Lecture 1.

A **binary operation** ϕ on a set A is a function from $A \times A$ to A, i.e., $\phi : A \times A \to A$.

$\phi((x, y))$ is also denoted by $x\phi y$. Often we use a symbol for ϕ: $+, \cdot, -, \star$.

A **field** $\mathbb{F}$ is a set with two operation, called addition and multiplication, which are denoted by $+$ and $\cdot$ (often omitted), respectively, and which satisfy the following properties:

1. Both operations are **commutative**: $a + b = b + a$ and $ab = ba$ for all $a, b \in \mathbb{F}$
2. Both operations are **associative**: $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$ for all $a, b, c \in \mathbb{F}$
3. There exists a unique **identity** element for each operation, denoted by 0 and 1, i.e., $0 + a = a + 0 = a$ and $1a = a1 = a$ for all $a \in \mathbb{F}$
4. For every $a \in \mathbb{F}$, there exists a unique $b \in \mathbb{F}$ such that $a + b = b + a = 0$. This element b is called the **additive inverse** of a in $\mathbb{F}$, and is denoted by $-a$.
5. For every $a \in \mathbb{F}^\times := \mathbb{F} \setminus \{0\}$, there exists a unique $b \in \mathbb{F}$ such that $ab = ba = 1$. This element b is called the **multiplicative inverse** of a in $\mathbb{F}$, and is denoted by a^{-1}.
6. Multiplication and addition are related by the **distributive** laws:

$$a(b + c) = ab + ac \quad \text{and} \quad (b + c)a = ba + ca.$$

The axiom system above is not the 'most economic' one. Check that it implies that $0a = a0 = 0$ for every $a \in \mathbb{F}$.

Examples of fields.

$\mathbb{Q}$ – the field of rational numbers;

$\mathbb{R}$ – the field of real numbers;

$\mathbb{C}$ – the field of complex numbers;

$\mathbb{F}_p$ – the finite field of p elements, p is prime. The field $\mathbb{F}_p$ is often denoted by $\mathbb{Z}/p\mathbb{Z}$ or $\mathbb{Z}_p$, and is thought as the set of p elements $\{0, 1, \ldots, p-1\}$ where addition and multiplication are done by modulo p.

We have $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, and the operations in the subfields are just restrictions of the corresponding operations in $\mathbb{C}$.

Suppose V is a set whose elements are called **vectors** and denoted by $\bar{v}$. Suppose there exists a binary operation on V, called addition and denoted by $+$, which is commutative, associative, there exists an identity element, denoted by $\bar{0}$ and called **zero** vector, and for every vector $\bar{v}$, there exists an additive inverse which we denote by $-\bar{v}$.

Suppose $\mathbb{F}$ is a field and there exists a function μ from $\mu : \mathbb{F} \times V \to V$ given by $(k, \bar{v}) \mapsto k\bar{v}$, which satisfies the following axioms:

1. $1v = v$ for all $\bar{v} \in V$
2. $k(\bar{v} + \bar{u}) = k\bar{v} + k\bar{u}$ for all $k \in \mathbb{F}$ and all $\bar{v}, \bar{u} \in V$
3. $(k + m)\bar{v} = k\bar{v} + m\bar{v}$ for all $k \in \mathbb{F}$ and all $\bar{v} \in V$
4. $k(m\bar{v}) = (km)\bar{v}$ for all $k, m \in \mathbb{F}$ and all $\bar{v} \in V$

An ordered triple $((V, +), \mathbb{F}, \mu)$ is called a vector space. In a simpler manner, we just say that V is a **vector space over the field** $\mathbb{F}$. The function μ is mentioned very rarely.

We say that μ defines a 'multiplication' of elements of $\mathbb{F}$ by vectors from V. Often in this context the elements of $\mathbb{F}$ are called **scalars**, and we say that μ defines a 'multiplication of vectors by scalars'. Note that this multiplication is *not* a binary operation on a set. Its result is an element of V, i.e., is always a vector. When we write $k(m\bar{v}) = (km)\bar{v}$, we mean that the scalars k and m are multiplied in $\mathbb{F}$, and the result of this multiplication km, which is a scalar, is 'multiplied' by a vector $\bar{v}$.

Examples of vector spaces.

• For a field $\mathbb{F}$, and positive integer n, let

$$V = \mathbb{F}^n := \{\bar{v} = (v_1, v_2, \ldots, v_n) : v_i \in \mathbb{F} \text{ for all } i = 1, 2, \ldots, n\}$$

The addition on $\mathbb{F}^n$ and the multiplication by scalars are defined as follows: for every $\bar{v}, \bar{u} \in V$, and every $k \in \mathbb{F}$,

$$\bar{v} + \bar{u} := (v_1 + u_1, \ldots, v_n + u_n) \quad \text{and} \quad k\bar{v} = k(v_1, \ldots, v_n) := (kv_1, \ldots, kv_n).$$

Then $V = \mathbb{F}^n$ is a vector space over $\mathbb{F}$, which can be (and has to be) checked easily. For $\mathbb{F} = \mathbb{R}$, we obtain the well familiar space $\mathbb{R}^n$.

If $\mathbb{F} = \mathbb{F}_p$, vector space $\mathbb{F}_p^n$ contains p^n vectors. E.g., $\mathbb{F}_2^3$ contains $2^3 = 8$ vectors.

• Let $V = C(0,1)$ be the set of all continuous real valued functions on $(0,1)$: $f : (0,1) \to \mathbb{R}$.

The addition on $C(0,1)$ and the scalar multiplication are defined as follows: for any $f, g \in C(0,1)$, and every $k \in \mathbb{R}$,

$$(f+g)(x) := f(x) + g(x) \quad \text{and} \quad (kf)(x) := kf(x).$$

Again, $C(0,1)$ is a vector space over $\mathbb{R}$, which has to be checked. Here the fact that $+$ is a binary operation on $C(0,1)$, or more precisely, that it is 'closed', which means $f + g \in C(0,1)$, is not a trivial matter. The same for kf, though it is a little easier. Similarly we can consider $C(R)$ or $C^1(0,1)$ – the vector space over $\mathbb{R}$ of all real valued differentiable functions from $(0,1)$ to $\mathbb{R}$ with continuous first derivative.

• $V = \mathbb{R}$ is a vector space over $\mathbb{Q}$. $V = \mathbb{C}$ is a vector space over $\mathbb{R}$. $V = \mathbb{C}$ is a vector space over $\mathbb{Q}$. In all these examples the addition in V is the usual addition in the field, and the multiplication of vectors by scalars is the usual multiplication of two numbers in V.

• $V = \mathbb{F}[x]$ – set of all polynomials of x with coefficients from a field $\mathbb{F}$. V is a vector space over $\mathbb{F}$ with respect to the usual addition of polynomials and the multiplication of polynomials by numbers (elements of $\mathbb{F}$).

• V is the set of all functions $f : \mathbb{R} \to \mathbb{R}$ which satisfy the differential equation

$$y'' - 5y' + 6y = 0.$$

V is a vector space over $\mathbb{R}$.

• V is the set of all sequences of real numbers (x_n), $n \geq 0$, defined by the recurrences:

$$x_0 = a,\ x_1 = b,\ a, b \in \mathbb{R}, \text{ and for all } n \geq 0,\ x_{n+2} = 2x_{n+1} - 3x_n.$$

V is a vector space over $\mathbb{R}$.

• $V = M_{m\times n}(\mathbb{F})$ – the set of all $m \times n$ matrices with entries from $\mathbb{F}$ with respect to usual addition of matrices and the multiplication of matrices by scalars.

• Let $A \in M_{m\times n}$, i.e., A is an $m \times n$ matrix over $\mathbb{R}$. Then the set V of all solutions of the homogeneous system of linear equations $A\bar{x} = \bar{0}$, i.e., the set of all vectors $\bar{x} \in \mathbb{R}^n$ such that $A\bar{x} = \bar{0}$, is a vector space over $\mathbb{R}$ with respect to usual addition of vectors and the scalar multiplication in $\mathbb{R}^n$. Note that in this example elements of $\mathbb{R}^n$ are thought of as the *column* vectors ($n \times 1$ matrices).

Proposition 1 *Let V be a vector space over a field $\mathbb{F}$. Then*

(i) $\bar{0}$ *is unique.*

(ii) for each $\bar{a} \in V$*, the inverse of* $\bar{a}$ *is unique.*

(iii) $0\bar{a} = \bar{0}$ *for every* $\bar{a} \in V$*.*

(iv) $(-1)\bar{a} = -\bar{a}$ *for every* $\bar{a} \in V$*.*

(v) $-(-\bar{a}) = \bar{a}$ *for every* $\bar{a} \in V$*.*

(vi) Cancellation Law: $\bar{a} + \bar{b} = \bar{a} + \bar{c}$ *if and only if* $\bar{b} = \bar{c}$*.*

Proof. (i) Indeed, if $\bar{0}'$ is a possible another identity element, then $\bar{0} + \bar{0}' = \bar{0}$ as $\bar{0}'$ is an identity, and $\bar{0} + \bar{0}' = \bar{0}'$ as $\bar{0}$ is an identity. So $\bar{0} = \bar{0}'$. This justifies the notation $\bar{0}$.

(ii) Indeed, let $\bar{b}$, $\bar{b}'$ be inverses of a. Then consider the element $(\bar{b} + \bar{a}) + \bar{b}'$. Since $\bar{b} + \bar{a} = \bar{0}$, then $(\bar{b} + \bar{a}) + \bar{b}' = \bar{0} + \bar{b}' = \bar{b}'$. Similarly, consider $\bar{b} + (\bar{a} + \bar{b}')$. Since $\bar{a} + \bar{b}' = \bar{0}$, then $\bar{b} + (\bar{a} + \bar{b}') = \bar{b} + \bar{0} = \bar{b}$. Due to the associativity, $(\bar{b} + \bar{a}) + \bar{b}' = \bar{b} + (\bar{a} + \bar{b}')$. So $\bar{b}' = \bar{b}$. This justifies the notation $-\bar{a}$.

(iii)

(iv)

(v)

(vi)

From now on:

We will NOT denote vectors by bars.

When we write kv, we will mean that $k \in \mathbb{F}$ and $v \in V$.

If we do not say otherwise, V will denote a vector space over an arbitrary field $\mathbb{F}$.

Lecture 2.

Let V be a vector space and $k_1, \ldots, k_m \in \mathbb{F}$ and $v_1, \ldots, v_m \in V$. Then the vector

$$k_1 v_1 + \ldots + k_m v_m$$

is called a **linear combination** of $v_1, \ldots, v_m$. At this time we do not define infinite linear combinations.

For a subset A of V the set of all (finite) linear combinations of vectors from A is called the **span** of A and is denoted by $Span(A)$ or $\langle A \rangle$. Clearly, $A \subseteq Span(A)$.

Sometimes it is convenient to use $\sum_{a \in A} k_a a$ as a notation for a general linear combination of finitely many elements of A. In this notation we always assume that only finitely many coefficients k_a are nonzero

A subset W of a vector space V over a field $\mathbb{F}$ is called a **subspace** or a **linear subspace** of V if it is a vector space over the the operation on V restricted to W and the multiplication of elements of V by elements from $\mathbb{F}$ restricted to W.

In order to check that a subset W is a subspace in V it is sufficient to check that it is "closed" with respect to the addition of V and with respect to the multiplication by scalars. All other axioms will be inherited from the ones in V. The existence of the zero vector in W and the additive inverses follows from the fact that in V $0a = 0\,(= \bar{0})$, $(-1)a = -a$ and that W is closed with respect to multiplication of vectors by scalars.

QUESTION: Given a subset A in a vector space V, what is the smallest subspace of V with respect to inclusion which is a superset of A?

It turns out that it is $Span(A)$! It is proved in the following theorem among other properties of subspaces and spans.

Theorem 2 *Let V be a vector space.*

1. *For any subset A of V, $Span(A)$ is a subspace of V*
2. *A subset W of V is a subspace if and only if $Span(W) = W$.*
3. *For every subset A of V, $Span(Span(A)) = Span(A)$.*
4. *Intersection of any collection of subspaces of V is a subspace of V.*

5. *For every subset W of V,*

$$Span(W) = \bigcap_{W \subseteq U} U,$$

where the intersection is taken for all subspaces U of V for which W is a subset.

Proof. Please complete. ∎

Comments. The first statement of the theorem describes subspaces "from within". It can be used to prove that a subset is a subspace.

It is clear that for every $A \subset V$, $A \subseteq Span(A)$. The second statement means that taking a span of a subset of V more than once does not produce a greater subspace.

The last statement describes a subspace "from outside".

Let V be a vector space. We say that $A \subseteq V$ **generates** V (or **spans** V) if $Span(A) = V$.

For example, the set $A = \{(3,1),(-1,2))\} \subset \mathbb{R}^2$ generates $\mathbb{R}^2$.

Problems.

When you do these problems, please do not use any notions or facts of linear algebra which we have not discussed in this course. You must prove all your answers.

1. Prove that if V is a vector space and $A \subseteq B \subseteq V$, then $Span(A) \subseteq Span(B)$.
2. Prove or disprove:
   (a) $A = \{(3,1,2),(-1,3,1),(2,4,3)\} \subset \mathbb{R}^3$ generates $\mathbb{R}^3$.
   (b) $A = \{(3,1,2),(-1,3,1),(2,4,-3)\} \subset \mathbb{R}^3$ generates $\mathbb{R}^3$.
   (c) $A = \{(t,t^2,t^3) : t \in \mathbb{R}\} \subset \mathbb{R}^3$ generates $\mathbb{R}^3$. This subset is known as **twisted cubic**.
   (d) $A = \{(3,1),(-1,2)\} \subset \mathbb{F}_5^2$ generates $\mathbb{F}_5^2$.
3. Is $(2,-1,0)$ in $Span(\{(1,1,1),(2,3,-1)\})$?
4. Let $v \in F_3^4$ and v is not a zero vector. How many vectors does $Span(\{v\})$ have?
5. Let P_3 be the set of all polynomials of x with real coefficients of degree at most 3. Show that P_3 is a vector space over $\mathbb{R}$. Show that the set $\{1,x,x^2,x^3\}$ spans P_3, and the set $\{1,x-1,(x-1)^2,(x-1)^3\}$ spans P_3.

6. Does the set of functions $A = \{1, e^x, e^{2x}, \dots e^{nx}, \dots\} \subset C(\mathbb{R})$ span $C(\mathbb{R})$? Here $C(\mathbb{R})$ is the vector space of all continuous functions $f : \mathbb{R} \to \mathbb{R}$. Will the answer change if we consider A as a subset of the vector space of all differentiable functions from $\mathbb{R}$ to $\mathbb{R}$?

7. Show that if U and W are subspaces of a vector space V, then $U \cup W$ need not be a subspace of V. However, $U \cup W$ is a subspace of V if and only if $U \subseteq W$ or $W \subseteq U$.

8. Let V be the set of all sequences of real numbers (x_n), $n \geq 0$, defined by the recurrences:

$$x_0 = a,\ x_1 = b, \text{ and for all } \ n \geq 0,\ \ x_{n+2} = -2x_{n+1} + 3x_n.$$

   Every such sequence is completely defined by a choice of a and b. Show that V is a vector space over $\mathbb{R}$ and that it can be spanned by a set of two vectors.

9. Let U and W be subspaces of a vector space V, and let

$$U + W = \{u + w : u \in U \text{ and } w \in W\}.$$

   Prove that $U + W$ is a subspace of V, and that U and W are subspaces of $U + W$. Prove that if a subspace X of V contains U and W as subspaces (equivalently as subsets), than $U + W \subseteq X$. Hence $U + W$ is the smallest (with respect to inclusion) subspace of V which contains both U and W.

10. Let X, Y and Z be subspaces of a vector space V and $X + Y = X + Z$. Does it imply that $Y = Z$? Prove your answer.

11. Consider $\mathbb{R}^\infty$ – the vector space of all infinite sequences of real numbers, with addition of vectors and multiplication of vectors by scalars defined similarly to the ones in $\mathbb{R}^n$. Consider a subset $l^2(\mathbb{R})$ of all those sequences (x_n) such that $\sum_{i=1}^{\infty} x_i^2$ converges. Does $l^2(\mathbb{R})$ span R^∞?

Lecture 3.

Usually, and in this course, a set can have only *distinct* elements. Otherwise we would refer to it as a multiset.

A set X is called **finite** if there exists an integer $n \geq 0$ and a bijection from X to $\{1, \ldots, n\}$. A set X is called **infinite** if there exists a bijection from a subset Y of X to $\mathbb{N} = \{1, 2, \ldots\}$.

Let $\{v_1, \ldots, v_m\} \subseteq V$ be a set of m vectors. We say that vectors $v_1, \ldots, v_m$ are **linearly independent** if $k_1 v_1 + \ldots + k_m v_m = 0$ implies that all $k_i = 0$. Equivalently, we say that $\{v_1, \ldots, v_m\}$ is a **linearly independent set**.

A set of vectors is called linearly independent, if every finite subset of it is linearly independent. Otherwise a set is called **linearly dependent**.

Examples.

- A vector $v \in V$ forms a linearly independent set $\{v\}$ if and only if $v \neq 0$. The set $\{0\}$ is linearly dependent.
- A set of two vectors $\{u, v\} \subseteq V$, $u \neq 0$, is linearly independent if and only if $v \neq ku$ for some $k \in \mathbb{F}$, i.e., they are not **colinear**.
- A set $\{(2,3), (-1,4), (5,-9)\} \subset \mathbb{R}^2$ is linearly dependent: $1(2,3) + (-3)(-1,4) = (5,-9)$.
- A set $\{1, x, e^x\} \subset C(\mathbb{R})$ is linearly independent. Indeed: let

  $$a1 + bx + ce^x = a + bx + ce^x = 0 \quad \text{for all} \quad x \in \mathbb{R}.$$

  Hence the equality holds, in particular, for $x = 0, 1, -1$. This leads to

  $$a + c = 0, \quad a + b + ce = 0, \quad a - b + ce^{-1} = 0.$$

  Hence $a(1-e)+b = 0$ and $a(1-e^{-1})-b = 0$. Adding these equalities we get $(2-e-e^{-1})a = 0$. Since $e = 2.7182818284590\ldots$, $2 - e - e^{-1} \neq 0$. Hence $a = 0$. Substituting back, we get $b = 0$ and $c = 0$. Hence $\{1, x, e^x\}$ is linearly independent.
- Let $i \in \mathbb{N}$ and $e_i = (x_1, \ldots, x_n, \ldots)$ be a vector (i.e., an infinite sequence) from the vector space $\mathbb{R}^\infty$ such that $x_i = 1$ and $x_j = 0$ for all $j \neq i$. An infinite set of all vectors e_i is linearly independent.

Theorem 3 *Let V be a vector space.*

1. *If $A \subset B$ and B is linearly independent, then A is linearly independent.*

2. *If $A \subset B$ and A is linearly dependent, then B is linearly dependent.*

3. *A set of vectors is linearly dependent if and only if there exists a vector in the set which is a linear combination of other vectors.*

   Or, equivalently,

   A set of vectors is linearly dependent if and only if, there exists a vector in the set which is a linear combination of some other vectors from the set.

   (This explains why the same definition of linear independence is not made for multiset).

   In particular, *no linearly independent set contains* 0 *(zero vector).*

4. *If A is linearly independent subset of V, then*

$$\sum_{a \in A} \beta_a a = \sum_{a \in A} \gamma_a a \quad \textit{implies} \quad \beta_a = \gamma_a \quad \textit{for all} \quad a \in A.$$

   Remark: when denote a linear combination of vectors from A by $\sum_{a \in A} k_a a$, we assume that only finitely many coefficients k_a are nonzero.

5. *Let A be a linearly independent subset of V which does not span V. Let $b \in V \setminus Span(A)$. Then $A \cup \{b\}$ is linearly independent subset of V.*

Proof. Please complete. ∎

Lecture 4.

A set of vectors of V is called a **basis** of V if it spans V and is linearly independent. We will show that every non-trivial vector space has a basis and all bases of V are of the same cardinality. This will justify the following definition.

The cardinality of a basis is called the **dimension of** V. If V has a basis of n vectors, for some $n \geq 1$, we say that V is **finite-dimensional**, or n-dimensional, or has dimension n, or write $\dim V = n$.

If V has no finite basis, it is called **infinite-dimensional**.

We assume that the trivial vector space $\{0\}$ has dimension zero, though it has no basis.

Examples

- Let $n \geq 1$, $1 \leq i \leq n$. Let e_i denote an vector from $\mathbb{F}^n$ having the i-th component 1 and all other components 0. Then $\{e_1, \ldots, e_n\}$ is a basis of $\mathbb{F}^n$, called the **standard** basis of $\mathbb{F}^n$.
- Let $V = \mathbb{Q}[\sqrt{2}] := \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Then V is a vector space over $\mathbb{Q}$ of dimension 2 and $\{1, \sqrt{2}\}$ is a basis.
- Let $P = \mathbb{F}[x]$ be the vector space of all polynomials of x over $\mathbb{F}$. Then P is infinite dimensional. Indeed, if it is not, then it has a finite basis B. Each element of the basis is a polynomial of some (finite) degree. Let m be the greatest of the degrees of the polynomials from the basis. Then $Span(B)$ contains only polynomials of degree at most m and hence $Span(B)$ is a proper subset of P. E.g., $x^{m+1} \notin Span(B)$. The obtained contradiction proves that P is infinitely dimensional over $\mathbb{F}$.

We wish to remind ourselves that the field $\mathbb{F}$ which we decided not to mention every time is there. The same set of objects can be a vector space over different fields, and the notion of dimension depends on the field. For example, $\mathbb{C}$ is vector space over $\mathbb{R}$ of dimension two: $\{1, i\}$ is a basis. The same $\mathbb{C}$ is a vector space over $\mathbb{Q}$ of infinite dimension. All these statements have to be, and will be, be proved.

Question: Why does every vector space have a dimension? In other words, why all its bases have the same cardinality?

This question is not trivial, especially for infinite-dimensional

spaces. Here we will answer it for finitely dimensional spaces.

Proposition 4 *If $Span(A) = V \neq \{0\}$ and $|A| = m$, then some subset of A is a basis of V.*

Proof. By Theorem 3, we can assume that A contains no zero vector. We proceed by induction on m. Let $m = 1$. Then $A = \{v\}$, where $v \neq 0$. Hence A is a basis.

Suppose the statement is proven for all sets A, $|A| = k$, $1 \leq k < m$. $Span(A) = V \neq \{0\}$ and $|A| = m$. If A is linearly independent, then A is a basis, and the proof is finished. Therefore we assume that A is linearly dependent. Then some $v \in A$ is a linear combination of other vectors from A (Theorem 3). Let $A' = A \setminus \{v\}$. Then $Span(A') = Span(A) = V$, and $|A'| = m - 1$. By induction hypothesis A' has a subset which is a basis of V. This subset is also the subset of A, and the proof is finished. ∎

Theorem 5 *Let A be a basis of V and $|A| = n$. Then*

(i) any set of $n + 1$ vectors from V is linearly dependent, and

(ii) any set of $n - 1$ vectors from V does not span V.

Proof. See the handout. ∎

Corollary 6 *In a finite-dimensional space every basis has the same number of vectors.*

Now our definition of dimension for finite-dimensional spaces is completely justified.

Corollary 7 *If a vector space contains an infinite linearly independent subset, then it is infinite-dimensional.*

In mathematics we often look for sets which satisfy a certain property and which are **minimal (maximal)** in the sense that no proper subset (superset) of them satisfy this property. E.g., a minimal set of axioms for a theory, a minimal independence set of a graph, maximal matching in graphs, a minimal generating set for a group, etc.. At the same time we often want to find

a **minimum (maximum)** sets which satisfy a certain property, which are defined as the sets of the smallest (largest) cardinalities which satisfy a given property. The two notions do not always coincide, though it is clear that a finite minimum (maximum) set is always minimal (maximal).

Example. Consider all families of non-empty subsets of the set $\{1, 2, 3, 4, 5\}$ such that the intersection of every two subsets of the family is empty. Examples of such families are $\mathcal{A} = \{\{1, 2, 3, 4\}, \{5\}\}$, or $\mathcal{B} = \{\{1, 5\}, \{2, 3\}, \{4\}\}$. $|\mathcal{A}| = 2$ and $|\mathcal{B}| = 3$. Both $\mathcal{A}$ and $\mathcal{B}$ are maximal, but none of them is maximum, since the family of singletons $\mathcal{C} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$ also possess the property and has 5 members. So a maximum family must contain at least five members.

Nevertheless, we have the following corollary.

Corollary 8 *In a finite-dimensional vector space V*

1. *every maximal linearly independent subset of vectors is a maximum linearly independent subset;*
2. *every minimal spanning subset of vectors is a minimum spanning subset.*

Proof. Complete! ∎.

Problems.

When you do these problems, please do not use any notions or facts of linear algebra which we have not discussed in this course. You must prove all your answers.

1. Let $V = F$ be a field considered as a vector space over itself. Find $\dim V$ and describe all bases of V.

2. Prove or disprove:

   (a) $A = \{(3,1,2), (-1,3,1), (2,4,3)\} \subset \mathbb{R}^3$ is a basis of $\mathbb{R}^3$.

   (b) $A = \{(3,1,2), (-1,3,1), (2,4,-3)\} \subset \mathbb{R}^3$ is a basis $\mathbb{R}^3$.

   (c) $A = \{(t, t^2, t^3) : t \in \mathbb{R}\} \subset \mathbb{R}^3$ is a basis of $\mathbb{R}^3$.

   (d) $A = \{(3,1), (-1,2), (1,0)\} \subset \mathbb{F}_5^2$ is a basis of $\mathbb{F}_5^2$.

3. How many bases does F_p^2 have? You can first try to answer this question for $p = 2, 3, 5$.

4. Let P_3 be the set of all polynomials of x with real coefficients of degree at most three. Show that P_3 is a vector space over $\mathbb{R}$. Show that the set $\{1, x-1, (x-1)^2, (x-1)^3\}$ is a basis of P_3.

   (Remember, that here P_3 is a space of polynomials (as formal sum of monomials), not polynomial functions.)

   Is there a basis of P_3 which contains no polynomial of degree one?

5. The fact that polynomials $1, x, x^2, x^3$ linearly independent as vectors in $\mathbb{F}[x]$ is trivial. It follows from the definition of polynomials.

   Let $1, x, x^2, x^3$ be functions (polynomial functions) from the vector space $C(R)$. Prove that they are linearly independent.

6. Let $V = \{(x, y, z, t) : x + y + 2z - t = 0\} \subseteq \mathbb{R}^4$. Prove that V is a vector space and find a basis of V over $\mathbb{R}$. What is $\dim V$?

7. Let $0 \neq (a_1, \ldots, a_n) \in \mathbb{F}^n$, and let

$$V = \{(x_1, \ldots, x_n) : \ a_1x_1 + \ldots + a_nx_n = 0\} \subseteq \mathbb{F}^n.$$

   Prove that V is a subspace of $\mathbb{F}^n$ and find a basis of V over $\mathbb{F}$. What is $\dim V$?

8. Let $V = \{(x, y, z, t) : x + y + 2z - t = 0 \quad \text{and} \quad x + iy + z - t = 0\} \subseteq \mathbb{C}^4$. Prove that V is a subspace of $\mathbb{C}^4$ and find a basis of V over $\mathbb{C}$. What is $\dim V$?

9. Let V is the vector space of all sequences of real numbers (x_n), $n \geq 0$, defined by the recurrences:

$$x_0 = a,\ x_1 = b, \text{ and for all } \ n \geq 0,\ \ x_{n+2} = x_{n+1} + x_n.$$

Find a basis in V consisting of two geometric progressions, i.e., of sequences of the form $(cr^n),\ n \geq 0$. Write the vector (i.e., sequence) of this space corresponding to $a = b = 1$ as linear combination of the vectors from this basis.

10. Let u, v, w be three distinct vectors in $\mathbb{F}_p^n$. How many vectors can $Span(\{u, v, w\})$ have?

11. Consider $V = \mathbb{R}^\infty$ – the vector space of all infinite sequences of real numbers, with addition of vectors and multiplication of vectors by scalars defined similarly as in $\mathbb{R}^n$. Prove that V is infinite-dimensional.

12. A complex (in particularly real) number α is called **transcendental**, if α is not a root of a polynomial equation with integer coefficients. For example, the famous numbers π and e are transcendental, though the proofs are hard. Explain that the existence of transcendental numbers implies that $\mathbb{R}$ is infinite-dimensional as a vector space over $\mathbb{Q}$.

13. Let $V = \mathbb{R}$ be the vector space over $\mathbb{Q}$. Prove that the set $\{1, 2^{1/3}, 2^{2/3}\}$ is linearly independent.

14. (Optional) Let $V = \mathbb{R}$ be the vector space over $\mathbb{Q}$. Prove that the infinite set

    (i) $\{1, 2^{1/2}, 2^{1/2^2}, \ldots, 2^{1/2^n}, \ldots\}$ is linearly independent

    (ii) $\{\sqrt{p} :\ p \in \mathbb{Z},\ \ p \geq 2,\ \ \text{and } p \text{ is prime}\}$ is linearly independent.

15. (Optional) Prove that the functions e^x, e^{2x} form a basis in the vector space of all solutions of the differential equation $y'' - 3y' + 2y = 0$.

Lecture 5.

We would like to add a few more facts about subspaces and their dimensions. The following notations are useful: by $W \leq V$ we will denote the fact that W is a subspace of V, and we write $W < V$ if W is a proper subspace of V (i.e., $W \neq V$).

Corollary 9 *If $W \leq V$ and V is finite-dimensional, then every basis of W can be extended to a basis of V.*

Proof. Let $\dim V = n$. As $W \leq V$, W has a basis. Denote it by B. If B is a basis of V, we are done. If $W < V$, then $Span(B) = W < V$. Take $v_1 \in V \setminus W$. Then $B_1 = B \cup \{v_1\}$ is linearly independent. If $Span(B_1) = V$, we are done. If not, take $v_2 \in V \setminus Span(B_1)$. Then $B_2 = B_1 \cup v_2 = B \cup \{v_1, v_2\}$ is linearly independent. If $Span(B_2) = V$, we are done. If not we proceed in a similar way. As V is n-dimensional, there are no $n+1$ linearly independent vectors in V. Hence the process will terminate with a basis of V containing B as a subset. ∎

Corollary 10 *If $W \leq V$ and $\dim W = \dim V = n$, then $W = V$.*

Proof. If $W < V$, then a basis of W, which has n vectors in it, is not a basis of V. By the previous corollary, it can be extended to a basis of V which will contain $n+1$ elements. This is impossible, since every basis of V contains n vectors. ∎

QUESTION: Let U and W be subspaces of finite-dimensional space V. What is $\dim(U+W)$?

Theorem 11 *For any finite-dimensional subspaces U and W of V,*

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W).$$

Moreover, $\dim(U+W) = \dim U + \dim W$ *if and only if* $U \cap W = \{0\}$.

Proof. Let $\dim U = p$, $\dim W = q$, and $\dim(U \cap W) = r$. As $U \cap W$ is a subspace of both U and W, a basis $\{v_1, \ldots, v_r\}$ of $U \cap W$ can be extended to a basis $\{v_1, \ldots, v_r, u_1, \ldots, u_{p-r}\}$ of U, and to a basis $\{v_1, \ldots, v_r, w_1, \ldots, w_{q-r}\}$ of W (by Corollary 9).

We claim that

$$B = \{v_1, \ldots, v_r, u_1, \ldots, u_{p-r}, w_1, \ldots, w_{q-r}\}$$

is a basis of $U + W$.

Indeed, every vector $u + w \in U + W$, where $u \in U$, $w \in W$, is, clearly, in $Span(B)$. What is left is to show that B is linearly independent. Let

$$\sum_{i=1}^{r} a_i v_i + \sum_{i=1}^{p-r} b_i u_i + \sum_{i=1}^{q-r} c_i w_i = 0. \tag{1}$$

This can be rewritten as

$$u := \sum_{i=1}^{p-r} b_i u_i = -\sum_{i=1}^{r} a_i v_i - \sum_{i=1}^{q-r} c_i w_i =: v.$$

As $u \in U$ and $v \in W$, then $u \in U \cap W$. Hence $u = d_1 v_1 + \ldots + d_r v_r$, and we have

$$\sum_{i=1}^{r} (d_i + a_i) v_i + \sum_{i=1}^{q-r} c_i w_i = 0.$$

Since $\{v_1, \ldots, v_r, w_1, \ldots, w_{q-r}\}$ is linearly independent (as a basis of W), then all the coefficients are zeros. In particular all c_i are zeros. A similar argument gives all b_i equal zero. Then (1) gives

$$\sum_{i=1}^{r} a_i v_i = 0,$$

and as $\{v_1, \ldots, v_r\}$ is linearly independent (as a basis of $U \cap W$, all a_i are zeros. Hence B is linearly independent and

$$\dim(U + W) = |B| = r - (p - r) + (q - r) = p + q - r.$$

The last statement is trivial, since $r = 0$ if and only if $U \cap W = \{0\}$. ∎

The space $U + W$ is the smallest subspace of V which contains $U \cup W$. If for every $v \in U + W$, there exist unique $u \in U$ and $w \in W$ such that $v = u + w$, then $U + V$ is called the **direct sum** of its subspaces U and V or the **internal direct sum** of U and W, and it is denoted by $U \oplus W$.

Proposition 12 *$U + W = U \oplus W$ if and only if $U \cap W = \{0\}$.*

Proof. Done in class. ∎

Let V_1 and V_2 be vector spaces over the same field $\mathbb{F}$. Consider the set

$$V_1 \times V_2 = \{(v_1, v_2) : \ v_1 \in V_1, v_2 \in V_2.\}$$

If we define

$$(v_1, v_2) + (v_1', v_2') = (v_1 + v_1', v_2 + v_2') \text{ and } k(v_1, v_2) = (kv_1, kv_2) \text{ for } k \in \mathbb{F},$$

then it is easy to check that $V_1 \times V_2$ is a vector space over $\mathbb{F}$. It is called the **direct product** or the **external direct product** of V_1 and V_2, and will be denoted by $V_1 \times V_2$. The direct product of more than two vector spaces over $\mathbb{F}$ is defined similarly.

A reader may get a feeling that the notions of the direct sum and direct product are very similar, and there is no distinction between the resulting vector spaces. Note also that, opposite to U and W being subspaces of $U \oplus W$, neither V_1 nor V_2 is a subspace of $V_1 \times V_2$. All this may be a little bothering. The notion of isomorphism, which we are going to define very soon, will help to discuss these issues in a precise way.

Problems.

When you do these problems, please do not use any notions or facts of linear algebra which we have not discussed in this course. You must prove all your answers.

1. It is clear how to generalize the definition of the direct product of two vector spaces to any finite (or infinite) collection of vector spaces over the same field $\mathbb{F}$.

   What about the direct sum of subspaces? Try $n \geq 2$ subspaces. Can you state and prove a statement similar to Proposition 12? If you do not see how to do it for any $n \geq 2$ subspaces, maybe do it first for three subspaces. Then state the result for n subspaces. You do not have to prove it.

   Can you extend the definition of the direct sum for infinitely many subspaces? If you can, please state it.

2. Extend Theorem 11 to three finite-dimensional subspaces X, Y, Z of a vector space V and prove it.

   Can you extend the statement of Theorem 11 to any $n \geq 2$ finite-dimensional subspaces? You do not have to prove the statement.

3. Proof that $C(\mathbb{R}) = E \oplus O$, where E is the subspace of all even functions and O is the subspace of all odd functions.

   (A function $f : \mathbb{R} \to \mathbb{R}$ is called **even** (resp. **odd**) if for all $x \in \mathbb{R}$, $f(-x) = f(x)$ (resp. $f(-x) = -f(x)$).)

4. How many k-dimensional subspaces, $k = 1, 2, 3$, does $\mathbb{F}_p^5$ have? You can first try to answer this question for $p = 2, 3$.

5. Can two 4-dimensional subspaces of $\mathbb{F}_2^6$ have exactly 20 vectors in common?

6. How many k-dimensional subspaces, $k \geq 1$ and fixed, does $C(\mathbb{R})$ (over $\mathbb{R}$) have?

7. Prove that functions $1, e^x, e^{2x}, \ldots, e^{nx}$, $n \geq 1$ and fixed, are linearly independent as vectors in $C^\infty(\mathbb{R})$.

8. Give an example of three functions $f_1, f_2, f_3 \in C^\infty(\mathbb{R})$, and three distinct real numbers a, b, c, such that f_1, f_2, f_3 are linearly independent, but the vectors $(f_1(a), f_2(a), f_3(a))$, $(f_1(b), f_2(b), f_3(b))$, $(f_1(c), f_2(c), f_3(c))$ are linearly dependent as vectors in $\mathbb{R}^3$.

9. Let $U = \langle \sin x, \sin 2x, \sin 3x \rangle$ and $W = \langle \cos x, \cos 2x, \cos 3x \rangle$ be two subspaces in $C^\infty(\mathbb{R})$. Find $\dim(U \cap W)$.

   (The symbol $\langle v_1, \ldots, v_n \rangle$ is just another common notation for $Span(\{v_1, \ldots, v_n\})$).

10. Prove that if $\dim V = n$ and $U \leq V$, then there exists $W \leq V$ such that $V = U \oplus W$. Does such W have to be unique for a given U and V?

11. Prove that $\mathbb{R}^n$ is the direct sum of two subspaces defined as:

$$U = \{(x_1, \ldots, x_n) : \ x_1 + \ldots + x_n = 0\} \quad \text{and}$$

$$W = \{(x_1, \ldots, x_n) : \ x_1 = x_2 = \cdots = x_n\}.$$

12. Let $U = \{(x_1, \ldots, x_4) : \ x_1 + x_2 + x_3 - x_4 = 0, \text{ and } x_1 - x_3 = 0, \text{ all } x_i \in \mathbb{R}\} \leq \mathbb{R}^4$. Find a basis of a subspace $W \leq \mathbb{R}^4$ such that $U \oplus W = \mathbb{R}^4$.

13. (Optional) Let $V = \mathbb{R}$ be the vector space over $\mathbb{Q}$. Prove that the vectors π and $\cos^{-1}(1/3)$ are linearly independent.

14. (Optional) A **hyperplane** in an n-dimensional space, $n \geq 1$, is any subspace of dimension $n - 1$. Is $\mathbb{R}^n$ a union of a finite number of hyperplanes?

Lecture 6.

Strange objects (vector spaces) have to be studied by strange methods. This is how we arrive to linear mappings. Since linear mappings are functions, we review related terminology and facts. We do it here in a very brief way.

Given sets A and B, a function f from A to B is a subset of $A \times B$ such that for every $a \in A$ the exists a unique $b \in B$ that $(a, b) \in f$. The fact that f is a function from A to B is represented by writing $f : A \to B$. The fact that $(a, b) \in f$ is represented by writing by $f : a \mapsto b$, or $fa = b$, or in the usual way: $f(a) = b$. We also say that f **maps** a to b, or that b is an **image of** a in f.

Let $\mathbf{im}\ f := \{b \in B : \ b = f(a) \text{ for some } a \in A\}$. $\mathbf{im}\ f$ is called the **the image of** f, or the **range** of f.

We say that f is **one-to-one** or **injective** if for every $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ implies $a_1 = a_2$. Or, equivalently, if for every $a_1, a_2 \in A$, $a_1 \neq a_2$ implies $f(a_1) \neq f(a_2)$.

We say that f is **onto** or **surjective** if $\mathbf{im}\ f = B$.

We say that f is **bijective** if it is one-to-one and onto, or, equivalently, if f is both injective and surjective.

Let $A_1 \subseteq A$ and $f : A \to B$. Then

$$f|_{A_1} = \{(a, b) : a \in A_1, f(a) = b\} \subseteq A_1 \times B$$

is a function from A_1 to B, called the **restriction** of f on A_1.

Given a function $f : A \to B$ and a function $g : B \to C$, one can consider the set

$$h = \{(a, g(f(a))) : a \in A\} \subseteq A \times C.$$

It is easy to show that h is a function from A to C, and $h(a) = g(f(a))$. It is called the **composition** of functions f and g, and denoted by $g \circ f$. It is easy to check that h is

- injective if and only if both f and $g|_{\mathbf{im}\ f}$ are injective.
- surjective if and only if $g|_{\mathbf{im}\ f}$ is surjective;
- bijective if and only if f is injective and $g|_{\mathbf{im}\ f}$ is bijective.

It may happen that for $f : A \to B$, the set $\{(b, a) : f(a) = b\} \subset B \times A$ is a function from B to A. This function is denoted by f^{-1} and called the **inverse** (function) of f. It is clear that this happens, i.e., f^{-1} exists, if and only if f is a bijection.

Let V and W be two vector spaces over the same field $\mathbb{F}$, and let $f : V \to W$ satisfy the following properties:

- for all $x, y \in V$, $f(x + y) = f(x) + f(y)$, and
- for all $x \in V$, and all $k \in \mathbb{F}$, $f(kx) = kf(x)$.

Then f is called a **linear map** (or mapping) from V to W. If $V = W$, a linear map from V to V is called a **linear operator** on V, or a **linear transformation** of V.

Here are some examples of linear maps. Verification that the maps are linear is left to the reader.

- Let $f : V \to V$, where $f(x) = x$ for all $x \in V$ – the **identity** map.
- Let $f : V \to W$, where $f(x) = 0$ for all $x \in V$ – the **zero** map.
- Let $V = W$. Fix $k \in F$, and let $f : V \to V$ be defined via $x \mapsto kx$.
- Let A be an $m \times n$ matrix with entries from $\mathbb{F}$, $V = \mathbb{F}^n$, $W = \mathbb{F}^m$, and let $f : V \to W$ be defined via $x \mapsto Ax$. (Here vectors of V or W are thought as columns.)
- Let $V = C^2(a, b)$, $W = C^1(a, b)$, and let $f : V \to W$ be defined via $f \mapsto f'$ – the derivative of f.
- Let $V = C^2(a, b)$, $W = C(a, b)$, and let $f : V \to W$ be defined via $f \mapsto 2f'' - 3f' + f$.
- Let $V = C[a, b]$, $W = \mathbb{R}$, and let $f : V \to W$ be defined via $f \mapsto \int_a^b f(x)\,dx$ – the definite integral of f on $[a, b]$.
- Let $V = W = \mathbb{R}^\infty$, and let $f : V \to W$ be defined via $(x_1, x_2, x_3 \ldots) \mapsto (x_2, x_3, \ldots)$ – the **backward shift** on $\mathbb{R}^\infty$.

An easy way to construct a linear map is the following. Let $\{v_1, \ldots, v_n\}$ be a basis of V. Chose arbitrary n vectors $\{w_1, \ldots, w_n\}$ in W and define $f : V \to W$ via $\sum_{i=1}^n k_i v_i \mapsto \sum_{i=1}^n k_i w_i$ for all possible choices of $k_i \in \mathbb{F}$. As every vector of V is a unique linear combination of vectors v_i, f is a function. It is easy to see that f is linear.

For a linear map $f : V \to W$ let $\mathbf{ker}\ f := \{v \in V : f(v) = 0\}$. The set $\mathbf{ker}\ f$ is called the **kernel** of f.

We collect several important properties of linear maps in the following theorem.

Theorem 13 *Let V and W be vector spaces over $\mathbb{F}$, and let $f : V \to W$ be a linear map. Then*

1. $f(0) = 0$.

2. $f(-v) = -f(v)$ *for every* $v \in V$.

3. $\mathbf{ker}\ f \leq V$, *and* $f(x) = f(y)$ *if and only if* $x - y \in \mathbf{ker}\ f$.

4. f *is injective if and only if* $\mathbf{ker}\ f = \{0\}$.

5. $\mathbf{im}\ f \leq W$

6. *If* V *is finite-dimensional, then* $\dim \mathbf{ker}\ f + \dim \mathbf{im}\ f = \dim V$.

Proof. Will be discussed in class. ∎

We introduce three more definitions. An injective linear map is called a **monomorphism**, a surjective linear map is called an **epimorphism**, and a bijective linear map is called an **isomorphism**.

Corollary 14 *Let* V *and* W *be finite-dimensional spaces over* $\mathbb{F}$, *and let* $f : V \to W$ *be a linear map.*

1. *If* f *is a monomorphism, then* $\dim V \leq \dim W$.

2. *If* f *is an epimorphism, then* $\dim V \geq \dim W$.

3. f *is an isomorphism, then* $\dim V = \dim W$.

4. *If* $\dim V = \dim W$, *then there exists an isomorphism from* V *to* W.

Proof.

1. Since f in injective, then $\mathbf{ker}\ f = \langle 0 \rangle$. Hence $\dim V = \dim \mathbf{ker}\ f + \dim \mathbf{im}\ f = \dim \mathbf{im}\ f \leq \dim W$.

2. Since f in surjective, then $\dim V = \dim \mathbf{ker}\ f + \dim \mathbf{im}\ f \leq \dim \mathbf{ker}\ f + \dim W \geq \dim W$.

3. Follows from the first two statements.

4. Let n be the common dimension of V and W, and let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ be the bases of V and W, respectively. Define $f : V \to W$ by $\sum_{i=1}^{n} k_i v_i \mapsto \sum_{i=1}^{n} k_i w_i$. Since $\{v_1, \ldots, v_n\}$ is a basis, f is well-defined. Also f is a linear map, and $\mathbf{ker}\ f = \langle 0 \rangle$. So f is injective. Also, $\dim \mathbf{im}\ f = n - \dim \mathbf{ker}\ f = n - 0 = n$, hence $\mathbf{im}\ f = W$. This implies that f is onto, and therefore is an isomorphism. ∎

We will denote the fact that V is isomorphic to U by $V \simeq U$.

So, why do we study distinct finite-dimensional spaces if we could concentrate on just $\mathbb{F}^n$? Yes, as just vector spaces over the same field, all n-dimensional spaces are isomorphic, and this is very helpful! But often we ask questions about vectors which are of interest for a given vector space only, and which are related to the properties not preserved by isomorphisms. We will see many such questions in this course.

Problems.

1. Let $A = \{v_1, \dots, v_n\}$ be a linearly dependent subset of V. Chose arbitrary n vectors $\{w_1, \dots, w_n\}$ in W and try to define a function f from $Span(A)$ to W by mapping $\sum_{i=1}^n k_i v_i \mapsto \sum_{i=1}^n k_i w_i$ for all possible choices of $k_i \in \mathbb{F}$. Will f necessarily be a function? Sometimes the same question is phrased as: "Will f be well-defined?"

2. Can you describe **ker** f and **im** f for all linear maps f from the examples of this lecture?

3. Let V and W be vector spaces over $\mathbb{F}$, and let $f : V \to W$ be a function satisfying one of the two conditions required for f being linear. For each of the two conditions, find an example of f which satisfies this condition, but not the other one.

4. Find an isomorphism between $\mathbb{F}^{n+1}$ and $P_n(\mathbb{F})$ – the vector space of all polynomials over $\mathbb{F}$ of degree at most n.

5. Find an isomorphism between $\mathbb{F}^{mn}$ and $M_{m\times n}(\mathbb{F})$ – the vector space of all $m \times n$ matrices over $\mathbb{F}$.

6. Let V be a finite dimensional space over $\mathbb{F}$. Decide whether the following statements are true or false? Explain.

   (i) If $V = V_1 \oplus V_2$, then $V \simeq V_1 \times V_2$.

   (ii) If $V \simeq V_1 \times V_2$, where V_1, V_2 are subspaces of V, then $V = V_1 \oplus V_2$.

   (iii) Let $f : V \to V$ be a linear operator on V. Then $V \simeq$ **ker** $f \times$ **im** f.

   (iv) Let $f : V \to V$ be a linear operator on V. Then $V =$ **ker** $f \oplus$ **im** f.

7. Let U, W be subspaces of V and $\dim V$ be finite. Reprove the formula

$$\dim(U + W) = \dim U + \dim W - \dim U \cap W$$

   by considering the map $f : U \times W \to V$ given by $f((u, w)) = u - w$. Hint: is f linear? what is **ker** f? **im** f?

8. (Optional) Suppose S is a set of $2n+1$ irrational real numbers. Prove that S has a subset T of $n + 1$ elements such that no nonempty subset of T has a rational sum.

When you come to a fork on the road take it.

– Yogi Berra

Lecture 7.

What do we do next? There are many natural questions we can ask about vector spaces at this point. For example, how do they relate to geometry or other parts of mathematics? Or to physics? Did they help Google or the national security?

I hope we will touch all these relations, but now we will talk about matrices, objects which are inseparable from vector spaces. We assume that the reader is familiar with basic definitions and facts about matrices. Below we wish to discuss some natural questions which lead to matrices.

One can arrive to matrices in many ways.

1. Trying to solve systems of linear equations is one of them. There matrices and vectors (the latter also can be viewed as $n \times 1$ or $1 \times n$ matrices) appear as very convenient *notations*. As all of you know, the problem of solving a system of m linear equations each with n variables can be restated as finding a column vector $x \in \mathbb{F}^n$ such that $Ax = b$, where $A = (a_{ij})$ is the matrix of the coefficients of the system, and $b \in \mathbb{F}^m$ is the column vector representing the right hand sides of the equations. Here the *definition* for the multiplication of A by x is chosen in such a way that $Ax = b$ is just a short way of rewriting the system. It seems that nothing is gained by this rewriting, but not quite. Somehow this way of writing reminds us about the simplest linear equation $ax = b$, where $a, b \in \mathbb{F}$ are given numbers and x

is the unknown number. We know that we can always solve it if $a \neq 0$, and the unique solution is $x = a^{-1}b$. The logic of arriving to this solution is as follows:

$$ax = b \;\Leftrightarrow\; a^{-1}(ax) = a^{-1}b \;\Leftrightarrow\; (a^{-1}a)x = a^{-1}b \;\Leftrightarrow\; 1(x) = a^{-1}b \;\Leftrightarrow\; x = a^{-1}b.$$

We also know that for $a = 0$, we either have no solutions (if $b \neq 0$), or every element of $\mathbb{F}$ is a solution (if $b = 0$). Analogy in appearance, suggests analogy in the approach, and we may try to invent something like 0, or 1, or A^{-1} for matrices. We may ask the question whether the product of matrices is associative, etc. .

Trying to push the analogy, we may say that, in $\mathbb{F}$, it does not matter whether we are solving $ax = b$ or $xa = b$ or $xa - b = 0$. Trying to see whether it is true for matrices, we immediately realize that xA (what is it ???) has little to do with already introduced Ax, and that the obvious candidate for zero-matrix, does not allow to claim that if A is not 'the' zero matrix, then $Ax = b$ is always solvable. Thinking about all this, one may come to the usual non-commutative (but associative) ring of square matrices $M_{n\times n}(\mathbb{F})$. Analyzing further, one realizes

that it is convenient to write the matrix (ca_{ij}) as cA: a great simplification of actual writing! Now we do not need to repeat c mn times. Doing this for a while, one may begin saying that one 'is multiplying a matrix by a number', and that every matrix can be written uniquely as a 'linear expression', or 'a linear form', or 'a linear combination' of some very simple mn matrices. That is how we can start viewing $M_{m\times n}(\mathbb{F})$ as a vector space over $\mathbb{F}$, and forget about the multiplication of matrices if it is not needed for what we are doing....

2. Another way to arrive to matrices, especially to the notion of matrix multiplication, can be through doing changes of variables. The method was used long before vectors or matrices were born. Mostly in number theory or in geometry, when the coordinates were used. If $x = 3a - 2b$ and $y = a+5b$, and $a = e-f-5g$, and $b = 2e+f+7g$, then $x = 3(e-f-5g)-2(2e+f+7g) = -e-5f-29g$, and $y = (e-f-5g)+5(2e+f+7g) = 11e+4f+30g$. The expression for x and y in terms of e, f, and g, can be obtain via the following computation with matrices:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix} \left(\begin{bmatrix} 1 & -1 & -5 \\ 2 & 1 & 7 \end{bmatrix} \begin{bmatrix} e \\ f \\ g \end{bmatrix} \right) = \tag{2}$$

$$\left(\begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & -1 & -5 \\ 2 & 1 & 7 \end{bmatrix} \right) \begin{bmatrix} e \\ f \\ g \end{bmatrix} = \begin{bmatrix} -1 & -5 & -29 \\ 11 & 4 & 30 \end{bmatrix} \begin{bmatrix} e \\ f \\ g \end{bmatrix} \tag{3}$$

Tracing how the coefficients are transformed, leads to the rule for matrix multiplication. This rule will become more visible if we use letters instead of numbers for the coefficients in our transformations.

3. A linear map $f : U \to V$ can be represented by a matrix in the following way. Let $\{u_1, \ldots, u_m\}$ be a basis of U, and let $\{v_1, \ldots, v_n\}$ be a basis of V. Then f is completely defined by a $m \times n$ matrix $M_f = (a_{ij})$, where $f(u_i) = a_{i1}v_1 + \ldots a_{in}v_n$, $i \in \{1, \ldots, m\}$. One can use matrix notation to write this map as

$$\begin{bmatrix} f(u_1) \\ \ldots \\ f(u_m) \end{bmatrix} = M_f \begin{bmatrix} v_1 \\ \ldots \\ v_n \end{bmatrix} \tag{4}$$

Note that in this notation the column matrices are not from $\mathbb{F}^n$ and $\mathbb{F}^m$, as they usually are. It is just a convenient way of expressing the set of equalities $f(u_i) = a_{i1}v_1 + \ldots a_{in}v_n$, $i \in \{1, \ldots, m\}$.

It is clear that the correspondence $f \mapsto M_f$ defined by (4) between the set of all linear maps from U to V and the set of all $m \times n$ matrices over $\mathbb{F}$ is a bijection, and depends heavily on the choices of two bases.

A linear map $g : V \to W$ can be represented by a matrix in a similar way. Let $\{w_1, \dots, w_q\}$ be a basis of W. Then g is completely defined by a $n \times q$ matrix $B = (b_{k,l})$, where $g(v_k) = b_{k1}w_1 + \dots b_{kq}w_q$, $k \in \{1, \dots, n\}$. By using matrix notation one can represent this map as

$$\begin{bmatrix} g(v_1) \\ \dots \\ g(v_n) \end{bmatrix} = M_g \begin{bmatrix} w_1 \\ \dots \\ w_q \end{bmatrix} \tag{5}$$

Then the composition $g \circ f$ of linear maps f and g, which is a linear map itself, is given by a $m \times q$ matrix $M_{g\circ f} = (c_{st})$, where $(g \circ f)(u_s) = c_{s1}w_1 + \dots + c_{sq}w_q$, $s \in \{1, \dots, m\}$. Let us express the coefficients c_{st} in terms of a_{ij} and b_{kl}.

$$(g \circ f)(u_s) = g(f(u_s)) = g\big(\sum_j a_{sj}v_j\big) = \sum_j a_{sj}g(v_j) =$$
$$\sum_j a_{sj}\big(\sum_t b_{jt}w_t\big) = \sum_j \big(\sum_t a_{sj}b_{jt}w_t\big) =$$
$$\sum_t \big(\sum_j a_{sj}b_{jt}\big)w_t.$$

Therefore $c_{st} = \sum_j a_{sj}b_{jt}$, and we obtain

$$M_{g\circ f} = M_f\, M_g.$$

Hence

$$\begin{bmatrix} (g \circ f)(u_1) \\ \dots \\ (g \circ f)(u_m) \end{bmatrix} = M_{g\circ f} \begin{bmatrix} w_1 \\ \dots \\ w_q \end{bmatrix} = (M_f M_g) \begin{bmatrix} w_1 \\ \dots \\ w_q \end{bmatrix} \tag{6}$$

Though equality (6) makes the relation between the matrix of a composition of linear maps and the product of the corresponding matrices very clear, one should not read more from it than what it displays. Contrary to the associativity of the product of matrices in (2) and (3), the right hand side of (6) is not the usual product of three matrices: trying to check the associativity, we have difficulties with the meaning of the 'products' involved. If the column vector of vectors w_i is denoted ny $\vec{w}$, what is the meaning of $M_g\vec{w}$? Or of $M_f\vec{w}$, if we hope that multiplying by M_f first may help?

Problems.

1. Check that $M_{m\times n}(\mathbb{F})$ is a vector space over $\mathbb{F}$ of dimension mn.

2. Check that the only matrix $X \in M_{m\times n}(\mathbb{F})$ with the property that $X + A = A$ for all $A \in M_{m\times n}(\mathbb{F})$ is the **zero matrix**, i.e. the one with all entries equal zero.

3. Check that the only matrix $X \in M_{m\times m}(\mathbb{F})$ with the property that $XA = A$ for all $A \in M_{m\times n}(\mathbb{F})$ is the identity matrix I_m =diag$(1, 1, \ldots, 1)$. Similarly, the only matrix $Y \in M_{n\times n}(\mathbb{F})$ with the property that $AY = A$ for all $A \in M_{m\times n}(\mathbb{F})$ is the identity matrix I_n =diag$(1, 1, \ldots, 1)$.

4. Check that if AB and BA are defined, then both AB and BA are square matrices. Show that the matrix multiplication of square matrices is not, in general, commutative.

5. Check that the following products of three matrices, $(AB)C$ and $A(BC)$, always exist simultaneously, and are equal. Hence the matrix multiplication is associative. Do this exercise in two different way. First by using the formal definition for matrix multiplication and manipulating sums. Then by using the correspondence between matrix multiplication and composition of linear maps. Prove the fact that composition of three functions is associative.

6. Check that for any three matrices A, B, C over the same field $\mathbb{F}$, $A(B + C) = AB + AC$, and $(A + B)C = AC + BC$, provided that all operations are defined. These are the distributive laws.

7. Show that there exist matrices $A, B, C \in M_{2\times 2}(\mathbb{F})$ such that $AB = AC = 0$, $A \neq 0$, $B \neq C$.

8. Matrices from $M_{n\times n}(\mathbb{F})$ are also referred to as **square matrices of order n.** Let A be a square matrix of order n. A matrix B is called **the inverse of A, if** $AB = BA = I_n$.

   (i) Prove that if the inverse matrix exists, then it is unique.

   The inverse matrix, if it exists, is denoted by A^{-1}, and A is called **nonsingular**, otherwise it is called **singular**.

   (ii) Give an example of a 2×2 matrix A such that A is not zero matrix, and A^{-1} does not exist.

9. Prove that if the Show that there exist matrices $A, B, C \in M_{2\times 2}(\mathbb{F})$ such that $AB = AC = 0$, $A \neq 0$, $B \neq C$.

10. Let $A = (a_{ij}) \in M_{n\times n}(\mathbb{F})$. Define **tr** $A := \sum_{i=1}^{n} a_{ii}$. The field element **tr** A is called the **trace** of A.

    Prove that $f : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ defined via $A \mapsto$ **tr** A is a linear map, and that **tr** $AB =$ **tr** BA for every two matrices $A, B \in M_{n\times n}(\mathbb{F})$.

11. Prove that the matrix equation $XY - YX = I_n$ has no solutions with $X, Y \in M_{n\times n}(\mathbb{F})$, where $\mathbb{F}$ is a subfield of $\mathbb{C}$.

12. Find the set of all matrices $C \in M_{n\times n}(\mathbb{F})$ such that $CA = AC$ for all $A \in M_{n\times n}(\mathbb{F})$.

Lecture 8.

Let $\dim V = n$, and let $f : V \to V$ be an isomorphism of V. Let $\alpha = \{u_1, \ldots, u_n\}$, and $\beta = \{v_1, \ldots, v_n\}$ be bases of V, and for each i, let $f(u_i) = a_{i1}v_1 + \ldots + a_{in}v_n$. Then $M_f = (a_{ij})$. Since f is an isomorphism of V, then f^{-1} exists (obvious), is linear (check), and therefore is also an isomorphism of V. Let $f^{-1}(v_i) = b_{i1}u_1 + \ldots + b_{in}u_n$. Then $M_{f^{-1}} = (b_{ij})$. As $= f^{-1} \circ f = id$ – the identity map on V, the matrix of id with respect to bases α and α is I_n. Similarly, $f \circ f^{-1} = id$, and the matrix of id with respect to bases β and β is I_n. Therefore we have

$$M_{f^{-1}} M_f = M_f M_{f^{-1}} = I_n.$$

This implies that

$$\boxed{M_{f^{-1}} = (M_f)^{-1}} \tag{7}$$

With α and β as above, let $g : V \to V$ be a linear map. Then we can represent g in two different ways:

$$\begin{bmatrix} g(u_1) \\ \ldots \\ g(u_n) \end{bmatrix} = M_{g,\alpha} \begin{bmatrix} u_1 \\ \ldots \\ u_n \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} g(v_1) \\ \ldots \\ g(v_n) \end{bmatrix} = M_{g,\beta} \begin{bmatrix} v_1 \\ \ldots \\ v_n \end{bmatrix} \tag{8}$$

As β is a basis, we have

$$\begin{bmatrix} u_1 \\ \ldots \\ u_n \end{bmatrix} = C \begin{bmatrix} v_1 \\ \ldots \\ v_n \end{bmatrix}, \tag{9}$$

for some matrix C. Since $C = M_{id}$ with respect to the bases α and β, and id is an isomorphism of V to V, then C is an invertible matrix, as we proved at the beginning of this lecture. Next we notice that (9) implies

$$\begin{bmatrix} g(u_1) \\ \ldots \\ g(u_n) \end{bmatrix} = C \begin{bmatrix} g(v_1) \\ \ldots \\ g(v_n) \end{bmatrix} \tag{10}$$

Equalities (8), (9), (10), and the associativity of matrix multiplication imply that

$$\begin{bmatrix} g(u_1) \\ \ldots \\ g(u_n) \end{bmatrix} = (M_{g,\alpha}\, C) \begin{bmatrix} v_1 \\ \ldots \\ v_n \end{bmatrix} = (C\, M_{g,\beta}) \begin{bmatrix} v_1 \\ \ldots \\ v_n \end{bmatrix} \tag{11}$$

As vectors $\{v_1, \ldots, v_n\}$ are linearly independent, we obtain $M_{g,\alpha}\, C = C\, M_{g,\beta}$, or, since C is invertible,

$$\boxed{M_{g,\beta} = C^{-1}\, M_{g,\alpha} C} \tag{12}$$

Equation (12) represents the change of a matrix of a linear map g on V with respect to the change of bases of V.

We conclude this lecture with another important result. Though it is also related to a change of basis of V, it is about a different issue.

Every vector $x \in V$ can be written uniquely as $x = x_1u_1 + \ldots + x_nu_n$, where all $x_i \in F$. The scalars x_i are called the **coordinates of x in basis α**. The column vector (x_i), i.e., $n \times 1$ matrix (x_{i1}) with all $x_{i1} = x_i$, is called the **coordinate vector of x in basis α**, and we denote it by $[x]_\alpha$. It is clear that the map $f : V \to \mathbb{F}^n$ defined via $x \mapsto [x]_\alpha$ is an isomorphism. The following theorem describes the change of $[x]_\alpha$ if we change basis α to $\beta = \{v_1, \ldots, v_n\}$.

Theorem 15 *Let V be vector spaces, and let $\alpha = \{u_1, \ldots, u_n\}$ and $\beta = \{v_1, \ldots, v_n\}$ be two bases of V, and let $A = (a_{ij})$ be a $n \times n$ matrix over $\mathbb{F}$, such that $u_j = \sum_i a_{ij}v_i$. Then*

$$[x]_\beta = A\,[x]_\alpha.$$

Proof. We have

$$x = \sum_j x_j u_j = \sum_j x_j \big(\sum_i a_{ij} v_i \big) = \sum_i \big(\sum_j a_{ij} x_j \big) v_i.$$

Now we observe that $(i1)$-th, or just the i-th, entry of the column vector $A\,[x]_\alpha$ is precisely $\sum_j a_{ij}x_j$. ∎

Remarks

- This theorem becomes very useful when we want to convert coordinates of *many* vectors from one fixed basis to another fixed basis.
- Note that according to our definition of the coordinate vector, $[u_j]_\beta$ is the j-th column of matrix A.
- If α is a standard basis of $\mathbb{F}^n$, and $x \in \mathbb{F}^n$, then the i-th components of x and $[x]_\alpha$ are equal.

For any $m \times n$ matrix $A = (a_{ij})$, let A^T denote the **transpose of A**, i.e., the $n \times m$ matrix (a'_{kl}), where $a'_{kl} = a_{lk}$ for all $k \in \{1, \ldots, n\}$ and $l \in \{1, \ldots, m\}$. If in the statement of Theorem 15, we wrote $u_i = \sum_j a_{ij}v_j$, then the result would change to

$$[x]_\beta = A^T\,[x]_\alpha.$$

By using the notion of the transpose matrix, we can write $[v_j]_\beta = [a_{1j}, a_{2j}, \ldots, a_{nj}]^T$. So the transpose notation is a good way to avoid writing column vectors!

One may ask why we do not just use row notations for vectors from $\mathbb{F}^n$. Many texts do it. In this case, to multiply a vector and a matrix, one would write xA. The approach has both advantage and disadvantages. One disadvantage is having x on the left of A when we want to consider a function defined by $x \mapsto xA$.

All this said, we will adopt the following convention:

No matter how we write vectors we think about them as columns.
Row vectors will be represented by means of the transposition symbol.

Problems.

I recommend that you check your computations with any CAP (Computer Algebra Package), i.e., Maple, or Mathematica, etc.

1. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear map defined as $f([x, y, z]) = [2x - y, x + y - 2z]$, and let α and β be bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively. Find matrix M_f of f if

   (a) α and β are the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively;

   (b) α is the standard basis, and $\beta = \{v_1 = [1, -1],\ v_2 = [1, 3]\}$;

   (c) $\alpha = \{u_1 = [1, -1, 0],\ u_2 = [0, 1, -1],\ u_3 = [0, 2, 1]$, and β is the standard basis;

   (d) $\alpha = \{u_1 = [1, -1, 0],\ u_2 = [0, 1, -1],\ u_3 = [0, 2, 1]$, and $\beta = \{v_1 = [1, -1],\ v_2 = [1, 3]\}$.

2. Let $\alpha = \{v_1 = [1, 1, 0],\ v_2 = [0, 1, 1],\ v_3 = [1, 0, 1]\} \subset \mathbb{R}^3$, and let f be a linear operator on $\mathbb{R}^3$ such that $[f(v_1)]_\alpha = [-1, -3, -3],\ [f(v_2)]_\alpha = [3, 5, 3],\ [f(v_3)]_\alpha = [-1, -1, 1]$. Find $M_{f,\alpha}$. Then find $M_{f,\beta}$ for $\beta = \{[1, 1, 1],\ [1, 0, 0],\ [1, 0, -3]\}$.

3. Let $V = P_3$ be the vector space of all polynomials with real coefficients of degree at most 3. View P_3 as the subspace of $\mathbb{C}^\infty(\mathbb{R})$. Let $d : V \to V$ defined as $f \mapsto f' = \frac{df}{dx}$. Find $M_{d,\alpha}$, $M_{d,\beta}$ for bases: $\alpha = \{1, x, x^2, x^3\}$ and $\beta = \{1, x - 2, (x - 2)^2/2!, (x - 2)^3/3!\}$. In which basis the matrix is simpler? Then compute $M_{d^i,\alpha}$, $M_{d^i,\beta}$, $i = 2, 3, 4$, where $d^i := d \circ d^{i-1}$ (the composition of d with itself i times.

4. Identify the Euclidean plane $\mathbb{E}^2$ with $\mathbb{R}^2$ by introducing a Cartesian coordinate system in $\mathbb{E}^2$, and matching a point with coordinates (x_1, x_2) with vector (x_1, x_2) (same as $[x_1, x_2]$). We can depict the vector as a directed segment from the origin $(0, 0)$ to point (x_1, x_2).

   Write the matrix M_f for the following linear operators on $\mathbb{R}^2$ with respect to the standard basis.

   (a) $f = s_l$ – the symmetry with respect to a line l of $\mathbb{E}^2$, where l passes through the origin. Do it for l being x-axis; y-axis; line $l : y = mx$.

(b) $f = s_{l_1} \circ s_{l_2}$, where $l_1 : y = m_1 x$ and $l_2 : y = m_2 x$.

(c) f is a rotation r_θ around the origin counterclockwise by angle θ.

(d) f is a rotation $r_{\theta_1+\theta_2}$ around the origin counterclockwise by angle $\theta = \theta_1 + \theta_2$. Then notice that $r_{\theta_1+\theta_2} = r_{\theta_2} \circ r_{\theta_1}$. Compute now $M_{r_{\theta_1+\theta_2}}$ as a product of two matrices. Surprised?

(e) Prove that a symmetry s_l is never a rotation r_θ, i.e., that they are always different linear maps. (When you explain, remember that "things which look different are not always different".)

(f) Prove that any rotation r_θ is a composition of two symmetries s_{l_1} and s_{l_2} with respect to some lines l_1 and l_2.

(g) Is it true that a composition of a rotation r_θ and a symmetry s_l, is always a symmetry with respect to a line?

5. (a) Let $V = U \oplus W$, where V is a finite-dimensional space. Consider a map $f : V \to U$ defined as follows: write every $v \in V$ in a unique way as $v = u + w$, where $u \in U$ and $w \in W$, and define $f(v) = u$. Prove that f is a linear map and $f^2 := f \circ f = f$. Such a map f is called the **projection of V on U in the direction W**.

   Write the matrix M_f for this map with respect to the basis $\{u_1, \ldots, u_k, w_1, \ldots w_l\}$ of V, where $\{u_1, \ldots, u_k\}$ is a basis of U.

   (b) Let V be a finite-dimensional space, and let $f : V \to V$ be a linear map satisfying the property $f \circ f = f$. Such a map is called an **idempotent**. Prove that f is a projection of V on some subspace U in some direction W.

   (c) If we just consider a function $f : V \to V$ such that $f \circ f = f$, does f have to be linear?

   (d) Let f be a projection of $\mathbb{R}^3$ on a hyperplane $U = \{(x, y, z) : x - y - 2z = 0\}$ in the direction $W = \langle(1, 1, 1)\rangle$. Find the matrix M_f with respect to the standard basis of $\mathbb{R}^3$ and the basis $\{(1, 1, 0), (0, 2, -1)\}$ of U.

6. (i) Let A, $B \in M_{m\times n}(\mathbb{F})$. Prove that $(A + B)^T = A^T + B^T$.

   (ii) Let $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$. Prove that $(AB)^T = B^T A^T$.

7. Let A and B be square matrices of order n. Prove that if $AB = I_n$, then $BA = I_n$. This means that the equality $AB = I_n$ alone implies $B = A^{-1}$, i.e., the second condition in the definition of a nonsingular matrix can be dropped.

8. (Optional) Let f be a rotation of $\mathbb{R}^3$ by angle $\pi/2$ (or θ) with the axis $\langle(1, 2, 3)\rangle$ (or $\langle(a, b, c)\rangle$). Find M_f with respect to the standard basis.

Lecture 9.

Now we wish to introduce the notion of the determinant of a square matrix A. Roughly speaking the determinant of A is an element of the field constructed by using all n^2 entries of the matrix in a very special way.

It is denoted by $\det A$. Though $\det A$ does not capture all the properties of A, it does capture a few very important ones. Determinants has always played a major role in linear algebra. They appeared when people tried to solve systems of linear equations. Then they found use in analysis, differential equations, geometry. The notion of a volume of a parallelepiped in $\mathbb{R}^n$, defined by n linearly independent vectors, is introduced as the absolute value of the the determinant of the matrix which has these vectors as rows.

There are several conventional expositions of determinants, each having its own merits. The one we chose employs the notion of an exterior algebra. So far we have discussed two main algebraic structures, namely fields and vector spaces. We briefly mentioned non-commutative rings, which are like fields, but the multiplication is not commutative and the existence of some multiplicative inverses is not required. The main example of non-commutative ring in this course is $M_{n\times n}(\mathbb{F})$. Now we introduce the definition of an algebra. In this context 'algebra' is a specific technical term, not the whole field of mathematics known as algebra.

Let V be a vector space over $\mathbb{F}$, and let a function $\star : V \times V \to V$ be a **multiplication on** V, $(u, w) \mapsto u \star w =: uw$, which satisfies the following axioms:

- Distributive laws: $u(v + w) = uv + uw$, and $(v + w)u = vu + wu$ for all $u, v, w \in V$
- $(ku)v = u(kv) = k(uv)$ for all $u, v \in V$, and all $k \in \mathbb{F}$.

Then V with such a multiplication is called an **algebra over** $\mathbb{F}$. If the multiplication is commutative or associative, we obtain a **commutative** or an **associative algebra**, respectively. If there exists a vector $e \in V$ such that $ev = ve = v$ for all $v \in V$, then it is called the **identity element** of V, and the algebra is called an algebra with the **identity**. It is clear, that if the identity exists, it is unique. If for every non-zero $v \in V$, there exists a v^{-1} – the inverse of v with respect to the multiplication on V, then an algebra with the identity is called a **division** algebra.

It is easy to understand that if $\{v_1, \dots, v_n\}$ is a basis of V, then, by defining all products $v_i v_j = \sum_{k=1}^n a_{ij}^k v_k$, we actually define the products of every two vectors $u, w \in V$ due to distributive laws: if $u = \sum_{k=1}^n t_k v_k$ and $w = \sum_{k=1}^n s_k v_k$, then

$$uv = \sum_{k=1}^n t_k v_k \sum_{k=1}^n s_k v_k = \sum_{k=1}^n \big(\sum_{1\le i,j\le n} t_i s_j a_{ij}^k \big) v_k$$

This will give us an algebra. If we wish to attain some additional properties, like associativity, or commutativity, or the existence of the identity element, etc,we have to chose the n^3 constants a_{ij}^k very carefully, and often such a choice is not possible. Of course, one may take all products to be zero, but who needs this algebra?

Here are some examples of algebras.

- Let $\mathbb{F}$ be a subfield of a field $\mathbb{E}$. Then E is a commutative division algebra. Hence $\mathbb{R}$ is an infinite dimensional algebra over $\mathbb{Q}$, and $\mathbb{C}$ is a 2-dimensional algebra over $\mathbb{R}$ ($\{1, i\}$ is a basis).

- The vector space $\mathbb{F}[x]$ of all polynomials over $\mathbb{F}$ is an infinite-dimensional algebra over $\mathbb{F}$ with respect to the usual product of polynomials. It is commutative, associative, with the identity, but not a division algebra.

- A similar example is the algebra $\mathbb{F}[x_1, \ldots, x_n]$ of all polynomials over $\mathbb{F}$ with n variables.

- There exists a 4-dimensional associative division algebra over $\mathbb{R}$, called the algebra of quaternions, or the **Quaternion** algebra (or the algebra of Hamilton Quaternions). It has a basis $\{1, i, j, k\}$, with multiplication of the basis elements defined as

  $$i^2 = j^2 = k^2 = -1 \text{ ; and } ij = -ji = k, jk = -kj = i, ki = -ik = j,$$

  and continued to the whole algebra by the distributive laws. As we see, this algebra is not commutative. 1 (the field element viewed as a vector) is the identity element.

  There exists one more division algebra over reals. It is the Graves-Cayley algebra of octonions, which is 8-dimensional, non-commutative and non-associative.

- $M_{n\times n}(\mathbb{F})$ - the n^2-dimensional associative algebra of all $n \times n$ matrices over $\mathbb{F}$. It has the identity, but is not commutative, and is not a division algebra.

- $\mathbb{R}^3$ with the usual cross product of vectors is a non-commutative, non-associative 3-dimensional algebra over $\mathbb{R}$ without the identity.

Now we define the exterior algebra which will be used to develop the theory of the determinants of square matrices.

Consider $V = \mathbb{F}^n$, and let $\{e_1, \ldots, e_n\}$ be the standard basis of V. We define the following vector spaces over $\mathbb{F}$.

$\Lambda^0 = \mathbb{F}$ with basis $\{1\}$.

$\Lambda^1 = V = \mathbb{F}^n$ with basis $\{e_1, \ldots, e_n\}$.

Λ^2 with basis $\{e_i \wedge e_j : 1 \le i < j \le n\}$. The expression $e_i \wedge e_j$ is just a symbol, and we read it 'e_i wedge e_j'.

Similarly, let Λ^3 with basis $\{e_i \wedge e_j \wedge e_k : 1 \le i \ < j < k \le n\}$,

Λ^k with basis $\{e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_k} : 1 \le i_1 < i_2 < \ldots < i_k \le n\}$,

Λ^n with basis $\{e_1 \wedge e_2 \wedge \ldots \wedge e_n\}$.

Let Λ be defined as the direct product of all Λ^i:

$$\Lambda = \Lambda^0 \times \Lambda^1 \times \ldots \times \Lambda^n$$

Note that according to our definition $\Lambda \neq \Lambda^1$.

Similarly to complex numbers or polynomials, we may, and we will, think about elements of Λ as formal sums of elements from Λ^i, something like

$$2 + 3\,e_2 + e_3 - e_1 \wedge e_2 + \frac{2}{15}\,e_1 \wedge e_3 - \sqrt{2}\,e_1 \wedge e_2 \wedge e_3.$$

We will not write any term with coefficient 0, since it is equal to zero vector.

It is clear that $\dim \Lambda^k = \binom{n}{k}$, and $\dim \Lambda = \sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n$.

Now we introduce the multiplication on Λ which will distributes over vector addition and thus it will suffice to define multiplication on basis elements. Our algebra is also going to be associative.

(i) We postulate that the basis element 1 for $\Lambda^0 (= \mathbb{F})$ is to be the identity element for multiplication.

(ii) Next we define the product of two basis elements e_i and e_j from Λ^1 to be an element of Λ^2:

$$e_i e_j = e_i \wedge e_j = -e_j \wedge e_i \ \text{ for all } \ i, j, \text{ even for } \ i = j.$$

This implies that $e_i e_i = -e_i e_i$, hence $2e_i \wedge e_i = 0$. When we consider exterior algebras, we will consider only those fields where $2 \neq 0$, e.g., $\mathbb{F}_2$ is prohibited. This allows to conclude that $2e_i \wedge e_i = 0$ is equivalent to $e_i \wedge e_i = 0$.

(iii) The product of basis elements $e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_k} \in \Lambda^k$ and $e_{j_1} \wedge e_{j_2} \wedge \ldots \wedge e_{j_m} \in \Lambda^m$ is defined as follows:

(a) it is zero vector if $k + m > n$, or if $\{e_{i_1}, \ldots e_{i_k}\} \cap \{e_{j_1}, \ldots e_{j_m}\} \neq \emptyset$, and

(b) it is $\epsilon\, e_{h_1} \wedge e_{h_2} \wedge \ldots \wedge e_{h_{k+m}} \in \Lambda^{k+m}$, if $\{e_{i_1}, \ldots e_{i_k}\} \cap \{e_{j_1}, \ldots e_{j_m}\} = \emptyset$, where $h_1 < \ldots < h_{k+m}$ is the increasing ordering of $\{i_1, \ldots, i_k, j_1, \ldots, j_m\}$, and $\epsilon = 1$ or -1, depending on whether the ordering requires even or odd number of interchanges of the subsequent elements (i.e., the parity of the permutation $i_1 \ \ldots \ i_k j_1 \ldots j_m$).

Let us make several comments on this definition.

1. Part (iii) of the definition can also be rephrased as follows. The products of two vectors from the basis of Λ is defined as

$$(e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_k}) (e_{j_1} \wedge e_{j_2} \wedge \ldots \wedge e_{j_m}) = e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_k} \wedge e_{j_1} \wedge e_{j_2} \wedge \ldots \wedge e_{j_m},$$

with the assumption that the latter can be simplified, by using associativity: either to 0, if some index i_s is equal to some j_t or $k + m > n$, or, otherwise, to a basis element of Λ^{k+m} or its opposite.

2. Why should we use a special notation for the product? Why instead of uw we cannot just write $u \wedge w$, same wedge as in other symbols? It certainly agrees with the multiplication of the basis vectors of Λ. That what we will do. And we will refer to the multiplication in Λ as the **wedge product**.

3. It is not obvious, that the rule of determining ϵ in the part (iii) will agree with the associativity of the product of the basis elements, but it is possible to show that it does.

4. Since zero vector $0_V = 0_{\mathbb{F}} v$, then the rule $(ku)v = u(kv) = k(uv)$ for all $u, v \in V$, and all $k \in \mathbb{F}$, implies that if among the factors in a wedge product one vector is 0, then the whole product is 0.

Examples.

- $e_2 \wedge (e_1 \wedge e_3) = (e_2 \wedge e_1) \wedge e_3 = (-e_1 \wedge e_2) \wedge e_3 = -e_1 \wedge e_2 \wedge e_3$.
- $e_2 \wedge (e_1 \wedge e_2) = (e_2 \wedge e_1) \wedge e_2 = (-e_1 \wedge e_2) \wedge e_2 = -(e_1 \wedge e_2) \wedge e_2 = -e_1 \wedge (e_2 \wedge e_2) = -e_1 \wedge 0 = 0$.
- $(e_1 \wedge e_5) \wedge (e_2 \wedge e_3 \wedge e_4) = -e_1 \wedge e_2 \wedge e_3 \wedge e_4 \wedge e_5$.
- $(e_1 + e_2 \wedge e_3) \wedge (e_1 + e_2 \wedge e_3) = 2e_1 \wedge e_2 \wedge e_3 \neq 0$. This example illustrates that it is not true that $w^2 = 0$ for all $w \in \Lambda$.
- $(a_0 + a_1\, e_1 + a_3\, e_3 + a_{13}\, e_1 \wedge e_3 + a_{123}\, e_1 \wedge e_2 \wedge e_3) \wedge (b_2\, e_2 + b_{12}\, e_1 \wedge e_2 + b_{13}\, e_1 \wedge e_3) = a_0 b_2\, e_2 + (a_0 b_{12} + a_1 b_2)\, e_1 \wedge e_2 + a_0 b_{13}\, e_1 \wedge e_3 -$

  $a_3 b_2\, e_2 \wedge e_3 + (a_3 b_{12} - a_{13} b_2)\, e_1 \wedge e_2 \wedge e_3$

Lecture 10.

We are ready to proceed with the properties of the exterior algebra.

Proposition 16 *If $x, y \in \Lambda^1$ are linearly dependent, then $x \wedge y = 0$. For every $x \in \Lambda^1$, $x \wedge x = 0$.*

Proof. Let $x = \sum_{i=1}^n x_i e_i$ and $y = \sum_{i=1}^n y_i e_i$. Linear dependence of x and y implies that one of them is a scalar multiple of another. Suppose $x = ky$, for some $k \in \mathbb{F}$. Then $(x_1, \ldots, x_n) = (ky_1, \ldots, ky_n)$ and

$$x \wedge y = \big(\sum_{i=1}^n x_i e_i\big) \wedge \big(\sum_{i=1}^n y_i e_i\big) = \sum_{1 \le i < j \le n} (x_i y_j - x_j y_i)(e_i \wedge e_j) =$$
$$\sum_{1 \le i < j \le n} (k y_i y_j - k y_j y_i)(e_i \wedge e_j) = \sum_{1 \le i < j \le n} (0)(e_i \wedge e_j) = 0.$$

The second statement follows from the fact that x and x are linearly dependent. ∎

Proposition 17 *If $v_1, \ldots, v_m \in \Lambda^1$ are linearly dependent, then $v_1 \wedge \ldots \wedge v_m = 0$.*

Proof. Linear dependence of v_i implies that one of them is a linear combination of others. Renumber them such that $v_1 = a_2 v_2 + \ldots + a_m v_m$. Then

$$v_1 \wedge v_2 \wedge \ldots \wedge v_m = (a_2 v_2 + \ldots + a_m v_m) \wedge v_2 \wedge \ldots \wedge v_m =$$
$$a_2 v_2 \wedge v_2 \wedge \ldots \wedge v_n + \ldots + a_m v_m \wedge v_2 \wedge \ldots \wedge v_m = 0,$$

since when we distribute the product every term will contain some v_i twice among its wedge factors, and, by Proposition 16, each such term is zero. ∎

The contrapositive to the first statement of the Proposition 17 is as interesting: if $v_1 \wedge \ldots \wedge v_m \neq 0$, then $v_1, \ldots, v_m \in \Lambda^1$ are linearly independent.

Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{F}^n$, and let $1 \le i_1 < i_2 < \ldots < i_p \le n$. For $A \in M_{p \times p}(\mathbb{F})$, we define the "product of A and e_{i_k} with respect to $(e_{i_1}, e_{i_2}, \ldots, e_{i_p})$" as follows:

$$A\, e_{i_k} := a_{1k} e_{i_1} + a_{2k} e_{i_2} + \ldots + a_{pk} e_{i_p} = \sum_{t=1}^p a_{tk} e_{i_t}.$$

Hence $[a_{1k}, \ldots, a_{nk}]^T$ is the k-th column of A.

Next we define the "product of A and $e_{i_1} \wedge \ldots \wedge e_{i_p}$ with respect to $(e_{i_1}, e_{i_2}, \ldots, e_{i_p})$" as

$$A\,(e_{i_1} \wedge \ldots \wedge e_{i_p}) = A\, e_{i_1} \wedge \ldots \wedge e_{i_p} := A\, e_{i_1} \wedge \ldots \wedge A\, e_{i_p} = \bigwedge_{k=1}^p \sum_{t=1}^p a_{tk} e_{i_t}.$$

Then $A\,e_{i_1} \wedge \ldots \wedge e_{i_p}$ lies in 1-dimensional space $\langle\, e_{i_1} \wedge \ldots \wedge e_{i_p} \,\rangle$. Hence it is a scalar multiple of the basis vector $e_{i_1} \wedge \ldots \wedge e_{i_p}$. Denote the corresponding scalar by $\lambda = \lambda(A; (i_1, i_2, \ldots, i_p))$. It is clear that for every increasing sequence of p integers $1 \le i_1 < i_2 < \ldots < i_p \le n$, the value of this scalar is the same! It is called the **determinant of** A, and it is denoted by $\det A$. Thus

$$Ae_{i_1} \wedge \ldots \wedge e_{i_p} = A\,e_{i_1} \wedge \ldots \wedge A\,e_{i_p} = \prod_{k=1}^{p} \sum_{t=1}^{p} a_{tk} e_{i_t} =: (\det A) e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

If $p = n$, then the there exists only one increasing sequence of length n in $\{1, 2, \ldots, n\}$. As $A \in M_{n\times n}(\mathbb{F})$, the coordinate vector of $A\,e_i$ in the basis $(e_1, \ldots, e_n)$ represents the i-th column of A, which is $[a_{1i}, \ldots, a_{ni}]^T$. In this case $A\,e_i$ can be thought as the genuine product of two matrices: A and the the column ($n \times 1$ matrix) e_i. Again, $A\,e_1 \wedge \ldots \wedge A\,e_n$ lies in 1-dimensional space Λ^n. Hence it is a scalar multiple of the basis vector $e_1 \wedge \ldots \wedge e_n$. Hence

$$A\,e_1 \wedge \ldots \wedge A\,e_n = (\det A)\, e_1 \wedge \ldots \wedge e_n. \tag{13}$$

It is easy to check that if $n = 1$ and $A = (a)$, then $\det A = a$. If $n = 2$, and $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $A\,e_1 \wedge A\,e_2 = (a_{11}\,e_1 + a_{21}\,e_2) \wedge (a_{12}\,e_1 + a_{22}\,e_2) = (a_{11}a_{22} - a_{12}a_{21})\,e_1 \wedge e_2$, hence,

$$\det A = a_{11}a_{22} - a_{12}a_{21}.$$

Let $A \in M_{n\times n}(\mathbb{F})$. For any $x = x_1e_1 + \ldots + x_nv_n$, $Ax = x_1A\,e_1 + \ldots + x_nA\,e_n$. For every $v_1, \ldots v_p \in \Lambda$, we now define

$$A\,(v_1 \wedge \ldots \wedge v_p) := A\,v_1 \wedge \ldots \wedge A\,v_p.$$

The following statements are easy to prove.

Proposition 18 *1.* $\det I_n = 1$.

2. $\det(kA) = k^n \det A$ *for* $k \in \mathbb{F}$.

3. If columns of A *are linearly dependent, then* $\det A = 0$.

4. For all A, $B \in M_{n\times n}(\mathbb{F})$, $\det AB = \det A \det B$.

5. For each $A \in M_{n\times n}(\mathbb{F})$, *the inverse matrix* A^{-1} *exists if and only if* $\det A \neq 0$.

6. If $C^{-1}AC = B$, *then* $\det A = \det B$.

Proof. Done in class. ∎

Lecture 11.

Here we wish to prove other important properties of the determinants.

One of them is often called the Laplace expansion. It allows to compute the determinant of a matrix by computing (many!) determinants of smaller matrices.

Let $A = (a_{ij})$ be an $n \times n$ square matrix over $\mathbb{F}$, and let A_{ij} denote the $(n-1) \times (n-1)$ square matrix obtained from A by deleting the i-th row and the j-th column of A.

Proposition 19

1. *(Expansion by the j-th column) For every $j = 1, \dots, n$,*
$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}. \tag{14}$$
2. *(Expansion by the i-th row) For every $i = 1, \dots, n$,*
$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}. \tag{15}$$
3. *(Expansion by permutations) Let $\pi \in S_n$, where S_n is the set of all $n!$ bijections on $\{1, \dots, n\}$. Then*
$$\det A = \sum_{\pi \in S_n} \operatorname{sgn}(\pi)\, a_{1\pi(1)} \cdots a_{n\pi(n)}, \tag{16}$$
where $\operatorname{sgn}(\pi) = 1$ if π is an even permutation, and $\operatorname{sgn}(\pi) = -1$ if π is an odd permutation.

Proof. As you remember, $\bigwedge\limits_{1 \le i \le n} A\, e_i = \det A \bigwedge\limits_{1 \le i \le n} e_i$.

We begin with the first statement. First we establish the result for $j = 1$, i.e.,

$$\det A = \sum_{1 \le t \le n} (-1)^{t+1} a_{t1} \det A_{t1}.$$

We have:

$$\begin{aligned}
\bigwedge_{1 \le i \le n} A\, e_i &= \bigwedge_{1 \le i \le n} \sum_{1 \le t \le n} a_{ti} e_t = \left(\sum_{1 \le t \le n} a_{t1} e_t \right) \wedge \left(\bigwedge_{2 \le i \le n} A\, e_i \right) = \sum_{1 \le t \le n} (a_{t1} e_t) \wedge \left(\bigwedge_{2 \le i \le n} \sum_{1 \le k \le n} a_{ki} e_k \right) \\
&= \sum_{1 \le t \le n} (a_{t1} e_t) \wedge \left(\bigwedge_{2 \le i \le n} \sum_{\substack{1 \le k \le n \\ k \ne t}} a_{ki} e_k \right) = \sum_{1 \le t \le n} a_{t1} e_t \wedge \left(\det A_{t1} \bigwedge_{\substack{1 \le i \le n \\ i \ne t}} e_i \right) \\
&= \sum_{1 \le t \le n} a_{t1} \det A_{t1} \left(e_t \wedge \bigwedge_{\substack{1 \le i \le n \\ i \ne t}} e_i \right) = \sum_{1 \le t \le n} a_{t1} \det A_{t1} \left((-1)^{t-1} \bigwedge_{1 \le i \le n} e_i \right) = \\
&= \left(\sum_{1 \le t \le n} (-1)^{t-1} a_{t1} \det A_{t1} \right) \bigwedge_{1 \le i \le n} e_i.
\end{aligned}$$

Since $(-1)^{t-1} = (-1)^{t+1}$, then $\det A = \sum\limits_{1\le t\le n} (-1)^{t+1} a_{t1} \det A_{t1}$.

In order to get a similar result for the expansion with respect to the j-th column, we prove the following lemma. It states that if two columns of a square matrix are interchanged, then the determinant of the matrix changes its sign.

Lemma 20 *Let A be a square matrix of order p, $1 \le p \le n$. Let matrix A' be obtained from matrix A by interchanging any two columns of A. Then* $\det A = -\det A'$.

Proof. Let $1 < i_1 < \ldots < i_p < n$, and $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{F}^n$. Then

$$\bigwedge_{1\le k\le p} A\, e_{i_k} = \det A \bigwedge_{1\le k\le p} e_{i_k}.$$

What happens when we interchange two *adjacent* columns of A, say the j-th and the $(j+1)$-th? As

$$A\, e_{i_j} \wedge A\, e_{i_{j+1}} = \sum_{1\le k\le p} a_{ki_j} e_{i_k} \ \wedge \sum_{1\le t\le p} a_{ti_{j+1}} e_{i_t} = \sum_{1\le k\le p} \sum_{1\le t\le p} (a_{ki_j} e_{i_k}) \wedge (a_{ti_{j+1}} e_{i_t}) =$$

$$\sum_{1\le k\le p} \sum_{1\le t\le p} a_{ki_j} a_{ti_{j+1}} (e_{i_k} \wedge e_{i_t}) = \sum_{1\le k\le p} \sum_{1\le t\le p} a_{ti_{j+1}} a_{ki_j} (-e_{i_t} \wedge e_{i_k}) = -(A\, e_{i_{j+1}} \wedge A\, e_{i_j}),$$

the interchange of two adjacent columns of A leads to a change of sign of the determinant. If we wish to interchange the 1-st and the j-th columns of A, we can use $(j-1)$ adjacent column interchanges to place the j-th column first, and then $j-2$ adjacent column interchanges to place the 1-th column of A to be the j-th column of A'. Since the total number of interchanges of adjacent columns is an odd integer $2j-3$, the sign of the determinant will change odd number of times. This proves the lemma. ∎

Now we are ready to prove the formula for the expansion of $\det A$ with respect to the j-th column for the arbitrary j. Consider the matrix A'' obtained from A by subsequent interchanges of the j-th column with the first $j-1$ columns. In other words, the first column of A'' is the j-th column of A, the k-the column of A'' is the $(k-1)$-th column of A for $1 < k \le j$, and it is the k-th column of A for $j < k \le n$. As it takes $(j-1)$ interchanges, then $\det A = (-1)^{j-1} \det A''$ by Lemma 20. At the same time, $\det A_{ij} = \det A''_{i1}$ for all i. Expanding $\det A''$ with respect to the 1-st column we obtain:

$$\det A = (-1)^{j-1} \det A'' = \sum_{1\le t\le n} (-1)^{t+1} a_{tj} \det A''_{tj} = \sum_{1\le t\le n} (-1)^{t+j} a_{tj} \det A_{tj},$$

and this ends the proof of part 1.

The proof of the second statement is similar.

We now prove the third statement. Recall that a permutation is even if it can be constructed from the identity permutation by an even number of transpositions, and it is odd otherwise. Therefore

$$\bigwedge_{1\le i\le n} e_{\pi(i)} = \operatorname{sgn}(\pi) \bigwedge_{1\le i\le n} e_i \ .$$

But $\bigwedge_{1\le i\le n} A\, e_i = \bigwedge_{1\le i\le n} \sum_{1\le j\le n} a_{ji} e_j \ =$

$$\sum_{\pi\in S_n} \left[\left(\prod_{1\le i\le n} a_{\pi(i)i} \right) \bigwedge_{1\le i\le n} e_{\pi(i)} \right] = \left(\sum_{\pi\in S_n} \operatorname{sgn}(\pi) \prod_{1\le i\le n} a_{\pi(i)i} \right) \bigwedge_{1\le i\le n} e_i$$

But $\bigwedge_{1\le i\le n} A\, e_i = (\det A) \bigwedge_{1\le i\le n} e_i$ by definition of the determinant. This ends the proof. ∎

We just wish to mention that for large n, and general A, the computation of $\det A$ using the expansion by permutations takes long, and, hence, is not practical.

Problems.

In solutions of the following problems you can use the techniques based on exterior algebra, as well as all other properties of determinants that we proved.

1. Prove the second statement of Proposition 19.
2. Check the third statement of Proposition 19 for $n = 3$.
3. A square matrix $D = (d_{ij})$ of order n is called **diagonal** if $d_{ij} = 0$ for all $i \neq j$. Prove that $\det D = d_{11}d_{22}\cdots d_{nn}$.
4. A square matrix $A = (a_{ij})$ of order n is called **upper triangular** (**lower triangular**) if $a_{ij} = 0$ for all $i > j$ ($i < j$). Prove that for both upper and lower triangular matrix A, $\det A = a_{11}a_{22}\cdots a_{nn}$.
5. Let A be a square matrix of order n. Prove that $\det A = \det A^T$.
6. Let A and B be square matrices (possibly of different size). Show that the determinant of $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is $(\det A)(\det B)$. Generalize the statement.
7. Let A be a square matrix of order n, A_1 (A_2) be a matrix obtained from A by interchanging two rows (columns) of A, and A_3 (A_4) be a matrix obtained from A by replacing the i-th row (column) of A by the sum of this row (column) with the j-th row (column) multiplied by a scalar.

   Prove that
   $$\det A = -\det A_1 = -\det A_2 = \det A_3 = \det A_4.$$
   The property $\det A = \det A_3 = \det A_4$ is very useful for computing determinants: if $a_{ij} \neq 0$, then applying the transformation several time one can make all entries of the j-th column or the i-th row of A, except a_{ij} zeros.
8. (Vandermonde's determinant) Let $\{x_1, \ldots, x_n\}$ be n distinct elements from a field $\mathbb{F}$. Prove that
   $$\det \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ {x_1}^2 & {x_2}^2 & \ldots & {x_n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ {x_1}^{n-1} & {x_2}^{n-1} & \ldots & {x_n}^{n-1} \end{bmatrix} = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$
9. Let $a, b \in \mathbb{F}$ and $a \neq b$. Compute the following determinant:
   $$\det \begin{bmatrix} a & b & b & \ldots & b \\ b & a & b & \ldots & b \\ b & b & a & \ldots & b \\ \vdots & \vdots & \ddots & \vdots & \\ b & b & b & \ldots & a \end{bmatrix}$$
   (all diagonal entries of the matrix are a, and all other entries are b).

10. Recall that a square matrix A is called nonsingular, if A^{-1} exists, or equivalently, $\det A \neq 0$. Compute the number of all non–singular $n \times n$ matrices over the field $\mathbb{F}_p$ (p is a prime number).

11. For $i = 1, \ldots, n$, let $f_i : \mathbb{R} \to \mathbb{R}$ be a continuous function. Let $f : \mathbb{R} \to \mathbb{R}^n$ defined by $t \to (f_1(t), f_2(t), \ldots, f_n(t))$ be a continuous curve in $\mathbb{R}^n$. Find f (i.e., find n continuous functions f_i) such that every n distinct points on the curve do not lie on a hyperplane of $\mathbb{R}^n$.

Lecture 12.

We are going to present a kind of an explicit formula for the inverse of a nonsingular square matrix. As before, we denote by A_{ij} the matrix obtained from a square matrix A by deleting the i-th row and the j-th column. Let

$$b_{ij} = (-1)^{i+j} \det A_{ij}$$

be the (ij)**-cofactor** of A, and let $\operatorname{adj} A = (b_{ij})^T$. The matrix $\operatorname{adj} A$ is called the **classical adjoint of** A. By $GL(n, \mathbb{F})$ we will denote the set of all nonsingular $n \times n$ matrices. Since it forms a group under multiplication (i.e., the multiplication is an associative operation on $GL(n, \mathbb{F})$, there exists an identity element, and every element has an inverse with respect to multiplication), $GL(n, \mathbb{F})$ is called the **general linear group**.

Theorem 21 *Let A be a square matrix of order n over $\mathbb{F}$. Then $A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)I_n$.*

Proof. Our proof is based on two facts: (i) the Laplace expansion formula, and (ii) that a matrix with two equal columns has zero determinant.

Let $B = (b_{ij})$. Then $\operatorname{adj} A = B^T$. Let $C = (c_{ij}) = (\operatorname{adj} A)A$. Then, for all j, we have

$$c_{jj} = [b_{1j}, b_{2j}, \ldots, b_{nj}]\, [a_{1j}, a_{2j}, \ldots, a_{nj}]^T = \sum_{k=1}^{n} b_{kj} a_{kj} = \det A,$$

due to the Laplace expansion with respect to the j-th column of A.

For $i \neq j$, we get

$$c_{ij} = [b_{1i}, b_{2i}, \ldots, b_{ni}]\, [a_{1j}, a_{2j}, \ldots, a_{nj}]^T = \sum_{k=1}^{n} b_{ki} a_{kj}.$$

Consider a matrix $A' = (a'_{ij})$ with two equal columns: A' is obtained from A by replacing the i-th column of A by the j-th column of A. Note that $\det A'_{ki} = \det A_{ki}$ for all k, and that $\det A' = 0$ since columns of A' are linearly dependent. Expanding $\det A'$ with respect to the i-th column we obtain:

$$0 = \det A' = \sum_{1 \leq k \leq n} a'_{ki} (-1)^{k+i} \det A'_{ki} = \sum_{1 \leq k \leq n} a_{kj} (-1)^{k+i} \det A_{ki} = \sum_{1 \leq k \leq n} a_{kj} b_{ki}.$$

Therefore $c_{ij} = 0$ for $i \neq j$. This proves that $(\operatorname{adj} A)A = (\det A)I_n$.

Similarly, we can show that $A(\operatorname{adj} A) = (\det A)I_n$. ∎

The proof of the following corollary is obvious. It gives an explicit formula for the inverse of a nonsingular square matrix.

Corollary 22 *Let $A \in GL(n, \mathbb{F})$. Then $A^{-1} = (\frac{1}{\det A}) \operatorname{adj} A$.*

We just wish to mention that for large n, and general A, the computation of A^{-1} using $\operatorname{adj} A$ takes very long, and, hence, is not practical.

Theorem 21 and Corollary 22 allow us to prove Cramer's rule for solutions of a system of n linear equations with n unknowns:

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n &= b_2 \\ \ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= b_n. \end{aligned}$$

Let b_{ij} be the (ij)-cofactor of $A = (a_{ij})$. Fix some j, $1 \le j \le n$. Multiplying both sides of the i-th equation by b_{ij}, and then adding all the results, we obtain:

$$\sum_{1\le i\le n} \left(\sum_{1\le k\le n} a_{ik}b_{ij} \right) x_k = \sum_{1\le k\le n} \left(\sum_{1\le i\le n} a_{ik}b_{ij} \right) x_k = \sum_{1\le i\le n} b_i b_{ij}.$$

By Theorem 21, the inner sum $\sum_{1\le i\le n} a_{ik}b_{ij}$ is equal to 0 if $k \ne j$, and is $\det A$ for $k = j$. Hence we have

$$\det A\, x_j = \sum_{1\le i\le n} b_i b_{ij}.$$

Let A_j be the matrix obtained from A by replacing its j-th column by the column $[b_1, \ldots, b_n]^T$. Then $\sum_{1\le i\le n} b_i b_{ij} = \det A_j$, and we have

$$(\det A)\, x_j = \det A_j$$

for all j. This implies that if the system has a solution x, and $\det A \ne 0$, then $x = [\frac{detA_1}{\det A}, \ldots, \frac{detA_n}{\det A}]^T$. It also shows that if $\det A = 0$, but $\det A_j \ne 0$ for at least one value of j, then the system has no solutions.

For $\det A \ne 0$, we can check that $x_i = \frac{detA_i}{\det A}$, $i = 1, \ldots, n$, are indeed the solutions of the system by substituting them into an arbitrary equation of the system and simplifying the right hand side.

If $b = 0$, i.e., the system is homogeneous, then all $\det A_j = 0$. Therefore, if $\det A \ne 0$, the system will have only the trivial solution $x = 0$. If $b = 0$ and the system has a non-trivial solution, then $\det A = 0$. Hence we proved the following theorem.

Theorem 23 (Cramer's Rule) *Let $Ax = b$ be a system of n linear equations with n unknowns, $x = [x_1, \ldots, x_n]^T$, and $b = [b_1, \ldots, b_n]^T$. Let A_j be the matrix obtained from A by replacing its j-th column by the column b.*

If $\det A \neq 0$, the system has unique solution $x = [\frac{detA_1}{\det A}, \ldots, \frac{detA_n}{\det A}]^T$. If $\det A = 0$, but $\det A_j \neq 0$ for at least one value of j, then the system has no solutions.

If $b = 0$, i.e., the system is homogeneous, then, if $\det A \neq 0$, the system has only the trivial solution $x = 0$. Equivalently, if $b = 0$ and the system has a non-trivial solution, then $\det A = 0$.

As applications of the Cramer's rule and Vandermonde's determinant, we can prove the following fundamental facts about polynomials.

Theorem 24 *Let $x_1, \ldots, x_{n+1}$ be $n+1$ distinct elements of a field $\mathbb{F}$, and $y_1, \ldots, y_{n+1}$ be arbitrary $n+1$ elements of $\mathbb{F}$. Then there exists a unique polynomial f over $\mathbb{F}$ of degree at most n such that $f(x_i) = y_i$ for all i.*

Proof. Let $f = a_n x^n + \ldots + a_1 x + a_0$. Then $f(x_i) = y_i$, $i = 1, \ldots, n+1$, is a system of $n+1$ linear equations with $n+1$ unknowns a_i. As the matrix A of the coefficients of the system is the Vandermonde's matrix, its determinant $\det A = \prod_{1 \leq i < j \leq n+1} (x_i - x_j)$. Since all x_i are distinct, $\det A \neq 0$, and all a_i are determined uniquely by Cramer's rule. ∎

Corollary 25 *Let $f = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \in \mathbb{F}[x]$, $a_n \neq 0$, and let $x_1, \ldots, x_{n+1}$ be $n+1$ distinct elements of $\mathbb{F}$. If $f(x_i) = 0$ for all $i = 1, \ldots, n+1$, then f is zero polynomial, i.e., all its coefficients are zeros.*

Proof. If A is the matrix of the coefficients of the corresponding system, then by Cramer's rule $a_i = \det A_i / \det A = 0 / \det A = 0$, $i = 0, \ldots n$, as every matrix A_i contains a column of all zeros. ∎

The corollary above can be restated this way: no polynomial of degree n over a field can have more than n distinct roots. It turns out that if roots are not distinct, but are counted with their multiplicities, then it is still true that no polynomial of degree n over a field can have more than n roots. But a proof in this case must be different.

The following corollary just restates the uniqueness part of Theorem 24. It generalizes the facts that there exists a unique line passing through two points, and a unique parabola passing through any three points, etc..

Corollary 26 *Let $f, g \in \mathbb{F}[x]$ be two polynomials of degree at most n. If $f(x_i) = g(x_i)$ for $n+1$ distinct elements x_i of $\mathbb{F}$, then $f = g$.*

Proof. Consider $h = f - g$. Then degree of h is at most n and $h(x_i) = 0$ for all i. Applying Corollary 25, we obtain that h is zero polynomial. Hence $f = g$. ∎

Again, let $x_1, \ldots, x_{n+1}$ be $n+1$ distinct elements of a field $\mathbb{F}$, and $y_1, \ldots, y_{n+1}$ be arbitrary $n+1$ elements of $\mathbb{F}$. Consider the following $n+1$ polynomials, each of degree n:

$$f_i(x) = \frac{(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_{n+1})}{(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_{n+1})}, \quad i = 1, \ldots, n+1,$$

where the factor $x - x_i$ is missing in the numerator and the factor $x_i - x_i = 0$ is missing in the denominator.

Then it is obvious that $f_i(x_i) = 1$ for all i, and $f_i(x_j) = 0$ for all $i \neq j$. This implies that the polynomial

$$L(x) = y_1 f_1(x) + \ldots + y_{n+1} f_{n+1}(x) =$$

$$\sum_{i=1}^{n+1} y_i \frac{(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_{n+1})}{(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_{n+1})} \tag{17}$$

has the property that its degree is at most n and $L(x_i) = y_i$ for all i. We wish to note that the polynomial L is exactly the same polynomial as the polynomial f obtained in the proof of Theorem 24. Though it is not clear from the ways these polynomials are defined, it is the case, as we showed in the proof of the theorem, and then again in Corollary 26, that such a polynomial is unique. The form (17) is just a representation of the polynomial in the basis f_i of the vector space of all polynomials over $\mathbb{F}$ of degree at most n. This form is often referred to as the **Lagrange Interpolation Formula**. For another view of the Lagrange interpolation formula via Linear Algebra, the one which uses the notion of the dual basis, see the text by Hoffman and Kunze.

We conclude with the following amazing fact. The proof is immediate and is left to the reader.

Corollary 27 *Every function $f : \mathbb{F}_p \to \mathbb{F}_p$ can be represented by a polynomial of degree at most $p-1$.*

Problems.

1. Prove Corollary 27.
2. Find a polynomial over the reals of degree at most 3 whose graph passes through points $(0, 1), (1, 2), 2, 4)$, and $(3, 2)$.

Lecture 13.

Sketch of the lecture.

Reminding the class the notions of three basic elementary row operations. The row-echelon form of a matrix, and the reduced row-echelon form.

Suggested reading for the review: the text by Dummit and Foote, pages 424 – 431 as the collection of all main facts, and favorite undergraduate texts.

We assume that we know that any matrix can be transformed into a row-echelon form by using the elementary row operations.

The **row space (column space)** of a matrix $A \in M_{m\times n}(\mathbb{F})$ is the span of all its row vectors (column vectors). The dimensions of these spaces are called, respectively, the **row-rank** and the **column-rank** of the matrix. Is there any relation between these spaces? The row space is a subspace of $\mathbb{F}^m$, and the column space is the subspace of $\mathbb{F}^n$, so, in general, they are distinct spaces. These two spaces can be distinct even for $m = n$, as the following example shows:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

The row space of A is $\langle (1, 1)\rangle$, and its column space is $\langle (1, 0)\rangle$. The spaces are distinct, though both are 1-dimensional. Experimenting with several more examples, like

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 1 & 1 & 2 & 7 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 0 \\ 3 & 2 & 2 \\ 6 & 4 & 4 \\ 1 & 0 & 1 \end{bmatrix},$$

we notice that the dimension of the row space and column space of each of these matrices are equal. It turns out that this is always true: the row-rank of any matrix is equal to its column-rank. This will be proved in Theorem 30 below. The common value of the row-rank of A and its column-rank, is called the **rank** of A, and it is denoted by $\operatorname{rank} A$.

Lemma 28 *Both row-rank and column-rank are preserved by (or invariant under) the elementary row operations and by the elementary column operations.*

Proof. Let $A = (a_{pq})$ be an $m \times n$ matrix. The fact that an elementary row operation does not change the row space of A, and so its row-rank, is easy to demonstrate, and we leave it to the reader. We will prove a less obvious fact, namely that an elementary row operation preserves the column-rank.

Let A' be a matrix obtained from A by an elementary row operation. It is obvious that interchange of two rows or multiplication of a row by a nonzero constant does not change the column-rank: if columns with indices $i_1, \ldots, i_k$ form a basis of column space of A, the columns with the same indices will form a basis for the column space of A'. Suppose $A' = (a'_{pq})$ is obtained from A by replacing the ith row of A by its sum with the jth row of A multiplied by a nonzero constant c, $i \neq j$. Let $C_1, \ldots, C_m$ be the columns of A, and $C'_1, \ldots, C'_m$ be the ones of A'.

Suppose $\sum_{k=1}^m \lambda_k C_k = 0$. This is equivalent to $\sum_{k=1}^m \lambda_k a_{tk} = 0$ for all $t \in [m]$. Then

$$\sum_{k=1}^m \lambda_k a'_{ik} = \sum_{k=1}^m \lambda_k (a_{ik} + c a_{jk}) = \sum_{k=1}^m \lambda_k a_{ik} + c \sum_{k=1}^m \lambda_k a_{jk} = 0 + c0 = 0.$$

As all rows of A and A', except maybe the ith, are equal, we get that $\sum_{k=1}^m \lambda_k C_k = 0$ implies $\sum_{k=1}^m \lambda_k C'_k = 0$.

Suppose $\sum_{k=1}^m \lambda_k C'_k = 0$. This is equivalent to $\sum_{k=1}^m \lambda_k a'_{tk} = 0$ for all $t \in [m]$. For $t = j$, we get, we get $\sum_{k=1}^m \lambda_k a'_{jk} = \sum_{k=1}^m \lambda_k a_{jk} = 0$. For $t = i$, we get

$$\sum_{k=1}^m \lambda_k a'_{ik} = 0 \;\; \Leftrightarrow \;\; \sum_{k=1}^m \lambda_k (a_{ik} + c a_{jk}) = 0 \;\; \Leftrightarrow \;\; \sum_{k=1}^m \lambda_k a_{ik} + c \sum_{k=1}^m \lambda_k a_{jk} = 0.$$

As $\sum_{k=1}^m \lambda_k a_{jk} = 0$, we obtain $\sum_{k=1}^m \lambda_k a_{ik} = 0$. This implies that $\sum_{k=1}^m \lambda_k C_k = 0$. Hence

$$\sum_{k=1}^m \lambda_k C_k = 0 \quad \text{if and only if} \quad \sum_{k=1}^m \lambda_k C'_k = 0.$$

Think about it. Does not this immediately imply what we need, i.e., that the column-ranks of A and A' are equal? Of course!

Let us state and prove this result in slightly greater generality.

Lemma 29 *Given two sets of n vectors $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_n\}$ such that*

$$\lambda_1 v_1 + \ldots + \lambda_n v_n = 0 \quad \textit{if and only if} \quad \lambda_1 u_1 + \ldots + \lambda_n u_n = 0.$$

Then $\dim \langle v_1, \ldots, v_n \rangle = \dim \langle u_1, \ldots, u_n \rangle$.

Proof. Let $k = \dim \langle v_1, \ldots, v_n \rangle$. Then there exists $v_{i_1}, \ldots, v_{i_k}$ which form a basis of $\langle v_1, \ldots, v_n \rangle$. The condition of the lemma implies that $\{u_{i_1}, \ldots, u_{i_k}\}$ is a basis of $\langle u_1, \ldots, u_n \rangle$. Indeed, the linear independence of $u_{i_1}, \ldots, u_{i_k}$ is clear. If $v_j = \sum_{t=1}^k \beta_t v_{i_t}$ for some β_t, then $(-1)v_j + \sum_{t=1}^k \beta_t v_{i_t} = 0$. Hence, $(-1)u_j + \sum_{t=1}^k \beta_t u_{i_t} = 0$, which is equivalent to $u_j = \sum_{t=1}^k \beta_t u_{i_t}$. This proves that $\{u_{i_1}, \ldots, u_{i_k}\}$ spans $\langle u_1, \ldots u_n \rangle$, and so it is a basis of $\langle u_1, \ldots, u_n \rangle$. ∎

By Lemma 29, the column-ranks of A and A' are equal.

Similarly one can show that both the column-rank and the row-rank is invariant with respect to the elementary column operations. ∎

As the row-rank and the column-rank of a matrix in its row-echelon form are equal, we get the following

Theorem 30 *Row-rank and column rank of an arbitrary matrix in $M_{m\times n}(\mathbb{F})$ are equal.*

Here are several more important facts related to $\operatorname{rank} A$.

Let $A \in M_{m\times n}(\mathbb{F})$. Deleting some (or possibly none) rows or columns of A, we can obtain a **submatrix of** A.

Theorem 31 *Let $A \in M_{m\times n}(\mathbb{F})$. Then the following holds.*

1. $\operatorname{rank} A = \operatorname{rank} A^T$
2. $r = \operatorname{rank} A$ *if and only if there exists a square $r \times r$ submatrix of A which is nonsingular, but every larger square submatrix of A is singular.*
3. *The solution set of a homogeneous system of linear equations $Ax = 0$, $A \in M_{m\times n}(\mathbb{F})$, is a subspace of $\mathbb{F}^n$ of dimension $n - \operatorname{rank} A$. In particular, every homogeneous system of linear equations which has less equations than the number of unknowns has a nontrivial solution.*
4. $\operatorname{rank} A - \operatorname{rank} B \le \operatorname{rank}(A + B) \le \operatorname{rank} A + \operatorname{rank} B$
5. *Let $B \in M_{n\times p}(\mathbb{F})$. Then* $\operatorname{rank} AB \le \min\{\operatorname{rank} A, \operatorname{rank} B\}$. *If $m > n$, then AB is singular.*

Proof. Here are hints for proofs. The reader should supply all missing details.

1. Follows from Theorem 30

2. Straightforward.

3. One may try two different approaches. For the first, use the row-echelon form of the matrix. For the second approach, consider the map $\phi : \mathbb{R}^n \to \mathbb{R}^m$, given by $x \mapsto Ax$. Then **im** ϕ is the span of the column space of A, and $\ker \phi$ is the solution space of the system $Ax = 0$.

4. Follows easily from the earlier result $\dim(U + W) = \dim U + \dim W - \dim U \cap W$, where U and W are subspaces of a finite-dimensional space V.

5. The inequality follows from the observation that a column space of AB is a subspace of the column space of A; and the row space of AB is a subspace of the row space of B. The second statement follows from it. ∎

For the linear map $\phi : \mathbb{F}^n \to \mathbb{F}^m$, given by $x \mapsto Ax$, $\ker \phi$ is exactly the solution space of the homogeneous system $Ax = 0$. Therefore this space is often referred to as the **kernel of the matrix** A, and is denoted by $\ker A$. As **im** ϕ is the span of the set of columns of A, $\dim(\textbf{im } \phi) = \operatorname{rank} A$. Hence $\dim(\ker A) = n - \operatorname{rank} A$.

Problems.

1. Prove all parts of Theorem 31.

Lecture 14.

In this lecture we begin to study how the notions of the Euclidean geometry in dimensions 2 and 3 are generalized to higher dimensions and arbitrary vector spaces. Our exposition is close to the one in [2].

Let V be a vector space over a field $\mathbb{F}$. A **bilinear form** over V is a map $f : V \times V \to \mathbb{F}$ such that f is linear in each variable:

$$f(\alpha u + \beta v, w) = \alpha f(u, w) + \beta f(v, w),$$

and

$$f(w, \alpha u + \beta v) = \alpha f(w, u) + \beta f(w, v)$$

for all $\alpha, \beta \in \mathbb{F}$ and $u, v, w \in V$.

A bilinear form is called **symmetric** if $f(x, y) = f(y, x)$ for all $x, y \in V$.

EXAMPLES.

- $V = \mathbb{F}^n,\ f(x, y) = x_1y_1 + \ldots + x_ny_n$.
- $V = \mathbb{F}^3,\ f(x, y) = x_1y_1 + 4x_2y_2 - 5x_3y_3$.
- $V = \mathbb{F}^n$, B is a matrix of order n over $\mathbb{F}$, and $f(x, y) = x^TBy$. The properties of multiplication of matrices immediately imply that f is a bilinear form. We say that f is **associated with** B, or is **defined by** B. In two previous examples the matrix B was I_n and $\mathrm{diag}(1, 4, -5)$.

  If $n = 3$ and
  $$A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 0 & 5 \\ -3 & 5 & 4 \end{pmatrix},$$
  then $f(x, y) = x^TAy = (x_1, x_2, x_3)\, A\, (y_1, y_2, y_3)^T =$
  $$x_1y_1 + 2x_1y_2 - 3x_1y_3 + 2x_2y_1 + 5x_2y_3 - 3x_3y_1 + 5x_3y_2 + 4x_3y_3.$$
- $V = C[a, b],\ f(u, v) = \int_a^b u(t)v(t)\, dt$.
- $V = C(\mathbb{R})$. Let $K(s, t)$ be a continuous function of two variables s and t. We define $f(u, v) = \int_a^b \int_a^b K(s, t)u(s)v(t)\, ds\, dt$.

The relation between bilinear forms on a finite-dimensional vector space and matrices is described in the theorem below. We remind the reader that a matrix B is called symmetric if $B = B^T$.

Proposition 32 *For every bilinear form f over $\mathbb{F}^n$, there exists a square matrix B of order n over $\mathbb{F}$, such that $f(x,y) = x^T By$. Conversely, every square matrix B of order n over $\mathbb{F}$ defines a bilinear form via $(x,y) \mapsto x^T By$. If f is symmetric, then B is a symmetric matrix. If B is a symmetric matrix, then the form $x^T By$ is symmetric.*

Proof. If $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{F}^n$, then $B = (b_{ij})$, where $b_{ij} = f(e_i, e_j)$. Then, using bilinearity of f,

$$f(x,y) = f\big(\sum_{1\le i\le n} x_i e_i, \sum_{1\le j\le n} y_j e_j\big) = \sum_{1\le i,j\le n} x_i y_j f(e_i, e_j) = x^T By.$$

The converse follows from the properties of matrix multiplication. ∎

This argument shows that B depends on the choice of the basis. If $\alpha = \{v_1, \ldots, v_n\}$ is another basis, let $A = (a_{ij})$, where $a_{ij} = f(v_i, v_j)$. Then $f(x,y) = [x]_\alpha^T A\,[y]_\alpha$.

If f is a symmetric bilinear form over $\mathbb{F}$, we call the pair (V, f) an **inner product space**, with f the **inner product**.

When $V = \mathbb{F}^n$ and $B = I_n$, $f(x,y) = x^T By = x_1 y_1 + \ldots + x_n y_n$ is called the **standard inner product** over $\mathbb{F}^n$.

For the rest of this lecture, f will denote an arbitrary inner product over $V = \mathbb{F}^n$, and B will denote the associated symmetric matrix.

The vectors u and v are called **orthogonal** (or **perpendicular**) if their inner product is zero. This often is denoted by $u \perp v$. For a subset $S \subseteq V$, we define the **perpendicular space** of S as

$$S^\perp = \{v \in V : f(u,v) = 0 \text{ for all } u \in S\}.$$

Two subsets $S, T \subseteq V$ are called **perpendicular** (denoted $S \perp T$) if $u \perp v$ for all $u \in S$ and $v \in T$.

Proposition 33 *Let S and T be two subsets of V. Then the following holds.*

1. $\{0\}^\perp = V$.
2. $S \perp T \Leftrightarrow S \subseteq T^\perp \Leftrightarrow T \subseteq S^\perp$.
3. $S^\perp$ *is a subspace of* V *and* $S^\perp = (Span(S))^\perp$.
4. *If* $S \subseteq T \subseteq V$ *then* $T^\perp \le S^\perp \le V$.
5. $S \subseteq S^{\perp\perp}$.

Proof. Left as an exercise. ∎

A nonzero vector $v \in V$ is called **isotropic** if $v \perp v$, or, equivalently, $f(v, v) = 0$. A subspace $U \leq V$ is called **isotropic** if it contains an isotropic vector; otherwise it is called **anisotropic**. U is called **totally isotropic** if $U \perp U$, i.e., every pair of vectors of U is orthogonal. (Equivalently, $U \leq U^{\perp}$.)

The **radical** of a subspace U, denoted $\operatorname{rad} U$, is defined as

$$\operatorname{rad} U = U \cap U^{\perp}.$$

The subspace U is called **singular** if $\operatorname{rad} U \neq \langle 0 \rangle$; and **nonsingular** otherwise. We call the **inner product singular** or **nonsingular** according to whether or not the whole space V itself is singular. In other words, (V, f) is nonsingular if the only vector of V orthogonal to the whole V is zero vector. The terminology suggests that the notion of singularity in inner product spaces is related to singularity of matrices. Indeed, this is the case.

Theorem 34 *Let (V, f) be an inner product space, $\dim V = n$, B be a matrix associated with f, and $U \leq V$ be an arbitrary subspace of V. Then the following holds.*

(a) $\dim U + \dim U^{\perp} \geq n$.

(b) If V is nonsingular, then $\dim U + \dim U^{\perp} = n$.

(c) V is nonsingular if and only if the matrix B is nonsingular.

Proof. Let $\{u_1, \dots u_k\}$ be a basis of U. Then $x \in U^{\perp}$ if and only if $x \perp u_i$ for all i, or, equivalently, x is a solution of the homogeneous system of k linear equations

$$u_i^T B x = 0 \quad (i = 1, \dots, k). \tag{18}$$

As the rank of this system is at most k, therefore its solution space $U^{\perp}$ has dimension at least $n - k$ (Theorem 31). This proves (a).

Suppose B is nonsingular. Then all vectors $u_i^T B$ are linearly independent. Indeed, let

$$\sum_{1 \leq i \leq k} \lambda_i u_i^T B = 0.$$

Then $(\sum_{1 \leq i \leq k} \lambda_i u_i^T) B = 0$. Let $u^T := \sum_{1 \leq i \leq k} \lambda_i u_i^T$. Then $u^T B = 0$, which is equivalent to $B^T u = 0$. As B is nonsingular, then so is B^T, and the only solution of $B^T u = 0$ is $u = 0$. Since $\{u_1, \dots u_k\}$ is a basis of U, we have $\lambda_i = 0$ for all i. Hence all vectors $u_i^T B$ are linearly independent. Note that these vectors $u_i^T B$ are the rows of the matrix of the coefficients of the

system (18). Therefore the rank of the matrix is exactly k, and the solution space of (18) has dimension $n-k$ by Theorem 31. This proves (b).

By part (b), if B is nonsingular, then $\dim(\operatorname{rad} V) = \dim(V^\perp) = n - \dim V = n - n = 0$. Hence, V is nonsingular. This proves one implication in (c).

If B is singular, then $\operatorname{rank} B < n$, and the system $By = 0$ has a nontrivial solution. Call it v, $v \neq 0$. So $Bv = 0$. Then for every $w \in V$, $w^T B\, v = w^T (B\, v) = w^T 0 = 0$. Hence $v \neq 0$ and v is orthogonal to the whole space V. Then $\operatorname{rad} V \neq \langle 0 \rangle$, and the inner space is singular. This proves the second implication in part (c). ∎

Corollary 35 *In a nonsingular inner product space of dimension n, every totally isotropic subspace has dimension at most $\lfloor n/2 \rfloor$.*

Proof. Since $U \leq U^\perp$, $\dim U \leq \dim U^\perp$. As $\dim U + \dim U^\perp = n$, we have $2 \dim U \leq \dim U + \dim U^\perp \leq n$. Hence $\dim U \leq n/2$. ∎

Problems.

1. Let (V, f) be an inner product space, and let u, w be two isotropic vectors in V. Prove that if $u \perp w$, then $Span(\{u, w\})$ is a totally isotropic subspace.

2. For each of the following inner product spaces determine whether it is singular, and if it is find a nonzero vector orthogonal to the whole space. Find isotropic vectors or show that they do not exist. Find a maximal totally isotropic subspace (if isotropic vectors exist).

   (a) $V = \mathbb{R}^2$, $f(x, y) = x_1 y_1$.

   (b) $V = \mathbb{R}^2$, $f(x, y) = x_1 y_2 + x_2 y_1$.

   (c) $V = \mathbb{R}^2$, $f(x, y) = x_1 y_1 - x_2 y_2$.

   (d) $V = C[a, b]$, $f(u, v) = \int_a^b u(t)v(t)\, dt$.

   (e) $V = \mathbb{C}^2$ with the standard inner product.

   (f) $V = \mathbb{R}^4$, $f(x, y) = x^T B y$, where

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

   (g) $V = \mathbb{F}_p^2$ with the standard inner product, for $p = 5$, and for $p = 7$.

   (h) $V = \mathbb{F}_2^5$ with the standard inner product.

   (i) $V = \mathbb{R}^4$ with the inner product $f(x, y) = x_1 y_1 + x_2 y_2 + x_3 y_3 - x_4 y_4$. This is the famous Minkowski space used in the theory of special relativity. Here (x_1, x_2, x_3) correspond to coordinate of an event in $\mathbb{R}^3$, and x_4 to its time coordinate.

3. Let U be a subspace in an inner product space over $\mathbb{F}$, where $\mathbb{F}$ is any field we have been using in this course except $\mathbb{F}_2$. Prove that if every vector of U is isotropic, then U is totally isotropic.

   Show that the statement is not necessarily correct for $\mathbb{F} = \mathbb{F}_2$.

4. Let V be an anisotropic inner product space. Let $\{v_1, \ldots, v_n\}$ be a set of nonzero vectors in V such that $v_i \perp v_j$ for all $i \neq j$. Prove that $\{v_1, \ldots, v_n\}$ are linearly independent.

5. Prove that the set of vectors $\{\cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\}$ is linearly independent in $V = C[0, 2\pi]$.

   (*Hint*: Consider the inner product $f(u, v) = \int_a^b u(t)v(t)\, dt$, and use the previous exercise.)

6. Prove Proposition 33.

7. Let f be the standard inner product on $\mathbb{F}^3$. Let $\alpha = \{[2, 1, -1], [1, 0, 1], [3, 1, 1]\}$ be another basis of $\mathbb{F}^n$. Find a matrix A such that

   (i) $A = (a_{ij}) = (f(v_i, v_j))$.

   (ii) $f(x, y) = [x]_\alpha^T A\, [y]_\alpha$.

8. (Optional) Prove that $\mathbb{F}_p^3$ with the standard inner product is isotropic for every prime p.

Lecture 15.

The goal of this lecture is to discuss the Gram-Schmidt orthogonalization process. Besides being a very useful tool, it will lead us to some striking conclusions.

Let (V, f) be an inner product space. A set S of vectors of (V, f) is called **orthogonal** if every two distinct vectors of S are orthogonal.

Proposition 36 *Let (V, f) be an n-dimensional inner product space, and let S be an orthogonal set of vectors, such that no vector of S is isotropic. Then S is a linearly independent set, and $|S| \leq n$.*

Proof. Consider some m vectors in S, denote them $v_1, \ldots, v_m$. If $\lambda_1 v_1 + \ldots + \lambda_m v_m = 0$, then, for each $i = 1, \ldots, m$,

$$0 = f(v_i, 0) = f(v_i, \lambda_1 v_1 + \ldots + \lambda_m v_m) = \lambda_1 f(v_i, v_1) + \ldots + \lambda_m f(v_i, v_m) = \lambda_i f(v_i, v_i).$$

So $\lambda_i f(v_i, v_i) = 0$ for all i. Since $f(v_i, v_i) \neq 0$ (S has no isotropic vectors), $\lambda_i = 0$ for all i. Hence S is linearly independent. Since $\dim V = n$, then every $n + 1$ vectors of V are linearly independent. So $|S| \leq n$. ∎

We would like to mention two applications of this simple result.

1. *Prove that the set $S = \{\cos x, \ldots, \cos nx, \sin x, \ldots, \sin nx\}$ of functions in $V = C[0, 2\pi]$ is linearly independent (over $\mathbb{R}$).*

Solution. Indeed, let $f(u, v) := \int_a^b u(t)v(t)\,dt$, where $u, v \in V$. Then f an inner product on V. Consider the inner space (V, f). As

$$f(\cos kt,\, \cos kt) = \int_0^{2\pi} \cos^2 kt\,dt = \int_0^{2\pi} \sin^2 kt\,dt = f(\sin kt,\, \sin kt) > 0,$$

for all nonzero integers k, and

$$f(\cos kt,\, \cos mt) = \int_0^{2\pi} \cos kt\, \cos mt\,dt = \int_0^{2\pi} \sin kt\, \sin mt\,dt = f(\sin kt,\, \sin mt) = 0$$

for all nonzero integers k, m, $k \neq m$, S is an orthogonal set of nonisotropic vectors. Hence it is linearly independent. ∎

2. *People in a city of* 100 *residents like to form clubs. The only restrictions on these clubs are the following:*

(i) each club must have an odd number of people;

(ii) every two clubs must share an even number of people.

What is the greatest number of clubs that they can form?

Solution. Let $\{p_1, \ldots, p_{100}\}$ be the set of all people in the city, and let $C_1, \ldots, C_m$ denote a set of clubs. For each club C_i consider a vector $v_i \in \mathbb{F}_2^{100}$, such that the k-th component of v_i is 1 if $p_k \in C_i$, and is 0 otherwise. Consider the standard inner product f on $\mathbb{F}_2^{100}$. As each $|C_i|$ is an odd integer, each $v_i = (v_{i1}, \ldots, v_{i100})$ contains an odd number of components equal to 1. Hence

$$f(v_i, v_i) = \sum_{k=1}^{100} v_{ik} v_{ik} = 1 + \ldots + 1 \ \ (|C_i| \text{ addends}) = 1.$$

As each $|C_i \cap C_j|$ is an even integer, the vectors v_i and v_j share 1's in even number of same components. Hence

$$f(v_i, v_j) = \sum_{k=1}^{100} v_{ik} v_{jk} = 1 + \ldots + 1 \ \ (|C_i \cap C_j| \text{ addends}) = 0.$$

Therefore the set of all v_i is an orthogonal set in $(\mathbb{F}_2^{100}, f)$, and it is linearly independent by Proposition 36. As we have m vectors, $m \leq 100$. Of course, it is possible to have 100 clubs: take, e.g., $C_i = \{p_i\}$ for $i = 1, \ldots, 100$, among many other constructions. ∎

OK, let's get serious. Everyone likes orthogonal bases. A good news is that often we can build one. A popular technology is called the **Gram-Schmidt Orthogonalization** process. And it is free! And Jorgen P. Gram (1850-1916) and Erhard Schmidt (1876-1959) are two different people. And the method seems to be known to Laplace (1749-1827), and used by Cauchy in 1846...

Theorem 37 (Gram-Schmidt Orthogonalization) *Let (V, f) be an anisotropic n-dimensional inner product space. Then V has an orthogonal basis.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis of V. Set $v_1' = v_1$, and try to find $v_2' \in Span(\{v_1, v_2\}$ such that $v_1' \perp v_2'$ and $Span(\{v_1', v_2'\}) = Span(\{v_1, v_2\})$.

In order to do this, search for v_2' in the form $v_2' = v_2 + x v_1'$, where the scalar x is unknown. Then

$$v_1' \perp v_2' \ \Leftrightarrow \ f(v_1', v_2') = 0 \ \Leftrightarrow \ f(v_1', v_2 + x v_1') = 0 \ \Leftrightarrow$$
$$f(v_1', v_2) + x f(v_1', v_1') = 0 \ \Leftrightarrow \ x = -\frac{f(v_1', v_2)}{f(v_1', v_1')}.$$

Note that the denominator is not zero, since $v_1' = v_1 \neq 0$ and (V, f) is anisotropic. Hence

$$v_2' = v_2 - \frac{f(v_1', v_2)}{f(v_1', v_1')}\, v_1'.$$

As $v_1' \perp v_2'$ and, as none of them is zero vector (why?), and as (V, f) is anisotropic, vectors v_1', v_2' are linearly independent by Proposition 36. As they are in $Span(\{v_1, v_2\})$, we get $Span(\{v_1', v_2'\}) = Span(\{v_1, v_2\})$.

If $n > 2$, we search for v_3' in the form $v_3' = v_3 + xv_1' + yv_2'$. We wish to have $v_1' \perp v_3'$ and $v_2' \perp v_3'$. Taking the inner product of v_1' with v_3', and of and v_2' with v_3', we obtain

$$x = -\frac{f(v_1', v_3)}{f(v_1', v_1')} \quad \text{and} \quad y = -\frac{f(v_2', v_3)}{f(v_2', v_2')}.$$

Hence

$$v_3' = v_3 - \frac{f(v_1', v_3)}{f(v_1', v_1')}\, v_1' - \frac{f(v_2', v_3)}{f(v_2', v_2')}\, v_2'.$$

Clearly v_1', v_2', v_3' are pairwise orthogonal, and each vector is nonzero. As (V, f) is anisotropic, vectors v_1', v_2', v_3' are linearly independent by Proposition 36. As $Span(\{v_1', v_2', v_3'\}) \leq Span(\{v_1, v_2, v_3\})$, we obtain $Span(\{v_1', v_2', v_3'\}) = Span(\{v_1, v_2, v_3\})$. Continue by induction, if needed. ∎

We call u a **unit vector** if $f(u, u) = 1\, (= 1_{\mathbb{F}})$. If every element of $\mathbb{F}$ is a square in $\mathbb{F}$ (not true in many fields), then every nonzero nonisotropic vector x in (V, f) is collinear to a unit vector. Indeed, we have $f(x, x) = a^2$ for some $a \in \mathbb{F}$, $a \neq 0$. Then for $x' := \frac{1}{a}x$, we have

$$f(x',\, x') = f(\frac{1}{a}\, x,\, \frac{1}{a}\, x) = \frac{1}{a}\frac{1}{a}\, f(x, x) = \frac{1}{a^2}\, a^2 = 1.$$

For the field of real numbers $\mathbb{R}$, every nonnegative number is a square. Therefore if f has the property that $f(x, x) > 0$ for all $x \neq 0$ (f is called **positive definite** in this case), then each vector has 'length' $\sqrt{f(x, x)}$, which is usually called the **norm** of x in (V, f), and which is denoted by $\| x \|$. If $x \neq 0$, then $\frac{1}{\|x\|}\, x$ is a unit vector. We say that a set of vectors in an inner product space is **orthonormal** if it is orthogonal and all vectors in the set are unit vectors.

Corollary 38 *Let (V, f) be an n-dimensional inner product space over $\mathbb{R}$, where f is positive definite. Then V has an orthonormal basis.*

We are ready to pass to the 'striking conclusions' promised at the beginning of the lecture.

Problems.

1. Find an orthogonal basis in the following spaces (V, f). Make the found basis orthonormal, if possible.

   (a) $V = Span(\{(1,1,3),(0,1,1)\}) \subseteq \mathbb{R}^3$, f is the standard inner product.

   (b) $V = Span(\{(1,1,3,-1,0),(0,1,1,-2,0),(1,2,3,4,5),(5,6,7,8,6)\}) \subseteq \mathbb{R}^5$, f is the standard inner product. (Use computer!)

   (c) $V = Span(\{(1,1,3),(0,1,1)\}) \subseteq \mathbb{F}_5^3$, f is the standard inner product.

   (d) $V = P_3 \subset C(R)$, where P_3 is the set of all polynomials of degree at most three with real coefficients. Let $f(u,v) = \int_{-1}^{1} u(t)v(t)\,dt$ define the inner product on V.

2. Let $P_0(x) = 1$, and $P_n(x) = \frac{d^n}{d\,x^n}\,(x^2-1)^n$ for $n \geq 1$. Prove that $\{P_n(x) : n \geq 0\}$ form an orthogonal basis in the space of all polynomial functions with real coefficients with inner product given by $f(u,v) = \int_{-1}^{1} u(t)v(t)\,dt$. These polynomials are called **Legendre polynomials**.

3. Consider the $V = Span(\mathcal{F}_n) \subset C([0, 2\pi])$, where

$$\mathcal{F}_n = \{1, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\}.$$

   Elements of V are called **Fourier polynomials of order at most** n. Consider the inner product $\int_0^{2\pi} u(t)v(t)\,dt$ on V. Check that $\mathcal{F}_n$ is an orthogonal basis of V. Turn it into an orthonormal basis of V. Let

$$f = a_0/2 + \sum_{k\geq 1}^{n} (a_k \cos kx + b_k \sin kx).$$

   Express the coefficients a_i and b_i (**Fourier coefficients of** f) as inner products of f and vectors from the orthonormal basis.

4. People in a city of 100 residents like to form clubs. The only restrictions on these clubs are the following:

   (i) each club must have an even number of people;

   (ii) every two clubs must share an even number of people;

   (iii) all clubs are distinct (as subsets of people).

   Prove that the greatest number of clubs that they can form is 2^{50}. (Note the surprising difference with the example discussed in the lecture.)

5. People in a city of 100 residents like to form clubs. The only restrictions on these clubs are the following:

   (i) each club must have exactly 20 people;

   (ii) every two clubs must share exactly 11 people.

(i) Prove that they cannot form more than 100 clubs.

Hint: Let $\{p_1, \ldots, p_{100}\}$ be the set of all people in the city, and let $C_1, \ldots, C_m$ denote a set of clubs. For each club C_i consider a vector $v_i \in \mathbb{R}^{100}$, such that the k-th component of v_i is 1 if $p_k \in C_i$, and is 0 otherwise. Consider the $m \times 100$ matrix A, whose rows are the vectors v_i. Prove that AA^T is nonsingular.

(ii) (Optional) What is the greatest number of clubs that they can form?

6. (Optional) Solve the previous problem where the seemingly strong condition (i) is replaced by a trivial conditions that all clubs are distinct (as subsets of people).

Lecture 16.

Thought there are many advantages in considering general inner product spaces, if one wants to generalize usual Euclidean geometry of 2- and 3-dimensional spaces to higher dimensions, one considers an inner product space (V, f) over real numbers with f being a positive definite form. We will call these spaces **Euclidean spaces**. If $V = \mathbb{R}^n$ and f is the standard inner product, we call (V, f) the **standard** n-dimensional Euclidean space. There are at least two ways of proceeding with such generalizations. One way is to study arbitrary Euclidean spaces axiomatically, i.e., based on the definition of a symmetric positive definite bilinear form. Another way is to explain that all n-dimensional Euclidean spaces are in a certain sense the same, and to work with just one of them, your favorite. We will discuss both these approaches.

Theorem 39 *Let (V, f) be a Euclidean space, and $\|x\| = f(x, x)$ be the norm.*

1. (The Pythagorean Theorem.) *For any mutually orthogonal vectors $v_1, \dots, v_k$, $k \geq 2$,*

$$\|v_1 + \dots + v_k\|^2 = \|v_1\|^2 + \dots + \|v_k\|^2.$$

2. (The Cauchy-Schwarz Inequality.) *For every two vectors x and y,*

$$|f(x, y)| \leq \|x\|\|y\|.$$

   The equality is attained if and only if x and y are colinear.

3. (The Triangle Inequality.) *For any vectors $v_1, \dots, v_k$, $k \geq 2$,*

$$\|v_1 + \dots + v_k\| \leq \|v_1\| + \dots + \|v_k\|.$$

   The equality is attained if and only if all vectors are colinear and of the 'same direction', i.e., there exists i such that for all j, $v_j = k_j v_i$ and $k_j \geq 0$.

Proof. 1. For the first statement, using $f(v_i, v_j) = 0$ for $i \neq j$, we obtain

$$\Big\| \sum_{1 \leq i \leq k} v_i \Big\|^2 = f\Big(\sum_{1 \leq i \leq k} v_i, \sum_{1 \leq i \leq k} v_i \Big) = \sum_{1 \leq i,j \leq k} f(v_i, v_j) = \sum_{1 \leq i \leq k} f(v_i, v_i) = \sum_{1 \leq i \leq k} \|v_i\|^2.$$

2. For $y = 0$, the statement is obvious. Let $y \neq 0$. As f is positive definite, we have

$$g(t) := \|x - ty\|^2 = f(x - ty, x - ty) = \|y\|^2 t^2 - 2f(x, y)\, t + \|x\|^2 \geq 0 \text{ for all } t \in \mathbb{R}.$$

Since $\|y\|^2 > 0$, $g(t)$ is a quadratic function on $\mathbb{R}$ which takes only nonnegative values. Hence its discriminant D is nonpositive, i.e.,

$$D = (-2f(x, y))^2 - 4\,\|y\|^2\|x\|^2 \leq 0 \;\Leftrightarrow\; |f(x, y)| \leq \|x\|\|y\|.$$

The equality sign in the last inequality is attained if and only if

$$D = 0 \;\Leftrightarrow\; \exists t_0\; g(t_0) = 0 \;\Leftrightarrow\; x = t_0 y.$$

As y is nonzero, the last condition is equivalent to colinearity of x and y.

3. For $k = 2$, we have

$$\|v_1 + v_2\|^2 = \|v_1\|^2 + 2f(v_1, v_2) + \|v_2\|^2 \leq \quad \text{(by Cauchy-Schwarz)}$$

$$\|v_1\|^2 + 2\|v_1\|\|v_2\| + \|v_2\|^2 = (\|v_1\| + \|v_2\|)^2.$$

Taking square roots, we obtain the result. The equality happens if and only if $f(v_1, v_2) = \|v_1\|\|v_2\|$. Hence the vectors are colinear (again by Cauchy-Schwarz). If $v_1 = tv_2$, we have $f(v_1, v_2) = f(tv_2, v_2) = tf(v_2, v_2) = t\|v_2\|^2$. At the same time, $\|v_1\|\|v_2\| = \|tv_2\|\|v_2\| = |t|\|v_2\|^2$. Hence $t \geq 0$. If $k > 2$, one proceeds by a straightforward induction. ∎

For the standard n-dimensional Euclidean space, the inequalities can be rewritten as

$$\Big|\sum_{1\leq i\leq n} a_i b_i\Big| \leq \Big(\sum_{1\leq i\leq n} a_i^2\Big)^{1/2}\Big(\sum_{1\leq i\leq n} b_i^2\Big)^{1/2},$$

and for $f(u, v) = \int_a^b u(t)v(t)\,dt$ on $C([a, b])$,

$$\Big|\int_a^b u(t)v(t)\,dt\Big| \leq \Big[\int_a^b u(t)^2\,dt\Big]^{1/2} \cdot \Big[\int_a^b v(t)^2\,dt\Big]^{1/2}.$$

The **distance between two vectors** x and y in a Euclidean space (V, f) is defined as $\|x - y\|$. This definition is, of course, motivated by the distance between two points. Note that the notion of a *point* Euclidean space has not been defined. Intuitively, we think about the vectors as directed segments which have initial points at the origin, and about points as the endpoints of these vectors, i.e., like in dimensions 2 and 3.

Now we turn to the second approach. We call two inner product spaces (V, f) and (V', f') **isometric** (or isomorphic), if there exists an isomorphism $\phi : V \to V'$, $x \mapsto x'$, such that $f(x, y) = f'(x', y')$ for every $x, y \in V$. Such inner product preserving isomorphism is called an **isometry**. When $V = V'$ and $f = f'$, the isometries are often called **orthogonal** maps. It is clear that an isometry also preserves norms associated with f and f'. The following theorem may, at first, look very surprising.

Theorem 40 *Every two n-dimensional Euclidean spaces are isometric. In particular, all such spaces are isometric to the standard Euclidean space.*

Proof. Let (V, f) and (V', f') be two n-dimensional Euclidean spaces. Then, by Corollary 36, each space has an orthonormal basis. Let $\{v_1, \ldots, v_n\}$ and $\{v'_1, \ldots, v'_n\}$ be such bases. Consider the map $\phi : V \to V'$ such that $\phi(v_i) = v'_i$ for all i, and continue it by linearity. In other words, let $\phi(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 v'_1 + \ldots + \lambda_n v'_n$ for all $\lambda_i \in \mathbb{R}$. Then $f(v_i, v_j) = f'(v'_i, v'_j) = 0$ for $i \neq j$ and 1 for $i = j$. Every two vectors $x, y \in V$, can be written as $x = \sum_{1 \leq i \leq n} x_i v_i$ and $y = \sum_{1 \leq i \leq n} y_i v_i$, where all $x_i, y_j \in \mathbb{R}$. Therefore

$$f(x, y) = f\big(\sum_{1 \leq i \leq n} x_i v_i\,,\ \sum_{1 \leq i \leq n} y_i v_i\big) = \sum_{1 \leq i \leq n} x_i y_i = f'\big(\sum_{1 \leq i \leq n} x_i v'_i\,,\ \sum_{1 \leq i \leq n} y_i v'_i\big) = f'(x', y'),$$

and ϕ is an isometry. ∎

Hence, there exists essentially only one n-dimensional Euclidean geometry (up to isometries). The theorem implies that every 'geometric' assertion (i.e., an assertion stated in terms of addition, inner product and multiplication of vectors by scalars) pertaining to vectors in the standard n-dimensional Euclidean space, is also true in all other n-dimensional Euclidean spaces. For $n = 2, 3$, it allows to claim certain geometric facts without a proof, as long as we know that they are correct in the usual Euclidean geometry. In particular, the inequality

$$\big[\int_0^{2\pi} (u + v)^2\, dt\,\big]^{1/2} \leq \big[\int_0^{2\pi} u^2\, dt\big]^{1/2} + \big[\int_0^{2\pi} v^2\, dt\big]^{1/2}$$

holds because the triangle inequality holds in usual plane geometry. No proof is necessary!

By now the reader understands that we use the word **"geometry"** quite freely in our discussions. Here we wish to add several other informal comments.

What we mean by saying "geometry" depends on the content. We certainly used the term much earlier, when we discussed just vector spaces. There we also had a way to identify different spaces by using isomorphisms (just linear bijections). Those mappings preserved the only essential features of vector spaces: addition and multiplication by scalars. This gave us geometry of vector spaces, or just linear geometry.

Now, in addition to linearity, we wanted to preserve more, namely the inner product of vectors. This is what the isometries do. This leads us to the geometries of inner product spaces.

So a geometry is defined by a set of objects and a set of transformations which preserve certain relations between the objects, or some functions defined on the sets of objects. We will return to this discussion later in the course, where new examples of geometries will be considered.

Sometimes, instead of a bilinear form f in the description of (V, f) which intuitively represents somehow both the lengths and angles, we can begin with a quadratic form Q, which correspond to the notion of length only.

Let $\mathbb{F}$ be any field we used in this course but $\mathbb{F}_2$. Let $Q : V \to \mathbb{F}$ be a function with the following properties:

(i) $Q(kx) = k^2 Q(x)$ for all $k \in \mathbb{F}$ and $x \in V$; and

(ii) the function $g(x, y) := \frac{1}{2}[Q(x + y) - Q(x) - Q(y)]$ is a symmetric bilinear function on V (i.e., g is an inner product on V).

Then Q is called a **quadratic form on** V. The function g in the definition of Q is referred to as the bilinear form **polar** to the quadratic form Q.

Notice that

$$g(x, x) = \frac{1}{2}[Q(2x) - 2Q(x)] = \frac{1}{2}[4Q(x) - 2Q(x)] = Q(x),$$

so $g(x, x) = Q(x)$. Hence Q can be "recovered" from g.

On the other hand, beginning with *any* symmetric bilinear form g on V, the function $H(x) := g(x, x)$ is a quadratic form on V with g being its polar bilinear form. Indeed,

$$H(kx) = g(kx, kx) = k^2 g(x, x) = k^2 H(x), \quad \text{and}$$

$H(x + y) = g(x + y, x + y) = g(x, x) + g(x, y) + g(y, x) + g(y, y) = H(x) + 2f(x, y) + H(y)$, which gives

$$g(x, y) = \frac{1}{2}[H(x + y) - H(x) - H(y)].$$

It is time to mention something about the word "form". Why forms?

In algebra a **form of degree** $k \geq 0$ **of** $x_1, \ldots, x_n$ (referred to as symbols, or indeterminates, or variables) over $\mathbb{F}$ is just a homogeneous polynomial of degree k over $\mathbb{F}$, i.e., a sum of monomials with coefficients from $\mathbb{F}$ such that the (total) degree of each monomial is k. For example, for $n = 4$ and $\mathbb{F} = \mathbb{R}$, $2x_1 - 3x_2 + 5x_3 + x_4$ is form of degree 1, $2x_1x_2 - x_3^2 - x_4^2 + x_1x_3$ is a form of degree 2, $x_1^3 + x_1x_2x_3 - x_2x_4^2$ is a form of degree 3.

An equivalent definition is that a polynomial $p = p(x_1, \ldots, x_n)$ is a form of degree k, if

$$p(\lambda x_1, \ldots, \lambda x_n) = \lambda^k p(x_1, \ldots, x_n)\ textforall\ \ \lambda \in \mathbb{F}.$$

For $k = 0$, assume $\lambda^0 = 1$ for all λ.

We have seen that, as soon as a basis α in V is chosen, every bilinear function on V can be represented as $x^T By = \sum_{1 \leq i,j \leq n} b_{ij} x_i y_j$, where $[x]_\alpha = [x_1, \ldots, x_n]$ and $[y]_\alpha = [y_1, \ldots, y_n]$. This is a homogeneous polynomial of $x_1, \ldots, x_n, y_1, \ldots, y_n$ of degree 2. Also every quadratic function can be represented as $x^T Bx$, where B is the symmetric matrix corresponding to the polar bilinear form. This is a homogeneous polynomial of $x_1, \ldots, x_n$ of degree 2. This is all to it.

Problems.

1. Prove that $a^2 + b^2 + c^2 \geq ab + bc + ca$ for all real a, b, c, and that the equality is attained if and only if they are equal.

2. Let $x = (x_1, \ldots, x_5) \in \mathbb{R}^5$ and the standard norm $\|x\| = 1$. Let σ be an arbitrary permutation (bijection) on $\{1, 2, 3, 4, 5\}$, and $x_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(5)})$. What is the greatest value of $x_1 x_{\sigma(1)} + \ldots + x_5 x_{\sigma(5)}$, and for which x it is attained.

3. Prove that if $a + b + c = 1$, then $a^2 + b^2 + c^2 \geq 1/3$, and that the equality is attained if and only if $a = b = c$.

4. Consider a plane $\alpha : 2x - 3y - z = 5$ in the usual 3-space (point space). Find the point in α which is the closest to the origin.

5. Let $a_1, \ldots, a_n$ be positive real numbers. Prove that

$$(a_1 + \ldots + a_n)\,(1/a_1 + \ldots + 1/a_n) \geq n^2.$$

   For which a_i the equality is attained?

6. Prove that in the usual 2- or 3-dimensional Euclidean space, the following geometric fact holds: in any parallelogram, the sum of squares of the diagonals is equal to the sum of squares of all sides. Does it remind you something from the lectures?

7. Prove that in every tetrahedron $ABCD$, if two pairs of opposite (skew) sides are perpendicular, then so is the third pair. Prove also that in such tetrahedron, the sums of squares of lengthes of every pair of opposite sides are equal.

8. The goal of this exercise is to show that an isometry can be characterized in a somewhat more economic way, namely as a map on V which just fixes zero vector and all distances between the vectors.

   Let (V, f) be a Euclidean n-dimensional space and $\phi : V \to V$ is such that

   (i) $\phi(0) = 0$ (ϕ fixes zero vector) and

   (ii) $\|\phi(x) - \phi(y)\| = \|x - y\|$ for all $x, y \in V$ (ϕ preserves the distances between vectors).

   Prove that the following hold.

   (a) ϕ preserves the norm on (V, f): $\|\phi(x)\| = \|x\|$ for all $x \in V$;

   (b) ϕ preserves the inner product f: $f(\phi(x), \phi(y)) = f(x, y)$ for all $x, y \in V$;

   (c) ϕ maps every orthonormal basis (of V) to an orthonormal basis;

   (d) ϕ is a liner map on V;

   (e) ϕ is an isomorphism on V.

   Show that only one of the conditions (i) or (ii) does not imply that the map is an isometry.

9. Describe all isometries in the standard 2-dimensional Euclidean space. You can do it in terms of matrices or without them.

10. Consider three non-coplanar rays $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$ in the usual Euclidean 3-dimensional space. Draw bisectors of each of three angles: $\angle BAC$, $\angle CAD$, $\angle BAD$. Prove that the angles formed by these bisectors are either all acute, all all right, or all obtuse.

    Give an explicit example of $\overrightarrow{AB}$, $\overrightarrow{AC}$, $\overrightarrow{AD}$ such that the three angles are all right angles. You can do by showing the coordinates of B, C, D, assuming that A is at the origin.

Lecture 17, 18.

The goals of this lecture are the following.

- To establish some ties between usual Euclidean geometry and Linear algebra.
- To demonstrate that using even the simplest facts from linear algebra enables us to answer some nontrivial questions from 2- and 3-dimensional Euclidean geometry.
- To generalize the geometric notions to Euclidean spaces of higher dimensions.

When we think about vector spaces geometrically, we draw diagrams imagining Euclidean 1-,2-, and 3-dimensional *point spaces*. Vectors are depicted as directed segments. Often we "tie" all vector to the origin, and say that their "endpoints" correspond to the points in the space. This creates an inconvenience when we want to draw them somewhere else in the space, which is desirable for many applications of vectors, in particular, in geometry and physics. Then we agree that two different directed segments define the same vector if they have equal lengths and "directed the same". We explain how to add and subtract vectors geometrically. This type of discussion is usually not precise, but, nevertheless, we got used to passing from points to vectors and back. Coordinates allow to discuss all this with greater rigor, but the relations between the definitions and the geometric rules for operations on vectors still have to be justified. It is not hard to make the relation between point spaces and vector spaces precise, but we will not do it here. See, i.e., [15], or [8], or [12] for rigorous expositions. Usually a point space built over a vector space V is referred to as **affine space** associated with V.

Instead dealing with affine Euclidean spaces, we will translate the usual notions of these spaces into the language of vector spaces, and try to discuss them by means of linear algebra.

For the rest of this lecture we will deal with the standard inner product space $(\mathbb{R}^n, f)$ only.

Let a be a nonzero vector in E^n. A **segment spanned by** a is defined as the set of vectors

$$\{ta \,:\, 0 \le t \le 1\}.$$

Let a, b be two vectors. We define an **affine segment spanned by** a **and** b as the following set of vectors:

$$\{a + (b-a)t \,:\, 0 \le t \le 1\} = \{(1-t)a + tb \,:\, 0 \le t \le 1\} = \{t_1 a + t_2 b \,:\, 0 \le t_i,\ t_1 + t_2 = 1\}.$$

The first representation exhibits a vector $c = (1-t)a + tb$ whose endpoint C lies on (point) segment AB and divides its length in proportion $AC/CB = t/(1-t) = t_1/t_2$, $t, t_1 > 0$. For $t = t_1 = 0$ we get b and for $t = t_2 = 1$ we get a. The second way or writing is more symmetric.

Clearly, the segment spanned by a and b is a shift, or, equivalently, a (parallel) translation of the segment spanned by $b-a$ by vector a. If we allow t to range through the whole $\mathbb{R}$, we get the set of vectors which correspond to the **affine line spanned by** a **and** b or the affine line through A and B.

We say that an affine segment (line) spanned by a and b is **parallel** to the affine segment (line) spanned by c and d if vectors $b-a$ and $d-c$ are colinear.

Let b, c be two non-colinear vectors. We define the **triangle spanned by** b **and** c as the following set of vectors:

$$\{tb+sc \,:\, t,s\geq 0,\, s+t\leq 1\}.$$

If the definition of the segment made sense to us, the definition of a triangle should too. If $0<k<1$, then the set of vectors $\{tb+sc \,:\, t,s\geq 0,\, s+t=k\}$ is equal to the set of vectors $\{t(kb)+s(kc) \,:\, t,s\geq 0,\, s+t=1\}$, which is the segment with the endpoints kb and kc. Hence the triangle is the union of all such segments.

An **affine triangle spanned by** a, b, c, where vectors $b-a$ and $c-a$ are non-colinear, is the set of vectors defined as follows:

$$\{a+t(b-a)+s(c-a) \,:\, t,s\geq 0,\;\; s+t\leq 1\},\quad \text{or}$$

$$\{t_1a+t_2b+t_3c \,:\, t_i\geq 0,\; t_1+t_2+t_3=1\}.$$

We leave the verification that these sets are equal to the reader. The affine triangle spanned by vectors a, b, c is a translation by vector a of the triangle spanned by $b-a$ and $c-b$.

Let b, c, d be three non-coplanar vectors. We define a **tetrahedron spanned by** b**,** c **and** d as the following set of vectors:

$$\{t_1b+t_2c+t_3d \,:\, 0\leq t_i,\; t_1+t_2+t_3\leq 1\}.$$

An **affine tetrahedron spanned by** a, b, c, d, where vectors $b-a$, $c-a$, $d-a$ are non-coplanar, is the set of vectors defined as follows:

$$\{a+t_2(b-a)+t_3(c-a)+t_4(d-a) \,:\, 0\leq t_i,\; t_2+t_3+t_4\leq 1\},\quad \text{or}$$

$$\{t_1a+t_2b+t_3c+t_4d \,:\, 0\leq t_i,\; t_1+t_2+t_3+t_4=1\}.$$

The affine tetrahedron spanned by vectors a, b, c, d is a translation by vector a of the tetrahedron spanned by vectors $b-a$, $c-a$ and $d-a$.

Let b, c be two non-colinear vectors. We define the **parallelogram spanned by** b **and** c as the following set of vectors:

$$\{sb + tc \,:\, 0 \le s, t \le 1\},$$

and an **affine parallelogram spanned by** a, b, c, as

$$\{a + t(b - a) + s(c - a) \,:\, 0 \le s, t \le 1\},$$

where the vectors $b - a$ and $c - a$ assumed to be non-colinear.

Let b, c, d be three non-coplanar vectors. We define the **parallelepiped spanned by** b**,** c **and** d as the following set of vectors:

$$\{t_1 b + t_2 c + t_3 d \,:\, 0 \le t_i \le 1\},$$

and an **affine parallelepiped spanned by** a, b, c, d, as

$$\{a + t_1(b - a) + t_2(c - a) + t_3(d - a) \,:\, 0 \le t_i \le 1\},$$

where the vectors $b - a$, $c - a$ and $d - a$ are assumed to be non-coplanar.

We call an arbitrary subset S of V a **figure**.

Theorem 41 *Every non-singular linear operator ϕ on V preserves the following properties of affine figures.*

1. *The property of a figure to be an affine segment, line, triangle, parallelogram, or parallelepiped.*
2. *The property of segments and lines being parallel.*
3. *The ratio of lengths of parallel segments. In particular, it preserves the ratio of lengths of segments on the same line. In particular, the midpoint of a segment is mapped to the midpoint of its image.*
4. *The ratio of areas or volumes of figures, where those are defined and nonzero.*

Proof. 1. Let us demonstrate the property for affine triangles only. Others can be done similarly. Let $T = \{t_1 a + t_2 b + t_3 c \,:\, 0 \le t_i,\, t_1 + t_2 + t_3 = 1\}$ be an affine triangle spanned by a, b, c. Then $b - a$, $c - a$ are non-colinear vectors. Then $\phi(T) = \{\phi(t_1 a + t_2 b + t_3 c) \,:\, 0 \le t_i,\, t_1 + t_2 + t_3 = 1\} = \{t_1\phi(a) + t_2\phi(b) + t_3\phi(c) \,:\, 0 \le t_i,\, t_1 + t_2 + t_3 = 1\}$. As ϕ is non-singular, vectors $\phi(b - a) = \phi(b) - \phi(a)$ and $\phi(c - a) = \phi(c) - \phi(a)$ are non-colinear, and, hence, $\phi(T)$ is the affine triangle spanned by $\phi(a), \phi(b)$ and $\phi(c)$. □

2. Let $S = \{a + (b-a)t \ : \ 0 \le t \le 1\}$ and $S' = \{a' + (b'-a')t \ : \ 0 \le t \le 1\}$ be affine segments spanned by a, b and a', b', respectively. As $\phi(x + t(y-x)) = \phi(x) + t\phi(y-x) = \phi(x) + t(\phi(y) - \phi(x))$, $\phi(S)$ is the affine segment spanned by $\phi(a)$ and $\phi(b)$. Similarly, $\phi(S')$ is the affine segment spanned by $\phi(a')$ and $\phi(b')$. As S and S' are parallel, $b-a$ and $b'-a'$ are collinear. As ϕ is linear, their images, namely $\phi(b-a) = \phi(a) - \phi(b)$ and $\phi(b'-a') = \phi(a') - \phi(b')$ are colinear. Hence $\phi(S)$ is parallel to $\phi(S')$. □

3. All these statements follow immediately from the relation $\phi(x+t(y-x)) = \phi(x)+t(\phi(y-x))$. □

4. Let us first restrict ourselves to areas only, and recall the notion of an area in $\mathbb{R}^2$.

Consider a grid of congruent unit squares in $\mathbb{R}^2$. As we have shown in parts 1,2,3, parallel lines are mapped to parallel lines, and midpoints of segments to midpoints of their images. Therefore the image of this grid will be a grid of *congruent* parallelograms.

Let F_1 and F_2 be two figures in $\mathbb{R}^2$ for which area exist. If the grid of squares is sufficiently fine, then the ratio of the number of squares in the interior of F_1 to the number of squares in the interior of F_2 can be made as close to the ratio of their areas as we wish. Actually it will be equal to the ratio of areas in the limit, as the length of the side of a square in the square grid decreases to zero.

Consider now $\phi(F_1)$ and $\phi(F_2)$. The ratio of the numbers of the parallelograms in the interiors of $\phi(F_1)$ and $\phi(F_2)$ will be *exactly the same* as the ratio of the numbers of square grids in the interiors of F_1 and F_2. When the sides of a square in the square grid decreases, so does the size of the parallelogram in the corresponding parallelogram grid. Passing to the limit will give

$$\frac{\text{area}\,(F_1)}{\text{area}\,(F_2)} = \frac{\text{area}\,(\phi(F_1))}{\text{area}\,(\phi(F_2))}.$$

It is clear that a similar argument can be applied to the ratio of volumes in $\mathbb{R}^3$. ∎

Remark. The statement of part 4 can be restated as follows: there exists a positive constant $c = c(\phi)$, such that $\text{area}\,(\phi(F)) = c\,\text{area}\,(F)$ for all figure F in E^2 which have areas. The argument used in the proof of part 4, applies to all figures with positive areas (volumes), and to all dimensions. It turns out that the the coefficient c is just $|\det M_\phi|$, where M_ϕ is the matrix representing ϕ in the standard basis. Let us show it for dimensions 2 and 3.

Recall the following fundamental facts from analytical geometry: the area of a parallelogram spanned by vectors a, b, whose coordinates in the standard basis of $\mathbb{R}^2$ are $[a] = [a_1, a_2]$ and $[b] = [b_1, b_2]$, respectively, is the absolute value of the determinant of the matrix

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix},$$

and the volume of a parallelepiped spanned by vectors a, b, c, whose coordinates in the standard basis of $\mathbb{R}^3$ are $[a] = [a_1, a_2, a_3]$, $[b] = [b_1, b_2, b_3]$ and $[c] = [c_1, c_2, c_3]$, respectively, is the absolute

value of the determinant of the matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

Let $n = 2$, let $\{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$, and let $[\phi(e_1), \phi(e_2)]^T = A[e_1, e_2]^T$. Hence $|\det A|$ is the area of the parallelogram spanned by $\phi(e_1)$ and $\phi(e_2)$. Since the determinant of the identity matrix is 1, the area of the unit square (the one spanned by e_1 and e_2) is 1. So $|\det A|$ is the ratio of the area of the parallelogram to the area of the square. Let's see that the area of *any* parallelogram will change by the same factor. Let x and y be two non-colinear vectors in $\mathbb{R}^2$. Vectors x, y span a parallelogram with the area $|\det B|$, where B is the matrix having $[x]$ and $[y]$ as its rows. Then vectors $\phi(x)$ and $\phi(y)$ span a parallelogram with area $|\det(BA)|$, since the rows of BA are the coordinate vectors of $[\phi(x)]$ and $[\phi(y)]$. Since $|\det(BA)| = |\det(B)|\,|\det A|$, we obtain that the area of every parallelogram changes by the same factor $|\det A|$. A similar argument works also for $n = 3$.

Proposition 42 *Every two segments, two triangles, two parallelograms, two tetrahedra, two parallelepipeds in V can be mapped to each other by a linear map.*

Proof. The statement follows from the definitions of the figures in the statement, and the fact that every bijection between two sets of linearly independent vectors can be continued to a linear map of $\mathbb{R}^n$. ∎

Let us present three examples of how Theorem 41 and Proposition 42 can be used to solve problems in Euclidean geometry.

Example 43 *Is there a non-regular pentagon with the property that each its diagonal is parallel to one of its sides?*

Solution. Yes. It is easy to show that a regular pentagon has this property (do it!). Consider any linear operator of $\mathbb{R}^2$ which maps a regular pentagon to a non-regular one. There are many such projections: e.g., just pick three consecutive vertices of the regular pentagon and map the corresponding triangle to an equilateral one. Then the image of the whole pentagon is not regular, since one of its angles has measure of $60°$. Since parallel segments are mapped to parallel segments, the image will satisfy the required property. □

Example 44 *Let A_1, B_1, C_1 be points on the sides $\overline{BC}, \overline{CA}, \overline{AB}$ of a $\triangle ABC$, respectively, such that*

$$BA_1/A_1C = CB_1/B_1A = AC_1/C_1B = 1/2.$$

Let A_2, B_2, C_2 be the points of intersections of the segments BB_1 and CC_1, CC_1 and AA_1, AA_1 and BB_1, respectively. Prove that

$$\frac{area\,(\triangle A_2B_2C_2)}{area\,(\triangle ABC)} = \frac{1}{7}.$$

Proof. Consider a linear map which maps $\triangle ABC$ to an equilateral $\triangle A'B'C'$. Points A', B', C' will divide the sides of $\triangle A'B'C'$ in the same ratio, and therefore, $A'C_1' = B'A_1' = C'B_1' = 1$ (we can just choose the scale this way). See Figure ***. Therefore it is sufficient to solve the problem in this case, since the ratio of areas does not change.

This can be done in many ways. Here is one of them. Rotating $\triangle A'B'C'$ counterclockwise by 120° around its center, we obtain that $A_1' \mapsto B_1' \mapsto C_1' \mapsto A_1'$, where $\mapsto$ means 'is mapped to'. This implies that

$$\overline{A'A_1'} \mapsto \overline{B'B_1'} \mapsto \overline{C'C_1'} \mapsto \overline{A'A_1'},$$

and therefore $A_2' \mapsto B_2' \mapsto C_2' \mapsto A_2'$. It implies that $\triangle A_2'B_2'C_2'$ is equilateral. Using the Cosine theorem for $\triangle A'C_1'C'$, we get

$$C'C_1' = \sqrt{1^2 + 3^2 - 2 \cdot 1 \cdot 3 \cdot \cos(\pi/3)} = \sqrt{7}.$$

Now, $\triangle A'B_2'C_1' \sim \triangle A'B'A_1'$, since they have two pairs of congruent angles. Therefore $B_2'C_1' = 1/\sqrt{7}$ and $A'B_2' = C'A_2' = 3/\sqrt{7}$. Therefore $A_2'B_2' = \sqrt{7} - 1/\sqrt{7} - 3/\sqrt{7} = 3/\sqrt{7}$. This implies that $A_2'B_2'/A'B' = 1/\sqrt{7}$, and therefore

$$\frac{\text{area}\,(\triangle A_2'B_2'C_2')}{\text{area}\,(\triangle A'B'C')} = \Big(\frac{A_2'B_2'}{A'B'}\Big)^2 = \Big(\frac{1}{\sqrt{7}}\Big)^2 = \frac{1}{7}.$$

Since the ratio of areas is an invariant of a parallel projection, $\text{area}(\triangle A_2B_2C_2)/\text{area}(\triangle ABC) = 1/7$. ∎

Example 45 *Let $ABCD$ be a tetrahedra, and let E and F be the midpoints of segments AB and CD, respectively. Let α be any plane passing through E and F. Prove that α divides $ABCD$ into two polyhedra of equal volumes.*

Proof. By using a linear operator map $ABCD$ to a regular tetrahedra $A'B'C'D'$. Then E and F are mapped to the midpoints E' and F' of segments $A'B'$ and $C'D'$, and plane α to a plane α' passing through E' and F'.

Now note that the line $E'F'$ is perpendicular to the sides $A'B'$ and $C'D'$ and lies in α'. Therefore a rotation around the line $E'F'$ by 180° maps each of the two polyhedra into which α' divides $A'B'C'D'$ to another one. Hence they are congruent, and their volumes are equal. But the ratio of volumes is preserved by any non-singular linear operator. ∎

Now we generalize some of the notions above to $\mathbb{R}^n$.

Let $a_1, \ldots, a_{m+1} \in \mathbb{R}^n$ be $m+1$ vectors which span a subspace of dimension m. The following set of vectors is called the **m-simplex in $\mathbb{R}^n$ spanned by** $a_1, \ldots, a_m$:

$$\{t_1 a_1 + \ldots + t_{m+1} a_{m+1} \,:\, 0 \le t_i,\ t_1 + \ldots + t_{m+1} = 1\}.$$

For $m = 0, 1, 2, 3$, we get a point, a segment, a triangle and a tetrahedron, respectively.

Similarly, let $a_1, \ldots, a_m$ be the set of linearly independent vectors in $\mathbb{R}^n$. The set of vectors

$$\{t_1 a_1 + \ldots + t_m a_m \,:\, 0 \le t_i \le 1\},$$

is called the **m-parallelepiped in $\mathbb{R}^n$ spanned by** $a_1, \ldots, a_m$. For $m = 1, 2, 3$, we get a segment, a parallelogram and a 3-dimensional parallelepiped, respectively.

We define the **volume of an n-parallelepiped spanned by** $a_1, \ldots, a_n$ in $\mathbb{R}^n$ as $|\det A|$, where A is the matrix such that the i-th rows of A is a_i^T. We denote this volume by $\mathrm{vol}\,(a_1, \ldots, a_n)$.

It takes some work to verify that this definition of the volume of a parallelepiped can be extended to to definitions of volumes of other figures, and the obtained function satisfies all the axioms imposed on volumes as objects of Measure theory and Euclidean geometry. What is obvious, at least, is that the number is positive, and it becomes zero when the n-parallelepiped degenerates into one of a smaller dimension, i.e., when defining vectors are linearly dependent. It also satisfies our expectation that two sets of n linearly independent vectors $\{a_i\}$ and $\{b_i\}$ span the same parallelepiped then the volumes defined by them should be the same. (Check!).

Before closing this section we wish to introduce another important matrix which determinant also can be used to compute volumes.

Let (V, f) be any n-dimensional Euclidean space, and let $a_1, \ldots, a_m$ be a sequence of vectors in V. Consider a square matrix G of order m define as: $G = G(a_1, \ldots, a_m) = (g_{ij})$, where $g_{ij} = f(a_i, a_j)$. Then G is called the **Gram matrix of the sequence of vectors** $a_1, \ldots, a_m$, and $\det G$ is called the **Gram determinant** of this sequence of vectors.

Theorem 46 *Let $G = G(a_1, \ldots, a_m)$ be the Gram matrix of $a_1, \ldots, a_m$ in the n-dimensional Euclidean space (V, f). Then the following holds.*

1. *G is singular if and only if $a_1, \ldots, a_m$ are linearly dependent.*
2. *$a_1, \ldots, a_m$ is an orthonormal set of vectors if and only if $G = I_m$.*
3. *(*Hadamard's inequality.*) Let $\{a_i'\}$ be the orthogonal basis constructed from $\{a_i\}$ by Gram-Schmidt procedure. Then*

$$G = \|a_1'\|^2 \cdot \cdots \cdot \|a_m'\|^2 \le \|a_1\|^2 \cdot \cdots \cdot \|a_m\|^2.$$

The equality in the inequality is attained if and only if they are orthogonal.

4. *Let $m = n$, and let α be an orthonormal basis of (V, f). If A is a matrix which i-th row is $[a_i]_\alpha$ for all i, then $G = AA^T$. Consequently,*

$$\det G = (\det A)(\det A^T) = (\det A)^2 \geq 0, \quad \textit{and}$$

$$\operatorname{vol}(a_1, \ldots, a_n) = \sqrt{\det G} = |\det A|.$$

Proof.

1. Handout was given in class.

2. Obvious.

3. Follows from Gram-Schmidt procedure and the properties of determinants. Handout was given in class. The inequality $\|a_i'\|^2 \leq \|a_i\|^2$ follows from the Pythagorean Theorem.

4. Straightforward. ∎

Problems.

1. Prove that the area of a parallelogram and the volumes a parallelepipeds in 2- and 3- dimensional Euclidean space, respectively, are equal to the absolute values of the determinants of the corresponding matrices.

   You are allowed to use the facts from elementary geometry: the area of a parallelogram is equal to the product of the length of its side and the length of the corresponding altitude, and that the volume of a parallelepiped is the product of the area of its base and the corresponding altitude.

2. Given a tetrahedron $AA_1A_2A_3$. Let A_i' be a point on the side AA_i such that $AA_i'/AA_i = \lambda_i$, $i = 1, 2, 3$. Prove that $\mathrm{vol}\,(AA_1'A_2'A_3') = \lambda_1\lambda_2\lambda_3\ \mathrm{vol}\,(AA_1A_2A_3)$.

3. A plane passes through the midpoints of two skew sides of a tetrahedron, and intersects two other sides at points M and N. Prove that the points M and N divide the sides (they belong to) in the same ratio.

4. Think about an ellipsoid with semi-axes of length a and b in $\mathbb{R}^2$ as

$$\{\ x = (x_1, x_2)\ :\quad \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \le 1.\ \}$$

   Prove that every ellipsoid can be viewed is an image of a circular disc with respect to a linear map. Conclude that the area of the ellipsoid with semi-axes of length a and b is πab. State and prove the generalization of this result to $\mathbb{R}^3$.

5. Given a tetrahedron $AA_1A_2A_3$. Let A_i' be a point on the side AA_i such that $AA_i'/AA_i = \lambda_i$, $i = 1, 2, 3$. Prove that $\mathrm{vol}\,(AA_1'A_2'A_3') = \lambda_1\lambda_2\lambda_3\ \mathrm{vol}\,(AA_1A_2A_3)$.

6. Let $a_1, \ldots, a_m$ be vectors in the standard inner product space $(\mathbb{R}^n, f)$ such that the distance between every two of them is equal to 1. Prove that $m \le n + 1$. Construct an example of $n + 1$ vectors with this property.

   Hint: Let $u_i = a_i - a_1$ for all i. Show that $G(u_2, \ldots, u_m)$ is nonsingular.

Lectures 19.

In this lecture we discussed the notion of the distance from a vector (point) in a Euclidean space (V, f) to its n-dimensional subspace. We proved the following theorem.

Theorem 47 *Let (V, f) be a Euclidean space, and W be its n-dimensional subspace. Then for every $v \in V$, there exists a unique vector w_0 in W such that*

(i) $v - w_0 \perp W$, *and*

(ii) $\|v - w_0\| = \min\{\|v - w\| : w \in W\}$.

Proof. Given in class. ∎

Vector w_0 described in the theorem is called the **orthogonal projection of v to W**, and is often denoted by $\mathrm{proj}_W v$, and the linear map $V \to W$ defined by $v \mapsto \mathrm{proj}_W v$ is called the **orthogonal projection of V to W**. The orthogonal projection is the projection of V to W in the direction $W^\perp$, which was the notion we defined before. The number $\|v - w_0\|$ is called the **distance from v to W**, and it is often denoted by $\mathrm{dist}\,(v, W)$.

Connections to the question of approximation of functions in functional spaces by functions from a given subspace were discussed.

Problems.

1. Let (V, f) be a Euclidean space, W be its subspace, $v \in V$ and is not orthogonal to W. Prove that the angle between v and $\text{proj}_W v$ is the smallest among all angles between v and w, $w \in W$. (The measure of the angle between two nonzero vectors a and b in V is, by definition, $\cos^{-1} \frac{f(v,w)}{\|v\| \, \|w\|}$.) Try to get a good feeling for this result in the standard Euclidean 2- and 3-dimensional spaces, with W being of dimension 1 and 2, respectively.

2. In $\mathbb{R}^4$, let $W = \langle (1,1,0,0), (1,1,1,2) \rangle$. Find $w \in W$ such that $\|w - (1,0,2,2)\|$ is as small as possible.

3. Find $\text{proj}_W v$, and $\text{dist}\,(v, W)$ for (V, f), W, and v defined below.

   (i) V is the standard Euclidean space $\mathbb{R}^4$, $v = (0,1,2,3)$ and

   $$W = \{ x = (x_1, x_2, x_3, x_4) \, : \, x_1 + x_2 + x_3 - 2x_4 = 0 \}.$$

   (ii) V is the standard Euclidean space $\mathbb{R}^4$, $v = (0,1,2,3)$ and

   $$W = \{ x = (x_1, x_2, x_3, x_4) \, : \, x_1 + x_2 = x_3 - x_4, \text{ and } x_1 = x_2 + x_3 = 0 \}.$$

   (iii) $V = C[-1,1]$, $f(u,v) = \int_{-1}^{1} u(t)v(t)\,dt$, $v = \cos x$ and $W = \langle 1, x, x^2, x^3 \rangle$.

Lectures 20.

Let V be a vector space over a field $\mathbb{F}$, and let $L(V)$ denote the set of all linear operators on V. For $\phi \in L(V)$, a subset $S \subseteq V$ such that $\phi(S) \subseteq S$ is called **stable** or **invariant with respect to** ϕ, or **ϕ-stable**, or **ϕ-invariant**. Clearly $\{0\}$ and V are ϕ-stable for every ϕ. Restricting ϕ on its invariant subspace W, we obtain a linear operator $\phi|_W$ on W. The study of the latter may be easier than of ϕ on the whole V, mainly because this subspace is of smaller dimensions than V. In particular, if V is a direct sum of its subspaces $W_1, W_2, \ldots$, then $\phi|_V$ is completely defined by all $\phi|_{W_i}$. This simple idea suggest the following approach for studying linear operators.

Given $\phi \in L(V)$,

(i) find its invariant subspaces $W_1, W_2, \ldots$ whose direct sum is V;

(ii) understand how ϕ acts on each of them, i.e., all $\phi|_{W_i}$.

This is exactly what we will try to do during the next several lectures..

As the action of ϕ on $\langle 0 \rangle$ is trivial, the first interesting case is when a ϕ-invariant space is 1-dimensional.

Let $\phi(v) = \lambda v$ for some nonzero vector $v \in V$ and a scalar $\lambda \in \mathbb{F}$. Then v is called an **eigenvector of** ϕ, and λ is called an **eigenvalue of** ϕ. We also say that λ and v **correspond** to each other. Every eigenvector v of ϕ spans a 1-dimensional ϕ-stable subspace $\langle v \rangle$.

If $\dim V = n$, and if V has a basis $\alpha = \{v_1, \ldots, v_n\}$ consisting of eigenvectors of ϕ with corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, the image of an arbitrary vector can be computed in a very simple way. Let $x = x_1 v_1 + \cdots + x_n v_n$. Then

$$\phi(x) = \lambda_1 x_1 v_1 + \cdots + \lambda_n x_n v_n \,.$$

In coordinate notations, if $[x]_\alpha = [x_1, \ldots x_n]$, then $[\phi(x)]_\alpha = [\lambda_1 x_1, \ldots \lambda_n x_n]$.

Therefore finding as many as possible linearly independent eigenvectors of ϕ is useful. How can one find them? In order to answer this question, we consider matrices which represent ϕ in different bases.

Let $\dim V = n$, $\phi \in L(V)$, and $\alpha = \{v_1, \ldots, v_n\}$ be a basis of V. Let $M_{\phi,\alpha}$, be the matrix of ϕ corresponding to α. We remind ourselves that $M_{\phi,\alpha} = (a_{ij})$, where the entries a_{ij} are defined by the equalities $\phi(v_i) = a_{i1} v_1 + \cdots + a_{in} v_n$, $i = 1, \ldots, n$. To simplify the presentation, we denote $M^T_{\phi,\alpha}$ by A. Then it is easy to check that

$$[\phi(x)]_\alpha = A\,[x]_\alpha$$

for all $x \in V$. This implies that the matrix which represents ϕ with respect to a basis consisting of eigenvectors of ϕ is diagonal!

As $\phi(v) = \lambda v$ if and only if $A\,[v]_\alpha = \lambda[v]_\alpha$, we wish also to define the notions of eigenvector and eigenvalues for matrices.

Let $V = \mathbb{F}^n$, A be a square matrix of order n, and $A\,v = \lambda v$ for some nonzero vector v and a scalar λ. Then v is called an **eigenvector of** A, and λ is called an **eigenvalue of** A. The equality $A\,v = \lambda v$ is equivalent to $(\lambda\,I - A)\,v = 0$, where $I - I_n$ is the identity matrix of order n. The system of homogeneous linear equations $(\lambda\,I - A)\,x = 0$ has a nontrivial solution if and only if its matrix of coefficients $\lambda\,I - A$ is singular, or if and only if $\det(\lambda\,I - A) = 0$. Therefore, each scalar λ which satisfies the equation $\det(\lambda\,I - A) = 0$, is an eigenvalue of A, and the corresponding eigenvector can always be found by solving the system of equations $(\lambda\,I - A)\,x = 0$. Therefore, for a short while, we will discuss how one finds the eigenvalues of A.

Until this moment, we considered determinants of matrices over fields only. A very similar theory of determinants exists for matrices over commutative rings. The only ring we will be concerned with is the ring of polynomials $\mathbb{F}[x]$. If we analyze the statements about determinants which do not refer to the notion of linear independence of vectors, those will hold for matrices over $\mathbb{F}[x]$, and the proofs can be carried verbatim. The expression $x\,I - A$ can be considered as a matrix over $\mathbb{F}[x]$, and its determinant as a polynomial of x over $\mathbb{F}$. We call $\det(x\,I - A)$ the **characteristic polynomial of** A, and denote it by $c_A(x)$. We proved the following important fact.

Theorem 48 *Let A be a square matrix of order n over $\mathbb{F}$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if λ is a root of $c_A(x) = \det(x\,I - A)$.*

This theorem implies that in order to find eigenvalues of a linear operator ϕ one can choose a basis α, consider the matrix $A = M^T_{\phi,\alpha}$ which represents ϕ in this basis and find its eigenvalues. A choice of another basis β leads to another matrix $B = M^T_{\phi,\beta}$. How do eigenvalues of B compare to the ones of A? It turns out that they are exactly the same! Let us call the the multiset of all eigenvalues of a matrix A the **spectrum of** A, and denote it by $\operatorname{spec} A$.

Corollary 49 *Let A and B be square matrices which represent $\phi \in L(V)$ in bases α and β, respectively. Then $c_A(x) = c_B(x)$. Hence* $\operatorname{spec} A = \operatorname{spec} B$.

Proof. Let $\alpha = \{v_1, \ldots, v_n\}$, $\beta = \{u_1, \ldots, u_n\}$, and C be the matrix whose i-th column is $[u_i]_\alpha$. Then C is nonsingular, and $B = C^{-1}AC$. This implies

$$c_B(x) = \det(x\,I - B) = \det(x\,I - C^{-1}AC) = \det[C^{-1}\,(x\,I - A)\,C] =$$

$$\det C^{-1} \det(x\,I - A) \det C = (\det C)^{-1} \det(x\,I - A) \det C =$$

$$\det(x\,I - A) = c_A(x). \qquad \blacksquare$$

The result of this corollary allows us to define the **characteristic polynomial** $c_\phi(x)$ **of** $\phi \in L(V)$ as $c_A(x)$, where A is the matrix representing ϕ in some basis.

We have understood that when V has a basis of n eigenvectors, the matrix of ϕ in this basis is diagonal. The following theorem provides a simple sufficient condition for existence of linearly independent eigenvectors. The condition is not necessary, as simple examples show (like $\phi = id$).

Theorem 50 *Let* $\dim V = n$*, and let* $v_1, \ldots, v_m$ *be eigenvectors of a linear operator* ϕ *on* V *which correspond to distinct eigenvalues* $\lambda_1, \ldots, \lambda_m$*, respectively. Then*

$v_1, \ldots, v_m$ *are linearly independent. If* $m = n$*, then* ϕ *has a basis consisting of eigenvectors, and, hence,* ϕ *is diagonalizable.*

Proof. Suggested to be read in a textbook. $\blacksquare$

As we know, not every polynomial in $\mathbb{F}[x]$ has roots in $\mathbb{F}$. And if it does, not all of the roots must be distinct. What can be said about ϕ if $c_\phi(x)$ has this property? We will be discussing this question in the next lecture.

We wish to finish this section with an example illustrating how useful can be to diagonalize a matrix. We will find an explicit formula for the Fibonacci sequence: $F_0 = F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$.

The example was presented in class, and was based on the observation that for $i \geq 2$,

$$\begin{bmatrix} F_i \\ F_{i-1} \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{bmatrix} F_{i-1} \\ F_{i-2} \end{bmatrix}.$$

Diagonalizing $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ lead to an explicit formula for F_n. Some details were left as exercises.

Problems.

1. (i) Let λ be an eigenvalue of $\phi \in L(V)$, and let $V_\lambda = \{v \in V : \phi(v) = \lambda v\}$. Prove that V_λ is a subspace of V. V_λ is called the **eigenspace of ϕ corresponding to λ**. A similar notion can be defined for a square matrix of order n.

   (ii) Let V_{λ_i}, $i = 1, \ldots, k$, be eigenspaces of $\phi \in L(V)$ corresponding to pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Let α_i be a basis of V_{λ_i}. Then union of all α_i is a linearly independent set of vectors of V.

2. For a matrix A below, find its characteristic polynomial $c_A(x)$; $\operatorname{spec} A$; for each $\lambda \in \operatorname{spec} A$, find a maximum set of linearly independent eigenvectors of A corresponding to λ, i.e. a basis for the eigenspace of A corresponding to λ. If A is diagonalizable, find C such that $C^{-1}AC$ is a diagonal matrix.

   Try to do it first without using computer. Then use computer if you have difficulties, and in order to check your results.

(i) $A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$ (ii) $A = \begin{pmatrix} 3 & -1 & 1 \\ 7 & -5 & 1 \\ 6 & -6 & 2 \end{pmatrix}$ (iii) $A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$

3. Let $A = J_n(\lambda)$, which is a square matrix of order n, having all diagonal elements equal to λ, all entries in $(i, i+1)$ positions equal to 1 ($i = 1, \ldots, n-1$), and all zero entrees everywhere else. Matrices of the form $J_n(\lambda)$ are called **Jordan matrices** or **Jordan blocks**. Find $\operatorname{spec} J_n(\lambda)$ and a maximum set of linearly independent corresponding eigenvectors of A.

   (This will generalize your computation for part (iii) of Problem 2 of this set.)

4. Let $a, b \in \mathbb{F}$ and $a \neq b$. Find eigenvalues and eigenvectors of a square matrix A, where

$$A = \begin{pmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{pmatrix}$$

   Generalize the result, if you wish.

5. Find the characteristic polynomial of a matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -d & -c & -b & -a \end{pmatrix}.$$

6. Given a polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, find a square matrix A such that $c_A(x) = f(x)$. Such a matrix is called the **companion matrix for f**.

   (*Hint:* try a matrix like the one in Problem 5 of this set.)

7. Supply all details for the computations used in class for an explicit formula for the n-th Fibonacci number F_n. Show that
$$F_n = \frac{1}{\sqrt{5}}\big(\lambda_1^{n+1} - \lambda_2^{n+1}\big),\ n \geq 0,$$
where $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$.

8. Let $A \in M_{3\times 3}(\mathbb{R})$ such that all entries of A are positive. Prove that A has an eigenvector having all its components positive.

9. (i) Is there a matrix $A \in M_{6\times 6}(\mathbb{R})$ with negative determinant and having no real eigenvalues?

   (ii) Is there a matrix $A \in M_{6\times 6}(\mathbb{R})$ with no real eigenvector?

   (iii) Is there a matrix $A \in M_{7\times 7}(\mathbb{R})$ with no real eigenvector?

10. Let λ be an eigenvalue of ϕ and v be a corresponding eigenvector. Let $p(x) \in \mathbb{F}[x]$. Then $p(\lambda)$ is an eigenvalue of $p(\phi)$ and v is a corresponding eigenvector, i.e., $p(\phi)\, v = p(\lambda)\, v$.

11. Let ϕ be a nonsingular operator on a finite-dimensional space V. Let W be a ϕ-stable subspace of V. Prove that W is a stable subspace of ϕ^{-1}.

    Does the statement hold if V is infinite-dimensional?

12. Let $\phi \in L(V)$. Is it possible for ϕ not to have any nontrivial invariant subspaces, but for ϕ^2 to have one?

13. Prove that if linear operators $\phi, \psi \in L(V)$ commute, i.e., $\phi\psi = \psi\phi$, then every eigenspace of ϕ is an invariant subspace of ψ (and vice versa).

Lectures 21.

In this lecture we continue our discussion of the question of how to find invariant subspaces for a linear operator ϕ on V, $\dim V = n$. Instead of operator ϕ we will deal with a matrix A which represents it in some basis. We assume that A acts as an operator on $\mathbb{F}^n$, by mapping x to Ax. A choice of another basis leads to a transformation of A to a similar matrix $C^{-1}AC$. Finding a basis where ϕ is presented in a simple way is equivalent to transforming A to a simple form by means of the similarity transformation. This is usually done by further investigations of the connections between matrices and polynomials.

For any square matrix A of order n, we can consider the set of all matrices of the form $p(A) = a_dA^d + a_{d-1}A^{d-1} + \ldots + a_1A + a_0I_n$, where all $a_i \in \mathbb{F}$. We add and multiply such matrices similarly to the polynomials in $\mathbb{F}[x]$. We can also think that $p(A)$ is obtained from $p(x) = x^d + a_{d-1}x^{d-1} + \ldots + a_1x + a_0$ by substituting A instead of x. In order to do it, we just have to interpret a_0 as $a_0I_n = a_0A^0$. If the reader is familiar with the notion of ring (or algebra) homomorphism, we can just say that $p(A)$ is the image of $p(x)$ under the homomorphism $\mathbb{F}[x] \to M_{n\times n}(\mathbb{F})$, where $p(x) \mapsto p(A)$. If $p(A) = 0$, we say that the polynomial p an **annihilating polynomial** of A.

Given A, is there always an annihilating polynomial of A different from zero polynomial? The answer is Yes, and the proof is surprisingly easy.

The algebra $M_{n\times n}(\mathbb{F})$ is a n^2-dimensional space over $\mathbb{F}$. Therefore matrices $A^d, A^{d-1}, \ldots A, I$ form a linearly dependent set in $M_{n\times n}(\mathbb{F})$ if $d \geq n^2$, as we get at least $n^2 + 1$ matrices in the set. If $\lambda_dA^d + \lambda_{d-1}A^{d-1} + \ldots + \lambda_1A + \lambda_0I_n = 0$, then $p(x) = x^d + a_{d-1}x^{d-1} + \ldots + a_1x + a_0$ is an annihilating polynomial of A. Hence we proved the following fact.

Proposition 51 *For every matrix $A \in M_{n\times n}(\mathbb{F})$, there exists an annihilating polynomial of A of degree at most n^2.* ∎

Dividing an annihilating polynomial of A by its leading coefficient, we get a monic annihilating polynomial. It is clear that among all annihilating polynomials of A, there exists a monic one of the smallest degree. It is called the **minimal polynomial of** A. We denote it by $m_A(x)$. The degree of the minimal polynomial is always at least 1.

Proposition 52 *Minimal polynomial of A divides every annihilating polynomial of A.*

Proof. Let $p(A) = 0$. Dividing $p(x)$ by $m_A(x)$ with remainder, we obtain

$$p(x) = q(x)m_A(x) + r(x), \text{ where } 0 \leq \deg r(x) < \deg m_A(x).$$

Substituting $x = A$, we obtain $0 = q(A)0 + r(A)$. Hence $r(x)$ is an annihilating polynomial of A of degree smaller than the degree of the minimal polynomial, and therefore $r(x)$ is zero polynomial. This proves that $m_A(x)$ divides $p(x)$. ∎

In order to move our investigations further, we have to recall several basic notions and facts about polynomials. We remind the readers that for every two polynomials $a = a(x)$ and $b = b(x)$ in $\mathbb{F}[x]$, not both zero polynomials, there exits a unique monic polynomial $d = d(x)$ such that d is a common divisor of a and b (i.e., d divides both a and b), and d is divisible by every other common divisor of a and b. It is called the **greatest common divisor of** a **and** b, and it is denoted by $\gcd(a, b)$. The $\gcd(a, b)$ can be found by the Euclidean algorithm applied to a and b, and it leads to the following fundamental fact: if $d(x) = \gcd(a(x), b(x))$, there exist polynomials $u(x), v(x)$ such that

$$d(x) = u(x)a(x) + v(x)b(x).$$

If $\gcd(a, b) = 1$, a and b are called **relatively prime**. In this case, the above equality becomes

$$1 = u(x)a(x) + v(x)b(x).$$

The following main theorem allows to reduce the question of finding invariant subspaces of A to the one of factoring of polynomials.

Theorem 53 (Splitting Theorem) *Let $p(x)$ be an annihilating polynomial of A, and suppose that*

$$p(x) = p_1(x)\, p_2(x),$$

where $p_1(x)$ and $p_2(x)$ are relatively prime. Then $V = \mathbb{F}^n$ can be represented as the direct sum

$$V = V_1 \oplus V_2,$$

where subspaces V_1 and V_2 are invariant with respect to A. Moreover,

$$V_1 = \textbf{ker } p_2(A), \quad \textit{and} \quad V_2 = \textbf{ker } p_1(A),$$

so $p_1(x)$ and $p_2(x)$ are annihilating polynomials of $A|_{V_2}$ and $A|_{V_1}$, respectively.

Proof. As p_1 and p_2 are relatively prime, there exist $q_1, q_2 \in \mathbb{F}[x]$ such that

$$q_1(x)p_1(x) + q_2(x)p_2(x) = 1,$$

and hence

$$q_1(A)p_1(A) + q_2(A)p_2(A) = I.$$

Let $V_i = \mathbf{im}\ p_i(A)(V) = \{p_i(A)\, v :\ v \in V\}$, $i = 1, 2$. For every $x \in V_i$, $x = p_i(A)\, y$ for some $y \in V$. Then, as $A\, x = A\, p_i(A)\, y = p_i(A)\, A\, y \in V_i$, as $A\, y \in V$. Hence each V_i is invariant with respect to A.

For every $x \in V_1$, $x = p_1(A)y$ for some $y \in V$. Then

$$p_2(A)\, x = p_2(A)(p_1(A)\, y = (p_2(A)p_1(A))\, y = p(A)\, y = 0\, y = 0.$$

Hence $V_1 \leq \mathbf{ker}\ p_2(A)$. A similar argument gives $V_2 \leq \mathbf{ker}\ p_1(A)$.

For every $v \in V$,

$$v = q_1(A)p_1(A)\, v + q_2(A)p_2(A)\, v = q_1(A)(p_1(A)\, v) + q_2(A)(p_2(A)\, v) = v_1 + v_2,$$

where $v_i = q_i(A)(p_i(A)\, v)$. Since $p_i(A)\, v \in V_i$, and V_i is A-stable, V_i is $q_i(A)$-stable. Hence $v_i \in V_i$. This proves that $V = V_1 + V_2$.

For every $v \in V_1 \cap V_2$,

$$v = q_1(A)p_1(A)\, v + q_2(A)p_2(A)\, v = q_1(A)(p_1(A)\, v) + q_2(A)(p_2(A)\, v) =$$

$$q_1(A)\, 0 + q_2(A)\, 0 = 0.$$

Hence $V_1 \cap V_2 = \langle 0 \rangle$, and the sum $V_1 + V_2$ is direct. Hence $V = V_1 \oplus V_2$.

The only statement remained to be proved is that

$$V_1 = \mathbf{ker}\ p_2(A), \ \text{ and } \ V_2 = \mathbf{ker}\ p_1(A).$$

As $V = V_1 \oplus V_2 = V_1 \oplus \mathbf{im}\ p_2(A)$, $\dim V_1 = \dim V - \dim \mathbf{im}\ p_2(A) = \mathbf{ker}\ p_2(A)$. We have already showed that $V_1 \leq \mathbf{ker}\ p_2(A)$. Hence $V_1 = \mathbf{ker}\ p_2(A)$. The equality $V_2 = \mathbf{ker}\ p_1(A)$ can be proven similarly. ∎

Theorem 53 explains how one can split V in the direct sum of A-stable subspaces. The strategy is simple:

(1) Find an annihilating polynomial $p(x)$ of A

(2) Represent it as a product of pairwise relatively prime factors: $p(x) = p_1(x) \cdot \ldots p_k(x)$, such that each factor $p_i(x)$ cannot be further split in this way.

(3) Consider $A|_{\mathbf{im}\ p_i(A)}$, $i = 1, \ldots, k$, and try to find a basis in $\mathbf{im}\ p_i(A)$, where the operator defined by $A|_{\mathbf{im}\ p_i(A)}$ can be easily described. The latter is equivalent to finding a matrix similar to $A|_{\mathbf{im}\ p_i(A)}$ and of a simple form.

How easy is to accomplish all these steps? It depends on $\mathbb{F}$ and on A.

Regarding (1). According to Proposition 51, there always exists an annihilating polynomial of A of degree n^2, but we do not have a good way of finding it, and if the degree is close to n^2, then even if it is found, it can be very hard to work with. For $n = 10$, it may be of degree 99.

It turns out that there exists a much better way. It was found by two of the creators of matrix theory in the 19-th century. It turns out that an annihilating polynomial of degree n exists, and we actually already know what it is!

Theorem 54 (Hamilton-Cayley Theorem). *Let $A \in M_{n\times n}(\mathbb{F})$, and let $c_A(x) = \det(x\,I - A)$ be the characteristic polynomial of A. Then $c_A(A) = 0$, i.e., every matrix is annihilated by its characteristic polynomial.*

The obvious "proof": $c_A(A) = \det(A\,I - A) = \det(0) = 0$, is, unfortunately, a nonsense. We will discuss a proof later, but none of the existing proofs is easy in general case.

Striving for more, namely for the minimal polynomial $m_A(x)$ of A which can have much smaller degree than n, we can look for it among the factors of $c_A(x)$, as it must divide it due to Proposition 52.

Regarding (2). We see that the success in accomplishing part 1 depends heavily on the property of polynomials over $\mathbb{F}$ and on particular matrix A. What do we know about factoring of polynomials?

There are several fundamental theorems in this regard. A polynomial $f \in \mathbb{F}[x]$ is called **irreducible in** $\mathbb{F}[x]$ if $\deg f \geq 1$ and f is not a product of two polynomials of smaller degrees from $\mathbb{F}[x]$.

Theorem 55 *Every polynomial $f \in \mathbb{F}[x]$ can be represented as a product of irreducible polynomials from $\mathbb{F}[x]$. Such representation is unique up to order of factors and multiplication of the factors by scalars from $\mathbb{F}$. In particular, for every monic polynomial $f \in \mathbb{F}[x]$ of degree at least 1 is either irreducible, or*

$$f(x) = f_1(x)^{e_1} f_2(x)^{e_2} \cdot \ldots \cdot f_k(x)^{e_k},$$

where $k \geq 2$, $f_i(x)$ are distinct irreducible monic polynomials, and e_i are positive integers. This representation is unique up to the order of factors.

Over an arbitrary field $\mathbb{F}$, Theorem 55 is all we can say.

But more can be said if $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$.

Theorem 56 *Every monic polynomial $f \in \mathbb{C}[x]$ of degree at leat 2 can be represented as a product of powers of distinct monic linear polynomials:*

$$f(x) = (x - \lambda_1)^{e_1}(x - \lambda_2)^{e_2} \cdot \ldots \cdot (x - \lambda_k)^{e_k},$$

where λ_i are all distinct roots of f in $\mathbb{C}$. Such representation is unique up to order of factors.

Theorem 57 *Every monic polynomial $f \in \mathbb{R}[x]$ of degree at leat 2 can be represented as a product of powers of distinct monic linear polynomials and monic quadratic polynomials:*

$$f(x) = (x - \lambda_1)^{e_1}(x - \lambda_2)^{e_2} \cdot \ldots \cdot (x - \lambda_k)^{e_k} \cdot q_1(x)^{t_1} \cdot \ldots \cdot q_s(x)^{t_s},$$

where λ_i are all distinct roots of f in $\mathbb{R}$, and $q_j(x)$ are irreducible monic quadratic polynomials over $\mathbb{R}$. Such representation is unique up to order of factors.

Regarding (3). This part is also far from easy. As two success stories, we present the Jordan canonical form, and The Rational Canonical Form. Jordan form can be used in all those cases when we have (or there exists) a factorization of an annihilating polynomial into the product of linear factors. In particular, it exists for matrices over $\mathbb{C}$. The The Rational Canonical Form can be used whenever we have factorization of an annihilating polynomial into the product of powers of distinct irreducible factors. Both forms cover the diagonal case, if such is possible.

For particular classes of matrices, more can be said. Those cover symmetric real matrices, hermitian matrices, and the orthogonal real matrices (the ones which correspond to the isometries inner product spaces). The list can be continued.

Problems.

1. Give an example of two non-similar square matrices A and B such that

   (i) $c_A(x) = c_B(x)$

   (ii) $m_A(x) = m_B(x)$

   (iii) $c_A(x) = c_B(x)$ and $m_A(x) = m_B(x)$.

2. What is wrong with the obvious "proof" of the Hamilton-Cayley Theorem: $c_A(A) = \det(A\,I - A) = \det(0) = 0$.

3. Assuming Hamilton-Cayley Theorem, prove that $c_A(x)$ divides $(m_A(x))^t$ for some positive integer t.

4. Let $A \in M_n(\mathbb{F})$ and $c_A(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be all eigenvalues of A (not necessarily distinct). Then

   $$c_{n-1} = \mathbf{tr}\ A = \lambda_1 + \lambda_2 + \ldots + \lambda_n \qquad \text{and} \qquad c_0 = (-1)^n \det A = \lambda_1 \lambda_2 \ldots \lambda_n.$$

   In general, using Viète's formuli, we obtain that $c_{n-k} = (-1)^k \sum \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k}$, where the summation is over all distinct sequences $i_1 < i_2 \ldots < i_k$ of distinct integers from $\{1, \ldots, n\}$.

Lecture 22.

In this lecture we describe a special basis for an operator $\psi \in L(V)$ having $(x-\lambda)^a$ as its annihilating polynomial. As $(\psi - \lambda\, id)^a = 0$, setting $\phi = \psi - \lambda\, id$, we obtain an even simpler equation $\phi^a = 0$. An operator ϕ with the property $\phi^a = 0$ for some $a \in \mathbb{N}$, is called **nilpotent**, and the smallest positive $b \in \mathbb{N}$ such that $\phi^b = 0$ is called the **order of nilpotency** of ϕ.

The most prominent nilpotent operator is, undoubtedly, the differential operator D on a vector space of polynomials over $\mathbb{R}$ (or $\mathbb{C}$) of degree at most $m-1$, which maps every polynomial to its derivative. Consider the following basis in this space: $\{v_i = \frac{1}{i!}\, x^i \ : i = 0, \ldots, m-1\}$. It is clear that

$$v_{m-1} \stackrel{D}{\longmapsto} v_{m-2} \stackrel{D}{\longmapsto} \cdots \stackrel{D}{\longmapsto} v_1 \stackrel{D}{\longmapsto} v_0 \stackrel{D}{\longmapsto} 0,$$

and that D is nilpotent with m being the order of nilpotency.

The $m \times m$ matrix of D in this basis, ordered $(v_{m-1}, \ldots, v_0)$, has a very simple form, having 1 in positions $(1,2), (2,3), \ldots, (m-1, m)$, and 0 everywhere else. It is denoted by $J_m(0)$. The matrix $J_m(\lambda) := \lambda\, I + J_m(0)$, which is obtained from $J_m(0)$ by putting a scalar λ on the main diagonal, is called a **Jordan matrix**, or a **Jordan block**. For $m = 4$,

$$J_4(0) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad J_4(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}$$

Observe that x^m and $(x-\lambda)^m$ are the annihilating polynomials of $J_m(0)$ and $J_m(\lambda)$, respectively. Moreover, they are the minimal polynomials.

Though we arrive to Jordan matrices via the example of a particular nilpotent operator D, it turns out that similar bases exist for other nilpotent operators. This explains the importance of Jordan matrices in linear algebra. Before we prove the existence of such a basis, we would like to mention another attractive computational feature of Jordan matrices, wildly admired in some 19-th and 20-th century pre-computer societies.

Theorem 58 *Let $p(x) \in \mathbb{C}[x]$. Then*

$$p(J_m(\lambda)) = \begin{bmatrix} p(\lambda) & \frac{p'(\lambda)}{1!} & \frac{p''(\lambda)}{2!} & \ldots & \frac{p^{(m-1)}(\lambda)}{(m-1)!} \\ 0 & p(\lambda) & \frac{p'(\lambda)}{1!} & \ldots & \frac{p^{(m-2)}(\lambda)}{(m-2)!} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & p(\lambda) \end{bmatrix}$$

Proof. Was discussed in class. See Gel'fand's book [6], p. 136 – 137, for details. ∎

We are ready to prove the main theorem of this lecture.

Theorem 59 *Let* $\dim V = n$*, and let* $\phi \in L(V)$ *with minimal polynomial* x^a*. Then there exists a basis* α *of* V *such that the matrix* $M_{\phi,\alpha}$ *is block-diagonal of the form*

$$M_{\phi,\alpha} = \operatorname{diag}[J_{a_1}(0), \ldots, J_{a_1}(0), J_{a_2}(0), \ldots, J_{a_2}(0), \ldots\ldots, J_{a_t}(0), \ldots, J_{a_t}(0)],$$

where $a = a_1 > a_2 > \ldots \geq a_t$*.*

Proof. Let $\phi^0 = id$, and $W_i := \mathbf{ker}\ \phi^{a-i}$ for $i = 0, \ldots, a$. First we show that

$$V = W_0 > W_1 > \ldots > W_a = \langle 0 \rangle, \tag{19}$$

where each W_i is ϕ-invariant, and all inclusions are strict.

Indeed, if $x \in W_i$, $1 \leq i \leq a$, then $\phi^{a-i+1}(x) = \phi(\phi^{a-i}(x)) = \phi(0) = 0$. Hence $x \in \mathbf{ker}\ \phi^{a-(i-1)} = W_{i-1}$, and

$$W_{i-1} \geq W_i. \tag{20}$$

For $1 \leq i \leq a-1$, $0 = p^{a-i}(x) = \phi^{a-i-1}(\phi(x))$, hence, $\phi(x) \in \mathbf{ker}\ \phi^{a-(i+1)} = W_{i+1}$. So $\phi(W_{i+1}) \geq W_i$. Together with (20), it gives $W_i \geq \phi(W_i)$, which proves that each W_i is ϕ-invariant. For $i = a$ the statement is obvious. Suppose that for some i, $0 \leq i \leq a-1$, $W_i = W_{i+1}$. Then, $W_i \neq \langle 0 \rangle$, as the order of nilpotency of ϕ is a. Hence $\mathbf{im}\ \phi^{a-i} = \mathbf{im}\ \phi^{a-i-1}$, since the former is a subspace of the latter and they have equal positive dimensions. This implies $\langle 0 \rangle = \mathbf{im}\ \phi^a = \mathbf{im}\ \phi^{a-1} = \ldots \mathbf{im}\ \phi^{a-i} = \mathbf{im}\ \phi^{a-i-1} \neq \langle 0 \rangle$, a contradiction. This proves that all inclusions in (19) are strict.

Let U be a proper subspace in V. A set of vectors $\{v_1, \ldots, v_p\}$ from $V \setminus U$ is called **U-independent** if $a_1 v_1 + \ldots + a_p v_p \in U$ implies $a_1 = \ldots = a_p = 0$. It is clear that uniting a U-independent set with a set of linearly independent vectors of U, we obtain a linearly independent set of vectors of V (check!). A U-independent set is called a **U-basis** if it is U-independent, and united with a basis of U gives a basis of V. It is clear, that $\{v_1, \ldots, v_p\}$ is a U-basis if and only if $p = \dim V - \dim U$ (check!). It is also clear that at least one U-basis always exists (check!).

Lemma 60 *Let* $\alpha_i = \{v_1, \ldots, v_p\} \subset W_{i-1} \setminus W_i$ *be a* W_i*-basis of* W_{i-1}*,* $1 \leq i \leq a-1$*. Then* $\phi(\alpha_i) = \{\phi(v_1), \ldots, \phi(v_p)\} \subseteq W_i$ *is* W_{i+1}*-independent.*

Proof. We have

$$a_1\phi(v_1) + \ldots + a_p\phi(v_p) \in W_{i+1} \quad \Leftrightarrow \quad \phi(a_1 v_1 + \ldots + a_p v_p) \in W_{i+1} \quad \Leftrightarrow$$

$$\phi^{a-(i+1)}(\phi(a_1 v_1 + \ldots + a_p v_p)) = 0 \quad \Leftrightarrow \quad \phi^{a-i}(a_1 v_1 + \ldots + a_p v_p) = 0 \quad \Leftrightarrow$$

$$a_1 v_1 + \ldots + a_p v_p \in W_i \quad \Rightarrow \quad a_1 = \ldots = a_p = 0,$$

since α_i is a W_i-basis of W_{i-1}. ∎

Having Lemma 60, the construction of the desirable basis is as follows. Let $d_i = \dim W_i$. First chose

$$\alpha_1 = \{e_1, \ldots, e_{s_1}\},$$

a W_1-basis of $W_0 = V$, with $s_1 = d_0 - d_1 = n - d_1$. Then $\{\phi(e_1), \ldots, \phi(e_{s_1})\}$ is a linearly independent set in W_1. Extend it to

$$\alpha_2 = \{\phi(e_1), \ldots, \phi(e_{s_1}), e_{s_1+1}, \ldots, e_{s_2}\},$$

a W_2-basis of W_1 of $d_2 - d_1$ elements. Continue until you get a basis α_a of W_{a-1} (over $W_a = \langle 0 \rangle$) of $d_{a-1} - d_a = d_{a-1} - 0 = d_{a-1}$ elements. Let us list all these relative basis in the following table.

$\alpha_1 :$ $\quad e_1 \quad \ldots \quad e_{s_1}$

$\alpha_2 :$ $\quad \phi(e_1) \quad \ldots \quad \phi(e_{s_1}), \qquad e_{s_1+1} \quad \ldots \quad e_{s_2}$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

$\alpha_a :$ $\quad \phi^{a-1}(e_1) \quad \ldots \quad \phi^{a-1}(e_{s_1}), \quad \phi^{a-2}(e_{s_1+1}) \quad \ldots \quad \phi^{a-2}(e_{s_2}), \quad e_{s_{a-1}+1} \ldots e_{s_a}.$

Now we collect vectors in this table which stand in the same column. Let

$$\beta_i = \{e_i, \ldots, \phi^{b_i - 1}(e_i)\}, \tag{21}$$

where $i = 1, \ldots, s_a$, and

$$b_1 = \ldots = b_{s_1} = a - 1,\ b_{s_1+1} = \ldots = b_{s_2} = a - 2, \ldots,\ b_{s_{a-1}+1} = \ldots = b_{s_a} = 1,$$

with $b_1 \geq b_2 \geq \ldots \geq b_{s_a}$.

Lemma 61

(i) Each subspace $\langle \beta_i \rangle$ is ϕ-invariant.

(ii) $\alpha = \bigcup_{j=1}^{a} \alpha_j = \bigcup_{i=1}^{s_a} \beta_i$ is a basis of V.

(iii) Each β_i is a linearly independent set. $\beta_i \cap \beta_j = \emptyset$ for $i \neq j$.

Proof. (i) This part is obvious, since $\phi^{b_i}(e_i) = 0$.

(ii) α is a basis of V due to its construction: $\bigcup_{j=a}^{i} \alpha_j$ is just a basis of W_{i-1}. Note that $|\alpha| = |\alpha_a| + \ldots + |\alpha_1| = (d_{a-1} - d_a) + (d_{a-2} - d_{a-1}) + \ldots + (d_0 - d_1) = d_0 - d_a = n - 0 = n$.

(iii) β_i is linearly independent as a subset of the basis α. If $i \neq j$, and $\beta_i \cap \beta_j \neq \emptyset$, then two vectors from some α_j are equal, or two vectors from distinct α_j and $\alpha_{j'}$ are equal. None of these cases is possible. ∎

To finish our proof of the theorem, we just observe that if $\phi_i = \phi|_{\langle \beta_i \rangle}$, then $M_{\phi_i,\beta_i} = J_{b_i}(0)$. ∎

Let us mention some important corollaries of of Theorem 59.

Theorem 62 (Upper-triangular Form) *Every square matrix over $\mathbb{C}$ is similar to an upper triangular matrix.*

Proof. Since the Jordan form is upper trianguar, the statement follows. ∎

Theorem 63 (Hamilton-Cayley) *For every matrix $A \in M_n(\mathbb{C})$, $c_a(A) = 0$.*

Proof. As we mentioned before, no 'easy' proof of this theorem exists. Instead of presenting a proof, we describe four different ideas on which a proof can be based, and refer the reader to the literature.

1. For a proof based on the Jordan form, see [14] p. 155, or

`http://www.blue-arena.com/mewt/entry.php?id=147` , or

`http://www.cs.ut.ee/~toomas_l/linalg/lin1/node19.html`

2. For a proof based on Theorem 62, see [13], or [1], p. 173. Of course, in this case Theorem 62 should be proved independently from Theorem 59. The latter can be done by induction and can be found in [1] p. 84, or [8] p. 64-65.

3. For a proof based on the density of matrices with distinct eigenvalues in the space of all matrices, see `http://planetmath.org/encyclopedia/ProofOfCayleyHamiltonTheorem.html` The density is understood relative to Zariski's topology. An advantage of this proof is that it works for many fields different from $\mathbb{C}$.

4. For a proof based on the isomorphism of rings $M_n(\mathbb{F})[x]$ and $M_n(\mathbb{F}[x])$, see [4] p. 94-95. Proofs based on this idea can be found in many other books, but many of them do not state the isomorphism clearly, and develop some weaker results instead. ∎

The following theorem provides more details on the block structure of the Jordan form of an operator. We remind the reader that the **algebraic multiplicity of an eigenvalue λ of ϕ** is the multiplicity of λ as a root of the characteristic polynomial $c_\phi(x)$. A **geometric multiplicity of an eigenvalue λ of ϕ** is the dimension of its eigenspace V_λ.

Theorem 64 *Let* $\dim V = n$*, and let* $\phi \in L(V)$ *with the characteristic polynomial* $c_\phi(x) = \prod_{i=1}^k (x-\lambda_i)^{e_i}$ *and the minimal polynomial* $m_\phi(x) = \prod_{i=1}^k (x-\lambda_i)^{m_i}$*. Then there exists a basis* β *of* V *such that the matrix* $M_{\phi,\beta}$ *is block-diagonal of the form*

$$M_{\phi,\beta} = \operatorname{diag}[B_1, B_2, \ldots, B_k],$$

where each

$$B_i = \operatorname{diag}[J_{m_{i,1}}(\lambda_i), J_{m_{i,2}}(\lambda_i), \ldots, J_{m_{i,l_i}}(\lambda_i),]$$

where $m_{i,1} \geq m_{i,1} \geq \ldots \geq m_{i,l_i} \geq 1$.

Moreover,

(i) Each B_i *is a* $e_i \times e_i$ *matrix.*

(ii) The number l_i *of Jordan blocks in* B_i *is equal to the geometric multiplicity of* λ_i*.*

(iii) $m_{i,1} = m_i$*,* $i = 1, \ldots, k$*.*

(iv) The total number of 1*'s above the diagonal in each* B_i *is* $e_i - l_i$*.*

(v) The Jordan form of a matrix is unique.

Proof. (i) Suppose B_i is a $b_i \times b_i$ matrix. Then

$$c_{M_{\phi,\beta}}(x) = \prod_{i=1}^k (x-\lambda_i)^{b_i},$$

since the matrix $xI - M_{\phi,\beta}$ is upper triangular and its determinant is equal to the product of its diagonal entries. Then $b_i = e_i$ from the uniqueness of a representation of a polynomial as a product of irreducible factors.

(ii) Each diagonal block of B_i has exactly one eigenvector corresponding to the eigenvalue λ_i. Therefore B_i has $\dim \mathbf{ker}\ (\phi - \lambda_i\, id)$ linearly independent eigenvectors, which is, by definition, l_i.

(iii) Each matrix B_i is annihilated by $(x-\lambda_i)^{m_i}$. Hence each block of B_i is annihilated by $(x-\lambda_i)^{m_i}$. Hence $m_{i,1} \leq m_i$. If $m_{i,1} < m_i$, then $(x-\lambda_i)^{m_{i,1}}$ would annihilate B_i, and the minimal polynomial $m_B(x)$ would have smaller degree than $m_\phi(x)$, a contradiction.

(iv) Each B_i has l_i Jordan blocks, and the number of 1's above the diagonal in each block is one less than the block's size.

(v) Most proofs are by induction, and are reduced to the uniqueness of the Jordan form of a nilpotent operator. The idea is to use the fact that the cardinalities for bases a_j (see the table

which precedes (21)) define the cardinalities of β_i given by (21) uniquely. For details, see, e.g., [8]. The uniqueness of the Jordan form can also be derived from a much more general theorem of the uniqueness of the elementary divisor form for finitely generated modules over a PID. The latter is usually presented in Algebra texts. See, e.g., [5], Chapter 12.3. ∎

Problems.

1. Find all possible Jordan forms for a real matrix A whose characteristic and minimal polynomials are as follows.

   (a) $c_A(x) = (x-1)^4(x-2)^2$, $m_A(x) = (x-1)^2(x-2)^2$

   (b) $c_A(x) = (x+5)^4(x-2)^4$, $m_A(x) = (x+5)^2(x-2)^2$

   (c) $c_A(x) = (x-1)^6$, $m_A(x) = (x-1)^2$

2. Given $c_A(x)$ and $m_A(x)$ for $A \in M_3(\mathbb{C})$, show that these completely determines the Jordan form of A.

3. Find the Jordan form of the following matrices.

$$\text{(i)}\quad A = \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix} \qquad \text{(ii)}\quad A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 2 & 1 \\ 2 & -1 & 4 \end{bmatrix} \qquad \text{(iii)}\quad A = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 5 & 2 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 5 & -2 \end{bmatrix}$$

$$\text{(iv)}\quad A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 1 & 3 \\ 1 & 2 & 0 \end{bmatrix} \qquad \text{(v)}\quad A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \text{(vi)}\quad A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

4. Find $p\left(J_4(3)\right)$ for $p(x) = x^5 - 4x^3 + x^2 - x - 1$.

5. Let $A \in M_n(\mathbb{F})$ be upper-triangular. Then the multiset of its diagonal entries is precisely $\operatorname{spec} A$.

6. Let $A \in M_n(\mathbb{F})$ and $c_A(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be all eigenvalues of A (not necessarily distinct). Then

$$c_{n-1} = \mathbf{tr}\ A = \lambda_1 + \lambda_2 + \ldots + \lambda_n \qquad \text{and} \qquad c_0 = (-1)^n \det A = \lambda_1 \lambda_2 \ldots \lambda_n.$$

   (This problem has appeared earlier, but we suggest that readers think about it again.)

7. Show that a square matrix A is diagonalizable if and only if the minimal polynomial $m_A(x)$ is a product of distinct linear factors (i.e., $m_A(x) = (x-\lambda_1) \cdot \ldots \cdot (x-\lambda_k)$ where all $\lambda_1, \ldots, \lambda_k$ are distinct).

8. Let $A \in M_n(\mathbb{C})$ such that $A^m = I$, $m \geq 1$. Prove that A is diagonalizable.

9. Let $A \in M_n(\mathbb{C})$. Prove that A is similar to A^T.

10. Let $A \in M_n(\mathbb{C})$ such that $\mathbf{tr}\ A = 0$. Prove that there exist $B, C \in M_n(\mathbb{C})$ such that $A = BC - CB$.

References

[1] S. Axler. Linear Algebra Done Right, 2nd edition, Springer-Verlag, 1997.

I disagree that this text does Linear Algebra "right", and I disagree with many methodological decisions of the author. But some pages and exercises are good.

[2] L. Babai, P. Frankl. Linear Algebra Methods in Combinatorics. Preliminary Version 2, Department of Computer Science, The University of Chicago, 1992. Great price.

An unfinished manuscrip. Excellent for the title, but also much beyod it. Some chapters are masterpieces.

[3] O. Bretcher, Linear Algebra with Applications, Prentice Hall, 1997.

A very good undergraduate text. If you find some sections of these notes to fast/hard, try to find the corresponding material in this book. Sometimes you will not succeed.

[4] M.L. Curtis. Abstract Linear Algebra, Springer-Verlag New York Inc., 1990.

Nice, rather algebraic. A friendly introduction to exterior algebras, those some details are missing (like in these notes).

[5] D.S. Dummit and R.M. Foote. Abstract Algebra, 3rd edition, John Wiley & Sons, Inc., 2004.

A quite complete text in Abstract Algebra. Good for references and examples.

[6] I.M. Gel'fand. Lectures on Linear Algebra, Dover, 1989.

A classics. The terminology and notations are sometimes out of fashion. No matter how many times I read this thin book, I often find something new in it. Great price.

[7] K. Hoffman, R. Kunze. Linear Algebra, 2nd edition, Prentice Hall, 1971.

Quite complete and thourough. Good as a reference, but it is not easy to use.

[8] A.I. Kostrikin, Yu. I. Manin. Linear Algebra and Geometry (Algebra, Logic and Applications), Gordon and Breach Science Publishers, 1989.

Outstanding and demanding. Details may be read elsewhere. Makes connections with many advanced mathematical topics. A stimulating exposition of linear algebra related to Quantum Mechanics.

[9] S. Lang. Linear Algebra, 3rd edition, Springer-Verlag, 1987.

Not great, but some chapter are fine.

[10] S. Lipschutz and M. Lipson. Linear Algebra, 3rd edition. Schaum's Outline Series, McGraw-Hill, 2001.

Very well written text. Many good examples and exercises.

[11] P. D. Lax. Linear Algebra, 3rd edition, John Wiley & Sons, Inc., 1997.

Stimulating, some great examples, rather unusual (for linear algebra texts) content.

[12] B.A. Rosenfeld. Multidimensional Spaces, Nauka, Moscow, 1966. (In Russian).

A very clearly written monograph, which discusses the use of linear algebra in high-dimensional geometries.

[13] G. Sewell, http://www.math.tamu.edu/ sewell/640/notes3b.pdf

[14] G.E. Shilov. Linear Algebra. Dover, 1977.

A classics. The terminology and notations are sometimes out of fashion. Complete. The best treatment of orthogonalization, Gram matrix/determinant, and volumes. Great price.

[15] H. Weyl. Space,Time, Matter, Springer, 1923.

A great book in general. Has definition of Euclidean point spaces based on linear algebra.