EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF NUMERICAL ANALYSIS

Computer Science BSc

Basic Mathematics – Linear Algebra

University material

István Csörgő

May 2021

Written by
assist. prof. Dr. István CSÖRGŐ

Vetted by
Zoltán István FILIPP

Contents

1. Matrices

In this chapter we will compute with real or complex number tables. Basically, these tables are what we call matrices.

Let us introduce the notation $\mathbb{K}$ which denotes one of $\mathbb{R}$ or $\mathbb{C}$. It will be useful, because the real and the complex cases can be discussed in parallel.

The algebraic structures of $\mathbb{R}$ or $\mathbb{C}$ is: field (number field). In this sense we can speak about the number field $\mathbb{K}$.

1.1. Theory

As it was written in the introduction, $\mathbb{K}$ denotes one of $\mathbb{R}$ or $\mathbb{C}$.

1.1.1. The Concept of Matrix

1.1. Definition Let m and n be positive integers. The functions

$$A : \{1, \dots, m\} \times \{1, \dots, n\} \to \mathbb{K}$$

are called $m \times n$ matrices (over the number field $\mathbb{K}$). The set of $m \times n$ matrices is denoted by $\mathbb{K}^{m\times n}$. The replacement value $A(i, j)$ of the matrix A at the place (i, j) is called the j-th entry in the i-th row (or the i-th entry of the j-th column), and it is denoted by a_{ij} or by $(A)_{ij}$.

The matrix is called square matrix if $m = n$, that is the number of rows equals the number of columns.

The $m \times n$ matrices are given as $m \times n$ tables (this is the origin of the names row, column):

$$A = \begin{bmatrix} A(1,1) & A(1,2) & \dots & A(1,n) \\ A(2,1) & A(2,2) & \dots & A(2,n) \\ & & \vdots & \\ A(m,1) & A(m,2) & \dots & A(m,n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} (A)_{11} & (A)_{12} & \dots & (A)_{1n} \\ (A)_{21} & (A)_{22} & \dots & (A)_{2n} \\ & & \vdots & \\ (A)_{m1} & (A)_{m2} & \dots & (A)_{mn} \end{bmatrix}.$$

The entries $a_{11}, \ a_{22}, \ \dots$ are called the diagonal entries of the matrix A, the line which connects them is called the main diagonal (or simply diagonal) of A. If the matrix is a square matrix then the diagonal is the same as the geometrical diagonal.

Let us mention some special matrices:

- Zero matrix: all its entries are zero. The zero matrix is often denoted by the symbol 0.
- Row matrix: it has only one row. In other words the elements of $\mathbb{K}^{1\times n}$. The row matrices are often called row vectors.
- Column matrix: it has only one column. In other words the elements of $\mathbb{K}^{m\times 1}$. The column matrices are often called column vectors.

  Later we will speak more about the reason of the names „row vector", „column vector" (see Remark 3.8).
- Lower triangular matrix: all the entries above the diagonal are 0, that is $a_{ij} = 0$ if $j > i$.
- Upper triangular matrix: all the entries below the diagonal are 0, that is $a_{ij} = 0$ if $j < i$.
- Diagonal matrix: all the entries outside of the diagonal are 0, that is $a_{ij} = 0$ if $i \neq j$.

Among the square matrices it is important the unit matrix (or identity matrix):

1.2. Definition The matrix $I \in \mathbb{K}^{n\times n}$ is called $(n \times n)$ unit matrix (or identity matrix), if:

$$(I)_{ij} := \begin{cases} 0 & \text{ha} \quad i \neq j, \\ 1 & \text{ha} \quad i = j \end{cases} \qquad (i, j = 1, \ldots, n).$$

1.3. Remark. It is obvious, that the identity matrix is diagonal matrix.

1.1.2. Operations with Matrices

There are several operations which can be made with matrices. The simplest ones are the addition and the multplication by scalar. these are performed „entries".

1.4. Definition Let $A, B \in \mathbb{K}^{m\times n}$. The matrix

$$A + B \in \mathbb{K}^{m\times n}, \qquad (A + B)_{ij} := (A)_{ij} + B_{ij}$$

is called the sum of the matrices A and B.

1.5. Definition Let $A \in \mathbb{K}^{m \times n}$ and $\lambda \in \mathbb{K}$. The matrix

$$\lambda A \in \mathbb{K}^{m \times n}, \qquad (\lambda A)_{ij} := \lambda \cdot (A)_{ij}$$

is called the λ-multiple of the matrix A.

1.6. Theorem *The main properties of the above defined matrix operations are as follows:*

I. *1.* $\forall\, A, B \in \mathbb{K}^{m \times n}: \quad A + B \in \mathbb{K}^{m \times n}$

2. $\forall\, A, B \in \mathbb{K}^{m \times n}: \quad A + B = B + A$

3. $\forall\, A, B, C \in \mathbb{K}^{m \times n}: \quad (A + B) + C = A + (B + C)$

4. $\exists\, 0 \in \mathbb{K}^{m \times n} \quad \forall\, A \in \mathbb{K}^{m \times n}: \quad A + 0 = A$
(namely, let 0 *be the zero matrix)*

5. $\forall\, A \in \mathbb{K}^{m \times n} \quad \exists\, (-A) \in \mathbb{K}^{m \times n}: \quad A + (-A) = 0$
(namely, let $(-A)_{ij} := -(A)_{ij}$*)*

II. *1.* $\forall\, \lambda \in \mathbb{K} \quad \forall\, A \in \mathbb{K}^{m \times n}: \quad \lambda A \in \mathbb{K}^{m \times n}$

2. $\forall\, A \in \mathbb{K}^{m \times n} \quad \forall\, \lambda, \mu \in \mathbb{K}: \quad \lambda(\mu A) = (\lambda \mu) A$

3. $\forall\, A \in \mathbb{K}^{m \times n} \quad \forall\, \lambda, \mu \in \mathbb{K}: \quad (\lambda + \mu) A = \lambda A + \mu A$

4. $\forall\, A, B \in \mathbb{K}^{m \times n} \quad \forall\, \lambda \in \mathbb{K}: \quad \lambda(A + B) = \lambda A + \lambda B$

5. $\forall\, A \in \mathbb{K}^{m \times n}: \quad 1A = A$

1.7. Remark. The 10 properties listed above are called vector space axioms (see definition 3.1). Thus the vector space axioms hold in $\mathbb{K}^{m \times n}$.

The following operation, the product of matrices is a bit more complicated.

1.8. Definition Let $A \in \mathbb{K}^{m \times n}$, $B \in \mathbb{K}^{n \times p}$. The matrix

$$AB \in \mathbb{K}^{m \times p}, \qquad (AB)_{ij} := a_{i1} b_{1j} + a_{i2} b_{2j} + \ldots + a_{in} b_{nj} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

is called the product of the matrices A and B (in this order).

The following properties of the matrix product can be proved by simple calculations:

1.9. Theorem *1. associative:*

$$(AB)C = A(BC) \qquad (A \in \mathbb{K}^{m\times n},\ B \in \mathbb{K}^{n\times p},\ C \in \mathbb{K}^{p\times q});$$

2. distributive:

$$A(B+C) = AB + AC \qquad (A \in \mathbb{K}^{m\times n},\ B,\, C \in \mathbb{K}^{n\times p});$$

$$(A+B)C = AC + BC \qquad (A,\, B \in \mathbb{K}^{m\times n},\ C \in \mathbb{K}^{n\times p});$$

3. multiplication by the identity matrix: let I be the unit matrix of right size. Then:

$$AI = A \quad (A \in \mathbb{K}^{m\times n}), \qquad IA = A \quad (A \in \mathbb{K}^{m\times n}).$$

4. multiplication of a product by a scalar:

$$(\lambda A)B = \lambda(AB) = A(\lambda B) \qquad (A \in \mathbb{K}^{m\times n},\ B \in \mathbb{K}^{n\times p},\ \lambda \in \mathbb{K}).$$

About the commutativity of the matrix product: Using the above notations BA is defined if and only if $p = m$. That is the sides of the equation $AB = BA$ are defined if and only if $A \in \mathbb{K}^{m\times n}$ and $B \in \mathbb{K}^{n\times m}$. The necessary condition for the equality is that the matrices on the both sides have the same sizes, that is $m = n$. Even in the case $m = n$ the equality is not true in every cases, as it turns out from the following example:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix}.$$

We can exponentiate the square matrices. Let $A \in \mathbb{K}^{n\times n}$. Then

$$A^0 := I, \quad A^1 := A, \quad A^2 := A \cdot A, \quad A^3 := A^2 \cdot A, \quad \dots .$$

Moreover, we can substitute a square matrix into a polynomial:

1.10. Definition Let $f(x) := c_k x^k + c_{k-1}x^{k-1} + \dots c_1 x + c_0$ be a polynomial, whose coefficients are in $\mathbb{K}$. Then for any $A \in \mathbb{K}^{n\times n}$

$$f(A) := c_k A^k + c_{k-1}A^{k-1} + \dots c_1 A + c_0 I$$

The transpose and the adjoint (Hermitian adjoint) of a matrix are important operations too.

1.11. Definition Let $A \in \mathbb{K}^{m\times n}$. The matrix

$$A^T \in \mathbb{K}^{n\times m}, \qquad (A^T)_{ij} := (A)_{ji}$$

is called the transpose of A, the matrix

$$A^* \in \mathbb{K}^{n\times m}, \qquad (A^*)_{ij} := \overline{(A)_{ji}}$$

is called the (Hermitian) adjoint of A.

The overline denotes the complex conjugate. It worths it to agree, that we define the conjugate for real numbers: the conjugate of a real number is itself. Thus it is obvious, that in the case $\mathbb{K} = \mathbb{R}$ the transpose and the adjoint are the same.

The properties of the above defined operations are as follows:

1.12. Theorem *1.*

$$(A+B)^T = A^T + B^T, \quad (A+B)^* = A^* + B^* \qquad (A,\, B \in \mathbb{K}^{m\times n})$$

2.

$$(\lambda A)^T = \lambda \cdot A^T, \quad (\lambda A)^* = \overline{\lambda} \cdot A^* \qquad (A \in \mathbb{K}^{m\times n},\, \lambda \in \mathbb{K})$$

3.

$$(AB)^T = B^T A^T, \quad (AB)^* = B^* A^* \qquad (A \in \mathbb{K}^{m\times n},\, B \in \mathbb{K}^{n\times p})$$

4.

$$(A^T)^T = A, \quad (A^*)^* = A \qquad (A \in \mathbb{K}^{m\times n}).$$

Sometimes we subdivide the matrix into smaller matrices by inserting imaginary horizontal or vertical straight lines between its selected rows and/or columns. These smaller matrices are called „submatrices” or „blocks”. The so decomposed matrices can be regarded as „matrices” whose elements are also matrices.

The algebraic operations can be made similarly to the learned methods but you must be careful to keep the following requirements:

1. If you regard the blocks as matrix elements the operations must be defined between the resulting „matrices”.

2. The operations must be defined between the blocks itselves.

In this case the result of the operation will be a partitioned matrix, that coincides with the block decomposition of the result of operation with the original (numerical) matrices.

Finally, we speak about an important matrix operation: the inverses of matrices. It corresponds to the reciprocal of real or complex numbers.

1.13. Definition Let $A,\ C \in \mathbb{K}^{n\times n}$. The matrix C is called the inverse of the matrix A, if

$$AC = CA = I$$

(Here I denotes the $n \times n$ identity matrix.) The inverse of A is denoted by A^{-1}.

1.14. Definition Let $A \in \mathbb{K}^{n\times n}$.

(a) The matrix A is called regular (invertible) if it has inverse, that is if $\exists\, A^{-1}$.

(b) The matrix A is called singular (non-invertible), if it has no inverse, that is if $\nexists A^{-1}$.

We can easily prove the uniqueness of the inverse:

1.15. Theorem *Let $A \in \mathbb{K}^{n\times n}$ be a regular matrix, and suppose that both $C \in \mathbb{K}^{n\times n}$ and $D \in \mathbb{K}^{n\times n}$ are the inverses of A, that is*

$$AC = CA = I \qquad \text{and} \qquad AD = DA = I\,.$$

Then $C = D$.

Proof.

$$D = DI = D(AC) = (DA)C = IC = C\,.$$

□

Thus a square matrix either has no inverse (singular case) or it has only one inverse (regular case).

We will deal later with the conditions of existence of the inverse and with the methods of its computation. Here we show an example.

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

because

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus we have proved that the matrix $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$ is regular.

1.1.3. Control Questions to the Theory

1. Define the concept of matrix
2. Define the addition of matrices and list the most important properties of this operation
3. Define the scalar multiplication of a matrix and list the most important properties of this operation
4. Define the product of matrices and list the most important properties of this operation
5. Define the Transpose and the Hermitian adjoint of a matrix and list the most important properties of these operations

1.2. Exercises

1.2.1. Exercises for Class Work

1. What are the sizes of the following matrices? Which of them is zero matrix, row matrix, column matrix, lower triangular matrix, upper triangular matrix, diagonal matrix, identity matrix?

$$A = \begin{bmatrix} 1 & -1 & 2 & 4 & 3 \end{bmatrix} \ ; \quad B = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \ ; \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ ;$$

$$D = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \end{bmatrix} \;; \quad E = \begin{bmatrix} 7 \\ 1 \end{bmatrix} \;; \quad F = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

2. Consider the following matrices:

$$A = \begin{bmatrix} -2 & 1 & 3 \\ 0 & 2 & 5 \end{bmatrix} \;; \quad B = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 3 & -1 \end{bmatrix} \;; \quad C = \begin{bmatrix} 2 & 4 \\ 5 & 4 \end{bmatrix} .$$

Determine (if the result exists):

$A+B\,;\quad A-B\,;\quad 2A-3B\,;\quad A+C\,;\quad A\cdot B\,;\quad A^\top\,;\quad A^\top\cdot C\,;\quad C^2.$

3. Let $A = \begin{bmatrix} 1 & 2 \\ -1 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$, and let f be the following polynomial:

$$f(x) := 2x^3 - x^2 - 5x + 3 \qquad (x \in \mathbb{R})\,.$$

Compute the matrix $f(A)$.

4. Decide whether C is the inverse of A or not, if

$$a) \qquad A = \begin{bmatrix} 3 & -8 \\ 4 & 6 \end{bmatrix} \;; \qquad C = \begin{bmatrix} 3 & 8 \\ 1 & 3 \end{bmatrix}$$

$$b) \qquad A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 5 & -3 \\ -3 & 2 & -4 \end{bmatrix} \;; \qquad C = \begin{bmatrix} 14 & -8 & -1 \\ -17 & 10 & 1 \\ -19 & 11 & 1 \end{bmatrix}$$

1.2.2. Additional Tasks

1. What are the sizes of the following matrices? Which of them is zero matrix, row matrix, column matrix, lower triangular matrix, upper triangular matrix, diagonal matrix, identity matrix?

$$A = \begin{bmatrix} 7 & 2 \\ -1 & 0 \end{bmatrix} ; B = \begin{bmatrix} 2 & 0 & 3 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} ; C = \begin{bmatrix} 0 \\ 0 \end{bmatrix} ;$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} ; \quad E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} ; F = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} .$$

2. Consider the following matrices:

$$A = \begin{bmatrix} 1 & 1 & 5 \\ -3 & 0 & 1 \\ 0 & 1 & 2 \\ 2 & -4 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -4 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 2 & 4 & 0 \\ -1 & 1 & 1 \\ 3 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Determine

$$A + 2B - C, \quad A^T B, \quad (AB^T)C$$

3. Let $A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3}$, and let f be the following polynomial:

$$f(x) := 4x^3 - 5x^2 + 7x + 2 \qquad (x \in \mathbb{R}) .$$

Compute the matrix $f(A)$.

4. Decide whether C is the inverse of A or not, if

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} ; \qquad C = \frac{1}{2} \begin{bmatrix} 1 & 1 & -3 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

5. Prove the statements of theorems 1.6, 1.9, 1.12.

2. Determinants

In this chapter we'll get to know the determinants as numbers ordered to square matrices. In the light of determinants we will return to the discussion of inverse matrices.

2.1. Theory

As we have agreed $\mathbb{K}$ denotes one of the number sets $\mathbb{R}$ or $\mathbb{C}$, that is $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

2.1.1. The Concept of Determinant

To the definition of determinant we need minor matrices via deleting one row and one column of a square matrix:

2.1. Definition Let $n \geq 2$ and $A \in \mathbb{K}^{n\times n}$ and (i, j) be a pair of row-column indices ($i, j \in \{1, \ldots, n\}$). Delete the i-th row and the j-th column from A. The remainder $(n-1) \times (n-1)$-size matrix is called the minor matrix of A related to the index pair (i, j). This minor matrix is denoted by A_{ij}.

After these preliminaries we define recursively the function $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$ as follows:

2.2. Definition

1. If $A = [a_{11}] \in \mathbb{K}^{1\times 1}$, then $\det(A) := a_{11}$.
2. If $A \in \mathbb{K}^{n\times n}$, then:

$$\det(A) := \sum_{j=1}^{n} a_{1j} \cdot (-1)^{1+j} \cdot \det(A_{1j}) = \sum_{j=1}^{n} a_{1j} \cdot a'_{1j},$$

   where $a'_{ij} := (-1)^{i+j} \cdot \det(A_{ij})$, and it is called: signed subdeterminant or cofactor.

We say that we have defined the determinant by expansion along the first row.

2.3. Examples

1. The determinant of a 2×2 matrix can be computed as follows:

$$\det(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = a \cdot (-1)^{1+1} \cdot \det([d]) + b \cdot (-1)^{1+2} \cdot \det([c]) = ad - bc,$$

   that is we obtain the determinant of a 2×2 matrix if we subtract from the product of its main diagonal entries the product of its side diagonal entries.

2. It follows immediately from the definition, that the determinant of a lower diagonal matrix (especially a diagonal matrix) equals the product of its diagonal elements. Consequently, the determinant of the identity matrix equals 1.

In addition to $\det(A)$ we will use the notation

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

also for the determinant. We speak in this sense about the rows, the columns, the entries etc. of the determinant.

Thus, using the above notation:

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

Hereinafter we list - without proof - some important and useful properties of the determinant.

1. The determinant can be expanded by its any row or its any column, that is for any r, $s \in \{1, \dots, n\}$ holds:

$$\det(A) = \sum_{j=1}^{n} a_{rj} \cdot a'_{rj} = \sum_{i=1}^{n} a_{is} \cdot a'_{is}.$$

2. The consequence of the previous statement is that $\det(A) = \det(A^T)$. It follows from here, that the determinant of an upper triangular matrix equals the product of its main diagonal entries.

3. If a determinant has only 0 entries in a row (or in a column), then its value equals 0.

4. If we swap two rows (or two columns) of a determinant, then its value will be the opposite of the original one.

5. If a determinant has two equal rows (or two equal columns), then its value equals 0.

6. If we multiply every entry of a row (or of a column) of the determinant by a number λ, then its value will be the λ-multiple of the original one.

7. $\forall\, A \in \mathbb{K}^{n\times n}$ and $\forall\, \lambda \in \mathbb{K}$ holds $\det(\lambda \cdot A) = \lambda^n \cdot \det(A)$.

8. If two rows (or two columns) of a determinant are proportional, then its value equals 0.

9. The determinant is additive in its any row (and by its any column). This means – in the case of additivity of its r-th row – that:

   $$\text{If} \qquad (A)_{ij} := \begin{cases} \alpha_j & \text{if} \quad i = r \\ a_{ij} & \text{if} \quad i \neq r, \end{cases} \qquad \text{and} \qquad (B)_{ij} := \begin{cases} \beta_j & \text{if} \quad i = r \\ a_{ij} & \text{if} \quad i \neq r, \end{cases}$$

   $$\text{and} \qquad (C)_{ij} := \begin{cases} \alpha_j + \beta_j & \text{if} \quad i = r \\ a_{ij} & \text{if} \quad i \neq r\,, \end{cases}$$

   then $\det(C) = \det(A) + \det(B)$.

10. If we add to a row of a determinant a scalar multiple of another row (or to a column a scalar multiple of another column), then the value of the determinant remains unchanged.

11. The determinant of the product of two matrices equals the product of their determinants:

    $$\det(A \cdot B) = \det(A) \cdot \det(B) \qquad (A,\, B \in \mathbb{K}^{n\times n})\,.$$

The determinant is in close connection with the calculation of length, of area and of volume. We write more about this topic in the Appendix, see there „The geometric meaning of the determinant”.

2.1.2. Inverses of Matrices

In the definition 1.13 the concept of the inverse matrix was defined, further it was proved its uniqueness.

In this section – using the determinant – we investigate in more detail the conditions of the existence of inverse matrix.

2.4. Theorem *[existence of the right-hand inverse]*

Let $A \in \mathbb{K}^{n \times n}$. Then there exists a matrix $C \in \mathbb{K}^{n \times n}$ for which holds $AC = I$ if and only if $\det(A) \neq 0$. Such a matrix C is called a right-hand inverse of A.

Proof. First suppose the existence of C with this property. Then $AC = I$, consequently:

$$1 = \det(I) = \det(A \cdot C) = \det(A) \cdot \det(C).$$

From here immediately follows $\det(A) \neq 0$.

Conversely, suppose $\det(A) \neq 0$. Define the following matrix:

$$C := \frac{1}{\det(A)} \cdot \tilde{A}\ , \qquad \text{ahol} \qquad (\tilde{A})_{ij} := a'_{ji}\,.$$

Now we will show that $AC = I$ holds for this C. Really:

$$\begin{aligned}(AC)_{ij} &= \left(A \cdot \frac{1}{\det(A)} \cdot \tilde{A}\right)_{ij} = \frac{1}{\det(A)} \cdot (A \cdot \tilde{A})_{ij} = \\ &= \frac{1}{\det(A)} \cdot \sum_{k=1}^{n} (A)_{ik} \cdot (\tilde{A})_{kj} = \frac{1}{\det(A)} \cdot \sum_{k=1}^{n} a_{ik} \cdot a'_{jk}.\end{aligned}$$

The last sum equals 1 if $i = j$, because – using the expansion of the determinant along the i-th row:

$$\frac{1}{\det(A)} \cdot \sum_{k=1}^{n} a_{ik} \cdot a'_{ik} = \frac{1}{\det(A)} \cdot \det(A) = 1.$$

Now suppose that $i \neq j$. In this case the above mentioned sum is the expansion of a determinant along its j-th row which can be obtained from $\det(A)$ by exchanging its j-th row to its i-th row. But this determinant has two equal rows (the i-th and the j-th), so its value equals 0. This means that

$$\forall\, i \neq j : \qquad (AC)_{ij} = 0\,.$$

We have proved that $(AC)_{ij} = (I)_{ij}$, consequently the product AC really equals the identity matrix. $\square$

Using the theorem about the right-hand inverse, it can be proved the existence of the inverse matrix.

2.5. Theorem *Let* $A \in \mathbb{K}^{n\times n}$. *Then*

$$\exists\, A^{-1} \qquad \Longleftrightarrow \qquad \det(A) \neq 0\,,$$

that is the matrix A *is regular if and only if* $\det(A) \neq 0$. *Consequently, the matrix* A *is singular if and only if* $\det(A) = 0$.

Proof. First suppose that A is regular, that is $\exists\, A^{-1}$. Then $A \cdot A^{-1} = I$, thus

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

This implies $\det(A) \neq 0$. Furthermore we have

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

Conversely, suppose $\det(A) \neq 0$. Then – usig the second half of the previous theorem – there exists a matrix $C \in \mathbb{K}^{n\times n}$, for which holds $AC = I$.

We will show, that this matrix C is the inverse of A. Since $AC = I$ is proved (see the previous theorem), it is enough to prove that $CA = I$.

This is proved as follows. Since $\det(A^T) = \det(A) \neq 0$, we can apply the second half of the previous theorem for the matrix A^T. Thus we have

$$\exists\, D \in \mathbb{K}^{n\times n}: \qquad A^T D = I\,.$$

Let us transpose both sides of the equality:

$$(A^T D)^T = I^T\,,$$

form where $D^T A = I$ follows. With the help of this fact the equality $CA = I$ will be proved easily:

$$CA = ICA = D^T ACA = D^T (AC) A = D^T I A = D^T A = I\,.$$

$\square$

2.6. Remarks.

1. We emphasize once more time, that the regularity of a matrix $A \in \mathbb{K}^{n\times n}$ is equivalent with the fact that its determinant is nonzero. In the regular case we have deduced an explicit formula for the inverse:

$$A^{-1} = \frac{1}{\det(A)} \cdot \tilde{A}\,, \qquad \text{where} \qquad (\tilde{A})_{ij} := a'_{ji}\,.$$

2. Let us apply the previous result for the 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2\times 2}\,.$$

Then A is regular if and only if $ad - bc \neq 0$. In this case the inverse matrix is

$$A^{-1} = \frac{1}{ad - bc} \cdot \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\,.$$

Expressing it in words:

We obtain the inverse of a 2×2 regular matrix if we interchange the entries in its main diagonal, then we change the signs of the entries in the side diagonal, finally, we multiply the obtained matrix with the reciprocal of the determinant of the original matrix.

3. It follows from our considerations, that to prove that the inverse of $A \in \mathbb{K}^{n\times n}$ is the matrix $C \in \mathbb{K}^{n\times n}$ it is enough to prove only one of the equalities $AC = I$ or $CA = I$. The other one holds automatically.

2.1.3. Control Questions to the Theory

1. Define the concept of the minor matrix assigned to the index pair $(i,\, j)$ of an $m \times n$ matrix, and give a numerical example for this

2. Define the concept of determinant

3. Define the concept of cofactor assigned to $(i,\, j)$

4. How can we compute the 2×2 determinants?

5. How can we compute the determinant of a triangular matrix?

6. State the following properties of the determinant:

   - expansion along any row/column
   - transpose-property
   - 0 row/column
   - row/column interchange property
   - two rows/two columns are equal
   - row/column homogeneous
   - the determinant of λA
   - proportional rows/columns
   - row/column additive
   - the determinant of AB

7. Define the right-hand inverse, the left-hand inverse and the inverse of a square matrix

8. Define the concept of singular matrix and regular matrix

9. State and prove the theorem about the existence and formula of the right-hand inverse

10. State and prove the theorem about the necessary and sufficient condition of the existence of the left-hand inverse (reducing the problem back to the right-hand inverse)

11. State and prove the theorem about the connection between the right-hand and the left-hand inverses

12. State and prove the statement about the existence and formula of the inverse

13. State and prove the formula of the inverse of a 2×2 matrix

2.2. Exercises

2.2.1. Exercises for Class Work

1. Let

$$a) \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 2} \qquad\qquad b) \quad A = \begin{bmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

   (1) Compute $\det A$ in different ways.

   (2) Determine whether the matrix A is regular or singular. In the regular case determine the inverse of A using cofactors.

   (3) Check $A \cdot A^{-1} = I$.

2. The following matrices are regular or singular? In regular case determine the inverse matrix.

$$A = \begin{bmatrix} -2 & 5 \\ -3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} -2 & 3 \\ -4 & 6 \end{bmatrix}$$

3. Illustrate the properties of determinants with concrete matrices.

2.2.2. Additional Tasks

1. Compute the following determinants:

$$a) \quad \begin{vmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{vmatrix} \qquad b) \quad \begin{vmatrix} 1 & 0 & 0 & -1 \\ 3 & 1 & 2 & 2 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & 0 & 1 \end{vmatrix}$$

2. Determine the inverse matrices:

$$a)\quad \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \qquad b)\quad \begin{bmatrix} 3 & 2 & -1 \\ 1 & 6 & 3 \\ 2 & -4 & 0 \end{bmatrix}$$

furthermore check the result by the definition of the inverse matrix.

3. Illustrate the properties of determinants with concrete matrices.

4. Let $A \in \mathbb{K}^{n\times n}$ be a diagonal matrix (that is $a_{ij} = 0$ if $i \neq j$). Prove that it is invertible if and only if none of the diagonal elements equals 0. Prove that in this case A^{-1} is a diagonal matrix with diagonal elements

$$\frac{1}{a_{11}},\ \frac{1}{a_{22}},\ \dots\ \frac{1}{a_{nn}}\,.$$

3. Vectors, Vector Spaces

In this chapter the concept of vector will be generalized.

3.1. Theory

3.1.1. The Concept of Vector Space

In the secondary school we got acquainted with the concept of vector, operations with vectors and their properties. We have found, that the vector addition has the following properties:

1. If $\underline{a}$ and $\underline{b}$ are vectors then $\underline{a} + \underline{b}$ is also a vector
2. $\underline{a} + \underline{b} = \underline{b} + \underline{a}$ (commutative law)
3. $(\underline{a} + \underline{b}) + \underline{c} = \underline{a} + (\underline{b} + \underline{c})$ (associative law)
4. $\underline{a} + \underline{0} = \underline{a}$ (the characterization of the zero vector)
5. $\underline{a} + (-\underline{a}) = \underline{0}$ (the characterization of the opposite vector)

The most important properties of the multiplication of vectors by a scalar are follows:

1. If λ is a real number, and $\underline{a}$ is a vector, then $\lambda\underline{a}$ is a vector
2. $\lambda(\mu\underline{a}) = (\lambda\mu)\underline{a}$ (multiply a product)
3. $(\lambda + \mu)\underline{a} = \lambda\underline{a} + \mu\underline{a}$ (distributive law)
4. $\lambda(\underline{a} + \underline{b}) = \lambda\underline{a} + \lambda\underline{b}$ (distributive law)
5. $1\underline{a} = \underline{a}$

We have seen the same properties in connection with matrices in Theorem 1.6.

Now we generalize the concept of vector in the following way: we take a nonempty set (whose elements will be called vectors), and we take a number set (which will be called scalar range and whose elements will be called scalars). Furthermore we take two operations (addition of vectors and multiplication of vectors by scalars), which have the above written 10 features. The resulting „structure" will be called vector space. The 10 features are called vector space axioms.

We will use the number fields $\mathbb{R}$ and $\mathbb{C}$ as scalar range, that is, the scalar range will be the number field $\mathbb{K}$. Thus we can investigate the real and the complex vector spaces „in parallel".

After this introduction let us see the definition of a vector space:

3.1. Definition Let $V \neq \emptyset$. We say, that V is a vector space over $\mathbb{K}$ if there exist the operations $x+y$ (addition) and $\lambda x = \lambda \cdot x$ (multiplication by scalar) so that the following axioms hold

I.
1. $\forall\,(x,y) \in V \times V: \quad x+y \in V$
2. $\forall\,x,y \in V: \quad x+y = y+x$
3. $\forall\,x,y,z \in V: \quad (x+y)+z = x+(y+z)$
4. $\exists\,0 \in V\ \forall\,x \in V: \quad x+0 = x$
   It can be proved that 0 is unique. Its name is: zero vector.
5. $\forall\,x \in V\ \exists\,(-x) \in V: \quad x+(-x) = 0$
   It can be proved that $(-x)$ is unique. Its name is: the opposite of x.

II.
1. $\forall\,(\lambda,\,x) \in \mathbb{K} \times V: \quad \lambda x \in V$
2. $\forall\,x \in V\ \forall\,\lambda,\mu \in \mathbb{K}: \quad \lambda(\mu x) = (\lambda\mu)x = \mu(\lambda x)$
3. $\forall\,x \in V\ \forall\,\lambda,\mu \in \mathbb{K}: \quad (\lambda+\mu)x = \lambda x + \mu x$
4. $\forall\,x,y \in V\ \forall\,\lambda \in \mathbb{K}: \quad \lambda(x+y) = \lambda x + \lambda y$
5. $\forall\,x \in V: \quad 1x = x$

The elements of V are called vectors, the elements of $\mathbb{K}$ are called scalars. $\mathbb{K}$ is called the scalar region (scalar range) of V.

Applying several times the associative law of addition we can define the sums of several terms:

$$x_1 + x_2 + \cdots + x_k = \sum_{i=1}^{k} x_i \qquad (x_i \in V)\,.$$

3.2. Remarks.

1. The vectors are often denoted by underlined lowercases , but it is not required.
2. It is in evidence, that the axioms are derived from the properties of geometric vectors studied in secondary school. Thus we have our first example for vector space:

   The plane vectors starting from a fixed point of the plane form a vector space over $\mathbb{R}$, with respect to the common vector addition and multiplication by scalar.

   The fixed starting point is necessary, so there won't be any problem with the equality of vectors.
3. Sometimes we use the operations „multiplication by scalar from the right", „division by a nonzero number", „subtraction" as follows:

   $$x \cdot \lambda := \lambda \cdot x, \qquad \frac{x}{\lambda} := \frac{1}{\lambda} \cdot x \qquad x - y := x + (-y).$$

   The properties of these operations follow from the axioms.
4. If the scalar region and the two operations are given by some default setting, then we simply say: „V is a vector space".

3.3. Examples

1. The vectors in the plane, with the usual vector operations form a vector space over $\mathbb{R}$. This is the vector space of plane vectors. Since the plane vectors can be identified with the points of the plane, instead of the vector space of the plane vectors we can speak about the vector space of the points in the plane.
2. The vectors in the space, with the usual vector operations form a vector space over $\mathbb{R}$. This is the vector space of space vectors. Since the space vectors can be identified with the points of the space, instead of the vector space of the space vectors we can speak about the vector space of the points in the space.
3. From the algebraic properties of the number field $\mathbb{K}$ immediately follows that $\mathbb{R}$ is vector space over $\mathbb{R}$, $\mathbb{C}$ is vector space over $\mathbb{C}$, that is $\mathbb{K}$ is vector space over $\mathbb{K}$.

   As a matter of fact, note that $\mathbb{C}$ is vector space over $\mathbb{R}$ too.
4. For fixed $m, n \in \mathbb{N}^+$ the set of $m \times n$ matrices, that is $\mathbb{K}^{m \times n}$ forms a vector space over $\mathbb{K}$. This follows immediately from Theorem 1.6.

5. The one-element-set is vector space over $\mathbb{K}$. Since the single element of this set must be the zero vector of the space, we will denote this vector space by $\{0\}$. The operations in this space are:

$$0 + 0 := 0, \qquad \lambda \cdot 0 := 0 \quad (\lambda \in \mathbb{K}).$$

The name of this vector space is: zero vector space.

If we don't say anything else, the symbol V will denote a vector space over $\mathbb{K}$.

In the following theorem some basic properties of vector spaces will be listed. They can be prove using the axioms.

3.4. Theorem *Let $x \in V$, $\lambda \in \mathbb{K}$. Then*

1. $0 \cdot x = 0$ *(the left hand side 0 denotes the number zero in $\mathbb{K}$, the right hand side 0 denotes the zero vector in V)*
2. $\lambda \cdot 0 = 0$ *(here both 0-s denote the zero vector in V)*
3. $(-1) \cdot x = -x$.
4. $\lambda \cdot x = 0 \quad \Longleftrightarrow \quad \lambda = 0 \quad \text{or} \quad x = 0.$

3.1.2. The Vector Space $\mathbb{K}^n$

This section will be about a very important vector space: about $\mathbb{K}^n$.

For a fixed $n \in \mathbb{N}^+$ the function

$$x : \{1, \ldots, n\} \to \mathbb{K}$$

is called an n-term sequence (in other words: ordered n-tuple) created from the elements of $\mathbb{K}$.

The number $x(i) \in \mathbb{K}$ is called the i-th component of the vector x, and it is denoted by x_i $(i = 1, \ldots, n)$. The n-tuple itself is denoted as follows:

$$x = (x_1, x_2, \ldots x_n).$$

E. g. $(1, -3, 5, 8)$ is an ordered 4-tuple.

Let us denote the set of all n-tuples constructed from the elements of $\mathbb{K}$ by $\mathbb{K}^n$:

$$\mathbb{K}^n := \{x = (x_1,\, x_2,\, \dots,\, x_n) \mid x_i \in \mathbb{K}\}$$

Following the previous example e.g. $(1,\, -3,\, 5,\, 8) \in \mathbb{R}^4$.

Let us define the operations „componentwise":

$$(x_1,\, x_2,\, \dots,\, x_n) + (y_1,\, y_2,\, \dots,\, y_n) := (x_1 + y_1,\, x_2 + y_2,\, \dots,\, x_n + y_n)\,;$$

$$\lambda \cdot (x_1,\, x_2,\, \dots,\, x_n) := (\lambda x_1,\, \lambda x_2,\, \dots,\, \lambda x_n) \qquad (\lambda \in \mathbb{K})\,.$$

In another way:

$$(x+y)_i := x_i + y_i; \quad \text{and} \quad (\lambda \cdot x)_i := \lambda \cdot x_i \quad (i = 1, \dots, n;\ x, y \in \mathbb{K}^n).$$

3.5. Theorem *$\mathbb{K}^n$ is a vector space over the number field $\mathbb{K}$.*
The zero vector of this vector space is the n-tuple $(0,\, 0,\, \dots,\, 0)$, the opposite of the vector $(x_1,\, x_2,\, \dots,\, x_n)$ is $(-x_1,\, -x_2,\, \dots,\, -x_n)$.

Proof. One can easily check the validity of the 10 axioms. □

The conventional way for writing the elements of $\mathbb{K}^n$ is:

$$x = (x_1,\, x_2,\, \dots,\, x_n) \in \mathbb{K}^n\,,$$

but sometimes it is useful to write them in column-mode:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n\,.$$

The column-mode writing is usefol e.g. in the case when we perform algebraic operations with elements in $\mathbb{K}^n$. In this case the components having the same indices stand at the same height. Thus they are better manageable. For example in $\mathbb{R}^4$ we have:

$$2 \cdot \begin{pmatrix} -1 \\ 2 \\ 5 \\ 1 \end{pmatrix} + 3 \cdot \begin{pmatrix} 1 \\ 2 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \cdot (-1) + 3 \cdot 1 \\ 2 \cdot 2 + 3 \cdot 2 \\ 2 \cdot 5 + 3 \cdot (-3) \\ 2 \cdot 1 + 3 \cdot 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \\ 1 \\ 8 \end{pmatrix}.$$

3.6. Remark. If we speak about the vector space $\mathbb{K}^n$ then the default scalar region is $\mathbb{K}$ and the default operations are the componentwise operations.

3.7. Remark. It is known that the points and the vectors in the plane can be characterized by number pairs. Similarly, the points in the space and the vectors in the space can be characterized by number triples. Thus $\mathbb{R}^2$ can be considered the vector space of the points of the plane, or of the vectors of the plane. Similarly, $\mathbb{R}^3$ can be considered the vector space of the points of the space, or of the vectors of the space.

Using similar justification, $\mathbb{R} = \mathbb{R}^1$ can be considered the vector space of the points of the line (number line).

3.8. Remark. The vector space $\mathbb{K}^n$ can be identified with the space of row matrices $\mathbb{K}^{1\times n}$, moreover, it can be identified with the space of column matrices $\mathbb{K}^{n\times 1}$ too. This is the reason, that the row matrices are sometimes called „row vectors", the column matrices are called sometimes „column vectors".

As a special case of matrix product, we can define the matrix-vector product operation as follows:

3.9. Definition Let $A \in \mathbb{K}^{m\times n}$, $x \in \mathbb{K}^n$. The vector

$$Ax \in \mathbb{K}^m, \qquad (Ax)_i := a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j \qquad (i = 1, \ldots, m)$$

is called the product of matrix A and vector x (in this order).

3.10. Remark. One can see, that we can compute vector Ax in the following way: multiply matrix A by the column matrix associated to x, and take of as a result the vector associated to the column matrix we computed. This will be Ax. Thus the matrix-vector product is essentially a product of a matrix and a column matrix. This identification implies in a natural way the properties of the matrix-vector product.

3.1.3. Subspaces

3.11. Definition Let W be a nonempty subset of the vector space V. We say that W is a subspace of V (or: W is a subspace in V), if W is a vector space regarding the operations of V.

The following theorem gives a useful method to decide whether a subset in V is a subspace or not.

3.12. Theorem *Let $W \subset V$, $W \neq \emptyset$.*

W is a subspace of V if and only if the following two assumptions hold:

1. $\forall\, x, y \in W: \quad x + y \in W,$
2. $\forall\, x \in W \quad \forall\, \lambda \in \mathbb{K}: \quad \lambda x \in W.$

The first assumption expresses that W is closed under addition. Similarly, the second assumption expresses that W is closed under scalar multiplication.

Proof. The two given assumptions are obviously necessary.

To prove that they are sufficient, let us realize that the vector space axioms I.1. and II.1. are exactly the given conditions so they are true. Moreover, axioms I.2., I.3., II.2., II.3., II.4., II.5. are identities, so they are inherited from V to W.

It remains us to prove only two axioms: I.4., I.5.

Proof of I.4.: Let $x \in W$ and 0 be the zero vector in V. Then – because of the second condition: $0 = 0x \in W$, so W really contains a zero vector, and the zero vectors in V and W are the same.

Proof of I.5.: Let $x \in W$ and $-x$ be the the opposite vector of x in V. Then – also because of the second condition: $-x = (-1)x \in W$, so W really contains an opposite of x and the opposite vectors in V and W are the same.

□

3.13. Corollary. It follows immediately from the above proof, that a subspace must contain the zero vector of V. In other words: if a subset does not contain the zero vector of V, then it is not a subspace. Similar considerations are valid for the opposite vector too.

Using the above theorem one can easily prove that the following examples are subspaces.

3.14. Examples

1. The zero vector space $\{0\}$ and V itself both are subspaces in V. They are called the trivial subspaces.
2. All the subspaces of the vector space of plane vectors ($\mathbb{R}^2$) are:

- the zero vector space $\{0\}$,
- the straight lines trough the origin,
- $\mathbb{R}^2$ itself.

3. All the subspaces of the vector space of space vectors ($\mathbb{R}^3$) are:

    - the zero vector space $\{0\}$,
    - the straight lines trough the origin,
    - the planes trough the origin,
    - $\mathbb{R}^3$ itself.

3.1.4. Control Questions to the Theory

1. Define the concept of a vector spaces
2. Give 2 examples for vector space
3. State the 4 elementary properties of vector spaces
4. Define the vector space $\mathbb{K}^n$ (its elements, operations in it)
5. Define the matrix-vector product operation
6. Define the subspace of a vector space
7. State and prove the theorem about the necessary and sufficient condition for a set to be a subspace
8. Give 2 examples for subspaces

3.2. Exercises

3.2.1. Exercises for Class Work

1. Let us give the following vectors in $\mathbb{R}^5$:

$$x = (-3, 4, 1, 5, 2) \qquad y = (2, 0, 4, -3, -1) \qquad z = (7, -1, 0, 2, 3)\,,$$

and the matrix

$$A = \begin{bmatrix} 5 & 1 & -4 & -2 & 1 \\ 0 & 2 & 4 & -3 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 5}\,.$$

Compute:

$$x + y\,, \qquad y - z\,, \qquad 4x\,, \qquad x + 3y - 2z\,, \qquad Ax\,.$$

2. Are the following sets subspaces in $\mathbb{R}^2$ or not? Which famous sets of points are they (give their geometric names)?

$$K := \{(x, y) \in \mathbb{R}^2 \mid x^2+y^2 = 1\} \quad ; \qquad N := \{(x, y) \in \mathbb{R}^2 \mid x \geq 0 \;\, y \geq 0\}\,.$$

3. Are the following sets subspaces in $\mathbb{R}^3$ or not? Give their geometric names.

   (a) $S_1 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$,
   (b) $S_2 = \{(x, y, z) \in \mathbb{R}^3 \mid x \geq 0, y \geq 0, z \geq 0\}$,
   (c) $S_3 = \{(x, y, z) \in \mathbb{R}^3 \mid 2x - 3y + z = 0\}$,
   (d) $S_4 = \{(x, y, z) \in \mathbb{R}^3 \mid 2x - 3y + z = 5\}$,
   (e) $S_5 = \{(x - y, 3x, 2x + y) \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\}$.

4. Prove, that the vector space axioms hold in $\mathbb{K}^n$.

3.2.2. Additional Tasks

1. Consider the following vectors

$$x = (1, -2, 3, 4) \in \mathbb{R}^4\,, \qquad y = (-4, 0, 2, 1) \in \mathbb{R}^4\,, \qquad z = (2, -1, 0) \in \mathbb{R}^3\,,$$

   and the matrix

$$A = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 1 & -2 \\ 0 & 2 & 4 \\ 2 & -1 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 3}$$

   Compute:

$$3x + y + 2Az$$

2. Are the following sets subspaces in $\mathbb{R}^2$ or not? Which famous sets of points are they (give their geometric names)?

$$K := \{(x, y) \in \mathbb{R}^2 \mid xy = 1\} \quad ; \qquad N := \{(x, y) \in \mathbb{R}^2 \mid xy \geq 0\}\,.$$

3. Are the following sets subspaces in $\mathbb{R}^3$ or not? Give their geometric names.

   (a) $S_1 = \{(x, y, z) \in \mathbb{R}^3 \mid z = x^2 + y^2\}$,
   (b) $S_2 = \{(x, y, z) \in \mathbb{R}^3 \mid z^2 = x^2 + y^2\}$,
   (c) $S_3 = \{(2x, x + y, y) \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\}$.
   (d) $S_4 = \{(x, y, z) \in \mathbb{R}^3 \mid 4x - y + 3z = 0\}$,
   (e) $S_5 = \{(x, y, z) \in \mathbb{R}^3 \mid 4x - y + 3z + 1 = 0\}$,

4. Prove, that the vector space axioms hold in $\mathbb{K}^n$ (continuation of a class-work-exercise).

4. Geneated Subspaces

In this section we will give subspaces with the help of some vectors of a finite number.

4.1. Theory

From now on, it will be often mentioned the concept of a (finite) vector system. We speak about a (finite) vector system, if we choose a finite number of vectors from a vector space in the way, that some vectors may be chosen several times. This „chosen several times” option is exactly the reason why we make a distinction between a vector system and a set of vectors.

4.1.1. Linear Combination

4.1. Definition Let $k \in \mathbb{N}^+$, $x_1, \ldots, x_k \in V$ be a vector system, $\lambda_1, \ldots \lambda_k \in \mathbb{K}$. The vector

$$\lambda_1 x_1 + \cdots + \lambda_k x_k = \sum_{i=1}^{k} \lambda_i x_i$$

(or the expression itself) is called the linear combination of the vector system $x_1, \ldots, x_k$, with the coefficients $\lambda_1, \ldots \lambda_k$.

The linear combination is called trivial, if all its coefficients are zero, and it is called nontrivial, if it has at least one nonzero coefficient.

Obviously, the result of a trivial linear combination is always the zero vector.

4.2. Remark. Using mathematical induction, it can be proved, that if W is a subspace in V, and $x_1, \ldots, x_k \in W$, $\lambda_1, \ldots \lambda_k \in \mathbb{K}$, then $\sum\limits_{i=1}^{k} \lambda_i x_i \in W$, that is: subspaces are closed under linear combination.

4.1.2. The Concept of Generated Subspace

Let $x_1,\, x_2,\, \ldots,\, x_k \in V$ be a vector system. Consider the following subset of V:

$$W^* := \{\sum_{i=1}^{k} \lambda_i x_i \in V \mid \lambda_1,\, \ldots,\, \lambda_k \in \mathbb{K}\}\,. \tag{4.1}$$

One can see that the elements of W^* are all the possible linear combinations of the vector system $x_1,\, x_2,\, \ldots,\, x_k$.

4.3. Theorem *1. W^* is a subspace in V.*

2. W^* covers the system $x_1,\, x_2,\, \ldots,\, x_k$, which means that

$$x_i \in W^* \quad (i=1,\ldots,k)\,.$$

3. For any subspace $Z \subseteq V$ which covers (in the above sense) the system $x_1,\, x_2,\, \ldots,\, x_k$, holds $W^* \subseteq Z$.

Note before the proof, that the statement of this theorem is shortly saying that W^* is the minimal subspace, which covers $x_1,\, x_2,\, \ldots,\, x_k$.
Proof.

1. Let $a = \sum\limits_{i=1}^{k} \lambda_i x_i \in W^*$ and $b = \sum\limits_{i=1}^{k} \mu_i x_i \in W^*$. Then

$$a+b = \sum_{i=1}^{k} \lambda_i x_i + \sum_{i=1}^{k} \mu_i x_i = \sum_{i=1}^{k} (\lambda_i + \mu_i) x_i \in W^*\,.$$

   Furthermore, for any $\lambda \in \mathbb{K}$ holds

$$\lambda a = \lambda \sum_{i=1}^{k} \lambda_i x_i = \sum_{i=1}^{k} (\lambda\lambda_i) x_i \in W^*\,.$$

   Consequently, W^* is really a subspace in V.

2. For any fixed index $i \in \{1,\, \ldots,\, k\}$ holds

$$x_i = 0x_1 + \ \ldots\ + 0x_{i-1} + 1x_i + 0x_{i+1} + \ \ldots\ + 0x_k \in W^*\,.$$

3. Let Z be a subspace given in the theorem, and let $a = \sum\limits_{i=1}^{k} \lambda_i x_i \in W^*$. Since Z covers the vector system, then

$$x_i \in Z \qquad (i=1,\, \ldots,\, k)\,.$$

   But Z is a subspace, therefore it is closed under linear combination. It implies that $a \in Z$. So really $W^* \subseteq Z$.

□

4.4. Definition The subspace W^* defined in (4.1) is called the subspace generated (or spanned) by the vector system $x_1, x_2, \ldots, x_k$. It is denoted by $\mathrm{Span}\,(x_1, x_2, \ldots, x_k)$.

4.5. Definition Let W be a subspace of V. We say that W has a finite generator system if

$$\exists\, k \in \mathbb{N}^+ \quad \exists\, x_1, x_2, \ldots, x_k \in V : \qquad \mathrm{Span}\,(x_1, x_2, \ldots, x_k) = W\,.$$

In this case the vector system $x_1, x_2, \ldots, x_k$ is called a (finite) generator system of the subspace W.

4.6. Definition In the case when $\mathrm{Span}\,(x_1, x_2, \ldots, x_k) = V$ the vector system $x_1, x_2, \ldots, x_k$ is called simply generator system.

4.7. Remark. The fact, that a vector x lies in the subspace generated by the vectors $x_1, \ldots, x_k$ is equivalent with the fact, that x can be written as a linear combination of vectors $x_1, \ldots, x_k$. We can also say that vector x linearly depends on vectors $x_1, \ldots, x_k$.

4.8. Examples

1. Let v be a fixed vector in the space of plane vectors. Then

   $$\mathrm{Span}\,(v) = \begin{cases} \{0\} & \text{if } v = 0, \\ v \text{ a straight line through the origin with direction vector } v & \text{if } v \neq 0. \end{cases}$$

   One can easily prove that in the space of the plane vectors any two nonparallel vectors are forming a generator system.

2. In the space of space vectors let v_1 and v_2 be two fixed vectors. Then

   $$\mathrm{Span}\,(v_1, v_2) = \begin{cases} \{0\} & \text{if } v_1 = v_2 = 0, \\ \text{the common straight line of } v_1 \text{ and } v_2 & \text{if } v_1 \parallel v_2 \\ \text{the common plane of } v_1 \text{ and } v_2 & \text{if } v_1 \nparallel v_2. \end{cases}$$

   It can be easily proved, that in the space of space vectors any three vectors, which are not contained in the same plane are forming a generator system.

3. The i-th canonical unit vector (or standard unit vector) in $\mathbb{K}^n$ – let us denote it by e_i – is defined as follows:

   Let the i-th component of e_i be 1, and all its other components to be 0 $(i = 1, \ldots, n)$.

Then the vector system $e_1, \ldots, e_n$ is a generator system in $\mathbb{K}^n$, because for any $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$ holds

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \cdot 1 + x_2 \cdot 0 + \cdots + x_n \cdot 0 \\ x_1 \cdot 0 + x_2 \cdot 1 + \cdots + x_n \cdot 0 \\ \vdots \\ x_1 \cdot 0 + x_2 \cdot 0 + \cdots + x_n \cdot 1 \end{pmatrix} =$$
$$= x_1 \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} x_i e_i$$

Thus x really can be written as a linear combination of the standard unit vectors $e_1, \ldots, e_n$.

4.9. Remark. It is clear, that if we enlarge a generator system in V, then it still remains a generator system. But if we leave vectors from a generator system, then the result system may not necessarily stay a generator system. The generator systems are – in this sense – the „large" systems. Later we will study the question of „minimal" generator systems.

4.1.3. Finite Dimensional Vector Space

The concept of a generator system can be extended into an infinite system as well. In this connection, we call the above defined generator system more precisely a finite generator system. An important class of vector spaces are the spaces having a finite generator system.

4.10. Definition The vector space V is called finite-dimensional if it has finite generator system, that is, if

$$\exists\, k \in \mathbb{N}^+ \quad \exists\, x_1,\, x_2,\, \ldots,\, x_k \in V : \qquad \operatorname{Span}(x_1,\, x_2,\, \ldots,\, x_k) = V\,.$$

The fact, that V is a finite dimensional space is denoted by $\dim V < \infty$.

4.11. Definition The vector space V is called infinite-dimensional if it does not have a finite generator system, that is, if

$$\forall\, k \in \mathbb{N}^+ \quad \forall\, x_1,\, x_2,\, \ldots,\, x_k \in V : \qquad \operatorname{Span}(x_1,\, x_2,\, \ldots,\, x_k) \neq V\,.$$

The fact, that V is infinite dimensional is denoted by $\dim V = \infty$.

4.12. Examples

1. The space of the plane vectors is finite dimensional. A finite generator system is i, j.

2. The space of space vectors is finite dimensional. A finite generator system is i, j, k.

3. The space $\mathbb{K}^n$ is finite dimensional. A finite generator system is the system of the n standard unit vectors.

4. You can find an example for an infinite dimensional vector space in the Appendix.

4.1.4. Control Questions to the Theory

1. Define the linear combination

2. State and prove the theorem about a generated subspace by a finite vector system (This is the theorem about W^*)

3. Give 2 examples for generated subspaces in $\mathbb{R}^2$

4. Give 2 examples for generated subspaces in $\mathbb{R}^3$

5. Define the standard unit vectors in $\mathbb{K}^n$. What is the subspace generated by them?

6. Define the notion of a finite dimensional vector space. Why is the vector space $\mathbb{K}^n$ finite dimensional?

7. Define the notion of an infinite dimensional vector space.

4.2. Exercises

4.2.1. Exercises for Class Work

1. Write the subspace

$$W := \text{Span}\left((1,2,-1),(-3,1,1)\right) \subseteq \mathbb{R}^3$$

as a set. Give several elements of this subspace.

Determine whether the vectors $(2,4,0)$ and $(5,-4,-1)$ are contained in this subspace or not.

2. Consider the following vectors in $\mathbb{R}^3$:

$$u = (1,2,-1)\ ; \quad v = (6,4,2)\ ; \quad x = (9,2,7)\ ; \quad y = (4,-1,8).$$

(a) Compute the result of the linear combination $-2u+3v$.

(b) Write up the elements of the subspace $\text{Span}\,(u,v)$.

(c) Determine whether $x \in \text{Span}\,(u,v)$ or not.

(d) Determine whether $y \in \text{Span}\,(u,v) ornot.$

3. Consider the subspaces

(a) $S_5 = \{(x-y, 3x, 2x+y) \in \mathbb{R}^3 \mid x,y \in \mathbb{R}\} \subseteq \mathbb{R}^3\,,$

(b) $S_3 = \{(x,y,z) \in \mathbb{R}^3 \mid 2x-3y+z=0\} \subseteq \mathbb{R}^3$

and the subspaces

(c) $W_1 = \{(x-y+5z, 3x-z, 2x+y-7z, -x) \in \mathbb{R}^4 \mid x,y,z \in \mathbb{R}\} \subseteq \mathbb{R}^4\,,$

(d) $W_2 = \{(x,y,z) \in \mathbb{R}^3 \mid x+3y=0\} \subseteq \mathbb{R}^3$

discussed on the previous practice. Determine (finite) generator systems to each of them.

Remark: If it is possible to write the elements of the set as linear combinations of a finite generator system, then it proves that the set is really a subspace. Thus this is a new possibility to prove, that a set is a subspace.

4. Determine (finite) generator systems for the following subspaces in $\mathbb{R}^3$:

(a) $W_1 = \{(x, y, z) \in \mathbb{R}^3 \mid \begin{bmatrix} 2 & -3 & 5 \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}\}$,

(b) $W_2 = \{(x, y, z) \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & -2 & 3 \\ 2 & 0 & -1 \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\}$,

(c) $W_3 = \{(x, y, z) \in \mathbb{R}^3 \mid \begin{bmatrix} 2 & -1 & -2 \\ -4 & 2 & 4 \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\}$.

Remark: Considering the remark of the previous exercise, it is unnecessary to justify forward that the given sets are really subspaces.

5. Determine a (finite) generator system of the following subspace in $\mathbb{R}^4$:

$$W = \{(2x-y+z,\ y+3z,\ x+y-2z,\ x-y) \in \mathbb{R}^4 \mid x, y, z \in \mathbb{R},\ x+2y+z = 0\} \subseteq \mathbb{R}^4 .$$

Remark: Here it is also unnecessary to justify forward that the given set is really a subspace.

4.2.2. Additional Tasks

1. Consider the vectors $a = (1, 2, -1),\ b = (-3, 1, 1) \in \mathbb{R}^3$.

(a) Compute the vector $2a - 4b$.

(b) Write up several elements of the subspace $\operatorname{Span}(a, b)$.

(c) Determine whether the vectors $x = (2, 4, 0),\ y = (2, 4, -3)$ are contained in $\operatorname{Span}(a, b)$ or not.

2. Determine (finite) generator systems for the following subspaces:

(a) $W_1 = \{(x - y, x + 2y, 3x, 4x + 3y) \in \mathbb{R}^4 \mid x, y \in \mathbb{R}\} \subseteq \mathbb{R}^4$

(b) $W_2 = \{(x, 5x, -4x, 7x) \in \mathbb{R}^4 \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^4$

(c) $W_3 = \{(x, y, z, u) \in \mathbb{R}^4 \mid x - 2y + 4z + 3u = 0\} \subseteq \mathbb{R}^4$

(d) $W_4 = \{(x, y, z, u) \in \mathbb{R}^4 \mid 2x - 3z = 0\} \subseteq \mathbb{R}^4$

(e) $W_5 = \{(x, y, z) \in \mathbb{R}^3 \mid \begin{bmatrix}4 & -1 & 2\end{bmatrix} \cdot \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\end{pmatrix}\} \subseteq \mathbb{R}^3$,

(f) $W_6 = \{(x, y, z) \in \mathbb{R}^3 \mid \begin{bmatrix}1 & -1 & 4\\3 & 1 & 2\end{bmatrix} \cdot \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}\} \subseteq \mathbb{R}^3$,

(g) $W_7 = \{(x, y) \in \mathbb{R}^2 \mid \begin{bmatrix}1 & 2\\3 & 6\\-1 & -2\end{bmatrix} \cdot \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\} \subseteq \mathbb{R}^2$.

3. Determine a (finite) generator system of the following subspace in $\mathbb{R}^4$:

$$W = \{\begin{pmatrix}x+y+u\\3x-2y+5z-u\\2x+2z-u\\3z+2u\end{pmatrix} \in \mathbb{R}^4 \mid x, y, z, u \in \mathbb{R},\ x+y+z+u = 0\} \subseteq \mathbb{R}^4\,.$$

5. Linear Independence

5.1. Theory

5.1.1. The Concept of Linear Independence

5.1. Definition Let $k \in \mathbb{N}^+$ and $x_1, \ldots, x_k \in V$ be a vector system in the vector space V. This vector system is called linearly independent (shortly: independent), if among its possible linear combinations only the trivial linear combination results the zero vector. That is, if

$$\sum_{i=1}^{k} \lambda_i x_i = 0 \quad \Longrightarrow \quad \lambda_1 = \lambda_2 = \ldots = \lambda_k = 0 .$$

The vector system is called linearly dependent (shortly: dependent) if it is not independent, that is if:

$$\exists\, \lambda_1, \lambda_2, \ldots \lambda_k \in \mathbb{K},\ \lambda_i \text{ not all } 0 : \quad \sum_{i=1}^{k} \lambda_i x_i = 0 .$$

5.2. Remarks.

1. The equation $\sum_{i=1}^{k} \lambda_i x_i = 0$ is called dependence equation (or: dependence relation).
2. One can simply prove, that a vector system containing the zero vector or containing identical vectors is linearly dependent. Consequently, a linearly independent system cannot contain neither the zero vector, nor identical vectors.
3. The one-term vector system is linearly independent if and only if its single vector is not the zero vector. This follows immediately from the basic properties of vector spaces.
4. Two nonzero vectors are linearly dependent if and only if they are a constant multiple of each other.
5. In a real vector space (vector space over $\mathbb{R}$) two vectors are called parallel, if they form a two-term linearly dependent system. These

vectors have the same direction, if they are a positive constant multiple of each other, and they have opposite direction, if they are a negative constant multiple of each other.

5.3. Examples

1. Using elementary geometry, one can easily prove, that in the space of space vectors:

- Two vectors lying on the same line are linearly dependent.
- Two vectors not lying on the same line are linearly independent.
- Three vectors lying in the same plane are linearly dependent.
- Three vectors not lying in the same plane are linearly independent.

2. The system of the standard unit vectors $e_1, \ldots, e_n$ (see examples 4.8) are forming a linearly independent system in $\mathbb{K}^n$. To prove this let us see the dependence relation:

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0 = \sum_{i=1}^{n} \lambda_i e_i = \begin{pmatrix} \lambda_1 \cdot 1 + \lambda_2 \cdot 0 + \cdots + \lambda_n \cdot 0 \\ \lambda_1 \cdot 0 + \lambda_2 \cdot 1 + \cdots + \lambda_n \cdot 0 \\ \vdots \\ \lambda_1 \cdot 0 + \lambda_2 \cdot 0 + \cdots + \lambda_n \cdot 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix},$$

which implies $\lambda_i = 0 \quad (i = 1, \ldots, n)$.

5.4. Remark. One can easily see, that if we tighten a linearly independent system in V then it remains linearly independent. But if we enlarge a linearly independent system, then the result system will not necessarily stay linearly independent. The linearly independent systems are – in this sense – the „small" systems. Later on, we will study the question of „maximal" linearly independent systems.

A characteristic property of a linearly independent system is, that if a vector can be written as a linear combination of its vectors, then this prescription is unique. This is expressed in the following theorem:

5.5. Theorem (the theorem of unique representation) *Let* $x_1, \ldots, x_k \in V$ *be a vector system,* $x \in \operatorname{Span}(x_1, \ldots, x_k)$*. Then*

a) If the system $x_1, \ldots, x_k$ *is linearly independent, then* x *is uniquely represented as the linear combination of the given system.*

b) If the system $x_1, \ldots, x_k$ *is linearly dependent, then* x *can be represented in infinitely many ways as the linear combination of the given system.*

Proof. a)
Suppose

$$x = \sum_{i=1}^{k} \lambda_i x_i = \sum_{i=1}^{k} \mu_i x_i \,,$$

and rearrange the right-hand side inequality (reduce it to 0, common sum, factoring out):

$$\sum_{i=1}^{k} (\lambda_i - \mu_i) x_i = 0 \,.$$

Hence – using the independence of the system $x_1, \dots x_k$ – it follows that $\lambda_i - \mu_i = 0$, that is:
$\lambda_i = \mu_i \quad (i = 1, \dots, k)$.

b)
Let a representation of x be:

$$x = \sum_{i=1}^{k} \lambda_i x_i \,.$$

Since the system is linearly dependent, there exist the not-all-zero coefficients α_i such that

$$0 = \sum_{i=1}^{k} \alpha_i x_i \,.$$

Let us multiply this equation by an arbitrary number $\beta \in \mathbb{K}$, and sum it up with the equation which produces x:

$$x + 0 = \sum_{i=1}^{k} \lambda_i x_i + \sum_{i=1}^{k} \beta \alpha_i x_i$$

$$x = \sum_{i=1}^{k} (\lambda_i + \beta \alpha_i) x_i \,.$$

By the linear dependence there exists an index j for which $\alpha_j \neq 0$. But in this case the coefficient $\lambda_j + \beta \alpha_j$ takes infinitely many values if β runs over $\mathbb{K}$.

□

5.1.2. Theorems about Vector Systems

Let us see some theorems about the connection between independent systems, dependent systems, and generator systems.

5.6. Theorem (Diminution of a dependent system)
Let $x_1, \ldots, x_k \in V$ be a linearly dependent system. Then

$$\exists\, i \in \{1,\, 2,\, \ldots\, k\}: \quad \mathrm{Span}\,(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = \mathrm{Span}\,(x_1, \ldots, x_k)\,.$$

In words: In a linearly dependent system there exits a vector, which can be omitted from the system and the generated subspace stays unchanged. In other words: at least one of the vectors in the system is redundant from the point of view of the spanned subspace.

Proof. By the dependence of the system there exist not-all-zero numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ such that

$$\lambda_1 x_1 + \ldots + \lambda_k x_k = 0\,.$$

Let i be an index for which $\lambda_i \neq 0$ holds. Furthermore let

$$W_1 := \mathrm{Span}\,(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \quad \text{és} \quad W_2 := \mathrm{Span}\,(x_1, \ldots, x_k)\,.$$

It must be proved that $W_1 = W_2$.

The comprehension $W_1 \subseteq W_2$ is obvious because, since

$$x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k \in \mathrm{Span}\,(x_1, \ldots, x_k) = W_2$$

the subspace W_2 covers the system $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$. But W_1 is the minimal covering subspace of this system, consequently $W_1 \subseteq W_2$.

To prove the converse comprehension $W_2 \subseteq W_1$, let us start from the following obvious fact:

$$x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k \in \mathrm{Span}\,(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = W_1\,.$$

Now we will prove that $x_i \in W_1$. To see this, rearrange x_i from the dependence relation

$$\lambda_1 x_1 + \ldots + \lambda_k x_k = 0\,.$$

(It is possible because $\lambda_i \neq 0$.)

$$x_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} \left(-\frac{\lambda_j}{\lambda_i} \right) \cdot x_j\,.$$

We get that x_i also can be expressed as the linear combination of vectors $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$, consequently x_i is really contained in the subspace $\operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) = W_1$.

So the subspace W_1 covers the vector system $x_1, \dots, x_k$. Since the subspace W_2 is the minimal covering subspace of this system, then $W_2 \subseteq W_1$.

The containing relationships $W_1 \subseteq W_2$ and $W_2 \subseteq W_1$ together imply, that $W_1 = W_2$.

□

5.7. Remark. It turned out from the proof that the redundant vector is the vector whose coefficient in a dependence equation is not zero.

5.8. Theorem (Extension to a dependent system) *Let* $x_1, \dots, x_k \in V$ *be a vector system, and let* $x \in V$*. Then*

$$x \in \operatorname{Span}(x_1, \dots, x_k) \qquad \Longrightarrow \qquad x_1, \dots, x_k, x \textit{ is linearly dependent.}$$

Proof. $x \in \operatorname{Span}(x_1, \dots, x_k)$, consequently x can be written as the linear combination of the generator system:

$$\exists\, \lambda_1, \dots, \lambda_k \in \mathbb{K} : \quad x = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k \,.$$

After a rearrangement we have:

$$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k + (-1) \cdot x = 0 \,.$$

$-1 \neq 0$, consequently the system is really dependent. □

5.9. Corollary. (Diminution of an independent system) Omitting any vector from a linearly independent system (suppose, that it has originally at least two terms), the remaining system does not generate the same subspace as the original one.

5.10. Theorem (Extension of an independent system) *Let* $x_1, \dots, x_k \in V$ *be a linearly independent system, and let* $x \in V$*. Then*

a) $x \in \operatorname{Span}(x_1, \dots, x_k) \qquad \Longrightarrow \qquad x_1, \dots, x_k, x$ *is linearly dependent*

b) $x \notin \operatorname{Span}(x_1, \dots, x_k) \qquad \Longrightarrow \qquad x_1, \dots, x_k, x$ *is linearly independent*

Proof. The part a) is a special case of the previous theorem.

To prove part b) let us start with the dependence equation:

$$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k + \lambda \cdot x = 0\,,$$

and let us show, that all the coefficients here are 0.

We will show first, that $\lambda = 0$. Suppose indirectly, that $\lambda \neq 0$. Then x can be expressed from the dependence equation:

$$x = -\frac{\lambda_1}{\lambda} x_1 - \ldots - \frac{\lambda_k}{\lambda} x_k\,.$$

This implies, that $x \in \operatorname{Span}(x_1, \ldots, x_k)$, which contradicts to assumption of part b). Thus $\lambda = 0$.

Let us substitute the obtained result into the dependence equation:

$$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k + 0x = 0\,.$$

Using the independence of the original system it follows

$$\lambda_1 = \lambda_2 = \ldots = \lambda_k = 0\,,$$

consequently the system $x_1, \ldots, x_k, x$ is really independent. □

5.11. Corollary. Let $x_1, \ldots, x_k, x \in V$. If the system $x_1, \ldots, x_k$ is linearly independent and $x_1, \ldots, x_k, x$ is linearly dependent, then

$$x \in \operatorname{Span}(x_1, \ldots, x_k)\,.$$

5.1.3. Control Questions to Theory

1. Define the linearly independence and the dependence of finite vector systems

2. Give two examples for linearly independent systems and two examples for linearly dependent systems

3. State the theorem about the diminution of a dependent system

4. State the theorem about the dependence of the system $x_1, \ldots, x_k, x$, where $x \in \operatorname{Span}(x_1, \ldots, x_k)$.

5. State the theorem about the extension of an independent system

5.2. Exercises

5.2.1. Exercises for Class Work:

1. Determine whether the following vector systems in $\mathbb{R}^4$ are linearly independent or dependent:

(a) $v_1 = (1, 2, 2, -1)$; $v_2 = (4, 3, 9, -4)$; $v_3 = (5, 8, 9, -5)$.

(b) $v_1 = (1, 2, 3, 1)$; $v_2 = (2, 2, 1, 3)$; $v_3 = (-1, 2, 7, -3)$.

2. Determine whether the vectors

$$v_1 = (1, 2, 3, 1), \quad v_2 = (2, 2, 1, 3), \quad v_3 = (-1, 2, 7, -3)$$

in $\mathbb{R}^4$ are linearly independent or dependent

Can one vector be omitted from the above system $v_1,\ v_2,\ v_3$, such that the generated subspace does not change? If the answer is „yes", then give such a vector.

3. Expand the linearly independent system $v_1 = (1, -2, 1)$, $v_2 = (2, 1, 0)$ in $\mathbb{R}^3$ by a vector $v_3 \in \mathbb{R}^3$, such that the expanded system v_1, v_2, v_3

(a) is linearly dependent.

(b) is linearly independent.

5.2.2. Additional Tasks

1. Let $v_1 = (1, -2, 3),\ v_2 = (5, 6, -1),\ v_3 = (3, 2, 1) \in \mathbb{R}^3$. Determine whether this system is linearly independent or dependent.

2. Determine whether the vectors

$$v_1 = (-1, 0, 2, 1), \quad v_2 = (3, 4, -1, -5), \quad v_3 = (1, 4, 3, -3)$$

in $\mathbb{R}^4$ are linearly independent or dependent.

Can one vector be omitted from the above system $v_1,\ v_2,\ v_3$, such that the generated subspace does not change? If the answer is „yes", then give such a vector.

3. Expand the linearly independent system $v_1 = (1, 4, -1, 3)$, $v_2 = (-1, 5, 6, 2)$ in $\mathbb{R}^4$ by a vector $v_3 \in \mathbb{R}^4$, such that the expanded system v_1, v_2, v_3

(a) is linearly dependent.

(b) is linearly independent.

6. Basis, Dimension

6.1. Theory

6.1.1. Basis

6.1. Definition The vector system $x_1, \ldots, x_k \in V$ is called basis (in V) if it is generator system and it is linearly independent system at the same time.

6.2. Remark. What is the advantage of a basis? Since it is generator system, each veactor of the space can be written as the linear combination of the basis vectors. By the independence of the basis vectors this production is unique. To summarize it:

Each vector of the space can be written uniquely as the linear combination of the basis vectors. This production is called the expansion of the vector relative to the given basis.

6.3. Definition The coefficients of the above expansion are called the coordinates of the given vector relative to the given basis.

We have seen examples for generator systems and for linearly independent systems respectively, thus the following examples for basis can be easily justified.

6.4. Examples

1. In the space of the plane vectors any two vectors not lying on the same line form a basis.
2. In the space of the space vectors any two vectors not lying in the same plane form a basis.
3. The system of the standard unit vectors in $\mathbb{K}^n$ form a basis. This basis is called the standard basis or the canonical basis of $\mathbb{K}^n$.

One can ask the: is there a basis in any vector space or not.

Since the zero vector space $\{0\}$ has no linearly independent system, then this vector space has no basis. The following theorem states, that apart from this case, every finite dimensional vector space has a basis.

6.5. Theorem (existence of a basis) *Every finite dimensional nonzero vector space V has a basis.*

Proof. Let $x_1, \dots, x_k$ be a finite generator system in V. If this system is linearly independent, then it is basis. If it is dependent then by Theorem 5.6 a vector can be left from it, such that the remainder system spans V. If this new system is linearly independent, then it is a basis. If it is dependent then we leave once more a vector from it, and so on.

Let us continue this process until it is possible.

Thus either in some step we obtain a basis or after $k-1$ steps we arrive to one-element system, that is generator system in V. Since $V \neq \{0\}$, this single vector is nonzero, that is linearly independent, consequently it is a basis. □

6.6. Remark. We have proved more than the statement of the theorem: we have proved, that one can choose bases from any finite generator system, moreover, we have given an algorithm to make this.

We will prove in the following part, that the number of vectors in any two bases of the space are equal. As a first step let us prove the following theorem:

6.7. Theorem (Exchange Theorem) *Let $x_1, \dots, x_k \in V$ be a linearly independent system, $y_1, \dots, y_m \in V$ be a generator system in the vector space V.*

Then for any index $i \in \{1, \dots, k\}$ there exists an index $j \in \{1, \dots, m\}$, such that the vector system

$$x_1,\ \dots,\ x_{i-1},\ y_j,\ x_{i+1},\ \dots,\ x_k$$

is linearly independent.

Proof. It is enough to discuss the case $i = 1$, the proof for the other i-s is similar.

Suppose indirectly that the system $y_j, x_2, \dots, x_k$ is linearly dependent for every $j \in \{1, \dots, m\}$. Since the system $x_2, \dots, x_k$ is linearly independent, then by Corollary 5.11 we have

$$y_j \in \operatorname{Span}(x_2, \dots, x_k) \qquad (j = 1, \dots, m)\,,$$

that is

$$\{y_1,\ \dots,\ y_m\} \subseteq \operatorname{Span}(x_2, \dots, x_k) \subseteq V\,.$$

From here follows, that

$$V = \operatorname{Span}(y_1, \dots, y_m) \subseteq \operatorname{Span}(x_2, \dots, x_k) \subseteq V\,.$$

Since the first and the last members of the above chain coincide, at every $\subseteq$ in it stands equality. This implies, that

$$\operatorname{Span}(x_2, \dots, x_k) = V\,.$$

But $x_1 \in V$, so $x_1 \in \operatorname{Span}(x_2, \dots, x_k)$. This means that x_1 is linear combination of $x_2, \dots, x_k$, in contradiction with the linear independence of $x_1, \dots, x_k$. $\square$

6.8. Theorem *The number of vectors in a linearly independent system is not greater than the number of vectors in a generator system. (Thus we have the precise meaning that the linearly independent systems are the „small" systems, and the generator systems are the „large" systems.*

Proof. Let $x_1, \dots, x_k$ be an independent system and $y_1, \dots, y_m$ be a generator system in V. Using the Exchange Theorem replace x_1 with a suitable y_{j_1} to obtain the linearly independent system $y_{j_1}, x_2, \dots, x_k$. Apply the Exchange Theorem for this new system: replace x_2 with a suitable y_{j_2}, thus we obtain the linearly independent system $y_{j_1}, y_{j_2}, x_3, \dots, x_k$. Continuing this process, we arrive after k steps to the linearly independent system $y_{j_1}, \dots, y_{j_k}$. This system contains different vectors (because of the independence). We have the conclusion, that among the vectors $y_1, \dots, y_m$ k pieces are different. Consequently $k \leq m$. $\square$

6.9. Theorem *Let V be a finite dimensional nonzero vector space. Then in V all bases have the same number of elements.*

Proof. Let $x_1, \dots, x_k$ and $y_1, \dots, y_m$ be two bases in V. By Theorem 6.8 we can deduce that

$$\left.\begin{array}{ll} x_1, \dots, x_k & \text{is independent} \\ y_1, \dots, y_m & \text{is a generator system} \end{array}\right\} \quad \Rightarrow \quad k \leq m$$

On the other hand

$$\left.\begin{array}{ll} y_1, \dots, y_m & \text{is independent} \\ x_1, \dots, x_k & \text{is a generator system} \end{array}\right\} \quad \Rightarrow \quad m \leq k$$

Consequently $k = m$. $\square$

6.10. Definition Let V be a finite-dimensional nonzero vector space. The common number of the bases in V is called the dimension of the space and is denoted by $\dim V$. We agree that by definition $\dim(\{0\}) := 0$. If $\dim V = n$ then V is called an n-dimensional vector space.

6.11. Examples

1. The space of the line vectors is 1 dimensional.
2. The space of the plane vectors is 2 dimensional.
3. The space of the space vectors is 3 dimensional.
4. $\dim \mathbb{K}^n = n \quad (n \in \mathbb{N})$.

The above examples follow immediately from examples 6.4.

6.12. Theorem *[„ 4 small statements"]*
Let $1 \leq \dim(V) = n < \infty$. *Then*

1. *If* $x_1, \ldots, x_k \in V$ *is a linearly independent vector system, then* $k \leq n$.

   Otherwise: Any linearly independent vector system contains up to as many terms as the dimension of the space.

   Even otherwise: Any vector system containing at least $\dim V + 1$ *terms is linearly dependent.*

2. *If* $x_1, \ldots, x_k \in V$ *is a generator system, then* $k \geq n$.

   Otherwise: Any generator system contains at least as many terms as the dimension of the space.

   Even otherwise: Any vector system containing at most $\dim V - 1$ *terms is not a generator system.*

3. *If* $x_1, \ldots, x_n \in V$ *is a linearly independent system, then it is a generator system (consequently: it is a basis).*

   Otherwise: If a linearly independent system contains as many terms as the dimension, then it is a generator system (consequently: it is basis).

4. *If* $x_1, \ldots, x_n \in V$ *is a generator system, then it is linearly independent (consequently: it is a basis).*

   Otherwise: If a generator system contains as many terms as the dimension, then it is linearly independent (consequently: it is basis).

Proof.

1. Let $e_1, \ldots, e_n$ be a basis in V. Then it is a generator system, thus by Theorem 6.8 we have:

$$k \leq n\,.$$

2. Let $e_1, \ldots, e_n$ be a basis in V. Then it is a linearly independent system, thus by Theorem 6.8 we have:

$$k \geq n\,.$$

3. Suppose indirectly that $x_1, \ldots, x_n$ is not a generator system. Then

$$V \setminus \operatorname{Span}(x_1, \ldots, x_n) \neq \emptyset\,.$$

Let $x \in V \setminus \operatorname{Span}(x_1, \ldots, x_n)$. Then by Theorem 5.10 the system $x_1, \ldots, x_n, x$ is linearly independent. This is a contradiction, because this system has $n+1$ terms, more than the dimension of the space.

4. Suppose indirectly that $x_1, \ldots, x_n$ is linearly dependent. Then by Theorem 5.6 we have

$$\exists\, i \in \{1, 2, \ldots n\} : \quad \operatorname{Span}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \operatorname{Span}(x_1, \ldots, x_n) = V\,.$$

This is a contradiction, because the system $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ has $n-1$ terms, less than the dimension of the space.

$\square$

6.1.2. Control Questions to the Theory

1. Define the concept of a basis in a vector space and give 3 examples for bases

2. Define the concept of coordinates

3. What theorem have we learned about the existence of a basis?

4. State the Exchange Theorem

5. State most important corollary of the Exchange Theorem (about the number of terms in a linearly independent and in a generator system)

6. Define the concept of the dimension, and give 3 examples for this concept

7. State and the „4 small statements" about the vector systems in an n-dimensional vector space

6.2. Exercises

6.2.1. Exercises for Class Work

1. Let in $\mathbb{R}^4$ be

$$x_1 = (3, 0, -2, 4), \quad x_2 = (2, 1, -1, 3), \quad x_3 = (-1, 4, 2, 0), \quad x_4 = (-1, 1, 1, -1)\,.$$

Select a basis from the generator system in the subspace generated by them. What is the dimension of this subspace?

2. Determine whether the following vector systems form a basis in $\mathbb{R}^4$ or not.

$$a)\;\; x_1,\; x_2 \qquad b)\;\; x_1,\; x_2,\; x_3,\; x_4,\; x_5 \qquad c)\;\; x_1,\; x_2,\; x_3,\; x_4,$$

where

$$x_1 = (2, 3, -2, 7), \qquad x_2 = (0, 1, 0, 1), \quad x_3 = (1, 2, -1, 0),$$
$$x_4 = (-1, -5, 2, 0), \quad x_5 = (3, -1, 1, 2).$$

3. Select a basis from the following vector systems in $\mathbb{R}^4$ in the subspace $W = \text{Span}\,(x_1, x_2, x_3)$. Determine $\dim W$ too.

a)

$$x_1 = (1, 2, 2, -1)\;; \quad x_2 = (4, 3, 9, -4)\;; \quad x_3 = (5, 8, 9, -5).$$

b)

$$x_1 = (1, 2, 3, 1)\;; \quad x_2 = (2, 2, 1, 3)\;; \quad x_3 = (-1, 2, 7, -3).$$

4. The following vector systems form bases in $\mathbb{R}^3$ or not?

(a) $(1, 0, 0)$, $(2, 2, 0)$, $(3, 3, 3)$

(b) $(3, 1, -4)$, $(2, 5, 6)$, $(1, 4, 8)$

(c) $(2, -3, 1)$, $(4, 1, 1)$, $(0, -7, 1)$, $(1, 6, 4)$

(d) $(2, 4, -1)$, $(-1, 2, 5)$

6.2.2. Additional Tasks:

1. Which of the following vector systems form a basis?

(a) $x_1 = (1, 0, 0)$, $x_2 = (2, 2, 0)$, $x_3 = (3, 3, 3)$ in $\mathbb{R}^3$.

(b) $y_1 = (3, 1, -4)$, $y_2 = (2, 5, 6)$, $y_3 = (1, 4, 8)$ in $\mathbb{R}^3$.

(c) $z_1 = (1, 2, -1, 0)$, $z_2 = (0, 1, 0, 1)$, $z_3 = (-1, -5, 2, 0)$, $z_4 = (2, 3, -2, 7)$ in $\mathbb{R}^4$.

(d) $v_1 = (1, 2, 1, 2)$, $v_2 = (2, 1, 0, -1)$, $v_3 = (-1, 4, 3, 8)$, $v_4 = (0, 3, 2, 5)$ in $\mathbb{R}^4$.

2. Using the data of the previous exercise, select a basis from the given vector systems in the subspace generated by them. What are the dimensions of these subspaces?

7. Rank, System of Linear Equations

7.1. Theory

7.1.1. The Rank of a Vector System

In this section we will characterize the measure of dependence of a vector system. For example we feel that in the vector space of the space vectors three vectors are „much more" if they lie on a straight line, than if they lie in a plane, but not in a line. This observation motivates the following definition.

7.1. Definition Let V be a vector space, $x_1, \ldots, x_k \in V$.

The dimension of the subspace generated by the system $x_1, \ldots, x_k$ is called the rank of the vector system. It is denoted by $\operatorname{rank}(x_1, \ldots, x_k)$. Thus

$$\operatorname{rank}(x_1, \ldots, x_k) := \dim \operatorname{Span}(x_1, \ldots, x_k) .$$

7.2. Remarks.

1. We see that $0 \leq \operatorname{rank}(x_1, \ldots, x_k) \leq k$.

2. The rank expresses really the measure of dependence. The smaller is the rank the more dependent are the vectors. Especially:

$$\operatorname{rank}(x_1, \ldots, x_k) = 0 \quad \Leftrightarrow \quad x_1 = \ldots = x_k = 0 \quad \text{and}$$

$$\operatorname{rank}(x_1, \ldots, x_k) = k \quad \Leftrightarrow \quad x_1, \ldots, x_k \text{ are linearly independent} .$$

3. $\operatorname{rank}(x_1, \ldots, x_k)$ is the maximal number of linearly independent vectors in the system $x_1, \ldots, x_k$.

7.1.2. The Rank of a Matrix

7.3. Definition Let $A \in \mathbb{K}^{m \times n}$. The entries in the i-th row of A form the i-th row vector of A:

$$s_i := (a_{i1},\, a_{i2},\, \dots,\, a_{in}) \in \mathbb{K}^n \qquad (i = 1, \dots, m)$$

The subspace generated by the row vectors $s_1,\, s_2,\, \dots,\, s_m$ is called the row vector space or simply the row space of A. It is denoted by $\mathrm{Row}(A)$. Thus

$$\mathrm{Row}(A) = \mathrm{Span}\,(s_1,\, \dots,\, s_m) \subseteq \mathbb{K}^n$$

7.4. Definition Let $A \in \mathbb{K}^{m \times n}$. The entries in the j-th column of A form the j-th column vector of A:

$$a_j := \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{K}^m \qquad (j = 1, \dots, n)$$

The subspace generated by the column vectors $a_1,\, a_2,\, \dots,\, a_n$ is called the column vector space or simply the column space of A. It is denoted by $\mathrm{Col}(A)$. Thus

$$\mathrm{Col}(A) = \mathrm{Span}\,(a_1,\, \dots,\, a_n) \subseteq \mathbb{K}^m$$

7.5. Remark. If $A \in \mathbb{K}^{m \times n}$ then we have the following simple observations:

$$\dim \mathrm{Row}(A) \le m, \qquad \dim \mathrm{Row}(A) \le n, \qquad \dim \mathrm{Col}(A) \le n, \qquad \dim \mathrm{Col}(A) \le m\,,$$

$$\mathrm{Row}(A^T) = \mathrm{Col}(A) \subseteq \mathbb{K}^m \quad \text{and} \quad \mathrm{Col}(A^T) = \mathrm{Row}(A) \subseteq \mathbb{K}^n\,.$$

7.6. Theorem *The dimensions of the row space and the column space are equal, that is for any matrix $A \in \mathbb{K}^{m \times n}$ holds*

$$\dim \mathrm{Col}(A) = \dim \mathrm{Row}(A)\,.$$

Proof. The statement is obviously true for $A = 0$. Suppose $A \neq 0$.

Let $r := \dim \mathrm{Col}(A) \ge 1$, and $b_1,\, \dots,\, b_r \in \mathbb{K}^m$ be a basis in $\mathrm{Col}(A)$. Denote by $B \in \mathbb{K}^{m \times r}$ the matrix that consists of the column vectors $b_1,\, \dots,\, b_r$:

$$B := [b_1 \ \dots \ b_r] \in \mathbb{K}^{m \times r}\,.$$

Then the column vectors of A can be written as the linear combinations of the vectors $b_1, \ldots, b_r$:

$$\exists\, d_{ij} \in \mathbb{K}: \quad a_j = \sum_{i=1}^{r} d_{ij} b_i \qquad (j = 1, \ldots, n)\,.$$

Let $D = (d_{ij}) \in \mathbb{K}^{r \times n}$. One can see – using the rule of the matrix product that

$$A = BD\,.$$

Let us consider this equation as each row vector of A can be expressed as the linear combinations of the row vectors of D (the coefficients are the the entries in the appropriate rows of B). By this reason, each row vector of A is contained in $\mathrm{Row}(D)$, consequently $\mathrm{Row}(A) \subseteq \mathrm{Row}(D)$. Hence it follows that

$$\dim \mathrm{Row}(A) \leq \dim \mathrm{Row}(D) \leq r = \dim \mathrm{Col}(A)\,.$$

We have proved that $\dim \mathrm{Row}(A) \leq \dim \mathrm{Col}(A)$. If we apply this result for A^T instead of A, then we obtain the opposite inequality:

$$\dim \mathrm{Col}(A) = \dim \mathrm{Row}(A^T) \leq \dim \mathrm{Col}(A^T) = \dim \mathrm{Row}(A)\,.$$

So the theorem has been proved. □

7.7. Remark. In this proof it was not necessary to choose the basis of $\mathrm{Col}(A)$ from the column vectors of A.

If we choose the basis from the column vectors of A, then the r columns of D corresponding to the basis indices are the r standard unit vectors. Roughly speaking we say that D contains an $r \times r$ identity matrix. In this case the factorization $A = BD$ is called a basis factorization of A (according to the chosen basis).

7.8. Definition Let $A \in \mathbb{K}^{m \times n}$.

The common value of $\dim \mathrm{Col}(A)$ and $\dim \mathrm{Row}(A)$ is called the rank of the matrix A, and it is denoted by $\mathrm{rank}\,(A)$. That is

$$\mathrm{rank}\,(A) := \dim \mathrm{Row}(A) = \dim \mathrm{Col}(A)\,.$$

7.9. Remarks.

Let $A \in \mathbb{K}^{m \times n}$. Then

1. $\mathrm{rank}\,(A) = \mathrm{rank}\,(a_1, \ldots, a_n) = \mathrm{rank}\,(s_1, \ldots, s_m)$, where $a_1, \ldots, a_n$ is the system of the columns of A, $s_1, \ldots, s_m$ is the system of the rows of A respectively.

2. $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

3. $0 \le \operatorname{rank}(A) \le \min\{m, n\}$,

$\operatorname{rank}(A) = 0 \iff A = 0$.

4. $\operatorname{rank}(A) = m$ if and only if the row vectors of A are linearly independent.

$\operatorname{rank}(A) = n$ if and only if the column vectors of A are linearly independent.

7.1.3. System of Linear Equations (Linear Systems)

7.10. Definition Let $m \in \mathbb{N}^+$ and $n \in \mathbb{N}^+$ be positive integers. The general form of the $m \times n$ system of linear equations (or: linear equation system, or simply: linear system) is:

$$\begin{array}{ccccccc} a_{11}x_1 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots \\ a_{m1}x_1 & + & \dots & + & a_{mn}x_n & = & b_m \end{array},$$

where the coefficients $a_{ij} \in \mathbb{K}$ and the right-side constants b_i are given. This form is called the scalar form of the linear system.

We are looking for all the possible values of the unknowns (or: variables) $x_1, \dots, x_n \in \mathbb{K}$ such that all the equations will be true. Such a system $x_1, \dots, x_n$ of values of the variables is called a solution of the linear system.

7.11. Definition The linear system is named consistent if it has a solution. It is named inconsistent if it has no solution.

Let us introduce the vectors

$$a_1 := \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad \dots, \quad a_n := \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}, \quad b := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

in $\mathbb{K}^m$. Using them the linear system can be written in the following simpler form:

$$x_1 a_1 + x_2 a_2 + \dots + x_n a_n = b\,. \tag{7.1}$$

This form is called the vector form of the linear system. Now we reformulate the question in the following way: can we expand vector b as a linear combination of the vectors $a_1, \dots, a_n$? If yes, then compute the coefficients of all the possible expansions.

Finally, if we introduce the matrix

$$A := [a_1 \dots a_n] := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{K}^{m \times n},$$

which is called the coefficient matrix or simply the matrix of the system, and the unknown vector $x := (x_1, \dots, x_n) \in \mathbb{K}^n$, then the shortest form of the linear system is

$$Ax = b\,. \tag{7.2}$$

This is called the matrix form of the linear system.

In this form the problem is to find all the possible vectors x in $\mathbb{K}^n$ for which the statement $Ax = b$ will be true. Such a vector (if it exists) is called a solution vector of the system.

7.12. Remark. It is easy to observe that

$$\text{the system is consistent} \quad \Longleftrightarrow \quad b \in \operatorname{Span}(a_1, \dots, a_n) = \operatorname{Col}(A)\,.$$

Thus the consistence of a linear system is equivalent with the question that b lies in the column space of A or not. Consequently the smaller the column space is, the greater is the chance of inconsistence. If rank (A) equals the number of rows m (in other words: the rows of A are linearly independent), then $\operatorname{Col}(A)$ is the possible greatest subspace that is $\operatorname{Col}(A) = \mathbb{K}^m$. In this case the system is surely consistent.

Let us denote by $\mathcal{M}$ the set of the solution vectors of the system $Ax = b$:

$$\mathcal{M} := \{x \in \mathbb{K}^n \mid Ax = b\} \subseteq \mathbb{K}^n\,.$$

It is called solution set.

7.13. Definition Two liner systems are said to be equivalent, if their solution sets are the same.

7.14. Remark. The following statements can be considered easily:

1. The linear system can be solved (is consistent) if and only if $\mathcal{M} \neq \emptyset$.

2. The solution set of the linear system equals the intersection of the solution sets of the equations contained in the system.

3. If the linear system contains at least one equation of the form

$$0x_1 + 0x_2 + \ldots + 0x_n = q \qquad (q \neq 0)\,,$$

then the system is inconsistent (it has no solution).

4. The following transformations result equivalent linear systems:

   (a) We multiply an equation by a nonzero constant.

   (b) We add the constant multiple of an equation to another equation.

   (c) We omit an equation of the form

$$0x_1 + 0x_2 + \ldots + 0x_n = 0$$

   from the system.

7.15. Definition Let $A \in \mathbb{K}^{m \times n}$. The linear system $Ax = 0$ is called homogeneous system. We often say that $Ax = 0$ is the homogeneous system associated with $Ax = b$.

Notice, that the homogeneous system is always solvable, because the zero vector is surely a solution of it.

7.16. Theorem *Let $\mathcal{M}_h$ be the solution set of the homogeneous system, that is*

$$\mathcal{M}_h := \{x \in \mathbb{K}^n \mid Ax = 0\} \subseteq \mathbb{K}^n\,.$$

Then $\mathcal{M}_h$ is a subspace in $\mathbb{K}^n$.

Proof. Since $0 \in \mathcal{M}_h$, then $\mathcal{M}_h \neq \emptyset$.

$\mathcal{M}_h$ is closed under addition, because if $x, y \in \mathcal{M}_h$, then $Ax = Ay = 0$, consequently

$$A(x + y) = Ax + Ay = 0 + 0 = 0\,.$$

Hence it follows that $x + y \in \mathcal{M}_h$.

Furthermore $\mathcal{M}_h$ is closed under scalar multiplication too, because if $x \in \mathcal{M}_h$ and $\lambda \in \mathbb{K}$, then $Ax = 0$, consequently

$$A(\lambda x) = \lambda A x = \lambda 0 = 0\,.$$

Hence it follows that $\lambda x \in \mathcal{M}_h$. □

7.17. Definition Let $A \in \mathbb{K}^{m\times n}$. The subspace $\mathcal{M}_h$ is called the null space or the kernel space or simply the kernel of matrix A. Its notation is $\mathrm{Ker}\,(A)$. So

$$\mathrm{Ker}\,(A) := \mathcal{M}_h = \{x \in \mathbb{K}^n \;\mid\; Ax = 0\} \subseteq \mathbb{K}^n\,.$$

We now move on to the investigation of the solution sets of consistent linear systems.

7.18. Theorem (Basic Theorem of Linear Systems)

Let $A \in \mathbb{K}^{m\times n}$, $r = \mathrm{rank}\,(A) \geq 1$ (that is $A \neq 0$), $b \in \mathbb{K}^m$, and let us consider the consistent linear system

$$Ax = b\,.$$

Let

$$A = [a_1 \;\ldots\; a_n]\,, \quad \textit{ahol} \quad a_j \in \mathbb{K}^m\,.$$

be the column-partitioned form of A.

We know that the column vectors of A form a generator system in $\mathrm{Col}(A)$*, furthermore* $\dim \mathrm{Col}(A) = r$*. By this reason we can choose r vectors from the columns of A, which r vectors form a basis in* $\mathrm{Col}(A)$*. Suppose for simplicity that these r vectors are $a_1, \ldots, a_r$, the first r columns of A.*

Let the unique expansion of b on this basis be

$$b = \sum_{i=1}^{r} c_i a_i\,.$$

Then

a) in the case $r = n$ the linear system has the unique solution:

$$x_i = c_i \qquad (i = 1, \dots, r = n)\,.$$

b) in the case $1 \le r < n$ let us consider the unique expansions of the column vectors $a_{r+1}, \dots, a_n \in \mathrm{Col}(A)$ relative to the basis $a_1, \dots, a_r$:

$$a_j = \sum_{i=1}^{r} d_{ij} a_i \qquad (j = r+1, \dots, n)\,.$$

Then all the solutions of the linear system are:

$$x_i = c_i - \sum_{j=r+1}^{n} d_{ij} x_j \qquad (i = 1, \dots, r)\,, \quad x_{r+1}, \quad \dots, \quad x_n\,,$$

where $x_{r+1}, \dots, x_n \in \mathbb{K}$ are arbitrary numbers. The above formula gives the general solution of the system.

Proof. a) Case $r = n$

It is trivial by the independence of the vectors $a_1, \dots, a_n$ and by the theorem of the unique expansion.

b) Case $1 \le r < n$

We arrive to the general solution by the following equivalent transformations:

$$\begin{aligned} Ax &= b \\ \sum_{i=1}^{n} x_i a_i &= b \\ \sum_{i=1}^{r} x_i a_i + \sum_{j=r+1}^{n} x_j a_j &= b \end{aligned}$$

we substitute b and a_j with their expansions:

$$\sum_{i=1}^{r} x_i a_i + \sum_{j=r+1}^{n} x_j \cdot \sum_{i=1}^{r} d_{ij} a_i = \sum_{i=1}^{r} c_i a_i$$

We interchange the order of summations, then we rearrange the equations:

$$\sum_{i=1}^{r}\left(x_i+\sum_{j=r+1}^{n} d_{ij}x_j\right)\cdot a_i=\sum_{i=1}^{r} c_i a_i$$

using the independence of the vectors a_i and the theorem of unique expansion we have

$$x_i+\sum_{j=r+1}^{n} d_{ij}x_j=c_i \quad (i=1,\ldots,r)$$

$$x_i=c_i-\sum_{j=r+1}^{n} d_{ij}x_j \quad (i=1,\ldots,r)\,.$$

□

7.19. Remarks.

1. If we agree that the value of the empty sum equals 0, then the cases $r=n$ and $r<n$ can be uniformly summarized in the formula

$$x_i=c_i-\sum_{j=r+1}^{n} x_j d_{ij} \quad (i=1,\ldots,r), \qquad x_{r+1},\,\ldots,\,x_n\in\mathbb{K}\,. \tag{7.3}$$

2. The variables $x_{r+1},\,\ldots,\,x_n$ are called free variables, they have arbitrary values. The variables $x_1,\,\ldots,\,x_r$ are called bound variables. They are uniquely depending on the free variables. The number $n-r$ is called the degree of freedom of the linear system. It gives the number of the free variables.

3. One can see, that in the case $r=n$ the degree of freedom is 0, there is no free variable, all the variables are bound variables, the solution is unique. In the case $r<n$ we have infinitely many solutions with $n-r$ free variables.

4. If we choose in $\mathrm{Col}(A)$ another basis instead of the first r columns, then we obtain a similar theorem, only the indexing will be more complicated.

Now we will arrange the solutions into a vector, thus we will arrive to the vector form of the solutions.

Using the equations of (7.3) we have:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 - d_{1,r+1}x_{r+1} - \ldots - d_{1n}x_n \\ c_2 - d_{2,r+1}x_{r+1} - \ldots - d_{2n}x_n \\ \vdots \\ c_r - d_{r,r+1}x_{r+1} - \ldots - d_{rn}x_n \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_{r+1} \cdot \begin{pmatrix} -d_{1,r+1} \\ -d_{2,r+1} \\ \vdots \\ -d_{r,r+1} \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + x_n \cdot \begin{pmatrix} -d_{1n} \\ -d_{2,n} \\ \vdots \\ -d_{rn} \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

This means shortly, that:

$$x = x^B + x_{r+1} \cdot v_{r+1} + \ \ldots \ + x_n \cdot v_n = x^B + \sum_{j=r+1}^{n} x_j v_j \,, \qquad (7.4)$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix}, \quad x^B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and

$$v_{r+1} = \begin{pmatrix} -d_{1,r+1} \\ -d_{2,r+1} \\ \vdots \\ -d_{r,r+1} \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad v_n = \begin{pmatrix} -d_{1n} \\ -d_{2,n} \\ \vdots \\ -d_{rn} \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \qquad (7.5)$$

Hence we have

$$\mathcal{M} = \left\{ x^B + \sum_{j=r+1}^{n} x_j v_j \ \middle| \ x_j \in \mathbb{K} \right\} \subseteq \mathbb{K}^n \,. \qquad (7.6)$$

7.20. Remark. One can see, that x^B is a solution of the linear system, moreover, in the case $r = n$ this is the unique solution.

The following theorem gives the structure of the solution set:

7.21. Theorem (The structure of the solution set)
Under the assumptions of theorem 7.18 we have

1. *The solution set $\mathcal{M}_h$ of the homogeneous system $Ax = 0$ is an $n - r$ dimensional subspace in $\mathbb{K}^n$. A basis of this subspace is the vector system $v_{r+1}, \ldots, v_n$ given by the formulas* (7.5).
2. *If the system $Ax = b$ is consistent (solvable), then its solution set $\mathcal{M}$ is a shifting of the subspace $\mathcal{M}_h$ with the vector x^B.*

Proof.

1. Since the system is homogeneous, then $b = 0$. Consequently $c_1 = \ldots = c_r = 0$, that is $x^B = 0$. Let us substitute this into formula (7.6):

$$\mathcal{M}_h = \left\{ \sum_{j=r+1}^{n} x_j v_j \mid x_j \in \mathbb{K} \right\}.$$

– In the case $r = n$ it means, that $\mathcal{M}_h = \{0\}$ (the case of the empty sum). Consequently $\dim \mathcal{M}_h = 0 = n - n = n - r$.

– In the case $r < n$ it means, that $v_{r+1}, \ldots, v_n$ is a generator system in the subspace $\mathcal{M}_h$. On the other hand – by the $0 - 1$ components – the vectors $v_{r+1}, \ldots, v_n$ are linearly independent. Thus the vector system $v_{r+1}, \ldots, v_n$ is a basis in the subspace $\mathcal{M}_h$, consequently $\dim \mathcal{M}_h = n - r$.

2. It follows immediately from formula (7.6).

□

7.22. Remarks.

1. In the case $r = n$ the sums in the formulas of $\mathcal{M}$ and $\mathcal{M}_h$ are empty, consequently we have

$$\mathcal{M}_h = \{0\}, \quad \dim \mathcal{M}_h = 0, \quad \mathcal{M} = \{x^B\}.$$

2. Using $\operatorname{Ker}(A) = \mathcal{M}_h$ and $\dim \mathcal{M}_h = n - r$ and $\dim \operatorname{Col}(A) = r$, we have the following important equality:

$$\dim \operatorname{Ker}(A) + \dim \operatorname{Col}(A) = n\,.$$

Our theory from above cannot be used to solve a linear system in practice, because it didn't give an algorithm
– to decide the solvability (consistence) of the linear system,
– to determine a basis in $\operatorname{Col}(A)$,
– to determine the numbers c_i and d_{ij}.

We will deal with the practical solution in the following section.

7.1.4. Solving a Linear System in Practice

In the secondary school we have learnt two methods for solving systems of linear equations: the method of substitution and the method of equal coefficients.

The essentiality of the substitution method is the following:

1. From one of the equations we express one of the unknowns. We mark the equality of expression (for example we frame it).

2. The expression resulting for this unknown we substitute into all the other equations. Thus we get a system involving one less equations and one less unknowns.

3. We repeat this process until it is possible.

4. After the process has stopped, first we find out whether the system has solution or not. If it has, then – using the marked equalities – we determine the values of the unknowns.

The essential step of the method of equal coefficients is as follows:

1. Choose two equations from the system in which the coefficients of the same unknowns are equal. If there are no such equations, then – by multiplying both sides of the chosen two equations – achieve that the coefficients of one of the unknowns to be equal.

2. Subtract this two equations from each other. Then the above mentioned unknown falls out in the resulted equation.

3. We replace one of the originally chosen equations by the resulted equation. The new system will be equivalent with the original one. But one of its equation surely does not contain the above mentioned unknown.

We can see that the essential step of the method of equal coefficients can be performed by the method of substitution too: let us express an unknown from one of the chosen two equations, and let us substitute it into another one. We illustrate this in the following example:

Suppose that the chosen two equations are:

$$\begin{array}{l} 5x_1 - 3x_2 + 2x_3 = 7 \\ \underline{5x_1 + \ x_2 \ + \ x_3 \ = 3} \end{array}$$

Let us subtract the second equation from the first one (method of equal coefficients):

$$-4x_2 + x_3 = 4\,.$$

If we apply the substitution method, then let us express x_1 from the first equation:

$$x_1 = \frac{3x_2 - 2x_3 + 7}{5}\,,$$

then let us substitute it into the second equation:

$$\begin{aligned} 5 \cdot \frac{3x_2 - 2x_3 + 7}{5} + x_2 + x_3 &= 3 \\ 4x_2 - x_3 &= -4 \\ -4x_2 + x_3 &= 4\,. \end{aligned}$$

We can see, that we have obtained the same equation like at the method of equal coefficients.

The essential step of the method of equal coefficients can be regarded as we have added a constant multiple of an equation to another one. Thus we have reached, that an unknown variable has vanished from the equation. If we improve this idea in the way that we add the appropriate constant multiples of a fixed equation to all the other equations, then we can remove (in Latin terminology: eliminate) an unknown variable from all, but one equation. This is the base of the elimination methods will be studied in the subject „Numerical Methods”.

In our subject Basic Mathematics we will investigate only one version, the Gauss-Jordan elimination method.

The first important remark is that instead of writing a complete linear system we will write only its coefficients and its right-hand side constant terms in a table in the following way:

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n = b_2 \\ \vdots \qquad \vdots \qquad\qquad \vdots \\ \underline{a_{m1}x_1 + \ldots + a_{mn}x_n = b_m} \end{array} \quad \Longleftrightarrow \quad \begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}.$$

If an unknown is absent from an equation, then the entry in the table equals 0, according to the term $0x_j$. This table is called the augmented matrix of the linear system.

One can easily see, that the representation by the augmented matrix requires less writing. The equations are identified by the rows, the unknowns are identified by the columns to the left of the vertical line, the right-hand-side constants are identified by the column to the right of the vertical line.

The linear system and the augmented matrix can be generated from each other easily. The operations with the equations correspond to the operations with the rows of the augmented matrix.

Suppose that the coefficient matrix of the linear system is not the zero matrix (in the opposite case the system can be discussed easily). The essentiality of the Gauss-Jordan method is the iterative repetition of the following steps, and after the process has stopped, reading the results out of the last table.

1. We choose to the left of the vertical line a nonzero element, which is neither in a marked row nor in a marked column (at the start of the process there is no marked row and there is no marked column). The chosen element will be named pivot element. Its row will be the pivot row, its column will be the pivot column, the unknown according to the pivot column will be the pivot unknown (pivot variable). If we cannot choose pivot element, then the process stops.

2. We divide the pivot row by the pivot element. We mark the 1 in the place of the pivot element (it becomes to a marked element).

3. We subtract from each rows except the pivot row the scalar multiple of the pivot row such that the element of this row in the pivot column becomes 0 (we reset the entries in the pivot column, except the pivot entry). Thus we have reached that the pivot column contains a marked 1, and all its other elements are 0-s. The short name of this step will be „the zeroing of the pivot column".

This means in connection with the linear system that we have eliminated an unknown from the equations according to the non-pivot rows, but we have left this unknown with coefficient 1 in the equation according to the pivot row.

4. We change the attributes „pivot" (element, row, column, unknown) into „marked", then we go to point 1.

The above process is still continued until it terminates by the stopping criterion in point 1., that is, until we find pivot element. The process is obviously finite, because the number of the marked elements (but the number of the marked rows and of columns) increases by 1 in every cycle.

Suppose at the termination the number of the marked elements is r. Then the numbers of the marked rows, columns and unknowns are also r respectively. Remember, that a marked row is a row which contains a marked element, and a marked column is a column which contains a marked element.

After the termination of the process we consider two cases:

Case 1., if $r = m$: In this case each row contains a marked element, that is each row is a marked row. By this reason, no more pivot element can be chosen. In this case we call the last table reduced table.

Case 2., if $r < m$: In this case there exist non-marked rows, but it stands only 0 elements in any non-marked row to the left of the vertical line. This is the reason, that we cannot choose a pivot element.

In this case the equation according to a non-marked row looks like:

$$0x_1 + 0x_2 + \ldots + 0x_n = q\,.$$

It is obvious that this equation has no solution if $q \neq 0$ (prohibited row), consequently the linear system also does not have solution. In the case $q = 0$ any vector $x \in \mathbb{K}^n$ is a solution of the equation, consequently the equation can be omitted from the system.

By this reason in the case $r < m$ we have:

- If there is a non-marked row whose last element (that is the element to the right of the vertical line) is nonzero (prohibited row), then the linear system is unsolvable (inconsistent). No reduced table is produced.

- If the last element (that is the element to the right of the vertical line) of each non-marked row is 0, then we omit the non-marked rows, and the remainder table will be called the reduced table.

We will show, that if the reduced table exists (as we have seen, it is possible in two ways), then the linear system is solvable (consistent).

The reduced table has the following properties:

1. It has r rows, and each of its rows is a marked row. This means that each of its rows has marked element, and we know that this marked element equals 1.

2. n columns stand to the left of the vertical line, and there are exactly r marked columns between them. Furthermore, these r marked columns are exactly the r r-dimensional standard unit vectors. 1 column stands to the right of the vertical line.

3. Hence it follows that in each row stands exactly 1 of the marked unknowns. The coefficient of this marked unknown is 1, and this marked unknown does not occur in another row. The r marked unknowns are separated in the r equations.

Finally, we express from each equation the single marked unknown with the help of the right-side constants and – if it exists – with the help of the non-marked unknowns. This is very simple, because

– in the case $r = n$ the value of the marked unknown is the right-side constant, so one can simply read out it from the table.

– in the case $r < n$ the terms containing non-marked unknowns must be put into the right side.

So, it can be seen, that if the reduced table exists, then the linear system has a solution. The solution is unique in the case when $r = n$, and there are infinitely many solutions with $n - r$ free parameters in the case when $r < n$.

It can be proved (see e.g.: István Csörgő: Linear algebra lecture schemes (2016) Lessons 7 and 8 about the Elementary Basis Transformation), that the Gauss-Jordan-method and the Elementary Basis Transformation Method are the same. Thus we have:

1. The rank of the coefficient matrix is r. Consequently, the number r determined in the Gauss-Jordan method is independent of choosing the pivot elements.

2. The degree of the freedom of the linear system is $n - r$.

3. The bound variables are identical with the marked variables (unknowns).

4. The free variables are identical with the non-marked unknowns. If there is no non-marked unknown, then there is no free variable.

5. The numbers c_i can be read out from the column to the right of the vertical line in the reduced table.

6. The numbers d_{ij} can be read out from the area to the left of the vertical line in the reduced table.

7.23. Remarks.

1. If the system is homogeneous, then in addition to the above, it is also true that the column to the right of the vertical line contains only 0-s in each table. By this reason, the reduced table exists, and its last column (the column to the right of the vertical line) is identically 0.

2. The Gauss-Jordan elimination method can be used also to determine the rank of a matrix. Then we copy the matrix A in the starting table, there is no vertical line, and there is no column to the right of the vertical line. We perform the elimination steps (repetitions of the steps 1-2-3-4 until it is possible). The rank of the matrix A will be the number of the marked elements: $\operatorname{rank}(A) = r$.

7.1.5. Three Computed Examples

In this section we will exemplify the Gauss-Jordan method and its applications via 3 computed examples. In all of the 3 examples the questions are as follows:

a) Determine the size, the coefficient matrix and the right-hand side vector of the linear system.

b) Solve the linear system using the Gauss-Jordan method (solvability, solutions, bound and free variables).

c) The rank of the coefficient matrix.

d) The vector form of the solution (if the system is solvable).

e) The solution set $\mathcal{M}$ (if the system is solvable).

f) Solve the corresponding homogeneous system (solutions, solution set $\mathcal{M}_h = \operatorname{Ker}(A)$, basis and dimension of $\mathcal{M}_h$).

In the process of the Gauss-Jordan elimination method the pivot elements will be denoted by framing, the marked elements will be denoted by underlining.

Example 1 *(the case of unique solution)*

$$\begin{aligned} x_2 - 3x_3 &= -5 \\ 4x_1 + 5x_2 - 2x_3 &= 10 \\ 2x_1 + 3x_2 - x_3 &= 7 \end{aligned}$$

Solution

$\boxed{\text{a)}}$

Obviously $m = 3$, $n = 3$, that is the size of the system is 3×3. Its coefficient matrix and right-hand side vector are:

$$A = \begin{bmatrix} 0 & 1 & -3 \\ 4 & 5 & -2 \\ 2 & 3 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 3}, \qquad b = \begin{pmatrix} -5 \\ 10 \\ 7 \end{pmatrix} \in \mathbb{R}^3 .$$

$\boxed{\text{b)}}$

Let us write the table (augmented matrix) corresponding to the original linear system. In step 1. let us choose the second entry of the first row as pivot element. Thus the first row will be the pivot row, the second column will be the pivot column, the unknown x_2 will be the pivot variable.

$$\begin{array}{ccc|c} 0 & \boxed{1} & -3 & -5 \\ 4 & 5 & -2 & 10 \\ 2 & 3 & -1 & 7 \end{array} .$$

Now it follows step 2. First we divide the pivot row (it is row 1) by the pivot element (it is 1), then we mark the 1 in the place of pivot element.

Then it follows step 3., the zeroing of the pivot column. First we subtract $\frac{5}{1} = 5$-times the first row from the second row, then we subtract $\frac{3}{1} = 3$-times the first row from the third row. The result is

$$\begin{array}{ccc|c} 0 & \underline{1} & -3 & -5 \\ 4 & 0 & 13 & 35 \\ 2 & 0 & 8 & 22 \end{array} .$$

In step 4. the first row will be marked, the second column will be marked, the unknown x_2 will be marked.

The first cycle is complete.

Now it follows again step 1. Let us choose a pivot element (we can choose between the 4, 13, 2, 8), let it be the first element of the third row:

$$\begin{array}{ccc|c} 0 & \underline{1} & -3 & -5 \\ 4 & 0 & 13 & 35 \\ \boxed{2} & 0 & 8 & 22 \end{array} .$$

Thus the third row is the pivot row, the first column is the pivot column, then unknown x_1 is the pivot variable.

It follows step 2., we divide the pivot row (row 3) by the pivot element (by 2), and we mark the 1 in the position of the pivot element. Thus we have already two marked elements.

After this it follows step 3., the zeroing of the pivot column. First we subtract $\frac{0}{2} = 0$-times the third row from the first row, then we subtract $\frac{4}{2} = 2$-times the third row from the second row. The result is

$$\begin{array}{ccc|c} 0 & \underline{1} & -3 & -5 \\ 0 & 0 & -3 & -9 \\ \underline{1} & 0 & 4 & 11 \end{array} .$$

By step 4. the third row will be marked, the first column will be marked, the unknown x_1 will be marked. Now we have two marked elements, two marked rows, two marked columns, and two marked unknowns.

The second cycle is complete.

Now it follows again step 1. Let us choose a pivot element (we can choose only the -3):

$$\begin{array}{ccc|c} 0 & \underline{1} & -3 & -5 \\ 0 & 0 & \boxed{-3} & -9 \\ \underline{1} & 0 & 4 & 11 \end{array} .$$

Thus the second row is the pivot row, the third column is the pivot column, then unknown x_3 is the pivot variable.

It follows step 2., we divide the pivot row (row 2) by the pivot element (by -3), and we mark the 1 in the position of the pivot element. Thus we have already three marked elements.

After this it follows step 3., the zeroing of the pivot column. First we subtract $\frac{-3}{-3} = 1$-times the second row from the first row, then we subtract $\frac{4}{-3} = -\frac{4}{3}$-times the second row from the third row. The result is

$$\begin{array}{ccc|c} 0 & \underline{1} & 0 & 4 \\ 0 & 0 & \underline{1} & 3 \\ \underline{1} & 0 & 0 & -1 \end{array} .$$

By step 4. the second row will be marked, the third column will be marked, the unknown x_3 will be marked. Now we have three marked elements, three marked rows, three marked columns, and three marked unknowns.

The third cycle is complete.

Now it follows again step 1. Let us choose a pivot element. We experience that all the rows are marked, it is impossible to choose a pivot element. The elimination process has stopped (it is terminated). We have three marked unknowns, consequently $r = 3$.

Since all the rows are marked (case $r = m$), then the system is solvable (consistent), and the reduced table is identical with the last table:

$$\begin{array}{ccc|c} 0 & \underline{1} & 0 & 4 \\ 0 & 0 & \underline{1} & 3 \\ \underline{1} & 0 & 0 & -1 \end{array} .$$

Now, let us compute the solutions. Since all the columns are marked (case $r = n$), then we have a unique solution, which can be read out from the last table in the following way:

By the first row we have: $x_2 = 4$.

By the second row we have: $x_3 = 3$.

By the third row we have: $x_1 = -1$.

We see, that if we move from top to bottom, then we obtain the values of the unknowns not in the natural order x_1, x_2, x_3. If we want to read out the values of the unknowns in natural order, then we have to rearrange the rows of the reduced table such that the part left to the vertical line will be the identity matrix:

$$\begin{array}{ccc|c} \underline{1} & 0 & 0 & -1 \\ 0 & \underline{1} & 0 & 4 \\ 0 & 0 & \underline{1} & 3 \end{array} .$$

If we disregard the explanations, the Gauss-Jordan elimination method means now the following sequence of tables:

$$
\begin{array}{ccc|c}
0 & \boxed{1} & -3 & -5 \\
4 & 5 & -2 & 10 \\
2 & 3 & -1 & 7 \\
\hline\hline
0 & \underline{1} & -3 & -5 \\
4 & 0 & 13 & 35 \\
\boxed{2} & 0 & 8 & 22 \\
\hline\hline
0 & \underline{1} & -3 & -5 \\
0 & 0 & \boxed{-3} & -9 \\
\underline{1} & 0 & 4 & 11 \\
\hline\hline
0 & \underline{1} & 0 & 4 \\
0 & 0 & \underline{1} & 3 \\
\underline{1} & 0 & 0 & -1 \\
\hline\hline
\underline{1} & 0 & 0 & -1 \\
0 & \underline{1} & 0 & 4 \\
0 & 0 & \underline{1} & 3
\end{array}
.
$$

The three marked unknowns x_1, x_2, x_3 are the bound variables. There is no free variable.

7.24. Remark. As it is clear from the example, we have to divide by the pivot element. By this reason if we compute on a paper, then we try to choose as pivot elements 1 or -1, to avoid fractions.

$\boxed{\text{c)}}$

We have three bound variables, thus the rank of the coefficient matrix is $\operatorname{rank}(A) = 3$.

$\boxed{\text{d)}}$

We have

$$x_1 = -1; \qquad x_2 = 4; \qquad x_3 = 3\,.$$

Let us arrange them into a vector. So we have the vector form solution:

$$x = (x_1,\, x_2,\, x_3) = (-1,\, 4,\, 3)\,.$$

$\boxed{\text{e)}}$
The solution set of the linear system is a one-element-set:

$$\mathcal{M} = \{(-1,\, 4,\, 3)\} \subset \mathbb{R}^3\,.$$

$\boxed{\text{f)}}$
The unique solution of the homogeneous system is:

$$x_1 = 0; \qquad x_2 = 0; \qquad x_3 = 0\,, \quad \text{in vector form:} \quad x = (0,\, 0,\, 0) \in \mathbb{R}^3\,,$$

its solution set is:

$$\mathcal{M}_h = \{(0,\, 0,\, 0)\} = \operatorname{Ker}(A) \subset \mathbb{R}^3\,.$$

This subspace has no basis, its dimension is $\dim \mathcal{M}_h = 0$, that is identical with the number of free variables.

Example 2 *(the case of infinitely many solutions)*

$$\begin{array}{rcrcrcrcrcl}
-3x_1 & + & x_2 & + & x_3 & - & x_4 & - & 2x_5 & = & 2 \\
2x_1 & - & x_2 & & & & & + & x_5 & = & 0 \\
-x_1 & + & x_2 & + & 2x_3 & + & x_4 & - & x_5 & = & 8 \\
 & & x_2 & + & x_3 & + & 2x_4 & & & = & 6 \\
\hline
\end{array}$$

Solution

$\boxed{\text{a)}}$
Obviously $m = 4$, $n = 5$, the size of our system is 4×5. Its coefficient matrix and its right-hand-side vector are:

$$A = \begin{bmatrix} -3 & 1 & 1 & -1 & -2 \\ 2 & -1 & 0 & 0 & 1 \\ -1 & 1 & 2 & 1 & -1 \\ 0 & 1 & 1 & 2 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 5}\,, \qquad b = \begin{pmatrix} 2 \\ 0 \\ 8 \\ 6 \end{pmatrix} \in \mathbb{R}^4\,.$$

$\boxed{\text{b)}}$

(The explanations are on the lectures, or you can apply the explanations of the first example.)

$$
\begin{array}{ccccc|c}
-3 & 1 & \boxed{1} & -1 & -2 & 2 \\
2 & -1 & 0 & 0 & 1 & 0 \\
-1 & 1 & 2 & 1 & -1 & 8 \\
0 & 1 & 1 & 2 & 0 & 6 \\
\hline\hline
-3 & 1 & \underline{1} & -1 & -2 & 2 \\
2 & -1 & 0 & 0 & \boxed{1} & 0 \\
5 & -1 & 0 & 3 & 3 & 4 \\
3 & 0 & 0 & 3 & 2 & 4 \\
\hline\hline
1 & -1 & \underline{1} & -1 & 0 & 2 \\
2 & -1 & 0 & 0 & \underline{1} & 0 \\
\boxed{-1} & 2 & 0 & 3 & 0 & 4 \\
-1 & 2 & 0 & 3 & 0 & 4 \\
\hline\hline
0 & 1 & \underline{1} & 2 & 0 & 6 \\
0 & 3 & 0 & 6 & \underline{1} & 8 \\
\underline{1} & -2 & 0 & -3 & 0 & -4 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}
.
$$

Now it follows again step 1, choosing a pivot element. We can see that although there is non-marked row, in each non-marked row we have 0-s to the left of the vertical line. By this reason, it is impossible to choose a pivot element. The elimination process has stopped (it is terminated). We have three marked unknowns, consequently $r = 3$.

Since the last elements of the non-marked rows are 0-s (only the fourth row is non-marked), then the system is solvable (consistent). The reduced table can be obtained by cleaning the fourth row:

$$
\begin{array}{ccccc|c}
0 & 1 & \underline{1} & 2 & 0 & 6 \\
0 & 3 & 0 & 6 & \underline{1} & 8 \\
\underline{1} & -2 & 0 & -3 & 0 & -4
\end{array}
.
$$

Now let us write the solution from the reduced table, by rearrangement:

By the first row we have: $x_3 = 6 - x_2 - 2x_4$.
By the second row we have: $x_5 = 8 - 3x_2 - 6x_4$.
By the third row we have: $x_1 = -4 + 2x_2 + 3x_4$.
Thus the general solution of the linear system is:

$$x_2 \in \mathbb{R}, \quad x_4 \in \mathbb{R}, \quad x_1 = -4+2x_2+3x_4, \quad x_3 = 6-x_2-2x_4, \quad x_5 = 8-3x_2-6x_4 \,.$$

The three marked unknowns x_1, x_3, x_5 are the bound variables. The two non-marked unknowns x_2, x_4 are the free variables.

$\boxed{\text{c)}}$

Since we have three bound variables, then the rank of the coefficient matrix is $\operatorname{rank}(A) = 3$.

$\boxed{\text{d)}}$

We have

$$x_1 = -4+2x_2+3x_4; \quad x_5 = 8-3x_2-6x_4; \quad x_3 = 6-x_2-2x_4 \quad (x_2,\, x_4 \in \mathbb{R}) \,.$$

Let us arrange them into a vector, and apply the separation technique. So we have the vector form solution:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -4 + 2x_2 + 3x_4 \\ x_2 \\ 6 - x_2 - 2x_4 \\ x_4 \\ 8 - 3x_2 - 6x_4 \end{pmatrix} = \begin{pmatrix} -4 \\ 0 \\ 6 \\ 0 \\ 8 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 2 \\ 1 \\ -1 \\ 0 \\ -3 \end{pmatrix} + x_4 \cdot \begin{pmatrix} 3 \\ 0 \\ -2 \\ 1 \\ -6 \end{pmatrix} .$$

It can be read out that:

$$x^B = (-4, 0, 6, 0, 8), \qquad v_2 = (2, 1, -1, 0, -3), \qquad v_4 = (3, 0, -2, 1, -6) \,.$$

$\boxed{\text{e)}}$

The solution set of the linear system is an infinite set:

$$\mathcal{M} = \{x^B + x_2 v_2 + x_4 v_4 \mid x_2, x_4 \in \mathbb{R}\} \,.$$

$\boxed{\text{f)}}$

The general solution of the homogeneous system is (consider 0-s to the right of the vertical line in the reduced table):

$$x_2 \in \mathbb{R}, \quad x_4 \in \mathbb{R}, \quad x_1 = 2x_2 + 3x_4, \quad x_3 = x_2 - 2x_4, \quad x_5 = 3x_2 - 6x_4 \,.$$

In vector form it is:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 2x_2 + 3x_4 \\ x_2 \\ -x_2 - 2x_4 \\ x_4 \\ -3x_2 - 6x_4 \end{pmatrix} = x_2 \cdot \begin{pmatrix} 2 \\ 1 \\ -1 \\ 0 \\ -3 \end{pmatrix} + x_4 \cdot \begin{pmatrix} 3 \\ 0 \\ -2 \\ 1 \\ -6 \end{pmatrix} = x_2 v_2 + x_4 v_4 \,.$$

the solution set is:

$$\mathcal{M}_h = \{x_2 v_2 + x_4 v_4 \mid x_2, x_4 \in \mathbb{R}\} = \operatorname{Span}(v_2,\, v_4) = \operatorname{Ker}(A) \,.$$

A basis of $\mathcal{M}_h$ is: $v_1,\ v_2$. Furthermore $\dim \mathcal{M}_h = 2$.

Example 3 *(the inconsistent case)*

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

Solution

$\boxed{\text{a)}}$

Obviously $m = 3$, $n = 4$, that is the size of the system is 3×4. Its coefficient matrix and right-hand side vector are:

$$A = \begin{bmatrix} 2 & 3 & -1 & 2 \\ 1 & 4 & -4 & 3 \\ 4 & 1 & 5 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 4}\,, \qquad b = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^3 \,.$$

$\boxed{\text{b)}}$

(The explanations are on the lectures, or you can apply the explanations of the first example.)

$$
\begin{array}{cccc|c}
2 & 3 & -1 & 2 & -1 \\
1 & 4 & -4 & 3 & 2 \\
4 & \boxed{1} & 5 & 0 & 1 \\
\hline\hline
-10 & 0 & -16 & \boxed{2} & -4 \\
-15 & 0 & -24 & 3 & -2 \\
4 & \underline{1} & 5 & 0 & 1 \\
\hline\hline
-5 & 0 & -8 & \underline{1} & -2 \\
0 & 0 & 0 & 0 & 4 \\
4 & \underline{1} & 5 & 0 & 1
\end{array}
$$

Now it follows again step 1, choosing a pivot element. We can see that although there is non-marked row, in each non-marked row stand 0-s to the left of the vertical line. By this reason, it is impossible to choose a pivot element. The elimination process has stopped (it is terminated). We have two marked unknowns, consequently $r = 2$.

Since we have such a non-marked row whose last element is nonzero (row 2, it is a prohibited row), then the linear system is unsolvable (inconsistent, antinomic).

$\boxed{\text{c)}}$

We have two marked unknowns, consequently $\operatorname{rank}(A) = 2$.

$\boxed{\text{d)}}$

There is no solution, consequently the vector form solution also does not exist.

$\boxed{\text{e)}}$

The solution set is the empty set: $\mathcal{M} = \emptyset$.

$\boxed{\text{f)}}$

The general solution of the homogeneous system is (consider 0-s to the right of the vertical line in the reduced table):

$$x_1 \in \mathbb{R}, \quad x_3 \in \mathbb{R}, \quad x_2 = -4x_1 - 5x_3, \quad x_4 = 5x_1 + 8x_3 .$$

The two marked unknowns x_2, x_4 are the bound variables, and the two non-marked unknowns x_1, x_3 are the free variables.

The vector form of the solution of the homogeneous system is:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ -4x_1 - 5x_3 \\ x_3 \\ 5x_1 + 8x_3 \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ -4 \\ 0 \\ 5 \end{pmatrix} + x_3 \cdot \begin{pmatrix} 0 \\ -5 \\ 1 \\ 8 \end{pmatrix} .$$

The solution set of the homogeneous system is:

$$\mathcal{M}_h = \{x_1 v_1 + x_3 v_3 \mid x_1, x_3 \in \mathbb{R}\} = \operatorname{Span}(v_1,\ v_3) = \operatorname{Ker}(A) .$$

A basis of $\mathcal{M}_h$ is: $v_1 = (1, -4, 0, 5), \quad v_3 = (0, -5, 1, 8)$. Furthermore $\dim \mathcal{M}_h = 2$.

7.1.6. Control Questions to the Theory

1. Define the rank of a vector system.

2. State the theorem about the connection between the dimensions of the row space and of the column space.

3. Define the rank of a matrix.

4. Write the scalar form and the vector form of a linear system.

5. Define the following sets: $\mathcal{M}$, $\mathcal{M}_h$, $\operatorname{Ker}(A)$.

6. State the theorem about the structure of the solution set of the homogeneous linear system.

7. State the theorem about the structure of the solution set of the general linear system.

7.2. Exercises

7.2.1. Exercises for Class Work

1. Consider the following systems of linear equations:

(a)

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

(b)

$$\begin{array}{rcrcrcrcrcr} -3x_1 & + & x_2 & + & x_3 & - & x_4 & - & 2x_5 & = & 2 \\ 2x_1 & - & x_2 & & & & & + & x_5 & = & 0 \\ -x_1 & + & x_2 & + & 2x_3 & + & x_4 & - & x_5 & = & 8 \\ & & x_2 & + & x_3 & + & 2x_4 & & & = & 6 \end{array}$$

(c)

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

Answer the following questions for each of them:

a) Give the size, the coefficient matrix and its size, and also the right hand side vector b of the system.

b) Solve the system of linear equations (it is consistent or not, find all solutions, give the solutions in scalar form, bound variables, free variables).

c) Find the rank of the coefficient matrix.

d) If the system is consistent, give the vector form of the solutions too.

e) If the system is consistent, give the solution set $\mathcal{M}$.

f) Give the solution set ($\mathcal{M}_h = \operatorname{Ker}(A)$) of the associated homogeneous system. Determine a basis in $\mathcal{M}_h$. Find the dimension of the subspace $\mathcal{M}_h$.

2. Let

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 3 & 1 & 1 \\ 1 & 0 & 8 & 2 \end{bmatrix} \in \mathbb{R}^{3\times 4}.$$

Determine a basis in the subspace

$$\mathrm{Ker}\,(A) := \{x \in \mathbb{R}^4 \mid Ax = 0\}\,.$$

Find the dimension of $\mathrm{Ker}\,(A)$.

7.2.2. Additional Tasks

1. Consider the following systems of linear equations:

$$a)\quad \begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & 3x_3 & = & 6 \\ 2x_1 & - & x_2 & + & 4x_3 & = & 1 \\ x_1 & - & x_2 & + & x_3 & = & 3 \\ \hline \end{array} \qquad\qquad b)\quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 5 \\ x_1 & & & + & x_3 & = & -2 \\ 2x_1 & + & x_2 & + & 3x_3 & = & 3 \\ \hline \end{array}$$

Answer the following questions for each of them:

a) Give the size, the coefficient matrix and its size, and also the right hand side vector b of the system.

b) Solve the system of linear equations (it is consistent or not, find all solutions, give the solutions in scalar form, bound variables, free variables).

c) Find the rank of the coefficient matrix.

d) If the system is consistent, give the vector form of the solutions too.

e) If the system is consistent, give the solution set $\mathcal{M}$.

f) Give the solution set ($\mathcal{M}_h = \mathrm{Ker}\,(A)$) of the associated homogeneous system. Determine a basis in $\mathcal{M}_h$. Find the dimension of the subspace $\mathcal{M}_h$.

2. Let

$$A = \begin{bmatrix} 3 & 1 & 9 \\ 1 & 2 & -2 \\ 2 & 1 & 5 \end{bmatrix}.$$

Determine a basis in the subspace

$$\mathrm{Ker}\,(A) := \{x \in \mathbb{R}^3 \mid Ax = 0\}$$

Find the dimension of $\mathrm{Ker}\,(A)$.

3. Determine the ranks of the following matrices:

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & 0 & -2 \\ 0 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 3 & -1 & -2 \\ 2 & 4 & 3 & 5 \\ -1 & 1 & 3 & 4 \end{bmatrix}$$

8. Connection with the Inverse Matrix

8.1. Theory

8.1.1. Linear System with Square Matrix

8.1. Theorem (Linear system with square matrix) *Consider the linear system $Ax = b$ given by the square matrix $A \in \mathbb{K}^{n\times n}$ and by the vector $b \in \mathbb{K}^n$. Then*

a) In the case $\operatorname{rank}(A) = n$ *the linear system has a unique solution-*

b) In the case $\operatorname{rank}(A) \leq n-1$ *the linear system either has no solution or the linear system has infinitely many solutions.*

Proof.
a) Suppose that $r = \operatorname{rank}(A) = n$.

Then the column vectors of A form an n-term linearly independent system in the n dimensional space $\mathbb{K}^n$. Consequently, they form a basis in $\mathbb{K}^n$, thus $\mathrm{Col}(A) = \mathbb{K}^n$. By this reason $b \in \mathrm{Col}(A)$, that is the linear system has a solution.

On the other hand, the degree of freedom is $n - r = n - n = 0$, thus the solution is unique.

b) Suppose that $r = \operatorname{rank}(A) \leq n-1$.

In this case by $\dim \mathrm{Col}(A) = r < n = \dim \mathbb{K}^n$ we have

$$\mathrm{Col}(A) \subset \mathbb{K}^n \qquad \mathrm{Col}(A) \neq \mathbb{K}^n\,.$$

If so $b \notin \mathrm{Col}(A)$, then the system has no solution. However, if $b \in \mathrm{Col}(A)$, then the system is solvable, and the degree of freedom is

$$n - r \geq n - (n-1) = 1\,,$$

consequently the system has infinitely many solutions. $\square$

8.1.2. Inverse Matrix and the Linear System

We can use our results about the linear systems with square matrix also for determination of the inverse matrix.

8.2. Theorem *Let $A \in \mathbb{K}^{n\times n}$ be a square matrix. Then*

a) $\operatorname{rank} A = n \quad \Longrightarrow \quad$ *A the matrix A is invertible (regular);*

b) $\operatorname{rank} A < n \quad \Longrightarrow \quad$ *A the matrix A is non-invertible (singular).*

Proof. Denote by I the $n \times n$ identity matrix. Its columns are the standard unit vectors:

$$I = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} .$$

We are looking for the inverse of A, that is, we are looking for the matrix

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \in \mathbb{K}^{n\times n}$$

such that $AX = I$ is true.

The matrix equation $AX = I$ can be written as follows:

$$A \cdot \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} ,$$

Which is equivalent the collection of the following linear systems:

$$Ax_1 = e_1, \quad Ax_2 = e_2, \quad \dots, \quad Ax_n = e_n . \tag{8.1}$$

Now we go to the proofs of the statements a) and b):

a) Since $r = n$ at each linear system, then – using the previous theorem – each linear system can be solved uniquely. It implies, that there exists A^{-1}, and that the columns of A^{-1} are the solution vectors $x_1, \dots, x_n$.

b) Since $\dim \operatorname{Col}(A) = r < n$, then all the standard unit vectors $e_1, \dots, e_n$ cannot be in $\operatorname{Col}(A)$. Thus – using the previous theorem – at least one of the above collection of linear systems is antinomic, it has no solution. Consequently A^{-1} does not exist. □

8.3. Remark. It follows from the theorem, that its parts a) and b) are actually equivalences

After all – considering also the connection between the determinants and the inverses – we can characterize the regular and the singular matrices as follows:

The 5 characterizations of a regular matrix $A \in \mathbb{K}^{n\times n}$:

1. $\exists A^{-1}$

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

3. $\operatorname{rank}(A) = n$

4. the columns of A are linearly independent

5. the rows of A are linearly independent

The 5 characterizations of a singular matrix $A \in \mathbb{K}^{n\times n}$:

1. $\nexists A^{-1}$

2. $\det(A) = 0$

3. $\operatorname{rank}(A) < n$

4. the columns of A are linearly dependent

5. the rows of A are linearly dependent

8.1.3. Computation of the Inverse Matrix with Gauss-Jordan Method

By the results of the previous section we can establish that to determine the inverse of an $n \times n$ matrix we have to solve n linear systems. From the point of view of the amount of arithmetic operations This method is more effective than the method of the cofactors (see in the chapter about the determinants). Thus we can compute the inverse matrix solving this n linear systems after each other.

But the coefficient matrices of these linear systems are the same, so we obtain a more effective method, if we solve the n linear systems not after each other, but simultaneously at the same time. The Gauss-Jordan method makes it possible. We need only the modification that in the starting table we put – by (8.1) – the n standard unit vectors, to the right of the vertical line. Then we perform the learnt elimination cycles with this $n \times 2n$ augmented table.

– If you don't have the n marked element at the stopping of the elimination cycles, then the matrix is singular, it has no inverse.

– If you have the n marked element at the stopping of the elimination cycles, then the matrix is regular, it has inverse. To read out the inverse, we have to rearrange the rows of the last table (it is the reduced table at the same time), such that the block to the left of the vertical line will be the identity matrix. Then the inverse matrix will be the block to the right of the vertical line.

The rank of the matrix will be in both cases the number of the marked elements.

Let us look at these two developed examples:

1. Example Using Gauss-Jordan method determine the inverse of the matrix

$$A = \begin{bmatrix} 5 & 2 & -3 \\ 3 & 1 & -2 \\ 2 & -3 & -4 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

Solution

$$\begin{array}{ccc|ccc}
5 & 2 & -3 & 1 & 0 & 0 \\
3 & \boxed{1} & -2 & 0 & 1 & 0 \\
2 & -3 & -4 & 0 & 0 & 1 \\
\hline\hline
-1 & 0 & \boxed{1} & 1 & -2 & 0 \\
3 & \underline{1} & -2 & 0 & 1 & 0 \\
11 & 0 & -10 & 0 & 3 & 1 \\
\hline\hline
-1 & 0 & \underline{1} & 1 & -2 & 0 \\
1 & \underline{1} & 0 & 2 & -3 & 0 \\
\boxed{1} & 0 & 0 & 10 & -17 & 1 \\
\hline\hline
0 & 0 & \underline{1} & 11 & -19 & 1 \\
0 & \underline{1} & 0 & -8 & 14 & -1 \\
\underline{1} & 0 & 0 & 10 & -17 & 1 \\
\hline\hline
\underline{1} & 0 & 0 & 10 & -17 & 1 \\
0 & \underline{1} & 0 & -8 & 14 & -1 \\
0 & 0 & \underline{1} & 11 & -19 & 1
\end{array}$$

We can read out here that the matrix is regular, its inverse is

$$A^{-1} = \begin{bmatrix} 10 & -17 & 1 \\ -8 & 14 & -1 \\ 11 & -19 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

Since there are three marked elements, the rank of the matrix equals 3.

2. Example Using Gauss-Jordan method determine the inverse of the matrix

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 3 \\ -3 & 1 & -7 \end{bmatrix} \in \mathbb{R}^{3\times 3}\,.$$

Solution

$$\begin{array}{ccc|ccc}
1 & 1 & -1 & 1 & 0 & 0 \\
2 & 0 & 3 & 0 & 1 & 0 \\
-3 & \boxed{1} & -7 & 0 & 0 & 1 \\
\hline\hline
4 & 0 & 6 & 1 & 0 & -1 \\
\boxed{2} & 0 & 3 & 0 & 1 & 0 \\
-3 & \underline{1} & -7 & 0 & 0 & 1 \\
\hline\hline
0 & 0 & 0 & 1 & -2 & -1 \\
\underline{1} & 0 & 3/2 & 0 & 1/2 & 0 \\
0 & \underline{1} & -5/2 & 0 & 3/2 & 1
\end{array}$$

The elimination process has stopped, because it was impossible to find pivot element. Since the number of the marked elements is less than 3, then the matrix is non-invertible, it is singular.

Since there are two marked elements, the rank of the matrix equals 2.

You find an example for the inverse of a 4×4 matrix in the Appendix.

8.1.4. Control Questions to the Theory

1. Stte the theorem about the linear systems with square matrices.
2. State the theorem about the connection between the rank and the invertibility of a matrix.
3. Write the 5 equivalent characterizations of the regular matrices.
4. Write the 5 equivalent characterizations of the singular matrices.

8.2. Exercises

8.2.1. Exercises for Class Work

1. Solve the following system of linear equations (the exercise was solved on the previous practice). Is the coefficient matrix regular or singular? Find the rank of the coefficient matrix.

$$\begin{array}{rcr} x_2 - 3x_3 &=& -5 \\ 4x_1 + 5x_2 - 2x_3 &=& 10 \\ 2x_1 + 3x_2 - x_3 &=& 7 \\ \hline \end{array}$$

2. Using Gauss-Jordan method determine the inverses of the matrices

$$a) \quad A = \begin{bmatrix} 5 & 2 & -3 \\ 3 & 1 & -2 \\ 2 & -3 & -4 \end{bmatrix} \in \mathbb{R}^{3\times 3} \qquad b) \quad A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 3 \\ -3 & 1 & -7 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

3. Find the rank of the following matrix. Is the coefficient matrix regular or singular?

$$A = \begin{bmatrix} 5 & 1 & -7 & -2 \\ 0 & 2 & 1 & 1 \\ 1 & 5 & 1 & 2 \\ -3 & -1 & 4 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}$$

4. Consider the matrix of the previous exercise:

$$A = \begin{bmatrix} 5 & 1 & -7 & -2 \\ 0 & 2 & 1 & 1 \\ 1 & 5 & 1 & 2 \\ -3 & -1 & 4 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4},$$

and let

$$b_1 = (-1, 3, 7, 0), \ b_2 = (0, 5, 7, -1) \in \mathbb{R}^4 .$$

Prove that
a) the linear system $Ax = b_1$ has infinitely many solutions,
b) the linear system $Ay = b_2$ has no solution.

8.2.2. Additional Tasks

1. Solve the following system of linear equations. Is the coefficient matrix regular or singular? Find the rank of the coefficient matrix.

$$a)\quad \begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & 3x_3 & = & 6 \\ 2x_1 & - & x_2 & + & 4x_3 & = & 1 \\ x_1 & - & x_2 & + & x_3 & = & 3 \\ \hline \end{array} \qquad\qquad b)\quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 5 \\ x_1 & & & + & x_3 & = & -2 \\ 2x_1 & + & x_2 & + & 3x_3 & = & 3 \\ \hline \end{array}$$

2. Using Gauss-Jordan method determine the inverses of the matrices

$$a)\quad A = \begin{bmatrix} 4 & 3 & 1 \\ -3 & -5 & -2 \\ 9 & 4 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3} \qquad\qquad b)\quad A = \begin{bmatrix} 2 & 1 & 4 \\ 3 & 2 & 5 \\ 0 & -1 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

3. Consider the matrix

$$A = \begin{bmatrix} 1 & 3 & 4 \\ -2 & -5 & -3 \\ 1 & 4 & 9 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

(a) Determine its rank. Is the matrix A regular or singular?

(b) Determine the vectors $b_1, b_2 \in \mathbb{R}^3 \setminus \{0\}$ such that the system $Ax = b_1$ is consistent and the system $Ax = b_2$ is inconsistent.

9. Eigenvalues and Eigenvectors of Matrices

In this chapter we investigate that in the case of a given matrix $A \in \mathbb{K}^{n\times n}$ in which directions in $\mathbb{K}^n$ will be the result of the multiplication by A parallel with the original direction. These directions (if they exist) will be called eigen-directions.

The question raised above is in close connection with the linear transformations, therefore let us speak some words about the linear transformations.

9.1. Theory

9.1.1. Linear Transformations in $\mathbb{K}^n$

9.1. Definition A function $\varphi : \mathbb{K}^n \to \mathbb{K}^n$ is called a linear transformation of the space $\mathbb{K}^n$ if

a) $\varphi(x + y) = \varphi(x) + \varphi(y) \qquad (x, y \in \mathbb{K}^n)$, and

b) $\varphi(\lambda x) = \lambda\varphi(x) \qquad (x \in \mathbb{K}^n,\ \lambda \in \mathbb{K})$.

A linear transformation in $\mathbb{R}^2$ is for example the line reflection about the x-axis. Also a linear transformation in $\mathbb{R}^2$ is the rotation about the origin through $+90°$.

If $A \in \mathbb{K}^{n\times n}$ then the function

$$\varphi : \mathbb{K}^n \to \mathbb{K}^n, \qquad \varphi(x) = Ax$$

is a linear transformation in $\mathbb{K}^n$. It can be proved that to each linear transformation in $\mathbb{K}^n$ there exist a unique matrix $A \in \mathbb{K}^{n\times n}$ such that $\varphi(x) = Ax$ $(x \in \mathbb{K}^n)$. Thus any linear transformation of $\mathbb{K}^n$ can be characterized by an $n \times n$ matrix in $\mathbb{K}^{n\times n}$. This matrix is called the matrix of the linear transformation.

9.2. Examples

1. The matrix of the above mentioned line reflection about the x-axis is:

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

2. The matrix of the above mentioned rotation about the origin through $+90°$ is:

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

An interesting question is, that a linear transformation in which directions behaves as central dilation, that is for which vectors $x \in \mathbb{K}^n \setminus \{0\}$ and numbers $\lambda \in \mathbb{K}$ holds

$$\varphi(x) = \lambda x \qquad \text{(the formulation of the eigenvalue problem with transformations)}\,.$$

Using the matrix of the transformation we have

$$Ax = \lambda x \qquad \text{(the formulation of the eigenvalue problem with matrices)}\,.$$

We will investigate this last question, in the formulation with matrices.

9.1.2. Basic Concepts

9.3. Definition Let $A \in \mathbb{K}^{n\times n}$ and $\lambda \in \mathbb{K}$. The number λ is called the eigenvalue of the matrix A if

$$\exists\, x \in \mathbb{K}^n,\, x \neq 0 : \quad Ax = \lambda x\,.$$

The above vector $x \in \mathbb{K}^n \setminus \{0\}$ is called an eigenvector of A associated with the eigenvalue λ. The equation $Ax = \lambda x$ is called the eigenvalue-equation.

The set of the eigenvalues is called the spectrum of the matrix A, its notation is: $\mathrm{Sp}\,(A)$. So

$$\mathrm{Sp}\,(A) := \{\lambda \in \mathbb{K} \mid \exists\, x \in \mathbb{K}^n \setminus \{0\} : \; Ax = \lambda x\} \subseteq \mathbb{K}\,.$$

After a simple rearrangement one can see, that the equation $Ax = \lambda x$ is equivalent with the following homogeneous system of linear equations for any fixed $\lambda \in \mathbb{K}$:

$$(A - \lambda I)x = 0\,, \tag{9.1}$$

where I denotes the identity matrix in $\mathbb{K}^{n\times n}$.

Using the theory of systems of linear equations hence follows that the number $\lambda \in \mathbb{K}$ is an eigenvalue of A if the above system of linear equations has infinitely many solutions. But this last statement – using the theory of square linear systems – is equivalent with the fact that the determinant of its coefficient matrix $A - \lambda I$ equals 0. So the parameter λ must satisfy the equation

$$\det(A - \lambda I) = 0\,.$$

The left-hand side of this equation is a polynomial of the variable λ, because at the expansion of the determinant we use only addition and multiplication. The roots in $\mathbb{K}$ of this polynomial are the eigenvalues. The eigenvectors associated with a fixed eigenvalue are the nontrivial solutions of the homogeneous system of linear equations (9.1).

9.4. Definition The polynomial

$$P(\lambda) = P_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} \qquad (\lambda \in \mathbb{K})$$

is called the characteristic polynomial of the matrix A. The equation $P_A(\lambda) = 0$ is called characteristic equation.

9.5. Remark. The expansion of the above determinant shows that the characteristic polynomial has the degree n, and the coefficient of λ^n (the leading coefficient) is $(-1)^n$. Moreover – since $P(0) = \det(A - 0I) = \det(A)$ – it follows that its constant term equals $\det(A)$. Thus the characteristic polynomial has the following form:

$$P(\lambda) = (-1)^n \cdot \lambda^n + \ldots + \det(A) \qquad (\lambda \in \mathbb{K})\,.$$

In the „Precalculus” part of the subject „Basic Mathematics” we have learnt about the multiplicity of a real root of a polynomial. Similarly we can define the multiplicity of complex roots. Using this concepts we can give the following definition:

9.6. Definition Let P be the characteristic polynomial of a matrix $A \in \mathbb{K}^{n\times n}$, and let $\lambda \in \mathbb{K}$ be an eigenvalue of A (that is a root of P). The multiplicity of the root λ is called the algebraic multiplicity of the eigenvalue λ, and is denoted by $a(\lambda)$.

Since the eigenvalues are the roots of the characteristic polynomial in $\mathbb{K}$, we can establish that:

- If $\mathbb{K} = \mathbb{C}$, then $\mathrm{Sp}\,(A) \neq \emptyset$, and has maximally n elements. If every eigenvalue is counted as many times as its algebraic multiplicity, then the number of eigenvalues is exactly n.

- If $\mathbb{K} = \mathbb{R}$, then $\mathrm{Sp}\,(A)$ may be the empty set, and it has maximally n elements. But even though every eigenvalue is counted as many times as its algebraic multiplicity, it is not sure, that the number of eigenvalues is exactly n. In the case when – counting the eigenvalues with their algebraic multiplicity – the number of eigenvalues is exactly n, then we say that all the eigenvalues of the matrix are real.

The eigenvalues of triangular matrices (especially of diagonal matrices) can be obtained easily, as it can be seen in the following remark.

9.7. Remark. Let $A \in \mathbb{K}^{n\times n}$ be a (lower or upper) triangular matrix. Then – e.g. in the lower triangular case – its characteristic polynomial is as follows:

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & 0 & \dots & 0 \\ a_{21} & a_{22}-\lambda & \dots & 0 \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} =$$

$$= (a_{11}-\lambda)\cdot(a_{22}-\lambda)\cdot\dots\cdot(a_{nn}-\lambda) \qquad (\lambda \in \mathbb{K})\,.$$

We have used, that the determinant of a triangular matrix equals the product of its diagonal entries.

Consequently, the eigenvalues of a triangular matrix are its diagonal entries. The algebraic multiplicity of each eigenvalue is the number of how many times it occurs in the diagonal.

Let us move on to the investigation of eigenvectors. First we will show, that infinitely many eigenvectors are associated with a fixed eigenvalue. Moreover, these infinitely many eigenvectors and the zero vector together form a subspace in $\mathbb{K}^n$.

9.8. Theorem *Let $A \in \mathbb{K}^{n\times n}$ and $\lambda \in \mathrm{Sp}\,(A)$. Then the set*

$$W_\lambda := W_\lambda(A) := \{x \in \mathbb{K}^n \mid Ax = \lambda x\}$$

consisting of the eigenvectors associated with λ and of the zero vector forms an $n - \mathrm{rank}\,(A - \lambda I)$ dimensional subspace in $\mathbb{K}^n$. Infinitely many eigenvectors are associated with the eigenvalue λ.

Proof.

$$W_\lambda = \{x \in \mathbb{K}^n \mid Ax = \lambda x\} = \{x \in \mathbb{K}^n \mid (A - \lambda I)x = 0\} = \mathcal{S}_h$$

Using the learned results about the homogeneous linear systems, the above set is really a subspace whose dimension is

$$\dim W_\lambda = \dim \mathcal{S}_h = n - \operatorname{rank}(A - \lambda I)\,.$$

Since $\dim W_\lambda = n - \operatorname{rank}(A - \lambda I) \geq 1$, then the set of the associated eigenvectors ($W_\lambda \setminus \{0\}$) is really infinite. □

At a fixed eigenvalue the real question is not the number of the associated eigenvectors, but the maximal number of *independent* associated eigenvectors, that is the dimension of the subspace W_λ.

9.9. Definition The subspace

$$W_\lambda := W_\lambda(A) := \{x \in \mathbb{K}^n \mid Ax = \lambda x\}$$

defined in the above theorem is called the eigenspace corresponding to the eigenvalue λ. The dimension of W_λ is called the geometric multiplicity of the eigenvalue λ. It is denoted by $g(\lambda)$. Thus we have $g(\lambda) = n - \operatorname{rank}(A - \lambda I)$.

The geometric multiplicity cannot exceed the algebraic one. This is expressed in the following theorem, which is given here without proof.

9.10. Theorem

$$\forall\, \lambda \in \operatorname{Sp}(A): \quad 1 \leq g(\lambda) \leq a(\lambda) \leq n\,.$$

9.1.3. Eigenvector Basis (E.B.)

The following theorem will be stated without proof. Essentially it states, that eigenvectors associated with different eigenvalues are linearly independent.

9.11. Theorem *Let $A \in \mathbb{K}^{n\times n}$, and $\lambda_1, \ldots, \lambda_k$ be k different eigenvalues of matrix A. Further let $s_i \in \mathbb{N}^+$, $1 \leq s_i \leq g(\lambda_i)$, and $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(s_i)}$ be a linearly independent system in the eigenspace W_{λ_i} ($i = 1, \ldots, k$). Then the united vector system*

$$x_i^{(j)} \in \mathbb{K}^n \qquad (i = 1, \ldots, k;\; j = 1, \ldots, s_i) \tag{9.2}$$

is linearly independent.

For each $\lambda \in \mathrm{Sp}\,(A)$ let us take the maximal number (that is $g(\lambda)$) linearly independent eigenvectors in the eigenspace W_λ. Their united system – in the sense of the above theory – is linearly independent, and the number of members in the united system equals $\sum\limits_{\lambda\in\mathrm{Sp}\,(A)} g(\lambda)$. Thus we can establish

$$\sum_{\lambda\in\mathrm{Sp}\,(A)} g(\lambda) \leq n\,.$$

In the case when equality holds in the above inequality, we have n independent vectors in the n dimensional space $\mathbb{K}^n$. Consequently this united vector system is a basis in $\mathbb{K}^n$.

9.12. Definition The basis described above is called Eigenvector Basis (E.B.)

We can easily conclude now that

$$\exists\ \text{E.B.} \quad \Longleftrightarrow \quad \sum_{\lambda\in\mathrm{Sp}\,(A)} g(\lambda) = n\,.$$

Using the connection between the algebraic and geometric multiplicities (see theorem 9.10) the following theorem can be proved:

9.13. Theorem *Let $A \in \mathbb{K}^{n\times n}$, $a(\lambda)$ be the algebraic multiplicity, and $g(\lambda)$ be the geometric multiplicity of the eigenvalue λ. Then*

$$\exists\ \textit{E.B.} \quad \Longleftrightarrow \quad \sum_{\lambda\in\mathrm{Sp}\,(A)} a(\lambda) = n \quad \textit{and} \quad \forall\lambda \in \mathrm{Sp}\,(A)\ : \quad g(\lambda) = a(\lambda)\,.$$

9.14. Remark. The condition $\sum\limits_{\lambda\in\mathrm{Sp}\,(A)} a(\lambda) = n$ expresses that the characteristic polynomial – counting the roots with their multiplicity – has n roots in $\mathbb{K}$. This condition

- in case $\mathbb{K} = \mathbb{C}$ is met „automatically”.
- in case $\mathbb{K} = \mathbb{R}$ is met if and only if all the roots of the characteristic polynomial are real.

9.1.4. Control Questions to the Theory

1. Define the eigenvalue and the eigenvector of a matrix

2. Define the characteristic polynomial

3. Define the algebraic multiplicity of an eigenvalue

4. Give the eigenvalues of a triangular matrix

5. Define the eigenspace

6. Define the geometric multiplicity of an eigenvalue

7. What is the connection between the algebraic and the geometric multiplicity of an eigenvalue?

8. State the theorem about the independence of eigenvectors

9. Define the concept of Eigenvector Basis (E. B.)

10. Give the theorems about the necessary and sufficient condition of the existence of Eigenvector Basis (2 theorems)

9.2. Exercises

9.2.1. Exercises for Class Work

1. Determine the eigenvalues and the eigenvectors of the following matrices. Determine the eigenspaces, the algebraic and the geometric multiplicities of the eigenvalues. Is there an eigenvector basis in the adequate vector space?

Solve the above problems in the case $\mathbb{K} = \mathbb{R}$ and in the case $\mathbb{K} = \mathbb{C}$ too.

a) $A = \begin{bmatrix} 2 & -1 & -1 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix}$ *b)* $A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

c) $A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 2 & -1 & 0 \end{bmatrix}$ *d)* $A = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$

9.2.2. Additional Tasks:

1. Determine the eigenvalues and the eigenvectors of the following matrices. Determine the eigenspaces, the algebraic and the geometric multiplicities of the eigenvalues. Is there an eigenvector basis in the adequate vector space?

Solve the above problems in the case $\mathbb{K} = \mathbb{R}$ and in the case $\mathbb{K} = \mathbb{C}$ too.

a) $\begin{bmatrix} 2 & -1 \\ 10 & -9 \end{bmatrix}$ *b)* $\begin{bmatrix} -2 & -7 \\ 1 & 2 \end{bmatrix}$ *c)* $\begin{bmatrix} 5 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

d) $A = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$ *e)* $\begin{bmatrix} 1 & 2 & -2 \\ -3 & 4 & 0 \\ -3 & 1 & 3 \end{bmatrix}$

10. Diagonalization of Matrices

10.1. Theory

10.1.1. Similarity of Matrices

10.1. Definition Let $A, B \in \mathbb{K}^{n\times n}$. We say that matrix B is similar to matrix A (in notation: $A \sim B$), if

$$\exists\, C \in \mathbb{K}^{n\times n}: \quad C \text{ invertible, and } B = C^{-1}AC\,.$$

Matrix C is called similarity matrix.

10.2. Remark. The similarity relation is symmetric, that is $A \sim B \Rightarrow B \sim A$. So we can refer to similarity as „A and B are similar (to each other)".

10.3. Theorem *If $A \sim B$, then $P_A = P_B$, that is, their characteristic polynomials coincide. Consequently their eigenvalues (with algebraic multiplicities) and their determinants coincide too.*

Proof. Let $A, B, C \in \mathbb{K}^{n\times n}$, and suppose $B = C^{-1}AC$. Then for any $\lambda \in \mathbb{K}$ holds

$$\begin{aligned} P_B(\lambda) = \det(B-\lambda I) &= \det(C^{-1}AC - \lambda C^{-1}IC) = \det(C^{-1}(A-\lambda I)C) = \\ &= \det(C^{-1}) \cdot \det(A-\lambda I) \cdot \det(C) = \det(C^{-1}) \cdot \det(C) \cdot \det(A-\lambda I) = \\ &= \det(C^{-1}C) \cdot \det(A-\lambda I) = \det(I) \cdot P_A(\lambda) = 1 \cdot P_A(\lambda) = P_A(\lambda)\,. \end{aligned}$$

This exactly means that $P_A = P_B$. □

10.1.2. Diagonalizability

In the following definition we define an important class of matrices.

10.4. Definition Let $A \in \mathbb{K}^{n\times n}$. We say that A is diagonalizable (sometimes the word „diagonable" is used) over $\mathbb{K}$, if it is similar to a diagonal matrix, that is, if

$$\exists\, C \in \mathbb{K}^{n\times n},\ C \text{ is invertible}: \quad C^{-1}AC \text{ is a diagonal matrix}\,.$$

Matrix C is called diagonalizing similarity matrix, and the diagonal matrix $D = C^{-1}AC$ is called the diagonal form of A.

10.5. Remark. If A is diagonalizable, then the diagonal entries of its diagonal form are the eigenvalues of A. Each eigenvalue stands in the diagonal as many times as its algebraic multiplicity is.

The essentiality of the following theorem is that the diagonalizability is equivalent with the existence of the Eigenvector Basis (E.B.)

10.6. Theorem *Let $A \in \mathbb{K}^{n\times n}$. The matrix A is diagonalizable over $\mathbb{K}$ if and only if there exists E.B. in $\mathbb{K}^n$.*

Proof. First suppose that A is diagonalizable. Let $c_1, \ldots, c_n \in \mathbb{K}^n$ be the column vectors of the diagonalizing similarity matrix C:

$$C = [c_1 \ \ldots \ c_n] \ .$$

We will show that $c_1, \ldots, c_n$ is an E.B.

By the invertibility of C the n-term system $c_1, \ldots, c_n$ is linearly independent, consequently it is a basis in $\mathbb{K}^n$.

To prove that the vectors c_j are eigenvectors, let us start from equation

$$C^{-1}AC = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Multiply this equation by C from the left:

$$A \cdot [c_1 \ldots c_n] = C \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [c_1 \ldots c_n] \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$[Ac_1 \ldots Ac_n] = [\lambda_1 c_1 \ldots \lambda_n c_n]$$

Using the column-wise equality we have

$$Ac_j = \lambda_j c_j \qquad (j = 1, \ldots, n)$$

So the basis really consists of eigenvectors.

Conversely, suppose $c_1, \ldots, c_n$ is an E.B. in $\mathbb{K}^n$. Let $C \in \mathbb{K}^{n\times n}$ be the matrix built from the columns $c_1, \ \ldots, \ c_n$.

C is obviously invertible, because its columns are linearly independent.

Let us write the eigenvalue-equations:

$$Ac_j = \lambda_j c_j \qquad (j = 1, \dots, n)\,,$$

then perform the steps written in the first half of the proof in the opposite direction. Then we have the following equality:

$$C^{-1}AC = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

We can see that A is really diagonalizable. □

10.7. Remarks.

1. One can see that the order of the eigenvector-columns in C is the same as the order of the corresponding eigenvalues in the diagonal of $C^{-1}AC$.

2. If a matrix $A \in \mathbb{K}^{n\times n}$ has n pairwise distinct eigenvalues in $\mathbb{K}$, then the corresponding n eigenvectors are linearly independent. Thus they form an E.B., consequently A is diagonalizable.

Considering our theorems, we suggest the following algorithm for discussing the diagonalizability of a matrix $A \in \mathbb{K}^{n\times n}$:

Step1.: Determine the eigenvalues with their algebraic multiplicities (solve the characteristic equation).

Step 2.: If the sum of the algebraic multiplicities is less than n, then stop, the matrix is non-digonalizable. Otherwise go to Step 3.

Step 3.: (The sum of the algebraic multiplicities equals n.) Take an eigenvalue and determine its geometric multiplicity.

Step 3A.: If the geometric multiplicity is less than the algebraic one, then stop, the matrix is non-diagonalizable. Otherwise go to Step 3B.

Step 3B.: If the geometric multiplicity equals the algebraic one, then determine the associated eigenvectors.

Repeat this process for all eigenvalues. If you performed Step 3B only, then continue with Step 4.

Step 4.: (The algebraic and the geometric multiplicities are equal for all eigenvalues.) The matrix is diagonalizable. The diagonalizing similarity matrix and the diagonal form can be obtained by the Remark 10.7.

10.1.3. Two Computed Examples

For illustration of eigenvalues, eigenvectors, diagonalizability let us see two computed examples.

In each of the two exercises the questions are as follows:

(a) Determine the eigenvalues, eigenvectors, eigenspaces, algebraic and geometric multiplicities.

(b) Determine the existence of E.B.

(c) Discuss the diagonalizability of the matrix (diagonalizing similarity matrix, diagonal form)

Answer these questions in the cases $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ too.

Example 1

$$A = \begin{bmatrix} 2 & -1 & -1 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix} \in \mathbb{K}^{3\times 3}.$$

Solution

The characteristic polynomial is

$$\begin{aligned} P(\lambda) = \det(A - \lambda I) &= \begin{vmatrix} 2-\lambda & -1 & -1 \\ 3 & -2-\lambda & -3 \\ -1 & 1 & 2-\lambda \end{vmatrix} = \\ &= (2-\lambda)\cdot[(-2-\lambda)(2-\lambda)+3] + 1\cdot[3(2-\lambda)-3] - 1\cdot[3-(2+\lambda)] = \\ &= (2-\lambda)(\lambda^2-1) - 2\lambda + 2 = (2-\lambda)(\lambda+1)(\lambda-1) - 2(\lambda-1) = \\ &= (\lambda-1)(\lambda-\lambda^2) = -\lambda(\lambda-1)^2 \qquad\qquad (\lambda \in \mathbb{K})\,. \end{aligned}$$

The eigenvalues are the roots of this polynomial:

$\lambda_1 = 0$, its algebraic multiplicity is $a(0) = 1$, because it is a single root.

$\lambda_2 = 1$, its algebraic multiplicity is $a(1) = 2$, because it is a double root.

So we have

$$\mathrm{Sp}\,(A) = \{0;\, 1\}.$$

If each eigenvalue is counted by its algebraic multiplicity, then we can also say that

$$\lambda_1 = 0, \quad \lambda_2 = 1, \quad \lambda_3 = 1.$$

Let us determine the eigenvectors.

In the case $\lambda_1 = 0$ the linear system to be solved is

$$\begin{bmatrix} 2 & -1 & -1 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Its nontrivial solutions are the eigenvectors:

$$x = \begin{pmatrix} x_1 \\ 3x_1 \\ -x_1 \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} \qquad (x_1 \in \mathbb{K} \setminus \{0\}).$$

The corresponding eigenspace is one dimensional (it is a line), whose basis (direction vector) is the vector $(1,\ 3,\ -1)$. By this reason the geometric multiplicity of the eigenvalue $\lambda_1 = 0$ is $g(0) = 1$.

In the case $\lambda_2 = 1$ the linear system to be solved is

$$\begin{bmatrix} 1 & -1 & -1 \\ 3 & -3 & -3 \\ -1 & 1 & 1 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Its nontrivial solutions are the eigenvectors:

$$x = \begin{pmatrix} x_2 + x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad (x_2, x_3 \in \mathbb{K},\ (x_2, x_3) \neq (0,0)).$$

The corresponding eigenspace is two dimensional (it is a plane), whose basis is the vector system $(1,1,0),\ (1,0,1)$. By this reason the geometric multiplicity of the eigenvalue $\lambda_2 = 1$ is $g(1) = 2$.

Now comes the discussion of the diagonalizability:

The sum of the algebraic multiplicities of the eigenvalues equals $a(1) + a(0) = 2 + 1 = 3$, thus we can go on. The geometric multiplicities are identical with the algebraic ones:

$$g(0) = a(0) = 1, \quad g(1) = a(1) = 2.$$

Therefore the matrix A is diagonalizable over $\mathbb{K}$. Let us note that here we cannot apply the sufficient condition „A has 3 different eigenvalues". The E.B. is the united system of eigenvectors:

$$c_1 = (1,3,-1), \qquad c_2 = (1,1,0) \qquad c_3 = (1,0,1).$$

The diagonalizing similarity matrix is built of the vectors of the E.B.

$$C = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \in \mathbb{K}^{3\times 3}.$$

Finally, the diagonal form of A is (we write the eigenvalues into the diagonal)

$$C^{-1}AC = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{K}^{3\times 3}.$$

Example 2

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} \in \mathbb{K}^{3\times 3}.$$

Solution

The characteristic polynomial (expansion by the first row is suggested) is:

$$\begin{aligned} P(\lambda) = \det(A-\lambda I) &= \begin{vmatrix} 1-\lambda & -1 & 1 \\ 1 & 1-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{vmatrix} = \\ &= (1-\lambda)\cdot[(1-\lambda)(2-\lambda)-1] - 1\cdot[(-1)(2-\lambda)+1] = \\ &= (1-\lambda)(\lambda^2-3\lambda+1)+1-\lambda = (1-\lambda)(\lambda^2-3\lambda+2) = \\ &= (1-\lambda)(\lambda-1)(\lambda-2) \qquad\qquad (\lambda\in\mathbb{K}). \end{aligned}$$

Hence the eigenvalues are $\lambda_1 = 1,\ \lambda_2 = 2$, their algebraic multiplicities are $a(1) = 2$, $a(2) = 1$. So we have

$$\mathrm{Sp}\,(A) = \{1;\, 2\}.$$

If each eigenvalue is counted by its algebraic multiplicity, then we can also say that

$$\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = 2.$$

Let us determine the eigenvectors.

<u>In the case $\lambda_1 = 1$</u> the linear system to be solved is

$$\begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Its nontrivial solutions are the eigenvectors:

$$x = \begin{pmatrix} x_2 \\ x_2 \\ x_2 \end{pmatrix} = x_2 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad (x_2 \in \mathbb{K} \setminus \{0\}).$$

The corresponding eigenspace is one dimensional (it is a line), whose basis (direction vector) is the vector $(1,\,1,\,1)$. By this reason the geometric multiplicity of the eigenvalue $\lambda_1 = 1$ is $g(1) = 1$.

In the case $\lambda_2 = 2$ the linear system to be solved is

$$\begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Its nontrivial solutions are the eigenvectors:

$$x = \begin{pmatrix} x_1 \\ 0 \\ x_1 \end{pmatrix} = x_1 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad (x_1 \in \mathbb{K} \setminus \{0\}).$$

The corresponding eigenspace is one dimensional (it is a line), whose basis (direction vector) is the vector $(1, 0, 1)$. By this reason the geometric multiplicity of the eigenvalue $\lambda_2 = 2$ is $g(2) = 1$.

Now comes the discussion of the diagonalizability:

The sum of the algebraic multiplicities of the eigenvalues equals

$$a(1) + a(2) = 2 + 1 = 3\,,$$

thus we can go on. But the identity between the algebraic and the geometric multiplicities is not true for any eigenvalues, because

$$g(1) = 1 < a(1) = 2.$$

Consequently we can establish – without discussing the other eigenvalues – that the matrix A is not diagonalizable over $\mathbb{K}$ and the E.B. does not exist in $\mathbb{K}^n$ neither in case $\mathbb{K} = \mathbb{R}$, nor in case $\mathbb{K} = \mathbb{C}$.

10.1.4. Control Questions to the Theory

1. Define the similarity of matrices

2. State and prove the theorem about the characteristic polynomials of similar matrices

3. Define the concept of a diagonalizable matrix

4. What are the diagonal entries of the diagonal form of a diagonalizable matrix?

5. State and prove the necessary and sufficient condition of diagonalizability

10.2. Exercises

10.2.1. Exercises for Class Work

1. The eigenvalues and the eigenvectors of the following matrices are discussed in the previous practice.

(a) Let us quote the results.

(b) Determine whether these matrices are diagonalizable or not. In the diagonalizable case determine the matrix C that diagonalizes A and the diagonal form $C^{-1}AC$.

Solve the above problems in the case $\mathbb{K} = \mathbb{R}$ and in the case $\mathbb{K} = \mathbb{C}$ too.

$$a) \quad A = \begin{bmatrix} 2 & -1 & -1 \\ 3 & -2 & -3 \\ -1 & 1 & 2 \end{bmatrix} \qquad b) \quad A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

$$c) \quad A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 2 & -1 & 0 \end{bmatrix} \qquad d) \quad A = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

10.2.2. Additional Tasks:

1. The eigenvalues and the eigenvectors of the following matrices are discussed in the previous practice.

(a) Let us quote the results.

(b) Determine whether these matrices are diagonalizable or not. In the diagonalizable case determine the matrix C that diagonalizes A and the diagonal form $C^{-1}AC$.

Solve the above problems in the case $\mathbb{K} = \mathbb{R}$ and in the case $\mathbb{K} = \mathbb{C}$ too.

$$a) \quad \begin{bmatrix} 2 & -1 \\ 10 & -9 \end{bmatrix} \qquad b) \quad \begin{bmatrix} -2 & -7 \\ 1 & 2 \end{bmatrix} \qquad c) \quad \begin{bmatrix} 5 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$$d) \quad A = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} \qquad e) \quad \begin{bmatrix} 1 & 2 & -2 \\ -3 & 4 & 0 \\ -3 & 1 & 3 \end{bmatrix}$$

11. Real Euclidean Spaces I.

In this and in the following chapter the elementary geometric vector operation the „scalar product of vectors" will be generalized for vector spaces. For simplicity it will only be about real (that is above $\mathbb{R}$) vector spaces.

11.1. Theory

11.1.1. The Concept of Real Euclidean Space

In the chapters so far we have generalized the concept of a vector, thus we have arrived to the concept of a vector space. In the secondary school we have learnt a third vector operation (outside of addition and scalar multiplication), namely the scalar product of vectors. We have established, that the scalar product has the following properties:

1. If $\underline{a}$ and $\underline{b}$ are vectors, then $\underline{a} \cdot \underline{b}$ is a real number (this is the origin of the name: scalar product)

2. $\underline{a} \cdot \underline{b} = \underline{b} \cdot \underline{a}$ (commutative law)

3. $(\lambda \underline{a}) \cdot \underline{b} = \lambda \cdot (\underline{a} \cdot \underline{b})$ (multiplication of a product by a number)

4. $\underline{a} \cdot (\underline{b} + \underline{c}) = \underline{ab} + \underline{ac}$ (multiplication of a sum, distributive law)

5. $\underline{a} \cdot \underline{a} \geq 0$, here stands equality if and only if $\underline{a} = \underline{0}$

We will generalize the concept of scalar product in the following way: we consider a real vector space, and out of the two vector space operations we introduce a third operation satisfying the above 5 properties. This „structure" will be called Euclidean space. The above 5 properties will be called the axioms of the scalar product.

After this short introduction, let us see the definition of the Euclidean space.

11.1. Definition Let V be a vector space over $\mathbb{R}$ with respect to the operations $x + y$ (addition) and λx (multiplication by scalar).

V is called Euclidean space (or inner product space) over $\mathbb{R}$, if there exist a third operation

$$xy = x \cdot y = \langle x,\, y\rangle$$

(it is called scalar product or inner product), for which the following axioms hold:

1. $\forall\, x, y \in V: \quad \langle x,\, y\rangle \in \mathbb{R}$

2. $\forall\, x, y \in V: \quad \langle x,\, y\rangle = \langle y,\, x\rangle$ (commutative law)

3. $\forall\, x, y \in V\ \forall\, \lambda \in \mathbb{R}: \quad \langle \lambda x,\, y\rangle = \lambda\langle x,\, y\rangle$ (multiplication of a product by a number)

4. $\forall\, x, y, z \in V: \quad \langle x,\, y+z\rangle = \langle x,\, y\rangle + \langle x,\, z\rangle$ (distributive law)

5. $\forall\, x \in V: \quad \langle x,\, x\rangle \geq 0$,

and here stands equality if and only if $x = 0$ (the scalar product is positive definite)

the other name of an Euclidean space over $\mathbb{R}$ is: real Euclidean space.

11.2. Examples

1. The plane vectors and the space vectors form a real Euclidean space with the well-known scalar product

$$\langle \underline{a},\, \underline{b}\rangle = \underline{a} \cdot \underline{b} = |\underline{a}| \cdot |\underline{b}| \cdot \cos\gamma\ ,$$

where γ denotes the angle between the vectors $\underline{a}$ and $\underline{b}$.

2. The vector space $\mathbb{R}^n$ is also an Euclidean space over $\mathbb{R}$ with the following scalar product:

$$\langle x,\, y\rangle := \sum_{i=1}^{n} x_i y_i \qquad (x, y \in \mathbb{R}^n)\,.$$

This is the default scalar product in $\mathbb{R}^n$.

The following theorem can be proved easily by the axioms:

11.3. Theorem (the basic properties of the scalar product)

Let V be an Euclidean space over $\mathbb{R}$. Then for any vector $x,\, x_i,\, y,\, y_j,\, z \in V$ and for any number $\lambda,\, \lambda_i,\, \mu_j \in \mathbb{R}$ hold

a) $\langle x,\, \lambda y\rangle = \lambda \cdot \langle x,\, y\rangle$;

b) $\langle x+y,\, z\rangle = \langle x,\, z\rangle + \langle y,\, z\rangle$;

c) $\langle \sum\limits_{i=1}^{n} \lambda_i x_i,\, \sum\limits_{j=1}^{m} \mu_j y_j\rangle = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} \lambda_i \mu_j \langle x_i,\, y_j\rangle$;

d) $\langle x,\, 0\rangle = \langle 0,\, x\rangle = 0$.

11.1.2. The Norm (Length) of a Vector

We obtain from the formula of the common geometrical scalar product

$$\langle \underline{a},\, \underline{b}\rangle = |\,\underline{a}| \cdot |\,\underline{b}| \cdot \cos\gamma$$

that

$$\langle \underline{a},\, \underline{a}\rangle = |\underline{a}| \cdot |\underline{a}| \cdot \cos 0 = |\underline{a}|^2 \quad \text{azaz} \quad |\underline{a}| = \sqrt{\langle \underline{a},\, \underline{a}\rangle}\,.$$

Using this observation we can generalize the concept of the „absolute value of a vector”:

11.4. Definition Let V be an Euclidean space over $\mathbb{R}$, and let $x \in V$. The norm of the vector x is defined as

$$\|x\| := \sqrt{\langle x,\, x\rangle}\,.$$

Other names for the norm are: the length of x, the absolute value of x.

The mapping $\|.\| : V \to \mathbb{R},\, x \mapsto \|x\|$ is called norm too.

11.5. Remark. The norm in an Euclidean space can be regarded as the abbreviation of the phrase „the square root of the scalar product of a vector with itself”.

11.6. Examples

1. In the Euclidean space of the plane vectors and of the space vectors the norm is identical with the well-known concept „length of a vector”:

$$\|\underline{a}\| = \sqrt{\langle a,\, a\rangle} = \sqrt{|\underline{a}| \cdot |\underline{a}| \cdot \cos(\underline{a}, \underline{a})} = |\underline{a}|\,.$$

2. In $\mathbb{R}^n$ the default scalar product generates the following norm:

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x_1^2 + x_2^2 + \ldots, x_n^2}\,,$$

that is – like what we have learned in secondary school – the square root from the square sum of the coordinates. This norm is the Euclidean vector norm in the space $\mathbb{R}^n$.

In the following theorem we will prove two simple properties of the norm.

11.7. Theorem (two simple properties of the norm)

1. $\|x\| \geq 0 \quad (x \in V)$. *Furthermore,* $\|x\| = 0 \iff x = 0$ *(the norm is positive definite)*
2. $\|\lambda x\| = |\lambda| \cdot \|x\| \qquad (x \in V;\ \lambda \in \mathbb{R})$ *(the norm is homogeneous)*

Proof. The first statement follows immediately from the fifth axiom of the scalar product.

The proof of the second statement is as follows:

$$\|\lambda x\| = \sqrt{\langle \lambda x,\, \lambda x\rangle} = \sqrt{\lambda\lambda\langle x,\, x\rangle} = \sqrt{\lambda^2 \cdot \|x\|^2} = \sqrt{\lambda^2} \cdot \sqrt{\|x\|^2} = |\lambda| \cdot \|x\|\,.$$

□

11.8. Remark. Another form of the first property is

$$\|\,0\| = 0 \quad \text{és} \quad \forall\, x \in V \setminus \{0\} : \|\,x\| > 0\,.$$

11.9. Definition Vector $x \in V$ is called a unit vector, if its norm equals 1, that is, if

$$\|x\| = 1$$

11.10. Remark. (normalization) Any nonzero vector can be transformed into a unit vector, which has the same direction as the original nonzero vector. Namely, if $x \in V \setminus \{0\}$, then the vector

$$x^0 := \frac{x}{\|x\|}$$

answers the purpose. Really, by $\dfrac{1}{\|x\|} > 0$ the directions of x and x^0 are identical, furthermore

$$\|x^0\| = \left\| \frac{x}{\|x\|} \right\| = \left\| \frac{1}{\|x\|} \cdot x \right\| = \frac{1}{\|x\|} \cdot \|x\| = 1\,.$$

This process (division by the norm) is called: normalization.

11.1.3. Orthogonality

It follows immediately from the formula of the common geometrical scalar product

$$\langle \underline{a},\, \underline{b} \rangle = |\underline{a}| \cdot |\underline{b}| \cdot \cos\gamma$$

that if neither $\underline{a}$ nor $\underline{b}$ is the zero vector, then these two vectors are perpendicular (orthogonal) to each other if and only if their scalar product equals 0. This observation will be used to the definition of orthogonality in Euclidean spaces.

In this section V denotes an Euclidean space over $\mathbb{R}$.

11.11. Definition The vectors $x, y \in V$ are called orthogonal (perpendicular) to each other, if their scalar product is 0, that is, if

$$\langle x,\, y \rangle = 0\,.$$

This relation (which is obviously symmetric) is denoted by $x \perp y$.

11.12. Remark. It is easy to see, that the zero vector is orthogonal to each vector in the space (to itself too). Furthermore, the zero vector is the unique vector which is orthogonal to itself.

11.13. Definition (Orthogonality to a set) Let $\emptyset \neq H \subseteq V$ and $x \in V$. Vector x is said to be orthogonal (perpendicular) to the set H (in notation: $x \perp H$), if it is orthogonal to any element of H, that is, if

$$\forall\, y \in H : \quad \langle x,\, y \rangle = 0\,.$$

The following theorem states, that the orthogonality to a finite dimensional subspace is equivalent to the orthogonality to one of its generator systems.

11.14. Theorem (orthogonality to a subspace) *Let $e_1, \ldots, e_n \in V$ be a vector system, $W := \operatorname{Span}(e_1, \ldots, e_n)$, and $x \in V$. Then*

$$x \perp W \quad \Longleftrightarrow \quad \langle x, e_i \rangle = 0 \quad (i = 1, \ldots, n).$$

Proof.
„$\Longrightarrow$": It is trivial by choosing $y := e_i$.
„$\Longleftarrow$": Let $y = \sum_{i=1}^{n} \lambda_i e_i \in W$ an arbitrary vector. Then

$$\langle x, y \rangle = \langle x, \sum_{i=1}^{n} \lambda_i e_i \rangle = \sum_{i=1}^{n} \lambda_i \langle x, e_i \rangle = \sum_{i=1}^{n} \lambda_i \cdot 0 = 0.$$

□

11.15. Definition Let $x_1, \ldots, x_n \in V$ be a finite vector system in the Euclidean space V.

1. The system $x_1, \ldots, x_n$ is said to be orthogonal system (O.S.), if each pair of different vectors in this system is orthogonal, that is, if

$$\forall\, i, j \in \{1, \ldots n\},\ i \neq j : \quad \langle x_i, x_j \rangle = 0.$$

2. The system $x_1, \ldots, x_n$ is said to be orthonormal system (O.N.S.), if it is an orthogonal system and each vector in this system is a unit vector too, that is, if

$$\forall\, i, j \in \{1, \ldots n\} : \qquad \langle x_i, x_j \rangle = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j. \end{cases}$$

3. If an orthogonal system is basis at the same time, then it is called an orthogonal basis (O.B.).

4. If an orthonormal system is basis at the same time, then it is called an orthonormal basis (O.N.B.).

11.16. Remarks.

1. One can simply see that

   - An O.S. can contain the zero vector.
   - An O.N.S. cannot contain the zero vector.

- In an O.S. the zero vector can occur several times, but any nonzero vector at most once.
- An O.N.S. cannot contain identical vectors.

2. (normalization of an O.S.) We can transform an O.S. easily into an O.N.S. which generates the same subspace as the original O.S.:

First let us omit the possible zero vectors from the system, then let us normalize each vector of the remainder system (see Remark 11.10).

11.17. Examples

1. In the Euclidean space of plane vectors the well-known basic vectors **i**, **j** form an O.N.B.

2. In the Euclidean space of space vectors the well-known basic vectors **i**, **j**, **k** form an O.N.B.

3. In the Euclidean space $\mathbb{R}^n$ the standard unit vectors $e_1, \ldots, e_n$ form an O.N.B.

As we can see in the world of space vectors, the orthogonality is a stronger concept, than the linear independence. This is expressed in the following theorem:

11.18. Theorem (independence of O.S.) *Let $x_1, \ldots, x_n \in V \setminus \{0\}$ be an O.S. Then this system is linearly independent. Consequently, every O.N.S. is linearly independent.*

Proof.

Consider the dependence equation

$$0 = \sum_{i=1}^{n} \lambda_i x_i \ ,$$

and multiply it by vector x_j (in the sense of scalar product), where $j = 1, \ldots, n$:

$$0 = \langle 0, \, x_j \rangle = \langle \sum_{i=1}^{n} \lambda_i x_i, \, x_j \rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, \, x_j \rangle = \lambda_j \langle x_j, \, x_j \rangle \, .$$

Since the zero vector is excluded from the given system, then $\langle x_j,\, x_j\rangle \neq 0$. Consequently $\lambda_j = 0$.

Thus all the coefficients in the upper dependence equation are 0, thus the system is really linearly independent. $\square$

Now comes the other basic theorem of orthogonal systems, which is the generalization of the Pythagorean theorem. The pythagorean theorem is taught in the elementary mathematics, as the square of the hypotenuse in a right-angled triangle equals the sum of the squares of the legs. This can be expressed in the language of vectors as follows:

If the plane or space vectors $\underline{a}$ and $\underline{b}$ are orthogonal to each other, then

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

This will be generalized for arbitrary but finite number of vectors.

11.19. Theorem (Pythagorean theorem) *Let $x_1, \ldots, x_n \in V$ be a finite O.S. Then*

$$\|\sum_{i=1}^n x_i\|^2 = \sum_{i=1}^n \|x_i\|^2 \,, \tag{11.1}$$

in more detail

$$\|x_1 + x_2 + \ldots + x_n\|^2 = \|x_1\|^2 + \|x_2\|^2 + \ldots \|x_n\|^2 \,.$$

Proof.

$$\|\sum_{i=1}^n x_i\|^2 = \langle\sum_{i=1}^n x_i,\, \sum_{j=1}^n x_j\rangle = \sum_{i=1}^n\sum_{j=1}^n \langle x_i,\, x_j\rangle = \sum_{\substack{i,j=1\\ i\neq j}}^n \langle x_i,\, x_j\rangle + \sum_{\substack{i,j=1\\ i=j}}^n \langle x_i,\, x_j\rangle =$$

$$= \sum_{\substack{i,j=1\\ i\neq j}}^n 0 + \sum_{i=1}^n \langle x_i,\, x_i\rangle = \sum_{i=1}^n \|x_i\|^2.$$

We have used that for $i \neq j$ holds $\langle x_i,\, x_j\rangle = 0$. $\square$

11.1.4. Fourier-expansion

Consider a finite dimensional subspace W in the Euclidean space V, and a finite generator system of W (attention: the basic space V is not assumed to be finite dimensional). We already know, that the vectors of the subspace W can be written as linear combinations of the generator system.

The basic question of this section is: how can we express the coefficients of this linear combination using the scalar product. We will see, that this expression is very simple if the generator system is orthonormal.

So let $e_1, \ldots, e_n \in V$ be a finite vector system, $W := \operatorname{Span}(e_1, \ldots, e_n)$ be the generated subspace, furthermore $x \in W$. Then

$$\exists\, \lambda_1, \ldots, \lambda_n \in \mathbb{R} : \qquad \sum_{j=1}^{n} \lambda_j e_j = x\,.$$

Multiply both sides of this equation by e_i (in the sense of scalar product) for $(i = 1, \ldots, n)$:

$$\langle \sum_{j=1}^{n} \lambda_j e_j,\, e_i \rangle = \langle x,\, e_i \rangle\,.$$

After some rearrangement we obtain:

$$\sum_{j=1}^{n} \lambda_j \langle e_j,\, e_i \rangle = \langle x,\, e_i \rangle \qquad (i = 1, \ldots, n)\,. \tag{11.2}$$

This is an $n \times n$ system of linear equations, with the unknown variables $\lambda_1, \ldots, \lambda_n$.

Thus the coefficients we are looking for are the solutions of the linear system (11.2).

Conversely, suppose that the numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are the solutions of the linear system (11.2). Now we will transform the upper equations in the „opposite" direction $(i = 1, \ldots, n)$:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \langle e_j,\, e_i \rangle &= \langle x,\, e_i \rangle \\ \langle x,\, e_i \rangle - \sum_{j=1}^{n} \lambda_j \langle e_j,\, e_i \rangle &= 0 \\ \langle x - \sum_{j=1}^{n} \lambda_j e_j,\, e_i \rangle &= 0\,. \end{aligned}$$

Since the system $e_1, \ldots, e_n$ is a generator system in the subspace W, then the last equation means that the vector $x - \sum\limits_{j=1}^{n} \lambda_j e_j \in W$ is orthogonal to the subspace W. But this vector lies in the subspace, consequently it is orthogonal to itself. We know that only the zero vector can be orthogonal to itself (see the basic properties of Euclidean spaces), thus we have

$$x - \sum_{j=1}^{n} \lambda_j e_j = 0 \qquad \text{that is} \qquad x = \sum_{j=1}^{n} \lambda_j e_j\,.$$

So the solutions of the linear system (11.2) really give the coefficients of the discussed linear combination.

The summarized result of the previous consideration is: finding the possible coefficients of the linear combinations which result the vector $x \in W$ is equivalent with solving the linear system (11.2).

11.20. Definition The linear equations (11.2) are called the Gaussian normal equations, and their system is called the Gaussian normal system.

The matrix form of the Gaussian normal system is as follows:

$$\begin{bmatrix} \langle e_1,\, e_1\rangle & \langle e_2,\, e_1\rangle & \ldots & \langle e_n,\, e_1\rangle \\ \langle e_1,\, e_2\rangle & \langle e_2,\, e_2\rangle & \ldots & \langle e_n,\, e_2\rangle \\ \vdots & \vdots & & \vdots \\ \langle e_1,\, e_n\rangle & \langle e_2,\, e_n\rangle & \ldots & \langle e_n,\, e_n\rangle \end{bmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \langle x,\, e_1\rangle \\ \langle x,\, e_2\rangle \\ \vdots \\ \langle x,\, e_n\rangle \end{pmatrix}. \tag{11.3}$$

One can see, that its coefficient matrix is depending only on the vectors $e_1, \ldots, e_n$, but it is independent of vector x. This leads us to the following definition:

11.21. Definition The matrix

$$G := G_n := G(e_1, \ldots, e_n) := \begin{bmatrix} \langle e_1,\, e_1\rangle & \langle e_2,\, e_1\rangle & \ldots & \langle e_n,\, e_1\rangle \\ \langle e_1,\, e_2\rangle & \langle e_2,\, e_2\rangle & \ldots & \langle e_n,\, e_2\rangle \\ \vdots & \vdots & & \vdots \\ \langle e_1,\, e_n\rangle & \langle e_2,\, e_n\rangle & \ldots & \langle e_n,\, e_n\rangle \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{11.4}$$

is said to be the Gram-matrix of the vector system $e_1, \ldots, e_n \in V$. The determinant of the Gram-matrix is called the Gram-determinant of the vector system $e_1, \ldots, e_n$.

11.22. Remark. As you can see, the entries of the Gram-matrix are given by

$$(G)_{ij} = \langle e_j,\, e_i \rangle \qquad (i, j = 1, \ldots, n\,.$$

In our subject we do not intended to investigate the Gram-matrix. You can read about an interesting property of the Gram-matrix in the Appendix, in section „The geometrical meaning of the determinant".

Once more we establish, that the coefficients of the expansion of $x \in W$ by a generator system can be obtained by solving the Gaussian normal system. This process requires generally a lot of computations.

However the generator system $e_1, \ldots, e_n$ is an orthonormal system (O.N.S.) at the same time, so its Gram-matrix is the identity matrix.

In this case the unique solution of the Gaussian system is

$$\lambda_i = \langle x,\, e_i \rangle \qquad (i = 1, \ldots, n)\,,$$

consequently, the unique expansion of x is

$$x = \sum_{i=1}^{n} \langle x,\, e_i \rangle \cdot e_i\,.$$

11.23. Definition The numbers

$$c_i = \langle x,\, e_i \rangle \qquad (i = 1, \ldots, n)$$

are said to be the Fourier-coefficients of x, the expansion

$$x = \sum_{i=1}^{n} c_i e_i = \sum_{i=1}^{n} \langle x,\, e_i \rangle \cdot e_i$$

is said to be the Fourier-expansion (or: Fourier-sum) of x with respect to the O.N.S. $e_1, \ldots, e_n$.

11.24. Remarks.

1. The Fourier-coefficients of $x \in W$ are identical with the coordinates of x with respect to the orthonormal basis $e_1, \ldots, e_n$ of the subspace W.

2. Applying our result to the Fourier-expansion of the zero vector $x = 0$ we obtain a new justification of the linear independence of an orthonormal system.

11.1.5. Control Questions to Theory

1. Define the concept of a real Euclidean space
2. State the theorem about the 4 basic properties of the scalar product
3. Define the norm of a vector in a real Euclidean space
4. Give the formula of the Euclidean norm in $\mathbb{R}^n$
5. State the theorem about the two simple properties of the norm
6. Define the unit vector.
7. What does normalization mean?
8. Define the following concepts: orthogonality of two vectors, orthogonality of a vector to a set
9. State the theorem about the orthogonality of a vector to a subspace
10. Define the orthogonal system and the orthonormal system
11. State the theorem about the linear independence of an orthogonal system
12. State the Pythagorean theorem in Euclidean spaces
13. What is the formula of the coefficients if we write a vector $x \in W$ as the linear combination of the O.N.S. $e_1, \ldots, e_n$, where $W := \mathrm{Span}\,(e_1, \ldots, e_n)$. What is the name of these coefficients?

11.2. Exercises

11.2.1. Exercises for Class Work

1. Let $w_1, \ldots, w_n$ be given positive numbers (weights). Prove that the vector space $\mathbb{R}^n$ is an Euclidean space over $\mathbb{R}$ with the scalar product

$$\langle x,\, y\rangle := \sum_{i=1}^{n} w_i x_i y_i$$

(In the case $w_i = 1$ the default scalar product is obtained.)

2. Consider the vectors

$$x := (1,-2,-3,5), \quad y := (-1,2,-1,0), \quad z := (2,-1,1,3) \in \mathbb{R}^4 .$$

Compute:

(a) $\langle x,\, y\rangle$

(b) $\| x\|$

(c) $\| x - z\|$

(d) $\dfrac{\langle x,\, z\rangle \cdot y - \langle y,\, z\rangle \cdot x}{\| y\|^2}$

(e) the unit vector in the direction of z and the unit vector in opposite direction of z

3. Consider the vector system

$$u_1 := (1,1,1,1),\ u_2 := (1,-1,-1,1),\ u_3 := (-1,0,0,1)$$

in the Euclidean space $\mathbb{R}^4$.

(a) Show that $u_1,\ u_2,\ u_3$ is an orthogonal system (O.R.).

(b) Check the Pythagorean theorem for the system $u_1,\ u_2,\ u_3$.

4. Prove the Parallelogram-identity in a real Euclidean space V:

$$\forall\, x, y \in V : \quad \|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

11.2.2. Additional Tasks

1. Let $x = (3, -2, 1, 1)$, $y = (4, 5, 3, 1)$, $z = (-1, 6, 2, 0) \in \mathbb{R}^4$, and let $\lambda = -4$. Prove that in this case:

a) $\langle x, y\rangle = \langle y, x\rangle$

b) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$

c) $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$

(We use the default operations in $\mathbb{R}^4$.)

2. Using the data of the previous exercise compute

$\dfrac{\langle x, x\rangle}{\langle y, z\rangle} \cdot x$, $\quad \| z - y\| \cdot x$, $\quad$ the unit vector in the direction of y

3. Let $x_1 = (0, 0, 0, 0)$, $x_2 = (1, -1, 3, 0)$, $x_3 = (4, 0, 9, 2) \in \mathbb{R}^4$. Determine whether $x = (-1, 1, 0, 2)$ is orthogonal to the subspace $\mathrm{Span}\,(x_1, x_2, x_3)$ or not.

4. Prove that

- the Gram matrix of a linearly independent vector system is regular

- the Gram matrix of a linearly dependent vector system is singular

(Hint: use the Gaussian normal equation system.)

12. Real Euclidean Spaces II.

12.1. Theory

12.1.1. The Projection Theorem

In the previous section we have investigated the expansion of a vector $x \in W$ in the case when the generator system of the subspace W is orthonormal. In this way we have arrived to the Fourier-expansion of x. Here W is a finite dimensional subspace of the Euclidean space V (V is not assumed to be finite dimensional).

But the Fourier-coefficients can be formed not only in the case $x \in W$, but for any $x \in V$ too. By this mean the following question arises naturally: For a vector $x \in V$ – especially for $x \in V \setminus W$ – which vector is given by the Fourier-sum

$$\sum_{i=1}^{n} \langle x,\, e_i \rangle \cdot e_i \ .$$

This question will be investigated in this section.

12.1. Theorem (Projection Theorem)

Let $e_1, \dots, e_n \in V$ be an O.N.S. and $W := \operatorname{Span}(e_1, \dots, e_n)$ the generated subspace. (Note that in this case $e_1, \dots, e_n$ is an O.N.B. in W.)

Then any vector $x \in V$ can be expressed uniquely in the form $x = x_1 + x_2$ where $x_1 \in W$ and $x_2 \perp W$. Namely

$$x_1 = \sum_{i=1}^{n} \langle x,\, e_i \rangle \cdot e_i \quad \textit{and} \quad x_2 = x - x_1 = x - \sum_{i=1}^{n} \langle x,\, e_i \rangle \cdot e_i \, .$$

Proof. Let us prove first the existence of the decomposition. We will show, that the given formulas give a correct decomposition. Let c_i denote the i-th Fourier-coefficient, that is let

$$c_i = \langle x,\, e_i \rangle \qquad (i = 1, \dots, n)\, .$$

Using this notation we have

$$x_1 = \sum_{i=1}^{n} c_i e_i \quad \text{and} \quad x_2 = x - x_1 = x - \sum_{i=1}^{n} c_i e_i \, .$$

Obviously $x_1 \in W$, because it is the linear combination of the e_i-s.

It is also obvious that $x = x_1 + x_2$, because $x_2 = x - x_1$.

It remains to prove only that $x_2 \perp W$. To prove this, we use the fact, that the orthogonality to the subspace W is equivalent with the orthogonality to its generator system $e_1, \ldots, e_n$ (see theorem 11.14). But this follows immediately from the following calculations:

$$\begin{aligned}\langle x_2, e_i\rangle &= \langle x - \sum_{j=1}^{n} c_j e_j, e_i\rangle = \langle x, e_i\rangle - \sum_{j=1}^{n} c_j \langle e_j, e_i\rangle = \\ &= \langle x, e_i\rangle - \sum_{\substack{i,j=1\\ i\neq j}}^{n} c_j \langle e_j, e_i\rangle - c_i\langle e_i, e_i\rangle = \\ &= \langle x, e_i\rangle - 0 - c_i \cdot 1 = 0 \qquad (i = 1, \ldots, n).\end{aligned}$$

As a second step let us prove the uniqueness.

Suppose that

$$x = x_1 + x_2 \quad \text{and} \quad x = x_1' + x_2'$$

both are decompositions corresponding to the requirements. Then

$$x_1 + x_2 = x_1' + x_2', \quad \text{after rearrangement:} \quad x_1 - x_1' = x_2' - x_2. \tag{12.1}$$

Using this we have

$$\begin{aligned}\langle x_1 - x_1', x_1 - x_1'\rangle &= \langle x_2' - x_2, x_1 - x_1'\rangle = \\ &= \langle x_2', x_1\rangle - \langle x_2, x_1\rangle - \langle x_2', x_1'\rangle + \langle x_2, x_1'\rangle = 0 - 0 - 0 + 0 = 0.\end{aligned}$$

From here – using the last axiom of the scalar product – we can conclude that $x_1 - x_1' = 0$, that is $x_1 = x_1'$. But in this case – using (12.1) – it follows directly, that $x_2 = x_2'$. □

12.2. Remarks.

1. Vector x_1 is said to be the parallel component of x relative to W. Its notation is $P(x)$. The name of x_2 is: the orthogonal component of vector x relative to the subspace W. Its notation is: $Q(x)$.

2. Another name of the vector $P(x) = x_1$ is: the orthogonal projection of the vector x onto the subspace W. If we want to emphasize this

content, then instead of $P(x)$ is better to use the notation $\mathrm{proj}_W(x)$. It follows from our theorem, that

$$\mathrm{proj}_W(x) = \sum_{i=1}^{n} \langle x,\, e_i\rangle \cdot e_i\,.$$

Thus we have answer now the question we have raised at the beginning of this section:

If $x \in V$, then the result of the Fourier-sum is the orthogonal projection of vector x onto the subspace $W = \mathrm{Span}\,(e_1, \dots, e_n)$.

In the case when $x \in W$, this projection is naturally vector x itself.

3. Later on (see corollary 12.6) we will show that any finite dimensional subspace can be generated by a finite O.N.S., thus the decomposition into parallel and orthogonal components can be made for any finite dimensional subspace.

The projection Theorem and its formulae can be easily generalized for the case of an orthogonal (not necessarily orthonormal) generator system. Suppose for simplicity, that the orthogonal generator system does not contain the zero vector.

So let $u_1, \dots, u_n \in V \setminus \{0\}$ be an orthogonal system (O.S.), $W := \mathrm{Span}\,(u_1, \dots, u_n)$ be the generated subspace, and $x \in V$. Then, by normalization we have the orthonormal system

$$\frac{u_1}{\|u_1\|},\ \frac{u_2}{\|u_2\|},\ \dots\ \frac{u_n}{\|u_n\|}\ ,$$

which obviously generates subspace W. For this normalized system we can apply the already proved formulae of the decomposition:

$$P(x) = \sum_{i=1}^{n} \langle x,\, \frac{u_i}{\|u_i\|}\rangle \cdot \frac{u_i}{\|u_i\|} = \sum_{i=1}^{n} \frac{1}{\|u_i\|^2} \cdot \langle x,\, u_i\rangle \cdot u_i = \sum_{i=1}^{n} \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \cdot u_i$$

$$Q(x) = x - \sum_{i=1}^{n} \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \cdot u_i$$

12.3. Theorem (Estimation of the Length of the Projection) *Preserving the conditions and notations of the Projection Theorem we have:*

$$\| P(x)\| \leq \| x\| \, .$$

Here stands equality if and only if $Q(x) = 0$ *(This last condition is equivalent with* $x \in W$*.)*

Proof. Since $P(x) \perp Q(x)$, let us apply the Pythagorean theorem, then let us omit the non-negative term $\| Q(x)\|^2$:

$$\| x\|^2 = \| P(x) + Q(x)\|^2 = \| P(x)\|^2 + \| Q(x)\|^2 \geq \| P(x)\|^2 \; .$$

After performing a square root we obtain the statement to be proved.

Obviously, in the last estimation equality stands if and only if $Q(x) = 0$. □

12.4. Remarks.

1. The inequality $\| Q(x)\| \leq \| x\|$ can be proved similarly. Here stands equality if and only if $P(x) = 0$, which is equivalent with $x \perp W$.
2. The estimations $\| P(x)\| \leq \| x\|$ and $\| Q(x)\| \leq \| x\|$ are the generalizations of the following statement in the elementary geometry: in a right-angled triangle the legs are no longer than the hypotenuse.

12.1.2. The Gram-Schmidt Process

Let $b_1, b_2, \ldots, b_n \in V$ be a finite linearly independent vector system. Let us describe now the Gram-Schmidt orthogonalization process, which – starting out from the above system – gives us an orthogonal system

$$u_1, u_2, \ldots, u_n \in V \setminus \{0\} \; ,$$

which is equivalent with the original system in the following sense:

$$\forall\, k \in \{1, 2, , \ldots, n\} : \quad \text{Span}\,(b_1, \ldots, b_k) = \text{Span}\,(u_1, \ldots, u_k) \, .$$

Especially (for $k = n$) the two systems generate the same subspace.

The process is as folows:

Step 1: $u_1 := b_1$

Step 2 $u_2 := b_2 - \dfrac{\langle b_2,\, u_1\rangle}{\langle u_1,\, u_1\rangle} \cdot u_1$

Step 3: $u_3 := b_3 - \dfrac{\langle b_3,\, u_1\rangle}{\langle u_1,\, u_1\rangle} \cdot u_1 - \dfrac{\langle b_3,\, u_2\rangle}{\langle u_2,\, u_2\rangle} \cdot u_2$

$\vdots$

Step n: $u_n := b_n - \dfrac{\langle b_n,\, u_1\rangle}{\langle u_1,\, u_1\rangle} \cdot u_1 - \dfrac{\langle b_n,\, u_2\rangle}{\langle u_2,\, u_2\rangle} \cdot u_2 - \ldots - \dfrac{\langle b_n,\, u_{n-1}\rangle}{\langle u_{n-1},\, u_{n-1}\rangle} \cdot u_{n-1}.$

It can be proved that this process produces a system $u_1,\, u_2,\, \ldots,\, u_n$ corresponding to the prescribed requirements of the original assignment.

12.5. Remarks.

1. The essence of this process is – and it also gives an illustrative justification of it – that

   - u_2 is the orthogonal component of b_2 relative to the subspace Span (u_1)
   - u_3 is the orthogonal component of b_3 relative to the subspace Span $(u_1,\, u_2)$
   - u_4 is the orthogonal component of b_4 relative to the subspace Span $(u_1,\, u_2, q, u_3)$

   $\vdots$

   - u_n is the orthogonal component of b_n relative to the subspace Span $(u_1,\, u_2,\, \ldots,\, u_{n-1})$

2. This process can be modified in the following way: multiply vector u_k obtained in the k-th step by a constant $c_k \neq 0$, and use vector $c_k u_k$ instead of u_k. One can easily see, that the modified process also gives us an equivalent orthogonal system.

   Using $c_k = \dfrac{1}{\|\, u_k\|}$ in the upper modification, the process will give us an equivalent orthonormal system. This is the normalized Gram-Schmidt process (Gram-Schmidt orthonormalization process).

12.6. Corollary. Let V be an Euclidean space over $\mathbb{R}$, $1 \leq \dim V = n < \infty$. Then there exists orthogonal basis and also orthonormal basis in V.

To get them take a basis $b_1, b_2, \ldots, b_n$ of the space, and apply the Gram-Schmidt process to it. Thus we obtain an orthogonal basis $u_1, u_2, \ldots, u_n$. Normalizing it, we obtain an orthonormal basis.

Thus we have the forecasted result from the Remark 12.2: Every finite dimensional nonzero subspace can be generated by an O.N.S.

Consequently, the decomposition into parallel and orthogonal components can be made in any finite dimensional subspace of V.

12.1.3. Triangle-inequality

In elementary geometry we have learnt, that the sum of the lengthes of two sides of a triangle is at least the length of the third side. We can express this fact with vectors as well so, that for any two vectors $\underline{a}$ and $\underline{b}$ we have

$$|\,\underline{a}+\underline{b}| \leq |\,\underline{a}| + |\,\underline{b}|\ .$$

In this section we prove the above inequality in an arbitrary Euclidean space.

To see this, first we prove the Cauchy-inequality, which is interesting by itself as well.

12.7. Theorem (Cauchy-inequality) *Let V be an Euclidean space, $x, y \in V$. Then*

$$|\,\langle x,\, y\rangle| \leq \|x\| \cdot \|y\|\,.$$

Proof. The statement is obviously true for $y = 0$ (in equality form). Suppose $y \neq 0$. Then we can apply the Projection Theorem for the one-term orthogonal system $u_1 := y$:

$$x = P(x) + Q(x), \quad \text{where} \quad P(x) = \sum_{i=1}^{1} \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \cdot u_i = \frac{\langle x,\, y\rangle}{\langle y,\, y\rangle} \cdot y = \frac{\langle x,\, y\rangle}{\|\,y\|^2} \cdot y$$

Let us use this formula of $P(x)$ in Theorem 12.3 about estimating the length of the projection. After some rearrangements we have

$$\left\| \frac{\langle x,\, y\rangle}{\|\, y\|^2} \cdot y \right\| \leq \|\, x\|$$

$$\left| \frac{\langle x,\, y\rangle}{\|\, y\|^2} \right| \cdot \|\, y\| \leq \|\, x\|$$

$$\frac{|\,\langle x,\, y\rangle|}{\|\, y\|^2} \cdot \|\, y\| \leq \|\, x\|$$

$$|\,\langle x,\, y\rangle| \leq \|x\| \cdot \|y\|\,.$$

□

12.8. Remark. It turns out from this proof, that equality holds if and only if x and y are linearly dependent. It can be easily verified, that

– if x and y have the same direction (that is $\exists\, \lambda > 0: \ x = \lambda y$), then the equality holds in the form

$$\langle x,\, y\rangle = \|x\| \cdot \|y\| \ ,$$

– and if x and y have the opposite direction (that is $\exists\, \lambda < 0: \ x = \lambda y$), then the equality holds in the form

$$\langle x,\, y\rangle = -\|x\| \cdot \|y\|$$

Using Cauchy's inequality we can prove the triangle-inequality. This will be the third basic property of the norm. The first and the second ones were proved in Theorem 11.7.

12.9. Theorem (Triangle-inequality)

$$\forall\, x,\, y \in V: \qquad \|x + y\| \leq \|x\| + \|y\|\,.$$

Proof.

$$\|x + y\|^2 = \langle x + y,\, x + y\rangle = \langle x,\, x\rangle + \langle x,\, y\rangle + \langle y,\, x\rangle + \langle y,\, y\rangle =$$

$$= \|x\|^2 + 2\langle x,\, y\rangle + \|y\|^2 \leq \|x\|^2 + 2 \cdot |\langle x,\, y\rangle| + \|y\|^2 \leq$$

$$\leq \|x\|^2 + 2 \cdot \|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\,.$$

By taking the square roots of both sides we obtain the desired inequality. In the last estimation we have used Cauchy's inequality. □

12.10. Remark. Taking into consideration the case of equality in the Cauchy-inequality, we can establish, that in the triangle inequality stands equality if and only if either $x = 0$ or $y = 0$ or none of them is the zero vector but they have the same direction (they are the positive constant-multiple of each other).

12.1.4. Control Questions to the Theory

1. State the Projection Theorem

2. State the theorem estimating the length of the projection

3. Describe the Gram-Schmidt orthogonalization process

4. State the theorem about the Cauchy-inequality (without the case of equality)

5. State the theorem about the Triangle-inequality (without the case of equality)

12.2. Exercises

12.2.1. Exercises for Class Work

1. Decompose vector $x = (2, 1, 3, 1) \in \mathbb{R}^4$ into parallel and orthogonal components by the subspace

$$W := \operatorname{Span}\left((1, -1, -1, 1),\ (1, 1, 1, 1),\ (-1, 0, 0, 1)\right) \subset \mathbb{R}^4 .$$

2. Use the Gram-Schmidt process to transform the linearly independent system

$$b_1 := (1,1,1,1), \quad b_2 := (3,3,-1,-1), \quad b_3 := (-2,0,6,8) \quad \in \mathbb{R}^4$$

into an equivalent orthogonal system. What is the rank of the system b_1, b_2, b_3?

3. Determine an orthogonal basis in the subspace generated by the vectors

$$b_1 := (1,1,1,1), \quad b_2 := (3,3,-1,-1), \quad b_3 := (-2,0,6,8) \quad \in \mathbb{R}^4 .$$

4. (a) Determine an orthogonal and an orthonormal basis in the subspace

$$W := \{y \in \mathbb{R}^4 \mid 3y_1+2y_2+y_3-2y_4 = 0,\ 5y_1+4y_2+3y_3+2y_4 = 0\} \subset \mathbb{R}^4$$

(b) Decompose vector $x := (3,4,-3,5) \in \mathbb{R}^4$ into parallel and orthogonal components by subspace W

(c) Determine the matrix, whose nullspace is W.

12.2.2. Additional Tasks

1. Determine the orthogonal projection of vector $x = (1,2,0,-2) \in \mathbb{R}^4$ onto the subspaces (from $\mathbb{R}^4$), generated by the following orthogonal systems:

a) $u_1 = (0,1,-4,-1)$, $u_2 = (3,5,1,1)$.

b) $u_1 = (1,-1,-1,1)$, $u_2 = (1,1,1,1)$, $u_3 = (1,1,-1,-1)$.

2. Using the Gram-Schmidt process, transform the basis

$$b_1 = (0,2,1,0),\ b_2 = (1,-1,0,0),\ b_3 = (1,2,0,-1),\ b_4 = (1,0,0,1) \in \mathbb{R}^4$$

(a) into an orthogonal basis in $\mathbb{R}^4$

(b) into an orthonormal basis in $\mathbb{R}^4$

3. Consider the following subspaces in $\mathbb{R}^4$:

(a) $W := \{y \in \mathbb{R}^4 \mid y_1 - y_2 + y_3 + y_4 = 0,\ 2y_1 - y_2 - y_3 = 0\}$

(b) $W := \{y \in \mathbb{R}^4 \mid y_1 - y_2 + y_3 + y_4 = 0\}$

Perform the following tasks for both subspaces:

(a) Determine an orthogonal and an orthonormal basis in W.

(b) Determine the orthogonal projection of $x := (0, 1, -1, 0) \in \mathbb{R}^4$ onto W.

(c) Determine the matrix whose nullspace is W.

4. Prove Bessel's inequality:

If $e_1, \ldots, e_n$ is an O.N.S. in a real Euclidean space V, then

$$\forall\, x \in V: \quad \sum_{i=1}^{n} |\, \langle x,\, e_i \rangle |^2 \leq \|x\|^2$$

13. Appendix

13.1. An Example of Infinite Dimensional Vector Space

Let V be the set of infinite sequences in $\mathbb{K}$, in which the number of nonzero terms is finite:

$$V := \{x : \mathbb{N} \to \mathbb{K} \mid \{i \mid x_i \neq 0\} \text{ is a finite set}\} .$$

Then V is a vector space over $\mathbb{K}$ relative to the common termwise addition and scalar multiplication of sequences.

We will show that $\dim V = \infty$.

To prove this, let $x^{(1)}, \ldots, x^{(k)} \in V$ be an arbitrary finite vector system. We will show that this system cannot be a generator system in V, that is $\operatorname{Span}(x^{(1)}, \ldots, x^{(k)}) \neq V$.

Let n_i be the index of the last nonzero term of the sequence $x^{(i)}$, where $i = 1, \ldots, k$. Furthermore, let

$$N := \max\{n_1, \ldots, n_k\} + 1 .$$

Then the N-th term in each of the sequences $x^{(1)}, \ldots, x^{(k)}$ equals 0, that is:

$$x_N^{(1)} = 0, \quad \ldots, \quad x_N^{(k)} = 0 .$$

Thus, the N-th term of any linear combination

$$\lambda_1 x^{(1)} + \cdots + \lambda_k x^{(k)}$$

also equals 0.

By this reason, if $y \in V$ is a sequence for which $y_N = 1$, then

$$y \notin \operatorname{Span}(x^{(1)}, \ldots, x^{(k)}), \quad \text{consequently} \quad \operatorname{Span}(x^{(1)}, \ldots, x^{(k)}) \neq V .$$

13.2. Examples of Invertibility of 4×4 Matrices

Example 1:

Using Gauss-Jordan method determine the inverse of the matrix

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

Solution:

$$\begin{array}{cccc|cccc}
2 & 1 & -1 & 0 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\
\boxed{1} & 0 & 1 & 2 & 0 & 0 & 1 & 0 \\
0 & 1 & -1 & -3 & 0 & 0 & 0 & 1 \\
\hline\hline
0 & \boxed{1} & -3 & -4 & 1 & 0 & -2 & 0 \\
0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\
\underline{1} & 0 & 1 & 2 & 0 & 0 & 1 & 0 \\
0 & 1 & -1 & -3 & 0 & 0 & 0 & 1 \\
\hline\hline
0 & \underline{1} & -3 & -4 & 1 & 0 & -2 & 0 \\
0 & 0 & -3 & -3 & 1 & 1 & -2 & 0 \\
\underline{1} & 0 & 1 & 2 & 0 & 0 & 1 & 0 \\
0 & 0 & 2 & \boxed{1} & -1 & 0 & 2 & 1 \\
\hline\hline
0 & \underline{1} & 5 & 0 & -3 & 0 & 6 & 4 \\
0 & 0 & \boxed{3} & 0 & -2 & 1 & 4 & 3 \\
\underline{1} & 0 & -3 & 0 & 2 & 0 & -3 & -2 \\
0 & 0 & 2 & \underline{1} & -1 & 0 & 2 & 1 \\
\hline\hline
0 & \underline{1} & 0 & 0 & 1/3 & -5/3 & -2/3 & -1 \\
0 & 0 & \underline{1} & 0 & -2/3 & 1/3 & 4/3 & 1 \\
\underline{1} & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & \underline{1} & 1/3 & -2/3 & -2/3 & -1
\end{array}$$

We have 4 marked elements, thus the inverse matrix exists.

Let us rearrange the rows such that the identity matrix stands to the left of the vertical line.

$$\begin{array}{cccc|cccc}
1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 1 & 0 & 0 & 1/3 & -5/3 & -2/3 & -1 \\
0 & 0 & 1 & 0 & -2/3 & 1/3 & 4/3 & 1 \\
0 & 0 & 0 & 1 & 1/3 & -2/3 & -2/3 & -1
\end{array}$$

Then the inverse of A stands on the area to the right of the vertical line:

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

One can see that the rank of the matrix A equals 4 (the number of marked elements).

Example 2:

Using Gauss-Jordan method determine the inverse of the matrix

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

Solution:

$$\begin{array}{cccc|cccc}
1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
2 & -1 & -1 & 1 & 0 & 1 & 0 & 0 \\
3 & -1 & 0 & 2 & 0 & 0 & 1 & 0 \\
-1 & 1 & \boxed{1} & 0 & 0 & 0 & 0 & 1 \\
\hline\hline
1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & \boxed{1} & 0 & 1 & 0 & 1 \\
3 & -1 & 0 & 2 & 0 & 0 & 1 & 0 \\
-1 & 1 & \underline{1} & 0 & 0 & 0 & 0 & 1 \\
\hline\hline
\boxed{1} & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & \underline{1} & 0 & 1 & 0 & 1 \\
1 & -1 & 0 & 0 & 0 & -2 & 1 & -2 \\
-1 & 1 & \underline{1} & 0 & 0 & 0 & 0 & 1 \\
\hline\hline
\underline{1} & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & \underline{1} & -1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & -1 & -2 & 1 & -2 \\
0 & 0 & \underline{1} & 0 & 1 & 0 & 0 & 1
\end{array}$$

No more pivot element can be chosen, thus the elimination has terminated with 3 marked elements. Since we have less than 4 marked elements, then A has no inverse. A is a singular matrix. One can see also that the rank of the matrix A equals 3 (the number of marked elements).

13.3. The Geometrical Meaning of the Determinant

In this section we will define the parallelepiped in a real Euclidean space, then we will discuss how can be calculated their „volume" with the help of the Gram-determinant. Since the letter V will be used for the volume, then the real Euclidean space will be denoted by E (instead of V) in this section.

One can see – using elementary geometry studies – that the parallelogram in the plane or in the space determined by the side-vectors $\underline{a}$ and $\underline{b}$ can be given as the point set

$$P = \{t_1\underline{a} + t_2\underline{b} \mid 0 \leq t_1 \leq 1,\ 0 \leq t_2 \leq 1\}.$$

Similarly can be given the parallelepiped (a parallelogram-based prism) determined by the edge-vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$ in the space:

$$P = \{t_1\underline{a} + t_2\underline{b} + t_3\underline{c} \mid 0 \leq t_1 \leq 1,\ 0 \leq t_2 \leq 1,\ 0 \leq t_3 \leq 1\}.$$

We know also that the area of the parallelogram equals the product of the length of any side and the altitude belonging to this side. Similarly, the volume of the parallelepiped equals the product of the area of any face and the altitude belonging to this face.

Using the above observations, we will define the „k-dimensional" parallelepiped (shortly named: k-parallelepiped, k-box) in the real Euclidean space E. Then we will define the k-dimensional measure of a k-parallelepiped. We will use the word „volume" instead of „measure" in each dimension.

Thus the 1-dimensional volume is the length, the 2-dimensional volume is the area, the 3-dimensional volume is the volume in the traditional sense.

13.1. Definition Let $a_1, \ldots, a_k$ be a vector system in the real Euclidean space E. The set

$$P(a_1, \ldots, a_k) := P_k := \left\{ \sum_{i=1}^{k} t_i a_i \in E \mid 0 \leq t_i \leq 1 \right\} \subset E$$

is called a k-box$or k$-parallelepiped in E spanned (or determined) by the vector system $a_1, \ldots, a_k$. This parallelepiped is named degenerate if the vectors $a_1, \ldots, a_k$ are linearly dependent, and it is named non-degenerate if the vectors $a_1, \ldots, a_k$ are linearly independent.

13.2. Remarks.

1. The 1-boxes are the line segments in E, the 2-boxes are the parallelograms in E.
2. If $k > \dim E$, then the vectors $a_1, \ldots, a_k$ are surely linearly dependent, thus the k-box is degenerate in this case.

Now we will define the volume of a k-box. The main idea of the definition is as follows:

The 1-dimensional volume (length) of a 1-box will be the norm (length) of its spanning vector.

The volume of a k-box will be the product of the $(k-1)$-dimensional volume of any „face" (which is a $(k-1)$-box) and the altitude belonging to this „face". Shortly:

$$\text{Volume} = \text{base} \cdot \text{altitude}$$

13.3. Definition Preserving the above notations let us define the (k-dimensional) volume of a k-box in the following recursive way:

1. $V_{P(a_1)} := \|\, a_1 \|$
2. $V_{P(a_1,\ldots,a_k)} := \underbrace{V_{P(a_1,\ldots,a_{k-1})}}_{\textit{base}} \cdot \underbrace{\|\, b \|}_{\textit{altitude}}$,

   where b denotes the orthogonal component of a_k relative to the subspace $\operatorname{Span}(a_1, \ldots, a_{k-1})$.

13.4. Remark. One can easily see that the volume of a degenerate parallelepiped equals 0.

The main result of this section is the following theorem:

13.5. Theorem *Preserving the above notations denote by*

$$G_k = G(a_1, \ldots, a_k)$$

the Gram-matrix generated by the vector system $a_1, \ldots, a_k \in E$ (see Definition 11.21). Then

$$\det G_k \geq 0, \quad \textit{and} \quad V_{P(a_1,\ldots,a_k)} = \sqrt{\det G_k}\,.$$

Proof. The proof is based on the following lemma:

13.6. Lemma Suppose that $k \geq 2$, and let

$$b := a_k - \sum_{i=1}^{k-1} \lambda_i a_i \,,$$

where the numbers $\lambda_1, \ldots, \lambda_{k-1} \in \mathbb{R}$ are arbitrary. Then

$$\det G(a_1, \ldots, a_{k-1}, a_k) = \det G(a_1, \ldots, a_{k-1}, b)\,.$$

Proof. The entries of the Gram-matrix G_k (see definition 11.21) are:

$$(G_k)_{ij} = \langle a_j, a_i \rangle\,.$$

Let us subtract from the k-th row of G_k the λ_i-multiple of the i-th row of G_k, for every $i = 1, \ldots k-1$. Thus the determinant and the first $k-1$ rows of the original matrix all are unchanged. The j-th entry of the k-th row will change to

$$(G_k)_{kj} - \sum_{i=1}^{k-1} \lambda_i \,(G_k)_{ij} = \langle a_j, a_k \rangle - \sum_{i=1}^{k-1} \lambda_i \langle a_j, a_i \rangle = \langle a_j, a_k - \sum_{i=1}^{k-1} \lambda_i a_i \rangle = \langle a_j, b \rangle\,.$$

Thus the entries of the resulted matrix $G_k^{'}$ are as follows:

$$\left(G_k^{'}\right)_{ij} = \begin{cases} (G_k)_{ij} & \text{if} \quad i = 1, \ldots, k-1 \\ \\ \langle a_j, b \rangle & \text{if} \quad i = k \end{cases}$$

Now let us make similar transformations with the last column of $G_k^{'}$. Let us subtract from the k-th column of $G_k^{'}$ the λ_j-multiple of the j-th column of $G_k^{'}$, for every $j = 1, \ldots k-1$. Thus the determinant and the first $k-1$ columns of the matrix $G_k^{'}$ all are unchanged. The i-th entry of the k-th column will change to

$$\left(G_k^{'}\right)_{ik} - \sum_{j=1}^{k-1} \lambda_j \left(G_k^{'}\right)_{ij} =$$

$$= \begin{cases} \langle a_k, a_i \rangle - \sum\limits_{j=1}^{k-1} \lambda_j \langle a_j, a_i \rangle = \langle a_k - \sum\limits_{j=1}^{k-1} \lambda_j a_j, a_i \rangle = \langle b, a_i \rangle & \text{if} \quad i = 1, \ldots, k-1 \\ \\ \langle a_k, b \rangle - \sum\limits_{j=1}^{k-1} \lambda_j \langle a_j, b \rangle = \langle a_k - \sum\limits_{j=1}^{k-1} \lambda_j a_j, b \rangle = \langle b, b \rangle & \text{if} \quad i = k \end{cases}$$

Thus the resulted matrix is as follows:

$$\begin{bmatrix} \langle a_1, a_1\rangle & \langle a_2, a_1\rangle & \dots & \langle a_{k-1}, a_1\rangle & \langle b, a_1\rangle \\ \langle a_1, a_2\rangle & \langle a_2, a_2\rangle & \dots & \langle a_{k-1}, a_2\rangle & \langle b, a_2\rangle \\ \vdots & \vdots & & \vdots & \vdots \\ \langle a_1, a_{k-1}\rangle & \langle a_2, a_{k-1}\rangle & \dots & \langle a_{k-1}, a_{k-1}\rangle & \langle b, a_{k-1}\rangle \\ \langle a_1, b\rangle & \langle a_2, b\rangle & \dots & \langle a_{k-1}, b\rangle & \langle b, b\rangle \end{bmatrix} = G(a_1, \dots, a_{k-1}, b)\,.$$

The determinant remained the same in each step, thus the proof of the lemma is completed. □

Let us return to the proof of the theorem. Since b is the orthogonal component of a_k relative to the subspace $\operatorname{Span}(a_1, \dots, a_{k-1})$, then b has the form

$$b := a_k - \sum_{i=1}^{k-1} \lambda_i a_i$$

with some coefficients λ_i. Thus – applying the previous lemma – we have:

$$\det G_k = \det G(a_1, \dots, a_{k-1}, a_k) = \det G(a_1, \dots, a_{k-1}, b)\,. \tag{13.1}$$

However, b is perpendicular to the subspace $\operatorname{Span}(a_1, \dots, a_{k-1})$, consequently

$$\langle a_j, b\rangle = 0 \quad (j = 1, \dots, k-1) \qquad \text{and} \qquad \langle b, a_i\rangle = 0 \quad (i = 1, \dots, k-1)\,.$$

By this reason we have:

$$G(a_1, \dots, a_{k-1}, b) = \begin{bmatrix} \langle a_1, a_1\rangle & \langle a_2, a_1\rangle & \dots & \langle a_{k-1}, a_1\rangle & 0 \\ \langle a_1, a_2\rangle & \langle a_2, a_2\rangle & \dots & \langle a_{k-1}, a_2\rangle & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \langle a_1, a_{k-1}\rangle & \langle a_2, a_{k-1}\rangle & \dots & \langle a_{k-1}, a_{k-1}\rangle & 0 \\ 0 & 0 & \dots & 0 & \langle b, b\rangle \end{bmatrix}.$$

Let us expand the determinant of this matrix along the last row. Then we have:

$$\det G(a_1, \dots, a_{k-1}, b) = \langle b, b\rangle \cdot \det G(a_1, \dots, a_{k-1}) = \|\, b\|^2 \cdot \det G_{k-1}\,.$$

After combining this with the equality (13.1) we obtain that:

$$\det G_k = \|\, b\|^2 \cdot \det G_{k-1}\,. \tag{13.2}$$

Since $\det G_1 = \| a_1\|^2$, then by mathematical induction we have

$$\det G_k \geq 0\,.$$

The formula for the volume can be obtained also by mathematical induction as follows:

The statement is true for $k = 1$, because

$$V_{P(a_1)} = \| a_1\| = \sqrt{\langle a_1,\, a_1\rangle} = \sqrt{G(a_1)}\,.$$

Then we step from $k-1$ to k in the following simple way:

$$V_{P(a_1,\,\dots,\,a_k)} = V_{P(a_1,\,\dots,\,a_{k-1})}\cdot\| b\| = \sqrt{\det G_{k-1}}\cdot\| b\| = \sqrt{\| b\|^2 \cdot \det G_{k-1}} = \sqrt{\det G_k}\,.$$

In the last step we have used the equality (13.2).

□

13.7. Remark. It can be proved that the value of the Gram-determinant $\det G(a_1,\, \dots,\, a_k)$ is independent of the order of the vectors $a_1,\, \dots,\, a_k$. By this reason, the volume of a k-box is independent of which its „face” is chosen as the base.

Let us apply our results in the real Euclidean space $E = \mathbb{R}^n$. First we have to investigate how can we calculate easily the Gram-matrix of a vector system $a_1, \dots,\, a_k \in \mathbb{R}^n$.

So let $a_1, \dots,\, a_k \in \mathbb{R}^n$, and denote by $G_k \in \mathbb{R}^{k\times k}$ the Gram-matrix of this vector system. Let $A \in \mathbb{R}^{n\times k}$ be the matrix whose column vectors are $a_1, \dots,\, a_k$:

$$A := [a_1, \dots,\, a_k] \in \mathbb{R}^{n\times k}\,.$$

One can prove easily – using the rules of the matrix product – that

$$(G_k)_{ij} = \langle a_j,\, a_i\rangle = \left(A^T A\right)_{ij}\,,$$

thet is $G_k = A^T A \in \mathbb{R}^{k\times k}$.

Thus we have the following formula for the (k-dimensional) volume of a k-box in $\mathbb{R}^n$:

$$V_{P_k} = \sqrt{\det(A^T A)}\,.$$

Now we highlight the special case $k = n$. Then the matrices A^T and A are both square matrices of the same size. Thus they both have determinants, not only their product $A^T A$. Applying the theorem about the determinant of the product of matrices we have

$$V_{P_n} = \sqrt{\det(A^T A)} = \sqrt{(\det A^T)\cdot(\det A)} = \sqrt{(\det A)\cdot(\det A)} = \sqrt{(\det A)^2} = |\det A|\ .$$

Now we have arrived to the geometrical meaning of the determinant:

The absolute value of an $n \times n$ determinant equals the volume of the n-dimensional volume of the n-dimensional parallelepiped in $\mathbb{R}^n$.

Let us see some numerical examples:

Example 1

Compute the area T of the parallelogram in $\mathbb{R}^3$ spanned by the side-vectors $a = (3, -1, 2)$ and $b = (-1, 2, 1)$.

Solution

In our example: $n = 3$, $k = 2$.

Let A be the matrix whose column vectors are a and b:

$$A = [a\ b] = \begin{bmatrix} 3 & -1 \\ -1 & 2 \\ 2 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}\,.$$

Then we can calculate easily that

$$A^T A = \begin{bmatrix} 14 & -3 \\ -3 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 2}\,, \qquad \det(A^T A) = 75\,.$$

Hence we have:

$$T = \sqrt{\det(A^T A)} = \sqrt{75} = 5\sqrt{2}\,.$$

13.8. Remark. The 4 vertices of the above parallelogram are:

$$0 = (0,0,0), \quad a = (3,-1,2), \quad b = (-1,2,1), \quad a+b = (2,1,3)\,.$$

Example 2

Compute the volume V of the parallelepiped in $\mathbb{R}^3$ spanned by the edge-vectors $a = (1, 0, -1)$, $b = (-1, 1, 3)$ and $c = (2, 4, 1)$.

Solution

In our example: $n = 3$, $k = 3$.

Let A be the matrix whose column vectors are a, b, c:

$$A = [a\ b\ c] = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 4 \\ -1 & 3 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}, \qquad \det A = -5\,.$$

Hence we have:

$$V = |\det A| = 5\,.$$

13.9. Remark. The 8 vertices of the above parallelogram are:

$$0 = (0,0,0), \quad a = (1,0,-1), \quad b = (-1,1,3), \quad c = (2,4,1),$$

$$a+b = (0,1,2), \quad a+c = (3,4,0), \quad b+c = (1,5,4), \quad a+b+c = (2,5,3)\,.$$