Linear Algebra Lecture Schemes
(with Control Questions)
(and with Homework)[1]

István CSÖRGŐ

October, 2016

[1]A jegyzet az ELTE Informatikai Kar 2016. évi Jegyzetpályázatának támogatásával készült

Contents

Preface

This work is a member of the author's series of lecture schemes published in the Digital Library of the Faculty of Informatics. These lecture schemes are addressed to the Computer Science BSc students of Linear Algebra and of Analysis. All these works are based on the lectures and practices of the above subjects given by the author for decades in the English Course Education.

The recent work contains the topics of the subject Linear Algebra. It starts with matrices and determinants, then it contains the following topics: vector spaces, system of linear equations, inner product spaces, self-adjoint matrices, quadratic forms.

It builds intensively on the preliminary subjects, as Mathematics in secondary school and Precalculus Practices.

This work uses the usual mathematical notations. The set of natural numbers ($\mathbb{N}$) will begin with 1. The symbol $\mathbb{K}$ will denote one of the sets of real numbers ($\mathbb{R}$) or of the complex numbers ($\mathbb{C}$).

The topics are explained on a weekly basis. Every chapter contains the material of an educational week. The control questions and the homework related to the topic can be found at the end of the chapter.

Thanks to my teachers and colleagues, from whom I learned a lot. I thank the lectors of this textbook – assoc. prof. Dr. Lajos László and assoc. prof. Dr. Ágnes Bércesné Novák – for their thorough work and valuable advice.

Budapest, October 2016

István CSÖRGŐ

1. Lesson 1

1.1. Complex Numbers

In our Linear Algebra studies we will use the real and the complex numbers as scalars. The real numbers are supposed to be familiar from the secondary school. Now we will collect shortly the most important knowledge about the complex numbers.

Axiomatic Definition:
Let i denote the "number" whose square equals -1. More precisely, we use $i^2 = -1$ about the symbol i.

1.1. Definition The set of complex numbers consists of the expressions $a + bi$ where a and b are real numbers:

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

The operations $+$ (addition) and $\cdot$ (multiplication) are defined as follows: let's compute with complex numbers as with binomial expressions and write in every case -1 instead of i^2. The number i is called: imaginary unit.

Let's collect the complex basic operations in algebraic form:

1. $(a + bi) + (c + di) = (a + c) + (b + d)i,$
2. $(a + bi) - (c + di) = (a - c) + (b - d)i,$
3. $(a + bi) \cdot (c + di) = ac + bci + adi + bdi^2 = (ac - bd) + (bc + ad)i,$
4. At the division multiply the numerator and the denominator by the complex conjugate (see below) of the denominator:

$$\frac{a+bi}{c+di} = \frac{(a+bi)\cdot(c-di)}{(c+di)\cdot(c-di)} = \frac{ac+bci-adi-bdi^2}{c^2-d^2i^2} = \\ = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}\cdot i$$

1.2. Definition Let $z = a + bi \in \mathbb{C}$. Then

1. $\operatorname{Re} z := a$ (real part),
2. $\operatorname{Im} z := b$ (imaginary part),

3. $\overline{z} := a - bi$ (complex conjugate),

4. $|z| := \sqrt{a^2 + b^2}$ (absolute value or modulus).

Some important properties of the introduced operations:

1.3. Theorem

1. *$\mathbb{C}$ is a field with respect to the operations $+$ and $\cdot$*

2. $\overline{z + w} = \overline{z} + \overline{w}$

3. $\overline{z - w} = \overline{z} - \overline{w}$

4. $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$

5. $\overline{\left(\dfrac{z}{w}\right)} = \dfrac{\overline{z}}{\overline{w}}$

6. $\overline{\overline{z}} = z$

7. $|\overline{z}| = |z|$

8. $|z + w| \leq |z| + |w|$

9. $|z \cdot w| = |z| \cdot |w|$

10. $\left|\dfrac{z}{w}\right| = \dfrac{|z|}{|w|}$

Proof. On the lecture. □

From now on $\mathbb{K}$ denotes the set $\mathbb{R}$ or $\mathbb{C}$.

1.2. Matrices

If we want to define the precise concept of matrix, then we have to define it as a special function:

1.4. Definition Let $m, n \in \mathbb{N}$. The $m \times n$ matrix (over the number field $\mathbb{K}$) is a mapping defined on the set $\{1, \ldots m\} \times \{1, \ldots n\}$ and maps into $\mathbb{K}$:

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

Denote by $\mathbb{K}^{m \times n}$ the set of $m \times n$ matrices. The number $A(i, j)$ is called the j-th element of the i-th row and is denoted by a_{ij} or $(A)_{ij}$. The elements of the matrix are called entries. The matrix is called square matrix (of order n) if $m = n$.

Usually the matrices are given as a rectangular array (hence the concept row and column):

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

The entries a_{11}, a_{22}, ... are called diagonal elements or simply diagonal. (main diagonal). Naturally, it coincides with the common concept of „diagonal" only for square matrices.

Some special matrices: zero matrix, row matrix, column matrix, triangular matrix (lower, upper), diagonal matrix, identity matrix.

1.5. Definition Operations with matrices:

1. Addition: Let $A, B \in \mathbb{K}^{m\times n}$. Then

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

2. Scalar multiple: Let $A \in \mathbb{K}^{m\times n}$ and $\lambda \in \mathbb{K}$. Then

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

3. Product: Let $A \in \mathbb{K}^{m\times n}$, $B \in \mathbb{K}^{n\times p}$. Then the product of A and B is as follows:

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

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

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

5. Adjoint or Hermitian adjoint: Let $A \in \mathbb{C}^{m\times n}$. Then

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

1.3. Properties of Matrix Operations

1.6. Theorem *[Sum and Scalar Multiple] Let A, B, $C \in \mathbb{K}^{m\times n}$, λ, $\mu \in \mathbb{K}$. Then*

1. $A+B = B+A$.
2. $(A+B)+C = A+(B+C)$.

3. $\exists\, 0 \in \mathbb{K}^{m\times n}\ \forall\, M \in \mathbb{K}^{m\times n}: \quad M + 0 = M.$

   It can be proved that 0 *is unique and it is the zero matrix.*

4. $\forall\, M \in \mathbb{K}^{m\times n}\ \exists\, (-M) \in \mathbb{K}^{m\times n}: \quad M + (-M) = 0.$

   It can be proved that $-M$ *is unique and its elements are the opposite ones of* M.

5. $(\lambda\mu)A = \lambda(\mu A) = \mu(\lambda A)$.

6. $(\lambda + \mu)A = \lambda A + \mu A$.

7. $\lambda(A + B) = \lambda A + \lambda B$.

8. $1A = A$.

Proof. Every statement can be easily verified by the help of „entry-vise" operations. □

This theorem shows us that $\mathbb{K}^{m\times n}$ is a vector space over $\mathbb{K}$. The definition and study of the vector space will follow later.

1.7. Theorem *[Product]*

1. *Associative law:*

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

2. *Distributive laws:*

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

3. *Multiplication with the identity matrix. Denote by* I *the identity matrix of suitable size. Then:*

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

Proof. On the lecture. □

You can easily consider that the multiplication of matrices is inner operation if and only if $m = n$ that is in the set of square matrices. In this case we can establish that $\mathbb{K}^{n\times n}$ is a ring with identity element. This ring is not commutative and it has zero divisors as the following examples show:

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

The connection between the product and the scalar multiple can be described by the following theorem:

1.8. Theorem

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

Proof. $\square$

This identity – and the ring and vector space structure of $\mathbb{K}^{n\times n}$ – shows us that $\mathbb{K}^{n\times n}$ is an algebra with identity element over $\mathbb{K}$.

1.9. Theorem *[Transpose, Adjoint] Let $A,\, B \in \mathbb{K}^{m\times n}$, $\lambda \in \mathbb{K}$. Then*

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.

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

Proof. On the lecture. $\square$

1.4. Control Questions

1. Starting out from $\mathbb{R}$ (the set of real numbers) define the set of complex numbers
2. Define the addition and multiplication in $\mathbb{C}$
3. How can we compute the difference and the quotient of the complex numbers $a + bi$ and $c + di$?
4. Define the followings: real part, imaginary part, conjugate and the absolute value of a complex number
5. List 4 properties of the operations in $\mathbb{C}$
6. Define the concept of matrix

7. Define the addition of matrices and list the most important properties of this operation

8. Define the scalar multiplication of a matrix and list the most important properties of this operation

9. Define the product of matrices and list the most important properties of this operation

10. Define the Transpose and the Hermitian adjoint of a matrix and list the most important properties of these operations

1.5. Homework

1. Let $z = 3 + 2i$, $w = 5 - 3i$, $u = -2 + i$. Compute:
$$z + w,\ z - w,\ zw,\ \frac{z}{w},\ \frac{2z^2 + 3w}{1 + u}.$$

2. Let
$$A = \begin{bmatrix} 1 & 1 & 5 \\ -3 & 0 & 1 \\ 0 & 1 & 2 \\ 2 & -4 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -4 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 2 & 4 & 0 \\ -1 & 1 & 1 \\ 3 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$
Compute:
$$A + 2B - C, \quad A^T B, \quad (AB^T)C$$

3. Let
$$A = \begin{bmatrix} 1 - i & 2 + i & 3 + i \\ 0 & 1 + i & 1 \\ 2 + i & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 + i & 2 + i & 1 + 3i \\ 4 - i & 0 & -i \\ 0 & 1 & i \end{bmatrix}.$$
Compute:
$$2A - B, \quad AB, \quad AB^*$$

2. Lesson 2

2.1. Decomposition of a matrix into Blocks

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 itself.

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.

2.2. Determinants

If we delete some rows and/or columns of a matrix then we obtain a submatrix of the original matrix. Now for us will be enough to delete one row and one column from a square matrix. The resulting submatrix will be called minor matrix.

2.1. Definition (Minor Matrix) Let $A \in \mathbb{K}^{n\times n}$ and $(i,j) \in \{1, \dots n\}\times\{1, \dots n\}$ a fixed index pair. The minor matrix of the position (i,j) is denoted by A_{ij} and is defined as follows:

$$(A_{ij})_{kl} := \begin{cases} a_{kl} & \text{if} \quad 1 \le k \le i-1, \;\; 1 \le l \le j-1 \\ a_{k,l+1} & \text{if} \quad 1 \le k \le i-1, \;\; j \le l \le n-1 \\ a_{k+1,l} & \text{if} \quad i \le k \le n-1, \;\; 1 \le l \le j-1 \\ a_{k+1,l+1} & \text{if} \quad i \le k \le n-1, \;\; j \le l \le n-1\,. \end{cases}$$

Obviously $A_{ij} \in \mathbb{K}^{(n-1)\times(n-1)}$. In words: the minor matrix is the remainder submatrix after deletion the i-th row and the j-th column of A.

2.2. Examples

$$\text{If} \qquad A = \begin{bmatrix} 3 & 5 & -2 & 8 & -1 \\ 0 & 3 & -1 & 1 & 2 \\ 2 & 1 & 2 & 3 & 4 \\ 7 & 1 & -3 & 5 & 8 \end{bmatrix} \qquad \text{then} \qquad A_{34} = \begin{bmatrix} 3 & 5 & -2 & -1 \\ 0 & 3 & -1 & 2 \\ 7 & 1 & -3 & 8 \end{bmatrix}$$

After this short preliminary let us define recursively the function $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$ as follows:

2.3. 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 the number $a'_{ij} := (-1)^{i+j} \cdot \det(A_{ij})$ is called signed subdeterminant or cofactor (assigned to the position (i, j).

The number $\det(A)$ is called the determinant of the matrix A and is denoted by

$$\det(A), \qquad \det A, \qquad |A|, \qquad \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

We say that we have defined the determinant by expansion along the first row. According to the last notation we can speak about the elements, rows, columns, e.t.c. of a determinant.

2.4. Examples

Let us study some important special cases:

1. The 1×1 determinant: for example $\det([5]) = 5$.

2. The 2×2 determinant:

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

   so a 2×2 determinant can be computed by subtracting from the product of the entries in the diagonal (a_{11}, a_{22}) the product of the entries of the other diagonal (a_{12}, a_{21}).

3. Applying $n-1$ times the recursive step of the definition we obtain that the determinant of a lower triangular matrix equals the product of its diagonal elements:
$$\begin{vmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{32} & a_{33} & \dots & 0 \\ & \vdots & & & \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} = a_{11} \cdot a_{22} \cdot \ \dots \ \cdot a_{nn} \,.$$

4. Immediately follows from the previous example that the determinant of the unit matrix equals 1.

2.3. The properties of the Determinants

2.5. Theorem 1. *The determinant can be expanded by its any row and by its any column that is for every* $r,\ s \in \{1, \dots, n\}$ *holds:*
$$\det(A) = \sum_{j=1}^{n} a_{rj} \cdot a'_{rj} = \sum_{i=1}^{n} a_{is} \cdot a'_{is} \,.$$

2. $\det(A) = \det(A^T)$ $(A \in \mathbb{K}^{n \times n})$. *An important corollary of this that the determinant of an upper triangular matrix equals the product of its diagonal elements.*

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

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

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:*

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

Proof.

1. It has a complicated proof, we don't prove it.

2. Immediately follows from the previous statement.

3. Expand the determinant by its 0-row.

4. Use mathematical induction by n. For $n = 2$ the statement can be checked immediately. To deduce from $n-1$ to n denote by r and s the indices of the two (different) rows that are interchanged in the $n \times n$ matrix A and denote by B the resulted matrix after interchanging. Expand $\det(A)$ and $\det(B)$ along their kth row where $k \neq r$, $k \neq s$. Then the elements are the same (a_{kj}) in both expansion but the cofactors – by the inductional assumption – are opposite. So the two expansions are opposite.

5. Interchange the two equal rows. This implies $\det(A) = -\det(A)$. After rearrangement we obtain $\det(A) = 0$.

6. Denote by r the index of the row in which every entry is multiplied by λ. Expand the new determinant by its r-th row and take out the common factor λ from the expansion sum.

7. Immediately follows from the previous property if you apply it for every row.

8. Immediately follows from the previous property and the „two rows are equal" property.

9. Expand the new determinant $\det(C)$ by its r-th row, apply the distributive law in every term of expansion sum and group this sum into two sub-sums. The sum of the first terms gives $\det(A)$, the sum of the second terms gives $\det(B)$.

10. Immediately follows from the previous two properties.

11. It has a complicated proof, we don't prove it.

□

2.4. The Inverse of a Matrix

In this section we will extend the concept of „reciprocal" and „division" from numbers to matrices. Instead of „reciprocal" will be used the name „inverse" and instead of „division" will be used the name „multiplication by inverse".

2.6. Definition Let $A \in \mathbb{K}^{n\times n}$ and denote by I the identity matrix in $\mathbb{K}^{n\times n}$. Then A is called

1. invertible from the right if $\exists\, C \in \mathbb{K}^{n\times n}$ such that $AC = I$. In this case C is called a right-hand inverse of A.

2. invertible from the left if $\exists\, D \in \mathbb{K}^{n\times n}$ such that $DA = I$. In this case D is called a left-hand inverse of A.

3. invertible if $\exists\, C \in \mathbb{K}^{n\times n}$ such that $AC = I$ and $CA = I$. In this case C is unique and is called the inverse of A and is denoted by A^{-1}.

2.7. Definition A matrix in $\mathbb{K}^{n\times n}$ is called regular if it is invertible. A matrix in $\mathbb{K}^{n\times n}$ is called singular if it is not invertible.

In the following part of the section we characterize the regular and the singular matrices with the help of their determinants.

2.8. Theorem *A matrix $A \in \mathbb{K}^{n\times n}$ is invertible from the right if and only if $\det(A) \neq 0$. In this case a right-hand inverse can be given as*

$$C := \frac{1}{\det(A)} \cdot \widetilde{A}\,, \qquad \textit{where} \qquad (\widetilde{A})_{ij} := a'_{ji}\,.$$

Remember that here a'_{ji} denotes the cofactor assigned to the position (j, i).

Proof. Assume first that A is invertible from the right and denote by C a right-hand inverse. Then:

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

From this equality it follows immediately that $\det(A) \neq 0$. Remark that we obtained another result too: $\det(C) = \dfrac{1}{\det(A)}$.

Conversely suppose that $\det(A) \neq 0$ and let C be the following matrix:

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

We will show that $AC = I$. Really:

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

First suppose that $i = j$. Then the last sum equals 1 because – using the expansion of the determinant along its i-th row– :

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

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 $AC = I$. $\square$

The existence of the left-hand inverse can reduce – with the help of the transpose – to the case of right-hand inverse:

2.9. Theorem *A matrix $A \in \mathbb{K}^{n\times n}$ is invertible from the left if and only if $\det(A) \neq 0$. In this case a left-hand inverse of A can be given as the transpose of a right-hand inverse of A^T.*

Proof.

$$\begin{aligned}\det(A) \neq 0 \quad &\Longleftrightarrow \quad \det(A^T) \neq 0 \quad \Longleftrightarrow \quad \exists\, D \in \mathbb{K}^{n\times n}: \quad A^T D = I \quad \Longleftrightarrow \\ &\Longleftrightarrow \quad \exists\, D \in \mathbb{K}^{n\times n}: \quad (A^T D)^T = D^T A = I^T = I.\end{aligned}$$

$\square$

Up to this point we have used intentionally the phrases „a right-hand inverse" and „a left-hand inverse" instead of „the right-hand inverse" and „the left-hand inverse" because their uniqueness was not proved. In the following theorem we state the uniqueness:

2.10. Theorem *Let $A \in \mathbb{K}^{n\times n}$ and $C \in \mathbb{K}^{n\times n}$ be a right-hand inverse of A, $D \in \mathbb{K}^{n\times n}$ be a left-hand inverse of A. Then $C = D$.*

Proof.

$$D = DI = D(AC) = (DA)C = IC = C, \quad \text{consequently} \quad C = D\,.$$

□

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

1. Suppose that $\det A = 0$. Then A has neither left-hand inverse nor right-hand inverse (it is invertible neither from the left nor from the right).

2. Suppose that $\det A \neq 0$. Then A is invertible from the left as well as it is invertible from the right. Any left-hand inverse equals any right-hand inverse, thus both inverses are unique and equal to each other. That means that A has a unique inverse and its inverse is

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

3. It follows immediately from the previous considerations that if we want to prove that a matrix C is the inverse of A then it is enough to check only one of the relations $AC = I$ or $CA = I$, the other one holds „automatically”.

4. It follows also from the previous considerations that if $A \in \mathbb{K}^{n\times n}$ then

   - A is regular if and only if $\det A \neq 0$,
   - A is singular if and only if $\det A = 0$.

Applying our results for 2×2 matrices we obtain easily the following theorem:

2.12. Theorem *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{K}^{2\times 2}$. *Then* A *is invertible if and only if* $ad - bc \neq 0$. *In this case:*

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

2.5. Control Questions

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

2.6. Homework

1. Compute the 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 of

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

and check that the products of the matrices with their inverses are really the identity matrices.

3. 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 no one 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. Lesson 3

3.1. Cramer's Rule

In this section we will study the solution of special system of linear equations. A system of linear equations having n equations and n unknowns can be written in the following form:

$$\begin{array}{ccccc} a_{11}x_1 & + \ldots & + a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + \ldots & + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & & \vdots & \\ \hline a_{n1}x_1 & + \ldots & + a_{nn}x_n & = & b_n \end{array},$$

where the coefficients $a_{ij} \in \mathbb{K}$ and the constants on the right side b_i are given. We are looking for the possible values of the unknowns $x_1, \ldots, x_n$ such that after substitution them in the equations each equation will be true.

We can abbreviate the system if we collect the coefficients, the constants on the right side and the unknowns into matrices:

$$A := \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \in \mathbb{K}^{n \times n}, \qquad B := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{K}^{n \times 1}, \qquad X := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^{n \times 1}.$$

Then the system of linear equations can be written as a matrix equation

$$AX = B\,.$$

3.1. Theorem *[Cramer's Rule]*

Suppose that $\det A \neq 0$*. Then there exists uniquely a matrix* $X \in \mathbb{K}^{n \times 1}$ *such that* $AX = B$*. The* k*-th element of the single column of this matrix is:*

$$x_k = \frac{\det(A_k)}{\det(A)}, \quad \textit{where} \quad (A_k)_{ij} := \begin{cases} a_{ij} & \textit{if} \quad j \neq k \\ \\ b_i & \textit{if} \quad j = k\,. \end{cases}$$

In words: the matrix A_k *can be obtained by replacing the* k*-th column of* A *to the column matrix* B*. Here* $k = 1, \ldots n$*.*

Proof. Since $\det(A) \neq 0$ so A is invertible. Moreover:

$$\begin{aligned} AX = B \quad &\iff \quad A^{-1}(AX) = A^{-1}B \quad \iff \quad (A^{-1}A)X = A^{-1}B \quad \iff \\ &\iff \quad Ix = A^{-1}B \quad \iff \quad X = A^{-1}B\,, \end{aligned}$$

that shows that the matrix equation (consequently the system of linear equations) has only one solution: $X = A^{-1}B$. Using the formula for the inverse matrix – the k-th component of X is:

$$x_k = (A^{-1}B)_{k1} = \frac{1}{\det(A)} \cdot (\widetilde{A}B)_{k1} = \frac{1}{\det(A)} \cdot \sum_{i=1}^{n} (\widetilde{A})_{ki} b_i =$$
$$= \frac{1}{\det(A)} \cdot \sum_{i=1}^{n} a'_{ik} b_i = \frac{1}{\det(A)} \cdot \det(A_k).$$

In the last step we have used the expansion of $\det(A_k)$ along its k-th column. Here $k = 1, \ldots n$. □

Remark that the Cramer's rule is effective only for systems of low sizes. For the systems of greater sizes there exist more effective methods. One of these method is given later (Elementary Basis Transformation). Other effective methods will be given in the subject „Numerical Methods".

3.2. Control Questions

1. State and prove the Cramer's Rule

3.3. Homework

1. Solve the linear equation systems using the Cramer's Rule

$$a) \quad \begin{array}{l} 7x - 2y = 3 \\ 3x + y = 5 \\ \hline \end{array} \qquad\qquad b) \quad \begin{array}{l} x - 4y + z = 6 \\ 4x - y + 2z = -1 \\ 2x + 2y - 3z = -20 \\ \hline \end{array}$$

4. Lesson 4

4.1. Vector Spaces

In this section we introduce the central concept of linear algebra: the concept of vector space. This is an extension of the concept of geometrical vectors.

4.1. Definition Let $V \neq \emptyset$ and let $V \times V \ni (x, y) \mapsto x + y$ (addition), $\mathbb{K} \times V \ni (\lambda, x) \mapsto \lambda \cdot x = \lambda x$ (multiplication by scalar) be two mappings (operations). Suppose that

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

II.
1. $\forall\, (\lambda,\, x) \in \mathbb{K} \times V : \quad \lambda x \in V$ (closure under multiplication by scalar)
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. $\forall\, x \in V : \quad 1x = x$

In this case we say that V is a vector space over $\mathbb{K}$ with the two given operations (addition and multiplication by scalar). The elements of V are called vectors, the elements of $\mathbb{K}$ are called scalars. $\mathbb{K}$ is called the scalar region of V. The above written ten requirements are the axioms of the vector space.

Remark that applying several times the associative law of addition we can define the sums of several terms:

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

Let us see some examples for vector space:

4.2. 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 $\mathbb{C}$ and $\mathbb{C}$ is vector space over $\mathbb{R}$.

4. 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.

5. Let
$$\mathbb{K}^n := \underbrace{\mathbb{K}\times\mathbb{K}\ldots\mathbb{K}} = \{x=(x_1,\,x_2,\,\ldots\,x_n)\ \mid\ x_i\in\mathbb{K}\}$$
be the set of n-term sequences (ordered n-tuples). Let us define the operations „componentwise”:
$$(x+y)_i := x_i+y_i \quad (i=1,\,\ldots\,n); \qquad\qquad (\lambda\cdot x)_i := \lambda\cdot x_i \quad (i=1,\,\ldots\,n)\,.$$
One can check that the axioms are satisfied, so $\mathbb{K}^n$ is a vector space over $\mathbb{K}$.

   Remark that

   - $\mathbb{R}^1$ can be identified with $\mathbb{R}$ or with the vector space of the points (vectors) in the straight line.
   - $\mathbb{R}^2$ can be identified with the vector space of the points (vectors) in the plane.
   - $\mathbb{R}^3$ can be identified with the vector space of the points (vectors) in the space.

6. It follows immediately from the properties of the matrix operations that (for any fixed $m,\,n\in\mathbb{N}$) the set of m by n matrices $\mathbb{K}^{m\times n}$ is a vector space

over $\mathbb{K}$. The operations are the usual matrix addition and multiplication by scalar.

Remark that

- $\mathbb{K}^{1\times 1}$ can be identified with $\mathbb{K}$.
- $\mathbb{K}^{m\times 1}$ (column matrices) can be identified with $\mathbb{K}^m$.
- $\mathbb{K}^{1\times n}$ (row matrices) can be identified with $\mathbb{K}^n$.

7. Now follows a generalization of $\mathbb{K}^n$ and $\mathbb{K}^{m\times n}$.

   Let $H \neq \emptyset$ and V be the set of all functions that are defined on H and map into $\mathbb{K}$. A common notation for the set of these functions is $\mathbb{K}^H$. So

   $$V = \mathbb{K}^H = \{f : H \to \mathbb{K}\} .$$

   Define the operations „pointwise”:

   $$(f{+}g)(h) := f(h){+}g(h); \qquad (\lambda f)(h) := \lambda f(h) \qquad (h \in H) \quad (f, g \in V;\ \lambda \in \mathbb{K}) .$$

   Then – one can check the axioms – V is a vector space over $\mathbb{K}$.

   Remark that

   - $\mathbb{K}^n$ can be identified with $\mathbb{K}^H$ if $H = \{1, 2, \dots n\}$.
   - $\mathbb{K}^{m\times n}$ can be identified with $\mathbb{K}^H$ if $H = \{1, 2, \dots m\} \times \{1, 2, \dots n\}$.

We can define other operations in the vector space V:

Subtraction: $x - y := x + (-y)$ $(x, y \in V)$.

Division by scalar: $\dfrac{x}{\lambda} := \dfrac{1}{\lambda} \cdot x$ $(x \in V,\ \lambda \in \mathbb{K},\ \lambda \neq 0)$.

In the following theorem we collect some simple but important properties of vector spaces.

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

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

4.2. Control Questions

1. Define the concept of vector space
2. Give 3 examples for vector space
3. State some elementary properties of a vector space and prove one of them

4.3. Homework

1. Let $V = \mathbb{R}^2$ with the following operations:

   $$x + y := (x_1 + y_1,\, x_2 + y_2) \quad \text{and} \quad \lambda x := (0, \lambda x_2)$$

   where $x = (x_1, x_2)$, $y = (y_1, y_2) \in V$, $\lambda \in \mathbb{K}$.

   Is V vector space or not? Find the vector space axioms that hold and find the ones that fail.

2. (An unusual vector space.) Let V be the set of positive real numbers:

   $$V := \mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}\,.$$

   Let us introduce the vector operations in V as follows:

   $$x + y := xy \quad (x, y \in V) \qquad\qquad \lambda x := x^\lambda \quad (\lambda \in \mathbb{R},\ x \in V)\,.$$

   (On the right sides of the equalities xy and x^λ are the usual real number operations.) Prove that V is a vector space over $\mathbb{R}$ with the above defined vector operations. What is the zero vector in this space? What is the opposite of $x \in V$? What do the statements in the last theorem of the section mean in this interesting vector space?

5. Lesson 5

5.1. Subspaces

The subspaces are vector spaces lying in another vector space. In this section V denotes a vector space over $\mathbb{K}$.

5.1. Definition Let $W \subseteq V$. W is called a subspace of V if W is itself a vector space over $\mathbb{K}$ under the vector operations (addition and multiplication by scalar) defined on V.

By this definition if we want to decide about a subset of V that it is a subspace or not, we have to discuss the ten vector space axioms. In the following theorem we will prove that it is enough to check only two axioms.

5.2. Theorem *Let $\emptyset \neq W \subseteq V$. Then W is a subspace of V if and only if:*

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

In words: the subset W is closed under the addition and multiplication by scalar in V.

Proof. The two given conditions 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 the 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 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 opposite of x and the opposite vectors in V and W are the same.

□

5.3. 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 no subspace. Similar considerations are valid for the opposite vector too.

Using the above theorem the following examples for subspaces can be easily verified.

5.4. Examples

1. The zero vector space $\{0\}$ and V itself both are subspaces in V. They are called 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.

4. In the vector space $\mathbb{K}^{\mathbb{K}}$ (the collection of functions $f : \mathbb{K} \to \mathbb{K}$) the following subsets form subspaces:

   - $\mathcal{P} := \mathcal{P}(\mathbb{K}) := \{f : \mathbb{K} \to \mathbb{K} \mid f \text{ is polynomial}\}$. This subspace $\mathcal{P}$ is called the vector space of polynomials.
   - Fix a nonnegative integer $n \in \mathbb{N} \cup \{0\}$ and let

     $$\mathcal{P}_n := \mathcal{P}_n(\mathbb{K}) := \{f \in \mathcal{P}(\mathbb{K}) \mid f = 0, \text{ or } \quad \deg f \leq n\}.$$

     Then $\mathcal{P}_n$ is a subspace that is called the vector space of polynomials of at most degree n. Remark that although the zero polynomial has no degree it is contained in $\mathcal{P}_n$.

   In connection with the polynomial spaces it is important to see that

   $$\{0\} \subset \mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}, \qquad \bigcup_{n=0}^{\infty} \mathcal{P}_n = \mathcal{P}.$$

5.2. Linear Combinations and Generated Subspaces

5.5. Definition Let $k \in \mathbb{N}$, $x_1, \ldots, x_k \in V$, $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$. The vector (and the expression itself)

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

is called the linear combination of the vectors $x_1, \ldots, x_k$ with coefficients $\lambda_1, \ldots, \lambda_k$. The linear combination is called trivial if every coefficient is zero. The linear combination is called nontrivial if at least one of its coefficients is nonzero.

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

One can prove simply by mathematical induction that a nonempty subset $W \subseteq V$ is subspace if and only if for every $k \in \mathbb{N}$, $x_1, \ldots, x_k \in W$, $\lambda_1, \ldots \lambda_k \in \mathbb{K}$:

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

In other words: the subspaces are exactly the subsets of V closed under linear combinations.

Let $x_1, x_2, \ldots, x_k \in V$ be a system of vectors. Let us define the following subset of V:

$$W^* := \left\{ \sum_{i=1}^{k} \lambda_i x_i \;\mid\; \lambda_i \in \mathbb{K} \right\}. \tag{5.1}$$

So the elements of W^* are the possible linear combinations of $x_1, x_2, \ldots, x_k$.

5.6. Theorem *1. W^* is subspace in V.*

2. W^* covers the system $x_1, x_2, \ldots, x_k$ that is $\quad \forall\, i : \quad x_i \in W^*$.

3. W^* is the minimal subspace among the subspaces that cover $x_1, x_2, \ldots, x_k$. More precisely:

$$\forall\, W \subseteq V, W \textit{ is subspace}, x_i \in W : \quad W^* \subseteq W\,.$$

Proof.

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

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

On the other hand for every $\lambda \in \mathbb{K}$:

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

So W^* is really a subspace in V.

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

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

3. Let W be a subspace described in the theorem and let $a = \sum_{i=1}^{k} \lambda_i x_i \in W^*$. Since W covers the system so

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

But the subspace W is closed under linear combination, which implies $a \in W$. So really $W^* \subseteq W$.

$\square$

5.7. Definition The above defined subspace W^* is called the subspace spanned (or generated) by the vector system $x_1, x_2, \ldots, x_k$ and is denoted by $\mathrm{Span}\,(x_1, x_2, \ldots, x_k)$. Sometimes we say shortly that W^* is the span of $x_1, x_2, \ldots, x_k$. The system $x_1, x_2, \ldots, x_k$ is called the generator system (or: spanning set) of the subspace W^*. Sometimes we say that $x_1, x_2, \ldots, x_k$ spans W^*.

Remark that a vector is contained in $\mathrm{Span}\,(x_1, x_2, \ldots, x_k)$ if and only if it can be written as linear combination of $x_1, x_2, \ldots, x_k$.

5.8. Examples

1. Let v be a vector in the vector space of plane vectors ($\mathbb{R}^2$). Then

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

Using geometrical methods one can prove that in the vector space of plane vectors any two nonparallel vectors form a generator system.

2. Let v_1 and v_2 be two vectors in the vector space of space vectors ($\mathbb{R}^3$). Then

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

Using geometrical methods one can prove that in the vector space of space vectors any three vectors that are not in the same plane form a generator system.

3. Let us define the standard unit vectors in $\mathbb{K}^n$ as

$$e_1 := (1, 0, 0, \ldots, 0), \quad e_2 := (0, 1, 0, \ldots, 0), \quad \ldots, \quad e_n := (0, 0, 0, \ldots, 1)\,.$$

Then the system $e_1, \dots, e_n$ is a generator system in $\mathbb{K}^n$. Really, if $x = (x_1, \dots, x_n) \in \mathbb{K}^n$, then

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

so x can be written as a linear combination of $e_1, \dots, e_n$.

4. A generator system in the vector space $\mathcal{P}_n$ is the so called power function system defined as follows:

$$h_0(x) := 1, \quad h_k(x) := x^k (x \in \mathbb{K},\ k = 1, \dots n)\,.$$

Really, if $f \in \mathcal{P}_n$, $\quad f(x) = a_0 + a_1 x + \cdots + a_n x^n \quad (x \in \mathbb{K})$ then $f = \sum\limits_{k=0}^{n} a_k h_k$.

It is clear that if we enlarge a generator system in V then it remains generator system. But if we leave vectors from a generator system then the resulted system will be not necessarily generator system. The generator systems are – in this sense – the „great” systems. Later we will study the question of „minimal” generator systems.

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

5.9. Definition The vector space V is called finite-dimensional if it has finite generator system. We denote this fact by $\dim V < \infty$.

If a vector space V does not have finite generator system then we call it infinite-dimensional. This fact is denoted by $\dim(V) = \infty$.

5.10. Examples

1. Some finite-dimensional vector spaces: $\{0\}$, the vector space of plane vectors, the vector space of space vectors, $\mathbb{K}^n$, $\mathbb{K}^{m \times n}$, $\mathcal{P}_n$.

2. Now we prove that $\dim \mathcal{P} = \infty$. (About the definition of $\mathcal{P}$ see Examples 5.4.)

Let $f_1, \dots, f_m$ be a finite polynomial system in $\mathcal{P}$. Let

$$k := \max\{\deg f_i \mid i = 1, \dots, m\}\,.$$

Then the polynomial $g(x) := x^{k+1} \quad (x \in \mathbb{K})$ cannot be expressed as linear combination of $f_1, \dots, f_m$ because the linear combination does not increase the degree of the maximally k-degree polynomials over k.

So $\mathcal{P}$ cannot be spanned by any finite polynomial system that is it does not have finite generator system.

5.3. Control Questions

1. Define the subspace of a vector space
2. State and prove the theorem about the necessary and sufficient condition for a set to be a subspace
3. Give 3 examples for subspaces
4. Define the linear combination
5. State and prove the theorem about generated subspace by a finite vector system (This is the theorem about W^*)
6. Give 3 examples for generated subspace in $\mathbb{R}^2$
7. Give 3 examples for generated subspace in $\mathbb{R}^3$
8. Define the standard unit vectors in $\mathbb{K}^n$. What is the subspace generated by them (with proof)?
9. Define the finite dimensional vector space. Prove that the vector space $\mathbb{K}^n$ is finite dimensional
10. Define the infinite dimensional vector space. Prove that the vector space of all polynomials ($\mathcal{P}$) is infinite dimensional

5.4. Homework

1. Let $A \in \mathbb{K}^{m\times n}$. Prove that the following subset of $\mathbb{K}^n$ is a subspace:

$$(A) := \{x \in \mathbb{K}^n \mid Ax = 0\}\,.$$

Here x is regarded as an $n \times 1$ matrix. The subspace (A) is called the nullspace (or kernel) of A.

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

   a) Compute $2a - 4b$.

   b) Determine whether the vector $x = (2, 4, 0)$ is in the subspace $\operatorname{Span}(a, b)$ or not.

3. Let

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 2 & 3 & 1 & 1 \\ 1 & 0 & 8 & 2 \end{bmatrix}\,.$$

Find a generator system in the subspace

$$\operatorname{Ker}(A) := \{x \in \mathbb{R}^4 \mid Ax = 0\}\,.$$

6. Lesson 6

6.1. Linear Independence

6.1. Definition Let $k \in \mathbb{N}$ and $x_1, \ldots, x_k \in V$ be a vector system. This system is called linearly independent (shortly: independent) if its every nontrivial linear combination results nonzero vector, that is:

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

The system is called linearly dependent (shortly: dependent) if it is no independent. That is

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

6.2. Remarks.

1. The equation $\sum\limits_{i=1}^{k} \lambda_i x_i = 0$ is called: dependence equation.

2. It can be simply shown that if a vector system contains identical vectors or it contains the zero vector then it is linearly dependent. In other words: a linearly independent system contains different vectors and it does not contain the zero vector.

3. From the simple properties of vector spaces follows that a one-element vector system is linearly independent if and only if its single element is a nonzero vector.

Let us see some examples for independent and dependent systems:

6.3. Examples

1. Using geometrical methods it can be shown that in the vector space of the space vectors:

    - Two parallel vectors are dependent;
    - Two nonparallel vectors are independent;
    - Three vectors lying in the same plane are dependent;
    - Three vectors that are not lying in the same plane are independent.

2. In the vector space $\mathbb{K}^n$ the system of the standard unit vectors $e_1, \dots, e_n$ is linearly independent, since
$$\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix} = 0 = \sum_{i=1}^{n} \lambda_i e_i = \begin{pmatrix}\lambda_1 \cdot 1 + \lambda_2 \cdot 0 + \dots + \lambda_n \cdot 0\\ \lambda_1 \cdot 0 + \lambda_2 \cdot 1 + \dots + \lambda_n \cdot 0\\ \vdots \\ \lambda_1 \cdot 0 + \lambda_2 \cdot 0 + \dots + \lambda_n \cdot 1\end{pmatrix} = \begin{pmatrix}\lambda_1\\\lambda_2\\\vdots\\\lambda_n\end{pmatrix},$$
which implies $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$.

3. It can be proved that in the vector space $\mathcal{P}_n$ the power function system
$$h_0(x) := 1, \quad h_k(x) := x^k \qquad (x \in \mathbb{K},\ k = 1, \ldots n)$$
is linearly independent.

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 resulted system will be not necessarily linearly independent. The linearly independent systems are – in this sense – the „small" systems. Later we will study the question of „maximal" linearly independent systems.

Now let us study some simple theorems about the connection between the independent, the dependent and the generator systems.

6.4. Theorem *[Diminution of a dependent system]*
Let $x_1, \ldots, x_k \in V$ be a linearly dependent system. Then
$$\exists\, i \in \{1, 2, \ldots k\}: \quad \operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) = \operatorname{Span}(x_1, \dots, x_k).$$
In words: at least one of the vectors in the system is redundant from the point of view of the spanned subspace.

Proof. The „$\subseteq$" relation is trivial, because
$$\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k\} \subseteq \{x_1, \dots, x_k\}.$$

To prove the relation „$\supseteq$" observe first that
$$\{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k\} \subseteq \operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k).$$
It remains the proof of
$$x_i \in \operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k).$$
Indeed, by the dependence of the system there exist the numbers $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ such that they are not all zero and
$$\lambda_1 x_1 + \ldots + \lambda_k x_k = 0.$$

Let i be an index with $\lambda_i \neq 0$. After rearrange the previous vector equation we obtain that:

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

That means that x_i can be expressed as linear combination of $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$, so it is really in the subspace $\operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$.

Thus the subspace $\operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$ covers the system $x_1, \dots, x_k$ which implies the relation „$\supseteq$".

$\square$

6.5. Remark. From the proof it turned out that the redundant vector is that vector whose coefficient in a dependence equation is nonzero.

6.6. Theorem *Let $x_1, \dots, x_k \in V$ be a vector system. If $x \in \operatorname{Span}(x_1, \dots, x_k)$, then the vector system $x_1, \dots, x_k, x$ is linearly dependent.*

Proof. Since $x \in \operatorname{Span}(x_1, \dots, x_k)$, then x can be written as a linear combination of the generator system, that is

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

After rearranging the equation we have

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

Since $-1 \neq 0$, the system is really linearly dependent. $\square$

6.7. Theorem *[Extension of an independent system] Let $x_1, \dots, x_k \in V$ be a linearly independent vector system. Furthermore let $x \in V$. Then*

a) If $x \in \operatorname{Span}(x_1, \dots, x_k)$, then the vector system $x_1, \dots, x_k, x$ is linearly dependent.

b) If $x \notin \operatorname{Span}(x_1, \dots, x_k)$, then the vector system $x_1, \dots, x_k, x$ is linearly independent.

Proof. Part a) is a special case of the previous theorem.

To prove part b) let us take the dependence equation

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

and show that each coefficients are 0. We 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 expression implies $x \in \operatorname{Span}(x_1, \ldots, x_k)$ in contradiction of the assumption of part b). Thus really $\lambda = 0$.

Substituting this result into the dependence equation we have

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

Using the independence of the original system we have

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

The proof is complete. □

An immediate consequence of part b) of the previous theorem is the following

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

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

Another consequence of Theorem 6.7 is that in an infinite dimensional space can be given independent systems of any sizes.

6.9. Theorem *Let V be a vector space over $\mathbb{K}$, and suppose that* $\dim V = \infty$. *Then*

$$\forall\, n \in \mathbb{N}\ \exists\, x_1, \ldots, x_n \in V : \quad x_1, \ldots, x_n \textit{ is a linearly independent system}\,.$$

Proof. Let us fix an arbitrary $n \in \mathbb{N}$.

Since $V \neq \{0\}$, we can choose a vector $x_1 \in V \setminus \{0\}$. Then the system x_1 is linearly independent.

Since $\dim V = \infty$, then $\operatorname{Span}(x_1) \neq V$ (there is not a finite generator system in V). Consequently

$$\exists\, x_2 \in V \setminus \operatorname{Span}(x_1)\,.$$

Using Theorem 6.7 we deduce that the system x_1, x_2 is linearly independent.

Since $\dim V = \infty$, then $\operatorname{Span}(x_1, x_2) \neq V$ (there is not a finite generator system in V). Consequently

$$\exists\, x_3 \in V \setminus \operatorname{Span}(x_1, x_2)\,.$$

Using again Theorem 6.7 we deduce that the system x_1, x_2, x_3 is linearly independent.

Repeating this process, we can construct a vector system $x_1, \ldots, x_n$ which is linearly independent. □

6.2. Basis

6.10. 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.

6.11. Remarks. Since in the zero vector space $\{0\}$ there is no linearly independent system, so this space has no basis. Later we will show that every other finite-dimensional vector space has basis.

The following examples can be easily to consider because we have studied them as examples for generator system and for linearly independent system.

6.12. Examples

1. - In the vector space of the plane vectors the system of every two non-parallel vectors is a basis.
   - In the vector space of the space vectors the system of every three vectors that are not lying in the same plane is a basis.
2. In $\mathbb{K}^n$ the system of the standard unit vectors is a basis. This basis is called the standard basis or the canonical basis of $\mathbb{K}^n$.
3. In the polynomial space $\mathcal{P}_n$ the power function system $h_0, h_1, \ldots h_n$ is a basis.

In the following part of the section we want to prove that every finite-dimensional nonzero vector space has basis.

6.13. Theorem *Every finite-dimensional nonzero vector space has a basis.*

Proof. Let $x_1, \ldots, x_k$ be a finite generator system of V. If this system is linearly independent then it is basis. If it is dependent then by Theorem 6.4 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 while it is possible.

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

6.14. Remarks.

1. 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.
2. It can be proved that every linearly independent system can be completed into a basis.

6.3. Dimension

The aim of this section is to show that in a vector space every basis has the same number of vectors. This common number will be called the dimension of the space.

6.15. Theorem *[Exchange Theorem] Let $x_1, \dots, x_k \in V$ be a linearly independent system and $y_1, \dots, y_m \in V$ be a generator system in V. Then*

$$\forall\, i \in \{1, \dots, k\}\ \exists\, j \in \{1, \dots, m\} : \quad x_1, \dots, x_{i-1}, y_j, x_{i+1}, \dots, x_k \text{ is independent.}$$

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 6.8 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 member of the above chain coincide, at every point in it stand equalities. 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 $x_1, \dots, x_k$. □

6.16. Theorem *The number of vectors in a linearly independent system is not greater than the number of vectors in a generator system.*

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 into 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 into 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 $k \leq m$. □

A simple consequence of the above theorem that the subspaces of a finite dimensional vector space are also finite dimensional, moreover, in a finite dimensional vector space it does not exist a linearly independent system of any size.

6.17. Theorem *Let V be a finite dimensional vector space. Then*

1. *Any subspace $W \subseteq V$ is also finite dimensional.*
2. *There exists $N \in \mathbb{N}$ such that any system of at least N terms is linearly dependent.*

Proof.

1. Since $\dim V < \infty$, there exists a finite generator system in V. Denote by m the number of terms in this generator system.

   Let W be a subspace, and suppose indirectly that $\dim W = \infty$. Then by Theorem 6.9 it must contain $m+1$ linearly independent vectors. Consequently – by the previous theorem – it follows $m+1 \leq m$. This is a contradiction.

2. Let $N = m+1$, and $n \geq N = m+1$. Suppose indirectly that $\dim V = \infty$. Then by Theorem 6.9 it must contain n linearly independent vectors. Consequently – by the previous theorem – it follows $n \leq m$. On the other hand we have
$$n \geq m+1 > m\,.$$
   This is a contradiction.

□

6.18. Corollary. The second statement of the theorem above and Theorem 6.9 say us the following equivalence:

A vector space is finite dimensional if and only if it has an n-term linearly independent system for every $n \in \mathbb{N}$.

Let us go back to the topic of bases.

6.19. 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 6.16 we can deduce that

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

On the other hand

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

Consequently $k = m$. □

6.20. 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 $\dim V$. By definition $\dim(\{0\}) := 0$. If $\dim V = n$ then V is called n-dimensional.

The statements of the following examples follow immediately from the examples for bases.

6.21. Examples

1. The space of the vectors on the straight line is one dimensional.
2. The space of the plane vectors is two dimensional.
3. The space of the space vectors is three dimensional.
4. $\dim(\mathbb{K}^n) = n \quad (n \in \mathbb{N})$.
5. $\dim \mathcal{P}_n = n + 1 \quad (n \in \mathbb{N} \cup \{0\})$.

In the following we will state and prove four useful statements about vector systems in an n-dimensional vector space.

6.22. 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 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*

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

Proof.

1. Let $e_1, \dots, e_n$ be a basis in V. Then it is a generator system, thus by Theorem 6.16 we have:
$$k \le n\,.$$

2. Let $e_1, \dots, e_n$ be a basis in V. Then it is a linearly independent system, thus by Theorem 6.16 we have:
$$k \ge n\,.$$

3. Suppose indirectly that $x_1,\ \dots,\ x_n$ is nor a generator system. Then
$$V \setminus \operatorname{Span}(x_1, \dots, x_n) \ne \emptyset\,.$$
Let $x \in V \setminus \operatorname{Span}(x_1, \dots, x_n)$. Then by Theorem 6.7 the system $x_1,\ \dots,\ x_n,\ x$ 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 $x_1, \dots, x_n$ is linearly dependent. Then by Theorem 6.4 we have
$$\exists\, i \in \{1,\, 2,\, \dots\, n\} : \quad \operatorname{Span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \operatorname{Span}(x_1, \dots, x_n) = V\,.$$
This is a contradiction, because the system $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ has $n-1$ terms, less than the dimension of the space.

□

6.23. Theorem *[Dimensions of subspaces]*
Let $\dim(V) = n < \infty$*. Then it holds for any subspace* $W \subseteq V$ *that*
$$\dim W \le n\,.$$
Furthermore, $\dim W = n$ *if and only if* $W = V$*.*

Proof. By Theorem 6.17 W is finite dimensional. Let $k := \dim W$. Then there exist a basis $e_1,\ \dots,\ e_k$ in W. The system $e_1,\ \dots,\ e_k$ is linearly independent in V. Consequently by Theorem 6.16 $k \le n$ holds.

In the second part of the theorem the direction $W = V \Rightarrow \dim W = n$ is obviously true. To prove the opposite direction, suppose that $\dim W = n$. Then there exist a basis $e_1,\ \dots,\ e_n$ in W. The system $e_1,\ \dots,\ e_n$ is a linearly independent system in V, having n terms. Consequently by Theorem 6.22 $e_1,\ \dots,\ e_n$ is a basis in V. Thus
$$W = \operatorname{Span}(e_1,\ \dots,\ e_n) = V$$

□

6.4. Control Questions

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

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

3. State and prove the independence of the standard unit vectors

4. Prove that

   - if a vector system contains the zero vector then it is dependent
   - if a vector system contains identical vectors then it is dependent

5. State and prove the theorem about the diminution of a dependent system

6. State and prove the theorem about the dependence of the system $x_1, \ldots, x_k, x$, where $x \in \text{Span}\,(x_1, \ldots, x_k)$.

7. State and prove the theorem about the extension of an independent system

8. State and prove the theorem about the arbitrarily large linearly independent systems in an infinite dimensional vector space

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

10. Prove that every finite dimensional nonzero vector space has a basis

11. State and prove the exchange theorem

12. State and prove the most important corollary of the exchange theorem (about the number of terms in a linearly independent and in a generator system)

13. State and prove the theorem about the subspaces of a finite dimensional vector space.

14. State and prove that a linearly independent system cannot be arbitrarily large in a finite dimensional vector space.

15. Prove that any two bases in a finite dimensional vector space have the same number of vectors

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

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

18. State and prove the theorem about the dimensions of subspaces

6.5. Homework

1. Let $x_1 = (1, -2, 3),\ x_2 = (5, 6, -1),\ x_3 = (3, 2, 1) \in \mathbb{R}^3$. Determine that this system is linearly independent or dependent.

2. Which of the following vector systems are bases in $\mathbb{R}^3$?

   a) $x_1 = (1, 0, 0)$, $x_2 = (2, 2, 0)$, $x_3 = (3, 3, 3)$.

   b) $y_1 = (3, 1, -4)$, $y_2 = (2, 5, 6)$, $y_3 = (1, 4, 8)$.

7. Lesson 7

7.1. Coordinates

In this section V is a vector space with $1 \leq \dim V = n \leq \infty$.

7.1. Theorem *Let* $e:\ e_1, \dots e_n$ *be a basis in* V*. Then*

$$\forall\, x \in V\ \exists!\ \xi_1,\ \dots,\ \xi_n \in \mathbb{K}: \quad x = \sum_{i=1}^{n} \xi_i e_i\,.$$

Proof. The existence of the numbers ξ_i is obvious because $e_1, \dots e_n$ is generator system. To confirm the uniqueness take two expansions of x:

$$x = \sum_{i=1}^{n} \xi_i e_i = \sum_{i=1}^{n} \eta_i e_i\,.$$

After rearrangement we obtain:

$$\sum_{i=1}^{n} (\xi_i - \eta_i) e_i = 0\,.$$

From here – using the linear independence of $e_1, \dots e_n$ – follows that $\xi_i - \eta_i = 0$ that is $\xi_i = \eta_i \quad (i = 1, \dots, n)$. □

7.2. Definition The numbers $\xi_1,\ \dots, \xi_n$ in the above theorem are called the coordinates of the vector x relative to the basis $e_1, \dots e_n$ (or shortly: relative to the ordered basis e). The vector

$$[x]_e := (\xi_1, \dots, \xi_n) \in \mathbb{K}^n$$

is called the coordinate vector of x relative to the ordered basis e.

7.3. Remark. If $V = \mathbb{K}^n$ and $e_1, \dots e_n$ is the standard basis in it then

$$\forall\, x \in \mathbb{K}^n: \quad [x]_e = x\,.$$

By this reason we call the components of $x \in \mathbb{K}^n$ the coordinates of x.

In the following theorem we will prove the simple geometrical fact, that if a coordinate of a vector with respect to a basis vector equals zero, then the vector lies in the subspace generated by the remainder basis vectors.

7.4. Theorem *Let* $e_1, \ldots, e_n$ *be a basis in* V*, and*

$$x = \sum_{j=1}^{n} \xi_j e_j \in V \,.$$

Let $i \in \{1, \ldots, n\}$ *be a fixed index, and let*

$$W := \mathrm{Span}\,(e_1,\, \ldots,\, e_{i-1},\, e_{i+1},\, \ldots,\, e_n)\,.$$

Then

$$x \in W \iff \xi_i = 0\,.$$

Proof. First suppose that $x \in W$. Then

$$\exists\, \eta_j \in \mathbb{K} : \quad x = \sum_{\substack{j=1 \\ j \neq i}}^{n} \eta_j e_j = \sum_{\substack{j=1 \\ j \neq i}}^{n} \eta_j e_j + 0 e_i \,.$$

Using the uniqueness of the basis expansion of x with respect to the basis $e_1, \ldots, e_n$ we have:

$$\xi_j = \eta_j \;\; (j \neq i) \quad \text{and} \quad \xi_i = 0\,.$$

The second part is the claim, which was to be demonstrated.

Conversely, suppose that $\xi_i = 0$. Substituting this fact into the basis expansion of x we have

$$x = \sum_{\substack{j=1 \\ j \neq i}}^{n} \xi_j e_j + 0 e_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \xi_j e_j \,.$$

This means that $x \in W$. □

7.5. Remark. The theorem can be extended to the case when more then one coordinates are 0.

Let $e_1, \ldots, e_n$ be a basis in V, and

$$x = \sum_{j=1}^{n} \xi_j e_j \in V \,.$$

Let $I \subseteq \{1, \ldots, n\}$ and $I^* := \{1, \ldots, n\} \setminus I$. Suppose that $I \neq \emptyset$, $I^* \neq \emptyset$, and let

$$W := \mathrm{Span}\,(e_i,\; i \in I^*)\,.$$

Then

$$x \in W \iff \forall\, j \in I :\; \xi_j = 0\,.$$

This extension can be proved similarly to the proof of the theorem.

The following theorem shows us how to compute the results of the vector space operations using coordinates.

7.6. Theorem *Let $e : e_1, \dots e_n$ be an ordered basis in V. Then for any $x, y \in V$ hold*

$$[x+y]_e = [x]_e + [y]_e \ ,$$
$$[\lambda x]_e = \lambda \, [x]_e \ .$$

Proof. To prove the first statement let

$$[x]_e = (\xi_1, \dots, \xi_n), \quad [y]_e = (\eta_1, \dots, \eta_n) \in \mathbb{K}^n \, .$$

Then

$$x + y = \sum_{i=1}^{n} \xi_i e_i + \sum_{i=1}^{n} \eta_i e_i = \sum_{i=1}^{n} (\xi_i + \eta_i) e_i \, ,$$

which implies that

$$[x+y]_e = (\xi_1 + \eta_1, \dots, \xi_n + \eta_n) = (\xi_1, \dots, \xi_n) + (\eta_1, \dots, \eta_n) = [x]_e + [y]_e \ .$$

So the first part is proved. The proof of the second part is similar. □

7.7. Theorem *[Change of Basis]*

Let $e : e_1, \dots, e_n$ and $e' : e'_1, \dots, e'_n$ two ordered basis in V. Define the $e \to e'$ transition matrix as follows:

$$C := \left[[e'_1]_e, \dots, [e'_n]_e \right] \in \mathbb{K}^{n \times n},$$

that is: the j-th column vector of C is the coordinate vector of e'_j relative to the basis e.

Then

$$\forall \, x \in V : \quad C \cdot [x]_{e'} = [x]_e \ .$$

Proof. Let $[x]_{e'} = (\xi'_1, \dots, \xi'_n)$. Then

$$C \cdot [x]_{e'} = \left[[e'_1]_e, \dots, [e'_n]_e \right] \cdot \begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \vdots \\ \xi'_n \end{pmatrix} = \sum_{j=1}^{n} \xi'_j \cdot [e'_j]_e = \sum_{j=1}^{n} \left[\xi'_j \cdot e'_j \right]_e = \left[\sum_{j=1}^{n} \xi'_j \cdot e'_j \right]_e = [x]_e \ .$$

□

7.8. Remark. The above theorem makes us possible to determine the coordinates of a vector if we know its coordinates in another basis. In this connection the basis e is called „old basis” and the basis e' is called „new basis”.

7.2. Elementary Basis Transformation (EBT)

In this section we will discuss the change of basis in the special case when the new basis differs from the old one in a single vector . We will compute the new coordinates in elementary way.

7.9. Theorem *[The basic theorem of the EBT]*
Let $1 \leq \dim V = n < \infty$ *and* $e_1, \ldots, e_n$ *be a basis in* V*. Furthermore let* b *be a vector in* V *with the basis expansion*

$$b = \beta_1 e_1 + \cdots + \beta_n e_n \in V \,.$$

Let $i \in \{1, \ldots, n\}$*, and denote by* e' *the vector system*

$$e_1, \ldots, e_{i-1}, b, e_{i+1}, \ldots, e_n \,.$$

Then e' *is a basis if and only if* $\beta_i \neq 0$.
The number β_i *is called: generator element.*

Proof. Let e' be a basis, and suppose indirectly that $\beta_i = 0$.
Then the term $\beta_i e_i$ is not contained in the basis expansion of b:

$$b = \sum_{\substack{j=1 \\ j \neq i}}^{n} \beta_j \cdot e_j \,.$$

After rearranging we have

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} \beta_j \cdot e_j + (-1) \cdot b = 0 \,.$$

The above equality shows that a nontrivial linear combination of e' equals 0, consequently e' is linearly dependent, it cannot be a basis. This is a contradiction.

To prove the converse statement, suppose that $\beta_i \neq 0$.

Since the number of the terms in e' is n, it is enough to prove (see Theorem 6.22) that e' is a generator system.

To show this let us express b_i from the basis expansion of b:

$$b = \sum_{\substack{j=1 \\ j \neq i}}^{n} \beta_j \cdot e_j + \beta_i e_i$$

$$e_i = \frac{1}{\beta_i} \left(b - \sum_{\substack{j=1 \\ j \neq i}}^{n} \beta_j \cdot e_j \right) \,.$$

Let $x = \xi_1 e_1 + \cdots + \xi_n e_n \in V$ be an arbitrary vector, and let us substitute in its basis expansion the above formula for e_i:

$$x = \sum_{\substack{j=1 \\ j \neq i}}^{n} \xi_j e_j + \xi_i e_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \xi_j e_j + \frac{\xi_i}{\beta_i} \left(b - \sum_{\substack{j=1 \\ j \neq i}}^{n} \beta_j e_j \right) =$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{n} \xi_j e_j + \frac{\xi_i}{\beta_i} b - \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\frac{\xi_i}{\beta_i} \beta_j \right) e_j = \frac{\xi_i}{\beta_i} b + \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\xi_j - \frac{\xi_i}{\beta_i} \beta_j \right) e_j .$$

Thus we have created x as a linear combination of the system e', which shows that e' is really a generator system. □

7.10. Remarks.

1. It turned out from the proof that if we denote by $\xi'_1, \ldots, \xi'_n$ the coordinates of x relative to the basis e' (these are the new coordinates), then

$$\xi'_i = \frac{\xi_i}{\beta_i}, \qquad \xi'_j = \xi_j - \frac{\xi_i}{\beta_i} \beta_j \qquad (j = 1, \ldots, i-1, i+1, \ldots, n) .$$

   These are the Transformation Formulas.

2. In manual calculations it is more useful to write the Transformation Formulas in the following way:

$$\xi'_i = \frac{\xi_i}{\beta_i}, \qquad \xi'_j = \xi_j - \frac{\beta_j}{\beta_i} \xi_i \quad (j = 1, \ldots, i-1, i+1, \ldots, n) .$$

3. Frequently it is necessary to compute the new coordinates of more than one vectors. In this case is suggested to compute the first coordinates of all vectors, then the second ones, etc. Thus we compute by coordinates, but not by vectors. The data are written in the Basis Table as follows:

	x	y	z	b
e_1	ξ_1	η_1	ζ_1	β_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_i	ξ_i	η_i	ζ_i	β_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
e_n	ξ_n	η_n	ζ_n	β_n

   Since the same coordinates stand in the same rows, the computation by coordinates is called computation by rows. The new Basis Table after the transformation is

$$
\begin{array}{c|cccc}
 & x & y & z & b \\
\hline
e_1 & \xi_1 - \dfrac{\beta_1}{\beta_i}\xi_i & \eta_1 - \dfrac{\beta_1}{\beta_i}\eta_i & \zeta_1 - \dfrac{\beta_1}{\beta_i}\zeta_i & \beta_1 - \dfrac{\beta_1}{\beta_i}\beta_i = 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
b & \dfrac{\xi_i}{\beta_i} & \dfrac{\eta_i}{\beta_i} & \dfrac{\zeta_i}{\beta_i} & \dfrac{\beta_i}{\beta_i} = 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
e_n & \xi_n - \dfrac{\beta_n}{\beta_i}\xi_i & \eta_n - \dfrac{\beta_n}{\beta_i}\eta_i & \zeta_n - \dfrac{\beta_n}{\beta_i}\zeta_i & \beta_n - \dfrac{\beta_n}{\beta_i}\beta_i = 0
\end{array}
$$

We can state the rule of filling the new Basis Table in words as follows:

- We obtain the row of b that we divide the row of e_i (named generator row) by the generator element β_i.
- For all $j \in \{1, \ldots, n\}$, $j \neq i$ we obtain the new row of e_j that we subtract from the old row of e_j the $\dfrac{\beta_j}{\beta_i}$-multiple of the generator row.
  Since this last step results 0-s under b, then the step can be stated as follows:
  For all $j \in \{1, \ldots, n\}$, $j \neq i$ we obtain the new row of e_j that we subtract from the old row of e_j as multiple of the generator row, which results 0-s under b.

4. It is natural that the coordinates of b relative to the basis e' are $0, \ldots, 0, 1, 0, \ldots, 0$, since
$$b = 0 \cdot e_1 + \cdots + 0 \cdot e_{i-1} + 1 \cdot e_i + 0 \cdot e_{i+1} + \cdots + 0 \cdot e_n \,.$$

5. We say that we built in the basis the vector b into the place of e_i using EBT. If we apply a sequence of EBT-s, then we can exchange a basis to a new basis.

6. Theorem 7.9 is a simple consequence of Theorem 7.4. Really, if
$$W := \mathrm{Span}\,(e_1,\, \ldots,\, e_{i-1},\, e_{i+1},\, \ldots,\, e_n)\,,$$
then
   - If $\beta_i = 0$, then by Theorem 7.4 $b \in W$. Consequently – using Theorem 6.6 – the system e' is linearly dependent.
   - If $\beta_i \neq 0$, then by Theorem 7.4 $b \notin W$. Consequently – using Theorem 6.7 – the system e' is linearly independent. Since e' has n terms, then e' is a basis.

The disadvantage of this proof is, that it does not give the new coordinates of the vectors, that is, it does not give the transformation formulas.

7.3. Control Questions

1. State and prove the theorem about the existence and uniqueness of coordinates

2. Define the coordinates and the coordinate vector

3. State and prove the theorem about the operations with coordinates

4. Define the transition matrix

5. State and prove the theorem about the Change of Basis

6. State and prove the basic theorem about the Elementary Basis Transformation (EBT)

7. What is a Basis Table?

7.4. Homework

1. The following basis is given in $\mathbb{R}^3$:

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

   Determine the coordinate vector of $x = (3,4,3)$ relative to the given basis.

2. It is given the following basis in $\mathcal{P}_2$:

$$P_1(x) = 1 + x,\ P_2(x) = 1 + x^2,\ P_3(x) = x + x^2\,.$$

   Determine the coordinate vector of $P(x) = 2 - x + x^2$ relative to the given basis.

8. Lesson 8

8.1. The Rank of a Vector System

In this section we try to characterize by a number the „measure of dependence". For example in the vector space of the space vectors we feel that a linearly dependent system is „better dependent" if it lies on a straight line than it lies in a plane. This observation motivates the following definition.

8.1. 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 this vector system. It is denoted by $\operatorname{rank}(x_1, \ldots, x_k)$. So

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

8.2. Remarks.

1. $0 \le \operatorname{rank}(x_1, \ldots, x_k) \le k$.
2. The rank expresses the „measure of dependence". The smaller is the rank the more dependent are the vectors. Especially:

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

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

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

8.2. The Rank of a Matrix

8.3. Definition Let $A \in \mathbb{K}^{m \times n}$. Then we can decompose it with horizontal straight lines into row submatrices. The entries of the ith row submatrix form the vector:

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

which is called the ith row vector of A. The subspace generated by the row vectors of A is called the row space of A and is denoted by $\operatorname{Row}(A)$.

8.4. Remark. Obviously $\dim \operatorname{Row}(A) \le m$.

8.5. Definition Let $A \in \mathbb{K}^{m\times n}$. Then we can decompose it with vertical straight lines into column submatrices. The entries of the jth column submatrix form the vector:

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

which is called the jth column vector of A. The subspace generated by the column vectors of A is called the column space of A and is denoted by $\mathrm{Col}(A)$.

8.6. Remarks.

1. Obviously $\dim \mathrm{Col}(A) \le n$.
2. Obviously

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

8.7. Theorem *Let $A \in \mathbb{K}^{m\times n}$, $B \in \mathbb{K}^{n\times p}$. Then*

$$\mathrm{Col}(AB) \subseteq \mathrm{Col}(A) \quad \textit{and} \quad \mathrm{Row}(AB) \subseteq \mathrm{Row}(B)\,.$$

Proof. Using the matrix multiplication with blocks we can establish that each column vector of AB is a linear combination of the column vectors of A (the coefficients are the entries in the current column of B), thus each column of AB lies in $\mathrm{Col}(A)$. This implies $\mathrm{Col}(AB) \subseteq \mathrm{Col}(A)$.

Similarly, each row vector of AB is a linear combination of the row vectors of B (the coefficients are the entries in the current row of A), thus each row of AB lies in $\mathrm{Row}(B)$. This implies $\mathrm{Row}(AB) \subseteq \mathrm{Row}(B)$. □

8.8. Corollary.

$$\dim \mathrm{Col}(AB) \le \dim \mathrm{Col}(A) \quad \text{and} \quad \dim \mathrm{Row}(AB) \le \dim \mathrm{Row}(B)\,.$$

8.9. Theorem *For any $A \in \mathbb{K}^{m\times n}$ holds*

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

Proof. For $A = 0$ the statement is trivially true.

Suppose that $A \ne 0$. Then $r := \dim \mathrm{Row}(A) \ge 1$. Let $b_1, \dots, b_r \in \mathbb{K}^m$ be a basis in $\mathrm{Col}(A)$, and denote by $B \in \mathbb{K}^{m\times r}$ the matrix whose columns are $b_1, \dots, b_r$. The columns of A can be written as linear combinations of $b_1, \dots, b_r$:

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

Let $D = (d_{ij}) \in \mathbb{K}^{r\times n}$. We have the following factorization of A (it is called: basis factorization):

$$A = BD\,.$$

Applying Corollary 8.8 we have

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

Thus $\dim \mathrm{Row}(A) \le \dim \mathrm{Col}(A)$. To obtain the opposite inequality let us apply this result for A^T instead of A:

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

The proof is complete. □

8.10. Definition The common value of $\dim \mathrm{Row}(A)$ and of $\dim \mathrm{Col}(A)$ is called the rank of the matrix A. Its notation is: $\mathrm{rank}\,(A)$. So

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

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

1. The rank of the matrix equals the rank of its row vector system and equals the rank of its column vector system.

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

3. $0 \le \mathrm{rank}\,(A) \le \min\{m, n\}$. $\qquad$ $\mathrm{rank}\,(A) = 0 \Leftrightarrow A = 0$.

4. $\mathrm{rank}\,(A) = m$ if and only if the row vectors of A are linearly independent. Furthermore – using Theorem 6.23:

   $$\mathrm{rank}\,(A) = m \quad \text{if and only if} \quad \mathrm{Col}(A) = \mathbb{K}^m\,.$$

   In this case necessarily $m \le n$.

5. $\mathrm{rank}\,(A) = n$ if and only if the column vectors A are linearly independent. Furthermore – using Theorem 6.23:

   $$\mathrm{rank}\,(A) = n \quad \text{if and only if} \quad \mathrm{Row}(A) = \mathbb{K}^n\,.$$

   In this case necessarily $m \ge n$.

8.3. System of Linear Equations

8.12. 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:

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

where the coefficients $a_{ij} \in \mathbb{K}$ and the right-side constants b_i are given. The system is called homogeneous if $b_1 = \dots = b_m = 0$.

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. These systems of values of the variables are called the solutions of the linear system.

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

Let us denote by $a_1, \dots, a_n$ the column vectors formed from the coefficients on the left side and by b the vector formed from the right-side constants as follows:

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

Using these notations our linear system can be written more succinctly as a vector equation in $\mathbb{K}^m$ as

$$x_1a_1 + x_2a_2 + \dots + x_na_n = b\,.$$

Let us introduce the following matrix (the so called coefficient matrix)

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

and the unknown vector $x := (x_1, \dots, x_n) \in \mathbb{K}^n$. Then the most succinct form of our system is:

$$Ax = b.$$

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

8.14. Remark. It is easy to observe that

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

Thus the consistence of a linear system is equivalent with the question that b lies in the column space of A or not. Consequently as smaller is the column space as greater is the chance of inconsistence. If $\operatorname{rank}(A)$ is equal to the number of rows m 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.

Denote by $\mathcal{S}$ the set of solution vectors of $Ax = b$ that is:

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

Naturally, the system is inconsistent if and only if $\mathcal{S} = \emptyset$.

8.15. Definition Let $A \in \mathbb{K}^{m\times n}$. Then the linear system $Ax = 0$ is called the homogeneous system of linear equations associated with the matrix A. It is called sometimes the homogeneous system associated with $Ax = b$.

Remark that the homogeneous system is always consistent because the zero vector is its solution.

8.16. Theorem *Denote by $\mathcal{S}_h$ the solution set of the homogeneous system, that is:*

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

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

Proof. Since the zero vector is contained in $\mathcal{S}_h$, then $\mathcal{S}_h \neq \emptyset$.

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

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

which implies $x + y \in \mathcal{S}_h$.

On the other hand $\mathcal{S}_h$ is closed under scalar multiplication, because if $x \in \mathcal{S}_h$ and $\lambda \in \mathbb{K}$, then $Ax = 0$, consequently

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

which implies $\lambda x \in \mathcal{S}_h$. □

8.17. Definition Let $A \in \mathbb{K}^{m\times n}$. The subspace $\mathcal{S}_h$ is called the null space or the kernel of the matrix A and is denoted by $\operatorname{Ker}(A)$. That is

$$\operatorname{Ker}(A) := \mathcal{S}_h = \{x \in \mathbb{K}^n \mid Ax = 0\} \subset \mathbb{K}^n\,.$$

How to solve a linear equation system? We have learnt in the secondary school the Substitution Method. It is applicable for systems having small sizes. For larger systems we need some algorithmic method. One of these algorithmic methods is discussed in the following section.

8.4. The Elementary Basis Transformation Method

In this section we will study an algorithmic method, called Elementary Basis Transformation Method (EBT-method) for solving system of linear equations. The method makes us possible to know the structure of the solution sets $\mathcal{S}$ and $\mathcal{S}_h$.

Suppose that $A \neq 0$. (The case $A = 0$ is trivial.) The essentiality of the EBT-method is that we construct a basis in $\mathrm{Col}(A)$.

Let us start with the standard basis in $\mathbb{K}^m$

$$e_1 := (1,\, 0,\, \dots,\, 0), \quad \dots, \quad e_m := (0,\, 0,\, \dots,\, 1)\,,$$

and write the coordinates of the column vectors of A and the coordinates of b relative to this basis. Thus we obtain the Starting Basis Table. Since the basis contains the standard unit vectors, the coordinates are the components itself, consequently we need simply to copy A and b into the table.

$$\begin{array}{c|cccc|c} & a_1 & a_2 & \dots & a_n & b \\ \hline e_1 & a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ e_2 & a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ e_m & a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \quad = \quad \begin{array}{c|c|c} & a_1\ a_2\ \dots\ a_n & b \\ \hline e_1 & & \\ \vdots & A & b \\ e_m & & \end{array}$$

Using a sequence of EBT-s put the vectors a_j into the basis up to it is possible. Denote by r the number of the column vectors a_j that can be built into the basis. Suppose for simplicity that the first r vectors $e_1,\, e_2,\, \dots,\, e_r$ of the basis are replaced the first r columns $a_1,\, a_2,\, \dots,\, a_r$ of A (simplification assumption). Then the last Basis Table is as follows:

$$\begin{array}{c|cccc|ccc|c} & a_1 & a_2 & \dots & a_r & a_{r+1} & \dots & a_n & b \\ \hline a_1 & 1 & 0 & \dots & 0 & d_{1,r+1} & \dots & d_{1n} & c_1 \\ a_2 & 0 & 1 & \dots & 0 & d_{2,r+1} & \dots & d_{2n} & c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_r & 0 & 0 & \dots & 1 & d_{r,r+1} & \dots & d_{rn} & c_r \\ \hline e_{r+1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 & q_{r+1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ e_m & 0 & 0 & \dots & 0 & 0 & \dots & 0 & q_m \end{array} \tag{8.1}$$

8.18. Remarks.

1. We can easily explain the 0 entries in the e-rows. If one of these entries were be nonzero, then we could make a further EBT, thus this table would not be the last one.

2. It is obvious that the system $a_1,\, \dots,\, a_r$ is basis in $\mathrm{Col}(A)$, because

- It is a subsystem of a basis, consequently it is linearly independent.
- The last Basis Table (8.5.) shows us that all the column vectors of A can be written as linear combinations of $a_1, \ldots, a_r$, consequently it is a generator system in $\mathrm{Col}(A)$. The coefficients of the linear combinations can be red out from the rows of the vectors $a_1, \ldots, a_r$.

By this reason we can establish that

$$\operatorname{rank}(A) = \dim \mathrm{Col}(A) = r\,.$$

3. In the case $r = m$ there are no e-rows but only a-rows. In this case all the vectors in the starting basis are replaced.

4. In the case $r = n$ all the column vectors of A are built into the basis.

5. We have assumed – for the simpler notations – that the first r columns of A are built in the basis to the place of the first r standard basis vectors. This resulted that the order of vectors in the first column of the last Basis Table is

$$a_1, \ldots, a_r, e_{r+1}, \ldots, e_m\,.$$

In general case the a-rows and the e-rows are in arbitrary way intermixed.

Let us see the system of linear equations represented by the last basis table:

$$\begin{array}{rcl}
1x_1 + 0x_2 + \ldots + 0x_r + d_{1,r+1}x_{r+1} + \ldots + d_{1n}x_n &=& c_1 \\
0x_1 + 1x_2 + \ldots + \quad + d_{2,r+1}x_{r+1} + \ldots + d_{2n}x_n &=& c_2 \\
&\vdots& \\
0x_1 + 0x_2 + \ldots + 1x_r + d_{r,r+1}x_{r+1} + \ldots + d_{rn}x_n &=& c_r \\
0x_1 + 0x_2 + \ldots + 0x_r + \quad 0x_{r+1} \quad + \ldots + \quad 0x_n &=& q_{r+1} \\
&\vdots& \\
0x_1 + 0x_2 + \ldots + 0x_r + \quad 0x_{r+1} \quad + \ldots + \quad 0x_n &=& q_m
\end{array} \tag{8.2}$$

This system is equivalent with the original system, because we have applied finitely times the following operations:

- Divide an equation by a constant
- Subtraction from an equation a constant multiple of another equation.

Thus the solution set of the system (8.5.) is $\mathcal{S}$. But this system is essentially simpler.

8.19. Theorem *If*

$$1 \le r < m \quad \textit{and} \quad \exists\, i \in \{r+1, \ldots, n\} : \; q_i \neq 0$$

then the original system $Ax = b$ has no solution (it is inconsistent), that is $\mathcal{S} = \emptyset$.

Proof. The system (8.5.) contains the antinomic equation

$$0x_1 + 0x_2 + \ldots + 0x_r + 0x_{r+1} + \ldots + 0x_n = q_i$$

□

8.20. Theorem *If*

- *or* $r = m$
- *or* $r < m$ *but* $q_{r+1} = \ldots = q_n = 0$

then the original system $Ax = b$ *is consistent, that is* $\mathcal{S} \neq \emptyset$*. All the solutions can be given by the formula*

$$x_i = c_i - \sum_{j=r+1}^{n} d_{ij}x_j \qquad (i = 1, \ldots, r)\,, \tag{8.3}$$

where $x_{r+1}, \ldots, x_n \in \mathbb{K}$ *are arbitrary numbers (they are called free variables).*

The variables $x_1, \ldots, x_r$ *are called bound variables. The number* $m - r$ *(that is the number of the free variables) is called the degree of freedom of the system* $Ax = b$*.*

Proof. The last $m - r$ equations

$$\begin{aligned} 0x_1 + 0x_2 + \ldots + 0x_r + 0x_{r+1} + \ldots + 0x_n &= q_{r+1} \\ &\vdots \\ 0x_1 + 0x_2 + \ldots + 0x_r + 0x_{r+1} + \ldots + 0x_n &= q_m \end{aligned}$$

can be left from the system (8.5.), because their solution sets are the whole $\mathbb{K}^n$. The remainder system

$$\begin{aligned} 1x_1 + 0x_2 + \ldots + 0x_r + d_{1,r+1}x_{r+1} + \ldots + d_{1n}x_n &= c_1 \\ 0x_1 + 1x_2 + \ldots + \qquad + d_{2,r+1}x_{r+1} + \ldots + d_{2n}x_n &= c_2 \\ &\vdots \\ 0x_1 + 0x_2 + \ldots + 1x_r + d_{r,r+1}x_{r+1} + \ldots + d_{rn}x_n &= c_r \end{aligned} \tag{8.4}$$

– that is called reduced system – is also equivalent with the original $Ax = b$. The bound variables can be expressed easily from the equations:

$$\begin{aligned} x_1 &= c_1 - d_{1,r+1}x_{r+1} - \ldots - d_{1n}x_n \\ x_2 &= c_2 - d_{2,r+1}x_{r+1} - \ldots - d_{2n}x_n \\ &\vdots \\ x_r &= c_r - d_{r,r+1}x_{r+1} - \ldots - d_{rn}x_n \end{aligned} \tag{8.5}$$

□

8.21. Remarks.

1. Let us complete the equations (8.5) with the trivial equations $x_i = x_i$ $(i = r+1, \ldots, n)$:

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

These equations are considered to componentwise equalities of the following vector equation:

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

Shortly:

$$x = x^B + x_{r+1} \cdot v_{r+1} + \ldots + x_n \cdot v_n = x^B + \sum_{j=r+1}^{n} x_j v_j \,. \tag{8.6}$$

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

$$v_{r+1} = \begin{pmatrix} -d_{1,r+1} \\ -d_{2,r+1} \\ \vdots \\ -d_{r,r+1} \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \dots, \ v_n = \begin{pmatrix} -d_{1n} \\ -d_{2,n} \\ \vdots \\ -d_{rn} \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{8.7}$$

Thus we have:

$$\mathcal{S} = \left\{ x^B + \sum_{j=r+1}^{n} x_j v_j \;\middle|\; x_j \in \mathbb{K} \right\} \subseteq \mathbb{K}^n . \tag{8.8}$$

2. Suppose – under the assumptions of the previous theorem – that $r = n$. In this case the numbers d_{ij} does not exist, the sums in (8.3) and in (8.6) are empty, and all the variables are bound (there is no free variable, the degree of freedom is 0). The solution of the system $Ax = b$ is unique:

$$x_1 = c_1, \ \dots, \ x_n = c_n \quad \text{and} \quad x = x^B \quad \text{and} \quad \mathcal{S} = \{x^B\} .$$

8.22. Theorem *Let $A \in \mathbb{K}^{m\times n}$ and $b \in \mathbb{K}^m$. Then*

1. *The solution set $\mathcal{S}_h$ of the homogeneous system of linear equations $Ax = 0$ is an $n-r$ dimensional subspace in $\mathbb{K}^n$. A basis in this subspace is the vector system $v_{r+1}, \dots, v_n$ defined in* (8.7)*.*

2. *If the system of linear equations $Ax = b$ is consistent, then its solution set $\mathcal{S}$ is a translation of the subspace $\mathcal{S}_h$ (a linear manifold), more precisely*

$$\mathcal{S} = x^B + \mathcal{S}_h .$$

Proof.

1. Let us see the homogeneous system $Ax = 0$. Since the zero vector has all zero coordinates relative to any basis, it follows that $c_1 = \dots = c_r = 0$, which means $x^B = 0$. Put this result into (8.8):

$$\mathcal{S}_h = \left\{ \sum_{j=r+1}^{n} x_j v_j \;\middle|\; x_j \in \mathbb{K} \right\},$$

which means that $v_{r+1}, \dots, v_n$ is a generator system in $\mathcal{S}_h$.

On the other hand – because of the 0-1 components – the system $v_{r+1}, \dots, v_n$ is linearly independent.

Thus the system $v_{r+1}, \dots, v_n$ is a basis in $\mathcal{S}_h$, consequently $\dim \mathcal{S}_h = n-r$.

2. It follows immediately using (8.8).

$\square$

8.23. Remarks.

1. In the case $r = n$ we have $\mathcal{S}_h = \{0\}$ and $\dim \mathcal{S}_h = 0$.
2. Since $\mathrm{Ker}(A) = \mathcal{S}_h$ and $\dim \mathcal{S}_h = n - r$ and $\dim \mathrm{Col}(A) = r$, then we have the important identity

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

8.5. Linear Equation Systems with Square Matrices

Let us study the linear equation system with square matrix:

$$Ax = b \qquad\qquad (A \in \mathbb{K}^{n\times n}, b \in \mathbb{K}^n)\,.$$

Denote by r the rank of A. We distinguish between the two basic cases as follows.

Case 1.: $r = n$.

In this case $\mathrm{rank}\,(A)$ equals the number of rows of A, consequently (see Remark 8.11) $\mathrm{Col}(A) = \mathbb{K}^n$, thus the system is consistent (see Remark 8.14).

On the other hand – because of $\mathrm{rank}\,(A)$ equals the number of columns – the solution is unique (see: Remark 8.21).

Thus in the case $\mathrm{rank}\, A = n$ the square system has a unique solution independently of b.

If we solve the square system using the EBT-algorithm, then the last Basis Table in this case is as follows:

$$\begin{array}{c|cccc|c} & a_1 & a_2 & \dots & a_n & b \\ \hline a_1 & 1 & 0 & \dots & 0 & c_1 \\ a_2 & 0 & 1 & \dots & 0 & c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ a_n & 0 & 0 & \dots & 1 & c_n \end{array}$$

which represents the following linear system:

$$\begin{aligned} 1x_1 + 0x_2 + \dots + 0x_n &= c_1 \\ 0x_1 + 1x_2 + \dots + \qquad &= c_2 \\ &\ \vdots \\ 0x_1 + 0x_2 + \dots + 1x_n &= c_n \end{aligned}$$

We can read out the unique solution easily:

$$x_1 = c_1, \quad x_2 = c_2, \quad \dots, \quad x_n = c_n\,.$$

Case 2.: $r < n.$

In this case rank (A) is less than the number of rows of A, consequently (see Remark8.11) $\mathrm{Col}(A) \subseteq \mathbb{K}^n$ but $\mathrm{Col}(A) \neq \mathbb{K}^n$. Thus the system may be consistent (if $b \in \mathrm{Col}(A)$) or it may be inconsistent (if $b \notin \mathrm{Col}(A)$).

If the system is consistent then – since r is less than the number of columns – the system has infinitely many solutions, the degree of freedom is $n - r \geq 1$.

8.6. Inverses with EBT

8.24. Theorem *Let $A \in \mathbb{K}^{n\times n}$ be a square matrix. Then*

1. $\operatorname{rank} A = n \quad \Longleftrightarrow \quad$ *A is invertible (regular);*

2. $\operatorname{rank} A < n \quad \Longleftrightarrow \quad$ *A is not invertible (singular).*

Proof. Denote by I the identity matrix of size $n \times n$. 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 $AX = I$ (see Section 2.4.). The matrix equation $AX = I$ can be written in the form

$$A \cdot \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} ,$$

which is equivalent with the following collection of linear systems:

$$Ax_1 = e_1, \quad Ax_2 = e_2, \quad \dots, \quad Ax_n = e_n\,. \tag{8.9}$$

Case 1.: $r = n.$ In this case – using the results of the previous section – all these systems can be solved uniquely. This implies that A^{-1} exists, and the columns of A^{-1} are the solution vectors $x_1, \dots, x_n$ of the above linear systems.

Case 2.: $r < n.$ Since $\dim \mathrm{Col}(A) = r < n$, then all the standard unit vectors $e_1, \dots, e_n$ cannot be in $\mathrm{Col}(A)$. Consequently at least one of the above collection of linear systems is inconsistent. This implies that A^{-1} does not exist. □

8.25. Remarks.

1. We determine the invertibility and the inverse of A as follows.

The linear systems in (8.9) have a common coefficient matrix, thus they can be solved simultaneously with EBT-algorithm using a common Basis Table. The starting table is:

$$\begin{array}{c|ccc|ccc} & a_1 & \dots & a_n & e_1 & \dots & e_n \\ \hline e_1 & & & & & & \\ \vdots & & A & & & I & \\ e_n & & & & & & \end{array}\,.$$

If not all the columns of A can be put into the basis, then $\operatorname{rank}(A) < n$, consequently A has no inverse, it is singular.

If all the columns of A can be put into the basis, then $\operatorname{rank}(A) = n$, consequently A has an inverse, it is regular. In this case suppose that the columns of A are put in the basis in their original order $a_1, \dots, a_n$. In this case A^{-1} can be read out simply from the last Basis Table:

$$\begin{array}{c|ccc|ccc} & a_1 & \dots & a_n & e_1 & \dots & e_n \\ \hline a_1 & 1 & \dots & 0 & c_{11} & \dots & c_{1n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_n & 0 & \dots & 1 & c_{n1} & \dots & c_{nn} \end{array} = \begin{array}{c|ccc|ccc} & a_1 & \dots & a_n & e_1 & \dots & e_n \\ \hline e_1 & & & & & & \\ \vdots & & I & & & A^{-1} & \\ e_n & & & & & & \end{array}\,.$$

2. Using the connection between the inverse and the determinant (see Section 2.4.), we can collect our results as follows:

$$\operatorname{rank}(A) = n \Leftrightarrow \exists A^{-1} \Leftrightarrow \det(A) \neq 0 \Leftrightarrow A \text{ is regular} \Leftrightarrow A \text{ is invertible}\,.$$

$$\operatorname{rank}(A) < n \Leftrightarrow \nexists A^{-1} \Leftrightarrow \det(A) = 0 \Leftrightarrow A \text{ is singular} \Leftrightarrow A \text{ is not invertible}\,.$$

8.7. Control Questions

1. Define the rank of a vector system

2. State and prove the theorem about the connection between the column spaces of AB and A

3. State and prove the theorem about the connection between the row spaces of AB and B

4. State and prove the theorem about the connection between the dimensions of the row space and the column space of A

5. Define the rank of a matrix

6. What is the scalar form, the vector form and the matrix form of a system of linear equations?

7. Define the sets $\mathcal{S}$, $\mathcal{S}_h$ and $\mathrm{Ker}(A)$. State and prove that $\mathcal{S}_h$ is a subspace

8. Give the starting Basis Table for $Ax = b$

9. Give the last Basis Table for $Ax = b$ (under the simplification assumption)

10. State (without proof) the theorem about the structure of solution sets of a homogeneous and a nonhomogeneous system of linear equations

11. State (without proof) the theorem about the system of linear equations with square matrices (case 1 and case 2)

12. State and prove the theorem about the connection between the rank and the invertibility of a matrix

8.8. Homework

1. Find the ranks of the matrices

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

2. Solve the systems of linear equations (with the Substitution Method):

$$a)\quad \begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & 3x_3 & = & 6 \\ 2x_1 & - & x_2 & + & 4x_3 & = & 1 \\ x_1 & - & x_2 & + & x_3 & = & 3 \\ \hline \end{array} \qquad\qquad b)\quad \begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 5 \\ x_1 & & & + & x_3 & = & -2 \\ 2x_1 & + & x_2 & + & 3x_3 & = & 3 \\ \hline \end{array}$$

9. Lesson 9

9.1. Eigenvalues and eigenvectors of Matrices

9.1. Definition Let $A \in \mathbb{K}^{n \times n}$. The number $\lambda \in \mathbb{K}$ is called the eigenvalue of A if there exists a nonzero vector in $\mathbb{K}^n$ such that

$$Ax = \lambda x$$

The vector $x \in \mathbb{K}^n \setminus \{0\}$ is called an eigenvector corresponding to the eigenvalue λ.

The set of the eigenvalues of A is called the spectrum of A and is denoted by $\mathrm{Sp}\,(A)$.

One can show by an easy rearrangement that the above equation is equivalent with the homogeneous square linear system

$$(A - \lambda I)x = 0$$

where I denotes the identity matrix in $\mathbb{K}^{n \times n}$.

So a number $\lambda \in \mathbb{K}$ is eigenvalue if and only if the above system has infinite many solutions that is if its determinant equals 0:

$$\det(A - \lambda I) = 0.$$

The left side of the equation is a polynomial whose roots are the eigenvalues.

9.2. 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 A. The multiplicity of the root λ is called the algebraic multiplicity of the eigenvalue λ and is denoted by $a(\lambda)$.

9.3. Remark. One can see by expansion along the first row that the coefficient of λ^n is $(-1)^n$. Furthermore from $P(0) = \det(A - 0I) = \det(A)$ follows that the constant term is $\det(A)$. So the form of the characteristic polynomial:

$$P(\lambda) = (-1)^n \cdot \lambda^n + \dots + \det(A) \qquad (\lambda \in \mathbb{K})\,.$$

Since the eigenvalues are the roots in $\mathbb{K}$ of the characteristic polynomial we can state as follows:

- If $\mathbb{K} = \mathbb{C}$ then $\mathrm{Sp}\,(A)$ is a nonempty set with at most n elements. Counting every eigenvalue with its algebraic multiplicity the number of the eigenvalues is exactly n.
- If $\mathbb{K} = \mathbb{R}$ then $\mathrm{Sp}\,(A)$ is a (possibly empty) set at most with n elements.

9.4. Remark. Let $A \in \mathbb{K}^{n\times n}$ be a (lower or upper) triangular matrix. Then – for example in lower triangular case – its characteristic polynomial is

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & 0 & \dots & 0 \\ a_{21} & a_{22}-\lambda & \dots & 0 \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} =$$

$$= (a_{11}-\lambda)\cdot(a_{22}-\lambda)\cdot\dots\cdot(a_{nn}-\lambda) \qquad (\lambda \in \mathbb{K})\,.$$

From here follows that the eigenvalues of a lower triangular matrix are the diagonal elements and the algebraic multiplicity of an eigenvalue is as many times as it occurs in the diagonal.

Let us discuss some properties of the eigenvectors. It is obvious that if x is eigenvector then αx is also eigenvector where $\alpha \in K \setminus \{0\}$ is arbitrary. So the number of the eigenvectors corresponding to an eigenvalue is infinite. The proper question is the maximal number of the linearly independent eigenvectors.

9.5. Definition Let $A \in \mathbb{K}^{n\times n}$ and $\lambda \in \mathrm{Sp}\,(A)$. The subspace

$$W_\lambda := W_\lambda(A) := \{x \in \mathbb{K}^n \mid Ax = \lambda x\}$$

is called the eigenspace of the matrix A corresponding to the eigenvalue λ. The dimension of W_λ is called the geometric multiplicity of the eigenvalue λ and is denoted by $g(\lambda)$.

9.6. Remarks.

1. The eigenspace consists of the eigenvectors and the zero vector as elements.
2. Since the eigenvectors are the nontrivial solutions of the homogeneous linear system $(A-\lambda I)x = 0$ it follows that
$$g(\lambda) = \dim W_\lambda = \dim \mathcal{S}_h = n - \mathrm{rank}\,(A-\lambda I)\,.$$
3. It can be proved that for every $\lambda \in \mathrm{Sp}\,(A)$ holds
$$1 \le g(\lambda) \le a(\lambda) \le n\,.$$

9.2. Eigenvector Basis

9.7. Theorem *Let $A \in \mathbb{K}^{n\times n}$ and $\lambda_1, \dots, \lambda_k$ be some different eigenvalues of the matrix A. Let $s_i \in \mathbb{N}$, $1 \le s_i \le g(\lambda_i)$ and $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(s_i)}$ be a linearly independent system in the eigenspace W_{λ_i} $(i = 1, \dots, k)$. Then the united system*

$$x_i^{(j)} \in \mathbb{K}^n \qquad (i = 1, \dots, k;\ j = 1, \dots, s_i)$$

is linearly independent.

Let us take from the eigenspace W_λ the maximal number of linearly independent eigenvectors (this maximal number equals $g(\lambda)$). The united system – by the previous theorem – is linearly independent and its cardinality is $\sum\limits_{\lambda\in \mathrm{Sp}\,(A)} g(\lambda)$. So we can establish that

$$\sum_{\lambda\in \mathrm{Sp}\,(A)} g(\lambda) \le n\,.$$

If here stands „=" then we have n independent eigenvectors in $\mathbb{K}^n$ so we have a basis consisting of eigenvectors. This basis will be called Eigenvector Basis (E. B.).

It follows simply from the previous results that

$$\exists\ \text{E.B.} \quad \Leftrightarrow \quad \sum_{\lambda\in \mathrm{Sp}\,(A)} g(\lambda) = n\,.$$

9.8. Theorem *Let $A \in \mathbb{K}^{n\times n}$ and denote by $a(\lambda)$ its algebraic and by $g(\lambda)$ its geometric multiplicity. Then there exists Eigenvector Basis in $\mathbb{K}^n$ if and only if*

$$\sum_{\lambda\in \mathrm{Sp}\,(A)} a(\lambda) = n \qquad \textit{and} \qquad \forall\lambda \in \mathrm{Sp}\,(A)\ :\quad g(\lambda) = a(\lambda)\,.$$

Proof. On the lecture. □

9.9. Remark. The meaning of the condition $\sum\limits_{\lambda\in \mathrm{Sp}\,(A)} a(\lambda) = n$ is that the number of roots in $\mathbb{K}$ of the characteristic polynomial – counted with their multiplicities – equals n. Therefore

- If $\mathbb{K} = \mathbb{C}$ then this condition is „automatically" true.

- If $\mathbb{K} = \mathbb{R}$ then this condition holds if and only if every root of the characteristic polynomial is real.

9.3. Diagonalization

9.10. Definition (Similarity) Let $A, B \in \mathbb{K}^{n\times n}$. We say that the matrix B is similar to the matrix A (notation: $A \sim B$) if

$$\exists\, C \in \mathbb{K}^{n\times n}: \quad C \text{ is invertible and } B = C^{-1}AC\,.$$

9.11. Remark. The similarity relation is an equivalence relation (it is reflexive, symmetric and transitive). So we can use the phrase: A and B are similar (to each other).

9.12. Theorem *If $A \sim B$ then $P_A = P_B$ that is their characteristic polynomials coincide. Consequently the eigenvalues, their algebraic multiplicities and the determinants are equal.*

Proof. Let $A, B, C \in \mathbb{K}^{n\times n}$ and suppose that $B = C^{-1}AC$. Then for every $\lambda \in \mathbb{K}$:

$$\begin{aligned} P_B(\lambda) &= \det(B-\lambda I) = \det(C^{-1}AC - \lambda C^{-1}IC) = \det(C^{-1}(A-\lambda I)C) = \\ &= \det(C^{-1}) \cdot \det(A-\lambda I) \cdot \det(C) = \det(C^{-1}) \cdot \det(C) \cdot \det(A-\lambda I) = \\ &= \det(C^{-1}C) \cdot \det(A-\lambda I) = \det(I) \cdot P_A(\lambda) = 1 \cdot P_A(\lambda) = P_A(\lambda)\,. \end{aligned}$$

□

The following definition gives us an important class of square matrices.

9.13. Definition Let $A \in \mathbb{K}^{n\times n}$. We say that the matrix A is diagonalizable (over the field $\mathbb{K}$) if

$$\exists\, C \in \mathbb{K}^{n\times n}: \quad C \text{ is invertible and } C^{-1}AC \text{ is diagonal matrix}\,.$$

The matrix C is said to diagonalize A. The matrix $D = C^{-1}AC$ is called the diagonal form of A.

9.14. Remarks.

1. Obviously A is diagonalizable if and only if it is similar to a diagonal matrix.
2. A matrix A can have more than one diagonal form.
3. If A is diagonalizable then the diagonal entries of its diagonal form are the eigenvalues of A. More precisely, each eigenvalue stands in the diagonal as much times as its algebraic multiplicity.

The diagonalizability of a matrix is in close connection with the Eigenvector Basis as the following theorem shows:

9.15. Theorem *Let $A \in \mathbb{K}^{n\times n}$. The matrix A is diagonalizable (over the field $\mathbb{K}$) if and only if there exists Eigenvector Basis (E. B.) in $\mathbb{K}^n$.*

Proof. First suppose that A is diagonalizable. Let $c_1, \dots, c_n \in \mathbb{K}^n$ be the column vectors of C to diagonalize A. So

$$C = [c_1 \; \dots \; c_n] \; .$$

We will show that $c_1, \dots, c_n$ is Eigenvector Basis.

Since C is invertible so $c_1, \dots, c_n$ is a linearly independent system having n members. Consequently it is a basis in $\mathbb{K}^n$.

To show that the vectors c_j are eigenvectors, set out from the relation

$$C^{-1}AC = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. Multiply by C from the left:

$$A \cdot [c_1 \dots c_n] = C \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [c_1 \dots c_n] \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$[Ac_1 \dots Ac_n] = [\lambda_1 c_1 \dots \lambda_n c_n]$$

Using the equalities of the columns:

$$Ac_j = \lambda_j c_j \qquad (j = 1, \dots, n)$$

so the basis $c_1, \dots, c_n$ really consists of eigenvectors.

Conversely suppose that $c_1, \dots, c_n$ is an Eigenvector Basis. Let C be the matrix whose columns are $c_1, \dots, c_n$. Then C is obviously invertible, moreover, setting out from the equations

$$Ac_j = \lambda_j c_j \qquad (j = 1, \dots, n)$$

and making the previous operations backward we obtain

$$C^{-1}AC = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} .$$

So A is really diagonalizable. □

9.16. Remarks.

1. You can see that the order of the vectors of E. B. in the matrix C is identical with the order of the corresponding eigenvalues in the diagonal of $C^{-1}AC$.

2. If the matrix $A \in \mathbb{K}^{n \times n}$ has n different eigenvalues in $\mathbb{K}$ then the corresponding eigenvectors (n vectors) are linearly independent. So they form an Eigenvector Basis and by this reason A is diagonalizable.

9.4. Control Questions

1. Define the eigenvalue and the eigenvector of a matrix

2. Define the characteristic polynomial

3. Prove that the eigenvalues are the roots of the characteristic polynomial

4. Define the algebraic multiplicity of an eigenvalue

5. State and prove the statement about the eigenvalues of a triangle matrix

6. Prove that the eigenvectors to an eigenvalue and the zero vector together form a subspace. What is the name of this subspace?

7. Define the geometric multiplicity of an eigenvalue

8. What is the connection between the algebraic and the geometric multiplicity of an eigenvalue?

9. State (without proof) the theorem about the independence of eigenvectors

10. Define the concept of Eigenvector Basis (E. B.)

11. What is the necessary and sufficient condition of the existence of Eigenvector Basis?

12. Define the similarity of matrices

13. State and prove the theorem about the characteristic polynomials of similar matrices

14. Define the concept of diagonalizable matrix

15. What are the diagonal elements in the diagonal form of a diagonalizable matrix?

16. State and prove the necessary and sufficient condition of diagonalizability

9.5. Homework

1. Find the eigenvalues and the eigenvectors of the following matrices:

$$a)\quad \begin{bmatrix} 2 & -1 \\ 10 & -9 \end{bmatrix} \qquad b)\quad \begin{bmatrix} -2 & -7 \\ 1 & 2 \end{bmatrix} \qquad c)\quad \begin{bmatrix} 5 & 1 & 3 \\ 0 & -1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

2. Determine whether the following matrices are diagonalizable or not. In the diagonalizable case determine the matrix C that diagonalizes A and the diagonal form $C^{-1}AC$.

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

10. Lesson 10

10.1. Inner Product Spaces

10.1. Definition Let V be a vector space over the number field $\mathbb{K}$.
Let $V \times V \ni (x, y) \mapsto \langle x,\, y\rangle$ be an operation, which will be named inner product or scalar product or dot product.

Suppose that

1. $\forall\,(x, y) \in V \times V: \quad \langle x,\, y\rangle \in \mathbb{K}$ (the value of the inner product is a scalar)
2. $\forall\, x, y \in V: \quad \langle x,\, y\rangle = \overline{\langle y,\, x\rangle}$ (if $\mathbb{K} = \mathbb{R}$: commutative law; if $\mathbb{K} = \mathbb{C}$: antisymmetry)
3. $\forall\, x \in V\ \forall\, \lambda \in \mathbb{K}: \quad \langle \lambda x,\, y\rangle = \lambda\langle x,\, y\rangle$ (homogeneous)
4. $\forall\, x, y, z \in V: \quad \langle x,\, y + z\rangle = \langle x,\, y\rangle + \langle x,\, z\rangle$ (distributive law)
5. $\langle x,\, x\rangle \geq 0 \quad (x \in V)$, furthermore $\langle x,\, x\rangle = 0 \Leftrightarrow x = 0$ (positive definite)

Then we call V inner product space or Euclidean space. More precisely in the case $\mathbb{K} = \mathbb{R}$ we call it real inner product space, in the case $\mathbb{K} = \mathbb{C}$ we call it complex inner product space.

10.2. Examples

1. The vector space of the plane vectors and the vector space of the space vectors are real inner product spaces if the inner product is the common dot product
$$\langle a,\, b\rangle = |a| \cdot |b| \cdot \cos\gamma$$
where γ denotes the angle of vectors a and b.

2. The vector space $\mathbb{K}^n$ is inner product space if the inner product is
$$\langle x,\, y\rangle := \sum_{k=1}^{n} x_k \overline{y_k} = \underline{y}^* \underline{x} \qquad (x, y \in \mathbb{K}^n)\,,$$
where $\underline{x}$ and $\underline{y}$ denote the column matrix corresponding to the vectors x and y:
$$\underline{x} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \underline{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\,.$$

This is the standard inner product in $\mathbb{K}^n$. Naturally in the case $\mathbb{K} = \mathbb{R}$ there is no conjugation:

$$\langle x,\, y\rangle := \sum_{k=1}^{n} x_k y_k = \underline{y}^T \underline{x} \qquad (x, y \in \mathbb{R}^n)\,.$$

3. Let $-\infty < a < b < +\infty$. The vector space $C[a,b]$ of all continuous functions defined on $[a,b]$ a mapping into $\mathbb{K}$ form an inner product space if the inner product is

   - in the case $\mathbb{K} = \mathbb{C}$: $\langle f,\, g\rangle := \int_a^b f(x)\overline{g(x)}\; dx$.
   - in the case $\mathbb{K} = \mathbb{R}$: $\langle f,\, g\rangle := \int_a^b f(x)g(x)\; dx$.

   This is the standard inner product in $C[a,b]$.

4. Since the polynomial vector spaces $\mathcal{P}[a,b]$, $\mathcal{P}_n[a,b]$ are subspaces of $C[a,b]$, so they are also inner product spaces with the inner product defined in the previous example.

Some basic properties of the inner product spaces follow.

10.3. Theorem *Let V be an inner product space over $\mathbb{K}$.*
Then for every $x,\, x_i,\, y,\, y_j,\, z \in V$ and for every $\lambda,\, \lambda_i,\, \mu_j \in \mathbb{K}$ hold

1. $\langle x,\, \lambda y\rangle = \overline{\lambda} \cdot \langle x,\, y\rangle$
2. $\langle x+y,\, z\rangle = \langle x,\, z\rangle + \langle y,\, z\rangle$
3. $\langle \sum_{i=1}^{n} \lambda_i x_i,\, \sum_{j=1}^{m} \mu_j y_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_i \overline{\mu_j} \langle x_i,\, y_j\rangle$ *Naturally in the real case $\mathbb{K} = \mathbb{R}$ there is no conjugation:* $\langle \sum_{i=1}^{n} \lambda_i x_i,\, \sum_{j=1}^{m} \mu_j y_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_i \mu_j \langle x_i,\, y_j\rangle$
4. $\langle x,\, 0\rangle = \langle 0,\, x\rangle = 0$

Proof.

1. $\langle x,\, \lambda y\rangle = \overline{\langle \lambda y,\, x\rangle} = \overline{\lambda \cdot \langle y,\, x\rangle} = \overline{\lambda} \cdot \overline{\langle y,\, x\rangle} = \overline{\lambda} \cdot \langle x,\, y\rangle$.
2. $\langle x+y,\, z\rangle = \overline{\langle z,\, x+y\rangle} = \overline{\langle z,\, x\rangle + \langle z,\, y\rangle} = \overline{\langle z,\, x\rangle} + \overline{\langle z,\, y\rangle} = \langle x,\, z\rangle + \langle y,\, z\rangle$.
3. Apply several times the axioms and the previous properties:

$$\langle \sum_{i=1}^{n} \lambda_i x_i,\, \sum_{j=1}^{m} \mu_j y_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \langle \lambda_i x_i,\, \mu_j y_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_i \overline{\mu_j} \cdot \langle x_i,\, y_j\rangle.$$

4. $\langle x,\, 0\rangle = \langle x,\, 0+0\rangle = \langle x,\, 0\rangle + \langle x,\, 0\rangle$. After subtraction $\langle x,\, 0\rangle$ from both sides we obtain the first statement. The other one can reduce to the first.

$\square$

10.2. The Cauchy's inequality

10.4. Theorem *[Cauchy's inequality] Let V be an inner product space and let $x, y \in V$. Then*

$$|\langle x,\, y\rangle| \le \sqrt{\langle x,\, x\rangle} \cdot \sqrt{\langle y,\, y\rangle}\,.$$

Here stands equality if and only if the vector system x, y is linearly dependent (x and y are parallel).

Proof. We will prove the statement of the theorem only in the case $\mathbb{K} = \mathbb{R}$. Let us observe that for any $\lambda \in \mathbb{R}$:

$$\begin{aligned} 0 \le \langle x+\lambda y,\, x+\lambda y\rangle &= \langle x,\, x\rangle + \lambda\langle y,\, x\rangle + \lambda\langle x,\, y\rangle + \lambda\lambda\langle y,\, y\rangle = \\ &= (\langle y,\, y\rangle)\lambda^2 + (2\langle x,\, y\rangle)\lambda + \langle x,\, x\rangle = P(\lambda)\,. \end{aligned}$$

So the above defined second degree polynomial P takes nonnegative values everywhere.

Suppose first that x and y are linearly independent. Then for any $\lambda \in \mathbb{R}$ holds $x + \lambda y \ne 0$ so $P(\lambda) > 0$ for any $\lambda \in \mathbb{R}$. That means that the discriminant of P is negative:

$$\text{discriminant } = (2\langle x,\, y\rangle)^2 - 4(\langle y,\, y\rangle)(\langle x,\, x\rangle) < 0\,.$$

After division by 4 and rearranging the inequality we obtain that

$$|\langle x,\, y\rangle| < \sqrt{\langle x,\, x\rangle} \cdot \sqrt{\langle y,\, y\rangle}\,.$$

Now suppose that x and y are linearly dependent. Then $x + \lambda y = 0$ holds for some $\lambda \in \mathbb{R}$. That means $P(\lambda) = 0$ so the nonnegative second degree polynomial P has a real root. Consequently its discriminant equals 0:

$$\text{discriminant } = (2\langle x,\, y\rangle)^2 - 4(\langle y,\, y\rangle)(\langle x,\, x\rangle) = 0\,.$$

After rearranging the equation we obtain that

$$|\langle x,\, y\rangle| = \sqrt{\langle x,\, x\rangle} \cdot \sqrt{\langle y,\, y\rangle}\,.$$

From the proved parts immediately follow the statements of the theorem. $\square$

10.5. Remark. Apply the Cauchy's inequality in $\mathbb{R}^n$:

$$(x_1y_1 + \cdots + x_ny_n)^2 \le (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) \qquad (i_i,\, y_i \in \mathbb{R})$$

and equality holds if and only if the vectors $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$ are linearly dependent (parallel). This is the well-known Cauchy-Bunyakovsky-Schwarz inequality.

10.3. Norm

In this section the concept of the length of vectors will be extended (in other words: the distances of points from the origin).

10.6. Definition Let V be an inner product space and let $x \in V$. Then its norm (or length or absolute value) is defined as

$$\|x\| := \sqrt{\langle x,\, x\rangle}\,.$$

The mapping $\|.\| : V \to \mathbb{R},\ x \mapsto \|x\|$ is called norm too.

10.7. Examples

1. In the inner product space of plane vectors or of the space vectors the norm of a vector a coincides with the classical length of a:

$$\|a\| = \sqrt{\langle a,\, a\rangle} = \sqrt{|a|\cdot|a|\cdot\cos(a,a)} = |a|\,.$$

2. In $\mathbb{C}^n$: $\|x\| = \sqrt{\sum\limits_{i=1}^{n} |x_i|^2}$.

   In $\mathbb{R}^n$: $\|x\| = \sqrt{\sum\limits_{i=1}^{n} x_i^2}$.

3. In $C[a,b]$: $\|f\| = \sqrt{\int\limits_a^b |f(x)|^2\,dx}$.

10.8. Remark. Using the notation of norm the Cauchy's inequality can be written as

$$|\langle x,\, y\rangle| \le \|x\|\cdot\|y\| \qquad\qquad (x, y \in V)\,.$$

10.9. Theorem *[the properties of the norm]*

1. $\|x\| \ge 0 \quad (x \in V)$. *Furthermore* $\|x\| = 0 \Leftrightarrow x = 0$
2. $\|\lambda x\| = |\lambda|\cdot\|x\| \qquad (x \in V;\ \lambda \in \mathbb{K})$
3. $\|x + y\| \le \|x\| + \|y\| \qquad (x, y \in V)$ *(triangle inequality)*

Proof. The first statement is obvious by the axioms of the inner product. The proof of the second statement is as follows:

$$\|\lambda x\| = \sqrt{\langle \lambda x,\, \lambda x\rangle} = \sqrt{\lambda\overline{\lambda}\langle x,\, x\rangle} = \sqrt{|\lambda|^2\cdot\|x\|^2} = |\lambda|\cdot\|x\|\,.$$

To see the triangle inequality let us see the following computations:

$$\begin{aligned}\|x+y\|^2 &= \langle x+y,\, x+y\rangle = \langle x,\, x\rangle + \langle x,\, y\rangle + \langle y,\, x\rangle + \langle y,\, y\rangle = \\ &= \|x\|^2 + \langle x,\, y\rangle + \overline{\langle x,\, y\rangle} + \|y\|^2 = \|x\|^2 + 2\mathrm{Re}\,(\langle x,\, y\rangle) + \|y\|^2 \leq \\ &\leq \|x\|^2 + 2\cdot|\langle x,\, y\rangle| + \|y\|^2 \leq \|x\|^2 + 2\cdot\|x\|\cdot\|y\| + \|y\|^2 = \\ &= (\|x\| + \|y\|)^2\,.\end{aligned}$$

(In the last estimation we have used the Cauchy's inequality.)

After taking square roots we obtain the triangle inequality. □

10.10. Remark. If we define on a vector space a mapping $\|.\| : V \to \mathbb{R}$ which satisfies the above properties then V is called (linear) normed space and the above properties are named the axioms of the normed space. So we have proved that every inner product space is a normed space with the norm indicated by the inner product $\|x\| = \sqrt{\langle x,\, x\rangle}$.

Other examples for norms and normed spaces will be studied in the subject Numerical Methods.

10.11. Definition (distance in the inner product space) Let V be an inner product space, $x, y \in V$. The number

$$d(x,y) := \|x-y\| = \sqrt{\langle x-y,\, x-y\rangle}$$

is called the distance between the vectors x and y.

10.12. Remark. The above defined distance in $\mathbb{R}^n$ is

$$d(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \ldots + (x_n-y_n)^2} \qquad (x,\, y \in \mathbb{R}^n)\,.$$

10.4. Orthogonality

Let V be an inner product space over the number field $\mathbb{K}$ all over in this section.

10.13. Definition The vectors $x, y \in V$ are called orthogonal (or: perpendicular) if their inner product equals 0 that is if

$$\langle x,\, y\rangle = 0\,.$$

The notation of orthogonality is $x \perp y$.

10.14. Definition Let $\emptyset \neq H \subset V$ and $x \in V$. We say that the vector x is orthogonal (or: perpendicular) to the set H (notation: $x \perp H$) if

$$\forall\, y \in H: \qquad \langle x,\, y\rangle = 0.$$

10.15. Theorem *Let $e_1, \dots, e_n$ be vector system in V, $W := \operatorname{Span}(e_1, \dots, e_n)$ and $x \in V$. Then*

$$x \perp W \quad \Leftrightarrow \quad x \perp e_i \ (i = 1, \dots, n).$$

Proof. „$\Rightarrow$": It is obvious if you choose $y := e_i$.
„$\Leftarrow$": Let $y = \sum\limits_{i=1}^{n} \lambda_i e_i \in W$ arbitrary. Then

$$\langle x, y \rangle = \langle x, \sum_{i=1}^{n} \lambda_i e_i \rangle = \sum_{i=1}^{n} \overline{\lambda_i} \langle x, e_i \rangle = \sum_{i=1}^{n} \overline{\lambda_i} \cdot 0 = 0.$$

$\square$

10.16. Definition Let $x_i \in V$ $(i \in I)$ a (finite or infinite) vector system.

1. This system $(x_i,\ i \in I)$ is said to be orthogonal system (O.S.) if any two members of them are orthogonal that is
$$\forall\, i, j \in I,\ i \neq j : \qquad \langle x_i, x_j \rangle = 0.$$
2. The system $(x_i,\ i \in I)$ is said to be orthonormal system (O.N.S.) if it is orthogonal system and each vector in it has the norm 1:
$$\forall\, i, j \in I : \qquad \langle x_i, x_j \rangle = \begin{cases} 0 \text{ ha } i \neq j \\ 1 \text{ ha } i = j. \end{cases}$$

10.17. Remarks.

1. One can simply see that
   - the zero vector can be contained in an orthogonal system
   - the zero vector cannot be contained in an orthonormal system
   - the zero vector may occur several times in an orthogonal system but any other vector may occur only one times in it.
   - the vectors in an orthonormal system are all different
2. (Normalization) One can construct orthonormal system from an orthogonal system such that the two systems generate the same subspace. Really, first leave the possible zero vectors from the orthogonal system, after it divide every vector in the remainder system by its norm.

10.18. Examples

1. In the inner product space of the plane vectors the system of the common basic vectors $\mathbf{i}$, $\mathbf{j}$ is O.N.S.
2. In the inner product space of the space vectors the system of the common basic vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ is O.N.S.
3. In the space $\mathbb{K}^n$ he system of the standard unit vectors $e_1, \dots, e_n$ is O.N.S.

10.5. Two important theorems for finite orthogonal systems

10.19. Theorem *If* $x_1, \dots, x_n \in V \setminus \{0\}$ *is an orthogonal system then it is linearly independent.*

Proof. Multiply the dependence equation

$$0 = \sum_{i=1}^{n} \lambda_i x_i$$

by the vector x_j where $j = 1, \dots, n$:

$$0 = \langle 0, x_j \rangle = \langle \sum_{i=1}^{n} \lambda_i x_i, x_j \rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, x_j \rangle = \lambda_j \langle x_j, x_j \rangle .$$

Since $\langle x_j, x_j \rangle \neq 0$ so $\lambda_j = 0$. □

10.20. Theorem *[Pythagorean Theorem] If* $x_1, \dots, x_n \in V$ *is an orthogonal system then*

$$\| \sum_{i=1}^{n} x_i \|^2 = \sum_{i=1}^{n} \|x_i\|^2 .$$

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 $\langle x_i, x_j \rangle = 0$ if $i \neq j$.) □

10.6. Control Questions

1. Define the concept of Inner Product Space (Euclidean Space)

2. Give 3 examples for Euclidean space

3. State and prove the 4 basic properties of an Euclidean Space

4. State and prove the theorem about the Cauchy's Inequality
5. Define the norm of a vector
6. State and prove the theorem about the 3 properties of the norm
7. Define the following concepts: orthogonality of two vectors, orthogonal system (O.S.), orthonormal system (O.N.S.)
8. Give 3 examples for orthogonal systems
9. State and prove the theorem about a vector that is perpendicular to a finite dimensional subspace
10. State and prove the theorem about the independence of an orthogonal system
11. State and prove the Pythagorean theorem

10.7. Homework

1. Let $x = (3, -2, 1, 1),\ y = (4, 5, 3, 1)\ z = (-1, 6, 2, 0) \in \mathbb{R}^4$ and let $\lambda = -4$. Verify the following identities:

   a) $\langle x,\, y\rangle = \langle y,\, x\rangle$
   b) $\langle x + y,\, z\rangle = \langle x,\, z\rangle + \langle y,\, z\rangle$
   c) $\langle \lambda x,\, y\rangle = \lambda\langle x,\, y\rangle$

   Remark that in $\mathbb{R}^4$ we use the usual operations.

2. Verify the Cauchy's inequality in $\mathbb{R}^4$ with the vectors
$$x = (0, -2, 2, 1) \quad \text{and} \quad y = (-1, -1, 1, 1)\,.$$

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 the vector $x = (-1, 1, 0, 2)$ is orthogonal to the subspace Span $(x_1,\, x_2,\, x_3)$ or not.

11. Lesson 11

11.1. The Projection Theorem

11.1. Theorem *[Projection Theorem] Let $u_1, \dots, u_n \in V \setminus \{0\}$ be an orthogonal system, $W := \operatorname{Span}(u_1, \dots, u_n)$. (It is important to remark that in this case $u_1, \dots, u_n$ is basis in W.) Then every $x \in V$ can be written uniquely as $x = x_1 + x_2$ where $x_1 \in W$ and $x_2 \perp W$.*

Proof. Look for x_1 as

$$x_1 := \sum_{j=1}^{n} \lambda_j \cdot u_j \quad \text{and let} \quad x_2 := x - x_1 .$$

Then obviously $x_1 \in W$ and $x = x_1 + x_2$ independently of the coefficients λ_i. It remains to satisfy the requirement $x_2 \perp W$. It is enough to discuss the orthogonality to the generator system $u_1, \dots, u_n$:

$$\begin{aligned} \langle x_2, u_i \rangle &= \langle x - \sum_{j=1}^{n} \lambda_j u_j, u_i \rangle = \langle x, u_i \rangle - \sum_{j=1}^{n} \lambda_j \langle u_j, u_i \rangle = \\ &= \langle x, u_i \rangle - \lambda_i \langle u_i, u_i \rangle \qquad (i = 1, \dots, n). \end{aligned}$$

This expression equals 0 if and only if

$$\lambda_i = \frac{\langle x, u_i \rangle}{\langle u_i, u_i \rangle} \qquad (i = 1, \dots, n) .$$

Since the numbers λ_i are obtained by a unique process and $u_1, \dots, u_n$ are linearly independent then x_1 and x_2 are unique. □

11.2. Remarks.

1. The vector x_1 is called the orthogonal projection of x onto W and is denoted by $\operatorname{proj}_W x$ or simply $P(x)$. From the theorem follows that

$$P(x) = \operatorname{proj}_W x = \sum_{i=1}^{n} \frac{\langle x, u_i \rangle}{\langle u_i, u_i \rangle} \cdot u_i .$$

   Another name for $P(x)$ is: the parallel component of x with respect to the subspace W.

2. The vector x_2 is called the orthogonal component of x with respect to the subspace W and is denoted by $Q(x)$. From the theorem follows that

$$Q(x) = x - P(x) = x - \sum_{i=1}^{n} \frac{\langle x,\, u_i \rangle}{\langle u_i,\, u_i \rangle} \cdot u_i \,.$$

If we introduce the subspace

$$W^{\perp} := \{x \in V \mid x \perp W\}$$

then $Q(x)$ can be regarded as the orthogonal projection onto $W^{\perp}$:

$$Q(x) = \operatorname{proj}_{W^{\perp}} x \,.$$

11.2. The Gram-Schmidt Process

Let $b_1,\, b_2,\, \ldots,\, b_n \in V$ be a finite linear independent system. The following process converts this system into an orthogonal system $u_1,\, u_2,\, \ldots,\, u_n \in V \setminus \{0\}$. The two system is equivalent in the sense that

$$\forall\, k =\in \{1, 2, , \ldots, n\} : \quad \operatorname{Span}(b_1,\, \ldots,\, b_k) = \operatorname{Span}(u_1,\, \ldots,\, u_k)\,.$$

Especially (for $k = n$) the generated subspaces by the two systems are the same.

The algorithm of the Gram-Schmidt process is as follows:

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 results the system $u_1,\, u_2,\, \ldots,\, u_n$ that satisfies all the requirements described in the introduction of the section. If we want to construct an equivalent orthonormal system then apply the normalization process for $u_1,\, u_2,\, \ldots,\, u_n$.

11.3. Remark. One can see that

- u_2 is the orthogonal component of b_2 with respect to the subspace $\operatorname{Span}(u_1)$
- u_3 is the orthogonal component of b_3 with respect to the subspace $\operatorname{Span}(u_1,\, u_2)$

  $\vdots$
- u_n is the orthogonal component of b_n with respect to the subspace $\operatorname{Span}(u_1,\, u_2,\, \ldots,\, u_{n-1})$.

11.3. Orthogonal and Orthonormal Bases

11.4. Definition A finite vector system in the inner product space V is called

- Orthogonal Basis (O.B.) if it is orthogonal system and basis.
- Orthonormal Basis (O.N.B.) if it is orthonormal system and basis.

11.5. Remarks.

1. An O.B. cannot contain the zero vector.
2. An orthogonal system that does not contain the zero vector is O.B. if and only if it is a generator system in V.
3. If we have an O.B. then we can construct from it – via normalization – an O.N.B..

On can easily verify that in $\mathbb{K}^n$ the standard basis is orthonormal basis.

It can be proved that every finite dimensional nonzero inner product space contains orthogonal and orthonormal basis. Moreover, every orthogonal system that does not contain the zero vector can be completed into orthogonal basis and every orthonormal system can be completed into orthonormal basis. The essential idea of the proof is:

Construct a basis and apply the Gram-Schmidt process for it.

11.6. Remark. The existence of the orthogonal basis implies that the projection theorem can be stated for every finite dimensional nonzero subspace of V.

In the remainder part of the section let us fix an orthonormal basis $e : e_1, \dots, e_n$ in the n-dimensional inner product space V. We will prove first that the inner product can be computed with the help of coordinates.

11.7. Theorem

$$\forall\, x, y \in V: \quad \langle x,\, y\rangle = \langle [x]_e,\, [y]_e\rangle = \sum_{i=1}^{n} \xi_i \overline{\eta_i}\,.$$

Here $[x]_e = (\xi_1, \dots, \xi_n)$ *and* $[y]_e = (\eta_1, \dots, \eta_n)$ *are the coordinate vectors of* x *and* y*.*

Proof. Since

$$x = \sum_{i=1}^{n} \xi_i e_i \quad \text{and} \quad y = \sum_{j=1}^{n} \eta_j e_j$$

so

$$\langle x,\, y\rangle = \langle \sum_{i=1}^{n} \xi_i e_i,\, \sum_{j=1}^{n} \eta_j e_j\rangle = \sum_{i,j=1}^{n} \xi_i \overline{\eta_j} \langle e_i,\, e_j\rangle = \sum_{i=1}^{n} \xi_i \overline{\eta_i} =$$

$$= \langle (\xi_1, \dots, \xi_n),\, (\eta_1, \dots, \eta_n)\rangle = \langle [x]_e,\, [y]_e\rangle .$$

□

11.8. Corollary. Apply the theorem for $y = x$. Then we obtain:

$$\| x\|^2 = \langle x,\, x\rangle = \langle [x]_e,\, [x]_e\rangle = \sum_{i=1}^{n} \xi_i \overline{\xi_i} = \sum_{i=1}^{n} |\, \xi_i|^2 \,,$$

especially in the case $\mathbb{K} = \mathbb{R}$:

$$\| x\|^2 = \sum_{i=1}^{n} \xi_i^2 \,.$$

The following theorem gives us the coordinates relative to an orthonormal basis.

11.9. Theorem *The i-th coordinate of a vector $x \in V$ relative to the orthonormal basis $e : e_1, \dots, e_n$ is*

$$\xi_i = \langle x,\, e_i\rangle \qquad (i = 1, \dots, n)$$

That is

$$x = \sum_{i=1}^{n} \langle x,\, e_i\rangle \cdot e_i \,.$$

This formula is the finite Fourier-expansion of x. The coefficients $\langle x,\, e_i\rangle$ are called the Fourier-coefficients of x relative to the orthonormal system (e_i).

Proof. Apply the previous theorem with $y = e_i$ $(i = 1, \dots, n)$. Then we obtain

$$\langle x,\, e_i\rangle = \langle [x]_e,\, [e_i]_e\rangle = \langle (\xi_1, \dots, \xi_n),\, (0, \dots, 1, \dots, 0)\rangle = \xi_i \qquad (i = 1, \dots, n)\,.$$

□

11.10. Remark. One can simply consider – using the normalization process – that the ith coordinate of a vector $x \in V$ relative to an orthogonal basis $u_1, \ldots, u_n$ is

$$\xi_i = \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \qquad (i = 1, \ldots, n)\,.$$

Consequently

$$x = \sum_{i=1}^{n} \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \cdot u_i\,.$$

This formula is the finite Fourier-expansion of x. The coefficients $\dfrac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle}$ are called the Fourier-coefficients of x relative to the orthogonal system (u_i).

11.4. Control Questions

1. State and prove the Projection Theorem

2. Give (without proof) the formulas of the parallel and of the orthogonal components of a vector relative to a subspace generated by a finite orthogonal system

3. Describe the Gram-Schmidt-process

4. Define the concept of orthogonal basis (O.B.) and of orthonormal basis (O.N.B.).

5. What a statement is known about the existence of the orthogonal basis (O.B.) in a finite dimensional Euclidean Space?

6. State and prove the theorem about the computation of scalar product with coordinates relative to an orthonormal basis

7. What is the (finite) Fourier-expansion of a vector in a finite dimensional Euclidean Space?

11.5. Homework

1. Find the orthogonal projections of the vector $x = (1, 2, 0, -2) \in \mathbb{R}^4$ onto the subspaces of $\mathbb{R}^4$ generated by the given 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. Use the Gram-Schmidt process to transform the given basis b_1, b_2, b_3, b_4 of $\mathbb{R}^4$ into an orthonormal basis.

$$b_1 = (0, 2, 1, 0),\ b_2 = (1, -1, 0, 0),\ b_3 = (1, 2, 0, -1),\ b_4 = (1, 0, 0, 1)$$

3. Show that the vectors

$$u_1 = (1, -2, 3, -4),\ u_2 = (2, 1, -4, -3),\ u_3 = (-3, 4, 1, -2),\ u_4 = (4, 3, 2, 1)$$

   form an orthogonal basis in $\mathbb{R}^4$. Find the coordinates and the coordinate vector of $x = (-1, 2, 3, 7)$ relative to the given basis.

   Answer the same questions if the basis is the orthonormal basis obtained from u_1, u_2, u_3, u_4 via normalization.

12. Lesson 12

12.1. Some statements in $\mathbb{C}^n$

Let us look at the Euclidean Space $\mathbb{C}^n$ with the standard inner product

$$\langle x,\, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i} = \underline{y}^* \underline{x} \qquad (x, y \in \mathbb{C}^n)\,.$$

Here $\underline{x}$ and $\underline{y}$ denote the column matrix corresponding to the vectors x and y (see Examples 10.2):

$$\underline{x} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \underline{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\,.$$

Using the short notation of the inner product we can easily prove the following theorem.

12.1. Theorem *Let $A \in \mathbb{K}^{m\times n}$ be an $m \times n$ matrix. Then*

$$\langle Ax,\, y\rangle = \langle x,\, A^* y\rangle \qquad (x \in \mathbb{K}^n,\ y \in \mathbb{K}^m)\,,$$

Here A^* denotes the (Hermitian) adjoint of A (see Section 1.2.).

Proof. For any $x \in \mathbb{K}^n,\ y \in \mathbb{K}^m$ holds

$$\langle Ax,\, y\rangle = y^*(Ax) = (y^*A)x = (y^*(A^*)^*)x = (A^*y)^*x = \langle x,\, A^*y\rangle\,.$$

□

12.2. Corollary. Let $A \in \mathbb{K}^{n\times m}$ be an $n \times m$ matrix. Then

$$\langle x,\, Ay\rangle = \langle A^*x,\, y\rangle \qquad (x \in \mathbb{K}^n,\ y \in \mathbb{K}^m).$$

Really, apply the previous theorem for $A^* \in \mathbb{K}^{m\times n}$. We have

$$\langle A^*x,\, y\rangle = \langle x,\, (A^*)^*y\rangle = \langle x,\, Ay\rangle\,.$$

In the following theorem we express an eigenvalue of a matrix with the help of an eigenvector associated with this eigenvalue:

12.3. Theorem *Let $A \in \mathbb{K}^{n\times n}$ and $\lambda \in \mathrm{Sp}\,(A)$ and $x \in \mathbb{K}^n$ an eigenvector associated with λ. Then*

$$\lambda = \frac{\langle Ax,\, x\rangle}{\|x\|^2}\,.$$

This fraction is called Rayleigh-Ritz-quotient.

Proof.

$$\langle Ax,\, x\rangle = \langle \lambda x,\, x\rangle = \lambda\langle x,\, x\rangle = \lambda\|x\|^2\,,$$

whereby the statement can be deduced by division with $||x||^2$. □

12.2. Self-adjoint matrices

12.4. Definition A square matrix $A \in \mathbb{C}^{n\times n}$ is said to be self-adjoint (or it is said to be Hermitian) if

$$A^* = A\,.$$

In other words:

$$a_{ij} = \overline{a_{ji}} \qquad (i, j = 1,\, \ldots, n)\,.$$

If the entries of A are all real numbers, then the self-adjoint matrices are often named real symmetric matrices. In this case A can be regarded as an element of $\mathbb{R}^{n\times n}$, and we say simply that A is symmetric. Naturally, in this case the above definition has the form

$$A^T = A, \quad \text{that is} \quad a_{ij} = a_{ji} \qquad (i, j = 1,\, \ldots, n)\,.$$

12.5. Theorem *Let $A \in \mathbb{C}^{n\times n}$ be a self-adjoint matrix. Then*

1. $\langle Ax,\, y\rangle = \langle x,\, Ay\rangle \qquad (x,\, y \in \mathbb{C}^n)\,.$
2. $\langle Ax,\, x\rangle \in \mathbb{R} \qquad (x \in \mathbb{C}^n)\,.$
3. $\mathrm{Sp}\,(A) \subset \mathbb{R}\,.$

   This statement implies that if $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, then all the roots of its characteristic polynomial are real. More precisely, the number of real roots of the characteristic polynomial – counted with their multiplicities – equals n.

4. *If $\lambda, \mu \in \mathrm{Sp}\,(A)$, $\lambda \neq \mu$, $x \in W_\lambda$, $y \in W_\mu$, then $\langle x,\, y\rangle = 0$.*

   This means that the eigenspaces of a self-adjoint matrix are pairwise orthogonal.

   The statement is naturally true if $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix. In other words, the eigenspaces of a real symmetric matrix are pairwise orthogonal.

Proof.

1. Apply Theorem 12.1 and that $A = A^*$:
$$\langle Ax,\, y\rangle = \langle x,\, A^*y\rangle = \langle x,\, Ay\rangle \qquad (x, y \in \mathbb{C}^n)\,.$$

2. Apply the previous part with $x = y$ and use the antisymmetry of the scalar product:
$$\langle Ax,\, x\rangle = \langle x,\, Ax\rangle = \overline{\langle Ax,\, x\rangle}\,,$$
which implies immediately that $\langle Ax,\, x\rangle \in \mathbb{R}$.

3. By the previous part of the theorem $\langle Ax,\, x\rangle \in \mathbb{R}$. Then by Theorem 12.3 λ is a quotient of two real numbers, consequently it is real:
$$\lambda = \frac{\langle Ax,\, x\rangle}{\|x\|^2} \in \mathbb{R}\,.$$

4. Use the proved parts of the theorem, then
$$\lambda\langle x,\, y\rangle = \langle \lambda x,\, y\rangle = \langle Ax,\, y\rangle = \langle x,\, Ay\rangle = \langle x,\, \mu y\rangle = \overline{\mu}\langle x,\, y\rangle = \mu\langle x,\, y\rangle\,.$$
After rearrangement we have
$$(\lambda - \mu)\cdot\langle x,\, y\rangle = 0\,.$$
Since $\lambda - \mu \neq 0$, then really $\langle x,\, y\rangle = 0$.

□

12.3. Unitary matrices

12.6. Definition A square matrix $A \in \mathbb{C}^{n\times n}$ is said to be unitary if
$$A^*A = I\,,$$
where I denotes the identity matrix of size $n \times n$.

If the entries of A are all real numbers, then the unitary matrices are often named real orthogonal matrices. In this case A can be regarded as an element of $\mathbb{R}^{n\times n}$, and we say simply that A is orthogonal. Naturally, in this case the above definition has the form
$$A^TA = I\,.$$

The following theorem expresses the alternative possibilities of the definition of the unitary matrix.

12.7. Theorem *Let $A \in \mathbb{C}^{n\times n}$. Then the following statements are equivalent.*

(a) *A is unitary.*

(b) *A is regular. In this case* $A^{-1} = A^*$ *.*

(c) $AA^* = I$ *, or equivalently:* A^* *is unitary.*

(d) *The column vectors of A form an orthonormal basis (O.N.B.) in* $\mathbb{C}^n$*.*

(e) *The row vectors of A form an orthonormal basis (O.N.B.) in* $\mathbb{C}^n$*.*

Proof. Corollary 2.11 shows us that $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

To prove the equivalence between (a) and (d) let $A = [a_1 \, \ldots \, a_n]$. We have

$$(A^*A)_{ij} = \Big(\begin{bmatrix} a_1^* \\ \vdots \\ a_n^* \end{bmatrix} \cdot \begin{bmatrix} a_1 \, \ldots \, a_n \end{bmatrix} \Big)_{ij} = \underline{a_i}^* \underline{a_j} = \langle a_j , \, a_i \rangle \qquad (i = 1, \ldots, n) \, ,$$

which implies that

$$\begin{aligned} A \text{ is unitary} &\Leftrightarrow A^*A = I \Leftrightarrow (A^*A)_{ij} = \delta_{ij} \Leftrightarrow \langle a_j , \, a_i \rangle = \delta_{ij} \Leftrightarrow \\ &\Leftrightarrow a_1, \ldots, a_n \text{ an orthonormal system with } n \text{ terms} \Leftrightarrow \\ &\Leftrightarrow a_1, \ldots, a_n \text{ is an O.N.B.} . \end{aligned}$$

Here we have used the Kronecker-symbol

$$\delta_{ij} := \begin{cases} 0 & \text{if} \quad i \neq j \\ \\ 1 & \text{if} \quad i = j \end{cases}$$

Finally, the equivalence between (a) and (e) can be proved similarly.. □

12.8. Theorem *Let* $A \in \mathbb{C}^{n \times n}$ *be an unitary matrix. Then*

1. $\langle Ax , \, Ay \rangle = \langle x , \, y \rangle \qquad (x, \, y \in \mathbb{C}^n)$.

2. $\| \, Ax \| = \| \, x \| \qquad (x \in \mathbb{C}^n)$.

3. $\forall \, \lambda \in \mathrm{Sp}\,(A) : \quad | \, \lambda | = 1$.

Remark that if $A \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix, then we do not have a particular result. Even in this case the eigenvalues are not necessarily real.*

4. If $\lambda \in \mathrm{Sp}\,(A), \; x \in W_\lambda$*, then* $A^*x = \overline{\lambda} x$ *.*

5. If $\lambda, \mu \in \mathrm{Sp}\,(A), \; \lambda \neq \mu, \; x \in W_\lambda, \; y \in W_\mu$*, then* $\langle x , \, y \rangle = 0$.

This means that the eigenspaces of a unitary matrix are pairwise orthogonal.

The statement is naturally true if $A \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix. In other words, the eigenspaces of a real orthogonal matrix are pairwise orthogonal.*

Proof.

1. Apply Theorem 12.1 and that $A^*A = I$:
$$\langle Ax,\, Ay\rangle = \langle x,\, A^*Ay\rangle = \langle x,\, Iy\rangle = \langle x,\, y\rangle \qquad (x, y \in \mathbb{C}^n)\,.$$
2. Apply the previous part with $x = y$ and use the definition of the norm:
$$\|\, Ax\|^2 = \langle Ax,\, Ax\rangle = \langle x,\, x\rangle = \|\, x\|^2\,,$$
which implies immediately $\|\, Ax\| = \|\, x\|$.
3. Using the previous part of the theorem we have
$$\|\, x\| = \|\, Ax\| = \|\, \lambda x\| = |\, \lambda| \cdot \|\, x\|\,.$$
After simplification with $\|\, x\|$ we obtain $|\, \lambda| = 1$.
4. $Ax = \lambda x$ implies $A^*Ax = \lambda A^*x$. Since $A^*A = I$, we have
$$\lambda A^*x = x\,.$$
Multiply this equation by $\overline{\lambda}$:
$$\overline{\lambda}\lambda A^*x = \overline{\lambda}x \qquad \text{that is} \quad |\, \lambda|^2 A^*x = \overline{\lambda}x\,.$$
In the previous part we have shown $|\, \lambda| = 1$, thus $A^*x = \overline{\lambda}x$.
5. Use the proved parts of the theorem, then
$$\lambda\langle x,\, y\rangle = \langle \lambda x,\, y\rangle = \langle Ax,\, y\rangle = \langle x,\, A^*y\rangle = \langle x,\, \overline{\mu}y\rangle = \mu\langle x,\, y\rangle\,.$$
After rearrangement we have
$$(\lambda - \mu) \cdot \langle x,\, y\rangle = 0\,.$$
Since $\lambda - \mu \neq 0$, then really $\langle x,\, y\rangle = 0$.

□

12.4. Unitary and orthogonal diagonalization

12.9. Definition 1. Let $A \in \mathbb{C}^{n\times n}$.

We say that A is unitarily diagonalizable if
$$\exists\, U \in \mathbb{C}^{n\times n} \quad \text{unitary matrix :} \quad U^*AU \text{ is a diagonal matrix}\,.$$
U is called a unitarily diagonalizing matrix to A

2. Let $A \in \mathbb{R}^{n\times n}$.

   We say that A is orthogonally diagonalizable if

   $$\exists\, Q \in \mathbb{R}^{n\times n} \quad \text{orthogonal matrix :} \quad Q^T AQ \text{ is a diagonal matrix}.$$

   Q is called an orthogonally diagonalizing matrix to A

Since $U^* = U^{-1}$ and $Q^T = Q^{-1}$, then the unitary and the orthogonal diagonalizability are special cases of the general diagonalizability, which was discussed in Section 9.3.. Consequently, all the theorems remain valid, that were proved in that section. Taking into account also Theorem 12.7 we have the following theorem

12.10. Theorem *1. Let $A \in \mathbb{C}^{n\times n}$. The matrix A is unitarily diagonalizable if and only if its eigenvectors form an orthonormal basis basis (shortly: there exists an E.O.N.B. in $\mathbb{C}^n$). In this case the columns of the unitarily diagonalizing matrix U are the terms of the E.O.N.B.*

2. Let $A \in \mathbb{R}^{n\times n}$. The matrix A is orthogonally diagonalizable if and only if there exists an orthonormal basis in $\mathbb{R}^n$ consisting of the eigenvectors of A (shortly: there exists a E.O.N.B. in $\mathbb{R}^n$). In this case the columns of the orthogonally diagonalizing matrix Q are the terms of the E.O.N.B.

12.5. Spectral Theorems

The spectral theorems are theorems about the unitary or orthogonal diagonalizability of some special matrices. Here in the basic linear algebra we will state spectral theorems only about the self-adjoint and about the real symmetric matrices.

12.11. Theorem *[Spectral Theorem of the self-adjoint matrices] Let $A \in \mathbb{C}^{n\times n}$ be a self-adjoint matrix. Then there exists E.O.N.B. in $\mathbb{C}^n$. Or equivalently: A is unitarily diagonalizable.*

Proof. We will prove the theorem only in the case when A has n distinct eigenvalues in $\mathbb{C}$. In this case A has n independent eigenvectors, that is it has an Eigenvector Basis (E.B.) in $\mathbb{C}^n$ (see Remark 9.16). Using the orthogonality of the eigenspaces (see Theorem 12.5, that the terms of this E.B. form an orthogonal system. After normalization we have an E.O.N.B.

The proof of the general case is not contained in our basic linear algebra studies.

□

12.12. Remark. The steps of the unitary diagonalization of a self-adjoint matrix are as follows:

1. Determine the eigenvalues of A.

   Note that the eigenvalues are real.

2. Determine the eigenvectors to each eigenvalue.

   We have to obtain n linearly independent eigenvectors, which form an E.B. in $\mathbb{C}^n$.

3. Apply the Gram-Schmidt Process for the eigenvectors in each eigenspace which has at least 2 dimension.

   Thus the terms of the E.B. become orthogonal.

4. Apply the normalization (divide each vector by its norm).

   Thus we obtain an E.O.N.B.

5. Put the vectors of the E.O.N.B. into the matrix U as columns.

   This matrix U will unitarily diagonalize the matrix A.

The case of real symmetric matrices can be easily reduced back to the case of self-adjoint matrices.

12.13. Theorem *[Spectral Theorem of real symmetric matrices] Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then there exists E.O.N.B. in $\mathbb{R}^n$. Or equivalently: A is orthogonally diagonalizable.*

Proof. The matrix A can be regarded as a self-adjoint matrix $A \in \mathbb{C}^{n\times n}$ with real entries. Applying the Spectral Theorem of the self-adjoint matrices we obtain that A has an E.O.N.B. in $\mathbb{C}^n$.

But following the process of the determination the E.O.N.B. (see Remark 12.12), we can establish that each step takes place in real arithmetic. Consequently, the constructed E.O.N.B. is in $\mathbb{R}^n$. $\square$

12.6. Quadratic Forms

12.14. Definition Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The function

$$Q : \mathbb{R}^n \to \mathbb{R}, \qquad Q(x) := \langle Ax,\, x\rangle = \underline{x}^T A \underline{x} = \sum_{i,j=1}^{n} a_{ij} \cdot x_i \cdot x_j$$

is called quadratic form associated with A. A is called the matrix of Q.

12.15. Remarks.

1. It can be proved that the connection between the $n \times n$ symmetric matrices and the quadratic forms is one-to-one.

2. The quadratic forms are exactly the homogeneous n-variable polynomials. This means that they are polynomials whose each term is of second degree.

12.16. Theorem *[Principal Axis Theorem] Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form associated with the symmetric matrix $A \in \mathbb{R}^{n \times n}$. Let $u_1, \dots, u_n$ be an E.O.N.B. of A in $\mathbb{R}^n$. Let $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ be the corresponding eigenvalues. Then*

$$Q(x) = \sum_{i=1}^{n} \lambda_i \xi_i^2 \qquad (x \in \mathbb{R}^n)\,, \tag{12.1}$$

where the numbers $\xi_1, \dots \xi_n$ are the coordinates of x relative to the basis $u_1, \dots, u_n$.

Proof. Since $x = \sum\limits_{i=1}^{n} \xi_i u_i$, then we have

$$\begin{aligned}
Q(x) = \langle Ax,\, x\rangle = \langle A \cdot \sum_{i=1}^{n} \xi_i u_i,\, \sum_{j=1}^{n} \xi_j u_j\rangle = \langle \sum_{i=1}^{n} \xi_i A u_i,\, \sum_{j=1}^{n} \xi_j u_j\rangle = \\
= \langle \sum_{i=1}^{n} \xi_i \lambda_i u_i,\, \sum_{j=1}^{n} \xi_j u_j\rangle = \sum_{i,j=1}^{n} \lambda_i \xi_i \xi_j \cdot \langle u_i,\, u_j\rangle = \\
= \sum_{i=1}^{n} \lambda_i \xi_i \xi_i \cdot 1 = \sum_{i=1}^{n} \lambda_i \xi_i^2 \qquad (x \in \mathbb{R}^n).
\end{aligned}$$

□

12.17. Remark. Since the coordinates can be computed by scalar product

$$\xi_i = \langle x,\, u_i\rangle \qquad (i = 1, \dots, n)\,,$$

then the Principal Axis Theorem can be written as

$$Q(x) = \sum_{i=1}^{n} \lambda_i (\langle x,\, u_i\rangle)^2 \qquad (x \in \mathbb{R}^n)\,.$$

12.18. Theorem *Using the notations of the Principal Axis Theorem suppose that the eigenvalues of A are denoted in nondecreasing order:*

$$\lambda_1 \le \lambda_2 \le \dots \le \lambda_n\,.$$

Then

$$\forall\, x \in \mathbb{R}^n : \quad \lambda_1 \|\, x\|^2 \le Q(x) \le \lambda_n \|\, x\|^2\,.$$

Especially if $\|\, x\| = 1$ (x is a point of the unit sphere of $\mathbb{R}^n$) we have:

$$\lambda_1 \le Q(x) \le \lambda_n\,.$$

Proof. Applying the Principal Axis Theorem and Corollary 11.8 we have for any $x \in \mathbb{R}^n$ the following:

$$Q(x) = \sum_{i=1}^{n} \lambda_i \xi_i^2 \leq \sum_{i=1}^{n} \lambda_n \xi_i^2 = \lambda_n \cdot \sum_{i=1}^{n} \xi_i^2 = \lambda_n \| x \|^2 ,$$

and

$$Q(x) = \sum_{i=1}^{n} \lambda_i \xi_i^2 \geq \sum_{i=1}^{n} \lambda_1 \xi_i^2 = \lambda_1 \cdot \sum_{i=1}^{n} \xi_i^2 = \lambda_1 \| x \|^2 .$$

□

12.19. Corollary. Using the Rayleigh-Ritz-quotient (see Theorem 12.3) we have

$$\lambda_1 = \frac{\langle Au_1, u_1 \rangle}{\|u_1\|^2} = \frac{Q(u_1)}{1} = Q(u_1) \quad \text{and similarly} \quad \lambda_n = Q(u_n) .$$

Thus we have proved

$$\min_{\|x\|=1} Q(x) = \lambda_1 \qquad \text{and} \qquad \max_{\|x\|=1} Q(x) = \lambda_n .$$

Let us classify the quadratic forms by the signs of their values.

12.20. Definition Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form associated with the symmetric matrix $A \in \mathbb{R}^{n \times n}$. We say that Q (or that A) is

(a) positive definite if $\forall\, x \in \mathbb{R}^n \setminus \{0\} : \quad Q(x) > 0$,

(b) negative definite if $\forall\, x \in \mathbb{R}^n \setminus \{0\} : \quad Q(x) < 0$,

(c) positive semidefinite if $\forall\, x \in \mathbb{R}^n : \quad Q(x) \geq 0$,

(d) negative semidefinite if $\forall\, x \in \mathbb{R}^n : \quad Q(x) \leq 0$,

(e) indefinite, if $\exists\, x, y \in \mathbb{R}^n : \quad Q(x) > 0,\ Q(y) < 0$.

Using the Principal Axis Theorem the above classification can be made using the signs of the eigenvalues of A. It will be stated without proof in the following theorem.

12.21. Theorem *Using the previous notations the quadratic form Q (and its matrix A) is*

(a) positive definite if and only if $\forall\, \lambda \in \mathrm{Sp}\,(A) :\ \lambda > 0$.

(b) negative definite if and only if $\forall\, \lambda \in \mathrm{Sp}\,(A) :\ \lambda < 0$.

(c) *positive semidefinite if and only if* $\forall\, \lambda \in \mathrm{Sp}\,(A):\ \ \lambda \geq 0\,.$

(d) *negative semidefinite if and only if* $\forall\, \lambda \in \mathrm{Sp}\,(A):\ \ \lambda \leq 0\,.$

(e) *indefinite if and only if* $\exists\, \lambda, \mu \in \mathrm{Sp}\,(A):\ \ \lambda > 0,\ \ \mu < 0\,.$

In the following theorem we will classify the two-variable quadratic forms.

12.22. Theorem *[classification of the two-variable quadratic forms]*
Let

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

and $Q : \mathbb{R}^n \to \mathbb{R}$ *be the quadratic form given by* A*. That is*

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 \qquad (x = (x_1, x_2) \in \mathbb{R}^2)\,.$$

Then Q *is*

- *positive definite if* $\det A = ac - b^2 > 0$ *and* $a > 0$,
- *negative definite if* $\det A = ac - b^2 > 0$ *and* $a < 0$.
  (The case $\det A = ac - b^2 > 0$ *and* $a = 0$ *is impossible.)*
- *indefinite if* $\det A = ac - b^2 < 0$.
- *semidefinite but not definite if* $\det A = ac - b^2 = 0$.

The semidefinite case is in detail as follows. Suppose that $\det A = ac - b^2 = 0$. *Then* Q *is*

- *positive semidefinite but not positive definite if* $a > 0$ *or if* $a = 0, c > 0$,
- *negative semidefinite but not negative definite if* $a < 0$ *or if* $a = 0, c < 0$,
- *the identical* 0*-function if* $a = c = 0$.

Proof. The proof is based on the following elementary identities:

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 = \begin{cases} \dfrac{(ax_1 + bx_2)^2 + (ac - b^2)x_2^2}{a} & \text{if } a \neq 0, \\[2ex] \dfrac{(bx_1 + cx_2)^2 + (ac - b^2)x_1^2}{c} & \text{if } c \neq 0, \\[2ex] 2bx_1x_2 & \text{if } a = c = 0. \end{cases}$$

Using these identities one can easily discuss the sign of the values of Q. □

12.7. Control Questions

1. State and prove the connection between $\langle Ax, y\rangle$ and $\langle x, A^*y\rangle$
2. State and prove the theorem about the Rayleigh-Ritz quotient
3. Define the concept of self-adjoint (Hermitian) and of real symmetric matrix
4. State and prove the 4 important properties of the self-adjoint matrices
5. Define the concept of unitary and of orthogonal matrix
6. State and prove the theorem about the equivalencies with „A is unitary”
7. State and prove the 5 important properties of the unitary matrices
8. Define the concept of unitary and of orthogonal diagonalizability of a matrix
9. State and prove the necessary and sufficient condition of unitary and of the orthogonal diagonalizability
10. State the Spectral Theorem of a self-adjoint matrix. Prove it in the case when the $n \times n$ self-adjoint matrix has n different eigenvalues
11. State and prove the Spectral Theorem of a real symmetric matrix (reducing back to self-adjoint matrices).
12. Define the quadratic form
13. State and prove the Principal Axis Theorem
14. What are the minimal and the maximal values of a quadratic form on the unit sphere? At which vectors it takes these extreme values?
15. Define the definiteness of a quadratic form
16. State and prove the theorem about the definiteness of a quadratic form using the eigenvalues
17. State and prove the theorem about the classification of the two-variables quadratic forms

12.8. Homework

1. Prove that the diagonal entries of a self-adjoint matrix all are real numbers.
2. Diagonalize orthogonally the following real symmetric matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} .$$

3. Classify the quadratic forms according to the following matrices by their sign (definiteness):

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} , \qquad B = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} , \qquad C = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} .$$