Linear Algebra lecture schemes
(with Homeworks)[1]

Csörgő István

November, 2014

[1]A jegyzet az ELTE Informatikai Kar 2014. évi Jegyzetpályázatának támogatásával készült

Contents

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:

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

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}, $\dots$ 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.

□

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. □

1.4. Homeworks

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 listen to the following requirements:

1. If you regard the blocks as matrix elements the operations must be defined between the so obtained „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 so obtained submatrix will be called minor matrix.

2.1. Definition (Minor Matrix) Let $A \in \mathbb{K}^{n\times n}$ and $(i,j) \in \{1, \ldots n\}\times\{1, \ldots n\}$ 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:

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

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

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

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

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

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

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:

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

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

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

So we have proved that $AC = I$. □

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.

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

□

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.

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

□

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

1. Suppose that $\det A = 0$. Then A has never left-hand inverse nor right-hand inverse (it is not invertible from the left and it is not invertible 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, so 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. A matrix $A \in \mathbb{K}^{n\times n}$ is regular if and only if $\det A \neq 0$.

5. A matrix $A \in \mathbb{K}^{n\times n}$ 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. Homeworks

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 $k = 1, \dots n$. $\square$

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 that will be studied in the subject „Numerical Methods".

3.2. Homeworks

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

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

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

We can define other operations in the vector space V:

4.3. Definition

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.

4.4. 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. Homeworks

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 $k \in \mathbb{N}$, $x_1, \dots, x_k \in W$, $\lambda_1, \dots \lambda_k \in \mathbb{K}$:

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

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

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

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

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

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

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

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

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

Proof.

1. Let $a = \sum_{i=1}^{k} \lambda_i x_i \in W^*$ and $b = \sum_{i=1}^{k} \mu_i y_i \in W^*$. 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 $i \in \{1, \dots, k\}$:

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

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 $\operatorname{span}(x_1, x_2, \ldots, x_k)$. 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 $\operatorname{span}(x_1, x_2, \ldots, x_k)$ 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

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

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

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

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

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \cdot 1 + x_2 \cdot 0 + \dots + x_n \cdot 0 \\ x_1 \cdot 0 + x_2 \cdot 1 + \dots + x_n \cdot 0 \\ \vdots \\ x_1 \cdot 0 + x_2 \cdot 0 + \dots + x_n \cdot 1 \end{pmatrix} =$$

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

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 $f \in \mathcal{P}_n, \quad f(x) = a_0 + a_1 x + \dots + a_n x^n \quad (x \in \mathbb{K})$ 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 $\dim \mathcal{P} = \infty$.

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

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

Then the polynomial $g(x) := x^{k+1} \quad (x \in \mathbb{K})$ cannot be expressed as linear combination of $f_1, \ldots, f_m$ 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. Homeworks

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

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

   Here x is regarded as an $n \times 1$ matrix. The subspace $\mathrm{null}(A)$ 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 that the vector $x = (2, 4, 0)$ is in the subspace $\mathrm{span}\,(a, b)$ 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 $\mathrm{null}(A)$.

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

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

   which implies $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$.

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.

6.2. Basis

6.4. 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.

6.5. 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.

6.6. 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. To this proof we need the following lemma:

6.7. Lemma Let $x_1, \ldots, x_k \in V$ be a linearly dependent system. Then

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

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

$$\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k\} \subseteq \{x_1, \ldots, x_k\}\,.$$

To prove the relation „$\supseteq$" observe first that

$$\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k\} \subseteq \operatorname{span}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)\,.$$

It remains the proof of

$$x_i \in \operatorname{span}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)\,.$$

Indeed, by the dependence of the system there exist the numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ 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 $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$, so it is really in the subspace $\operatorname{span}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$.

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

$\square$

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

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

Proof. Let $x_1, \dots, x_k$ be a finite generator system of V. If this system is linearly independent then it is basis. If it is dependent then – by the lemma – 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. $\square$

6.10. 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. Using the theorem it can be proved that every linearly independent system can be completed into 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.

6.11. 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 $j \in \{1, \dots, m\}$. Then there exist the coefficients $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ such that they are not all zero and $\lambda_1 y_j + \lambda_2 x_2 + \dots + \lambda_k x_k = 0$. If it were be $\lambda_1 = 0$ then it were be $\lambda_2 x_2 + \dots + \lambda_k x_k = 0$ with coefficients that are not all zero. This were be in contradiction with the linear independence of the subsystem $x_2, \dots, x_k$. So $\lambda_1 \neq 0$.

Since $\lambda_1 \neq 0$, y_j can be expressed from the dependence equation:

$$y_j = -\frac{\lambda_2}{\lambda_1} x_2 + \dots - \frac{\lambda_k}{\lambda_1} x_k\,.$$

This expression implies $y_j \in \operatorname{span}(x_2, \dots, x_k)$ $(j = 1, \dots, m)$. 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 $x_1, \dots, x_k$. □

6.12. Corollary. The number of vectors in a linearly independent system is not greater than the number of vectors in a generator system.

To prove this 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}, so we 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}, so 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). So we have the conclusion that among the vectors $y_1, \dots, y_m$ k piece are different. So really $k \le m$.

6.13. 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.

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

On the other hand

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

Consequently $k = m$. □

6.14. 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.

6.15. 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. $\dim \mathcal{P}_n = n + 1 \quad (n \in \mathbb{N} \cup \{0\})$.

6.16. Theorem *Let $1 \leq \dim(V) = n < \infty$. Then*

1. *If $x_1, \dots, x_k \in V$ and $k \geq n + 1$ then $x_1, \dots, x_k$ is linearly dependent. In other words: the number of vectors in a linearly independent system is at most the dimension of the space.*

2. *If $k \leq n - 1$ then $x_1, \dots, x_k$ is not generator system in V (it does not span V). In other words: the number of vectors in a generator system in V is at least the dimension of the space.*

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

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

Proof.

1. Suppose indirectly that $x_1, \dots, x_k$ is linearly independent and let $e_1, \dots, e_n$ be a basis in V. Then it is generator system, so by the corollary of the Exchange Theorem:
$$n + 1 \leq k \leq n\,,$$
which is an obvious contradiction.

   The proofs of the remainder statements are left as exercises.

□

6.4. Homeworks

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 $e:\, e_1, \ldots e_n$ be a basis in V. Then*

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

Proof. The existence of the numbers ξ_i is obvious because $e_1, \ldots e_n$ 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 $e_1, \ldots e_n$ – follows that $\xi_i - \eta_i = 0$ that is $\xi_i = \eta_i \quad (i = 1, \ldots, n)$. □

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

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

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

7.3. Remark. If $V = \mathbb{K}^n$ and $e_1, \ldots e_n$ 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 $x \in \mathbb{K}^n$ coordinates.

7.4. Theorem *Let $e: e_1, \ldots e_n$ be an ordered basis in V. Then for every $x, y \in V$ 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. □

7.5. 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:

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

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

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

□

7.6. 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. Homeworks

1. It is given the following basis 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. $0 \leq \operatorname{rank}(x_1, \ldots, x_k) \leq k$.
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:

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

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. 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:

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

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 $\mathrm{col}(A)$.

8.5. Remark. Obviously

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

8.6. Theorem $\dim \mathrm{row}(A) = \dim \mathrm{col}(A)$.

Proof. On the lecture. □

8.7. 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: $\mathrm{rank}\,(A)$. So

$$\mathrm{rank}\,(A) := \dim \mathrm{row}(A) = \dim \mathrm{col}(A)\,.$$

8.8. Remarks.

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

8.3. System of Linear Equations

8.9. 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) 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 from $\mathbb{K}$ of the unknowns $x_1, \dots, x_n$ such that all the equations will be true. These systems of the unknowns are called the solutions of the linear system. The linear equation system is called consistent if it has solution. It is called inconsistent if it has no solution.

Let us denote by $a_1, \ldots, a_n$ the column vectors of A that and by b the vector formed from the right-side constants:

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

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

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

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

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

and the unknown vector $x := (x_1, \ldots, x_n) \in \mathbb{K}^n$. 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.

8.10. Remark. It is easy to observe that

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

So 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 $\operatorname{col}(A)$ is the possible greatest subspace that is $\operatorname{col}(A) = \mathbb{K}^m$ then the system is consistent.

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

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

Naturally if the system is inconsistent then $\mathcal{S} = \emptyset$.

8.11. Definition Let $Ax = b$ be a system of linear equations. Then the system $Ax = 0$ is called the homogeneous system associated with $Ax = b$. Denote by $\mathcal{S}_h$ the set of solution vectors of the homogeneous system that is:

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

Remark that the homogeneous system is always consistent because the zero vector is its solution. So $\mathcal{S}_h \neq \emptyset$. Moreover $\mathcal{S}_h$ is a subspace in $\mathbb{K}^n$.

About the structures of the set of solutions we tell the following theorem without proof:

8.12. Theorem *Let $Ax = b$ be a consistent linear equation system and let $r = \operatorname{rank} A$. Then*

1. *If $r = n$ then the system has a unique solution.*

2. *If $r < n$ then the system has infinitely many solutions. In this case the solution set $\mathcal{S}_h$ of the associated homogeneous system is an $n-r$ dimensional subspace of $\mathbb{K}^n$. If $v_1, \ldots, v_{n-r}$ denotes a basis of $\mathcal{S}_h$ and x_0 is a particular solution of $Ax = b$ then the general solution of $Ax = b$ is:*

$$x = x_0 + \sum_{j=1}^{n-r} \lambda_j v_j \qquad (\lambda_j \in \mathbb{K}) .$$

where the constants $\lambda_j \in \mathbb{K}$ are arbitrary. So the solution set $\mathcal{S}$ is a translation of the $n - r$ dimensional subspace $\mathcal{S}_h$.

We will solve only low-measure systems and will use the „Substitution Method" studied in the secondary school. In the process of solving the system we will see the validity of the above theory.

For higher-measure systems the decision of consistence and the discussion of all solutions requires an algorithmic method for example: Elementary Basis Transformation Method, Gaussian Elimination, Gauss-Jordan Elimination. The algorithmic methods will be studied in the subject Numerical Methods.

8.4. Linear Equation Systems with Square Matrices

Let us study the linear equation system with square matrix:

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

Denote by r the rank of A. In the following discussion plays important role the fact that A is invertible if and only if all the linear systems $Ax = e_i$ are consistent $(i = 1, \ldots, n)$.

We distinguish between the two basic cases as follows.

Case 1.: $r = n$.

In this case – because of r equals the number of rows – the system is consistent. On the other hand – because of r equals the number of columns – the solution is unique. So in the case $\operatorname{rank} A = n$ the square system has a unique solution independently of b.

If we apply this result for $b = e_i$ where e_i denotes the ith standard unit vector we obtain that in the case $\operatorname{rank} A = n$ the matrix A is invertible (it is regular) consequently $\det A \neq 0$.

Case 2.: $r < n$.

In this case – since r is less than the number of rows – the system may be consistent (if $b \in \operatorname{col}(A)$) or inconsistent (if $b \notin \operatorname{col}(A)$). If the system is consistent then – since r is less than the number of columns – the system has infinitely many solutions.

Since $\operatorname{col}(A) \neq \mathbb{K}^m$ so there exists a standard unit vector e_i such that the system $Ax = e_i$ is inconsistent. Consequently in the case $\operatorname{rank} A < n$ the matrix A is not invertible (it is singular) and by this reason $\det A = 0$.

Let us collect the our results in the following theorem:

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

1. $\operatorname{rank} A = n \quad \Leftrightarrow \quad \det A \neq 0 \quad \Leftrightarrow \quad$ *A is invertible (regular);*
2. $\operatorname{rank} A < n \quad \Leftrightarrow \quad \det A = 0 \quad \Leftrightarrow \quad$ *A is not invertible (singular).*

8.5. Homeworks

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):

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

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

$$g(\lambda) = \dim W_\lambda = \dim \mathcal{S}_h = n - \mathrm{rang}\,(A-\lambda I)\,.$$

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 every eigenvalue stands in the diagonal as many 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. Homeworks

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 $V \times V \ni (x, y) \mapsto \langle x,\, y\rangle$ (inner product) be a mapping (operation). 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 (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. The operation defined above is the inner product (or dot product or scalar product).

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
$$\langle x,\, y\rangle := \sum_{k=1}^{n} x_k \overline{y_k}\,.$$
This is the standard inner product in $\mathbb{K}^n$. Naturally in the case $\mathbb{K} = \mathbb{R}$ there is no conjugation:
$$\langle x,\, y\rangle := \sum_{k=1}^{n} x_k y_k\,.$$

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 $\mathbb{K} = \mathbb{C}$: $\langle f,\, g\rangle := \int\limits_a^b f(x)\overline{g(x)}\; dx$.

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

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. $\langle \sum\limits_{i=1}^{n} \lambda_i x_i,\, \sum\limits_{j=1}^{m} \mu_j y_j\rangle = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} \lambda_i\overline{\mu_j}\langle x_i,\, y_j\rangle$ *Naturally in the real case $\mathbb{K} = \mathbb{R}$ there is no conjugation:* $\langle \sum\limits_{i=1}^{n} \lambda_i x_i,\, \sum\limits_{j=1}^{m} \mu_j y_j\rangle = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} \lambda_i\mu_j\langle x_i,\, y_j\rangle$

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*

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

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 $\lambda \in \mathbb{R}$:

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

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 $x + \lambda y \neq 0$ 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. □

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

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

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

10.3. Norm

In this section will be extended the concept of the length of vectors (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

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

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

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

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

□

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 can occur several times in an orthogonal system but any other vector can 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* $x_1, \ldots, x_n \in V \setminus \{0\}$ *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 $j = 1, \ldots, n$:

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

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

10.20. Theorem *[Pythagoras] If* $x_1, \ldots, x_n \in V$ *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 $i \neq j$.) $\square$

10.6. Homeworks

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 $u_1, \ldots, u_n \in V \setminus \{0\}$ be an orthogonal system, $W := \operatorname{span}(u_1, \ldots, u_n)$. (It is important to remark that in this case $u_1, \ldots, u_n$ 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 $u_1, \ldots, u_n$:

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

This expression equals 0 if and only if

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

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

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 $W^{\perp}$:

$$Q(x) = \mathrm{proj}_{W^{\perp}} x\,.$$

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

$$\forall\, k =\in \{1, 2, , \ldots, n\}: \quad \mathrm{span}\,(b_1,\, \ldots,\, b_k) = \mathrm{span}\,(u_1,\, \ldots,\, u_k)\,.$$

Especially (for $k = n$) the generated subspaces by the two systems are the same.

The Gram-Schmidt process sounds 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 span (u_1)
- u_3 is the orthogonal component of b_3 with respect to the subspace span $(u_1,\, u_2)$

  $\vdots$

- u_n is the orthogonal component of b_n with respect to the subspace 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 : $e_1, \ldots, e_n$ 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

$$\forall\, x, y \in V : \quad \langle x,\, y\rangle = \langle [x]_e,\, [y]_e\rangle = \sum_{i=1}^{n} \xi_i \overline{\eta_j}\,.$$

Here $[x]_e = (\xi_1, \ldots, \xi_n)$ *and* $[y]_e = (\eta_1, \ldots, \eta_n)$ *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

$$\langle x,\, y\rangle = \langle \sum_{i=1}^{n} \xi_i e_i,\, \sum_{j=1}^{n} \eta_j e_j\rangle = \sum_{i,j=1}^{n} \xi_i \overline{\eta_j} \langle e_i,\, e_j\rangle = \sum_{i=1}^{n} \xi_i \overline{\eta_i} =$$

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

□

The following theorem gives us the coordinates.

11.8. Theorem *The ith 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

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

Proof. Apply the previous theorem with $y = e_i$ $(i = 1, \dots, n)$. Then we obtain

$$\langle x,\, e_i\rangle = \langle [x]_e,\, [e_i]_e\rangle = \langle (\xi_1, \dots, \xi_n),\, (0, \dots, 1, \dots, 0)\rangle = \xi_i \qquad (i = 1, \dots, n)\,.$$

□

11.9. Remark. One can simply consider – using the normalization process – that the ith coordinate of a vector $x \in V$ relative to an orthogonal basis $u_1, \dots, u_n$ is

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

Consequently

$$x = \sum_{i=1}^{n} \frac{\langle x,\, u_i\rangle}{\langle u_i,\, u_i\rangle} \cdot u_i\,.$$

11.4. Homeworks

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.