n}) = x_1 \circ' l(\overrightarrow{b_1}) \sharp' x_2 \circ' l(\overrightarrow{b_2}) \sharp' \dots \sharp' x_n \circ' l(\overrightarrow{b_n}) \\
&= x_1 \circ' (a_{11} \circ' \overrightarrow{b'_1} \sharp' a_{21} \circ' \overrightarrow{b'_2} \sharp' \dots \sharp' a_{m1} \circ' \overrightarrow{b'_m}) \\
&\quad \sharp' x_2 \circ' (a_{12} \circ' \overrightarrow{b'_1} \sharp' a_{22} \circ' \overrightarrow{b'_2} \sharp' \dots \sharp' a_{m2} \circ' \overrightarrow{b'_m}) \sharp' \dots \\
&\quad \sharp' x_n \circ' (a_{1n} \circ' \overrightarrow{b'_1} \sharp' a_{2n} \circ' \overrightarrow{b'_2} \sharp' \dots \sharp' a_{mn} \circ' \overrightarrow{b'_m}) \\
&= (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) \circ \overrightarrow{b'_1} \sharp' (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n) \circ \overrightarrow{b'_2} \sharp' \dots \\
&\quad \sharp' (a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n) \circ \overrightarrow{b'_m} \\
&= y_1 \circ' \overrightarrow{b'_1} \sharp' y_2 \circ' \overrightarrow{b'_2} \sharp' \dots \sharp' y_m \circ' \overrightarrow{b'_m}.
\end{aligned}$$

Comparing now the corresponding coefficients, we obtain the required formula

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \qquad (i \le m).$$

□

3.3.7. Example. Consider the following linear transformation $l\colon R^3 \to \mathscr{P}$, ($\mathscr{P}$ stands for the linear space of all polynomials whose degrees are less than or equal to 1). The transformation is given by its matrix: The matrix

$$\mathbf{A} = \left\| \begin{matrix} 0, & \frac{1}{2}, & -\frac{3}{2} \\ -1, & -\frac{1}{2}, & \frac{3}{2} \end{matrix} \right\|$$

is known to be the matrix of l with respect to B and B', where

$$B = \{(0,0,1), (0,1,1), (1,-1,-1)\}, \quad B' = \{1-x, 1+x\}.$$

Find $l(-1,3,4)$ and $\operatorname{Ker}(l)$.

Solution: We will use the previous Prop. 3.3.4 (we could work the example using the definition of the matrix of l only, of course). We first have to find the coordinates of $(-1,3,4)$ with respect to B. We set

$$(-1,3,4) = x_1(0,0,1) + x_2(0,1,1) + x_3(1,-1,-1),$$

obtaining $x_1 = 1, x_2 = 2$ and $x_3 = -1$. Using the formula derived in Prop. 3.3.5,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

we see that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \left\| \begin{matrix} 0, & \frac{1}{2}, & -\frac{3}{2} \\ -1, & -\frac{1}{2}, & \frac{3}{2} \end{matrix} \right\| \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{5}{2} \\ -\frac{7}{2} \end{pmatrix}.$$

It follows that the coordinates of $l(-1,3,4)$ with respect to B' are $\left(\frac{5}{2}, -\frac{7}{2}\right)$. This means that

$$l(-1,3,4) = \frac{5}{2}(1-x) - \frac{7}{2}(1+x) = -1-6x.$$

The second part of this problem will again be approached via Prop. 3.3.4. Since the vector $(0,0)$ has coordinates $(0,0)$ with respect to B', we see that $\overrightarrow{x} \in \mathrm{Ker}(l)$ if and only if its coordinates with respect to B, (x_1, x_2, x_3), satisfy the equation

$$\mathbf{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This implies that

$$\mathbf{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \left\| \begin{matrix} 0, & \frac{1}{2}, & -\frac{3}{2} \\ -1, & -\frac{1}{2}, & \frac{3}{2} \end{matrix} \right\| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} & & \frac{1}{2}x_2 & - & \frac{3}{2}x_3 \\ -x_1 & - & \frac{1}{2}x_2 & + & \frac{3}{2}x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives the system

$$\begin{array}{rcrcrcl} & & \frac{1}{2}x_2 & - & \frac{3}{2}x_3 & = & 0 \\ -x_1 & - & \frac{1}{2}x_2 & + & \frac{3}{2}x_3 & = & 0 \end{array}.$$

This system has a one-parameter solution which can easily be guessed: $(x_1, x_2, x_3) = \mathrm{Span}\{(0,3,1)\}$. Thus,

$$\mathrm{Ker}(l) = \mathrm{Span}\{(0\cdot(0,0,1) + 3(0,1,1) + 1(1,-1,-1)\} = \mathrm{Span}\{(1,2,2)\}.$$

The following result moves the analogy of linear transformations with matrices further. The proof of the result is rather technical but essentially simple. For the sake of relative simplicity, let us agree to compose the transformations "*from the right to the left*", i.e. let us write $l_2 \bullet l_1 \colon L \to L''$ to denote the composite transformation of the mappings $l_1 \colon L \to L'$ and $l_2 \colon L' \to L''$. This convention is customary in algebra (as we know, in analysis we usually compose the other way round).

3.3.8. Theorem Let $L = (L, \sharp, \circ), L' = (L', \sharp', \circ')$ and $L'' = (L'', \sharp'', \circ'')$ be linear spaces. Let $l_1 \colon L \to L'$ and $l_2 \colon L' \to L''$ be linear transformations. Let $B = \{\overrightarrow{b_1}, \overrightarrow{b_2}, \dots, \overrightarrow{b_n}\}, B' = \{\overrightarrow{b_1}', \overrightarrow{b_2}', \dots, \overrightarrow{b_m}'\}$ and $B'' = \{\overrightarrow{b_1}'', \overrightarrow{b_2}'', \dots, \overrightarrow{b_p}''\}$ be bases of L, L' and L'', respectively. Let $\mathbf{A}_1 = \mathbf{A}_1(l_1, B, B') \in \mathscr{M}$ and $\mathbf{A}_2 = \mathbf{A}_2(l_2, B', B'') \in \mathscr{M}$ be the respective matrices of the transformations l_1 and l_2. Then

> the matrix $\mathbf{A}_2\mathbf{A}_1$ is the matrix of the composite transformation $l_2 \bullet l_1$ with respect to the bases B and B''.

Proof: Let us write $\mathbf{A}_1 = \|\bar{a}_{ij}\|, i \le m, j \le n, \mathbf{A}_2 = \|\bar{\bar{a}}_{rs}\|, r \le p, s \le m$. By our assumption,

$$l_1(\overrightarrow{b}_j) = \bar{a}_{1j} \circ' \overrightarrow{b_1}' \ \sharp' \ \bar{a}_{2j} \circ' \overrightarrow{b_2}' \ \sharp' \ \dots \ \sharp' \ \bar{a}_{mj} \circ' \overrightarrow{b_m}' \ (j \le n),$$
$$l_2(\overrightarrow{b}'_s) = \bar{\bar{a}}_{1s} \circ'' \overrightarrow{b_1}'' \ \sharp'' \ \bar{\bar{a}}_{2s} \circ'' \overrightarrow{b_2}'' \ \sharp'' \ \dots \ \sharp'' \ \bar{\bar{a}}_{ps} \circ'' \overrightarrow{b_p}'' \ (s \le m).$$

Then we have

$$\begin{aligned}(l_2 \bullet l_1)(\overrightarrow{b}_j) &= l_2(\bar{a}_{1j} \circ' \overrightarrow{b_1}' \,\sharp\, '\bar{a}_{2j} \circ' \overrightarrow{b_2}' \,\sharp\, ' \ldots \,\sharp\, '\bar{a}_{mj} \circ' \overrightarrow{b_m}')\\
&= \bar{a}_{1j} \circ'' l_2(\overrightarrow{b_1}') \,\sharp\, ''\bar{a}_{2j} \circ'' l(\overrightarrow{b_2}') \,\sharp\, '' \ldots \,\sharp\, ''\bar{a}_{mj} \circ'' l(\overrightarrow{b_m}')\\
&= \bar{a}_{1j} \circ'' \left(\bar{\bar{a}}_{11} \circ'' \overrightarrow{b}''_1 \,\sharp\, ''\bar{\bar{a}}_{21} \circ'' \overrightarrow{b}''_2 \,\sharp\, '' \ldots \,\sharp\, ''\bar{\bar{a}}_{p1} \circ'' \overrightarrow{b}''_p\right) \,\sharp\, ''\\
&\qquad \bar{a}_{2j} \circ'' \left(\bar{\bar{a}}_{12} \circ'' \overrightarrow{b}''_1 \,\sharp\, ''\bar{\bar{a}}_{22} \circ'' \overrightarrow{b}''_2 \,\sharp\, '' \ldots \,\sharp\, ''\bar{\bar{a}}_{p2} \circ'' \overrightarrow{b}''_p\right) \,\sharp\, ''\\
&\quad \vdots\\
&\qquad \bar{a}_{mj} \circ'' \left(\bar{\bar{a}}_{1m} \circ'' \overrightarrow{b}''_1 \,\sharp\, ''\bar{\bar{a}}_{2m} \circ'' \overrightarrow{b}''_2 \,\sharp\, '' \ldots \,\sharp\, ''\bar{\bar{a}}_{pm} \circ'' \overrightarrow{b}''_p\right)\\
&= (\bar{a}_{1j}\bar{\bar{a}}_{11} + \bar{a}_{2j}\bar{\bar{a}}_{12} + \cdots + \bar{a}_{mj}\bar{\bar{a}}_{1m}) \circ'' \overrightarrow{b}''_1 \,\sharp\, ''\\
&\qquad (\bar{a}_{1j}\bar{\bar{a}}_{21} + \bar{a}_{2j}\bar{\bar{a}}_{22} + \cdots + \bar{a}_{mj}\bar{\bar{a}}_{2m}) \circ'' \overrightarrow{b}''_2 \,\sharp\, ''\\
&\quad \vdots \quad (\bar{a}_{1j}\bar{\bar{a}}_{p1} + \bar{a}_{2j}\bar{\bar{a}}_{p2} + \cdots + \bar{a}_{mj}\bar{\bar{a}}_{pm}) \circ'' \overrightarrow{b}''_p.\end{aligned}$$

It follows that if we denote by $\mathbf{A}$ the matrix of the composite transformation $l_2 \bullet l_1$ with respect to B and B'' we see that $\mathbf{A} = \mathbf{A}_2\mathbf{A}_1$ (the coefficients of the last linear combination of the vectors $\overrightarrow{b}''_1, \overrightarrow{b}''_2, \ldots, \overrightarrow{b}''_p$ are exactly the components of the j-th column of $\mathbf{A}_2\mathbf{A}_1$). The proof is complete. □

A direct consequence of Theorem 3.3.8 is the following result.

3.3.9. Theorem Let $l\colon L \to L'$ be a linear transformation and let B and B' be bases of L and L'. Let $\mathbf{A}$ be the matrix of l with respect to B and B'. Suppose that l is an isomorphism (i.e., suppose that l is both injective and surjective). Then the inverse mapping $l^{-1}\colon L' \to L$ is also a linear transformation and the matrix of l^{-1} with respect to B' and B is $\mathbf{A}^{-1}$.

Proof: Let us first check that l^{-1} is linear. Suppose that $\overrightarrow{y}_1, \overrightarrow{y}_2 \in L'$. Then there are *unique* elements $\overrightarrow{x}_1, \overrightarrow{x}_2 \in L$ such that $l(\overrightarrow{x}_1) = \overrightarrow{y}_1$ and $l(\overrightarrow{x}_2) = \overrightarrow{y}_2$. Thus, $l^{-1}(\overrightarrow{y}_1) = \overrightarrow{x}_1$, and $l^{-1}(\overrightarrow{y}_2) = \overrightarrow{x}_2$. Since $l(\lambda \circ \overrightarrow{x}_1) = \lambda \circ' \overrightarrow{y}_1$, and $l(\overrightarrow{x}_1 \,\sharp\, \overrightarrow{x}_2) = \overrightarrow{y}_1 \,\sharp\, '\overrightarrow{y}_2$ and since l is injective, we see that $l^{-1}(\lambda \circ' \overrightarrow{y}_1) = \lambda \circ \overrightarrow{x}_1 = \lambda \circ l^{-1}(\overrightarrow{y}_1)$ and $l^{-1}(\overrightarrow{y}_1 \,\sharp\, '\overrightarrow{y}_2) = \overrightarrow{x}_1 \,\sharp\, \overrightarrow{x}_2 = l^{-1}(\overrightarrow{y}_1) \,\sharp\, l^{-1}(\overrightarrow{y}_2)$. It follows that l is a linear transformation.

The rest follows from Theorem 3.3.8. Denote by $\tilde{\mathbf{A}}$ the matrix of l^{-1} with respect to B' and B. Since both l and l^{-1} are linear transformations and since $l^{-1} \cdot l = \mathrm{id}$ ($\mathrm{id} : L \to L$ denotes the identity mapping), we see that $\tilde{\mathbf{A}}\mathbf{A} = \mathbf{E}$ (the unit matrix $\mathbf{E}$ is obviously the matrix of $\mathrm{id} : L \to L$ with respect to B). This means that $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$ and the proof is complete. □

In the series of previous results we saw a striking similarity in dealing with transformations and matrices. Let us express this similarity in precise terms. Let L and L' be linear spaces and let $\dim L = n, \dim L' = m$. Let $\mathscr{T}(\mathscr{L}, \mathscr{L}')$ denote the set of all linear transformations from L into L'. The set $\mathscr{T}(\mathscr{L}, \mathscr{L}')$ can naturally be given the structure of a linear space. Indeed, if $l\colon L \to L'$ and $k\colon L \to L'$ are elements of $\mathscr{T}(\mathscr{L}, \mathscr{L}')$, then we define $(l \,\tilde{\sharp}\, k)\colon L \to L'$ by setting $(l \,\tilde{\sharp}\, k)(\overrightarrow{x}) = l(\overrightarrow{x}) \,\sharp'\, k(\overrightarrow{x})$ for any $\overrightarrow{x} \in L$. Analogously, we define $\lambda \tilde{\circ} l\colon L \to L'$ by setting $(\lambda \tilde{\circ} l)(\overrightarrow{x}) = \lambda \circ' l(\overrightarrow{x})$ for any $\overrightarrow{x} \in L$. It is obvious that the triple $(\mathscr{T}(\mathscr{L}, \mathscr{L}'), \tilde{\sharp}, \tilde{\circ})$ is a linear space. We will simply denote this space by $\mathscr{T}(\mathscr{L}, \mathscr{L}')$. The harvest of the former results, which we now formulate in a single theorem, establishes the full equivalence of the space $\mathscr{T}(\mathscr{L}, \mathscr{L}')$ with the space $\mathscr{M}$:

3.3.10. Theorem Let $\dim L = n$ and $\dim L' = m$. Then there is a linear isomorphism $i\colon \mathscr{L}(\mathscr{L}, \mathscr{L}') \to \mathscr{M}$. Moreover, this isomorphism enjoys the following properties:

1. If $l \in \mathscr{L}(\mathscr{L}, \mathscr{L}')$, then $r(l) = r(i(l))$, ($i(l)$ is the "image matrix" under the mapping i),
2. If $l \in \mathscr{L}(\mathscr{L}, \mathscr{L}')$ and l is an isomorphism, then $i(l^{-1}) = (i(l))^{-1}$ (the latter $^{-1}$ means the matrix inverse),
3. If L'' is a finite dimensional space, $\dim L'' = p$, then there is a linear isomorphism $j\colon \mathscr{L}(\mathscr{L}', \mathscr{L}'') \to \mathscr{M}$ such that, for any $l_1 \in \mathscr{L}(\mathscr{L}, \mathscr{L}')$ and $l_2 \in \mathscr{L}(\mathscr{L}', \mathscr{L}'')$ the following identity holds true:

   $$(j \cdot i)(l_2 \cdot l_1) = j(l_2)i(l_1)$$

   (the right-hand side means the matrix multiplication).

PROOF: Choose bases B, B' and B'' of the respective spaces L, L' and L''. Let the mapping i assign to any l its matrix with respect to B and B', and let the mapping j assign to any $k \in \mathscr{L}(\mathscr{L}', \mathscr{L}'')$ its matrix with respect to B' and B''. Then the mappings i and j are obviously linear and satisfy all the properties of Theorem 3.3.10 (this follows from Theorem 3.3.5 and Theorem 3.3.9). □

Isomorphic spaces (only dimension matters!)

In this paragraph we explicitly formulate one important fact which the reader may have guessed on his/her own. The fact is that the linear spaces of equal dimension are identical as far as "linear" considerations are concerned.

3.3.11. Theorem. Let L and L' be finite dimensional linear spaces. Then the following statements are equivalent:

1. The spaces L and L' are isomorphic (i.e., there is a linear isomorphism $l\colon L \to L'$),
2. $\dim L = \dim L'$.

PROOF: Suppose that there is a linear isomorphism $l\colon L \to L'$. Since l is injective, $\dim \mathrm{Im}(l) = \dim L$ (Prop. 3.1.10). Since l is surjective, $\mathrm{Im}(l) = L'$. Thus, $\dim L' = \dim L$.

Suppose on the other hand that $\dim L = \dim L'$. Let $B = \{\overrightarrow{b}_1, \overrightarrow{b}_2, \ldots, \overrightarrow{b}_n\}$ be a basis of L and $B' = \{\overrightarrow{b}'_1, \overrightarrow{b}'_2, \ldots, \overrightarrow{b}'_n\}$ be a basis of L'. Let $l\colon L \to L'$ be the linear transformation which is determined by the equalities $l(\overrightarrow{b}_i) = \overrightarrow{b}'_i$ $(i \leq n)$. Then l is obviously an isomorphism (Prop. 3.1.3 and 3.1.10). The proof is complete. □

The above result establishes that the *linear spaces of equal dimension are intrinsically undistinguishable*. Thus, the space R^n $(n \in N)$ may well serve as a model of any n-dimensional space. The reader then obviously wonders why we bothered so much with so many examples of finite dimensional spaces and why we dealt so thoroughly with the theory of *general* finite dimensional spaces (Chap. 1). The exclusive study of R^n may seem sufficient and more digestible for a beginner.

The answer is that we wanted to address the student who would further proceed with his individual study of mathematical methods of modern engineering. His/her quest for expertise will naturally require a postgraduate study. A more in-depth study obviously needs a full grasp of basic algebraic methods. It is *the general algebraic aspect* that would partially vanish if we restricted ourselves to R^n. Secondly, in more advanced applications of linear algebra we often equip a linear space with other operations (matrix multiplication in $\mathscr{M}$, the vector product of geometric vectors, differentiation in the linear space of differentiable functions, etc.) This involves individual properties of vectors and brings new phenomena to the game. We could hardly manage to deal with all the "parameters" without having *the principles* of linearity in our blood.

Miscellaneous examples

Example 1. Check which of the following mappings are linear transformations

a) $l\colon R^2 \to R^2,\ l(x_1, x_2) = (|x|, |y|)$

b) $l\colon R^2 \to R^2,\ l(x_1, x_2) = (0, x_1x_2)$

c) $l\colon R^2 \to R^2,\ l(x_1, x_2) = (x_1 + x_2, 2x_1 - x_2)$.

Use the definition of linear transformation.

Solution: a) Take, for instance, $\overrightarrow{x} = (1,1)$ and $\overrightarrow{y} = (-1,-1)$. Then $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(0,0) = (0,0)$, but $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y}) = (1,1) + (1,1) = (2,2)$. This mapping is not linear.

b) Take, for instance, $\overrightarrow{x} = (1,1)$ and $\overrightarrow{y} = (1,1)$. Then $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(2,2) = (0,4)$, but $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y}) = (0,1) + (0,1) = (0,2)$. This mapping is not linear.

c) Consider two vectors $\overrightarrow{x} = (x_1, x_2)$ and $\overrightarrow{y} = (y_1, y_2)$. Then $l(\overrightarrow{x} \sharp \overrightarrow{y}) = l(x_1 + y_1, x_2 + y_2) = ((x_1 + y_1) + (x_2 + y_2), 2(x_1 + y_1) - (x_2 + y_2))$. We must check that $l(\overrightarrow{x}) \sharp l(\overrightarrow{y}) = l(\overrightarrow{x}) \sharp' l(\overrightarrow{y})$. Compute $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y})$. We obtain $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y}) = (x_1 + x_2, 2x_1 - x_2) + (y_1 + y_2, 2y_1 - y_2) = ((x_1 + x_2) + (y_1 + y_2), (2x_1 - x_2) + (2y_1 - y_2))$. We see that $l(\overrightarrow{x}) \sharp' l(\overrightarrow{y}) = l(\overrightarrow{x}) \sharp l(\overrightarrow{y})$.

Consider now $l(\lambda \circ \overrightarrow{x})$. We obtain

$$l(\lambda \circ \overrightarrow{x}) = l(\lambda x_1, \lambda x_2) = \lambda x_1 + \lambda x_2, 2\lambda x_1 - \lambda x_2).$$

We must check $l(\lambda \circ \overrightarrow{x}) = \lambda \circ' l(\overrightarrow{x})$. Compute $\lambda \circ' l(\overrightarrow{x})$. We obtain $\lambda \circ' l(\overrightarrow{x}) = \lambda(x_1 + 2x_2, 2x_1 - x_2) = (\lambda(x_1 + 2x_2), \lambda(2x_1 - x_2))$. We see that $l(\lambda \circ \overrightarrow{x}) = \lambda \circ' l(\overrightarrow{x})$. Summarizing all we have checked, we conclude that l is a linear transformation.

Example 2. Let the linear transformation $l\colon R^2 \to R^4$ be given by the following conditions:

$$l(1,-1) = (1,1,2,-1)$$
$$l(0,1) = (0,1,1,2).$$

Find a basis of $\mathrm{Ker}(l)$ and $\mathrm{Im}(l)$. Compute $d(l)$ and $r(l)$. Find the matrix of this transformation with respect to B and B', where B is the standard basis in R^2 and B' is the standard basis in R^4.

Solution: Suppose that $\overrightarrow{x} \in \mathrm{Ker}(l)$. Write $\overrightarrow{x} = x_1(1,-1) + x_2(0,1)$. Then $l(\overrightarrow{x}) = l(x_1(1,-1) + x_2(0,1)) = x_1(1,1,2,-1) + x_2(0,1,1,2) = (0,0,0,0)$. Since the vectors $(1,1,2,-1)$,

$(0,1,1,2)$ are obviously linearly independent, we infer that $x_1 = x_2 = 0$. Thus, $\mathrm{Ker}(l) = \{\overrightarrow{o}\}$. This (trivial) space does not have a basis. Looking for basis of $\mathrm{Im}(l)$, we see immediately (Prop. 4.1.10), that $\{(1,1,2,-1),(0,1,1,2)\}$ does the job. It follows that $d(l) = 0$ and $r(l) = 2$.

Let us compute the matrix $\mathbf{A} = (l, B, B')$, of l. We have $B = \{(1,0),(0,1)\}$. According to the definition of l, we obtain

$$l(1,0) = l(1(1,-1) + 1(0,1)) = 1(1,1,2,-1) + 1(0,1,1,2) = (1,2,3,1),$$
$$l(0,1) = l(0(1,-1) + 1(0,1)) = 0(1,1,2,-1) + 1(0,1,1,2) = (0,1,1,2).$$

Since $B' = \{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\}$, we immediately see that the desired matrix $\mathbf{A}$ reads

$$\mathbf{A} = \begin{Vmatrix} 1, & 0 \\ 2, & 1 \\ 3, & 1 \\ 1, & 2 \end{Vmatrix}.$$

Example 3. Let us suppose that a linear transformation $l\colon R^3 \to R^3$ be subject to the following requirements:

$$l(1,1,0) = (1,0,-1)$$
$$l(0,-1,1) = (0,2,1), \quad \mathrm{Ker}(l) = \mathrm{Span}\{(1,1,1)\}.$$

Convince yourself that by these three condition the transformation is completely determined. Then find $l(4,2,3)$.

SOLUTION: We know that $l(1,1,1) = (0,0,0)$. Since the vectors $(1,1,0),(0,-1,1),(1,1,1)$ are LI, for $\det \begin{Vmatrix} 1, & 1, & 0 \\ 0, & -1, & 1 \\ 1, & 1, & 1 \end{Vmatrix} \neq 0$, we see that l is determined "in vectors of a basis". Looking for $l(4,2,3)$, we first express the vector $(4,2,3)$ in the basis we can map. We have $(4,2,3) = x_1(1,1,0) + x_2(0,-1,1) + x_3(1,1,1)$. Solving the equation, which is a matter of simple algebra, we obtain $x_1 = 3, x_2 = 2, x_1 = 1$. Thus,

$$l(4,2,3) = 3 \cdot l(1,1,0) + 2 \cdot (0,-1,1) + 1 \cdot l(1,1,1) = 3(1,0,-1) + 2(0,2,1) + 1(1,1,1) = (4,5,0).$$

Example 4. Let $B = \{(1,1,0),(0,-1,1),(0,0,1)\}$ and $B' = \{(0,-1),(1,1)\}$. Let $l\colon R^3 \to R^2$ be given by its matrix, $\mathbf{A}$, with respect to B and B'. Let $\mathbf{A} = \begin{Vmatrix} 1, & 0, & 2 \\ -1, & 1, & 3 \end{Vmatrix}$. Compute $l(1,0,0)$.

SOLUTION: As one checks easily, $(1,0,0) = 1(1,1,0) + 1(0,-1,1) + (-1)(0,0,1)$. Since we know that

$$l(1,1,0) = 1(0,-1) - 1(1,1) = (-1,-2)$$
$$l(0,-1,1) = 0(0,-1) + 1(1,1) = (1,1)$$
$$l(0,0,1) = 2(0,-1) + 3(1,1) = (3,1)$$

we see that

$$l(1,0,0) = 1l(1,1,0) + 1l(0,-1,1) + (-1)(0,0,1) = 1(-1,-2) + 1(1,1) + (-1)(3,1) = (-3,-2).$$

Thus, $l(1,0,0) = (-3,-2)$.

Example 5. Suppose that $\mathbf{A} = \mathbf{A}(l, B, N)$, $l\colon R^m \to R^n$ is a linear transformation. Show that there exist a matrix,

$$\mathbf{A} = \begin{Vmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & & & \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} \end{Vmatrix} \quad \in \quad \mathscr{M}',$$

so that $l(x_1, x_2, \dots, x_m) = \|x_1, x_2, \dots, x_m\| = \mathbf{A}$. Can we associate this matrix $\mathbf{A}$ with a matrix of l with respect to bases (compare with Th. 3.3.8)?

SOLUTION: Take an arbitrary vector $\overrightarrow{x} = (x_1, x_2, \dots, x_m)$. Then $\overrightarrow{x} = x_1(1,0,0,\dots,0) + x_2(0,1,0,\dots,0) + \cdots + x_m(0,0,0,\dots,0,1)$. Suppose that

$$l(1,0,0,\dots,0) = (a_{11}, a_{12}, \dots, a_{1n}), l(0,1,0,\dots,0) = (a_{21}, a_{22}, \dots, a_{2n}), \dots, l(0,0,\dots,0,1) \\ = (a_{m1}, a_{m2}, \dots, a_{mn}).$$

Since l is linear, we easily see that

$$l(\overrightarrow{x}) = x_1(a_{11}, a_{12}, \dots, a_{1n}) + x_2(a_{21}, a_{22}, \dots, a_{2n}) + \cdots + x_m(a_{m1}, a_{m2}, \dots, a_{mn}).$$

When we express this equation in the matrix form, we see that $l(\overrightarrow{x}) = \|x_1, x_2, \dots, x_m\|\mathbf{A}$. This is what we were to show. (As regard the link of this matrix $\mathbf{A}$ with Th. 3.3.8, it is easily seen that $\mathbf{A} = \mathbf{A}^T$, where $\mathbf{A}$ is the matrix of l with respect to the standard bases B and B' respectively.)

Problems

(A problem indicated with * is supposed to be a challenge for an interested student)

1. Find out which of the following mappings are linear transformations
   a) $l\colon R^2 \to R^3, l(\overrightarrow{x}) = l(x_1, x_2) = (2x_1 - x_2, x_1, x_2)$
   b) $l\colon R^3 \to R^2, l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 - x_2 + x_3, 5)$
   c) $l\colon R^2 \to R^3, l(\overrightarrow{x}) = l(x_1, x_2) = (x_1, x_1x_2, x_2^2)$
   d) $l\colon \mathscr{M}' \to \mathscr{P}, (\overrightarrow{x}) = \left(\begin{Vmatrix} x_{11}, & x_{12} \\ x_{21}, & x_{22} \end{Vmatrix}\right) = \ + \ + \ -$
   e) $l\colon C\langle 0,1\rangle \to C\langle 0,1\rangle, l(\overrightarrow{x}) = l(f(t)) = \int_0^t f(p)\mathrm{d}p = g(t)$.

   (Answer: a) linear, b) not linear, c) not linear, d) linear, e) linear.)

2. Let $l\colon V_2 \to V_2$, where V_2 is the linear space of all free geometric vectors of E_2, be defined as follows (φ is a given angle, $\varphi \in \langle 0, 2\pi\rangle$):

$$l(\overrightarrow{v}) = l(v_1, v_2) = (v_1 \cos\varphi - v_2 \sin\varphi,\ v_1 \sin\varphi + v_2 \cos\varphi).$$

Is l linear? What is the geometric meaning of this mapping?

(Answer: l is linear; this mapping presents the rotation of $\overrightarrow{v}$ with the angle φ)

3. Check that the following mappings $l\colon R^3 \to R^3$ are linear transformations. Find $\mathrm{Ker}(l)$ and decide whether l is an isomorphism.

   a) $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3,\ 4x_1 + 5x_2 + 6x_3,\ x_1 + x_2 + x_3)$

   b) $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (x_2 + x_3,\ 2x_1 + x_3,\ 3x_1 - x_2 + x_3)$

   (Answers: a) $\mathrm{Ker}(l) = \mathrm{Span}\{(1, -2, 1)\}$; l is not an isomorphism; b) $\mathrm{Ker}(l) = (0, 0, 0)$; l is an isomorphism.)

4. The linear transformation $l\colon R^3 \to R$ is defined by the formula $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = 2x_1 - x_2 + 3x_3$. Find a basis of $\mathrm{Ker}(l)$. What is $d(l)$ and $r(l)$?

   (Answer: the set $\{(1, 2, 0), (3, 0, -2)\}$ is a basis of $\mathrm{Ker}(l)$; $d(l) = 2, r(l) = 1$).

5. * Suppose that the mapping $l\colon R^n \to R$ is linear. Prove that there are coefficients $a_1, a_2, \dots, a_n$ such that, for any $\overrightarrow{x} \in R^n$, $l(\overrightarrow{x}) = l(x_1, x_2, \dots, x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n$.

6. Suppose that the linear transformation $l\colon R^2 \to R^2$ is determined as follows:

$$\begin{aligned} l(2, \quad 1) &= (1, 1), \\ l(2, -2) &= (2, 2). \end{aligned}$$

   Find $l(3, 3)$ and also find $d(l)$ and $r(l)$.

   (Answer: $l(3, 3) = (1, 1); d(l) = 1, r(l) = 1$)

7. Suppose that the linear transformation $l\colon R^3 \to R$ is determined as follows:

$$\begin{aligned} l(1, 0, 0) &= 1, \\ l(1, 4, 0) &= -1, \\ l(0, 0, 1) &= 1. \end{aligned}$$

   Find $\mathrm{Ker}(l), d(l)$ and $r(l)$.

   (Answer: $\mathrm{Ker}(l) = \mathrm{Span}\{(0, 2, 1), (1, 2, 0)\}, d(l) = 2, r(l) = 1$)

8. Suppose that the linear transformation $l\colon R^3 \to R^2$ is determined as follows:

$$\begin{aligned} l(4, 0, -1) &= (1, 10), \\ l(1, 2, \quad 3) &= (14, 32), \\ l(3, -1, 1) &= (4, 13). \end{aligned}$$

   Find a basis of $\mathrm{Ker}(l)$.

   (Answer: $(1, -2, 1)$)

9. The linear transformation $l\colon \mathscr{M}' \to \mathscr{R}$ is determined as follows:

$$l\left(\begin{Vmatrix} 1, & 1 \\ -1, & 2 \end{Vmatrix}\right) = (1, 1, -1), \ l\left(\begin{Vmatrix} 0, & 1 \\ 2, & -1 \end{Vmatrix}\right) = (2, 0, 1),$$

$$l\left(\begin{Vmatrix} 0, & 0 \\ 1, & 2 \end{Vmatrix}\right) = (3, 1, 0), \quad l\left(\begin{Vmatrix} 1, & 1 \\ -1, & 3 \end{Vmatrix}\right) = (-1, 1, -2).$$

Find a basis of $\mathrm{Ker}(l)$.

$$\left(\text{Answer:}\ \left\{\left\|\begin{matrix}-1, & -2\\ 0, & 1\end{matrix}\right\|, \left\|\begin{matrix}1, & 3\\ 2, & -1\end{matrix}\right\|\right\}.\right)$$

10. Suppose that $l\colon R^3 \to R^3$ is the linear transformation that satisfies the following conditions:

$$\begin{aligned} l(1,1,1) &= (2,1,2),\\ l(3,0,2) &= (6,0,3),\\ \mathrm{Ker}(l) &= \mathrm{Span}\{(0,0,1)\}. \end{aligned}$$

Find $l(2,3,4)$.

(Answer: $l(2,3,4) = (4,3,5)$)

11. Suppose that $l\colon R^3 \to R^3$ is the linear transformation that satisfies the following conditions:

$$\begin{aligned} l(1,1,0) &= (0,-1,1),\\ \mathrm{Ker}(l) &= \mathrm{Span}\{(1,0,0),(2,2,3)\}. \end{aligned}$$

Find $l(0,0,3)$ and $r(l)$.

(Answer: $l(0,0,3) = (0,2,-2)$, $r(l) = 1$)

12. Suppose that $l\colon L \to L'$ is a linear transformation. Prove the following propositions:
    a) The transformation maps LD families of vectors in L into LD families of vectors in L'.
    b) The following equivalence holds true: l is injective $\Leftrightarrow$ l maps LI families of vectors in L into LI families of vectors in L'.
13. The linear transformation $l\colon \mathscr{P} \to \mathscr{P}$ is defined by the following formula:

$$l(a_0 + a_1x + a_2x^2 + a_3x^3) = a_1 + 2a_2x + 3a_3x^2.$$

Find $\mathrm{Ker}(l)$ and $r(l)$. (Observe that l can alternatively be defined as $l(f) = f'$ where f' denotes the derivative of f.)

(Answer: $\mathrm{Ker}(l) = \mathrm{Span}\{1\}, r(l) = 3$)

14. * Let L, L' be finite dimensional linear spaces. Let $\dim L \geq L'$. Prove: For any linear subspace M of L' there is a linear mapping $l\colon L \to L'$ such that $l(L) = M$ (i.e. $\mathrm{Im}(l) = M$). (Hint: Consider bases of L and M.)
15. * Let L be a linear space. A linear transformation $l\colon L \to L$ is called a projection if $l \bullet l = l$ ($l \bullet l$ means the composition of l with itself). Prove: If L is finite dimensional and M is a linear subspace of L, then there is a projection $l\colon L \to L$ such that $l(L) = M$ (i.e. $\mathrm{Im}(l) = M$).
16. Let $l\colon R^3 \to R^3$ be the linear transformation defined by the formula $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (0, x_1, 2x_2)$. Show that the composition mapping $l^3 = l \bullet l \bullet l\colon R^3 \to R^3$ is identically zero (i.e. $l^3(\overrightarrow{x}) = 0$ for any $\overrightarrow{x} \in R^3$).
17. It is assumed in this example that the reader is familiar with the construction of the scalar product and the vector product in V_3. (For a detailed treatment of these constructions, see Chap. 4. Use the coordinate expression for the scalar and vector product.)

a) Let $\overrightarrow{a} \in V_3$ be a given vector, $\overrightarrow{a} \neq \overrightarrow{0}$. Let $l\colon V_3 \to R$ be the mapping defined by the formula $l(\overrightarrow{x}) = \overrightarrow{a} \cdot \overrightarrow{x}$, where $\overrightarrow{a} \cdot \overrightarrow{x}$ denotes the scalar product. Show that l is linear and find $\mathrm{Ker}(l), d(l)$ and $r(l)$.

b) Let $\overrightarrow{a} \in V_3$ be a given vector, $\overrightarrow{a} \neq 0$. Let $l\colon V_3 \to V_3$ be the mapping defined by the formula $l(\overrightarrow{x}) = \overrightarrow{a} \times \overrightarrow{x}$ where $\overrightarrow{a} \times \overrightarrow{x}$ denotes the vector product. Show that l is linear and find $\mathrm{Ker}(l), d(l)$ and $r(l)$.

(Answer: a) $\mathrm{Ker}\, l = \{\overrightarrow{x} | \overrightarrow{x}$ is perpendicular to $\overrightarrow{a}\}, d(l) = 2, r(l) = 1$; b) $\mathrm{Ker}\, l = \mathrm{Span}\, \overrightarrow{a}, d(l) = 1, r(l) = 2$.)

18. Let $l\colon R^3 \to R^2$ be the linear transformation defined by the formula $l(\overrightarrow{x}) = l(x_1, x_2, x_3) = (2x_1 - x_3, x_2 + 4x_3)$. Find the matrix $\mathbf{A}$ of l with respect to the bases $B = \{(1,1,0), (0,2,0), (0,-1,1)\}$ and $B' = \{(0,1),(1,1)\}$.

$$\left(\text{Answer: } \mathbf{A} = \begin{Vmatrix} -1, & 2, & 4 \\ 2, & 0, & -1 \end{Vmatrix}\right).$$

19. Let $l\colon R \to R^2$ be the linear transformation defined by the formula $l(\overrightarrow{x}) = l(x) = (3x, -2x)$. Find the matrix $\mathbf{A}$ of l with respect to $B = \{-\dfrac{2}{3}\}$ and $B' = \{(1,-1),(1,0)\}$.

$$\left(\text{Answer: } \mathbf{A} = \begin{Vmatrix} -\frac{4}{3} \\ -\frac{2}{3} \end{Vmatrix}\right).$$

20. Let $l\colon R^3 \to R$ be the linear transformation defined as follows:

$$\begin{aligned} l(1,-1,0) &= -2, \\ l(0,\ \ 2,1) &= \ \ 1, \\ l(1,\ \ 0,1) &= \ \ 3. \end{aligned}$$

Find the matrix $\mathbf{A}$ of l with respect to the following bases:

a) $B = \{(0,2,1),(1,0,1),(1,-1,0)\}, B' = \{\frac{1}{2}\}$,

b) B is the standard basis of R^3 and B' is the standard basis of R.

(Answer: a) $\mathbf{A} = \|2, 6, -4\|$, b) $\|-6, -4, 9\|$)

21. The linear transformation $l\colon R^3 \to R^3$ is defined by the formula $l(\overrightarrow{x}) = l(x,y,z) = (2x - z, x + y + z, -y + z)$. Show that l is an isomorphism and find l^{-1}. Find the matrix $\mathbf{A}$ of l with respect to standard bases (i.e., both B and B' are the standard bases of R^3). Check that $\mathbf{A}^{-1}$ is the matrix of l^{-1} with respect to standard bases.

(Answer: $\mathrm{Ker}(l) = \overrightarrow{o} \Rightarrow l$ is an isomorphism; $l^{-1}(\overrightarrow{x}) = l^{-1}(x,y,z) =$

$$\left(-\frac{1}{5}(2x - 2y - 2z), -\frac{1}{5}(x - 2y - 3z), -\frac{1}{5}(x - 2y - 2z)\right), \mathbf{A} = \begin{Vmatrix} 2, & 0, & -1 \\ 1, & 1, & 1 \\ 0, & -1, & 1 \end{Vmatrix}\Bigg)$$

22. Consider two linear transformations $l\colon R^3 \to R^3$ and $h\colon R^3 \to R^3$. They are defined as follows:

$$\begin{aligned} l(\overrightarrow{x}) &= l(x_1, x_2, x_3) = (x_1 + x_3, 2x_2 + x_3, 2x_2) \\ k(\overrightarrow{y}) &= k(y_1, y_2, y_3) = (y_1 + 4y_2 + 3y_3, y_1 + 2y_2, \ \ -y_2 - y_3). \end{aligned}$$

Find the composite mapping $k \bullet l \colon R^3 \to R^3$ (we compose "l first, k afterwords"). Convince yourself that $k \bullet l$ is an isomorphism. Find $(k \bullet l)^{-1}$. Find the matrix $\mathbf{A}$ of $k \bullet l$ with respect to standard bases. Use $\mathbf{A}^{-1}$ to find $l^{-1}(20, 8, -5)$.

$\Big($Answer: $(k \bullet l)(\overrightarrow{x}) = (k \bullet l)(x_1, x_2, x_3) = (x_1 + 14x_2 + 5x_3, x_1 + 4x_2 + 3x_3, -4x_2 - x_3)$; $\mathrm{Ker}(k \bullet l) = \overrightarrow{o} \Rightarrow k \bullet l$ is an isomorphism; $(k \bullet l)^{-1}(\overrightarrow{z}) = (4z_1 - 3z_2 + 11z_3, \frac{1}{2}z_1 - \frac{1}{2}z_2 + z_3, -2z_1 + 2z_2 - 5z_3)$; $\mathbf{A} = \left\| \begin{matrix} 1, & 14, & 5 \\ 1, & 4, & 3 \\ 0, & -4, & -1 \end{matrix} \right\|$; $l^{-1}(20, 8, 5) = (1, 1, 1).\Big)$

Chapter 4

The linear space V_3 of “free” vectors and applications in geometry

In this chapter we make a short study of vectors in the three dimensional space. The principal aim is to introduce basic notions and constructions to be used in the calculus of more variables and in applications. The reader is assumed to have been acquainted with geometry in the extent of a high school curriculum.

4.1. Testing one's memory – how does one pass from “geometric vectors” to “free” vectors?

Let us adopt a rather practical attitude by assuming that we know the entity (set, space) that surrounds us. This may not be perfectly justified and rigorous (in fact, it may sound quite vulgar to many mathematicians and philosophers) but we feel we can fulfil the objective we have set up in this text reasonably well. Let us denote our “three-dimensional universum” by E_3 and call E_3 the (point) *Euclidean three space* (or, shortly, a three space). Let us call an ordered pair (A, B) of points in E_3 a *geometric vector*. It can be said that a geometric vector is na oriented segment in E_3 (see Fig. 4.1.1 below). If a geometric vector (A, B) is given, we can consider the set (class) of all geometric vectors which are parallel with (A, B) and have an equal magnitude and orientation. Let us denote this set by $\overrightarrow{AB}$ and call it a *free vector*. In other words, let us denote by $\overrightarrow{AB}$ the set of all (A', B') such that the centres of the segments AB' and $A'B$ coincide (see Fig. 4.1.1 again).

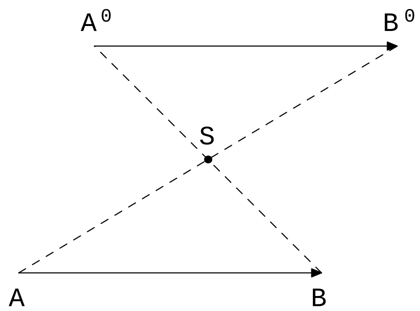


Figure 1:

Let us denote the set of all $\overrightarrow{AB}$ $(A, B \in E_3)$ by V_3. Let us now observe that V_3 can be naturally endowed with a structure of linear space. This is seen quite easily as soon as we accommodate to the strange nature of the elements in V_3 .These elements are sets. We therefore have to cope with a rather unusual problem of "operating" with sets as elements. Let us verify that the operations $\sharp$ and $\circ$ required for making V_3 a linear space can be simply defined by extending standard geometric operations $+$ and $\cdot$ on the geometric vectors in E_3. Indeed, let us first agree to call $\overrightarrow{AA}$ ($A \in E_3, A$ arbitrary) the null vector in V_3 (this vector is denoted by $\overrightarrow{o}$). The free vector $\overrightarrow{BA}$ is defined to be the inverse vector to $\overrightarrow{AB}$. Thus, $\overrightarrow{BA} = -\overrightarrow{AB}$. Let us see that this natural extension works. Before, let us agree to call free vectors simply vectors and use the customary notion $\overrightarrow{u}$, $\overrightarrow{v}$, etc., for these vectors.

Let us first take up the operation $\sharp$. Suppose we are given two vectors in V_3, $\overrightarrow{u}$ and $\overrightarrow{v}$. The (free) vector $\overrightarrow{u} \sharp \overrightarrow{v}$ will be defined as follows. Choose an arbitrary geometric vector (AB) so that $\overrightarrow{AB} = \overrightarrow{u}$ and then take a point $C \in E_3$ such that $\overrightarrow{BC} = \overrightarrow{v}$. Let us set $\overrightarrow{u} \sharp \overrightarrow{v} = \overrightarrow{AC}$ (see Fig 4.1.2 below). Obviously, it must be verified that the definition of $\overrightarrow{u} \sharp \overrightarrow{v}$ is correct, i.e., it must be shown that the definition does not depend upon the choice of A. This will be demonstrated later.

As for the "external" operation $\circ$, we can define it geometrically, too. Suppose $\overrightarrow{u} \in V$ and $\lambda \in R$. If $\overrightarrow{u} = \overrightarrow{o}$, we put $\lambda \circ \overrightarrow{u} = \overrightarrow{o}$. If $\overrightarrow{u} \neq \overrightarrow{o}$, let us choose $A \in E_3$ such that $\overrightarrow{AB} = \overrightarrow{u}$. Consider the geometric vector (A, B). It is easy to see that there is exactly one point, C, on the straight line determined by the points A, B such that

(i) $|(A, C)| = |\lambda||(A, B)|$ (the symbol $|\ \ |$ denotes the length),

(ii) if $\lambda \geq 0$, then C lies on the half-line AB (A being the origin); if $\lambda < 0$, then C does not lie on the half-line AB. Let us set $\lambda \circ \overrightarrow{u} = \overrightarrow{AC}$ (see Fig. 4.1.2).

Figure 2:

The foregoing definition of $\overrightarrow{u} \sharp \overrightarrow{v}$ and $\lambda \circ \overrightarrow{u}$ suffers a considerable weakness – it is based on the choice of concrete geometric representation of $\overrightarrow{u}$. What happens if the procedure of defining $\overrightarrow{u} \sharp \overrightarrow{v}$ and $\lambda \circ \overrightarrow{u}$ is performed with a different initial representation of $\overrightarrow{u}$? We are going to show that the resulting vectors, i.e., the resulting sets of geometric vectors $\overrightarrow{u} \sharp \overrightarrow{v}$ and $\lambda \circ \overrightarrow{u}$ will be always the same, giving the proof of correctness of our definition. We will also recall the coordinates of points in E_3.

Let us choose a Cartesian coordinate system in E_3. That is, let us agree on what the length unit is and let us choose three mutually perpendicular unit geometric vectors, $\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k}$, which are situated at a given point P. We will obtain standard coordinate axes with the origin in P (see Fig. 4.1.3 below).

Let $X \in E_3$. Then X can be naturally assigned a triple of real numbers, $[x_1, x_2, x_3]$, which are called *the coordinates of* X *in the coordinate system* $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$. These coordinates $[x_1, x_2, x_3]$ are determined as follows: The geometric vector (P, X) can be uniquely expressed as a (geometric) linear combination of the geometric vectors $\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k}$ and $[x_1, x_2, x_3]$ is the ordered triple for which we can write $(P, X) = x_1 \overrightarrow{i} + x_2 \overrightarrow{j} + x_3 \overrightarrow{k}$.

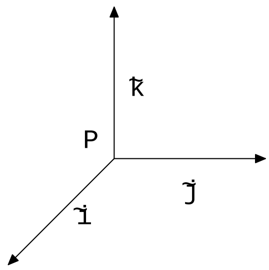


Figure 3:

Suppose that two geometric vectors $(A, B), (A', B')$ define the same free vector $\overrightarrow{AB}$. This means that the centres of the segments $\overline{AB'}$ and $\overline{A'B}$ coincide. Expressed in coordinates, if we write $A = [a_1, a_2, a_3], B = [b_1, b_2, b_3], A' = [a'_1, a'_2, a'_3]$ and $B' = [b'_1, b'_2, b'_3]$, we see that

$$\left[\frac{a_1 + b'_1}{2}, \frac{a_2 + b'_2}{2}, \frac{a_3 + b'_3}{2}\right] = \left[\frac{a'_1 + b_1}{2}, \frac{a'_2 + b_2}{2}, \frac{a'_3 + b_3}{2}\right].$$

We obtain the coordinate equalities

$$\frac{a_1 + b'_1}{2} = \frac{a'_1 + b_1}{2}, \quad \frac{a_2 + b'_2}{2} = \frac{a'_2 + b_2}{2}, \quad \frac{a_3 + b'_3}{2} = \frac{a'_3 + b_3}{2},$$

which implies $b_1 - a_1 = b'_1 - a'_1, b_2 - a_2 = b'_2 - a'_2, b_3 - a_3 = b'_3 - a'_3$.

It follows that the ordered triple $(b_1 - a_1, b_2 - a_2, b_3 - a_3)$ does *not* depend upon the choice of the geometric representation of the free vector $\overrightarrow{AB}$. Let us call this ordered triple *the coordinates of* $\overrightarrow{AB}$ *in the coordinate system* $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$. Further, let us agree to write $\overrightarrow{AB} = (b_1 - a_1, b_2 - a_2, b_3 - a_3)$. (Observe that this assignement provides a one-to-one mapping between V_3 and R^3!)

Let us finally verify the correctness of the definition of $\overrightarrow{u} \sharp \overrightarrow{v}$ (the correctness of the definition of $\lambda \circ \overrightarrow{u}$ can be verified analogously). Let $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$. Suppose that (A, B) is a geometric representation of $\overrightarrow{u}$. Then $A = [a_1, a_2, a_3]$ and $B = [a_1 + u_1, a_2 + u_2, a_3 + u_3]$. If (B, C) is a geometric representation of $\overrightarrow{v}$, $C = [a_1 + u_1 + v_1, a_2 + u_2 + v_2, a_3 + u_3 + v_3]$, then $\overrightarrow{u} \sharp \overrightarrow{v} = u_1 + v_1, u_2 + v_2, u_3 + v_3)$. We see that the vector $\overrightarrow{u} \sharp \overrightarrow{v}$ does not depend upon the concrete geometric representation of $\overrightarrow{u}$ (the coordinates of $\overrightarrow{u} \sharp \overrightarrow{v}$ do not involve the point coordinates of A!).

It can be easily shown that $(V_3, \sharp, \circ)$ with the operation defined above forms a linear space ("a vector space"). Indeed, all the axioms of Def. 1.1.1 are readily seen to be satisfied. In fact, the following result is in force.

4.1.1. Theorem: If we endow the set V_3 of free vectors in E_3 by the above defined operations $\sharp$ and $\circ$, then the triple $(V_3, \sharp, \circ)$ becomes a linear space. If we choose a Cartesian coordinate system $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ in E_3, then the mapping *coord* : $V_3 \to R^3$ which assigns to any $\overrightarrow{u} \in V_3$ its coordinates in the coordinate system $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ is a linear isomorphism. A corollary: $\dim V_3 = 3$ and $(\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ understood as a family of free vectors is a basis of V_3.

In what follows, we will simply call the elements of V_3 vectors, always meaning the free vectors. Since we will deal with the space V_3 only in this chapter, we will use the ordinary symbols $+, \cdot$ for the linear space operations $\sharp, \circ$ in V_3.

4.2. Scalar product of vectors

Suppose that $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ is the Cartesian coordinate system as defined in the previous section. Thus, $(\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ is a basis of V_3 and the length of all (free) vectors $\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k}$ is 1. Then we can naturally associate to any $\overrightarrow{u} \in V_3$ its *length* $|\overrightarrow{u}|$. Indeed, we simply represent $\overrightarrow{u}$ in E_3, $\overrightarrow{u} = \overrightarrow{AB}$, and we let $|\overrightarrow{u}|$ be the distance of the points A, B. This definition is meaningful since the unit length 1 has been previously defined. Also, this definition is correct since it obviously does not depend upon the geometric representation of $\overrightarrow{u}$.

Suppose that $\overrightarrow{u}, \overrightarrow{v} \in V_3$. We want to define the angle of $\overrightarrow{u}$ and $\overrightarrow{v}$. Let $\overrightarrow{u} = \overrightarrow{AB}$ and $\overrightarrow{v} = \overrightarrow{AC}$ (such representations of $\overrightarrow{u}$ and $\overrightarrow{v}$ are obviously possible). Then the *angle* of $\overrightarrow{u}$ and $\overrightarrow{v}$ is defined to be the angle of the half-lines AB and AC (see Fig. 4.2.1 below). Naturally, the angle, α, of $\overrightarrow{u}$ and $\overrightarrow{v}$ is expressed in the arc measure. Thus, $0 \le \alpha \le \pi$. (Observe that this definition of the angle of $\overrightarrow{u}$ and $\overrightarrow{v}$ is correct since it does not depend on the choice of the point A.)

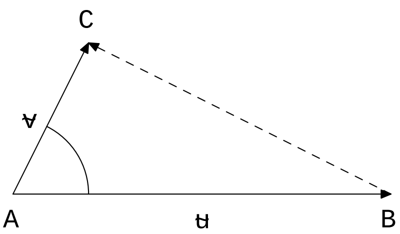


Figure 4:

The following important definition introduces "the degree of orthogonality" of two vectors.

4.2.1. Definition (the geometric definition of scalar product): Let $\overrightarrow{u}, \overrightarrow{v} \in V_3$. If vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ are nonzero, let us define $\overrightarrow{u} \cdot \overrightarrow{v}$ by setting $\overrightarrow{u} \cdot \overrightarrow{v} = |\overrightarrow{u}| \cdot |\overrightarrow{v}| \cdot \cos\alpha$, where α is the angle of the vectors $\overrightarrow{u}, \overrightarrow{v}$. If at least one of the vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ is $\overrightarrow{o}$, let us define $\overrightarrow{u} \cdot \overrightarrow{v} = 0$. The real number $\overrightarrow{u} \cdot \overrightarrow{v}$ is called *the scalar product* of $\overrightarrow{u}$ and $\overrightarrow{v}$.

As the name of the operation $\overrightarrow{u} \cdot \overrightarrow{v}$ suggests, the result of $\overrightarrow{u} \cdot \overrightarrow{v}$ is a scalar. An important fact is that the vectors $\overrightarrow{u}, \overrightarrow{v}$ are orthogonal (= perpendicular) if and only if $\overrightarrow{u} \cdot \overrightarrow{v} = 0$. Moreover, $|\overrightarrow{u}| = \sqrt{\overrightarrow{u} \cdot \overrightarrow{u}}$. Note that a typical physical interpretation of scalar product can be illustrated on Fig. 4.2.1. If a mass point moves from A to B "along the vector $\overrightarrow{u}$" and if this mass point is subject to the force $\overrightarrow{v}$, then the resulting work is equal to the scalar product $\overrightarrow{u} \cdot \overrightarrow{v}$.

Thus, the scalar product $\overrightarrow{u} \cdot \overrightarrow{v}$ can be easily figured out if we know $|\overrightarrow{u}|, |\overrightarrow{v}|$ and the angle of the vectors $\overrightarrow{u}$ and $\overrightarrow{v}$. However, in most problems we do not know these three quantities. We are rather given the coordinates of the vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ in the Cartesian system $(P, \vec{i}, \vec{j}, \vec{k})$ (the origin P is obviously not important in this consideration, the coordinates of $\vec{u}$ and $\vec{v}$ do not depend upon P since $\vec{u}$ and $\vec{v}$ are expressed with respect to the "free" basic vectors $\vec{i}, \vec{j}, \vec{k}$). As it turns out, the scalar product $\vec{u} \cdot \vec{v}$ can be conveniently computed "in coordinates" – a fact which enjoys several practical consequences. Here is the formulation of this result.

4.2.2. Theorem(on computing the scalar product in coordinates): Let $(P, \vec{i}, \vec{j}, \vec{k})$ be the Cartesian coordinate system in E_3. Let $\vec{u}, \vec{v} \in V_3$ and let $\vec{u} = (u_1, u_2, u_3)$, $\vec{v} = (v_1, v_2, v_3)$ be the coordinates of $\vec{u}$ and $\vec{v}$ with respect to the basis $\{\vec{i}, \vec{j}, \vec{k}\}$. Then

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + u_3v_3.$$

Proof: Assume that none of the vectors $\vec{u}$ and $\vec{v}$ is the zero vector (if either of them is the zero vector, than the verification reduces to triviality). By our assumption, $\vec{u} = u_1 \cdot \vec{i} + u_2 \cdot \vec{j} + u_3 \cdot \vec{k}$ and $\vec{v} = v_1 \cdot \vec{i} + v_2 \cdot \vec{j} + v_3 \cdot \vec{k}$. The Pythagorian theorem gives us $|\vec{u}| = \sqrt{u_1^2 + u_2^2 + u_3^2}$ and $|\vec{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2}$. We will now make use of the cosin theorem. Considering Fig. 4.2.1 on the previous page, we obtain

$$|\overrightarrow{BC}|^2 = |\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}| \cdot |\vec{v}| \cdot \cos\alpha$$

(here α means the angle of $\vec{u}$ and $\vec{v}$). Since $\overrightarrow{BC} = \vec{v} - \vec{u}$, we infer that $\overrightarrow{BC} = (v_1 - u_1, v_2 - u_2, v_3 - u_3)$. We finally obtain the equality

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

After a bit of manipulating, we see that

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + u_3v_3.$$

This is what we wanted to show. □

The above coordinate expression allows us to obtain the following properties of scalar product.

4.2.3. Proposition. Let $\vec{u}, \vec{v}, \vec{w} \in V_3$. Then

(i) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$,

(ii) $(\lambda \cdot \vec{u}) \cdot \vec{v} = \lambda \cdot (\vec{u} \cdot \vec{v})$ (the sign $\cdot$ takes the respective appropriate meaning),

(iii) $\vec{u} \cdot \vec{u} = \vec{u}^2 = |\vec{u}|^2 \geq 0$, and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{o}$,

(iv) $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$.

Proof: Only the part (iv) is nontrivial. In this case the equality amounts to showing that

$$(u_1 + v_1)w_1 + (u_2 + v_2)w_2 + (u_3 + v_3)w_3 = (u_1w_1 + u_2w_2 + u_3w_3) + (v_1w_1 + v_2w_2 + v_3w_3).$$

But this is easy. □

4.2.4. Remark. Observe that usually $(\vec{u} \cdot \vec{v})\vec{w} \neq \vec{u}(\vec{v}\,\vec{w})$.

Let us conclude this section with a few examples. In these examples (and in all that follows) we will automatically assume that the coordinates of vectors are expressed in our (fixed!) Cartesian system $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$. Also, we will often omit the dot in the scalar product $\overrightarrow{u} \cdot \overrightarrow{v}$, writing simply $\overrightarrow{u}\,\overrightarrow{v}$.

4.2.5. Example. Suppose that $|\overrightarrow{a}| = 1, |\overrightarrow{b}| = 3$ and suppose that the angle of $\overrightarrow{a}$ and $\overrightarrow{b}$ is $\dfrac{\pi}{3}$. Compute $|\overrightarrow{a} - 2\overrightarrow{b}|$.

Solution:

$$\begin{aligned} |\overrightarrow{a} - 2\overrightarrow{b}| &= \sqrt{(\overrightarrow{a} - 2\overrightarrow{b})(\overrightarrow{a} - 2\overrightarrow{b})} = \sqrt{\overrightarrow{a}^2 - 4\overrightarrow{a}\,\overrightarrow{b} + 4\overrightarrow{b}^2} \\ &= \sqrt{1 - 4 \cdot 1 \cdot 3 \cdot \cos\frac{\pi}{3} + 4 \cdot 9} = \sqrt{31}. \end{aligned}$$

Thus, $|\overrightarrow{a} - 2\overrightarrow{b}| = \sqrt{31}$.

4.2.6. Example. Compute the angle, α, of the vectors $\overrightarrow{u}, \overrightarrow{v}$, where $\overrightarrow{u} = (2, 2, -1)$ and $\overrightarrow{v} = (3, 0, 4)$.

Solution: Since $\overrightarrow{u} \cdot \overrightarrow{v} = |\overrightarrow{u}| \cdot |\overrightarrow{v}| \cdot \cos\alpha$, we obtain

$$\alpha = \arccos \frac{\overrightarrow{u} \cdot \overrightarrow{v}}{|\overrightarrow{u}| \cdot |\overrightarrow{v}|}.$$

Obviously, $\overrightarrow{u} \cdot \overrightarrow{v} = 2 \cdot 3 + 2 \cdot 0 + (-1) \cdot 4 = 2, |\overrightarrow{u}| = \sqrt{2^2 + 2^2 + (-1)^2} = 3$ and $|\overrightarrow{v}| = \sqrt{3^2 + 0^2 + 4^2} = 5$. Thus, $\alpha = \arccos \dfrac{2}{3 \cdot 5} \approx 1.4371$.

4.2.7. Example. Suppose that $\overrightarrow{a} = (-1, -2, 3)$ and $\overrightarrow{b} = (2, \alpha, 1)$ $(\alpha \in R)$. Find $\alpha \in R$ such that the vector $\overrightarrow{a}$ is orthogonal (=perpendicular) to the vector $\overrightarrow{b}$.

Solution: We are seeking $\alpha \in R$ so that $\overrightarrow{a}\,\overrightarrow{b} = 0$. Computing $\overrightarrow{a}\,\overrightarrow{b}$, we obtain $\overrightarrow{a}\,\overrightarrow{b} = -2 - 2\alpha + 3 = 1 - 2\alpha$. Thus, $\overrightarrow{a} \perp \overrightarrow{b}$ if and only if $\alpha = \dfrac{1}{2}$.

4.3. The vector product in V_3

In a number of problems coming from the analytic geometry we need construct a vector which is orthogonal to two given vectors. We will now develop an efficient procedure for finding such a vector. Though the essential novelty of this section will again involve the coordinate formula, let us first specify what the geometric idea of the construction is. (It is again recommended to become first aquainted with the following definition, then read the comments afterwards, and then carefully read the definition again.)

4.3.1. Definition (the geometric definition of vector product): Let $\overrightarrow{u}, \overrightarrow{v} \in V_3$. Let $\overrightarrow{w} \in V_3$ be the (only!) vector with the following properties:

(i) if $\overrightarrow{u}, \overrightarrow{v}$ are linearly dependent, then $\overrightarrow{w} = \overrightarrow{o}$,

(ii) if $\overrightarrow{u}, \overrightarrow{v}$ are linearly independent, then $\overrightarrow{w}$ is determined by the conditions *Vecpro* $_{1,2,3}$:

Vecpro $_1$ *:* $\overrightarrow{u} \cdot \overrightarrow{w} = \overrightarrow{v} \cdot \overrightarrow{w} = 0$ (i.e., $\overrightarrow{w}$ is orthogonal to both $\overrightarrow{u}$ and $\overrightarrow{v}$)

Vecpro $_2$ *:* $|\overrightarrow{w}| = |\overrightarrow{u}| \cdot |\overrightarrow{v}| \cdot \sin\alpha$, where α is the angle of the vectors $\overrightarrow{u}, \overrightarrow{v}$,

Vecpro $_3$ *:* When we take a geometric representation of $\overrightarrow{u}, \overrightarrow{v}$ and $\overrightarrow{w}$ so that $\overrightarrow{u} = \overrightarrow{AB}, \overrightarrow{v} = \overrightarrow{AC}$ and $\overrightarrow{w} = \overrightarrow{AD}$, then a hypothetical observer situated in D views the acute angle rotation of $\overrightarrow{u}$ into $\overrightarrow{v}$ in the anticlockwise orientation.

Then $\overrightarrow{w}$ is called *the vector product of* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ (in this order!). We write $\overrightarrow{w} = \overrightarrow{u} \times \overrightarrow{v}$.

The formulation of the above definition may look rather intimidating but the idea is rather simple. Considering the geometric vectors in E_3 , it is obvious that any pair of independent vectors allows for exactly one "orthogonal direction". It is this direction which is taken care of by the condition *Vecpro* $_1$ of Def. 4.3.1. What remains to be specified is the length (the condition *Vecpro* $_2$ and the orientation (the condition *Vecpro* $_3$. Thus, Def. 4.3.1 determines the vector $\overrightarrow{u} \times \overrightarrow{v}$ uniquely.

We will now formulate the coordinate expression of vector product.

4.3.2. Theorem(the coordinate expression of vector product): Let $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ be the Cartesian coordinate system. Let $\overrightarrow{u}, \overrightarrow{v} \in V_3$ and let $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$ be the coordinate expression of $\overrightarrow{u}$ and $\overrightarrow{v}$ with respect to the basis $\{\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k}\}$. Then

$$\overrightarrow{u} \times \overrightarrow{v} = \left(\det \begin{Vmatrix} u_2, & u_3 \\ v_2, & v_3 \end{Vmatrix}, \quad -\det \begin{Vmatrix} u_1, & u_3 \\ v_1, & v_3 \end{Vmatrix}, \quad \det \begin{Vmatrix} u_1, & u_2 \\ v_1, & v_2 \end{Vmatrix}\right).$$

PROOF: If the vectors $\overrightarrow{u}, \overrightarrow{v}$ are linearly dependent, then the coordinate formula gives $\overrightarrow{u} \times \overrightarrow{v} = \overrightarrow{o}$ which is in accord with the geometric definition of vector product. Suppose that $\overrightarrow{u}, \overrightarrow{v}$ are linearly independent. Let us now verify the properties of Def. 4.3.1 in order.

Vecpro $_1$: We have to show that the coordinate formula for $\overrightarrow{u} \times \overrightarrow{v}$ satisfies $\overrightarrow{u} \cdot (\overrightarrow{u} \times \overrightarrow{v}) = 0$ and $\overrightarrow{v} \cdot (\overrightarrow{u} \times \overrightarrow{v}) = 0$. Let us verify the former equality, the latter being analogous. We employ the following neat idea. We are to show that

$$u_1 \cdot \left(\det \begin{Vmatrix} u_2, & u_3 \\ v_2, & v_3 \end{Vmatrix}\right) + u_2 \cdot \left(-\det \begin{Vmatrix} u_1, & u_3 \\ v_1, & v_3 \end{Vmatrix}\right) + u_3 \cdot \left(\det \begin{Vmatrix} u_1, & u_2 \\ v_1, & v_2 \end{Vmatrix}\right) = 0.$$

But the left-hand side is nothing but the expansion of $\det \begin{Vmatrix} u_1, & u_2, & u_3 \\ u_1, & u_2, & u_3 \\ v_1, & v_2, & v_3 \end{Vmatrix}$ along the first row. Since this determinant is obviously 0 (the matrix in question has two rows identical), we have verified *Vecpro* $_1$.

Vecpro $_2$: Since $(|\overrightarrow{u}||\overrightarrow{v}|\cos\alpha)^2 + (|\overrightarrow{u}||\overrightarrow{v}|\sin\alpha)^2 = |\overrightarrow{u}|^2|\overrightarrow{v}|^2$, we have to check that $|\overrightarrow{u} \times \overrightarrow{v}|^2 = |\overrightarrow{u}|^2 \cdot |\overrightarrow{v}|^2 - (\overrightarrow{u} \cdot \overrightarrow{v})^2$. The coordinate formula for $\overrightarrow{u} \times \overrightarrow{v}$ gives

$$\begin{aligned}\overrightarrow{u} \times \overrightarrow{v} &= \left(\det \begin{Vmatrix} u_2, & u_3 \\ v_2, & v_3 \end{Vmatrix}, -\det \begin{Vmatrix} u_1, & u_3 \\ v_1, & v_3 \end{Vmatrix}, \det \begin{Vmatrix} u_1, & u_2 \\ v_1, & v_2 \end{Vmatrix}\right) \\ &= (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)\end{aligned}$$

and therefore

$$|\overrightarrow{u} \times \overrightarrow{v}|^2 = (u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2.$$

It remains to be checked that the following equality holds true:

$$(u_2v_3 - u_2v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2 = (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_2 + u_2v_2 + u_3v_3)^2.$$

We allow ourselves to leave it to the reader as an excercise in elementary algebra.

Vecpro $_3$: Since we have already shown the validity of the conditions *Vecpro* $_1$ and *Vecpro* $_2$, it is obvious that either the vector $\overrightarrow{u} \times \overrightarrow{v}$ or the vector $-(\overrightarrow{u} \times \overrightarrow{v})$ must satisfy the condition *Vecpro* $_3$. We will show that "the right vector" is the vector $\overrightarrow{u} \times \overrightarrow{v}$. We will provide an intuitive reasoning for this fact (a courious reader may consult more advanced books in geometry). Let us consider the vectors $\overrightarrow{u}$ and $\overrightarrow{v}$. By choosing an appropriate Cartesian coordinate system $(P, \overrightarrow{i}', \overrightarrow{j}', \overrightarrow{k}')$ by "transforming" our original system $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ so that $\mathrm{Span}\{\overrightarrow{i}'\} = \mathrm{Span}\{\overrightarrow{u}\}$, $\mathrm{Span}\{\overrightarrow{i}', \overrightarrow{j}'\} = \mathrm{Span}\{\overrightarrow{u}, \overrightarrow{v}\}$ and $\overrightarrow{j}' \cdot \overrightarrow{v} > 0$, we see that the coordinate expression of $\overrightarrow{u}$ and $\overrightarrow{v}$ in the basis $\{\overrightarrow{i}', \overrightarrow{j}', \overrightarrow{k}'\}$ is as follows:

$$\overrightarrow{u} = (u_1', 0, 0), \quad \overrightarrow{v} = (v_1', v_2', 0), \quad u_1' > 0, v_2' > 0.$$

In these coordinates we obtain $\overrightarrow{u} \times \overrightarrow{v} = (0, 0, u_1'v_2')$. The latter vector obviously fulfils the condition *Vecpro* $_3$ (we have $\overrightarrow{u} \times \overrightarrow{v} = \lambda \overrightarrow{k}$ in the new coordinates, where $\lambda > 0$). This completes the proof of Th. 4.3.2. □

4.3.3. Example. Let $\overrightarrow{u} = (1, -2, 2)$ and $\overrightarrow{v} = (3, 1, -2)$ be a coordinate expression of $\overrightarrow{u}, \overrightarrow{v}$ in the Cartesian coordinate system. Find $\overrightarrow{u} \times \overrightarrow{v}$.

SOLUTION: For practical reasons, it is handy to write (at least mentally) the vectors so that " $\overrightarrow{v}$ is written below $\overrightarrow{u}$ ". This allows us to compute the required determinants easily. Thus, "covering" the respective columns and computing the determinants, we obtain

$$\begin{matrix} (1, & -2, & 2) \\ (3, & 1, & -2) \end{matrix} \implies \overrightarrow{u} \times \overrightarrow{v} = \left(\det \begin{Vmatrix} -2, & 2 \\ 1, & -2 \end{Vmatrix}, -\det \begin{Vmatrix} 1, & 2 \\ 3, & -2 \end{Vmatrix}, \det \begin{Vmatrix} 1, & -2 \\ 3, & 1 \end{Vmatrix} \right) = (2, 8, 7).$$

Let us list some useful properties of vector product.

4.3.4. Proposition. Let $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w} \in V_3$. Then

(i) $\overrightarrow{u} \times \overrightarrow{v} = (-\overrightarrow{v}) \times \overrightarrow{u} = -(\overrightarrow{v} \times \overrightarrow{u})$,

(ii) $(\lambda \overrightarrow{u}) \times \overrightarrow{v} = \lambda \cdot (\overrightarrow{u} \times \overrightarrow{v})$,

(iii) $(\overrightarrow{u} + \overrightarrow{v}) \times \overrightarrow{w} = \overrightarrow{u} \times \overrightarrow{w} + \overrightarrow{v} \times \overrightarrow{w}$,

(iv) $(\overrightarrow{u} \times \overrightarrow{v}) \cdot \overrightarrow{w} = \overrightarrow{u} \cdot (\overrightarrow{v} \times \overrightarrow{w})$,

(v) $(\overrightarrow{u} \times \overrightarrow{v}) \times \overrightarrow{w} = (\overrightarrow{u} \cdot \overrightarrow{v}) \cdot \overrightarrow{v} - (\overrightarrow{v} \cdot \overrightarrow{w}) \cdot \overrightarrow{u}$.

PROOF: The proof of each of these equalities can be easily done in coordinates (see Th. 4.3.2).

The formula $(\overrightarrow{u} \times \overrightarrow{v}) \cdot \overrightarrow{w}$ is sometimes called *the mixed product of* $\overrightarrow{u}, \overrightarrow{v}$ *and* $\overrightarrow{w}$. The mixed product has two interesting properties. The algebraic side worthwhile observing is the expression of mixed product in coordinates. If $\overrightarrow{u} = (u_1, u_2, u_3), \overrightarrow{v} = (v_1, v_2, v_3)$ and $\overrightarrow{w} = (w_1, w_2, w_3)$, then

$$(\overrightarrow{u} \times \overrightarrow{v}) \cdot \overrightarrow{w} = \det \begin{Vmatrix} u_1, & u_2, & u_3 \\ v_1, & v_2, & v_3 \\ w_1, & w_2, & w_3 \end{Vmatrix}.$$

This follows immediately by expanding the determinant along the third row. (As a by-product, we see that $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}$ are linearly dependent if and only if $(\overrightarrow{u} \times \overrightarrow{v})\overrightarrow{w} = 0$.) An interesting geometric feature is that for linearly independent vectors $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}$, the number $|(\overrightarrow{u} \times \overrightarrow{v}) \cdot \overrightarrow{w}|$ equals to the volume of the parallelepiped $ABCD$, where $\overrightarrow{u} = \overrightarrow{AB}, \overrightarrow{v} = \overrightarrow{AC}$ and $\overrightarrow{w} = \overrightarrow{AD}$. We can write $|(\overrightarrow{u} \times \overrightarrow{v}) \cdot \overrightarrow{w}| = |(\overrightarrow{u} \times \overrightarrow{v})| \cdot |\overrightarrow{w}| \cdot |\cos\alpha|$, where α is the angle of $\overrightarrow{u} \times \overrightarrow{v}$ and $\overrightarrow{w}$. Since $|\overrightarrow{u} \times \overrightarrow{v}| = |\overrightarrow{u}||\overrightarrow{v}| \cdot \sin\alpha$, then $|\overrightarrow{u} \times \overrightarrow{v}$ is exactly the area of the bottom of the parallelepiped. Since $|\overrightarrow{w}| \cdot |\cos\alpha|$ is its height, we are done with the proof of the last proposition.

4.4. Linear varieties in E_3 (applications of scalar product and vector product)

By a linear variety in E_3 we mean either a point in E_3, or a straight line in E_3, or a plane in E_3 or the whole of E_3. We want to take up some questions of analytic geometry involving linear varieties in E_3. We will first provide a parametric description of varieties.

Let us start with the following convention. Let $A \in E_3$ and $\overrightarrow{v} \in V_3$. Then the (formal) sum $A + \overrightarrow{v}$ will mean the point B such that $\overrightarrow{AB} = \overrightarrow{v}$. We symbolically write $B = A + \overrightarrow{v}$, resp. $\overrightarrow{AB} = B - A$. In our Cartesian system of coordinates $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$, we can alternately write $[b_1, b_2, b_3] = [a_1+v_1, a_2+v_2, a_3+v_3]$, where $B = [b_1, b_2, b_3], A = [a_1, a_2, a_3]$ and $\overrightarrow{v} = (v_1, v_2, v_3)$ (we will always use a different type of brackets for points and vectors). Recall that the meaning of "point" coordinates is given by the equality $A = P + a_1\overrightarrow{i} + a_2\overrightarrow{j} + a_3\overrightarrow{k}, B = P + b_1\overrightarrow{i} + b_2\overrightarrow{j} + b_3\overrightarrow{k}$, etc.

The simplest variety is *a point*. It is usually given by its coordinates in $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$. Slightly more complex variety is a *straight line* (or simply a line). If a line, p, is determined by two points $A, B \in E_3$, it is obvious that p consists of all points X that allow for the expression $X = A + t\overrightarrow{v}$, where $t \in R$ and $\overrightarrow{v} = B - A$. We write $p: \ X = A + t \cdot \overrightarrow{v}, t \in R$.

The latter expression is called a *parametric equation of straight line.* In coordinates, the parametric equation reads

$$\begin{aligned} x_1 &= a_1 + tv_1 \\ x_2 &= a_2 + tv_2 \\ x_3 &= a_3 + tv_3, \quad t \in R. \end{aligned}$$

(The above equations are replaced in some textbooks with a seemingly simpler form $\dfrac{x_1 - a_1}{v_1} = \dfrac{x_2 - a_2}{v_2} = \dfrac{x_3 - a_3}{v_3}$. We will not use this.)

A plane is the next variety in the hierarchy. Typically, a plane is given by three points. Suppose that ρ is a plane in E_3 and suppose that A, B, C are points in E_3 that belong to

ρ. If the vectors $\overrightarrow{u} = B - A$ and $\overrightarrow{v} = C - A$ are linearly independent in V_3, then the plane ρ is exactly the set of all points $X \in E_3$ such that $X = A + t_1\overrightarrow{u} + t_2\overrightarrow{v}$ for some parameters $t_1, t_2 \in R$. We write

$$\rho\colon X = A + t_1\overrightarrow{u} + t_2\overrightarrow{v}, \qquad t_1, t_2 \in R,$$

and we call the above analytic definition of the plane *a parametric equation (a parametric form)* of ρ. In coordinates, the parametric equation reads

$$\begin{aligned} x_1 &= a_1 + t_1u_1 + t_2v_1 \\ x_2 &= a_2 + t_1u_2 + t_2v_2 \\ x_3 &= a_3 + t_1u_3 + t_2v_3, \quad t_1, t_2 \in R. \end{aligned}$$

If we are given a point $A \in \rho$ and two LI vectors $\overrightarrow{u}, \overrightarrow{v}$ such that $\overrightarrow{u} = B - A$ and $\overrightarrow{v} = C - A$ for some points $B, C \in \rho$, then we can immediately write down the parametric form of ρ. Very often it is handy to pass to a *point-normal equation* for ρ. Suppose that $A \in \rho$ and suppose that $\overrightarrow{n}$ is a vector that is normal to ρ. Thus, $\overrightarrow{n} \in V_3$ such that, for a point $A \in \rho$, the vector $\overrightarrow{XA}$ is orthogonal to $\overrightarrow{n}$ for any $X \in \rho$. Using the scalar product, we see that the equation

$$(X - A) \cdot \overrightarrow{n} = 0\,, \qquad A \in \rho,$$

determines all points $X \in \rho$. The above formula is called the point-normal equation of ρ. The coordinate alternative of this form reads, for $\overrightarrow{n} = (n_1, n_2, n_3)$,

$$n_1x_1 + n_2x_2 + n_3x_3 - d = 0,$$

where $d = a_1n_1 + a_2n_2 + a_3n_3$. On the other hand, every equation

$$a_1x_1 + a_2x_2 + a_3x + b = 0,$$

where a_1, a_2, a_3, b are real numbers, determines uniquely a plane in the above point-normal sense (in particular, the vector (a_1, a_2, a_3) is a normal vector to the plane in question).

4.4.1. Example. Suppose that the plane ρ is given by the points $A = [1, -1, 2], B = [0, 2, 1]$ and $C = [2, 0, 5]$. Find the parametric and point-normal description of ρ.

Solution: The parametric form is

$$\rho\colon X = A + t_1(B - A) + t_2(C - A), \quad t_1, t_2 \in (-\infty, +\infty).$$

In our concrete setup,

$$\rho\colon X = [1, -1, 2] + t_1(-1, 3, -1) + t_2(1, 1, 3), \quad t_1, t_2 \in R.$$

For the point-normal equation of ρ we need to construct a normal vector to ρ. The vector product suggests itself – the vector $\overrightarrow{n} = (B - A) \times (C - A)$ is certainly one of the normal vectors to ρ. We have $\overrightarrow{n} = (-1, 3, -1) \times (1, 1, 3)$. Thus, writing

$$\begin{matrix} (-1, & 3, & -1), \\ (\ \ 1, & 1, & 3), \end{matrix}$$

we see that $\overrightarrow{n} = (10, 2, -4)$. Since the "shorter" vector $\overrightarrow{n_1} = (5, 1, -2)$ is equally appropriate, we infer that the equation

$$(X - [1, -1, 2]) \cdot (5, 1, -2) = 0$$

is the required point-normal equation of ρ. Thus,

$$\rho\colon\ 5x_1 + x_2 - 2x_3 = 0.$$

4.4.2. Example. Suppose that

$$\rho\colon\ 2x_1 - 3x_2 + x_3 + 1 = 0.$$

Find a parametric equation of ρ.

Solution: Obviously, there is a lot of parametric equations of ρ. To obtain one, we just need a point belonging to ρ and two vectors from the "direction" of ρ. All these three "unknowns" we can easily guess. Since the vector $\overrightarrow{n} = (2, -3, 1)$ is a normal vector to ρ, all we need to obtain are two LI vectors $\overrightarrow{u}, \overrightarrow{v}$ such that $\overrightarrow{n} \cdot \overrightarrow{u} = 0$ and $\overrightarrow{n} \cdot \overrightarrow{v} = 0$. These vectors can be guessed from the coordinate expression of scalar product. The conditions are $(2, -3, 1)(u_1, u_2, u_3) = 0$ and $(2, -3, 1)(v_1, v_2, v_3) = 0$. We may therefore take $\overrightarrow{u} = (3, 2, 0)$ and $\overrightarrow{v} = (0, 1, 3)$ (the LI of $\overrightarrow{u}$ and $\overrightarrow{v}$ is obviously guaranteed). Thus,

$$\rho\colon\ X = [0, 0, 1] + t_1(3, 2, 0) + t_2(0, 1, 3), \quad t_1, t_2 \in R.$$

Sometimes a line is given as an intersection of planes. Let us illustrate it by example. Obviously, in the equations of lines or planes, we may use (x, y, z) instead of (x_1, x_2, x_3).

4.4.3. Example. Find the parametric form of the line p which is given as the intersection of the planes ρ_1 and ρ_2:

$$\begin{aligned} \rho_1\colon&\ x + y - 2z - 2 = 0 \\ \rho_2\colon&\ 2x - y + z + 1 = 0. \end{aligned}$$

Solution: Since $p = \rho_1 \cap \rho_2$, we easily obtain a point $A \in p$ and a direction vector of p. The coordinates of A can be found as a common solution of the equations for ρ_1 and ρ_2. After simple elimination of the equations, we have

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

and the point $A = [x_A, y_A, z_A]$ can be easily guessed:

$$A = [\,\cdot\,, 0, -1] \quad \Longrightarrow \quad A = [0, 0, -1].$$

The direction vector can be computed on the ground of the following observation. Since the direction vector $\overrightarrow{u}$ of p is orthogonal to the normal vectors $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ of ρ_1 and ρ_2, where $\overrightarrow{n_1} = (1, 1, -2)$ and $\overrightarrow{n_2} = (2, -1, 1)$, we have $\overrightarrow{u} = \overrightarrow{n_1} \times \overrightarrow{n_2}$. Thus, $\overrightarrow{u} = (-1, -5, -3)$ and therefore $p\colon\ X = [0, 0, -1] + t(-1, -5, -3)$, $t \in R$.

4.5. Analytic geometry in E_3

In this section we will give a short account of some geometric problems in E_3. We will essentially deal with concrete problems, minimizing general theoretical considerations. As a rule, the problems we want to pursue will require a space imagination and inventiveness rather than anything else.

We always prefer the so called synthetic-analytic approach. This is to say, we always want to acquire the insight of the problem rather than use recipes. We first proceed as if we intended to solve it syntheticaly (i.e., we approach the problem as if it was a descriptive geometry problem). This enables us to obtain the necessary understanding in order to chose the appropriate analytic apparatus.

Let us first take up problems on *the distance of linear varieties.* The simplest case is *the distance of two points.* Then, of course, the distance of these points is simply the magnitude (=the length) of the vector which they determine. In coordinates, if $A = [a_1, a_2, a_3]$ and $B = [b_1, b_2, b_3]$, then the distance of A and B, denoted by $d(A, B)$, is computed as follows:

$$d(A,B) = |\overrightarrow{AB}| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}.$$

4.5.1. Example. Consider the line $p\colon X = A + t\overrightarrow{v}, t \in R$, where $A = [1, -1, 2]$ and $\overrightarrow{v} = (1, 2, -2)$. Find a point, B, such that B lies on the positive half-line of p and such that $d(A, B) = 6$. Then compute $d(C, B)$ for $C = [0, 0, 0]$ and $C = [3, 0, 5]$.

Solution: Since B lies on the positive half-line of p, it will correspond to a positive parameter t. Obviously, we could take a general point B on p with a positive parameter and settle the parameter by making use of the formula for the distance. More "geometric" approach is this. Since $|\overrightarrow{v}| = \sqrt{(1)^2 + (2) + (-2)^2} = 3$ then the required parameter must be 2. Thus,

$$B = A + 2 \cdot \overrightarrow{v} = [1, -1, 2] + 2(1, 2, -2) = [3, 3, -2].$$

Finally,

$$d\left([0, 0, 0], [3, 3, -2]\right) = |(3, 3, -2)| = \sqrt{(3)^2 + (3^2) + (-2)^2} = \sqrt{22}$$

and

$$d\left([3, 0, 5], [3, 3, -2]\right) = |(0, 3, -7)| = \sqrt{0 + 9 + 49} = \sqrt{58}.$$

The next question is finding *the distance of a point, P, from a straight line p.* The distance is denoted by $d(A, p)$. If we wanted to solve this question syntheticaly, we would project the point A into the line p first and then measure the distance of the point A from its projection, A'. But the projection into a line can be easily found analyticaly. Let us see how one proceeds.

4.5.2. Example. Let $A = [-3, 1, 2]$ and $p\colon\ X = [1, 3, 2] + t(2, 1, -2), t \in R$. Find the projection of A into p. Then compute the distance $d(A, p)$.

Solution: It is easily seen that if we take a plane which passes through A and which is orthogonal to p, then the desired projection A' must lie in this plane. Thus, if we denote by

ρ this plane, we have $A' = p \cap \rho$. But the point-normal equation of ρ can be found easily – the direction vector of p must be the normal vector of ρ.

After having completed the "strategy", we may set up for the "tactics". The plane ρ passes through $A = [-3, 1, 2]$. The vector $\overrightarrow{n} = (2, 1, -2)$ is the normal to ρ. Thus, the point-normal equation for ρ is $(X - [-3, 1, 2])(2, 1, -2) = 0$. In coordinates, we obtain

$$\rho\colon\ 2x_1 + x_2 - 2x_3 + 9 = 0.$$

We want to intersect this plane with the line p the coordinate expression of which is

$$p\colon \quad \begin{aligned} x_1 &= 1 + 2t \\ x_2 &= 3 + \ \ t \\ x_3 &= 2 - 2t. \end{aligned}$$

The intersection point is then the projection A' we look for. Thus, if we write $A' = [a_1', a_2', a_3']$, where

$$\begin{aligned} a_1' &= 1 + 2u \\ a_2' &= 3 + \ \ u \\ a_3' &= 2 - 2u \ \ , \end{aligned}$$

we also have $A' \in \rho$, and therefore

$$2(1 + 2u) + (3 + u) - 2(2 - 2u) + 9 = 0.$$

This gives us the parameter $u \in R$ we look for:

$$\begin{aligned} 2 + 4u + 3 + u - 4 + 4u + 9 &= \quad 0 \quad \Longrightarrow \\ 9u &= -10 \quad \Longrightarrow \\ u \ &= -\frac{10}{9}. \end{aligned}$$

Substituting back, we have

$$A' = [a_1', a_2', a_3'] = \left[1 - 2 \cdot \frac{10}{9}, 3 - \frac{10}{9}, 2 - 2 \cdot (-\frac{10}{9})\right] = \left[-\frac{11}{9}, \frac{17}{9}, \frac{38}{9}\right].$$

By our analysis, $d(A, p) = d(A, A') = |\overrightarrow{AA'}|$. Since

$$\overrightarrow{AA'} = \left[-\frac{11}{9}, \frac{17}{9}, \frac{38}{9}\right] - [-3, 1, 2] = \left(\frac{16}{9}, \frac{8}{9}, \frac{20}{9}\right) = \frac{2}{9}(8, 4, 10),$$

we obtain $|\overrightarrow{AA'}| = \frac{2}{9} \cdot \sqrt{8^2 + 4^2 + 10^2} = \frac{4}{3}\sqrt{5}$. Thus, $d(A, p) = \frac{4}{3}\sqrt{5}$.

We have observed that not only we have solved our distance problem but we have also found the point with respect to which the distance should be measured (the projection point). Many problems concerning the distance require finding this projection point as much as finding the distance itself.

If the distance is the only "unknown" we want to discover, we can use the following general formula. It should be noticed that one can happily live without knowing this formula as the example of the author of this text acknowledges. We nevertheless decided to include it here (and the other analogous formulas, too). The reasons are twofold. First, the formula is a nice application of vector product. And second, if we do not do so, we will be soon exhausted explaining to students that we are aware of these formulas but we are afraid of mechanical application. In geometric problems *the understanding of the procedure is usually much more important than the result* (though we do not deny that the correctness of the final numerical outcome also matters).

4.5.3. Proposition (the distance of a point from a straight line). Let $A \in E_3$ and $p\colon X = B + t \cdot \overrightarrow{a}$ $(t \in R)$, for $\overrightarrow{a} \in V_3, \overrightarrow{a} \neq \overrightarrow{o}$. Then the distance, $d(A,p)$, of A from p can be computed by the formula

$$d(A,p) = \frac{|\overrightarrow{BA} \times \overrightarrow{a}|}{|\overrightarrow{a}|}.$$

Proof: Consider the figure below.

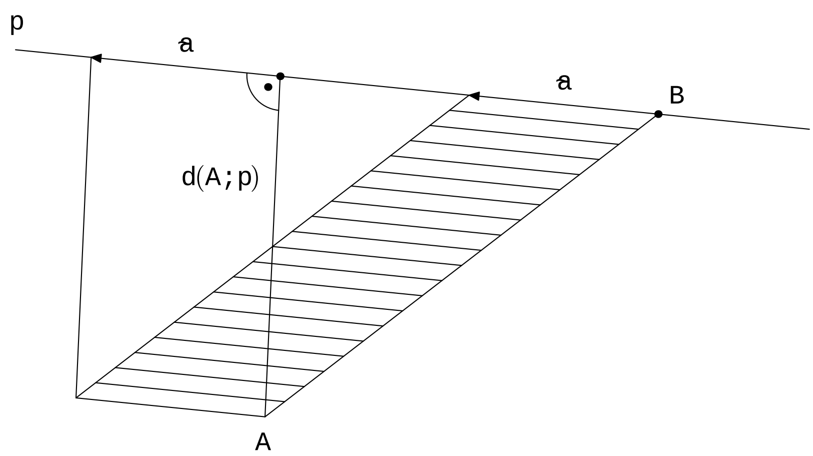


Figure 5:

By the definition of vector product, the area of the parallelogram indicated by the figure is $|\overrightarrow{BA} \times \overrightarrow{a}|$. Thus, $|\overrightarrow{BA} \times \overrightarrow{a}| = |\overrightarrow{a}| \cdot d(A,p)$. This proves the formula.

Let us test the correctness of the solution of the previous example. The formula gives

$$d(A,p) = \frac{|(-4,-2,0) \times (2,1,-2)|}{|(2,1,-2)|} = \frac{|(4,-8,0)|}{3} = \frac{4\sqrt{5}}{3}.$$

Let us take up the problem of finding *the distance of a point P from a plane ρ*. Let us denote this distance by $d(P,\rho)$. The problem then obviously reduces to constructing the orthogonal projection P' of the point into the plane ρ. The following example illustrates how one proceeds.

4.5.4. Example. Find the distance of the point $P = [-2,-4,3]$ from the plane $\rho\colon\ 2x - y + 2z + 6 = 0$.

Solution: If we denote by P' the orthogonal projection of P into the plane ρ, then $d(P,\rho) = d(p,P')$. We therefore need find P'. The point P' is obviously the intersectioon of ρ with the straight line, p, which passes through P and which is orthogonal to ρ. But p can be found easily – the normal to ρ must be the direction vector of p. Thus, p: $X = [-2,-4,3]+t(2,-1,2)$, $t \in R$. In coordinates, if $X=[x,y,z]$, then

$$\begin{aligned} x &= -2+2t \\ y &= -4-\ \ t \\ z &= \ \ 3+2t, \quad t \in R. \end{aligned}$$

Since $P' = p \cap \rho$, we substitute into the equation of the plane and obtain

$$2(-2+2t)-(-4-t)+2(3+2t)+6=0.$$

This gives $12+9t=0$. Hence $t=-\frac{4}{3}$. It follows that $P'=[-\frac{14}{3},-\frac{8}{3},\frac{1}{3}]$. Since $d(P,M)=d(P,P')$, we obtain

$$\begin{aligned} d(P,M) &= d([-2,-4,3],[-\frac{14}{3},-\frac{8}{3},\frac{1}{3}]) = (-\frac{8}{3},\frac{4}{3},-\frac{8}{3}) = |\frac{4}{3}\cdot(-2,1,-2)| = \\ &= \frac{4}{3}\sqrt{(2)^2+(-1)^2+(2)^2} = 4. \end{aligned}$$

The distance of a point from a plane can be also computed by a general formula. We will derive it in the following proposition.

4.5.5. Proposition. Let $P=[x_0,y_0,z_0]$ be a point and let ρ: $ax+by+cz+d=0$ be a plane. Let $d(P,\rho)$ describe the distance of P from ρ. Then

$$d(P,\rho)=\frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}.$$

PROOF: Take an arbitrary point, $Q=[q_1,q_2,q_3]$, from the plane ρ. Since the vector $\overrightarrow{n}=(a,b,c)$ is a normal to ρ, we see that ρ: $(X-Q)\overrightarrow{n}=0$. Since $d=aq_1+bq_2+cq_3$, our formula is equivalent with the formula

$$d(P,\rho)=\frac{|(P-Q)\overrightarrow{n}|}{|\overrightarrow{n}|}.$$

We will prove the latter formula.

Consider the figure below. We easily see that $\pm d(P,\rho)\cdot\frac{\overrightarrow{n}}{|\overrightarrow{n}|}=P-P'$, where P' is the projection of P into ρ (note that $\left|\frac{\overrightarrow{n}}{|\overrightarrow{n}|}\right|=1$). It follows that $\pm d(P,\rho)\cdot\frac{\overrightarrow{n}}{|\overrightarrow{n}|}=(P-Q)+(Q-P')$. Forming now the scalar product of $\overrightarrow{n}$ with both sides of this equation, we obtain

$$\pm d(P,\rho)\cdot\frac{\overrightarrow{n}^2}{|\overrightarrow{n}|}=(P-Q)\overrightarrow{n}+(Q-P')\overrightarrow{n}.$$

Since $(Q-P')\overrightarrow{n}=0$ and $\overrightarrow{n}^2=|\overrightarrow{n}|^2$, we have $\pm d(P,\rho)\cdot|\overrightarrow{n}|=(P-Q)\overrightarrow{n}$. Thus,

$$d(P,\rho)=\frac{|(P-Q)\overrightarrow{n}|}{|\overrightarrow{n}|}.$$

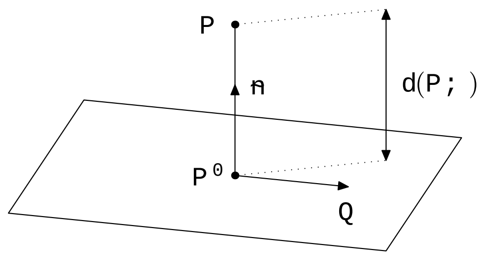


Figure 6:

Let us check the correctness of the solution to Ex. 4.5.4 by using the above formula. We have $P=[-2,-4,3]$ and ρ: $2x-y+2z+6=0$. Substituting, we obtain

$$d(P,Q)=\frac{|-4+4+6+6|}{\sqrt{4+1+4}}=4.$$

The next question is *finding the distance of two straight lines.* Consider two straight lines, p and q and denote by $d(p,q)$ their distance. If p and q intersect, then $d(p,q)=0$. If p and q are parallel, then the problem reduces to finding the distance of a point from a line ($d(p,q)=d(P,q)$ for any $P\in p$). The only case remain when p and q are neither parallel nor intersecting. In this case the lines p and q are called *the skew lines.* Finding the *distance* of skew lines then presents an interesting question. Again, we can do without any theory, the recomended synthetic-analytic approach is usually sufficient to suggest the appropriate strategy for the solution. The following example illustrates how one can proceed.

4.5.6. Example. Suppose that we are given two straight lines a and b. Find our whether they are skew lines. Then find their distance.

$$a: X=[5,2,-5]+t(2,2,-1), \quad t\in R,$$
$$b: X=[2,-1,1]+u(1,-2,-2), \quad u\in R.$$

Solution: Since the straight lines a,b are not parallel (the vectors $(2,2,-1)$ and $(1,-2,-2)$ are not linearly independent), we have checked that they are skew lines as soon as we have show that they are not intersecting. This would mean to show that they do not have any point in common. It is better to make the following observation: Two lines $a: X=A+t\overrightarrow{a}$ and $b: X=B+t\overrightarrow{b}$ are skew if and only if the vectors $\overrightarrow{a}, \overrightarrow{b}, B-A$ are linearly independent. Indeed, if the latter condition is valid then the straight lines a,b cannot be either parallel or intersecting. Expressing the linear independence of $\overrightarrow{a}, \overrightarrow{b}, B-A$ in coordinates, if $\overrightarrow{a}=(a_1,a_2,a_3)$, $\overrightarrow{b}=(b_1,b_2,b_3)$ and $B-A=(c_1,c_2,c_3)$, then the straight lines a,b are skew exactly when

$$\det\begin{Vmatrix} a_1, & a_2, & a_3\\ b_1, & b_2, & b_3\\ c_1, & c_2, & c_3\end{Vmatrix}\neq 0.$$

In our case we are to test the determinant $D = \det \begin{Vmatrix} 2, & 2, & -1 \\ 1, & -2, & -2 \\ -3, & -3, & 6 \end{Vmatrix}$. Since $D = 2(-12-6) - 2(6-6) - 1(-3-6) = -27 \neq 0$, we have verified that the lines a, b are skew.

In the attemp to settle the strategy of the solution, let us again imagine the problem as if it was solved. By the standard geometric definition, the distance $d(a,b)$ of these two lines can be found as the minimum of the distances $d(P,Q)$, where $P \in a$ and $Q \in b$. Thus, $d(a,b) = \min\{d(P,Q), P \in a,\ Q \in b\}$. It is intuitively obvious that $d(a,b) = d(P_0,Q_0)$, where $P_0 \in a$ and $Q_0 \in b$ are the (unique) points so that the vector $P_0 - Q_0$ *is orthogonal* (=perpendicular) to both a and b (one makes use of Pythagorian theorem). Our problem then reduces to finding such points P_0 and Q_0. These points P_0 and Q_0 can be found easily. Let us demonstrate two ways which are possible, but there are many others.

1. We know that $P_0 = [5,2,-5] + t_0(2,2,-1)$ for some (yet unknown) parameter t_0 ($t_0 \in R$) and $Q_0 = [2,-1,1] + u_0(1,-2,-2)$ for some (yet unknown) parameter u_0 ($u_0 \in R$). Form the vector $P_0 - Q_0 = [3,3,-6] + t_0(2,2,-1) - u_0(1,-2,-2)$ and look for the parameters t_0 and u_0 so that $P_0 - Q_0$ is orthogonal to both a and b. We shall use the scalar product. The conditions are

$$((3,3,-6) + t_0(2,2,-1) - u_0(1,-2,-2))\,(2,2,-1) = 0$$

and

$$((3,3,-6) + t_0(2,2,-1) - u_0(1,-2,-2))\,(1,-2,-2) = 0.$$

This gives the following equalities:

$$\begin{aligned} -6+6+\ 6+t_0(4+4+1)-u_0(2-4+2)&=0,\\ 3-6+12\ +t_0(2-4+2)-u_0(1+4+4)&=0. \end{aligned}$$

After simple manipulation we obtain

$$\begin{array}{ccccccc} 18 & + & 9t_0 & & & = & 0 \\ 9 & & & - & 9u_0 & = & 0 \end{array}$$

which means that $t_0 = -2$ and $u_0 = 1$. Thus,

$$P_0 = [5,2,-5] - 2(2,2,-1) = [1,-2,-3] \quad \text{and} \quad Q_0 = [2,-1,1] + 1(1,-2,-2) = [3,-3,-1].$$

Since $d(a,b) = d(P_0,Q_0)$, we conclude that $d(a,b) = d([1,-2,-3],[3,-3,-1]) = |(-2,1,-2)| = 3$.

2. Since $\overrightarrow{P_0Q_0}$ must be orthogonal to both a and b, we see that $\overrightarrow{P_0Q_0} = \alpha((2,2,-1) \times (1,-2,-2))$. Thus $\overrightarrow{P_0Q_0} = \alpha(-6,3,-6) = \alpha'(2,-1,2)$. The problem then reduces to finding a straight line, c, which intersects both a and b and which is parallel with the "direction" $(2,-1,2)$. In order to solve the latter problem, let us notice that the point Q_0 we look for must lie in the plane which contains the line a and which is parallel with $(2,-1,2)$. Let us denote this plane by ρ. Thus,

$$\rho\colon\ X = [5,2,-5] + v(2,2,-1) + w(2,-1,2), \quad v,w \in R.$$

Since $Q_0 = b \cap \rho$, we have completed the strategy (the rest of the strategic line is routine).

Let us embark on the "tactics". Seeking the point $Q_0 = b \cap \rho$, we write

$$\begin{aligned} b: x &= 2 + u \\ y &= -1 - 2u \\ z &= 1 - 2u \end{aligned}$$

and we express ρ in the point-normal form:

$$\rho\colon\ ((x,y,z) - [5,2,-5])\,((2,2,-1) \times (2,-1,2)) = 0.$$

Then $\rho\colon\ x - 2y - 2z - 11 = 0$. After substituting we have the following equation for the unknown parameter u:

$$(2+u) - 2(-1-2u) - 2(1-2u) - 11 = 0.$$

This gives $u = 1$ and therefore $Q_0 = [3,-3,-1]$. We have essentially solved the problem. After computing P_0 (i.e., after visualizing P_0 as the intersection of the line a with a suitably chosen plane σ), we obtain $P_0 = [1,-2,-3]$ (we leave this to the reader).Thus, $d(a,b) = d(P_0, Q_0) = 3$.

The previous type of problem (i.e., the problem of distance of two skew lines) allows for a general formula, too. We will derive the formula in the following theorem. Again, it is good to be aware of this formula but, as shown above, it is not at all necessary to memorize it.

4.5.7. Theorem: Let a and b be two straight lines. Let $a\colon\ X = A + \alpha\overrightarrow{u}$, $(\alpha \in R)$ and $b\colon\ X = B + \beta\overrightarrow{v}$, $(\beta \in R)$ be the parametric equations of the straight lines. Let $A = [a_1, a_2, a_3]$, $B = [b_1, b_2, b_3]$, $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$. If a and b are not parallel and if $d(a,b)$ denotes the distance of a and b, then

$$d(a,b) = \frac{\left| \det \begin{Vmatrix} u_1, & u_2, & u_3 \\ v_1, & v_2, & v_3 \\ a_1 - b_1, & a_2 - b_2, & a_3 - b_3 \end{Vmatrix} \right|}{|\overrightarrow{u} \times \overrightarrow{v}|}.$$

Proof: Suppose that the lines are skew (if they are not skew, then $d(a,b) = 0$ and this agrees with the above formula for $d(a,b)$). The situation is shown by the figure below.

Consider the paralleleppipet indicated by the figure. Let V be its volume. Then

$$V = \left| \det \begin{Vmatrix} u_1, & u_2, & u_3 \\ v_1, & v_2, & v_3 \\ a_1 - b_1, & a_2 - b_2, & a_3 - b_3 \end{Vmatrix} \right| = d(a,b) \cdot |\overrightarrow{u} \times \overrightarrow{v}|.$$

This proves the formula.

Let us check the result of the previous Example 4.5.6. There we had

$$a\colon\ X = [5,2,-5] + t(2,2,-1),\ (t \in R), \qquad b\colon\ X = [2,-1,1] + u(1,-2,-2),\ (u \in R).$$

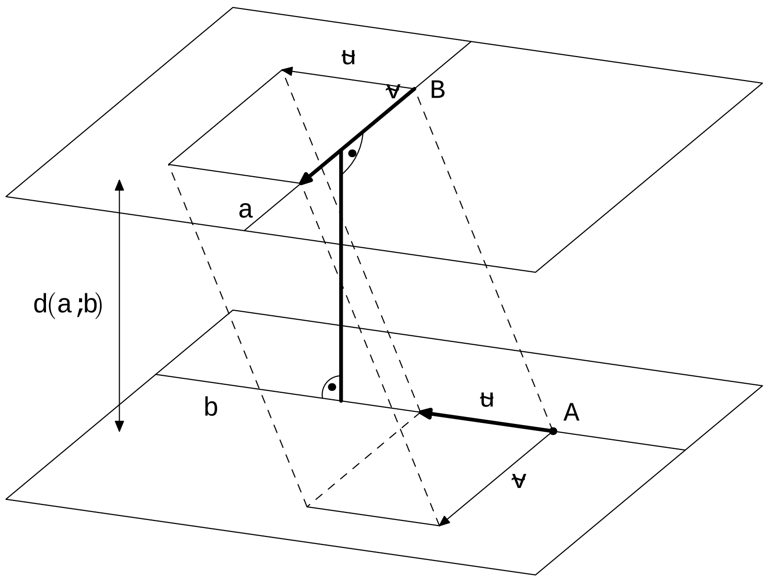


Figure 7:

Substituting into the formula derived in Theorem 4.5.7, we obtain

$$d(a,b) = \frac{\left| \det \begin{Vmatrix} 2, & 2, & -1 \\ 1, & -2, & -3 \\ 3, & 3, & -6 \end{Vmatrix} \right|}{|(2,2,-1) \times (1,-2,-2)|} = \frac{|2(12+9) - 2(-6+9) - 1(3+6)|}{|(-6,3,-6)|} =$$

$$= \frac{|2 \cdot 21 - 2 \cdot 3 - 1 \cdot 9|}{3 \cdot |(-2,1,-2)|} = \frac{|42 - 6 - 9|}{3 \cdot 3} = \frac{27}{9} = 3$$

What remains in our discussion of distance is the case of two planes. But this is an easy case to handle – two planes are either intersecting, in which case their distance is zero, or they are parallel, in which case we just take an arbitrary point of one of those planes and find its distance from the other plane. Thus, the problem of distance of two varieties has been fully clarified.

The next type of problem we want to deal with is *the problem of constructing a variety with preassigned properties.* This thema is not supplied with any profound theory – a bit of imagination and a friendly attitude towards geometry will do. Let us consider some typical examples.

4.5.8. Example. Find a plane which passes through two given points A and B and which is orthogonal to a given plane ρ. Solve this problem for $A = [-1,-2,0], B = [1,1,2]$ and ρ: $X = [0,1,1] + \alpha(2,-1,0) + \beta(0,1,-1),\ \alpha, \beta \in R$.

Solution: Let σ be the plane we are looking for. Then the normal vector to σ, $\vec{n}$, must be perpendicular to both $\overrightarrow{AB}$ and the normal, $\vec{n'}$, to ρ. As a consequence, we can put

$\overrightarrow{n} = \overrightarrow{AB} \times \overrightarrow{n'}$. Thus,

$$\overrightarrow{n} = (2,3,2) \times ((2,-1,0) \times (0,1,-1)) = (-2,2,1).$$

It follows that $\sigma\colon\ (x - [-1,-2,0])(-2,2,1) = 0$. The equation of the plane reads $\sigma\colon\ 2x - 2y + z - 2 = 0$.

4.5.9. Example. Find the straight line which passes through a given point, M , and which intersects two given straight lines, p, q . Solve this problem for $M = [-1,-7,4]$ and $p\colon\ X = [2,-5,3] + t(3,-2,-1),\ t \in R$; $q\colon$ it passes through the points $[7,0,-3]$ and $[5,-3,2]$.

Solution: It is easy to check that p, q are skew lines. Thus, the problem has a unique solution. Let us denote by r the line we look for. Let us think of the problem as if we previously solved it synthetically. Then the point M and the line p would determine a plane, ρ , so that r must pass through the intersection of ρ with q . Writing $Q = q \cap \rho$, we see that the equation of the desired line r reads as follows:

$$r\colon\ X = M + t(Q - M), \qquad t \in R.$$

Let us take up the analytic solution. It will follow the strategy we just set up. Let us first write the equation of the plane ρ . By the definition of ρ , we see that the parametric form of ρ reads $\rho\colon\ X = M + \alpha \overrightarrow{a} + \beta(M - A),\ \alpha, \beta \in R$, where $\overrightarrow{a} = (3,-2,1)$ and $A = [2,-5,3]$. Thus,

$$\rho\colon\ X = [-1,-7,4] + \alpha(3,-2,-1) + \beta(-3,-2,1), \qquad \alpha, \beta \in R.$$

We want to intersect ρ with q . To this aim, rewrite ρ first in the point-normal form. We obtain $\rho\colon\ (X - M)\overrightarrow{n} = 0$, where $M = [-1,-7,4]$ and $\overrightarrow{n} = (3,-2,-1) \times (-3,-2,1)$. Since $\overrightarrow{n} = (-4,0,-12) = -4(1,0,3)$, we easily find the point-normal equation of ρ :

$$\rho\colon\ x + 3z - 11 = 0.$$

The last step is finding the intersection point $Q,\ Q = q \cap \rho$. Since $q\colon\ X = [7,0,-3] + t'(-2,-3,5),\ t \in R$, we have to solve the equation $(7 - 2t') + 3(-3 + 5t') - 11 = 0$. This gives us $t' = 1$. Thus, $Q = [5,-3,2]$ and therefore

$$r\colon\ X = [-1,-7,4] + w(6,4,-2), \quad w \in R.$$

A final type of geometric problem we want to visit is *the discussion of intersection of given planes.* The algebraic background has been established in Chapter 2.

4.5.10. Example. Discuss the intersection of the following three planes ρ_1, ρ_2, ρ_3 with respect to the parameters $a, b (a, b \in R)$:

$$\begin{array}{lrrrrcl} \rho_1: & 2x & -y & +3z & -1 & = & 0 \\ \rho_2: & x & +2y & -z & +b & = & 0 \\ \rho_3: & x & +ay & -6z & +10 & = & 0 \end{array}$$

Solution: The intersection $\rho_1 \cap \rho_2 \cap \rho_3$ consists of exactly one point if, and only if, the corresponding system of three equations with three unknowns possesses exactly one solution.

This is the case if, and only if, the matrix of the system is regular (Th. 2.4.15). The latter fact is equivalent to $\det \mathbf{A} \neq 0$, where $\mathbf{A}$ is the matrix of the system. Writing the system in the matrix form,

$$\left\|\begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 1, & 2, & -1 & -b \\ 1, & a, & -6 & -10 \end{array}\right\|,$$

we therefore ask when $\det \mathbf{A} = \det \left\|\begin{array}{rrr} 2, & -1, & 3 \\ 1, & 2, & -1 \\ 1, & a, & -6 \end{array}\right\| \neq 0$. Since $\det A = 5a - 35$, we see that the intersection consists of a singleton exactly when $a \neq 7$. (Observe that this circumstance does not depend upon b!) The intersection point, P, then reads $P = [p_1, p_2, p_3]$, where (Cramer's rule)

$$\begin{aligned} p_1 &= \frac{\det C_1}{\det \mathbf{A}} = \frac{a + 6b - 3ab + 38}{5a - 35} \\ p_2 &= \frac{\det C_2}{\det \mathbf{A}} = \frac{15b - 45}{5a - 35} \\ p_3 &= \frac{\det C_3}{\det \mathbf{A}} = \frac{a + b + 2ab - 52}{5a - 35} \end{aligned}$$

Let us go on with our analysis. The remaining case is $a = 7$. If this occurs, the intersection must be either a straight line or an empty set. Consider now the situation with $a = 7$. We obtain

$$\left\|\begin{array}{rrr|r} 2, & -1, & 3 & 1 \\ 1, & 2, & -1 & -b \\ 1, & a, & -6 & -10 \end{array}\right\| \sim \left\|\begin{array}{rrr|r} 1, & 2, & -1 & -b \\ 0, & 5, & -5 & b-10 \\ 0, & 0, & 0 & 3b-9 \end{array}\right\|$$

If $b = 3$, then the system has infinitely many solutions. Thus, the intersection $\rho_1 \cap \rho_2 \cap \rho_3$ of the planes is a straight line, p. Its equation can be easily guessed:

$$p\colon\ X = \left[-\frac{8}{5}, 0, \frac{7}{5}\right] + t(-1, 1, 1), \qquad t \in R.$$

If $b \neq 3$, then the system does not have any solution. Thus, if $a = 7$ and $b = 3$, then $\rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$.

Summary:

I. If $a \neq 7$, then $\rho_1 \cap \rho_2 \cap \rho_3$ consists of a singleton, i.e., in this case $\rho_1 \cap \rho_2 \cap \rho_3 = P$, where P is a point. Moreover, $P = [p_1, p_2, p_3]$, where

$$\begin{aligned} p_1 &= \frac{a + 6b - 3ab + 38}{5a - 35} \\ p_2 &= \frac{15b - 45}{5a - 35} \\ p_3 &= \frac{a + b + 2ab - 52}{5a - 35} \end{aligned}$$

The situation is illustrated by the figure below (observe that for any $a, b \in R$ each pair of the planes ρ_1, ρ_2, ρ_3 intersects in a straight line)

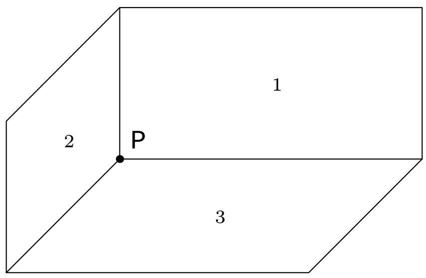


Figure 8:

II. If $a = 7$ and $b = 3$, then $\rho_1 \cap \rho_2 \cap \rho_3$ is a straight line, p. Moreover,

$$p\colon\ X = \left[-\frac{8}{5}, 0, \frac{7}{5}\right] + t(-1, 1, 1), \quad t \in R.$$

III. If $a = 7$ and $b \neq 3$, then $\rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$ ("a tent").

Miscellaneous examples

4.5.11. Example. Suppose that we are given a point, P, and a plane, ρ. Find a point, R, which is symmetric to P with respect to the plane ρ. Solve this problem for $P = [-4, 5, 8]$ and $\rho\colon\ 3x - 3y - 4z + 25 = 0$.

SOLUTION: We will first find the projection P' of P into ρ. For this, take the straight line, p, which passes through P and which is orthogonal to ρ. Then $P' = p \cap \rho$. The rest is easy – we have $\dfrac{P+R}{2} = P'$ and this equation determines the point R we seek.

The equation of p is not hard to write down – it suffices to observe that the direction vector of p is the normal, $\overrightarrow{n}$, to ρ. But $\overrightarrow{n} = (3, -3, 4)$ and therefore $p\colon\ X = [-4, 5, 8] + t(3, -3, 4),\ t \in R$. Since $P' = p \cap \rho$, we look for the parameter t $(t \in R)$ such that

$$3(-4 + 3t) - 3(5 - 3t) - 4(8 - 4t) + 25 = 0.$$

This gives $t = 1$. Thus, $P' = [-1, 2, 4]$. Since $P' = \dfrac{P+R}{2}$, we infer that $R = P' + (P' - P)$. After substituting we see that $R = [2, -1, 0]$.

4.5.12. Example. Let p be a straight line and let P be a point. Find a point, R, which is symmetric to the point P with respect to the straight line p. (Solve for $p\colon\ X = [7, 14, 18] + t(2, 4, 5)$ and $P = [4, 3, 10]$.)

SOLUTION: In order to find the required point P, we first need find the projection, P', of P into p. This can be done several ways. We will demonstrate three ways.

1. First solution (when one wants to safe mental energy for more engineering-like subjects then this ...).

We first want to find the projection point P'. Obviously, P' lies on the line p and the vector PP' is orthogonal to p. Thus, if we write $P' = [7,14,18] + t'(2,4,5)$, then the orthogonality condition translates as follows ($P = [4,3,10]$ and the vector $(2,4,5)$ is the direction vector of p):

$$(\ [4,3,10] - (\ [7,14,18] + t'(2,4,5)\ )\ (2,4,5) = 0.$$

This gives the following sequence of implications:

$$((-3,-11,-8) - t'(2,4,5))\ (2,4,5) = 0 \Longrightarrow (-6-44-40-t'(4+16+25)) = 0 \Longrightarrow$$

$$-90 - t'(45) = 0 \qquad\qquad t' = -2.$$

It follows that the projection point P' reads $P' = [7,14,18] - 2(2,4,5) = [3,6,8]$. The required symmetric point, R, must satisfy the equation $2 \cdot \overrightarrow{PP'} = \overrightarrow{PR}$. Substituting, this means ($R = [r_1, r_2, r_3]$ $2([4,3,10] - [3,6,8]) = R(\ [4,3,10] - [r_1,r_2,r_3]\ )$. We see that $2 \cdot (1,-3,2) = (4-r_1, 3-r_2, 10-r_3)$ and therefore $2 = 4-r_1$, $-6 = 3-r_2$ and $4 = 10-r_3$. Thus, $R = [2,9,6]$.

2. Second solution (when one feels like a bit of good old descriptive geometry ...).

We will again look for the projection point P' of P into p. Let σ be the plane which passes through P and which is orthogonal to the line p. Obviously, $P' = p \cap \sigma$. Since the point-normal equation of σ is easily available (its normal vector is known!), we can proceed to concrete computations.

Let us first find the equation of σ. We have

$$\sigma\colon\ ([x,y,z] - [4,3,10])(2,4,5) = 0 \quad \Longrightarrow \quad \sigma\colon\ 2x + 4y + 5z - 70 = 0.$$

Looking for $P' = p \cap \sigma$, we need solve for t the following equation:

$$2(7+2t) + 4(14+4t) + 5(18+5t) - 70 = 0.$$

We further obtain

$$14 + 4t + 56 + 16t + 90 + 25t - 70 = 0,$$

which gives the equation $90 + 45t = 0$. This yields $t = -2$, and therefore $P' = [3,6,8]$. As a consequence, $R = [2,9,6]$.

3. Third solution (when one had a dream on the vector product ...).

This solution starts with the following question: Can we write the equation of the "projection" straight line? The idea is that if we can find the equation of this line, then the projection point, P', is simply the intersection of this line with p. Let us denote by q the projection straight line. Obviously, q passes through the point $P = [4,3,10]$. We want to find a direction vector, $\overrightarrow{q}$, of q. Let us denote by $\tilde{P}$ the "origin" of p (in our case, $\tilde{P} = [7,14,18]$) and by $\overrightarrow{p}$ the given direction vector (in our case, $\overrightarrow{p} = (2,4,5)$). Then one easily sees that it suffices to let $\overrightarrow{q} = (\overrightarrow{p} \times (\tilde{P} - P)) \times \overrightarrow{p}$. This elegant geometric idea is worth remembering – it is applicable in many situations.

Let us compute the vector $\overrightarrow{q}$. We have

$$\begin{aligned}\overrightarrow{q} &= ((2,4,5) \times (3,11,8)) \times (2,4,5) = (-23,-1,10) \times (2,4,5)\\ &= (-45,135,-90) = -45 \cdot (1,-3,2).\end{aligned}$$

Thus, the vector $(1,-3,2)$ can be taken for the direction vector of the line q. We see that

$$q\colon\ X=[4,3,10]+u(1,-3,2),\quad (u\in R).$$

Further, the projection point P' is the intersection of q and p ($P'=q\cap p$). We therefore have to solve the following system for u and t (recall that $p\colon\ X=[7,14,18]+t(2,4,5),\ t\in R$):

$$\begin{array}{rcrcrcr} 4 & + & u & = & 7 & + & t \\ 4 & - & 3u & = & 14 & + & 4t \\ 10 & + & 2u & = & 18 & + & 5t \end{array}$$

After simple algebra we obtain $u=1$ and $t=-2$. As a consequence, $P'=[4,3,10]+t(1,-3,2)=[3,6,8]$ and therefore $R=[2,9,6]$.

4.5.13. Example. Find a point, P, which lies on a given straight line, a, and which fulfils the following condition: The distance of P from a given point, M, is 5 (i.e., $d(P,M)=5$).

$$\begin{aligned} a\colon\ X&=[-3,2,2]+t(-7,3,4),\quad t\in R,\\ M&=[1,-1,2] \end{aligned}$$

Solution: If $X=[x,y,t]$ is a general point on the line a corresponding to the parameter t, then

$$\begin{aligned} d(X,M)&=d([-3-7t,2+3t,2+4t],[1,-1,2])=|(-4-7t,3_3t,4t)|=\\ &=\sqrt{(-4-7t)^2+(3+3t)^2+(4t)^2}=\sqrt{16+56t+49t^2+9+18t+9t^2+16t^2}=\\ &=\sqrt{25+74t+74t^2}. \end{aligned}$$

We are looking for X which satisfies the condition $d(X,M)=5$. Thus, we have the equation $\sqrt{25+74t+74t^2}=5$, which gives the following quadratic equation for t:

$$74t^2+74t=0.$$

The equation has two roots – $t_1=0$ and $t_2=-1$. It follows that we obtain two points with the required property – $P_1=[-3,2,2]$ and $P_2=[4,-1,-2]$.

4.5.14. Appendix: A note on transforming the point (=Cartesian) coordinates

Let $(P,\overrightarrow{i},\overrightarrow{j},\overrightarrow{k})$ and $(P',\overrightarrow{i}',\overrightarrow{j}',\overrightarrow{k}')$ be two Cartesian coordinate systems in E_3. Let us call the former system "old" and the latter "new". We want to derive the transformation formula between the old and new coordinates. This formula is useful in a number of engineering problems.

Let the relationship of these two systems is expressed in terms of the following two formulas:

$$P'=P+p_1\overrightarrow{i}+p_2\overrightarrow{j}+p_3\overrightarrow{k},$$

$$\begin{aligned} \overrightarrow{i}'&=a_{11}\overrightarrow{i}+a_{12}\overrightarrow{j}+a_{13}\overrightarrow{k},\\ \overrightarrow{j}'&=a_{21}\overrightarrow{i}+a_{22}\overrightarrow{j}+a_{23}\overrightarrow{k},\\ \overrightarrow{k}'&=a_{31}\overrightarrow{i}+a_{32}\overrightarrow{j}+a_{33}\overrightarrow{k}. \end{aligned}$$

Take an arbitrary point $X = [x_1, x_2, x_3] = [x'_1, x'_2, x'_3]$, where $[x_1, x_2, x_3]$ are the coordinates of X with respect to $(P, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$ and $[x'_1, x'_2, x'_3]$ are the coordinates of X with respect to $(P', \overrightarrow{i}', \overrightarrow{j}', \overrightarrow{k}')$. In other words, we can write

$$X = P + x_1 \overrightarrow{i} + x_2 \overrightarrow{j} + x_3 \overrightarrow{k} = P' + x'_1 \overrightarrow{i}' + x'_2 \overrightarrow{j}' + x'_3 \overrightarrow{k}'.$$

Substituting, we further obtain

$$\begin{aligned} X &= P + x_1 \overrightarrow{i} + x_2 \overrightarrow{j} + x_3 \overrightarrow{k} \\ &= P + p_1 \overrightarrow{i} + p_2 \overrightarrow{j} + p_3 \overrightarrow{k} + x'_1(a_{11} \overrightarrow{i} + a_{12} \overrightarrow{j} + a_{13} \overrightarrow{k}) \\ &\quad + x'_2(a_{21} \overrightarrow{i} + a_{22} \overrightarrow{j} + a_{23} \overrightarrow{k}) + x'_3(a_{31} \overrightarrow{i} + a_{32} \overrightarrow{j} + a_{33} \overrightarrow{k}). \end{aligned}$$

After a simple manipulation we conclude that

$$\begin{aligned} x_1 &= a_{11}x'_1 + a_{21}x'_2 + a_{31}x'_3 + p_1, \\ x_2 &= a_{12}x'_1 + a_{22}x'_2 + a_{32}x'_3 + p_2, \\ x_3 &= a_{13}x'_1 + a_{23}x'_2 + a_{33}x'_3 + p_3. \end{aligned}$$

In the matrix form, we finally have the required transformation formula (note that we need *the transpose* to the transition matrix)

$$\begin{Vmatrix} x_1 \\ x_2 \\ x_3 \end{Vmatrix} = \begin{Vmatrix} a_{11}, & a_{21}, & a_{31} \\ a_{12}, & a_{22}, & a_{32} \\ a_{13}, & a_{23}, & a_{33} \end{Vmatrix} \quad \begin{Vmatrix} x'_1 \\ x'_2 \\ x'_3 \end{Vmatrix} + \begin{Vmatrix} p_1 \\ p_2 \\ p_3 \end{Vmatrix}.$$

We could derive the transformation formula from the new coordinates to the old analogously.

Problems

1. Find the equation by the straight line p which passes through the points $A = [4, -5, 7]$ and $B = [3, 1, 2]$.

   [Answer: $p\colon\ X = [4, -5, 7] + t(-1, 6, -5), t \in R$]

2. Consider the straight line $p\colon\ X = A + t\overrightarrow{a}, t \in R$, where $A = [3, -2, 1]$ and $\overrightarrow{a} = (-1, 2, 3)$. Which of the points $B = [1, 2, 7], C = [2, 3, -1]$ and $D = [4, -4, -2]$ lie on the line p? Which lie on the positive half-line?

   [Answer: B, D do, C does not; only B lies on the positive half-line]

3. Find the point-normal equation of the plane ρ which is determined by the points $A = [-1, 2, 3]$, $B = [2, 3, -2]$ and $C = [0, 5, 4]$.

   [Answer: $\rho\colon\ 2x - y + z = 1$]

4. a) Find the angle α of the straight lines p and q.

   $p\colon\ X = [1, 1, -1] + t(1, 2, 2),\ t \in R,$

   $q\colon$ it passes through the points $[1, 0, 2]$ and $[3, -2, 3]$.

b) Find the angle φ of the planes ρ and σ:

$$\begin{aligned}\rho&:\ x-2y+2z+3=0,\\ \sigma&:\ X=[1,1,-1]+t(1,-1,0)+u(2,2,1),\ t,u\ \in R.\end{aligned}$$

[Answer: a) $\alpha=\frac{\pi}{2}$, b) $\varphi=\frac{\pi}{4}$]

5. Show that the following straight lines a and b lie in a plane. Find the point P in which the lines a and b intersect.

$$\begin{aligned}a&:\ X=[0,1,0]+t(1,2,-1),\ t\in R,\\ b&:\ X=[2,4,1]+u(1,1,2),\ u\in R.\end{aligned}$$

[Answer: $P=[1,3,-1]$]

6. Find the distance of the point P from the plane ρ:

$$\begin{aligned}&P=[-1,1,2],\\ &\rho:\ X=[1,-1,1]+t(-3,2,2)+u(3,-4,3),\ t,u\in R.\end{aligned}$$

[Answer: $d(P,\rho)=\frac{4}{7}$]

7. Find the distance of the planes ρ and σ:

$$\begin{aligned}\rho&:\ 2x+3y-6z+10=0,\\ \sigma&:\ X=\left[\frac{15}{4},0,0\right]+t(3,-2,0)+u(0,2,1),\ t,u\in R.\end{aligned}$$

[Answer: $d(\rho,\sigma)=\frac{5}{2}$]

8. Find the distance of the point A from the straight line p:

$$\begin{aligned}A&=[2,3,-1],\\ p&:\ X=[2,-2,-23]+t(3,2,-2),\ t\in R.\end{aligned}$$

[Answer: $d(A,p)=21$]

9. Find the distance of the point A from the straight line p:

$$\begin{aligned}A&=[1,2,0],\\ p&:\ X=[3,0,0]+t(0,1,0),\ t\in R.\end{aligned}$$

[Answer: $d(A,p)=2$]

10. Show that the straight lines a and b are skew lines and find their distance:

$$\begin{aligned}a&:\ X=[-5,-5,1]+t(3,2,-2),\ t\in R,\\ b&:\ X=[9,0,2]+u(6,-2,-1),\ u\in R.\end{aligned}$$

[Answer: $d(a,b)=7$]

11. Show that the straight lines a and b are skew lines. Find their distance $d(a,b)$. Find also the point $A \in a$ and $B \in b$ such that $d(a,b) = d(A,B)$.

$$a\colon\ X = [1,2,1] + t(5,-2,-30,\ t \in R,$$
$$b\colon\ X = [-2,6,3] + u(4,-3,-1),\ u \in R.$$

[Answer: $d(a,b) = \sqrt{3}, A = [1,2,1], B = [2,3,2]$]

12. Find the straight line, which has the following three properties:
   (i) a passes through a given point M,
   (ii) a is parallel with a given plane ρ,
   (iii) a intersects a given straight line p.

$$M = [3,-2,-4]$$
$$\rho\colon\ 3x - 2y - 3z - 7 = 0$$
$$p\colon\ X = [2,-4,1] + t(3,-2,2),\ (t \in R).$$

[Answer: $a\colon\ X = [3,-2,-4] + t(5,-6,9),\ (t \in R)$]

13. Suppose that we are given a plane, ρ, and two points, M and N. Consider the straight line p determined by the points M and N. Find the orthogonal projection q of p into ρ. Find also the angle, α, of p and ρ.

$$\rho\colon\ x + 2y + 2z - 4 = 0,\ M = [-1,2,0],\ N = [1,1,2].$$

[Answer: $q\colon\ X = [2,1,0] + t(2,1,-2),\ t \in R$, $\alpha = \arcsin \frac{4}{9}$]

14. Consider the following point M and the straight lines a and b. Check that a and b are skew lines. Then find the straight line, c, which passes through the point M and intersects both a and b.

$$M = [-4,-5,3],$$
$$a\colon\ X = [-1,-3,2] + t(3,-2,-1),\ t \in R,$$
$$b\colon\ X = [2,-1,1] + u(2,3,-5),\ u \in R.$$

[Answer: $c\colon\ X = [-4,-5,3] + t(3,2,-1),\ t \in R$]

15. Consider two skew lines a and b, and consider also two planes ρ and σ. Find the straight line, c, which intersects both a and b and which is parallel with each of the planes ρ and σ.

$$a\colon\ X = [4,6,-1] + t(1,2,-2),\ t \in R,$$
$$b\colon\ X = [3,5,-1] + s(2,2,1),\ s \in R,$$
$$\rho\colon\ x + y + z + 2 = 0,$$
$$\sigma\colon\ 3x - 2y + 1 = 0.$$

[Answer: $X = [3,4,1] + u(2,3,-5),\ (u \in R)$]

16. Let ρ be the plane of a mirror and let M and N be two given points. Suppose that a ray r passes through M. In which point on ρ the ray r must hit the mirror so that it is reflected into N?

$$\rho\colon\ 2x - 2y + z = 0,$$
$$M = [3, -1, 1], N = [8, 3, 8].$$

[Answer: $P = [2, 3, 2]$, the angle is $\frac{\pi}{4}$]

17. Suppose that we are given a straight line, p, and two points A and B. Find a point, M, on p which is equally far away from A and B.

p : it is determined as the intersection of the following two planes ρ and σ;

$$\begin{array}{lrcrcrcrcl} \rho\colon & x & + & y & - & z & - & 3 & = & 0, \\ \sigma\colon & & & 3y & - & 2z & + & 1 & = & 0 \end{array}$$

$A = [1, 0, -1],\ B = [3, 4, 5]$.

[Answer: $M = [4, 1, 2]$]

18. Find a point, P on the axis z so that the distance of P from a given point M and a given plane ρ is equal.

$$M = [1, -2, 0], \qquad \rho\colon\ 3x - 2y + 6z - 9 = 0.$$

[Answer: There are two solutions:
$P_1 = [0, 0, -2],\ P_2 = [0, 0, -\frac{82}{13}]$]

19. Find a point, Q, which is symmetric to a given point P with respect to a given plane, σ:

$$P = [-3, 2, 5],$$
σ: it contains the following two straight lines a and b;

$$a = \alpha \cap \beta,\ \text{where } \begin{cases} \alpha\colon\ x - 2y + 3z - 5 = 0, \\ \beta\colon\ x - 2y - 4z + 3 = 0, \end{cases}$$

$$b = \gamma \cap \delta,\ \text{where } \begin{cases} \gamma\colon\ 3x + \ \ y + 3z + z = 0, \\ \delta\colon\ 5x - 3y + 2z + 5 = 0. \end{cases}$$

[Answer: $Q = [1, -6, 3]$]

20. Find a point, Q, which is symmetric to a given point P with respect to a given straight line, p:

p: it passes through $A = [5, 4, 6]$ and $B = [-2, -17, -8]$,
$$P = [2, -5, 7].$$

[Answer: $Q = [4, 1, -3]$]

21. Find the coordinates of the vertices of the square S which is subject to the following requirements. One vertice of S, A, is given and a diagonal of S lies on a given straight line, p $(A \notin p)$. What is the area of S?

$$A = [4, 1, 6],$$
p: it is the intersection of ρ and σ, where

$$\begin{aligned}\rho\colon&\quad x - y - 4z + 12 = 0,\\ \sigma\colon&\ 2x + y - 2z + \ \ 3 = 0\end{aligned}$$

[Answer: $S = ABCD$, where $A = [4,1,6]$, $B = [5,-3,5]$, $C = [2,-3,2]$, $D = [1,1,3]$, the area of S is 18]

22. Let a and L be real parameters. Discuss the possibilities for the intersection of the following planes ρ_1, ρ_2, ρ_3. Illustrate the respective situations by simple figures (see Ex. 4.5.10):

$$\begin{array}{llllllllll}\rho_1: & 2x & - & y & - & z & + & 2 & = & 0,\\ \rho_2: & x & + & ay & + & 3z & - & 1 & = & 0,\\ \rho_3: & 7x & + & 4y & + & 7z & + & b & = & 0.\end{array}$$

[Answer: $a \neq 2 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = P$ (P is a point), where $P = [x,y,z]$ with the coordinates

$$x = \frac{-14a + 3b - ab + 27}{21(a-2)}, \qquad y = \frac{7(b-1)}{21(a-2)}, \qquad z = \frac{14a - b - 2ab - 23}{21(a-2)};$$

$a = 2, b = 1 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = p$ (p is a straight line), $p\colon X = [-\frac{3}{5}, \frac{4}{5}, 0] + t(1,7,-5), t \in R$;

$a = 2, b \neq 1 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$ ("a tent" – any two planes out of ρ_1, ρ_2, ρ_3 are not parallel).]

23. Let a and b be real parameters. Discuss the possibilities for the intersection of the following planes ρ_1, ρ_2, ρ_3. Illustrate the respective situations by simple figures (see Ex. 4.5.10):

$$\begin{array}{llllllllll}\rho_1: & 2x & - & y & + & 3z & + & 1 & = & 0,\\ \rho_2: & 6x & - & 5y & + & z & + & a & = & 0,\\ \rho_3: & -8x & + & 4y & + & by & + & 9 & = & 0.\end{array}$$

[Answer: $b \neq -12 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = P$ (P is a point), where $P = [x,y,z]$ with the coordinates

$$x = \frac{-12a + 5b - 1b - 122}{-4(b+12)}, \qquad y = \frac{-2(12a - 3b + ab + 72}{-4(b+12)}, \qquad z = \frac{52}{-4(b+12)};$$

$b = -12 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$ (in this case ρ_1 is parallel with ρ_3).]

24. Let a be a real parameter. Discuss the possibilities for the intersection of the following planes ρ_1, ρ_2, ρ_3. Illustrate the respective situations by simple figures (see Ex. 4.5.10):

$$\begin{array}{llllllllll}\rho_1: & x & + & y & + & z & - & 1 & = & 0,\\ \rho_2: & x & + & ay & + & z & - & a & = & 0,\\ \rho_3: & x & + & y & + & az & - & a & = & 0.\end{array}$$

[Answer: $a \neq 1 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = [-1,1,1]$;

$a = 1 \Longrightarrow \rho_1 = \rho_2 = \rho_3$ and therefore $\rho_1 \cap \rho_2 \cap \rho_3 = \rho$ (ρ is a plane, $\rho = \rho_1 = \rho_2 = \rho_3$).]

25. Let a be a real parameter. Discuss the possibilities for the intersection of the following planes ρ_1, ρ_2, ρ_3 . Illustrate the respective situations by simple figures (see Ex. 4.5.10):

$$\begin{array}{rrcrcrcrcl} \rho_1 : & x & + & y & + & az & - & (a-1)^2 & = & 0, \\ \rho_2 : & x & + & ay & + & z & + & 1 & = & 0, \\ \rho_3 : & ax & + & y & + & z & - & 1 & = & 0. \end{array}$$

[Answer: $a \neq 1, a \neq -2 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = P$ (P is a point), where $P = [x, y, z]$ with the coordinates

$$x = \frac{-a^2 + 3a + 1}{(a-1)(a+2)}, \qquad y = \frac{-a^2 + a - 3}{(a-1)(a+2)}, \qquad z = \frac{a^2 - 1}{a + 2};$$

$a = 1 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$ (the planes ρ_1, ρ_2, ρ_3 are mutually different and parallel);
$a = -2 \Longrightarrow \rho_1 \cap \rho_2 \cap \rho_3 = \emptyset$ ("a tent" – the planes mutually intersect in a straight line but these three straight lines are all parallel).]

Chapter 5

Appendix 1: Mathematical Induction

Mathematical induction is a procedure (a logical principle) that can be used to prove a certain kind of mathematical propositions. As a rule, we use mathematical induction to show that a certain statement holds for all positive integers (or for all positive integers greater than a given integer). Formally, the procedure can be given the form of the following theorem. (As usual, we denote by N the set of all positive integers.)

5.1. Theorem (the principle of mathematical induction): Let $S(n)$ be a mathematical statement which depends upon integers ($n \in N$). Let the following two propositions hold true:

Step 1: The statement $S(n_0)$ is true for some $n_0 \in N$.

Step 2: If $S(k)$ is true ($k \geq n_0$), then $S(k+1)$ is true.

Then $S(n)$ is true for each n such that $n \geq n_0$.

PROOF: Suppose that $S(n)$ is false for some $n,\ n \geq n_0$. Put $M = \{n \in N | S(n) \text{ is false}\}$. Then M is nonempty. Take *the least number*, m , so that $m \in M$. Then $S(m)$ is false. If $m = n_0$, then we have a contradiction with Step 1. Thus, $m > n_0$. Then $m-1 \geq n_0$ and, by our assumption, $S(m-1)$ is true (the number m was the least number for which $S(m)$ is false!). By Step 2, $S(m)$ is true. This is a contradiction. We have proved that $S(n)$ is true for any $n,\ n \geq n_0$.

5.2 Example. Prove that for any $n \in N$,

$$1 + 3 + 5 + \cdots + (2n-1) = n^2.$$

SOLUTION: If $n_0 = 1$, then the statement holds true ($1 = 1^2$). We have verified Step 1. Suppose now that the statement is true for some $k \in N$. We want to show that $S(k+1)$ is also true. Thus, we want to show that

$$1 + 3 + 5 + \cdots + (2k-1) + (2(k+1)-1) = (k+1)^2,$$

having assumed (inductive assumption!) that

$$1 + 3 + 5 + \cdots + (2k-1) = k^2.$$

Let us consider the left-hand side of $S(k+1)$,

$$L = 1 + 3 + 5 + \cdots + (2k-1) + (2(k+1)-1),$$

and the right-hand side of $S(k+1)$, $R = (k+1)^2$. Let us substitute into L from our inductive assumption. We consecutively obtain

$$L = 1+3+5+\cdots+(2k-1)+(2(k+1)-1) = k^2 + (2(k+1)-1) = k^2+2k+1 = (k+1)^2 = R.$$

We see that $L = R$ and the proof is complete.

5.3. Example. Let $n \in N$, $n \geq 4$. Prove that the number of all diagonals in a convex plane figure with n vertices is $\frac{1}{2}n(n-3)$.

Solution: If $n_0 = 4$ then the figure obviously possesses 2 diagonals (see Fig. A1 a) below). This agrees with the formula. Thus Step 1 has been verified. Suppose that for the figure with k vertices we have $\frac{1}{2}k(k-3)$ diagonals. Compute now the number of diagonals provided a figure with $n+1$ vertices is given (see Fig. A1 b) below). Consider a vertice, A, of this figure and restrict ourselves to the k-verticed figure $\ldots B'D'BCC'\ldots$. Applying our induction assumption to this smaller figure, we easily see that the number of diagonals in the original figure is $\frac{1}{2}k(k-3)+(k-2)+1$.

Working with this formula, we obtain

$$\frac{1}{2}k(k-3)+(k-2)+1 = \frac{1}{2}k^2-\frac{3}{2}k+k-1 = \frac{1}{2}k^2-\frac{1}{2}k-1 = \frac{1}{2}(k+1)(k-2) = \frac{1}{2}(k+1)((k+1)-3).$$

This verifies Step 2 and we are done.

Fig. A1

5.4 Example. Suppose that a costumer has only the 3 crown coins and the 5 crown coins (he, however, has an unlimited supply of both types of coins $\cdots$). Prove that with those coins only he can pay any cost which is greater than or equal to 8 crowns.

Solution: Reformulating our problem in the formal mathematical language, we have to show that each number greater than or equal to 8 can be expressed in the form of $3m+5n$, where m, n are (not necessarily positive) integers. We shall prove this by induction. Here

$n_0 = 8$ and, indeed, $8 = 3 + 5$ (Step 1). Suppose that k is a given integer with $k > 8$. Suppose moreover that for this k our proposition is valid. Thus, $k = 3m_1 + 5n_1$ for some integers m_1, n_1. Let us consider the number $k + 1$. If $n_1 \neq 0$, then

$$k + 1 = 3n_1 + 5n_1 + 1 = 3n_1 + 5n_1 + 3 \cdot 2 - 5 = 3(n_1 + 2) + 5(n_1 - 1).$$

If $n_1 = 0$, then $k + 1 = 3n_1$, where $n_1 \geq 3$ (we know that $k > 8$). Thus,

$$k + 1 = 3n_1 + 1 = 3(n_1 - 3) + 9 + 1 = 3(n_1 - 3) + 10 = 5 \cdot 2 + 3(n_1 - 3).$$

We have verified Step 2 and the proof is complete.

Let us try to sort out our experiences from the examples we worked out. The difficulty of mathematical induction lies in Step 2 (Step 1 is usually easy to verify). The logical subtlety is that we *assume that the statement is valid for* $n = k$ (we do not prove it!). This assumption – called often the induction hypothesis – is then used to *show that the statement is valid for* $n = k + 1$. Thus, in Step 2 of mathematical induction we verify that *the logical implication* $[\,S(k) \Rightarrow S(k+1)\,]$ *is true.*

The appeal and the power of mathematical induction is that we do not have to prove each case of $S(n)$ separately. Instead, we prove $S(n)$ for a first case, assume $S(k)$ for an arbitrary case and then prove $S(k+1)$. As a result, these two cases take care of infinitely many cases. This is quite an extraordinary phenomenon.

Since mathematical induction needs quite a bit of practicing, and no more theory is needed, we allow ourselves to conclude this section by a collection of problems. Let us only make a final impressimistic remark. As powerful and universal mathematical induction may be, many problems concerning ordinals can be solved by other methods, too. Sometimes it is shorter and more elegant. Here is one example.

5.5 Example. Let n be an ordinal. Prove that the number $n^3 - 4n$ is always divisible by 3.

SOLUTION:

First way: Let us prove it by induction. If $n_0 = 1$, then $n_0^3 - 4n_0 = 0$ and 0 is divisible by any number. So much for Step 1. Considering Step 2, assume that $k^3 - 4k$ is divisible by 3. We must prove then $(k+1)^3 - 4(k+1)$ is divisible by 3. We consecutively obtain

$$(k+1)^3 - 4(k+1) = k^3 + 3k^2 + 3k + 1 - 4k - 4 = (k^3 - 4k) + (3k^2 + 3k - 3).$$

This is a sum of two numbers each of which is divisible by 3. Thus, $(k+1)^3 - 4(k+1)$ is divisible by 3 and the proof is complete.

Second way: Since an arbitrary integer can be expressed in either of the forms $3k - 1, 3k$ or $3k + 1$ $(k \in N)$, it is sufficient to show that $n^3 - 4n$ is divisible by 3 when we substitute either $3k - 1, 3k$ or $3k + 1$ for n. This is evident for $n = 3k$. Let us show it for $n = 3k - 1$ (the case of $n = 3k + 1$ verifies analogously). We obtain

$$\begin{aligned} n^3 - 4n &= (3k-1)^3 - 4(3k-1) = (3k)^3 - (3k)^2 + 3k - 1 - 4 \cdot 3k + 4 \\ &= (3k)^3 - (3k)^2 + (3k) - 3 \cdot 4k - 3 = 3\ell \qquad\qquad (\ell \in N). \end{aligned}$$

This completes the proof.

Third way. Since the number $(n-1)n(n+1)$ is always divisible by 3, for the set $\{n-1, n, n+1\}$ must obviously contain an integer divisible by 3, we easily see that the number $n^3 - 4n$ is divisible by 3 provided the number $(n^3 - 4n) - ((n-1)n(n+1))$ is divisible by 3. Let us compute the output of the latter formula. We obtain

$$n^3 - 4n - ((n-1)n(n+1)) = n^3 - 4n - ((n^2-1)n) = n^3 - 4n - n^3 + n = -3n.$$

The proof is complete.

Problems

In problems 1–18 prove the given formula using mathematical induction. Unless stated otherwise, assume that n is a positive integer.

1. $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$
2. $2 + 5 + 8 + \cdots + (3n-1) = \frac{n(3n+1)}{2}$
3. $1 + 3 + 9 + \cdots + 3^n = \dfrac{3^{n+1} - 1}{2}$
4. $1 \cdot 2 + 3 \cdot 4 + 5 \cdot 6 + \cdots + (2n-1)2n = \dfrac{n(n+1)(4n-1)}{3}$
5. $n^2 + 3n$ is even
6. $n^3 + 11n$ is divisible by 6
7. $n^3 + (n+1)^3 + (n+2)^3$ is divisible by 9
8. $7^{10n+1} + 4$ is divisible by 11
9. $3n^5 + 5n^3 - 8n$ is divisible by 15
10. $7^{2n} + 4$ is never divisible by 3
11. $(ab)^n = a^n b^n$ $(a, b \in R)$
12. The binomial theorem: $(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + \binom{n}{n}b^n$ $(a, b \in R)$
13. The Moivre's theorem (j denotes the imaginary unit):

$$(\cos\varphi + j\sin\varphi)^n = \cos n\varphi + j\sin n\varphi$$

14. $\sin x + \sin 2x + \cdots + \sin nx = \dfrac{\sin\frac{n+1}{2}x}{\sin\frac{x}{2}} \cdot \sin\dfrac{nx}{2}$ $(x \neq +2k\pi,\ k \in N \cup \{0\})$
15. $2^n > n$
16. $2^n < n!$ for $n \geq 4$
17. $n - 2 < \dfrac{n^2 - n}{12}$ for $n \geq 11$
18. Prove that there are exactly 2^n subsets of a set containing n elements.

19. Prove that there is exactly m^n mappings from a set containing n elements into a set containing m elements.

20. What is wrong with the following proof that each girl in a group of n girls has the same dress as every other girl in the group?

    Step 1: For $n = 1$ the statement is obviously true.

    Step 2: Suppose that the statement is true for $n = k$. That is, suppose that each girl in a group containing k girls has the same dress as every other girl in the group. Let $\{g_1, g_2, \ldots, g_k, g_{k+1}\}$ denote the $k+1$ girls in a group G. Let $G_1 = \{g_1, g_2, \ldots, g_k\}$ and $G_2 = \{g_2, g_3, \ldots, g_{k+1}\}$. The both G_1 and G_2 contain k girls, so the girls in each set have the same dress. But the groups G_1 and G_2 overlap. As a result, all girls $g_1, g_2, \ldots, g_k, g_{k+1}$ have the same dress. This completes the alledged proof.

21. This is an example of a problem when a simple solution is available but mathematical induction does not seem to help: Prove that $n^4 + 4$ is never a prime number provided $n \geq 2$. (Hint: Add and subtract the factor $4n^2$.)

Chapter 6

Appendix 2: Polynomials and rational functions

6.1. Preliminaries

In this chapter we study some algebraic properties of polynomials and rational functions. We restrict ourselves to *real polynomials* (we will, however, often deal with their complex roots). The results are applicable to many problems of engineering.

6.1.1. Definition. By a polynomial p we mean a real-valued function $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = \sum_{i=0}^{n} a_ix^i$ defined on R, where $a_0, a_1, \ldots, a_n$ are real constants. These constants are called *the coefficients of the polynomial* $p(x)$. The number n (i.e. , the greatest index for which the corresponding coefficient is nonzero) is called *the degree of* $p(x)$ (in symbols, $n = \deg p(x)$). If $p(x) = 0$, we say that $p(x)$ does not have a degree, and we call this polynomial the zero polynomial.

Let us endow the set P of all polynomials with the algebraic structure of addition, multiplication and scalar multiplication. We will do it in a natural way, obtaining a linear space (see Chap. 1). For simplicity, let us write p and q for the polynomials (functions) $p(x)$ and $q(x)$. Let us define, for polynomials p and q, the functions $p+q,\ p\cdot q$ and $c\cdot p\ (c \in R)$ as follows (the variable x varies over all real numbers):

$$\begin{aligned}(p+q)(x) &= p(x) + q(x),\\ (p\cdot q)(x) &= p(x)\cdot q(x), \text{ and}\\ (c\cdot p)(x) &= c\cdot p(x).\end{aligned}$$

It is easy to see that the functions $p+q, p\cdot q$ and $c\cdot p$ are again polynomials. Indeed, if $p(x) = \sum_{i=0}^{n} a_ix^i$ and $q(x) = \sum_{i=0}^{m} b_ix^i$ and if $n \geq m$, then

$$\begin{aligned}(p+q)(x) &= \sum_{i=1}^{m}(a_i + b_i)x^i + \sum_{i=m}^{n} a_ix^i,\\ (p\cdot q)(x) &= \sum_{r=0}^{m+n}(a_rb_0 + a_{r-1}b_1 + \cdots + a_1b_{r-1} + a_0b_r)x^r, \quad \text{and}\\ (c\cdot p)(x) &= \sum_{i=0}^{n}(ca_i)x^i.\end{aligned}$$

(If p is the zero polynomial, then $p+q=q$, $p\cdot q=0$ and $c\cdot p=0$.) It is easily seen that $\deg(p+q)\le\max(\deg p,\deg q)$, $\deg(p\cdot q)=\deg p+\deg q$ and $\deg(c\cdot p)=\deg p$ $(c\neq 0)$.

6.2. Dividing a polynomial by a polynomial

Let us take up the dividing of polynomials. Obviously, sometime we cannot divide "without a remainder". Let us say that p *is divisibly by* q (in symbols, $q|p$), if we can write $p=q\cdot r$ for a polynomial r. If $p=0$, then p is obviously divisible by any polynomial. If $p\neq 0$ and if p is divisible by q, then the polynomial r is determined uniquely.

Let us first recall the algorithm of dividing a polynomial by a polynomial, finding also an explicit method for computing the remainder. The procedure very much resembles the ordinary division of integers. Let us work some examples.

6.2.1. Example. Let $p=2x^5-2x^3-x+3$ and $q=x^3-x^2+x-1$. Divide p by q and find the remainder.

Solution: The following computation suggests how we proceed. If the reader does not know this procedure, let us only notice that in the procedure one always deals with the highest powers under consideration. Note also that if a polynomial does not explicitely have a power lower than its degree, it is because the corresponding coefficient is zero, of course. Let us now embark on the dividing manual.

$$
\begin{array}{llllll}
(2x^5 & & -3x^3 & & -x & +3)\ :\ (x^3-x^2+x-1)=2x^2+2x-3 \\
\ \ 2x^5-2x^4 & & +2x^3 & +2x^2 & & \\
\hline
\multicolumn{3}{l}{(\underline{\text{subtraction}}\ )} & & & \\
 & 2x^4 & -5x^3 & +2x^2 & -x & +3 \\
 & 2x^4 & -2x^3 & +2x^2 & -2x & \\
\hline
\multicolumn{3}{l}{\quad(\underline{\text{subtraction}})} & & & \\
 & & -3x^3 & & +x & +3 \\
 & & -3x^3 & +3x^2 & -3x & +3 \\
\hline
 & & \multicolumn{2}{l}{(\underline{\text{subtraction}}\ )} & & \\
 & & & -3x^2 & +4x &
\end{array}
$$

We see that the remainder is $-3x^2+4x$. The result of what we have done can be summarized as follows:

$$(2x^5-3x^3-x+3)\ :\ (x^3-x^2+x-1)=2x^2+2x-3+\frac{-3x^2+4x}{x^3-x^2+x-1}.$$

Let us remark that we usually automatically change the signs after settling the respective right-hand side coefficient and performing "the reverse multiplication". This economizes the procedure to a certain extent. The procedure then approximately looks as follows:

6.2.2. Example. Let $p = x^4 + 4$ and $q = x^2 + 2x + 2$. Divide p by q and find the remainder.

SOLUTION:

$$
\begin{array}{rrrrrl}
(x^4 & & & & +4) & : \quad (x^2+2x+2) = x^2 - 2x + 2 \\
-x^4 & -2x^3 & -2x^2 & & & \\
\cline{1-3}
 & -2x^3 & -2x^2 & & +4 & \\
 & 2x^3 & +4x^2 & +4x & & \\
\cline{2-5}
 & & 2x^2 & +4x & +4 & \\
 & & -2x^2 & -4x & -4 & \\
\cline{3-5}
 & & & & 0 &
\end{array}
$$

We have found out in the latter example that the remainder is 0. Thus, the polynomial $x^4 + 4$ is divisible by the polynomial $x^2 + 2x + 2$ and we have

$$(x^4 + 4) = (x^2 + 2x + 2)(x^2 - 2x + 2).$$

Let us now formulate the above procedure in a more rigorous language (the skill we acquired from the previous examples is useful but it may be too thin an ice to safely scate on). The proof is a nice example of a mathematical induction reasoning.

6.2.3. Theorem. Let p, q be two polynomials and let $q \neq 0$. Then we can find polynomials r and s so that the following two properties hold:

(i) $p = q \cdot r + s$,

(ii) $\deg s < \deg q$.

Moreover, the polynomials r and s are uniquely determined by the above properties (i) and (ii) (i.e. if $p = q \cdot r_1 + s_1$ and $\deg s_1 < \deg q$, then $r = r_1$ and $s = s_1$).

PROOF: We will proceed by induction over the degree of p. The idea of the proof is based on the dividing procedure examined above.

Suppose that p and q are the given polynomials. We are looking for the polynomials r and s. Let $S(n)$ be the following statement ($n \geq \deg q$). If $\deg p < n$, then there exist polynomials r and s with the properties from the above theorem. Let $n_0 = \deg q$. Then $S(n_0)$ is obviously true (we put $r = 0$ and $s = q$. Suppose that $k > n_0$ and $S(k)$ is true. We are to show that $S(k+1)$ is also true. In other words, we are to show that if the polynomials r and s exist for any p with $\deg p \leq k$, then such polynomial exist for any p with $\deg p = k + 1$. Let p be a polynomial with $\deg p = k + 1$. Put $l = n - \deg q$ and set $c = a/b$, where a (resp. b) is the coefficient by the highest power of p (resp. q). Then the polynomials p and $c \cdot x^k \cdot q$ have the degree equal to $k + 1$ and, also, they have the same coefficient by the highest power. It follows that the polynomial $\tilde{p} = p - c \cdot x^k \cdot q$ has the degree less than or equal to k. According to the induction assumption, there exist polynomials $\tilde{r}$ and $\tilde{s}$ so that $\deg s < \deg q$ and moreover, $\tilde{p} = p - c \cdot x^k \cdot q = q\tilde{r} + \tilde{s}$. This yields the equality $p = q \cdot (\tilde{r} + cx^k) + \tilde{s}$. We see that the polynomial r defined by $r = \tilde{r} + cx^k$ and the polynomial s defined by $s = \tilde{s}$ are polynomials we look for.

It remains to be proved that the polynomials r and s are defined uniquely. But this is easy. If r_1 and s_1 are other polynomials with the required property, then $q \cdot (r - r_1) = s_1 - s$

and $\deg q + \deg(r - r_1) = \max\{\deg s, \deg s_1\} < \deg q$. This inequality is obviously satisfied only if $r - r_1 = 0$. Thus, $r = r_1$ and $s = s_1$, which we wanted to show.

6.3. Horner's schema

There is an efficient procedure for computing the values $p(x)$ of a polynomial p in a given "independent variable" x. This procedure is called *Horner's schema.* Since this procedure also gives, as a by-product, the remained after dividing by the elementary polynomial $x - a$, where a is a real or a complex number, we will in addition find this schema handy in simplifying a given polynomial. (We will frequently use the following fact: If two polynomials are equal as functions on R, then they have equal coefficients. We will provide a rigorous proof of this fact later.)

Suppose that p is a given polynomial. Suppose that $\deg p > 0$. Suppose that a is a real or complex number. Then $x - a$ is a polynomial with $\deg(x - a) = 1$ and we therefore can write, for some polynomials r and s, $p = (x - a)r + s$, where $\deg s = 0$. Thus, s is a constant polynomial. If we now substitute $x = a$, we see that $s = p(a)$. As a consequence, for any $x \in R$ we obtain the equality $p(x) = (x - a)r(x) + p(a)$. Equivalently, we have $x \cdot r(x) + p(a) = p(x) + a \cdot r(x)$.

Consider the latter formua and suppose that $p(x) = \sum_{i=0}^{n} a_i x^i$ and $r(x) = \sum_{i=0}^{n-1} c_i x^i$. Then $x\left(\sum_{i=0}^{n-1} c_i x^i\right) + p(a) = \sum_{i=0}^{n} a_i x^i + a\left(\sum_{i=0}^{n-1} c_i x^i\right)$ and therefore

$$\sum_{i=0}^{n-1} c_i x^{i+1} + p(a) = \sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n-1} a c_i x^i = \sum_{i=0}^{n-1} (a_i + a c_i) x^i + a_n x^n.$$

Comparing the coefficients, we obtain a series of equalities:

$$\begin{array}{cl}
(n) & c_{n-1} = a_n, \\
(n-1) & c_{n-2} = a_{n-1} + a c_{n-1}, \\
(n-2) & c_{n-3} = a_{n-2} + a c_{n-2}, \\
 & \vdots \\
(0) & p(a) = a_0 + a c_0.
\end{array}$$

The above series of equalities can be transparently rewritten in the following way (Horner's schema):

$$\begin{array}{c|cccccc}
 & a_n & a_{n-1} & a_{n-2} & \dots & a_1 & a_0 \\
 & & + & + & & + & + \\
a & & a \cdot c_{n-1} & a \cdot c_{n-2} & \dots & a \cdot c_1 & a \cdot c_0 \\
\hline
 & c_{n-1} & c_{n-2} & c_{n-3} & \dots & c_0 & p(a)
\end{array}$$

The columns of the schema correspond to the equations $(n), (n-1), \dots, (0)$ above (i.e., $c_{n-1} = a_n, a_{n-1} + a \cdot c_{n-1} = c_{n-2}$, etc.). As we see, in computing $p(a)$ we do not have to compute any high powers – we can replace them with a series of simple first-degree formulas which makes the computation more convenient.

Let us illustrate Horner's schema by examples.

6.3.1. Example. Let $p(x) = 5x^8 - 54x^7 - 122x^5 + 12x^4 - 9x^3 - 21x^2 - 13x + 23$. Find $p(a)$ for $a = 11$.

SOLUTION: The Horner's schema for $p(11)$ reads:

$$\begin{array}{c|ccccccccc} & 5 & -54 & 0 & -122 & 12 & -9 & -21 & -13 & 23 \\ 11 & & 55 & 11 & 121 & -11 & 11 & 22 & 11 & -22 \\ \hline & 5 & 1 & 11 & -1 & 1 & 2 & 1 & -2 & 1 \end{array}$$

It follows that $p(11) = 1$. (Using Horner's schema, one can check that $p(10) = -52001207$ and $p(12) = 184863467$. Thus, $p(x)$ must have a root in the interval $(10, 12)$.)

It is obvious that for a polynomial p, the value of $p(a)$ is exactly the remainder which one obtains after dividing p by $(x - a)$. Moreover, when we compare Horner's schema with the mechanism of dividing a polynomial by a polynomial, we see that the coefficients in the bottommost line are the coefficients of "the partial polynomial" after dividing by $(x-a)$. Thus, Horner's schema also allows for the following application.

6.3.2. Example. Find the remainder after dividing the polynomial

$$p(x) = 4x^7 - 5x^6 + 4x^4 - 12x^3 + 2x^2 - x - 3$$

by the polynomial $(x - 2)$.

SOLUTION: In view of the remark above, the remainder is exactly the value of $p(a)$. Let us apply Horner's schema.

$$\begin{array}{c|cccccccc} & 4 & -5 & 0 & 4 & -12 & 2 & -1 & -3 \\ 2 & & 2\cdot 4 & 2\cdot 3 & 2\cdot 6 & 2\cdot 16 & 2\cdot 20 & 2\cdot 42 & 2\cdot 83 \\ \hline & 4 & 3 & 6 & 16 & 20 & 42 & 83 & 163 \end{array}$$

It follows that

$$p(x) = (x - 2)(4x^6 + 3x^5 + 6x^4 + 16x^3 + 20x^2 + 42x + 83) + 163.$$

Thus, $p(2) = 163$.

6.3.3. Example. Let $p(x) = x^4 - 7x^3 + 16x^2 - 15x + 9$. Check that $a = 3$ is a root of $p(x)$.

SOLUTION: We have to check that $p(3) = 0$. Let us use Horner's schema for $p(x)$ and $a = 3$. We obtain

$$\begin{array}{c|ccccc} & 1 & -7 & 16 & -15 & 9 \\ 3 & & 3 & -12 & 12 & -9 \\ \hline & 1 & -4 & 4 & -3 & 0 \end{array}$$

Thus, $p(3) = 0$ and therefore $a = 3$ is a root of $p(x)$.

Since the nature of Horner's schema is purely algebraic, it can be equally well applied to testing the complex numbers for roots of polynomials. (In fact, it can be applied to polynomials with complex coefficients in the same way as to polynomials with real coefficients. In this text, however, we will not systematicaly study the polynomials with complex coefficients.)

6.3.4. Example. Let $p(x) = x^3 - 2x^2 + x + 1$. Find $p(2 - j)$.

SOLUTION:

	3	-2	1	1
$2-j$		$6-3j$	$5-10j$	$-2-26j$
	3	$4-3j$	$6-10j$	$-1-26j$

We see that $p(2 - j) = -1 - 26j$.

6.3.5. Example. Let $p(x) = x^5 + 2x^4 - 16x^3 + 40x^2 - 44x + 24$. Show that the complex number $a = 1 + j$ is a root of $p(x)$.

SOLUTION: Let us use Horner's schema. We obtain

	1	2	-16	40	-44	24
$1+j$		$1+j$	$2+4j$	$-18-10j$	$32+12j$	-24
	1	$3+j$	$-14+4j$	$22-10j$	$-12+12j$	0

It follows that $p(1 + j) = 0$ and therefore $1 + j$ is a root of $p(x)$.

6.4. Roots of polynomials (multiplicity of roots)

Let us first recall some more notions of the algebra of polynomials.

6.4.1. Definition. By *algebraic equation* of the n-th degree we mean the equation

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

where $a_0, a_1, \ldots, a_n$ are real coefficients, $a_n \neq 0$.

By a *root* of an algebraic equation we mean any real or complex number ξ which satisfies the equation when substitute $x = \xi$.

By our definition, an algebraic equation of the n-th degree arises when we take a polynomial of the n-th degree and let it equal to 0. If $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$ is the algebraic equation, then each real or complex number ξ which satisfies $a_n \xi^n + a_{n-1} \xi^{n-1} + \cdots + a_1 \xi + a_0 = 0$ is then a root. Sometimes we simply call ξ a root of the polynomial $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$.

Looking for roots of an algebraic equation is a problem as old as mathematics itself. The most important result obtained is the theorem that every algebraic equation of the n-th order

(= every polynomial of the n-th order) has exactly n roots if one admits the complex numbers and if one countes the multiplicities. This result – the fundamental theorem of algebra – will be discussed in more detail in the next paragraph.

The actual computing of roots may be connected with fairly technical problems. As the degree increases, the technical difficulties change into a principal impossibility to find the roots analyticaly. The situation is as follows:

For $n = 1$, there are no problems at all – if we are given the equation $a_1x + a_0 = 0$ to solve, it is obvious that the only root ξ reads $\xi = -\dfrac{a_0}{a_1}$.

For $n = 2$ we obtain the quadratic equation $a_2x^2 + a_1x + a_0 = 0$. As known, its (only) two solutions are $\xi_{1,2} = \dfrac{1}{2a_2}(-a_1 \pm \sqrt{a_1^2 - 4a_0a_2})$. If the discriminant $D = a_1^2 - 4a_0a_2$ is strictly negative, we obtain complex roots. These roots are then the complex conjugate.

For $n = 3$ we have a cubic equation $a_3x^3 + a_2x^2 + a_1x + a_0 = 0$ to solve. The exact solution can be found (the so called Cardan formulas, found already in 1545) but very often does not give any reasonable information. Sometimes we cannot get rid of radicals of complex numbers even if all roots are known to be real.

The situation is similar for $n = 4$ – the formulas for roots are known but they are even more useless in practise than in the case of $n = 3$. For $n > 4$ there are no formulas for roots (in fact, it is even known that they cannot be analyticaly found at all). Usually we must be content with numerical solutions for roots.

On the other hand, there are some special types of algebraic equations which can always be solved. One of these types is a *binomial equation.* By binomial equation (of the n-th degree) we mean the equation $x^n - a = 0$ ($a \in R$). For these equation we can find a simple formula for all roots ξ_k ($k \leq n$):

$$\xi_{k+1} = \sqrt[n]{|a|} \cdot \left(\cos\frac{\alpha + 2k\pi}{n} + j \cdot \sin\frac{\alpha + 2k\pi}{n}\right),$$

where $k = 0, 1, \ldots, n-1$ and α is the argument of a (thus, α is either 0 or π). The formula si valid for complex a, too (a simple proof follows from Moivre's theorem).

6.4.2. Example. Solve the binomial equation $x^4 + 16 = 0$.

Solution: Here we have $a = -16$ and therefore the argument of a is π. Moreover, $|a| = 16$. The roots then read

$$\xi_{k+1} = 2\left(\cos(\frac{\pi}{4} + k\frac{\pi}{2}) + j \cdot \sin(\frac{\pi}{4} + k\frac{\pi}{2})\right), \quad k = 0, 1, 2, 3.$$

After substituting for k and after simple manipulation we can easily find all roots of our equation:

$$\begin{aligned}
\xi_0 &= \sqrt{2}(1 + j), \\
\xi_1 &= \sqrt{2}(-1 + j), \\
\xi_2 &= \sqrt{2}(-1 - j), \\
\xi_3 &= \sqrt{2}(1 - j).
\end{aligned}$$

Let us finally observe the following simple fact which leaves for a moment the central line of our investigation but which may be useful when we conjecture that an algebraic equation has a rational root. The assumption is that all coefficients of the equation are rational. In this case the equation can be "lifted" to an equation with whole coefficients and then the following result can be applied.

6.4.3. Theorem. Let $\xi = \dfrac{r}{s}$ be a rational number written in the irreducible form. Let $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, where $a_0, a_1, \ldots, a_n$ are whole numbers with $a_0 \neq 0$ and $a_n \neq 0$. If $p(\xi) = 0$, then the (whole) numbers p and q must satisfy the following conditions:

(i) a_0 is divisible by both r and s (in symbols, $r\,|a_0$ and $s\,|a_0$),

(ii) the number $p(1)$ (resp. $p(-1)$) is divisible by $r-s$ (resp. $r+s$) (in symbols, $(r-s)|p(1)$ and $(r+s)|p(-1)$.

Proof: We will prove the first part of the theorem, the second being analogous. Suppose that $\xi = \dfrac{r}{s}$ is a root of $p(x)$. Then

$$a_0 + a_1(\frac{r}{s}) + \cdots + a_{n-1}(\frac{r}{s})^{n-1} + a_n(\frac{r}{s})^n = 0.$$

Multiplying this equation with s^n, we obtain

$$a_0s^n + a_1rs^{n-1} + \cdots + a_{n-1}r^{n-1}s + a_nr^n = 0.$$

This gives us

$$\begin{aligned} a_0s^n &= -r(a_1s^{n-1} + a_2rs^{n-2} + \cdots + a_{n-1}r^{n-2}s + a_nr^{n-1}) \\ a_nr^n &= -s(a_0s^{n-1} + a_1rs^{n-2} + \cdots + a_{n-2}r^{n-2}s + a_{n-1}r^{n-1}). \end{aligned}$$

It follows that r divides a_0s^n and s divides a_nr^n. But r and s are irreducible whole numbers and therefore r divides a_0 and s divides a_n.

6.4.4. Example. Let $p(x) = x^4 + \dfrac{1}{6}x^3 + \dfrac{5}{3}x^2 + \dfrac{1}{3}x - \dfrac{2}{3}$. Find all rational roots of $p(x)$.

Solution: To obtain a polynomial with the whole coefficients, let us pass to $6p(x)$. Let $q(x) = 6p(x)$. Then $q(x) = 6x^4 + x^3 + 10x^2 + 2x - 4$ and the roots of $q(x)$ coincide with the roots of $p(x)$. The polynomial $q(x)$ satisfies the assumptions of the previous Theorem. Suppose that $\xi = \dfrac{r}{s}$ is a root of $q(x)$. Then $\dfrac{r}{4}$ and $\dfrac{s}{6}$. This implies that $r \in \{1, -1, 2, -2, 4, -4\}$ and $s \in \{1, 2, 3, 6\}$. As a result, each rational root ξ of $q(x)$ must belong to the set

$$\{1, -1, 2, -2, 4, -4, \frac{1}{2}, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}, \frac{4}{3}, -\frac{4}{3}, \frac{1}{6}, -\frac{1}{6}\}.$$

We can further shrink the potential set of rational roots by applying the condition (ii) of the previous Theorem. Indeed, $q(1) = 15$ and $q(-1) = 9$ and therefore $(r-s)15$ and $(r+s)9$. This obviously restricts the set of potential rational roots to the set

$$\{2, -2, -4, \frac{1}{2}, -\frac{1}{2}, -\frac{2}{3}\}.$$

By a direct substitution we immediately see that the only rational roots of $q(x)$ are the numbers $\frac{1}{2}$ and $-\frac{2}{3}$. Thus, the answer to our problem is that the only rational roots of the polynomial $x^4 + \frac{1}{6}x^3 + \frac{5}{3}x^2 + \frac{1}{3}x - \frac{2}{3}$ are the numbers $\frac{1}{2}$ and $-\frac{2}{3}$.

Let us carry on the study of roots of polynomials. The following result on complex roots will turn out important in the investigation of *real* polynomials and rational functions.

6.4.5. Theorem. Let $p(x)$ be a (real) polynomial and let ξ be a complex root of $p(x)$. Then the complex conjugate to ξ, $\bar{\xi}$, is also a root of $p(x)$. (A consequence: Each real polynomial of odd degree has at least one real root.)

PROOF: Let $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$. Let us suppose that all coefficients $a_0, a_1, \ldots, a_n$ are real. By elementary properties of complex numbers we have

$$\begin{aligned} p(\bar{\xi}) &= a_0 + a_1\bar{\xi} + a_2\bar{\xi}^2 + \cdots + a_n\bar{\xi}^n = \bar{a_0} + \bar{a_1}\bar{\xi} + \cdots + \bar{a_n}\bar{\xi}^n \\ &= \overline{a_0} + \overline{a_1\xi} + \cdots + \overline{a_n\xi}^n = \overline{a_0 + a_1\xi + a_2\xi^2 + \cdots + a_n\xi^n} = \overline{p(\xi)}. \end{aligned}$$

Since $p(\xi) = 0$, then $p(\bar{\xi}) = 0$.

6.4.6. Proposition. Let $p(x)$ be a polynomial and let ξ be a (real or complex) number. Then ξ is a root of $p(x)$ if and only if $p(x)$ is divisible by the polynomial $x - \xi$.

PROOF: By Theorem 6.2.3 we can write $p = (z - \xi)r + s$, where s is a constant polynomial ($\deg(z - \xi) = 1$). But $p(\xi) = s$ and the proof is complete.

6.4.7. Theorem. Let $p(x)$ be a polynomial and let $\deg p(x) = n$. Then $p(x)$ has at most n mutually distinct roots.

PROOF: Let us prove this theorem by induction over the degree of $p(x)$. Let us denote by $S(n)$ the following statement: Every polynomial whose degree is n possesses at most n roots. Obviously, $S(1)$ is true. Let us suppose that $S(k-1)$ is true ($k \geq 2$). Consider a polynomial $q(x)$ such that $\deg q(x) = k$. Let $R = \{\xi_1, \xi_2, \ldots, \xi_s\}$ be the set of all roots of $q(x)$. If $R = \emptyset$, then $S(k)$ is true. Otherwise there exist a polynomial $r(x)$ so that $q(x) = (x - \xi_1)r(x)$. Then $\deg r(x) = k - 1$ and therefore, by induction assumption, $r(x)$ has at most $(k-1)$ mutually distinct roots. But the numbers $\xi_2, \ldots, \xi_s$ must be roots of $r(x)$. Thus $s \leq r$. This proves the theorem.

6.4.8. Corollary . Let $p(x)$ and $q(x)$ be polynomials both of degree not greater than n. Let $p(y) = q(y)$ for at least $n+1$ mutually distinct (real or complex) numbers y. Then $p = q$. A consequence: If $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + a_1x + \cdots + a_mx^m$ and if $p(y) = q(y)$ for any real number $y \in R$, then $n = m$ and $a_0 = b_0, a_1 = b_1, \ldots, a_n = b_n$.

PROOF: Consider the polynomial $p(x) - q(x)$. This polynomial cannot have more than n roots. This proves the theorem.

Let us come to the case of multiple roots.

6.4.9. Definition. Let $p(x)$ be a polynomial and let ξ be a root of $p(x)$. Let us say that ξ *is a root of multiplicity* k $(k \in N)$ if $p(x)$ is divisible by the polynomial $(x-\xi)^k$ but $p(x)$ is not divisible by $(x-\xi)^{k+1}$. The polynomial $(x-\xi)^k$ is called the k-multiple factor of $p(x)$. The roots of multiplicity 1 are called simple roots, the roots of multiplicity k, $k>1$ are called multiple roots.

It is worth observing that the multiplicity of a root can be also checked by Horner's schema.

6.4.10. Example. Applying Horner's schema, check that $\xi = 2$ is a root of the polynomial

$$p(x) = x^4 - x^3 - 18x^2 + 52x - 40.$$

Find the multiplicity of this root.

Solution: Let us write down Horner's schema as if we wanted to compute $p(2)$. We obtain

$$\begin{array}{c|rrrrr} & 1 & -1 & -18 & 52 & -40 \\ 2 & & 2 & 2 & -32 & 40 \\ \hline & 1 & 1 & -16 & 20 & 0 \end{array} = p(2).$$

Thus, $p(2) = 0$ and therefore $\xi = 2$ is a root of $p(x)$. It follows that $k \geq 1$. Moreover, Horner's schema gives us the equality

$$(x^4 - x^3 - 18x^2 + 52x - 40) : (x-2) = x^3 + x^2 - 16x + 20$$

(observe that the coefficients in the last line of Horner's schema coincide with the coefficients of the polynomial obtained after dividing the original polynomial with the factor $x-2$). If the multiplicity of the root 2 is higher, the number 2 must be a root of the polynomial $x^3 + x^2 - 16x + 20$. This can be again found out by Horner's schema.

$$\begin{array}{c|rrrr} & 1 & 1 & -16 & 20 \\ 2 & & 2 & 6 & -20 \\ \hline & 1 & 3 & -10 & 0 \end{array} = p_1(2)$$

We see that the number 2 is again a root and therefore the multiplicity, k, of the root $\xi = 2$ is at least 2. Repeating the procedure ones more, we obtain

$$\begin{array}{c|rrr} & 1 & 3 & -10 \\ 2 & & 2 & 10 \\ \hline & 1 & 5 & 0 \end{array} = p_2(2)$$

It follows that $k \geq 3$. Going on in the analogous way, we again make use of Horner's schema to get

$$\begin{array}{c|rr} & 1 & 5 \\ 2 & & 2 \\ \hline & 1 & 7 \end{array} = p_3(2)$$

Since $p_3(2) \neq 0$, we see that $k = 3$. We conclude that the number $\xi = 2$ is a root of the polynomial $x^4 - x^3 - 18x^2 + 52x - 40$ of multiplicity $k = 3$. When we understand the mechanism of the repeated use of Horner's schema, we can write it more economicaly:

$$
\begin{array}{r|rrrrr}
 & 1 & -1 & -18 & 52 & -40 \\
2 & & 2 & 2 & -32 & 40 \\
\hline
 & 1 & 1 & -16 & 20 & \underline{0} \\
2 & & 2 & 6 & -20 & \\
\hline
 & 1 & 3 & -10 & \underline{0} & \\
2 & & 2 & 10 & & \\
\hline
 & 1 & 5 & \underline{0} & & \\
2 & & 2 & & & \\
\hline
 & 1 & \underline{7} & & &
\end{array}
$$

We see that

$$(x^4 - x^3 - 18x^2 + 52x - 40) : (x-2)^3 = (x+5).$$

Let us go on with the theory of real polynomials. The following proposition generalizes Theorem 6.4.5.

6.4.11. Proposition. Let $p(x)$ be a real polynomial. If ξ is a root of $p(x)$ of multiplicity k, then the complex conjugate, $\bar{\xi}$, is a root of $p(x)$ of multiplicity k.

PROOF: It suffices to prove that if $p(x)$ is divisible by $(x-\xi)^k$, then $p(x)$ is also divisible by $(x-\bar{\xi})^k$. Suppose that

$$p(x) = (x-\xi)^k(b_0 + b_1x + b_2x^2 + \cdots + b_nx^n),$$

where $b_0, b_1, \ldots, b_n$ are (possibly complex) coefficients. Then by obvious properties of taking the complex conjugate we obtain for any $x \in R$

$$
\begin{aligned}
p(x) = \overline{p(x)} &= \overline{(x-\xi)^k(b_0 + b_1x + b_2x^2 + \cdots + b_mx^m)} \\
&= \overline{(x-\xi)^k}\,\overline{(b_0 + b_1x + b_2x^2 + \cdots + b_mx^m)} \\
&= \overline{(x-\xi)^k}(\overline{b_0} + \overline{b_1}\,\overline{x} + \overline{b_2}\,\overline{x^2} + \cdots + \overline{b_m}\,\overline{x^m}) \\
&= (\overline{x-\xi})^k(\overline{b_0} + \overline{b_1}x + \overline{b_2}x^2 + \cdots + \overline{b_m}x^m) \\
&= (x-\bar{\xi})^k(\overline{b_0} + \overline{b_1}x + \overline{b_2}x^2 + \cdots + \overline{b_m}x^m).
\end{aligned}
$$

6.4.12. Proposition. Let $p(x)$ be a real polynomial. Then the generally complex number ξ is a root of $p(x)$ of complexity k ($k \in N,\ k \geq 1$) if and only if there is a polynomial $q(x)$ such that $q(\xi) \neq 0$ and $p(x) = (x-\xi)^k \cdot q(x)$ for all $x \in R$.

PROOF: The condition is obviously necessary. Let us check the sufficiency of the condition. Suppose that l is the multiplicity of ξ. Then, obviously, $k \leq l$ and there is a polynomial $q_1(x)$ so that, for all $x \in R$, we have the equality $p(x) = (x-\xi)^l q_1(x)$. It follows that $(x-\xi)^l q_1(x) = (x-\xi)^k q(x)$ and therefore $q(x) = (x-\xi)^{l-k} q_1(x)$. If $l > k$, then $q(\xi) = 0$ which we have excluded. This completes the proof.

The fundamental theorem of algebra and its consequences (the decomposition into irreducible polynomials)

In every field of mathematics we can find a result of the status of fundamental theorem. Probably the most fundamental theorem of all is the fundamental theorem of algebra. It states that each (real or complex) polynomial has a (generaly complex) root.

6.4.13. Theorem(*Fundamental theorem of algebra – FTA*). Every (real or complex) polynomial of degree at least one has a (generally complex) root. In other words, for each algebraic equation $a_0+a_1x+a_2x^2+\cdots+a_nx^n=0$, where $a_0,a_1,\ldots,a_n$ are real or complex numbers, $a_n\neq 0$, $n\geq 1$, there exists a (generaly complex) number $\xi\in C$ such that $a_0+a_1\xi+a_1\xi^2+\cdots+a_n\xi^n=0$.

The proof of this theorem is highly nontrivial and we will not provide it here (see a specialized algebraic literature).

Let us explicitely formulate an important consequence of the fundamental theorem of algebra for real polynomials.

6.4.14. Theorem. Let $p(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$ be a real polynomial of degree n and let $\xi_1,\xi_2,\ldots,\xi_r$ be the set of all roots of $p(x)$. Let $k_1,k_2,\ldots,k_r$ be the multiplicities of the respective roots. Then

$$k_1+k_2+\cdots+k_r=n=\deg p(x)$$

and, moreover, for each $x\in R$ we have the identity:

$$p(x)=a_n(x-\xi_1)^{k_1}(x-\xi_2)^{k_2}\ldots(x-\xi_r)^{k_r}.$$

Proof: According to Prop. 6.4.12 and the FTA, a simple inductive argument gives us the identity $p(x)=a_n(x-\xi_1)^{k_1}(x-\xi_2)^{k_2}\ldots(x-\xi_r)^{k_r}$. The rest is obvious.

The following result refers to real polynomials only. It is most important consequence of FTA as far as the investigation of real polynomials (resp. real rational functions) is concerned. (Prior to the theorem, observe that if η is a complex number, then the polynomial $x^2-2(Re\ \eta)x+(\eta)^2$ is a *real* polynomial which has η for a root. Indeed, $x^2-2(Re\ \eta)x+(\eta)^2=(x-\eta)(x-\bar{\eta})$.)

6.4.15. Theorem. Let $p(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$ $(a_n\neq 0)$ be a real polynomial. Let $\xi_1,\xi_2,\ldots,\xi_p$ be all real roots of $p(x)$. Let $\eta_1,\overline{\eta_1},\eta_2,\overline{\eta_2},\ldots,\eta_q,\overline{\eta_q}$ be all complex roots of $p(x)$. Let k_i $(i\leq p)$ be the multiplicities of the respective real roots $\xi_1,\xi_2,\ldots,\xi_p$ and let l_j $(j\leq q)$ be the multiplicities of the respective complex roots. Then

$$k_1+k_2+\cdots+k_p+2(l_1+l_2+\cdots+l_q)=n=\deg p(x)$$

and, moreover, for each $x\in R$ we have the following identity:

$$\begin{aligned} p(x)=&a_n(x-\xi_1)^{k_1}(x-\xi_2)^{k_2}\ldots(x-\xi_p)^{k_p}\cdot\Big(x^2-2(\mathrm{Re}\,\eta_1)x+|\eta_1|^2\Big)^{l_1}\cdot\\ &\cdot\Big(x^2-2(\mathrm{Re}\,\eta_2)x+|\eta_2|^2\Big)^{l_2}\ldots\Big(x^2-2(\mathrm{Re}\,\eta_q)x+|\eta_q|^2\Big)^{l_q2}.\end{aligned}$$

PROOF: By Prop. 6.4.11, the conjugate complex roots have the same multiplicity. Thus, we can organize them in pairs and construct the *real* polynomials $\left(x^2 - 2(\operatorname{Re}\eta_j)x + |\eta_j|\right)^{l_j}$ $(i \le j \le q)$. The rest follows from the previous theorem.

The latter expression of the real polynomial $p(x)$ is called *the decomposition of a polynomial into the product of real irreducible polynomials.* In many problems coming from applied algebra and analysis we need find this decomposition. Obviously, the polynomials occuring in this decomposition are the simplest real polynomials which allow for such a product expression.

Let us now work out some examples in order to acquire certain skill in computing the decomposition into irreducible polynomials. Again, Horner's schema often proves very useful.

6.4.16. Example. Decompose the following polynomial $p(x)$ into the product of real irreducible polynomials:

$$p(x) = x^7 - 4x^6 - 2x^5 + 13x^4 + 5x^3 - 2x^2 + 12x + 9.$$

Find all roots of $p(x)$.

SOLUTION: The polynomial $p(x)$ has a fairly high degree. Without a bit of good luck we would find ourselves in quite a hopeless situation. Attempting to find a root of $p(x)$, which might help us in simplifying the problem, let us conjecture that $p(x)$ has a rational root. According to Theorem 6.4.3, if $\xi = \frac{r}{s}$ and ξ is a root of $p(x)$, then $r|9$ and $s|1$. It follows that $\xi \in \{1, -1, 3, -3, 9, -9\}$. Checking the numbers in order, we first see that $p(1) = 32 \ne 0$. Then we observe that $p(-1) = 0$. Thus, -1 is a root of $p(x)$. Let us check the multiplicity of this root. Making use of Horner's schema, we obtain

	1	−4	−2	13	5	−2	12	9
−1		−1	5	−3	−10	5	−3	−9
	1	−5	3	10	−5	3	9	$\underline{0}$
−1		−1	6	−9	−1	6	−9	
	1	−6	9	1	−6	9	$\underline{0}$	
−1		−1	7	−16	15	−9		
	1	−7	16	−15	9	$\underline{0}$		
−1		−1	8	−24	39			
	1	−8	24	−39	$\underline{48}$			

We see that the number -1 is a root of $p(x)$ of multiplicity 3. As a consequence, there is a (real) polynomial $p_1(x)$ so that $p(x) = (x+1)^3 p_1(x)$. Using the above Horner's schema (or dividing $p(x)$ by $(x+1)^3$), we obtain $p_1(x) = x^4 - 7x^3 + 16x^2 - 15x + 9$. Again, let us assume that $p_1(x)$ has a rational root, some number η. Since this root must be also a root of $p(x)$, we know that $\eta \in \{3, -3, 9, -9\}$. Testing the number 3 by Horner's schema, we obtain

	1	−7	16	−15	9
3		3	−12	12	−9
	1	−4	4	−3	$\underline{0}$
3		3	−3	3	
	1	−1	1	$\underline{0}$	
3		3	6		
	1	2	$\underline{7}$		

We have been lucky again – the number 3 is also a root of $p(x)$ and its multiplicity is 2. As a consequence, we can write $p_1(x) = (x-3)^2 p_2(x)$, where $p_2(x)$ is a (real) polynomial. Using the above Horner's schema (or using an obvious dividing), we see that $p_2(x) = x^2 - x + 1$. Thus, the decomposition reads

$$p(x) = (x+1)^3(x-3)^2(x^2-x+1).$$

Since the roots of the polynomial $x^2 - x + 1$ are the complex numbers $-\frac{1}{2} \pm \frac{\sqrt{3}}{2} j$, we see that all roots of $p(x)$ constitute the set $\{-1, 3, -\frac{1}{2} + \frac{\sqrt{3}}{2} j, -\frac{1}{2} - \frac{\sqrt{3}}{2} j\}$, the respective multiplicities being $3, 2$ and 1.

6.4.17. Example. Decompose the following polynomial $p(x)$ into a product of real irreducible polynomials:

$$p(x) = x^5 + 2x^4 - 16x^3 + 40x^2 - 44x + 24.$$

Find all roots of $p(x)$. A hint: Observe that the complex number $1+j$ is a root of $p(x)$.

Solution: We of course appreciate it very much that we were told a root – if not, there is only a very little chance that we could advance in solving the problem. Knowing that $1+j$ is a root of $p(x)$, we can proceed as before (Horner's schema works for complex numbers, too!). If we do not feel like leaving the area of real numbers, we can also proceed by dividing the polynomial $p(x)$ by the real irreducible polynomial associated with the complex number $1+j$. This polynomial is the polynomial $p_1(x) = (x-(1+j))(x-(1-j))$. Thus, $p_1(x) = x^2 - 2x + 2$. Let us divide $p(x)$ by $p_1(x)$ ($p(x)$ must obviously be divisible by $p_1(x)$). We obtain

$$\begin{array}{l}
(x^5 + 2x^4 - 16x^3 + 40x^2 - 44x + 24) \;:\; (x^2 - 2x + 2) = x^3 + 4x^2 - 10x + 12 \\
\underline{-x^5 + 2x^4 - 2x^3} \\
\qquad 4x^4 - 18x^3 + 40x^2 \\
\qquad \underline{-4x^4 + 8x^3 - 8x^2} \\
\qquad\qquad -10x^3 + 32x^2 - 44x \\
\qquad\qquad \underline{10x^3 - 20x^2 + 20x} \\
\qquad\qquad\qquad 12x^2 - 24x + 24 \\
\qquad\qquad\qquad \underline{-12x^2 + 24x - 24} \\
\qquad\qquad\qquad\qquad 0
\end{array}$$

Thus, $p(x) = (x^2 - 2x + 2)(x^3 + 4x^2 - 10x + 12)$. We will now tackle the polynomial $x^3 + 4x^2 - 10x + 12$. Hoping to detect a rational root of this polynomial (if there is any!), we use Theorem 6.4.3 and find out that the number -6 is a root. This brings us very close to a complete solution. Since the polynomial $x^3 + 4x^2 - 10x + 12$ must be divisible by $(x+6)$, we can again simplify the problem. Performing the division, we obtain

$$\begin{array}{l}
(x^3 + 4x^2 - 10x + 12) \;:\; (x+6) = x^2 - 2x + 2 \\
\underline{-x^3 - 6x^2} \\
\qquad -2x^2 - 10x \\
\qquad \underline{2x^2 + 12x} \\
\qquad\qquad 2x + 12 \\
\qquad\qquad \underline{-2x - 12} \\
\qquad\qquad\qquad 0
\end{array}$$

We see that the decomposition reads $p(x) = (x^2 - 2x + 2)^2(x + 6)$. This implies that the numbers $1 \pm j$ and 6 are all roots of $p(x)$, the former having multiplicity 2 and the latter having multiplicity 1.

6.5. Decomposition of a rational function into partial fractions

By a rational function, $r(x)$, we mean a fraction of two real polynomials (ie. $r(x) = \dfrac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are polynomials). A distinguished position among rational functions enjoy two types of rational functions. These functions called *partial fractions* are rational functions of the form

$$\frac{a}{(x-\xi)^n} \qquad \text{and} \qquad \frac{bx+c}{(x^2+px+q)^n},$$

where a, ξ, b, c are arbitrary real numbers and p, q are real numbers with $p^2 - 4q < 0$ ($m, n \in N$). The following important result is in force.

6.5.1. Theorem. Every rational function $r(x) = \dfrac{p(x)}{q(x)}$, where $\deg p(x) < \deg q(x)$, can be written as a sum of partial fractions.

The above theorem, which we have first formulated in loose terms to indicate the general idea, plays a considerable rôle in several areas of algebra and analysis. Let us first rewrite this theorem slightly more rigorously in order to find a concrete link of the rational function $\dfrac{p(x)}{q(x)}$ in question with the form of the corresponding partial fractions. We will find that the construction of the partial fractions in the division of $\dfrac{p(x)}{q(x)}$ will depend upon the roots of $q(x)$. The general formulation of the theorem looks perhaps intimidating but it is essentially simple.

6.5.2. Theorem(Decomposition of rational function into partial fractions): Let $r(x) = \dfrac{p(x)}{q(x)}$ be a rational function and let $\deg p(x) < \deg q(x)$. Let

$$q(x) = b \cdot \prod_{i=1}^{r} (x - \xi_i)^{k_i} \cdot \prod_{j=1}^{s} (x^2 + p_j x + q_j)^{\ell_j}$$

be an expression of $q(x)$ in the form of a product of irreducible real polynomials. In other words, let $\xi_0, \xi_1, \dots, \xi_r$ be the set of all mutually distinct real roots of $q(x)$ and let $x^2 + p_j x + q_j = (x - \eta_j)(x - \bar{\eta}_j)$, where $\eta_1, \bar{\eta}_1, \eta_2, \bar{\eta}_2, \dots, \eta_s, \bar{\eta}_s$ is the set of all mutually distinct complex roots of $q(x)$. Then we can find *real* constants

$$\begin{array}{cccccccccc}
A_{11}, & \dots, & A_{1k_1}, & A_{21}, & \dots, & A_{2k_2}, & \dots, & A_{r1}, & \dots, & A_{rk_r} \\
B_{11}, & \dots, & B_{1\ell_1}, & B_{21}, & \dots, & B_{2\ell_2}, & \dots, & B_{r1}, & \dots, & B_{r\ell_s} \\
C_{11}, & \dots, & C_{1\ell_1}, & C_{21}, & \dots, & C_{2\ell_2}, & \dots, & C_{r1}, & \dots, & C_{r\ell_s}
\end{array}$$

so that

$$r(x)=\frac{p(x)}{q(x)}=\sum_{i=1}^{r}\left(\frac{A_{i1}}{x-\xi_i}+\frac{A_{i2}}{(x-\xi_i)^2}+\cdots+\frac{A_{ik_i}}{(x-\xi_i)^{k_i}}\right)$$
$$+\sum_{i=1}^{s}\left(\frac{B_{j_1}x+C_{j1}}{x^2+p_jx+q_j}+\frac{B_{j2}x+C_{j2}}{(x^2+p_jx+q_j)^2}+\cdots+\frac{B_{j\ell_j}x+C_{j\ell_j}}{(x^2+p_jx+q_j)^{\ell_j}}\right).$$

Moreover, this decomposition is unique (i.e., the respective coefficients occuring in the numerators are determined uniquely).

We will leave the proof to later on (see the appendix to this chapter). Instead, we will first develop our skill in computing the decomposition into partial fractions. Let us work out some examples to acquire the skill in the "manual" one uses.

The case of real roots without multiplicities

6.5.3. Example. Let $r(x)=\dfrac{3x+13}{x^2+4x+3}$. Write the rational function $r(x)$ as a sum of partial fractions.

SOLUTION: We must first express the denominator of $r(x)$, the polynomial x^2+4x+3, as a product of (real) irreducible polynomials. To this aim, let us find the roots of the polynomial x^2+4x+3. We easily see that $\xi_{1,2}=\dfrac{-16\pm\sqrt{16-12}}{6}$ and therefore the roots are $\xi_1=-1$ and $\xi_2=-3$. Thus, $x^2+4x+3=(x+1)(x+3)$. It follows (Th. 6.5.2) that there are real coefficients A and B so that, for all $x\in r-\{-1,-3\}$,

$$\frac{3x+13}{x^2+4x+3}=\frac{A}{x+1}+\frac{B}{x+3}.$$

It remains to compute the unknown coefficients A and B. Multiplying the above equation by the polynomial x^2+4x+3, we obtain $3x+13=A(x+3)+B(x+1)$. Since this equality is supposed to hold *identically*, we have (Cor. 6.4.8)

$$\begin{aligned}3 &= \ \ A+B\\ 13 &= 3A+B.\end{aligned}$$

This system has a unique solution $A=5$ and $B=-2$. We see that

$$\frac{3x+13}{x^2+4x+3}=\frac{5}{x+1}-\frac{2}{x+3}.$$

6.5.4. Example. Let $r(x)=\dfrac{x^4-6x^3+12x^2-4x+2}{x^3-6x^2+11x-6}$. Simplify $r(x)$ using partial fractions.

SOLUTION: Observe that in $r(x)$ the degree of numerator is greater than the degree of denominator. We therefore first apply the partial division to pass to a "purely" rational function. We easily see (compute on your own!) that

$$(x^4-6x^3+12x^2-4x+2):(x^3-6x^2+11x-6)=x+\frac{x^2+2x+2}{x^3-6x^2+11x-6}.$$

Let $s(x) = \dfrac{x^2+2x+2}{x^3-6x^2+11x-6}$. Since $s(x) = \dfrac{p(x)}{q(x)}$ is a rational function with $\deg p(x) < \deg q(x)$, we can express $s(x)$ as a sum of partial fractions. We are first interested in the roots of the denominator. Guessing the root $\xi_1 = 1$, we can apply Horner's schema to obtain $x^3-6x^2+11x-6 = (x-1)(x^2-5x+6)$. This gives us the other roots: $\xi_2 = 2$ and $\xi_3 = 3$. As a consequence, we can write

$$\frac{x^2+2x+2}{x^3-6x^2+11x-6} = \frac{A}{x-1} + \frac{B}{x-2} + \frac{C}{x-3},$$

where A, B and C are (yet unknown) real coefficients. "Lifting" both sides with $x^3-6x^2+11x-6$, we see that

$$x^2+2x+2 = A(x-2)(x-3) + B(x-1)(x-3) + C(x-1)(x-2).$$

The equation must be valid identically and therefore the coefficients of the polynomial on the left-hand side must agree with the coefficients of the polynomial on the right-hand side. We infer that

$$\begin{array}{rcrcrcr} 1 & = & A & + & B & + & C \\ 2 & = & -5A & - & 4B & - & 3C \\ 2 & = & 6A & + & 3B & + & 2C. \end{array}$$

Eliminating (resp. using Cramer's rule if we know it or if we have read this chapter after Chap. 2), we obtain

$$\begin{array}{rcrcrcr} 1 & = & A & + & B & + & C \\ 7 & = & & & B & + & 2C \\ 17 & = & & & & & 2C. \end{array}$$

We can now read the solution in the bottom-up way: $C = \dfrac{17}{2}$, $B = -10$ and $A = \dfrac{5}{2}$. We conclude that

$$\frac{x^4-6x^3+12x^2-4x+2}{x^3-6x^2+11x-6} = x + \frac{5}{2(x-1)} - \frac{10}{x-2} + \frac{17}{2(x-3)}.$$

The case of real roots with multiplicities

6.5.5. Example. Let $r(x) = \dfrac{3x^3+2x^2+x+1}{x^4-2x^2+1}$. Write $r(x)$ as a sum of partial fractions.

SOLUTION: We are first interested in the roots of the denominator polynomial. Since $x^4-2x^2+1 = (x^2-1)^2 = (x-1)^2(x+1)^2$, we may apply Th. 6.5.2. Let $A_{11} = A_1, A_{12} = A_2, A_{21} = B_1, A_{22} = B_2$. Then we can write

$$\frac{3x^3+2x^2+x+1}{x^4-2x^2+1} = \frac{A_1}{x-1} + \frac{A_2}{(x-1)^2} + \frac{B_1}{x+1} + \frac{B_2}{(x+1)^2}.$$

After multiplying with x^4-2x^2+1, we obtain the following polynomial identity:

$$3x^3+2x^2+x+1 = A_1(x-1)(x+1)^2 + A_2(x+1)^2 + B_1(x-1)^2(x+1) + B_2(x-1)^2.$$

Let us now compare the coefficients standing by the respective powers. We obtain

$$\begin{array}{rcrcrcrcrcrcr} x^3 & : & 3 & = & A_1 & & & + & B_1 & & \\ x^2 & : & 2 & = & A_1 & + & A_2 & - & B_1 & + & B_2 \\ x^1 & : & 1 & = & -A_1 & + & 2A_2 & - & B_1 & - & 2B_2 \\ x^0 & : & 1 & = & -A_1 & + & A_2 & + & B_1 & + & B_2 . \end{array}$$

After applying the standard elimination, which we allow ourselves to leave to the reader, we will find out that $A_1 = \frac{7}{4}, A_2 = \frac{7}{4}, B_1 = \frac{5}{4}$ and $B_2 = -\frac{1}{4}$. It follows that the required expression reads

$$\frac{3x^3 + 2x^2 + x + 1}{x^4 - 2x^2 + 1} = \frac{7}{4(x-1)} + \frac{7}{4(x-1)^2} + \frac{5}{4(x+1)} - \frac{1}{4(x+1)^2}.$$

6.5.6. Example. Let $r(x) = \dfrac{2x^3 - 3x^2 + 4x + 6}{2x^4 + 3x^3}$. Write $r(x)$ as a sum of partial fractions.

Solution: Obviously, $2x^4 + 3x^3 = 2x^3(x + \frac{3}{2}) = 2(x-0)^3(x + \frac{3}{2})$. Thus, 0 is a root of multiplicity 3 and $-\frac{3}{2}$ is a root of multiplicity 1. According to Th. 6.5.2, let us look for coefficients A, B, C, D so that

$$\frac{2x^3 - 3x^2 + 4x + 6}{2x^4 + 3x^3} = \frac{A}{x + \frac{3}{2}} + \frac{B}{x} + \frac{C}{x^2} + \frac{D}{x^3}.$$

After simple algebra we have the following polynomial identity:

$$2x^3 - 3x^2 + 4x + 6 = 2Ax^3 + Bx^2(2x+3) + Cx(2x+3) + D(2x+3).$$

This gives us the following system of linear equations:

$$\begin{array}{rcrcrcrcr} 2 = & 2A & + & 2B & & & & \\ -3 = & & & 3B & + & 2C & & \\ 4 = & & & & & 3C & + & 2D \\ 6 = & & & & & & & 3D. \end{array}$$

It follows that $D = 2, C = 0, B = -1$ and $A = 2$. As a consequence, we obtain

$$\frac{2x^3 - 3x^2 + 4x + 6}{2x^4 + 3x^3} = \frac{2}{x + \frac{3}{2}} - \frac{1}{x} + \frac{2}{x^3}.$$

The case of complex roots without multiplicities

6.5.7. Example. Let $r(x) = \dfrac{2x^3 + 3x^2 + 3x + 4}{x^4 + 3x^2 + 2}$. Write $r(x)$ as a sum of partial fractions.

Solution: As always, we are first interested in the roots of the denominator. Putting $x^2 = y$, we have $x^4 + 3x^2 + 2 = y^2 + 3y + 2 = (y+1)(y+2) = (x^2+1)(x^2+2)$. The latter formula is obviously the expression of $x^4 + 3x^2 + 2$ as a product of (real) irreducible polynomials. Thus, the polynomial $x^4 + 3x^2 + 2$ does not have real roots and the desider decomposition into partial fractions reads as follows (Th. 6.5.2):

$$\frac{2x^3 + 3x^2 + 3x + 4}{x^4 + 3x^2 + 2} = \frac{Ax + B}{x^2 + 1} + \frac{Cx + D}{x^2 + 2}.$$

From this equation we infer that

$$2x^3 + 2x^2 + 3x + 4 = (Ax + B)(x^2 + 2) + (Cx + D)(x^2 + 1).$$

Comparing the adequate coefficients we further obtain

$$\begin{array}{rccccc} 2 = & A & + & C & & \\ 3 = & & B & & + & D \\ 3 = & 2A & + & C & & \\ 4 = & & 2B & & + & D. \end{array}$$

After simple computation (use the Gaussian elimination procedure!) we have $D = 2, C = 1, B = 1$ and $A = 1$. This yields the following result:

$$\frac{2x^3 + 3x^2 + 3x + 4}{x^4 + 3x^2 + 2} = \frac{x + 1}{x^2 + 1} + \frac{x + 2}{x^2 + 2}.$$

The case of complex roots with multiplicities

6.5.8. Example. Let $r(x) = \dfrac{x^3 + x^2 + x + 2}{x^4 + 2x^2 + 1}$. Write $r(x)$ as a sum of partial fractions.

SOLUTION: Since $x^4 + 2x^2 + 1 = (x^2 + 1)^2$, we see that the only root of the denominator is the complex unit j (its multiplicity is 2). It follows that the "trial" sum into partial fractions must be written as follows (Th. 6.5.2):

$$\frac{x^3 + x^2 + x + 2}{x^4 + 2x^2 + 1} = \frac{Ax + B}{(x^2 + 1)} + \frac{Cx + D}{(x^2 + 1)^2}.$$

After multiplying with $(x^2 + 1)^2$ we obtain the equality

$$x^3 + x^2 + x + 2 = Ax + B + (Cx + D)(x^2 + 1).$$

Comparing the coefficients standing by the corresponding powers we obtain the following system:

$$\begin{array}{lrccccccc} x^3: & 1 = & & & & C & & \\ x^2: & 1 = & & & & & & D \\ x^1: & 1 = & A & & + & C & & \\ x^0: & 2 = & & B & & & + & D. \end{array}$$

The solution of this system is $C = D = B = 1$, $A = 0$. We conclude that the required expression reads

$$\frac{x^3 + x^2 + x + 2}{x^4 + 2x^2 + 1} = \frac{x + 1}{x^2 + 1} + \frac{1}{(x^2 + 1)^2}.$$

The general case of decomposition into partial fractions

6.5.9. Example. Let $r(x) = \dfrac{5x^2 - 7x + 12}{x^3 - x^2 + 4x - 4}$. Decompose $r(x)$ into a sum of partial fractions.

Solution: Observing that $\xi = 1$ is a root of the denominator and dividing by $x - 1$, we see that $(x^3 - x^2 + 4x - 4) = (x-1)(x^2+4)$. According to Theorem 6.5.2, we can write

$$\frac{5x^2 - 7x + 12}{(x-1)(x^2+4)} = \frac{A}{x-1} + \frac{Bx + C}{x^2+4}.$$

This gives us the equality

$$5x^2 - 7x + 12 = A(x^2+4) + (Bx+C)(x-1).$$

Computing the corresponding coefficients, we obtain

$$\begin{array}{lrclll} x^2: & 5 & = & A & +B & \\ x^1: & -7 & = & & -B & +C \\ x^0: & 12 & = & 4A & & -C \end{array}$$

Simple algebra yields $A = 2, B = 3$ and $C = -4$. We have therefore found the following decomposition:

$$\frac{5x^2 - 7x + 12}{x^3 - x^2 + 4x - 4} = \frac{2}{x-1} + \frac{3x-4}{x^2+4}$$

6.5.10. Example. Let $r(x) = \dfrac{2x^4 + 4x^3 + 6x + 3}{(x-1)^3(x^2+2x+3)}$. Decompose $r(x)$ into a sum of partial fractions.

Solution: Since the polynomial $x^2 + 2x + 3$ does not have real roots, the trial setup reads

$$\frac{2x^4 + 4x^3 + 6x + 3}{(x-1)^3(x^2+2x+3)} = \frac{A}{(x-1)} + \frac{B}{(x-1)^2} + \frac{C}{(x-1)^3} + \frac{Dx+E}{x^2+2x+3}.$$

After simple manipulation we obtain

$$\begin{aligned} 2x^4 + 5x^2 + 6x + 5 = {} & A(x-1)^2(x^2+2x+3) + B(x-1)(x^2+2x+3) + C(x^2+2x+3) + \\ & + (Dx+E)(x-1)^3. \end{aligned}$$

Comparing the corresponding coefficients of the polynomials on the left-hand and right-hand side, we get the following system:

$$\begin{array}{lrcl} x^4: & 2 & = & A \qquad\qquad\quad + D \\ x^3: & 0 & = & \qquad + B \qquad - 3D \; + E \\ x^2: & 5 & = & \qquad + B \; + C + 3D - 3E \\ x^1: & 6 & = & -4A \; + B + 2C \; - D + 3E \\ x^0: & 5 & = & 3A - 3B + 3C \qquad - E. \end{array}$$

Solving this system of linear equations, which we leave to the reader as a useful excercise in Gaussian elimination, we obtain $E = 1, D = 1, C = 3, B = 2$ and $A = 1$. Thus,

$$\frac{2x^4 + 5x^2 + 6x + 5}{(x-1)^3(x^2+2x+3)} = \frac{1}{x-1} + \frac{2}{(x-1)^2} + \frac{3}{(x-1)^3} + \frac{x+1}{x^2+2x+3}.$$

6.5.11. Example. Let $r(x) = \dfrac{x^5 + x^3 + x - 1}{(x^3 - x^2 + x - 1)(x^2 + x + 1)^2}$. Decompose $r(x)$ into a sum of partial fractions.

Solution: Since the polynomial $x^2 + x + 1$ does not have any real roots and since we can write $x^3 - x^2 + x - 1 = (x-1)(x^2+1)$, the decomposition of $r(x)$ reads

$$\frac{x^5 + x^3 + x - 1}{(x-1)(x^2+1)(x^2+x+1)^2} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1} + \frac{Dx+E}{x^2+x+1} + \frac{Fx+G}{(x^2+x+1)^2}.$$

This is equivalent with the equality

$$\begin{aligned} x^5 + x^3 + x - 1 = {}& A(x^2+1)(x^2+x+1)^2 + (Bx+C)(x-1)(x^2+x+1)^2 + \\ & + (Dx+E)(x-1)(x^2+1)(x^2+x+1) + (Fx+G)(x-1)(x^2+1). \end{aligned}$$

Equivalently, the latter equality can be restated in the form

$$\begin{aligned} x^5 + x^3 + x - 1 = {}& (A+B+D)x^6 + (2A - B + C + E)x^5 + (4A - B + C + D + F)x^4 + \\ & + (4A - B + C - D + E - F + G)x^3 + (4A - B - C + 2D - E + F - G)x^2 + \\ & + (2A - B - C - D - F + G)x + (A - C - E - G). \end{aligned}$$

Comparing the coefficients, we obtain

$$\begin{array}{lrl} x^6: & 0 & = A + B + D \\ x^5: & 1 & = 2A - B + C + E \\ x^4: & 0 & = 4A - B + C + D + F \\ x^3: & 1 & = 4A - B + C - D + E - F + G \\ x^2: & 0 & = 4A - B - C + 2D - E + F - G \\ x^1: & 1 & = 2A - B - C - D - F + G \\ x^0: & -1 & = A - C - E - G \end{array}$$

Solving this system by elimination, which is also an excercise in patience and will power, we find

$$G = \frac{1}{3},\ F = \frac{2}{3},\ E = \frac{16}{9},\ D = -\frac{1}{9},\ C = -1,\ B = 0 \quad \text{and} \quad A = \frac{1}{9}.$$

Substituting back and arranging the coefficients, we can finally write

$$\frac{x^5 + x^3 + x - 1}{(x^3 - x^2 + x - 1)(x^2+x+1)^2} = \frac{1}{9(x-1)} - \frac{1}{x^2+1} + \frac{-x+16}{9(x^2+x+1)} + \frac{2x+1}{3(x^2+x+1)^2}.$$

6.5.12. Remark. The latter example showed how tedious the computation of coefficients may be. Thus, after having grasped the strategy of the decomposition into partial fractions, we may think of simplifying the tactics. Here are some useful advices.

First, the number of unknowns in the system can be lowered by directly computing the coefficients corresponding to the simple real roots (i.e. , the roots of multiplicity 1). Let us consider the last example. There we had the equality

$$\frac{x^5+x^3+x-1}{(x-1)(x^2+1)(x^2+x+1)^2}=\frac{A}{x-1}+\frac{Bx+C}{x^2+1}+\frac{Dx+E}{x^2+x+1}+\frac{Fx+G}{(x^2+x+1)^2}.$$

Trying to determine A, let us multiply with the factor $(x-1)$. We obtain the equality

$$\frac{x^5+x^3+x-1}{(x-1)(x^2+1)(x^2+x+1)^2}=A+(x-1)\frac{Bx+C}{x^2+1}+(x-1)\frac{Dx+E}{x^2+x+1}+(x-1)\frac{Fx+G}{(x^2+x+1)^2}.$$

We see that the coefficient A can be now easily computed by substituting $x=1$. Indeed,

$$A=\frac{x^5+x^3+x-1}{(x-1)(x^2+1)(x^2+x+1)^2}\mid x=1\mid=\frac{1}{9}.$$

Further, observing that it is easy to guess the root of x^2+1, we can compute the coefficients B and C by multiplying the original equality by x^2+1. We then obtain

$$\begin{aligned}\frac{x^5+x^3+x-1}{(x-1)(x^2+1)(x^2+x+1)^2}=&(x^2+1)\frac{A}{x-1}+Bx+C+(x^2+1)\frac{Dx+E}{x^2+x+1}\\&+(x^2+1)\frac{Fx+G}{(x^2+x+1)^2}.\end{aligned}$$

If we substitute $x=j$, we obtain the equality $-1=jB+C$. But this can be viewed as an equality of complex numbers. This gives us $B=0$ and $C=-1$. We have already lowered the number of unknowns considerably. In the rest we can either find out the roots of the polynomial x^2+x+1, computing thus the coefficients D and E as a consequence, or go on with elimination. (Let us finally remark that sometimes it is practical to use differentiation of the polynomial equation in question. Indeed, when two polynomials agree everywhere, then so do their derivatives. This can be utilized in the case of real multiple roots though, as the reader may discover, the advantage over the standard elimination procedure is not always persuasive.)

6.6. Appendix to Appendix 2 – The proof of Theorem 6.5.2

In this paragraph we will present the proof of the decomposition theorem into partial fractions. We will proceed by induction, indicating basic ideas (the reader can provide the details of a formal proof). We will first deal with the real roots, then we will consider the complex roots.

Let $r(x)=\dfrac{p(x)}{q(x)}$ be a rational function and let $\deg p(x)<\deg q(x)$. Suppose that ξ is a root of $q(x)$ and suppose that the multiplicity of ξ is k. By Prop. 6.4.12, we can write $q(x)=(x-\xi)^k\cdot s(x)$, where $s(x)$ is a polynomial with the property $s(\xi)\neq 0$. Let $a_0=\dfrac{p(\xi)}{s(\xi)}$. Write $t(x)=\dfrac{p(x)-a_0s(x)}{q(x)}$. Then $t(x)$ is a rational function and $t(\xi)=0$. Thus, ξ is a root of the polynomial $p(x) - a_0 s(x)$. By Prop. 6.4.12 again, there is a polynomial $u(x)$ such that $p(x) - a_0 s(x) = (x-\xi)u(x)$. It follows that

$$t(x) = \frac{u(x)}{q_1(x)},$$

where $q_1(x) = (x-\xi)^{k-1}s(x)$. Thus, $\deg q_1 < \deg q$. Since $t(x) = \dfrac{p(x)}{q(x)} - \dfrac{a_0 s(x)}{q(x)} - r(x) - \dfrac{a_0 s(x)}{q(x)}$, we can write $r(x) = \dfrac{a_0}{(x-\xi)^k} + t(x)$. The rational function $t(x)$ can be again written in the form $\dfrac{a_1}{(x-\xi)^{k-1}} + t_1(x)$, etc. After finitely many steps we obtain the required expression

$$r(x) = \frac{a_0}{(x-\xi)^k} + \frac{a_1}{(x-\xi)^{k-1}} + \cdots + \frac{a_{k-1}}{(x-\xi)} + r_1(x),$$

where $r_1(x)$ is a rational function the denominator of which does not have ξ for its root. In this way we exhaust all real roots, obtaining the part of the fractions corresponding to the real roots. It remains to be shown that a rational function allows for the required decomposition provided all roots of the denominator polynomial are complex. This will be proved in the sequel.

Let $r(x) = \dfrac{p(x)}{q(x)}$ and let $\deg p(x) < \deg q(x)$. Let a complex number, η, be a root of $q(x)$ and let the multiplicity of η be k ($k \geq 1$). By Prop. 6.4.5, the complex conjugate $\bar{\eta}$ is also a root of $q(x)$ of multiplicity k and we can write

$$q(x) = (x^2 - 2(\operatorname{Re}\eta)x + |\eta|^2)^k \cdot s(x),$$

where $s(x)$ is a polynomial with the property $s(\eta) \neq 0$. If we can find such real numbers $a_0, b_0 \in R$ that the complex number η is a root of the polynomial $p(x) - (a_0 x + b_0)s(x)$, then the complex number $\bar{\eta}$ is also a root of $p(x) - (a_0 x + b_0)s(x)$ and we can write

$$p(x) - (a_0 x + b_0)s(x) = (x^2 - 2(\operatorname{Re}\eta)x + |\eta|^2) \cdot u(x).$$

Then we can go on, writing

$$\frac{p(x) - (a_0 x + b_0)s(x)}{q(x)} = \frac{p(x)}{q(x)} - \frac{a_0 x + b_0}{(x^2 - 2(\operatorname{Re}\eta)x + |\eta|^2)^k}.$$

Thus, $r(x) = \dfrac{p(x)}{q(x)} = \dfrac{a_0 x + b_0}{(x^2 - 2(\operatorname{Re}\eta)x + (\eta)^2)} + \dfrac{p(x) - (a_0 x + b_0)s(x)}{q(x)}$ and by the properties of a_0 and b_0, we would have

$$\frac{p(x) - (a_0 x + b_0)s(x)}{q(x)} = \frac{p_1(x)}{q_1(x)},$$

where $q_1(x) = (x^2 - 2(\operatorname{Re}\eta)x + (\eta)^2)^{k-1} \cdot s_1(x)$ for some $s_1(x)$ with the property $s_1(\eta) \neq 0$. Since $\deg q_1 < \deg q$, we can complete the proof because, after finitely many steps, we have

$$r(x) = \frac{a_0 x + b_0}{(x^2 - 2(\operatorname{Re}\eta)x + (\eta)^2)^k} + \frac{a_1 x + b_1}{(x^2 - 2(\operatorname{Re}\eta)x + (\eta)^2)^{k-1}} + \cdots + \frac{a_{k-1}x + b_{k-1}}{(x^2 - 2(\operatorname{Re}\eta)x + (\eta)^2)} + r_1(x),$$

where η is not a root of the denominator polynomial of $r_1(x)$. In this way we can exhaust all roots, concluding the proof.

It remains to be verified that the desired real numbers a_0, b_0 can be found. These numbers must satisfy the condition $p(\eta) - (a_0\eta + b_0)s(\eta) = 0$. We know that $s(\eta) \neq 0$. Put $\delta = \dfrac{p(\eta)}{s(\eta)}$. Then it can be easily shown that it suffices to set

$$a_0 = \frac{\operatorname{Im}\delta}{\operatorname{Im}\eta}, \qquad b_0 = \operatorname{Re}\delta - \operatorname{Re}\eta\frac{\operatorname{Im}\delta}{\operatorname{Im}\eta}$$

(this rather technical but essentialy simple computation is left to the reader). With these numbers a_0 and b_0 we will have $p(\eta) - (a_0\eta + b_0)s(\eta) = 0$, and this completes the last step of the proof.

Problems

1. Divide the polynomial $p(x)$ by the polynomial $q(x)$ and find the remainder.
   a) $p(x) = x^4 - x^3 + x^2 - 2x,\ \ q(x) = x - 1$;
   b) $p(x) = 2x^5 - x^4 + 4x^3 + 3x^2 - x + 1,\ \ q(x) = x^3 + x^2 - x + 1$.
2. Using Horner's schema, find $p(a)$ and $p(b)$, where $p(x) = 2x^4 - 3x^3 + 5x^2 - x + 5$ and $a = 3,\ b = \dfrac{1}{2}$.
3. Using Horner's schema, find $p(a)$, where $p = x^5 - 6x^4 + 10x^3 + 12x^2 - 55x + 49,\ a = 2 + j$.
4. Guessing one of the roots and applying a consecutive lowering of degree, express the given polynomial $p(x)$ as a product of real irreducible polynomials. Find all roots of $p(x)$ together with their multiplicities:
   $$p(x) = x^5 - 4x^3 - 2x^2 + 3x + 2.$$
5. Guessing one root first, find all roots of the following algebraic equations:
   a) $3x^3 - 7x^2 - 7x + 3 = 0$,
   b) $-2x^4 + 3x^3 + 4x^2 - 3x - 2 = 0$.
6. Find all roots of the following algebraic equation. (A hint: The complex number $\dfrac{3}{4} - \dfrac{5}{4}j$ is a root of the equation).
   $$8x^4 - 12x^3 - 9x^2 - 17 = 0.$$
7. Using Moivre's theorem, write the following polynomial as a product of real irreducible polynomials:
   a) $x^4 + 4$, b) $x^6 - 1$.
8. Find all rational roots of the following polynomials (use Theorem 6.4.3):
   a) $x^3 - 6x^2 + 12x - 9$,
   b) $36x^4 - 5x^2 - 1$,
   c) $6x^5 - 11x^4 - 33x^3 + 33x^2 + 11x - 6$.

9. Find the multiplicity of the root ξ of the polynomial $p(x)$:
   a) $\xi = -2,\ p(x) = x^6 + x^5 - 11x^4 - 13x^3 + 26x^2 + 20x - 24$,
   b) $\xi = 1 - 2j,\ p(x) = x^5 - 5x^4 + 18x^3 - 34x^2 + 45x - 25$.

In the following examples, decompose the given rational function into partial fractions (perform first the partial division if the degree of numerator is greater than the degree of denominator).

10. $\dfrac{-2x+5}{x^2-3x+2}$

11. $\dfrac{x^4+5x-14}{x^2-4}$

12. $\dfrac{2x^2+8x-1}{x^3+4x^2+x-6}$

13. $\dfrac{5x^4-15x^2+4}{x^5+5x^3+4x}$

14. $\dfrac{x^3+9x^2+21x+17}{x^4+7x^3+17x^2+17x+6}$

15. $\dfrac{x^5+2x^4+2x^3+6x^2+4}{x^3+2x^2}$

16. $\dfrac{-x^2+x}{x^3+3x^2+3x+1}$

17. $\dfrac{5x^2-7x+12}{x^3-x^2+4x+4}$

18. $\dfrac{x^2-x}{x^5-2x^4+2x^3-x^2+2x-2}$
    (Hint: $\xi = 1 - j$ is a root of denominator.)

19. $\dfrac{3x^2+4x-1}{x^3-1}$

20. $\dfrac{2x^3+3x^2+3x+4}{x^4+3x^2+2}$

21. $\dfrac{-x^4+2x^2-5x+4}{x^5-x^4+x^3-x-1}$

22. $\dfrac{x^3}{(x^2+x+1)^2}$

23. $\dfrac{x^5+6x^3-3x^2+3x+1}{x^6-2x^5+3x^4-4x^3+3x^2-2x+1}$
    (Hint: Observe that 1 and j are roots of denominator.)

Answers

1. a) $p(x) = q(x)(x^3 + x - 1) - 1$, b) $p(x) = q(x)(2x^2 - 3x + 9) - 11x^2 + 11x - 9$.
2. $p(3) = 128,\ p(\frac{1}{2}) = \frac{11}{2}$
3. $p(2 + j) = 1$.
4. $p(x) = (x - 1)(x + 1)^3(x - 2)$. The roots 1 and 2 are simple, 3 is a root of multiplicity 3.
5. a) $-1,\ \frac{1}{3},\ 3;$ b) $1,\ -1,\ -\frac{1}{2},\ 2$.
6.
7. a) $(x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1)$, b) $(x + 1)(x - 1)(x^2 + x + 1)(x^2 - x + 1)$
8. a) 3; b) $\frac{1}{2}, -\frac{1}{2}$; c) $1, -2, -\frac{1}{2}, 3, \frac{1}{3}$.
9. a) $\xi = -2$ is a root of multiplicity 3, b) $\xi = 1 - 2j$ is a root of multiplicity 2.
10. $\dfrac{1}{x-2} - \dfrac{3}{x-1}$
11. $x^2 + 4 + \dfrac{2}{x+2} + \dfrac{3}{x-2}$
12. $\dfrac{1}{x-1} + \dfrac{2}{x+2} - \dfrac{1}{x+3}$
13. $\dfrac{1}{x} + \dfrac{1}{x+1} + \dfrac{1}{x-1} + \dfrac{1}{x+2} + \dfrac{1}{x-2}$
14. $\dfrac{3}{x+2} + \dfrac{2}{(x+1)^2} - \dfrac{2}{x+3}$
15. $x^2 + 2 + \dfrac{3}{x+2} - \dfrac{1}{x} + \dfrac{2}{x^2}$
16. $-\dfrac{1}{x+1} + \dfrac{3}{(x+1)^2} - \dfrac{2}{(x+1)^3}$
17. $\dfrac{2}{x-1} + \dfrac{3x-4}{x^2+4}$
18. $\dfrac{x-3}{13(x^2+x+1)} + \dfrac{-x+6}{13(x^2-2x+2)}$
19. $\dfrac{2}{x-1} + \dfrac{x+3}{x^2+x+1}$
20. $\dfrac{x+1}{x^2+1} + \dfrac{x+2}{x^2+2}$
21. $\dfrac{-1+x}{x^2-x+1} + -2x - 3x^2 + x + 1$
22. $\dfrac{x-1}{x^2+x+1} + \dfrac{1}{(x^2+x+1)^2}$

Acknowledgement

I am grateful to my colleagues V. Bartík, V. Maňasová, P. Olšák, J. Tkadlec and D. Valášek for valuable comments on the textbook. Also, I am indepted to my students R. Asimeng and F. A. Mensah for correcting many mistakes in the preliminary version. I also thank to Mrs. E. Fiedlerová for typing the textbook in LaTeX and to V. Zýka for drawing figures. Finally, my deep thank to B. Williams for several suggestions on the organization and the style of the textbook.