Eidgenössische Technische Hochschule Zürich

Linear Algebra I/II

Autumn Semester 2025 / Spring Semester 2026

Lecture Notes

Contents

Lineare Algebra I/II

Fynn Krebser–fkrebser@student.ethz.ch

May 29, 2026

1 Introduction

Lec 1

1.1 Fibonacci Sequences

Definition 1.1:
The fibonacci sequence is defined as follows:

$$a_n = \begin{cases} 0, & n = 0 \\ 1, & n = 1 \\ a_{n-1} + a_{n-2}, & n \geq 2 \end{cases}.$$

The first few terms of the sequence are: 0, 1, 1, 2, 3, 5, 8, 13, ...

Today we try to find a explicit formula for a_n.

From high school we know both the arithmetic and geometric sequences:

Definition 1.2: Arithmetic Sequence
An arithmetic sequence is defined as follows:

$$a_n = \begin{cases} a, & n = 0 \\ a_{n-1} + d, & n \geq 1 \end{cases}.$$

For an arithmetic sequence we have the explicit formula $a_n = a + nd$.

Definition 1.3: Geometric Sequence
A geometric sequence is defined as follows:

$$a_n = \begin{cases} a, & n = 0 \\ a_{n-1} \cdot q, & n \geq 1 \end{cases}.$$

In this case, we aswell have an explicit formula: $a_n = a \cdot q^n$.

However looking at the fibonacci sequence, we see that it is diffrent from theese two. So we will do a mathematical trick to solve it. This trick is generalizing.

Definition 1.4:
Let $a_0, a_1 \in \mathbb{R}$ we define $a_n := a_{n-1} + a_{n-2}$ for $n \geq 2$.

Denote the sequence we get $\mathcal{F}_{a_0,a_1}$.

If $S = (s_0, s_1, \cdots)$ is any sequence of numbers, we say that S is Fibonacci if there exists $a_0, a_1 \in \mathbb{R}$ such that $S = \mathcal{F}_{a_0,a_1}$. Denote by `Fib` the set of all Fibonacci sequences.

`Fib` has some algebraic structures.

Claim 1.5:
If $\mathcal{F}' = (a_0, a_1, \cdots)$ and $\mathcal{F}'' = (b_0, b_1, \cdots)$ are Fibonacci then

$$\mathcal{F}' + \mathcal{F}'' := (a_0 + b_0, a_1 + b_1, \cdots, a_n + b_n, \cdots)$$

is also Fibonacci.

Proof. Lets write $c_0 := a_0 + b_0, c_1 := a_1 + b_1$ and $c_n := a_n + b_n$. Then $\mathcal{F}' + \mathcal{F}'' = (c_0, c_1, \cdots)$. To show this is a fibonacci sequence we need to show that $c_n = c_{n-1} + c_{n-2}$ for $n \geq 2$.

Indeed,

$$\begin{aligned} c_{n-1} + c_{n-2} &= (a_{n-1} + b_{n-1}) + (a_{n-2} + b_{n-2}) \\ &= (a_{n-1} + a_{n-2}) + (b_{n-1} + b_{n-2}) \\ &= a_n + b_n \\ &= c_n. \end{aligned}$$

□

This is somewhat special, as not every set of sequences is closed under addition. In fact, we showed that

$$\mathcal{F}_{a_0,a_1} + \mathcal{F}_{b_0,b_1} = \mathcal{F}_{a_0+b_0,a_1+b_1}.$$

But there is more structure.

Claim 1.6:
Let $\mathcal{A} = (a_0, a_1, \cdots)$ be a Fibonacci sequence and let $\alpha \in \mathbb{R}$. Define $\alpha\mathcal{A} := (\alpha a_0, \alpha a_1, \cdots)$. Then $\alpha\mathcal{A}$ is also Fibonacci. In fact, $\alpha\mathcal{F}_{a_0,a_1} = \mathcal{F}_{\alpha a_0, \alpha a_1}$.

Proof. We need to check that $\alpha a_n = \alpha a_{n-1} + \alpha a_{n-2}$ for $n \geq 2$. But this is "obviously" true, because we know that $a_n = a_{n-1} + a_{n-2}$. □

Note: The sequence $(0, 0, \cdots)$ is also Fibonacci, it is actually the sequence $\mathcal{F}_{0,0}$.

In fact, LINEAR COMBINATIONS of Fibonaccis are also Fibonacci.

$$\alpha\mathcal{F}_{a_0,a_1} + \beta\mathcal{F}_{b_0,b_1} = \mathcal{F}_{\alpha a_0 + \beta b_0, \alpha a_1 + \beta b_1},$$

for $\alpha, \beta \in \mathbb{R}$.

Corollary 1.7:
In order to find a formula for $\mathcal{F}_{a_0,a_1}$ it is enough to find a formula for $\mathcal{F}_{0,1}$ and $\mathcal{F}_{1,0}$.

Proof. This is because $\mathcal{F}_{a_0,a_1} = a_0\mathcal{F}_{1,0} + a_1\mathcal{F}_{0,1}$.

□

Out of curiosity, could it be that for some, very special $a_0, a_1 \neq (0, 0)$, the sequence $\mathcal{F}_{a_0,a_1}$ is arithmetic or geometric?

Since arithmetic sequences are characterized by a constant difference $a_n = a_{n-1} + d$, we would need $a_{n-1} + a_{n-2} = a_{n-1} + d$ for $n \geq 2$. This means that the sequence would need to be constant, i.e. $a_n = d$ for all $n \geq 1$. This cannot happen for any a_0, a_1 except $(0,0)$.

Perhaps it could be geometric? Let's try to see if $(1, q, \cdots)$ is Fibonacci for some $q \in \mathbb{R}$.

For such a sequence to be Fibonacci we need

$$q^n = q^{n-1} + q^{n-2} \text{ for } n \geq 2.$$

As $q \neq 0$ we get $q^2 = q + 1$. This is a quadratic equation with solutions

$$q = \frac{1 \pm \sqrt{5}}{2}.$$

Denote $\varphi := \frac{1+\sqrt{5}}{2}$ and $\psi := \frac{1-\sqrt{5}}{2}$.

We have found

$$\mathcal{F}_{1,\varphi} = (1, \varphi, \varphi^2, \varphi^3, \cdots)$$
$$\mathcal{F}_{1,\psi} = (1, \psi, \psi^2, \psi^3, \cdots)$$

Perhaps a linear combination of $\mathcal{F}_{1,\varphi}$ and $\mathcal{F}_{1,\psi}$ will give us $\mathcal{F}_{0,1}$?

We need to find $\alpha, \beta \in \mathbb{R}$ such that

$$\begin{cases} \alpha \cdot 1 + \beta \cdot 1 = 0 \\ \alpha\varphi + \beta\psi = 1 \end{cases}.$$

Solving this system of equations gives

$$\alpha = \frac{1}{\varphi - \psi} = \frac{1}{\sqrt{5}}, \quad \beta = \frac{1}{\psi - \varphi} = -\frac{1}{\sqrt{5}}.$$

So we found

$$\mathcal{F}_{0,1} = \frac{1}{\sqrt{5}}\mathcal{F}_{1,\varphi} - \frac{1}{\sqrt{5}}\mathcal{F}_{1,\psi}$$

$$= \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Lec 2

1.2 Logic and Math language

What is a mathematical/logical statement? A statement that can be either true or false, but not both.

Example 1.8:

A: "For every real number x, we have $x^2 \geq 0$".

B: "Every prime number $p \geq 3$ must be odd".

C: "Every odd number $p \geq 3$ must be prime".

A is True, B is True, C is False.

Negation of a statement: If A is a statement, then the statement "A is NOT ture" is called the negation of A, denoted by $\neg A$ ("not A").[1]

Such a statement can be represented in a truth table:

A	$\neg A$
T	F
F	T

[1] Sometimes we write $\bar{A}$ instead of $\neg A$.

Definition 1.9: And operation

Let A, B be statements. Then the statement $A \wedge B$ is the statement "A is true AND B is true". Sometimes we write $A\&B$ instead.

The corresponding truth table is:

A	B	$A \wedge B$
T	T	T
T	F	F
F	T	F
F	F	F

Example 1.10:

For real numbers x, ("x^2=1") $\wedge$ ("$x \geq 0$") is true. This is equivalent to "$x = 1$".

Similarly we can define the "or" operation.

Definition 1.11: Or operation

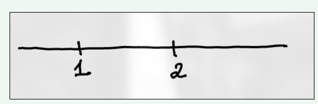
Let A, B be statements. Then the statement $A \vee B$ is the statement "A is true OR B is true", i.e. "at least one of the statements A or B are true". Sometimes we write $A|B$ instead.

Thus the corresponding truth table is:

A	B	$A \vee B$
T	T	T
T	F	T
F	T	T
F	F	F

Example 1.12:

("$x > 1$") $\vee$ ("$x < 2$"). We can draw a number line to visualize this situation:



We see that this statement is true for all $x \in \mathbb{R}$.

Example 1.13:

Consider the statement ("$x < 1$") $\vee$ ("$x > 2$"). Again looking at the number line, we see that this statement is equivalent to "$\neg(1 \leq x \leq 2)$".

Definition 1.14: Logical Implications

Let A, B be statements. We can form a new statement $A \Rightarrow B$ (A implies B). The statement "$A \Rightarrow B$" is "If A is true, then B is also true".

Depicted in a truth table:

A	B	$A \Rightarrow B$
T	T	T
T	F	F
F	T	T
F	F	T

As an exercise: Convice yourself that the statement $A \Rightarrow B$ is equivalent to $(\neg A \vee B)$.

Example 1.15:
("$0 = 1$") $\Rightarrow$ ("Earth is Flat") is a true statement. Notice that in logical terms, implies does not mean because A is true, B is true, no matter if there is no causality.

Similarly there is logical equivalence:

Definition 1.16: Logical Equivalence
$A \Leftrightarrow B$ is the statement $(A \Rightarrow B) \wedge (B \Rightarrow A)$. In diffrent words: A holds if and only if (iff) B holds.

We say that A and B are EQUIVALENT STATEMENTS if $A \Leftrightarrow B$ is true. This happens if A and B have the same truth table.

Example 1.17:

$$x^2 > 0 \Leftrightarrow x \neq 0,$$

$$0 = 1 \Leftrightarrow \text{Earth is flat.}$$

Let A, B be statements. Then there is a statement $A \Rightarrow B$ and $\neg B \Rightarrow \neg A$. These two statements are actually equivalent. This can be shown by truth tables.

As an exercise, state in words, why

$$\neg(\neg A) = A$$

and

$$\neg(A \wedge B) = (\neg A) \vee (\neg B)$$

.

Definition 1.18: Predicate
A PREDICATE is a statement involving some variables taken from a set.

$$P(x), P(x, y).$$

For example, $P(n) =$"$n = n^2$".

"for all integers $n : n = n^2$". This is definelty a false statement. Meanwhile, "There exists an integer $n : n = n^2$" is a true statement.

We denote $\forall$ for "for all/all" and $\exists$ for "there exists". The previous statements can be written as:

$$\forall n \in \mathbb{Z}, P(n) \quad , \quad \exists n \in \mathbb{Z}, P(n).$$

The symbols $\forall$ and $\exists$ are called QUANTIFIERS.

$\forall n \in \mathbb{Z} : n = n^2 \Rightarrow (n = 0) \vee (n = 1)$ is a true statement.
$\forall y \in \mathbb{R} : (y \geq 0) \Rightarrow (\exists x \in \mathbb{R} : x^2 = y)$ is true.

In practice, the second statement would be written as:

$$\forall 0 \leq y \in \mathbb{R}, \exists x \in \mathbb{R} s.t. x^2 = y.$$

Be careful about the order of quantifiers:

$$\forall x \in X, \exists y \in Y : A(x, y),$$

is not the same as

$$\exists y \in Y, \forall x \in X : A(x, y).$$

Example 1.19:
Let X be the set of students at ETH and Y the set of all courses given in HS25. Let $A(x, y)$ be the statement "student x attends course y". Then the first statement means "For every student, there is at least one course that the student attends". Which is true, whilst the second statement means "There is a course that every student attends", which is false.

In practice, instead of writing $\exists x \in X, \exists y \in X$, one rather writes $\exists x, y \in X$, similarly for $\forall$.

When negating quantifiers, we have the following rules:

$$\neg(\forall x \in X : A(x)) \Leftrightarrow \exists x \in X :\neq A(x).$$

$$\neg(\exists x \in X : A(x)) \Leftrightarrow \forall x \in X : \neg A(x).$$

There is also a quantifier "there exists a unique" denoted by $\exists!$. Lec 3

$$\exists! x \in X : A(x).$$

Example 1.20:
"$\exists! x \in \mathbb{R} : x^2 = 4$". This is a false statement as there exist two such x (namely 2 and -2).

"$\exists! x \in \mathbb{R} : x \geq 0 \wedge x^2 = 4$". This is a true statement as there exists exactly one such x (namely 2).

1.3 Set theory

The fundamental object in set theory is the set.

Definition 1.21: Set
A SET is a collection of different objects. These objects are called the ELEMENTS of the set.

For example, $M = \{1, 3, \text{unicorn}\} = \{3, 1, \text{unicorn}\}$ is a set with three elements. As we can see, the order of elements does not matter. Furthermore, repetition of elements does not matter, i.e. $\{1, 1, 1\} = \{1\}$.

If x is an element of a set M, we write $x \in M$. If it is not, we write $x \notin M$.

Another way to write sets is by specifying a property that the elements satisfy.

$$M = \{x : A(x)\} = \{x \mid A(x)\},$$

represents the set of all x such that $A(x)$ holds.

Example 1.22:

$$\{x \mid x \in \mathbb{Z} \wedge x^2 \leq 9\} = \{-3, -2, -1, 0, 1, 2, 3\}.$$

Another important set is the empty set $\{\}$, which contains no elements. It is denoted by $\varnothing$.

A set is allowed to contain other sets as elements one of its elements. For example $M = \{1, \{2, 3\}\}$ is a set with two elements, 1 and $\{2, 3\}$. Similarly, $\{\varnothing\}$ is a set with one element, the empty set.

Definition 1.23:

Let P, Q be sets.

1. We say that P is a SUBSET of Q if $x \in P \Rightarrow x \in Q$ We write $P \subseteq Q$ (Also: $P \subset Q$).
2. $P \subsetneq Q$ means $P \subseteq Q$ but $P \neq Q$.
3. $P \nsubseteq Q$ means $\neg(P \subseteq Q)$.

Careful: Every set is a subset of itself. Furthermore, the empty set $\varnothing \subseteq M$ for every set M.

We define the following operations for sets.

$$\begin{aligned}
P \cap Q &:= \{x : x \in P \wedge x \in Q\} && \text{(intersection)}\\
P \cup Q &:= \{x : x \in P \vee x \in Q\} && \text{(union)}\\
P \setminus Q &:= \{x : x \in P \wedge x \notin Q\} && \text{(complement of } Q \text{ in } P)\\
P \Delta Q &:= (P \cup Q) \setminus (P \cap Q) && \text{(symmetric difference)}
\end{aligned}$$

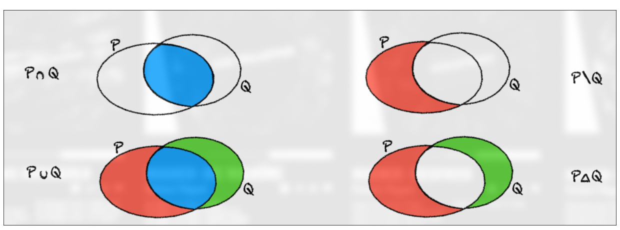


Figure 1: Set operations

Sometimes our sets will be subsets of some ambient/underlying set X.

In this case we define the COMPLEMENT of P in X as

$$P^c := X \setminus P = \{x \in X : x \notin P\}.$$

Definition 1.24:

Let $\mathcal{A}$ be a family of sets, $\mathcal{A} \neq \varnothing$. We can look at the union of the members of $\mathcal{A}$:

$$\bigcup_{A \in \mathcal{A}} A := \{x \mid \exists A \in \mathcal{A} \text{s.t.} x \in A\}.$$

Similarly, we can look at the intersection of the members of $\mathcal{A}$:

$$\bigcap_{A \in \mathcal{A}} A := \{x \mid \forall A \in \mathcal{A} : x \in A\}.$$

When working with big unions and intersections, it is often useful to translate them into set statements involving quantifiers, as this greatly improves intuition.

Example 1.25:

Let $\mathcal{A} = \{2, 3, 4, \cdots\}$.

Let $A_n = \{x \mid x \in \mathbb{Z} x \geq 1, n^2 | x\}$.

Then

$$\bigcup_{n \in \mathcal{A}} A_n = \{a \in \mathbb{Z} \mid \exists k \geq 2, \exists r \in \mathbb{Z} \text{s.t.} r \geq 1 \wedge a = k^2 r\}$$

and

$$\bigcap_{n \in \mathcal{A}} A_n = \{\}.$$

Definition 1.26: Cartesian Product

Let X, Y be sets. The CARTESIAN PRODUCT of X and Y is the set of ordered pairs (x, y).

$$X \times Y := \{(x, y) \mid x \in X, y \in Y\}.$$

Definition 1.27: n-Fold Cartesian Product

Let X be a set, $n \geq 1$. The n-FOLD CARTESIAN PRODUCT of X is defined as

$$X^n := \underbrace{X \times X \times \cdots \times X}_{n \text{ times}}.$$

An element of $X^n = (x_1, x_2, \cdots, x_n)$ is called an N-TUPLE.

The classic example is $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ – the Cartesian plane.

Definition 1.28: Power Set

Let X be a set. The POWER SET of X is the set of all subsets of X.

$$\mathcal{P}(X) := \{A \mid A \subseteq X\}.$$

Sometimes the notation 2^X is also used.

Example 1.29:

Let $X = \{1, 2, 3\}$. Then

$$\begin{aligned}
\mathcal{P}(X) = \{&\varnothing, \{1\}, \{2\}, \{3\},\\
&\{1, 2\}, \{1, 3\}, \{2, 3\},\\
&\{1, 2, 3\}\}
\end{aligned}$$

Notice that if X has n elements, then $\mathcal{P}(X)$ has 2^n elements.

The CARDINALITY of a set X is the number of elements in X. For example, if $X = \{1, 2, \cdots, n\}$, then $|X| = n$. This operation is only defined for finite sets at the moment.

1.4 Functions/Maps

Definition 1.30: Maps

Let X, Y be sets. A MAP $f : X \to Y$ is an assignment to *every* element $x \in X$ a uniquely defined element $f(x) \in Y$.

Lec 4

Roughly speaking, for an input $x \in X$, the map f produces an output output $f(x) \in Y$.

Given a map $f : X \to Y$, The set of inputs of a map is called the DOMAIN OF DEFINITION of f, whilst Y is called the TARGET of f.

Given two functions $f : X \to Y$ and $g : X' \to Y'$, we say that f and g are equal if $X = X'$, $Y = Y'$ and $\forall x \in X : f(x) = g(x)$.

Example 1.31:

$$f(x) := x^2.$$

This is lacking some information. Much better would be

$$X := \{x \in \mathbb{R} \mid 0 \leq x\}, \quad f : X \to \mathbb{R}, \quad f(x) := x^2.$$

Here is another function:

$$Z := \{x \in \mathbb{R} \mid -12 \leq x\}, \quad g : Z \to \mathbb{R}, \quad g(x) := x^2.$$

According to our definition, $f \neq g$ since $X \neq Z$.

Functions can be described by a formula, but do not necessarily have to be.

About notation: $f : X \to Y$ is the most standard notation. Sometimes $X \ni x \to f(x) \in Y$ is used.

Definition 1.32:

Let $f : X \to Y$ be a map. The IMAGE of f is the set

$$f(X) := \{y \in Y \mid \exists x \in X : f(x) = y\}.$$

Sometimes it's also written as $\text{image}(f)$. It holds that $f(X) \subseteq Y$.

Definition 1.33: Restriction

Let $f : X \to Y$, Let $A \subseteq X$. We can define a new map

$$f|_A : A \to Y, \quad f|_A(a) = f(a) \quad \forall a \in A$$

called the RESTRICTION of f to A.

A restriction restricts the domain of definition of a function.

Example 1.34: Characteristic Function

Let X be a set, $A \subseteq X$. The CHARACTERISTIC FUNCTION of A is defined as

$$\mathbb{1}_A : X \to \{0, 1\}, \quad \mathbb{1}_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

Example 1.35: Projection Function

Let X, Y be sets. The PROJECTION FUNCTION on X is defined as

$$\pi_X : X \times Y \to X, \quad \pi_X(x, y) := x.$$

Similarly, we can define $\pi_Y : Y \times X \to Y$.

Functions can have diffrent properties.

Definition 1.36:

Let $f : X \to Y$ be a map.

- f is called INJECTIVE if
$$\forall x_1, x_2 \in X : f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$
Equivalently, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$.
- f is called SURJECTIVE if
$$\forall y \in Y, \exists x \in X : f(x) = y.$$
Equivalently, $f(X) = Y$.
- f is called BIJECTIVE if it is both injective and surjective. Equivalently,
$$\forall y \in Y, \exists! x \in X : f(x) = y.$$

Definition 1.37:

Let $f : X \to Y$ be a bijection. The INVERSE map $f^{-1} : Y \to X$ is the map that assigns to every $y \in Y$ the unique $x \in X$ such that $f(x) = y$.

An example would be the map $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, f(x) = x^2$. Its inverse is $f^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, f^{-1}(y) = \sqrt{y}$.

Definition 1.38: Composition

Let $f : X \to Y, g : Y \to Z$ be maps. We define $g \circ f : X \to Z$ as the COMPOSITION of f and g:

$$g \circ f(x) := g(f(x)).$$

Somtimes this is written as

$$X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Example 1.39:

Let $f : X \to Y$ be a bijection and $f^{-1} : Y \to X$ its inverse map. We have $f \circ f^{-1} = \text{id}_Y$ and $f^{-1} \circ f = \text{id}_X$, where $\text{id}_X : X \to X, \text{id}_X(x) = x$ is the identity map on X.

The order of composition is important. In general, $f \circ g$ and $g \circ f$ are different maps. Even worse, one of them might not even be defined.

Lemma 1.40:

Let $f : X \to Y, g : Y \to Z$ be maps.

1. If f, g are injective, then $g \circ f$ is injective.
2. If f, g are surjective, then $g \circ f$ is surjective.
3. If f, g are bijective, then $g \circ f$ is bijective and
$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

Proof.

1. Let $x_1, x_2 \in X$ such that $(g \circ f)(x_1) = (g \circ f)(x_2)$. Then $g(f(x_1)) = g(f(x_2))$. Since g is injective, $f(x_1) = f(x_2)$. Since f is injective, $x_1 = x_2$. Thus $g \circ f$ is injective.
2. Let $z \in Z$. Since g is surjective, there exists $y \in Y$ such that $g(y) = z$. Since f is surjective, there exists $x \in X$ such that $f(x) = y$. Thus $(g \circ f)(x) = g(f(x)) = g(y) = z$. Hence $g \circ f$ is surjective.
3. Follows from 1. and 2.

□

This lemma is useful when showing a certain map is injective/surjective/bijective, as it allows us to work with simpler maps.

Definition 1.41: Image

Let $f : X \to Y$ be a map, $A \subseteq X$. Then
$$f(A) := \{y \in Y \mid \exists x \in A : f(x) = y\}$$
is called the IMAGE of A under f.

Definition 1.42: Inverse Image

Let $B \subseteq Y$. Define the INVERSE IMAGE of B under f as
$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

Lec 5

1.4.1 Graphs

Consider the following.

Definition 1.43:

Let $f : X \to Y$ be a map. The **graph** of f is the subset $\text{graph}(f) \subseteq X \times Y$ defined by
$$\text{graph}(f) := \{(x, f(x)) \mid x \in X\}.$$

Example 1.44:

If $X \subseteq \mathbb{R}$ and $Y = \mathbb{R}$, then the graph can be visualized in the Cartesian plane $\mathbb{R}^2$.

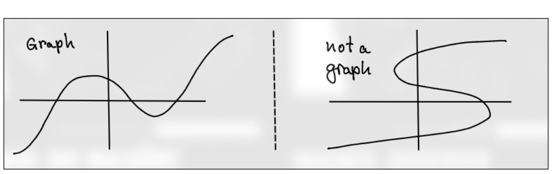


Figure 2: Graphs of some functions $f : \mathbb{R} \to \mathbb{R}$.

Sometimes injectivity and surjectivity can be read off from the graph.

2 Linear Systems of Equations

2.1 Fields

A FIELD (Körper) is a set K with two operations $+$ and $\cdot$. Together they form a commutative group $(K, +)$ with identity 0 and inverses $-a$ for $a \in K$. The operation $\cdot$ is associative and commutative with identity $1 \neq 0$ and inverses a^{-1} for $a \in K \setminus \{0\}$. The operations are linked by the distributive law
$$a \cdot (b + c) = a \cdot b + a \cdot c.$$
Furthermore, we ask that $1 \neq 0$.

Example 2.1: Examples of fields

$$\mathbb{Q} = \left\{a \mid a = \frac{m}{n}, m, n \in \mathbb{Z}, n \neq 0\right\}$$
with the usual addition and multiplication is a field.

$$\mathbb{R} = \text{the real numbers.}$$

$$\mathbb{F}_2 = \{0, 1\},$$
with the follwoing addition and multiplication tables:

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}.$$

Similarly, we can define $\mathbb{F}_3 = \{0, 1, 2\}$ and $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$, where the operations are defined modulo 3 and 5, respectively.

2.2 Systems of Linear Equations

Consider the following
$$\begin{cases} x + 3y + 4z &= -1 \\ 2x - y + z &= 5 \end{cases}.$$

To solve this, there exists a systematic method which is called ELIMINATION.

$$\begin{cases} x + 3y + 4z = -1 \\ 2x - y + z = 5 \end{cases} \xrightarrow[\to E_2]{-2E_1 + E_2} \begin{cases} x + 3y + 4z = -1 \\ -7y - 7z = 7 \end{cases}$$
$$\xrightarrow[\to E_2]{-\frac{1}{7}E_2} \begin{cases} x + 3y + 4z = -1 \\ y + z = -1 \end{cases}$$
$$\xrightarrow[\to E_1]{-3E_2 + E_1} \begin{cases} x + z = 2 \\ y + z = -1 \end{cases}$$

Staring at this for a while, we can see that z is a free variable. Thus, the set of solutions is given by:

choose z arbitrarily, $z = c \in \mathbb{R}$

then $y = -1 - c$ and $x = 2 - c$

$$(x, y, z) = (2 - c, -1 - c, c), c \in \mathbb{R} \subseteq \mathbb{R}^3.$$

A general system of many equations and matrix notation:

Fix a field $\mathbb{F}$. A system of m equations in n unknowns over $\mathbb{F}$ with unknowns $x_1, \dots, x_n$ is of the form

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

We call $a_{ij} \in \mathbb{F}$ the coefficients. Furthermore $b_i \in \mathbb{F}$. Such a system is called LINEAR because the unknowns appear only to the first power and are not multiplied together.

We notice the objects can be written as matrices and vectors.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

This object is called an $m \times n$ matrix with entries in $\mathbb{F}$.

Sometimes we write $A = (a_{ij})_{1 \le i \le m}^{1 \le j \le n}$.

Further we organize

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

These are $n \times 1$ VECTORS.

The system can now be written as

$$(S) : A \cdot x = b.$$

And the solution set as

$$L(S) = \{(x_1, x_2, \cdots) \mid x_i \in \mathbb{F} \forall 1 \le i \le n \wedge A \cdot x = b\}.$$

We define the matrix multiplication as follows:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix}.$$

If we now have $A \cdot x = b$, this is equivalent to the system of equations. We call this the MATRIX NOTATION.

Example 2.2:

The system from the beginning can be written as

$$\begin{pmatrix} 1 & 3 & 4 \\ 2 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}.$$

We can also write the EXTENDED MATRIX as

$$(A|b) = \left(\begin{array}{ccc|c} 1 & 3 & 4 & -1 \\ 2 & -1 & 1 & 5 \end{array}\right).$$

2.3 Elementary Row Operations

We define the following operations on matrices.

1. Choose $c \neq 0 \in \mathbb{F}$; multiply eq./row i by c.

   $(c \cdot R_i \to R_i)$

2. Choose $c \in \mathbb{F}$, choose $1 \le i, j \le m, i \neq j$; replace eq./row j by eq./row $j + c\cdot$ eq./row i.

   $(c \cdot R_i + R_j \to R_j)$

3. Interchange eq./row i and eq./row j.

   $(R_i \leftrightarrow R_j)$

Lec 6

Definition 2.3: Row-Equivalence

Let C and C' be $r \times s$ matrices with entries in $\mathbb{F}$. We say that C' is ROW-EQUIVALENT to C if C' can be obtained from C by a finite sequence of elementary row operations.

If we denote $C = C_0 \to C_1 \to \ldots \to C_k = C'$, then the following holds

Theorem 2.4:

Let $A \cdot x = b$ and $A' \cdot x = b'$ be two systems of m linear equations in n unknowns. Let's call first system (S) and second system (S'). Suppose the extended matrix $(A'|b')$ is row-equivalent to $(A|b)$. Then

$$L(S') = L(S).$$

Proof. We begin with the follwoing claim:

Claim 1: If S' is obtained form S by one elementary row operation, then $L(S) \subseteq L(S')$.

Let $(x_1, \cdots, x_n) \in L(S)$. We need to show that $(x_1, \cdots, x_n) \in L(S')$.

Operation 1: $c(\neq 0) \cdot R_i \to R_i$. In this case all the equations of S' coincide with the equations of S except equation i.

$$\begin{aligned} & a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \\ \xrightarrow{c \cdot R_i \to R_i} \quad & ca_{i1}x_1 + ca_{i2}x_2 + \ldots + ca_{in}x_n = cb_i. \end{aligned}$$

If $(x_1, \cdots, x_n)$ satisfies the first equation, it also satisfies the second equation, since we can multiply both sides of the first equation by c.

Operation 2: $R_j + c \cdot R_i \to R_j$. Again, all equations of S and S' coincide except equation j. We have

$$\begin{aligned} & a_{j_1}x_1 + \cdots + a_{j_n}x_n = b_j \\ \xrightarrow{R_j + c \cdot R_i \to R_j} \quad & (a_{j1} + ca_{i1})x_1 + \cdots + (a_{jn} + ca_{in})x_n = b_j + cb_i \\ & \Rightarrow (a_{j1}x_1 + \cdots + a_{jn}x_n) + c(a_{i1}x_1 + \cdots + a_{in}x_n) \\ & \quad = b_j + cb_i. \end{aligned}$$

So if $(x_1, \cdots, x_n)$ satisfies the first equation, it also satisfies the second equation, since we can add c times the i-th equation to the j-th equation.

Operation 3: $R_i \leftrightarrow R_j$. In this case, clearly we have that

$$(x_1, \cdots, x_n) \in L(S) \Rightarrow (x_1, \cdots, x_n) \in L(S').$$

This concludes the proof of Claim 1.

Claim 2: If S' is obtained from S by one elementary row operation, then S can be obtained from S' by one elementary row operation.

(1): If $S \xrightarrow[c \neq 0]{c \cdot R_i \to R_i} S'$, then $S' \xrightarrow{\frac{1}{c} \cdot R_i \to R_i} S$.

(2): If $S \xrightarrow[i \neq j]{R_j + c \cdot R_i \to R_j} S'$, then $S' \xrightarrow{R_j - c \cdot R_i \to R_j} S$.

(3): If $S \xrightarrow{R_i \leftrightarrow R_j} S'$, then $S' \xrightarrow{R_i \leftrightarrow R_j} S$.

This concludes the proof of Claim 2.

Claim 3: If S' is obtained from S by one elementary row operation, then $L(S) = L(S')$.

By Claim 1, we have $L(S) \subseteq L(S')$. By Claim 2, S is obtained by S' by one elementary row operation. Hence by Claim 1 again (with the roles of S and S' interchanged), we have $L(S') \subseteq L(S)$.

It follows that $L(S) = L(S')$.

Proof of Theorem: We are now in position to prove the theorem. By assumption, there is a finite sequence of elementary row operations

$$S = S_0 \to S_1 \to \cdots \to S_k = S'.$$

By Claim 3, we have $L(S_0) = L(S_1) = \cdots = L(S_k)$, i.e.

$$L(S) = L(S').$$

□

Definition 2.5: Row-Reduced Matrix

A $m \times n$ matrix A is called ROW-REDUCED if the following two conditions hold:

1. The first non-zero entry in each non-zero row of A is 1. This entry is called the LEADING ENTRY/PIVOT of that row.
2. Each column of A which contains the leading non-zero entry of some row has all its other entries equal to 0.

Example 2.6:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This matrix is row-reduced.

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

This matrix is not row-reduced, because the second condition is violated in column 3.

$$C = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix is not row-reduced, because the first condition is violated in row 1.

It might not be clear yet, why row-reduced matrices are important, however, we will see this soon. However, we will first prove the following theorem.

Theorem 2.7:

Every $m \times n$ matrix A is row-equivalent to a row-reduced $m \times n$ matrix.

Proof. Let $A = (a_{ij})_{1 \le i \le m,\, 1 \le j \le n}$, with $a_{ij} \in \mathbb{F}$.

If all entries in the 1st row of A are 0, the condition (1) is satisfied for row 1.

If row 1 does have a non-zero entry, let k be the smallest index j for which $a_{1j} \neq 0$. Multiply row 1 by a_{1k}^{-1} to make the leading entry 1. Now, condition (1) is satisfied for row 1.

For each row $2 \le i \le m$ add $(-a_{ik})$ times row 1 to row i. Formally

$$R_i + (-a_{ik}) \cdot R_1 \to R_i.$$

The result of these operations is a matrix A' which looks like

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & 0 & \cdots \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ik-1} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Notice that in row i of resulting matrix, the elements standing to the right of entry k (which is now 0) are unchanged. This is because to the left of the pivot in row 1, there are only 0s.

Summary: Condition (2) is now satisfied for the column of the pivot of row 1. Lec 7

We now turn to row number 2. If all elements in row 2 are 0, we just leave it as it is. In case not all elements are 0, we find the pivot in that row, say at entry $2, k'$.

Note that $k' \neq k$, because in column k of row 2, we have 0. Divide row 2 by the element lying at $(2, k')$ to make the pivot equal to 1.

Now we take row two and add suitable multiples of it to every other zero. Notice that for the critical row 1, this operation does not change the pivot, because $(2, k) = 0$. To proof that the pivot stays in row 1, we distinguish two cases:

If $k' < k$, the entry in the first row is already 0, so the multiple is $c = 0$.

If $k' > k$, by definition of k', all entries in row 2 to the left of k' and thus also k are 0. Thus they won't change.

We can continue this process for rows $3, 4, \ldots, m$. In the end, we arrive at a row-reduced matrix. □

Tip 2.8:

Only divide by pivot elements at the end of the algorithm, to avoid fractions.

Definition 2.9:

An $m \times n$ matrix is called ROW-REDUCED ECHELON if the following holds:

(a) It is row-reduced.

(b) Every row which has only 0 entries appears below every row which has a non-zero entry.

(c) If rows $1, \ldots, r$ are the non-zero rows, and if the pivot of row i occurs in column k_i, then

$$k_1 < k_2 < \cdots < k_r.$$

A row-reduced echelon matrix looks as follows:

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & * & 0 & * & 0 & * \\ 0 & \cdots & 0 & 0 & 0 & 1 & * & 0 & * \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 & * \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Theorem 2.10:

Every $m \times n$ matrix is row equivalent to a row reduced echelon matrix.

Proof. Apply the previous theorem + permutation between the rows. □

This is useful, for solving $A \cdot x = b$. Take $(A|b)$ and apply a sequence of row operations to bring it to row-reduced echelon form $(A'|b')$. Then $L(A|b) = L(A'|b')$.

Example 2.11:

We solve the following system of equations:

$$\begin{aligned} -9x_2 + 3x_3 + 4x_4 &= 9 \\ x_1 + 4x_2 - x_4 &= 5 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= -5. \end{aligned}$$

We transform this into an extended matrix

$$(A|b) = \begin{pmatrix} 0 & -9 & 3 & 4 & | & 9 \\ 1 & 4 & 0 & -1 & | & 5 \\ 2 & 6 & -1 & 5 & | & -5 \end{pmatrix}.$$

The corrseponding row reduced matrix is

$$(A'|b') = \begin{pmatrix} 0 & 1 & 0 & -\frac{5}{3} & | & \frac{12}{5} \\ 1 & 0 & 0 & \frac{17}{3} & | & \frac{-23}{5} \\ 0 & 0 & 1 & -\frac{11}{3} & | & \frac{51}{5} \end{pmatrix}.$$

Thus the row reduced echelon is

$$\begin{pmatrix} 1 & 0 & 0 & \frac{17}{3} & | & \frac{-23}{5} \\ 0 & 1 & 0 & -\frac{5}{3} & | & \frac{12}{5} \\ 0 & 0 & 1 & -\frac{11}{3} & | & \frac{51}{5} \end{pmatrix}.$$

We can see that we can take $x_4 \coloneqq a \in \mathbb{R}$ as a free variable. Then we have

$$x_3 = \frac{11}{3}a + \frac{51}{5}, \quad x_2 = \frac{5}{3}a + \frac{12}{5}, \quad x_1 = -\frac{17}{3}a - \frac{23}{5}.$$

It may happen that the system has a row with only zeros. For example

$$0 = b_3 - b_2 + 2b_1.$$

Then if $b_3 - b_2 + 2b_1 \neq 0$, the system has no solution. Otherwise, the system is equivalent to the system without that equation.

3 Vector Spaces

Vector spaces are essentially the playground of linear algebra, the following definition is thus very central for the subject.

Definition 3.1: Vector Space

A Vector Space over a field $\mathbb{F}$ (scalars) is a set V (vecotrs) endowed with two operations:

$$\begin{aligned} + &: V \times V \to V, (v_1, v_2) \mapsto v_1 + v_2 \\ \cdot &: \mathbb{F} \times V \to V, (\lambda, v) \mapsto \lambda \cdot v. \end{aligned}$$

such that the following axioms are satisfied:

V1-4) $(V, +)$ is an abelian group.

V5) $\forall a, b \in \mathbb{F}, \forall v \in V : a \cdot (b \cdot v) = (ab) \cdot v$

V6) $\forall v \in V, 1 \cdot v = v$

V7) $\forall a \in F, v_1, v_2 \in V : a \cdot (v_1 + v_2) = a \cdot v_1 + a \cdot v_2$

V8) $\forall a_1, a_2 \in \mathbb{F}, v \in V : (a_1 + a_2) \cdot v = a_1 \cdot v + a_2 \cdot v$

Tip 3.2:

Vector spaces are essentially ordered lists with component-wise operations.

Theese lists don't have to be finite.

Example 3.3: Coordinate Space

The coordinate space K^n for $n \in \mathbb{N}_0$ then

$$\underbrace{K^n}_{K \times K \cdots \times K} = \{(a_1, \ldots, a_n) \mid a_i \in K, 1 \leq i \leq n\}.$$

We turn $V = K^n$ into a vector space. Let $v = (a_1, \ldots, a_n), w = (b_1, \ldots, b_n) \in V$. Then

$$v + w = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n) \in V.$$

Let $a \in K$, then

$$a \cdot v \coloneqq (a \cdot a_1, a \cdot a_2, \ldots, a \cdot a_n) \in V.$$

$$0 \coloneqq (0, 0, \ldots, 0) \in V.$$

As an exercise, show that K^n with theese operations is a vector space.

Example 3.4: Matrix Space

Consider $M_{m \times n}$. The elemnts are

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad a_{ij} \in K.$$

This is basiclly the same as $K^{m \cdot n}$.

Lemma 3.5:
Let V be a vector space over K. Then:

(a) The zero vector is unique.

(b) Let $v \in V$. The element $v' \in V$ with $v + v' = 0$ is unique. It is denoted by $-v$.

(c) $\forall v \in V : 0 \cdot v = 0_V$

(d) $\forall a \in K : a \cdot 0_V = 0_V$

(e) $\forall v \in V : (-1) \cdot v = -v$

(f) $\forall v \in V, -(-v) = v$

(g) If $a \cdot v = 0$ for some $a \in K, v \in V$, then either $a = 0$ or $v = 0_V$.

Lec 8

Proof. **(a)**: Since $(V, +)$ is an abelian group, there is an unique identity element, which we denote by 0_V.

(b): Again, since $(V, +)$ is an abelian group, every element has an unique inverse.

(c): Let $v \in V$. We have to show that $0 \cdot v = 0_V$. Indeed,

$$0 \cdot v = (0 + 0) \cdot v \overset{\text{V8}}{=} 0 \cdot v + 0 \cdot v.$$

Add $(-0 \cdot v)$ to both sides. This yields

$$0_V = 0 \cdot v + (0 \cdot v + (-0 \cdot v)) = 0 \cdot v + 0_V = 0 \cdot v.$$

(g): Let $a \in K, v \in V$ such that $a \cdot v = 0_V$. We need to prove that either $a = 0$ or $v = 0_V$. Indeed, assume that $a \neq 0$. By the axioms of a field, $\exists$ an element $a^{-1} \in K$ such that $a^{-1}a = 1 = aa^{-1}$.

Now:

$$v \overset{\text{V6}}{=} 1 \cdot v = (a^{-1}a) \cdot v \overset{\text{V5}}{=} a^{-1} \cdot (a \cdot v) = a^{-1} \cdot 0_V \overset{\text{(d)}}{=} 0_V.$$

$\square$

Definition 3.6: Linear Subspace
Let V be a vector space over K. A subset $W \subseteq V$ is called a LINEAR SUBSPACE (Unterraum) of V if the following holds:

LSS1) $W \neq \varnothing$,

LSS2) $\forall w_1, w_2 \in W : w_1 + w_2 \in W$,

LSS3) $\forall w \in W, \forall a \in K : a \cdot w \in W$.

Lemma 3.7:
Let V be a vector space over K and $W \subseteq V$ a subspace. Then W is a vector space on its own when endowed with the operations coming from V.

Proof. From LSS2 and LSS3, it follows that the operations $+$ and $\cdot$ from V indeed define such operations on W.

Now we have to check the axioms V1-V8. This can be done as an exercise.

For example, for V2, Pick any $w \in W$ (Possible by LSS1). We have

$$0_V \overset{(c)}{=} 0_K \cdot w \overset{\text{LSS3}}{\in} W.$$

Thus $0_V \in W$. Take now $0_W := 0_V$. Clearly, $\forall w \in W$ we have

$$0_W + w = 0_V + w \overset{\text{V2}}{=} w.$$

$\square$

The above definition is often cumbersome to use. The following lemma gives a more handy criterion to check if a subset is a subspace.

Lemma 3.8: Criterion for Subspaces
Let V be a vector space ove K and $W \subseteq V$ a subset. Then W is a subspace of V iff the following holds:

(1) $0_V \in W$,

(2) $\forall a_1, a_2 \in K, \forall w_1, w_2 \in W : a_1 \cdot w_1 + a_2 \cdot w_2 \in W$.

Proof. $\Rightarrow$: Assume that W is a subspace of V. This follow at once from the fact that W is in itself a vector space.

$\Leftarrow$: Assume that (1) and (2) hold. We have to show that W is a subspace. Clearly, LSS1 holds by (1). To show LSS2, pick $w_1, w_2 \in W$. Then by (2) (with $a_1 = a_2 = 1$), we have

$$w_1 + w_2 = 1 \cdot w_1 + 1 \cdot w_2 \in W.$$

Similarly for LSS3, pick $w \in W, a \in K$. Then by (2) (with $a_1 = a, a_2 = 0, w_1 = w, w_2 = 0_V$), we have

$$a \cdot w = a \cdot w + 0 \cdot 0_V \in W.$$

$\square$

Example 3.9: Trivial Subspace
Let V be a vector space over K. Then

$$\{0_V\} \subseteq V,$$

is a subspace of V. It is called the ZERO SUBSPACE of V.

Example 3.10:
Fix $b \in K$. Consider the subset

$$\mathcal{U}_b := \{(x_1, x_2, x_3) \in K^3 \mid x_1 + x_2 + x_3 = b\} \subseteq K^3.$$

Claim: $\mathcal{U}_b$ is a subspace of K^3 iff $b = 0$.

Proof. $\Rightarrow$: Suppose $\mathcal{U}_b$ is a linear subspace of K^3. If (x_1, x_2, x_3) and $(y_1, y_2, y_3) \in \mathcal{U}_b$, then

$$x_1 + x_2 + x_3 = b = y_1 + y_2 + y_3.$$

Since $\mathcal{U}_b$ is a linear subspace, we have

$$(x_1 + y_1, x_2 + y_2, x_3 + y_3) \in \mathcal{U}_b.$$

And thus it follows that

$$(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = b.$$

But from before, we get that this sum is equal to $2b$. Thus $2b = b$, which implies that $b = 0$.

The other direction is left as an exercise. $\square$

Let A be an $m \times n$ matrix with entries in K. From now on we abbreviate this with

$$A \in \mathrm{Mat}_{m \times n}(K) \text{ or } A \in M_{m \times n}(K).$$

Fix $b \in K^m$ such that $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ is viewed as a column vector. Consider the subset of K^n defined by

$$L := \{x \in K^n | A \cdot x = b\}.$$

Claim 3.11:
$L \subseteq K^n$ is a linear subspace iff $b = 0_V$.

For the proof we need preparations.

Lemma 3.12:
$\forall a \in K, u, v \in K^n$ we have

(1) $A \cdot (u + v) = A \cdot u + A \cdot v$,

(2) $A \cdot (a \cdot u) = a \cdot (A \cdot u)$.

Proof. Write $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

$$A \cdot u + A \cdot v =$$
$$\begin{pmatrix} \vdots \\ a_{i1}u_1 + a_{i2}u_2 + \cdots + a_{in}u_n \\ \vdots \end{pmatrix} + \begin{pmatrix} \vdots \\ a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n \\ \vdots \end{pmatrix}$$
$$= \begin{pmatrix} \vdots \\ a_{i1}(u_1 + v_1) + a_{i2}(u_2 + v_2) + \cdots + a_{in}(u_n + v_n) \\ \vdots \end{pmatrix}$$
$$= A \cdot \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}$$
$$= A \cdot (u + v).$$

The second statement is left as an exercise. □

Lec 9

Proof. [Claim] $\Leftarrow$: Assume $b = 0$, we want to show that L is a subspace. Indeed $0 \in L$ because $A \cdot 0 = 0$.

If $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in L$, then

$$A \cdot (x + y) \overset{(1)}{=} A \cdot x + A \cdot y = 0 + 0 = 0.$$

Thus $x + y \in L$.

Let $x \in L, a \in K$. Then

$$A \cdot (a \cdot x) \overset{(2)}{=} a \cdot (A \cdot x) = a \cdot 0 = 0.$$

Thus $a \cdot x \in L$.

Together theese three properties show that $L \subseteq K^n$ is a subspace.

$\Rightarrow$: Assume that L is a subspace, and we we'll show that $b = 0$. Indeed since $L \subseteq K^n$ is a subspace, $0_V \in L$. Thus $A \cdot 0_V = b$. But $A \cdot 0_V = 0_V$. Thus $b = 0_V$.

□

Example 3.13:
Let K be a field. Define

$$K^\infty := \{(a_1, a_2, \ldots) \mid a_i \in K, i \in \mathbb{N}\}.$$

We turn K^∞ into a vector space as follows:

$$\begin{aligned}(a_1, a_2, \ldots) + (b_1, b_2, \ldots) &:= (a_1 + b_1, a_2 + b_2, \ldots) \\ c \cdot (a_1, a_2, \ldots) &:= (c \cdot a_1, c \cdot a_2, \ldots)\end{aligned}$$
.

Recalling the first lecture, taking $K = \mathbb{R}$, we have

$$\mathrm{Fib} := \{(a_1, a_2, \ldots) \mid a_k = a_{k-1} + a_{k-2}, k \ge 3\}.$$

As an exercise, show that Fib is a subspace of $\mathbb{R}^\infty$. More generally, fix $\alpha, \beta \in \mathbb{R}$. Then

$$\mathcal{U}_{\alpha,\beta} := \{(a_1, a_2, \ldots) \in \mathbb{R}^\infty \mid a_k = \alpha a_{k-1} + \beta a_{k-2}, k \ge 3\},$$

is a subspace of $\mathbb{R}^\infty$.

3.1 Spaces of Functions

Fix a field K and S is a non-empty set. Define

$$K^S := \{f : S \to K\}.$$

This denotes the set of all functions from S to K. We define the following operations on K^S:

$$\begin{aligned}(f + g)(s) &:= f(s) + g(s) \\ (a \cdot f)(s) &:= a \cdot (f(s)).\end{aligned}$$

With these operations, K^S is a vector space over K. (Exercise)

Exercise 3.14:
Let $S = [0, 1]$. Define $C(S) := \{f \in \mathbb{R}^S \mid f\}$ such that f is continuous. Then $C(S)$ is a subspace of $\mathbb{R}^S$.

3.2 Polynomials

Let K be a field. Fix a letter called FORMAL VARIABLE x. A POLYNOMIAL over K (or polynomial with coefficients in K) is an expression of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

where $n \in \mathbb{N}_0, a_i \in K \,\forall 0 \le i \le n$. We allow to omit 0-terms like $0 \cdot x^i$.

We call $f(x) = 0$ the ZERO POLYNOMIAL.

We can add and multiply polynomials as follows: Consider polynomials a and b we take $r = \max(n, m)$ and define

$$(a_0 + a_1 x + \cdots + a_n x^n) + (b_0 + b_1 x + \cdots + b_m x^m) := (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_r + b_r)x^r.$$

The multiplication is defined as

$$c \cdot (a_0 + a_1 x + \cdots + a_n x^n) := (c \cdot a_0) + (c \cdot a_1)x + \cdots + (c \cdot a_n)x^n.$$

We write $K[x]$ for the set of all polynomials with coefficients in K.

Claim 3.15:
$K[x]$ is a vector space over K with the above operations.

Every $f(x) \in K[x]$ defines also a function $\tilde{f} : K \to K$. The subset of K^K which comes from polynomials is a subspace of K^K.

But: from $\tilde{f}$ we cannot always recover f.

Example 3.16:

Take $K = \mathbb{F}_2 = \{0,1\}$. Consider $f(x) = x$ and $g(x) = x^2$. Then theese are diffrent polynomials, but $\tilde{f} = \tilde{g}$.

Definition 3.17: Degree of a polynomial

Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in K[x]$ with $a_n \neq 0$. Then we define the DEGREE of f as $\deg(f) = n$. We define $\deg(0) = -\infty$.

Fix $d \in \mathbb{N}_0 \cup \{-\infty\}$. Define $K[x]_d$ to be all the polynomials of degree at most d. Then $K[x]_d$ is a subspace of $K[x]$.

Question 3.18:

Is $\{f(x) \in k[x] \mid \deg(f) = d\}$ also a subspace of $K[x]$?

Cosnider $\{p \in k[x]_5 \mid \sum_{n=0}^{\infty} a_n = 0\}$. Is this a subspace of $K[x]_5$?

Spaces of matrices: Consider $M_{m\times n}(K)$ to be the set of all $m \times n$ matrices with entries in K. Define $+$ and $\cdot$ as

$$(a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij})$$
$$c \cdot (a_{ij}) := (c \cdot a_{ij}).$$

With theese operations, $M_{m\times n}(K)$ is a vector space over K.

As a motivation for the next few lectures: Let V be a vector space, let $S \subseteq V$ be a subset. What is the smallest subspace of V containing S?

Lec 10

3.3 Span

Consider the following lemma

Lemma 3.19:

Let V be a vector space and let $\{W_i\}_{i\in I}$ be a family of subspaces of V (i.e. $W_i \subseteq V$ is a subspace for all $i \in I$). Then the intersection

$$W := \bigcap_{i\in I} W_i$$

is a subspace of V.

Proof. Note that $0_V \in W_i$ for all $i \in I$, thus $0_V \in W$.

Let now $a, b \in K, v, w \in W$. We'll show that $a \cdot v + b \cdot w \in W$.

Let $j \in I$ be any index in I. Since $v, w \in W$, they must also be in W_j. Since W_j is a subspace of V, it follows that

$$a \cdot v + b \cdot w \in W_j.$$

But this holds $\forall j \in I$, therefore,

$$a \cdot v + b \cdot w \in \bigcap_{i\in I} W_i = W.$$

By a previous lemma, this shows that W is a subspace of V. □

Let $S \subseteq V$ be a subset.

Definition 3.20: Span

Define the SPAN of S as

$$\mathrm{Sp}(S) := \bigcap_{W\in\mathcal{N}} W,$$

where

$$\mathcal{N} := \{W \subseteq V \mid W \text{ is a subspace of } V, S \subseteq W\}.$$

Note that $\mathcal{N} \neq \varnothing$ since $V \in \mathcal{N}$.

By the previous lemma, $\mathrm{Sp}(S)$ is a subspace of V. Sometimes, we also call it the subspace of V generated by S.

Lemma 3.21:

Among all subspaces of V that contain S, $\mathrm{Sp}(S)$ is the smallest one, i.e.,

1. $\mathrm{Sp}(S) \subseteq V$ is a subspace of V.
2. If $W \subseteq V$ is a subspace and $W \supseteq S$, then $W \supseteq \mathrm{Sp}(S)$.

Proof. 1. This was shown above.

2. Let $W \subseteq V$ be a subspace with $W \supseteq S$. By definition of $\mathcal{N}$, we have $W \in \mathcal{N}$. Thus by definition of $\mathrm{Sp}(S)$, we have

$$\mathrm{Sp}(S) = \bigcap_{U\in\mathcal{N}} U \subseteq W.$$

□

3.3.1 Linear Combinations

This definition is good from a theoretical point of view, but not very handy when we actually want to compute $\mathrm{Sp}(S)$. We thus introduce the following concept.

Definition 3.22: Linear Combination

Let $n \in \mathbb{Z}_{\geq 1}$, $a_1, \ldots, a_n \in K$ and $v_1, \ldots, v_n \in V$. A LINEAR COMBINATION of the vectors $v_1, \ldots, v_n$ with coefficients $a_1, \ldots, a_n$ is the vector

$$a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n.$$

We only look at finite linear combinations, i.e., n is always finite.

Lemma 3.23:

Let $\varnothing \neq S \subseteq V$. Then

$$\mathrm{Sp}(S) = \{a_1 v_1 + \cdots + a_n v_n \mid n \in \mathbb{Z}_{\geq 1}, v_i \in S, a_i \in K\}.$$

i.e. $\mathrm{Sp}(S)$ is the subset of V obtained by taking all possible linear combinations of vectors in S.

Proof. Denote $\widetilde{\mathrm{Sp}}(S) := \{a_1 v_1 + \cdots + a_n v_n \mid n \in \mathbb{Z}_{\geq 1}, v_i \in S, a_i \in K\}$. We will show that $\widetilde{\mathrm{Sp}}(S) = \mathrm{Sp}(S)$.

We first show that 1) $S \subseteq \widetilde{\mathrm{Sp}}(S)$, 2) $\widetilde{\mathrm{Sp}}(S)$ is a subspace of V and 3) $\forall$ subspace $W \subseteq V$ with $S \subseteq W$ we have $\widetilde{\mathrm{Sp}}(S) \subseteq W$.

Note that this is useful since 1)+2) would imply that $\mathrm{Sp}(S) \subseteq \widetilde{\mathrm{Sp}}(S)$ by the previous lemma, and 3) would imply $\widetilde{\mathrm{Sp}}(S) \subseteq \bigcap_{W\in\mathcal{N}} W = \mathrm{Sp}(S)$. Together, this implies $\widetilde{\mathrm{Sp}}(S) = \mathrm{Sp}(S)$.

It remains to prove 1), 2) and 3).

1) If $v \in S$, then $v = 1 \cdot v \in \widetilde{\mathrm{Sp}}(S)$.

2) We'll show that $0_V \in \widetilde{\mathrm{Sp}}(S)$ and that $\forall \alpha, \beta \in K, \forall v, w \in \widetilde{\mathrm{Sp}}(S)$: $\alpha \cdot v + \beta \cdot w \in \widetilde{\mathrm{Sp}}(S)$.

Clearly, $0_V = 0 \cdot v \in \widetilde{\mathrm{Sp}}(S)$ for any $v \in S$.

Write $v = a_1 v_1 + \cdots + a_n v_n$ and $w = b_1 w_1 + \cdots + b_m w_m$. Then

$$\begin{aligned}\alpha \cdot v + \beta \cdot w &= \alpha(a_1 v_1 + \cdots + a_n v_n) + \beta(b_1 w_1 + \cdots + b_m w_m) \\ &= (\alpha a_1) v_1 + \cdots + (\alpha a_n) v_n + (\beta b_1) w_1 + \cdots + (\beta b_m) w_m.\end{aligned}$$

This is a linear combination of vectors in S, thus $\alpha \cdot v + \beta \cdot w \in \widetilde{\mathrm{Sp}}(S)$.

3) Let $W \subseteq V$ be a subspace with $S \subseteq W$. Let v be in $\widetilde{\mathrm{Sp}}(S)$. We need to prove that $v \in W$. Indeed, write $v = a_1 v_1 + \cdots + a_n v_n$ with $v_i \in S$. By assumption, $S \subseteq W$, thus $v_i \in W$. As W is a linear subspace, it follows that $v \in W$. □

We have shown that the span of a subset S of a vector space V can be written in two ways. The first one is more theoretical, the second one more practical.

The span of the empty set is $\{0_V\}$. From now on, let's agree that linear combinations of elements of the empty set is just 0_V.

As a notation: Many times $S = \{v_1, \ldots, v_n\}$ is a finite set. We write $\mathrm{Sp}\{v_1, \ldots, v_n\}$ or $\mathrm{Sp}(v_1, \ldots, v_n)$ instead of $\mathrm{Sp}(\{v_1, \ldots, v_n\})$.

Definition 3.24: Generating/Spanning Set

Let V be a vector space and $S \subseteq V$ a subset. We say that S GENERATES or SPANS V if $\mathrm{Sp}(S) = V$.

If $W \subseteq V$ is a subspace, we say that S generates or spans W if $\mathrm{Sp}(S) = W$.

Definition 3.25: Finite Dimensional Vector Space

A vector space V is FINITE DIMENSIONAL if there exists a finite set $S \subseteq V$ that generates V.

If such a finite S does not exist, then V is called INFINITE DIMENSIONAL.

Example 3.26: Subspaces of $\mathbb{R}^2$

Let $V = \mathbb{R}^2, K = \mathbb{R}$. Some subspaces of V are:

1. $\{0_{\mathbb{R}^2}\} = \{(0,0)\} = \mathrm{Sp}(\varnothing) = \mathrm{Sp}\{0_{\mathbb{R}^2}\}$
2. Let $0 \neq v = (a,b) \in \mathbb{R}^2$. Then

$$\mathrm{Sp}\{v\} = \{\alpha \cdot (a,b) \mid \alpha \in \mathbb{R}\} = \{(\alpha a, \alpha b) \mid \alpha \in \mathbb{R}\}.$$

Geometrically speaking, this is the line through the origin and v.

3. Let $0 \neq v_1, w \in \mathbb{R}^2$ If $\mathrm{Sp}\{v\} = \mathrm{Sp}\{w\}$, this is equivalent to saying that $\exists \alpha \in \mathbb{R} \setminus \{0\}$ such that $w = \alpha v$.
4. Let $W \subseteq \mathbb{R}^2$ be a subspace and assume that $W = \mathrm{Sp}\{v, w\}$ where $v \neq 0, w \notin \mathrm{Sp}\{v\}$. Then $W = \mathbb{R}^2$.

Together this tells us that there are only three types of subspaces of $\mathbb{R}^2$: the zero subspace, lines through the origin, and the whole space $\mathbb{R}^2$.

Question: Let $v_1, \ldots, v_n \in V$ and $w \in K^m$. How can we determine whether or not $w \in \mathrm{Sp}\{v_1, \ldots, v_n\}$?

e.g. let $v_1 = (1,3), v_2 = (7,73)$ and $w = (11,137)$. In this case yes, $w \in \mathrm{Sp}\{v_1, v_2\}$ as $-3 \cdot v_1 + 2 \cdot v_2 = w$. Lec 11

Solution: Take the column vectors $v_1, \ldots, v_n$ and make a $m \times n$ matrix out of them:

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

If $x = (x_1, \ldots, x_n) \in K^n$, then

$$A \cdot x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = \sum_{i=1}^n x_i v_i.$$

Given $w \in K^m$, the question is wheter or not the system of equations $A \cdot x = w$ has a solution. If a solution exists, then $w \in \mathrm{Sp}\{v_1, \ldots, v_n\}$. If a solution does not exist, then $w \notin \mathrm{Sp}\{v_1, \ldots, v_n\}$.

In the first case, any solution x will give us a choice of coefficients $x_1, \ldots, x_n$ such that $w = \sum_{i=1}^n x_i v_i$.

Example 3.27:

1) Let $V = K^n$. Denote by $e_1 := (1,0,0,\ldots,0), e_2 := (0,1,0,\ldots,0), \ldots, e_n := (0,0,0,\ldots,1)$. Then

$$\mathrm{Sp}\{e_1, e_2, \ldots, e_n\} = K^n.$$

2) Let $W = k[x]_d$ (polynomials in x of degree at most d). Then

$$W = \mathrm{Sp}\{1, x, x^2, \ldots, x^d\}.$$

Similarly, take $V = K[x]$. Then

$$V = \mathrm{Sp}\{1, x, x^2, x^3, \ldots\} \text{ (infinite set) }.$$

3) Let $M = M_{m \times n}(K)$ be the set of all $m \times n$ matrices. For every $1 \leq i \leq m, 1 \leq j \leq n$, define the matrix

$$E_{ij} := \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}.$$

i.e. the matrix with a 1 in the (i,j)-th position and 0 elsewhere.

Then

$$M = \mathrm{Sp}\{E_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}.$$

All of these examples are finite dimensional vector spaces except $K[x]$. As an exercise, show that $K[x]$ is infinite dimensional.

3.4 Linear Independence

Let V be a vector space over K

Definition 3.28:

Let $v_1, \ldots, v_n$ be a list of n vectors in V. We say that that $v_1, \ldots, v_n$ are LINEARLY INDEPENDENT if the only linear combination of $v_1, \ldots, v_n$ that equals 0_V is the one with all coefficients equal to zero.

If the vectors are not linearly independent, they are called LINEARLY DEPENDENT.

Alternatively, $v_1, \ldots, v_n$ are linearly independent iff

$\forall a_1, \ldots, a_n \in K$ for which $a_1 v_1 + \cdots + a_n v_n = 0_V$, we must have $a_1 = a_2 = \cdots = a_n = 0$.

Let's agree that $\varnothing$ is linearly independent.

Remark 3.29:

1) 0_V can always be written as a linear combination of any list of vectors as $0_V = 0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n$. This is called the TRIVIAL linear combination.

2) Suppose that in our list there is a repetition, say $v_i = v_j$ for some $i \neq j$. Then the list is linearly dependent, since

$$0_V = 1 \cdot v_i + (-1) \cdot v_j + 0 \cdot v_1 + \cdots + 0 \cdot v_n.$$

3) The order of the elements in the list does not matter.

4) If 0 is one of the elements in the list, then this list cannot be linearly independent.

As a generalization for families of vectors consider the following definition.

Definition 3.30:

Let $\mathbb{F} = \{v_i\}_{i \in \mathcal{I}}$ be a family of vectors in V. We say that $\mathbb{F}$ is LINEARLY INDEPENDENT if $\forall n \in \mathbb{Z}_{\geq 1}$ and any sequence of distinct indices $i_1, \ldots, i_n \in \mathcal{I}$, the list of vectors $v_{i_1}, \ldots, v_{i_n}$ is linearly independent.

Let $\varnothing \neq S \subseteq V$ be a subset. We say that S is linearly independent if every finite list of distinct vectors in S is linearly independent.

Lemma 3.31:

If $\varnothing \neq S \subseteq V$, is linearly independent, then every subset of it is also linearly independent.

Proof. Assume $\exists a_1, \ldots, a_n \in K$ and $v_1, \ldots, v_n \in W \subseteq S$ such that $a_1 v_1 + \cdots + a_n v_n = 0_V$ with not all a_i equal to zero. Since $W \subseteq S$, this contradicts the linear independence of S. $\square$

Example 3.32:

1) $e_1, \ldots, e_n \in K^n$ are linearly independent.

Solution. Suppose $a_1 e_1 + \cdots + a_n e_n = 0$. This means that

$$a_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + a_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

But since we are working component-wise, this means that

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In particular, $a_i = 0$ for all $1 \leq i \leq n$.

2) The set $\{e_1, e_2, \cdots\} \subseteq K^\infty$ is linearly independent.

3) The set $\{1, x, x^2, x^3, \ldots, x^d\} \subseteq K[x]_d$ is linearly independent.

4) The set $\{1, x, x^2, x^3, \ldots\} \subseteq K[x]$ is linearly independent.

5) Let V be a vector space, let $v \in V$. Show that $\{v\}$ is linearly independent iff $v \neq 0_V$.

3.5 Basis

The following is one of the most central definitions in linear algebra.

Definition 3.33: **Basis**

A subset $S \subseteq V$ is called a BASIS of V if the following holds:

1) S is linearly independent.

2) S spans V, i.e. $\mathrm{Sp}(S) = V$.

Proposition 3.34:

A subset $S \subseteq V$ is a basis of V iff every $v \in V$ can be written in a *unique* way as a linear combination of vectors from S.

Lec 12

The meaning of "unique" in the above proposition is the following: suppose $v \in V$ is written as $v =$ linear combination #1 with vectors from S, and also as $v =$ linear combination #2 with vectors from S.

Define C_1 to be all the vectors that appear in linear combination #1 without zero coefficient. Similarly, define C_2.

Then $C_1 = C_2$ and $\forall u \in C_1$, the coefficient of u in #1 equals the coefficient of u in #2.

In case, S is a finite set, say $S = \{u_1, \ldots, u_n\}$, then uniqueness means that if $v = a_1 u_1 + \cdots + a_n u_n = b_1 u_1 + \cdots + b_n u_n$, then $a_i = b_i$ for all $1 \leq i \leq n$.

Proof. [of proposition] For simplicity assume that S is finite. (The general case is an exercise.)

$\Leftarrow$ Assume every $v \in V$ can be written in a unique way. Because every $v \in V$ can be written, this implies that $\mathrm{Sp}(S) = V$. We'll show now that S is linearly independent.

Write $S = \{v_1, \dots, v_n\}$. Consider the 0_V. Clearly, $0 = 0 \cdot v_1 + \cdots + 0 \cdot v_n$. But by assumption, there is a unique linear combination of $v_1, \dots, v_n$ that equals 0_V. It follows that if

$$a_1 v_1 + \cdots + a_n v_n = 0_V.$$

Then all a_i must be zero. Thus, S is linearly independent. Thus, S is a basis of V.

$\Rightarrow$ Assume S is a basis of V. Therefore, $\mathrm{Sp}(S) = V$. This implies that every $v \in V$ can be written as a linear combination of vectors from S. It remains to show uniqueness.

Assume by contradiction that $\exists v \in V$ can be written as

$$v = a_1 v_1 + \cdots + a_n v_n = b_1 v_1 + \cdots + b_n v_n.$$

Where $a_k \neq b_k$ for some $1 \leq k \leq n$.

Now, subtracting both expressions, we get

$$0_V = (a_1 - b_1)v_1 + \cdots + \underbrace{(a_k - b_k)}_{\neq 0} v_k + \cdots + (a_n - b_n)v_n.$$

Thus 0_V can be written as a non-trivial linear combination of vectors from S. Thus S is not linearly independent, which contradicts our assumption that S is a basis. □

Example 3.35:

1) The set $\{e_1, \dots, e_n\}$ is a basis of K^n.

2) The set $\{1, x, x^2, \dots, x^d\}$ is a basis of $K[x]_d$.

3) The set $\{E_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis of $M_{m \times n}(K)$.

In case of K^n, we call the basis $\{e_1, \dots, e_n\}$ the STANDARD BASIS or CANONICAL BASIS of K^n.

Lemma 3.36:

Let $v_1, \dots, v_m \in V$ be a list of linearly dependent vectors. Then $\exists 1 \leq j \leq m$ such that

1) $v_j \in \mathrm{Sp}(v_1, \dots, v_{j-1})$

2) $\mathrm{Sp}(v_1, \dots, v_m) = \mathrm{Sp}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m)$

Proof. Since the list is linearly dependent, $\exists a_1, \dots, a_m \in K$ not all of them zero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0_V.$$

Define $j := \max\{i \mid a_i \neq 0\}$ (i.e. $a_j \neq 0$ and $a_{j+1} = \cdots = a_m = 0$).

Then $a_1 v_1 + \cdots + a_{j-1} v_{j-1} + a_j v_j = 0_V$. Since $a_j \neq 0$, we can solve for v_j:

$$v_j = -\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1}. \tag{3.1}$$

This implies $v_j \in \mathrm{Sp}(v_1, \dots, v_{j-1})$.

To prove 2), let $v \in \mathrm{Sp}(v_1, \dots, v_m)$. Then

$$v = c_1 v_1 + \cdots + c_m v_m.$$

Substituting (3.1) into this expression, we obtain, that v is a linear combination of $v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m$. Hence,

$$\mathrm{Sp}(v_1, \dots, v_m) \subseteq \mathrm{Sp}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m).$$

But clearly, we also have the other inclusion, thus

$$\mathrm{Sp}(v_1, \dots, v_m) = \mathrm{Sp}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m).$$

□

Lemma 3.37:

Let $v_1, \dots, v_m$ be a list of linearly dependent vectors such that $v_1, \cdots, v_k$ are linearly independent for some $1 \leq k < m$. Then j from Lemma 3.36 satisfies $k < j$.

Proof. Assume by contradiction that $j \leq k$. We have:

$$v_j = a_1 v_1 + \cdots + a_{j-1} v_{j-1},$$

for some $a_1, \dots, a_{j-1} \in K$.

But this implies that

$$0 = a_1 v_1 + \cdots + a_{j-1} v_{j-1} - 1 \cdot v_j.$$

Since there is a -1 coefficient, this is a non-trivial linear combination of $v_1, \dots, v_k$ that equals 0_V. This contradicts the linear independence of $v_1, \dots, v_k$. □

Lemma 3.38:

Let $w_1, \dots, w_n \in V$ such that $\mathrm{Sp}(w_1, \dots, w_n) = V$ and let $v \in V$. Then the list $v, w_1, \dots, w_n$ is linearly dependent.

Proof. Write $v = a_1 w_1 + \cdots + a_n w_n$ for some $a_1, \dots, a_n \in K$. Then

$$0_V = (-1) \cdot v + a_1 w_1 + \cdots + a_n w_n.$$

This is a non-trivial linear combination of $v, w_1, \dots, w_n$ that equals 0_V. Thus, the list is linearly dependent. □

Lemma 3.39:

Let $v_1, \dots, v_n \in V$ such that $\mathrm{Sp}(v_1, \dots, v_n) = V$. Let $u_1, \dots, u_m$ be a list of linearly independent vectors in V. Then $m \leq n$.

Proof. The proof will have m steps.

Step 1: We will replace u_1 by one of the vectors $v_1, \dots, v_n$. How? Consider the list $u_1, v_1, \dots, v_n$. By Lemma 3.38, this list is linearly dependent. By Lemma 3.36, one of the vectors in this list is in the span of the vectors that appear before it, and if we drop this vector from the list, the overall span does not change. By Lemma 3.37, this vector cannot be u_1 (because u_1 is linearly independent).

We now drop that vector and get a list of n vectors that still span V and the first vector is u_1.

Step j ($2 \leq j \leq m$): From step $j-1$, we should have a list of n vectors that span V and the first $j-1$ vectors are $u_1, \dots, u_{j-1}$ and the other rest $n-(j-1)$ of vectors are taken from the list $v_1, \dots, v_n$.

Let's write this list of n vectors as $u_1, \dots, u_{j-1}, w_1, \dots, w_{n-(j-1)}$. Consider now the list $u_1, \dots, u_{j-1}, u_j, w_1, \dots, w_{n-(j-1)}$. Lec 13

This list with $n+1$ vectors spans V as the span cannot be reduced by adding more vectors. By Lemma 3.38, we know that this list is linearly dependent. By Lemma 3.36, one of the vectors in this list is in the span of the vectors that appear before it, and if we drop this vector from the list, the overall span does not change. By Lemma 3.37, this vector can be not of $u_1, \dots, u_{j-1}$ (because they are linearly independent). Thus, the vector that we drop is in the set $\{w_1, \dots, w_{n-(j-1)}\}$.

So, the new list is $u_1, \dots, u_j, w_1, \dots, w_{n-j}$ and it still spans V.

Assume now, by contradiction, that $m > n$. Then, we can perform the steps $j = 1, \dots, j = n$ and obtain at step $j = n$ a list that looks like

$$u_1, u_2, \dots, u_n, u_n.$$

This list spans V as we have shows. But there are more u's which are not in the list, in particular u_{n+1}. If we now add this vecotr to the list, by lemma 3.38, the list is linearly dependent. But this contradicts the linear independence of $u_1, \dots, u_{n+1}$. □

In fact, we proved a bit more than Lemma 3.39, namely: We can remove m vectors from the list $v_1, \dots, v_n$ such that the remaining $n-m$ vectors $v_{i_1}, \dots, v_{i_{n-m}}$ when put together with $u_1, \dots, u_m$ still span V.

In particular, if $m = n$, then $\mathrm{Sp}(u_1, \dots, u_n) = V$.

Theorem 3.40:

Let V be a finite dimensional vector space over K. Then V has a finite basis. Moreover, every basis of V is finite and has the same number of elements.

Lemma 3.41:

Let $w_1, \ldots, w_l \in V$ and assume $\forall 1 \leq j \leq l$, $w_j \notin \mathrm{Sp}(w_1, \ldots, w_{j-1})$. Then $w_1, \ldots, w_l$ are linearly independent.

Proof. If $w_1, \ldots, w_l$ were linearly dependent, then by Lemma 3.36, $\exists 1 \leq j \leq l$ such that $w_j \in \mathrm{Sp}(w_1, \ldots, w_{j-1})$, contradicting our assumption. □

Lemma 3.42:

Assume $\mathrm{Sp}(v_1, \ldots, v_n) = V$. Then $\exists$ a subset of $\{v_1, \ldots, v_n\}$ that is a basis of V.

Proof. We'll have n steps, and each step we'll have consider one of the vectors form $v_1, \ldots, v_n$ and decide wether or not to drop it.

Step 1: If $v_1 = 0_V$, we drop it. If not, we keep it.

Step $2 \leq j \leq n$: If $v_j \in \mathrm{Sp}(v_1, \ldots, v_{j-1})$, we drop it. Otherwise, we keep it. After performing n steps as above, we get a possibly shorter list $v_{i_1}, \ldots, v_{i_m}$ with

$$1 \leq i_1 < i_2 < \cdots < i_m \leq n.$$

We claim that $\mathrm{Sp}(v_{i_1}, \ldots, v_{i_m}) = V$. Indeed, in any of the steps $1 \leq j \leq n$, we dropped the vector v_j only if $v_j \in \mathrm{Sp}(v_1, \ldots, v_{j-1})$. From this it follows that dropping v_j does not change the overall span.

$$\mathrm{Sp}(v_{i_1}, \ldots, v_{i_m}) = \mathrm{Sp}(v_1, \ldots, v_n) = V.$$

We now claim, that $v_{i_1}, \ldots, v_{i_m}$ are linearly independent. Indeed, $\forall 1 \leq k \leq m$ we have $v_{i_k} \notin \mathrm{Sp}(v_{i_1}, \ldots, v_{i_k - 1})$. Thus in particular,

$$v_{i_k} \notin \mathrm{Sp}(v_{i_1}, \ldots, v_{i_{k-1}}).$$

By Lemma 3.41 applied to the list $v_{i_1}, \ldots, v_{i_m}$, we deduce that they are linearly independent. So $v_{i_1}, \ldots, v_{i_m}$ is a basis of V. □

At last, we can prove Theorem 3.40:

Proof. [of Theorem 3.40] We need to prove three things:

1) V has a finite basis. Take any finite set $S \subseteq V$, such that $\mathrm{Sp}(S) = V$ (this is possible by definition of finite dimensional). By Lemma 3.42, there is a subset of S' of S that is a basis of V.

2) Every basis of V has finitely many elements. Let $\mathcal{C}$ be any basis for V. We'll show that $\mathcal{C}$ is a finite set. Suppose by contradiction that $\mathcal{C}$ is infinite. Also, take any finite basis, say S, for V (we know such a basis exists by 1), say $S = \{v_1, \ldots, v_n\}$. Choose $n + 2025$ vectors, $u_1, \ldots, u_{n+2025} \in \mathcal{C}$. By Lemma 3.39, we have that

$$n + 2025 \leq n.$$

This is a contradiction. Thus, every basis of V is finite.

3) Any two bases $\mathcal{A}, \mathcal{B}$ of V have the same number of elements. By statement 2), both $\mathcal{A}$ and $\mathcal{B}$ are finite sets. Now $\mathrm{Sp}(\mathcal{B}) = V$ and $\mathcal{A}$ is linearly independent. By Lemma 3.39, we have that

$$|\mathcal{A}| \leq |\mathcal{B}|.$$

But interchanging the roles of $\mathcal{A}$ and $\mathcal{B}$, we also have that

$$|\mathcal{B}| \leq |\mathcal{A}|.$$

Thus, $|\mathcal{A}| = |\mathcal{B}|$. □

This inspires the following definition:

Definition 3.43: Dimension

Let V be a finite dimensional vector space over K. We define the DIMENSION of V to be the unique number $n \in \mathbb{Z}_{\geq 0}$ of elements in any basis of V. We write

$$\dim(V) = n.$$

Sometimes, people also write $\dim_K(V)$ to emphasize the field K.

Example 3.44:

1) $\dim\{0_V\} = 0$.

2) $\dim(K^n) = n$.

3) $\dim(K[x]_d) = d + 1$.

4) $\dim(M_{m \times n}(K)) = m \cdot n$.

As a summary of the algorithm we developped, we have: Let V be a finite dimensional vector space over K. Then:

1) Every finite list of vectors that spans V will contain a sublist which is a basis of V.

2) Every linearly independent list of vectors in V can be extended to get a basis of V. Lec 14

Theorem 3.45:

Let V be a finite dimensional vector space of dimension $n := \dim(V)$. The follwoing statements are equivalent:

1. $v_1, \ldots, v_n \in V$ are linearly independent.
2. $v_1, \ldots, v_n$ span V.
3. $v_1, \ldots, v_n$ form a basis of V.

Proof. $1 \Rightarrow 3$: By what was done last time, we can extend the list $v_1, \ldots, v_n$ and get a basis for V. But this list has already n vectors in it. So there is no extension needed. Thus, $v_1, \ldots, v_n$ is a basis of V.

$3 \Rightarrow 1\&2$: By definition of basis.

$2 \Rightarrow 3$: By again, what we did last time, we can possibly drop some vectors from the list $v_1, \ldots, v_n$ and get a basis for V. But the length of the list $v_1, \ldots, v_n$ is already n. So dropping is impossible. Thus $v_1, \ldots, v_n$ was a basis already from the beginning. □

Theorem 3.46:

Suppose V is finite dimensional and $n = \dim(V)$. Let $v_1, \ldots, v_k \in V$ be a list. Then:

1. If $k < n$ then $v_1, \ldots, v_k$ do not span V.
2. If $k > n$ then $v_1, \ldots, v_k$ are linearly dependent.

Proof. 1) Assume $k < n$. Suppose by contradiction that this list spans V. As we have seen, we can thin this list to a sublist which is a basis of V. But any basis of V has n elements. So we have a contradiction.

2) Assume $k > n$. Assume by contradiction that $v_1, \ldots, v_k$ are linearly independent. By what we did last time, we can extend this list to a basis of V. But any basis of V has n elements. So we have a contradiction. □

Proposition 3.47:

Let V be a finite dimensional vector space. Then every subspace $U \subseteq V$ is also finite dimensional and we have

$$\dim(U) \leq \dim(V).$$

Moreover, $\dim(U) = \dim(V) \Leftrightarrow U = V$.

Proof. Let $n = \dim(V)$. We'll show first that U is finite dimensional.

If $U = \{0_V\}$, then $\dim(U) = 0$ and we are done. Assume now, that $U \neq \{0_V\}$. Take any vector $v_1 \in U \setminus \{0_V\}$. If $U = \mathrm{Sp}(v_1)$, then $\dim(U) = 1$ and we are done.

If $U \neq \mathrm{Sp}(v_1)$, take $v_2 \in U \setminus \mathrm{Sp}(v_1)$. By lemma 3.41, v_1, v_2 are linearly independent.

We continue this process and after j steps, we obtain a list $v_1, \ldots, v_j \in U$ that are linearly independent.

However, this process must stop after at most n steps, since otherwise we would have a list of $n+1$ linearly independent vectors in V, contradicting the fact that $\dim(V) = n$.

For the equality case, conversely suppose $k = n$. Thus the list of vectors $v_1, \ldots, v_n$ is a basis for U, hence these are linearly independent vectors. They continue to be linearly independent even when viewed as vectors in V. By the above theorem, $v_1, \ldots, v_n$ also span V. Thus $U = \mathrm{Sp}(v_1, \ldots, v_n) = V$. □

3.6 Row and Column Spaces

Consider K^m. Let $v_1, \ldots, v_n \in K^m$. The follwoing questions arise:

1. Let $w \in K^m$. How to determine whether or not $w \in \mathrm{Sp}(v_1, \ldots, v_n)$?
2. How do we describe $\mathrm{Sp}(v_1, \ldots, v_n)$?
3. How to determine if $v_1, \ldots, v_n$ are linearly independent?

Let us start with question 1).

For $w \in \mathrm{Sp}(v_1, \ldots, v_n)$, this is equivalent to the existence of $a_1, \ldots, a_n \in K$ such that

$$w = a_1 v_1 + \cdots + a_n v_n.$$

We can write this as

$$\begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} | \\ w \\ | \end{pmatrix}. \tag{3.2}$$

So the question is, wether or not the system of equations (3.2) with the unknowns $a_1, \ldots, a_n$ has a solution.

For the second question, apply answer 1) to a general $b = w$. So do elimination to the augmented matrix $(A|b)$, untill we get a reduced row echelon matrix $(A'|b')$. And $A'x = b'$ has a solution iff $Ax = b$ has a solution.

In case A' does not have any 0-rows (i.e each row of A' has a pivot). In this case, $A'x = b'$ will always have a solution, which implies that $\mathrm{Sp}(v_1, \ldots, v_n) = K^m$.

If A' does have some 0-rows, then the entries in b' corresponding to these 0-rows must also be 0 for $A'x = b'$ to have a solution. Thus

$$b \in \mathrm{Sp}(v_1, \ldots, v_n) \Leftrightarrow b'_{r+1} = b'_{r+2} = \cdots = b'_m = 0.$$

Example 3.48:

Take the following in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

The span of v_1, v_2 is given by all vectors $b \in \mathbb{R}^3$ such that

$$\begin{pmatrix} 1 & 1 & b_1 \\ 1 & 0 & b_2 \\ 1 & -1 & b_3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & b_1 \\ 0 & 1 & b_1 - b_2 \\ 0 & 0 & b_1 - 2b_2 + b_3 \end{pmatrix}.$$

From this, we see that for $\mathrm{Sp}(v_1, v_2)$, we must have

$$\mathrm{Sp}(v_1, v_2) = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3 \mid b_1 - 2b_2 + b_3 = 0 \right\}.$$

For question three, we see that $v_1, \ldots, v_n$ are linearly independent iff the system of equations

$$\begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} | \\ 0_V \\ | \end{pmatrix}.$$

Has only the trivial solution $a_1 = a_2 = \cdots = a_n = 0$. So Take $A = (v_1 \cdots v_n)$, do elimination and get A'. Lec 15

In other words, $v_1, \ldots, v_n$ are linearly independent iff the reduced row echelon form A' of A has percisely n pivots (i.e no zero rows). So the matrix is of the form

$$A' = \begin{pmatrix} 1 & * & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}.$$

Let us revisit theorem 3.46. Take a list $v_1, \ldots, v_n \in K^m$, where $n < m$. Form the matrix

$$A = \begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix}.$$

We can have at most n pivots but $m > n$. When trying to solve $Ax = b$, we will get at least one zero row in the reduced row echelon form of A. Now, the span is given by

$$\mathrm{Sp}(v_1, \ldots, v_n) = \{b \in K^m \mid b'_{n+1} = b'_{n+2} = \cdots = b'_m = 0\}.$$

Assume now that $n > m$. Why is $v_1, \ldots, v_n$ linearly dependent? Well, we need to check if the system

$$\begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} | \\ 0_V \\ | \end{pmatrix}$$

has non-trivial solutions. After elimination, we get $A'x = 0$ and A' can at most have m pivots, since there are only m rows. But $n > m$, so there are at least $n - m$ free variables. Thus, there are non-trivial solutions.

Definition 3.49: Row and Column Spaces

Let $A \in M_{m\times n}(K)$. Denote by $u_1, \dots, u_m \in K^n$ the rows of A and by $v_1, \dots, v_n \in K^m$ the columns of A.

$$A = \begin{pmatrix} - & u_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_m & - \end{pmatrix} = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Define two spaces:

$$\begin{aligned} \mathrm{RowS}(A) &:= \mathrm{Sp}(u_1, \dots, u_m) \subseteq K^n \\ \mathrm{ColS}(A) &:= \mathrm{Sp}(v_1, \dots, v_n) \subseteq K^m. \end{aligned}$$

Where we call $\mathrm{RowS}(A)$ the ROW SPACE of A and $\mathrm{ColS}(A)$ the COLUMN SPACE of A.

Lemma 3.50:

If $A, B \in M_{m\times n}(K)$ are row equivalent, then

$$\mathrm{RowS}(A) = \mathrm{RowS}(B).$$

Notice that it is NOT true that $\mathrm{ColS}(A) = \mathrm{ColS}(B)$

Proof. Let $M \in M_{m\times m}(K)$.

1) When we exchange two rows of M, the span of the rows does not change. So $\mathrm{RowS}(M)$ does not change under this operation.

2) The operation $\lambda R_i \to R_i$ for $\lambda \in K \setminus \{0\}$. Here too, the span of the rows does not change. So $\mathrm{RowS}(M)$ does not change under this operation.

3) The operation $R_i + \lambda R_j \to R_i$ for $\lambda \in K, i \neq j$. We need to show $\mathrm{Sp}(R_1, \dots, R_m) = \mathrm{Sp}(R_1, \dots, R_i + \lambda R_j, \dots, R_m)$.

Indeed, $\mathrm{Sp}(R_1, \dots, R_i + \lambda R_j, \dots, R_m) \subseteq \mathrm{Sp}(R_1, \dots, R_m)$ since $R_i + \lambda R_j \in \mathrm{Sp}(R_1, \dots, R_m)$.

We also have $\mathrm{Sp}(R_1, \dots, R_m) \subseteq \mathrm{Sp}(R_1, \dots, R_i + \lambda R_j, \dots, R_m)$. Indeed, $R_i = (R_i + \lambda R_j) - \lambda R_j$, and since $i \neq j$, R_j is in the span on the right hand side.

So we conclude that after one elementary row operation on M, the $\mathrm{RowS}(M)$ is unchanged.

If A and B are row equivalent, then $\exists$ a finite sequence of matrices $A = A_0, A_1, \dots, A_k = B$ such that each A_{i+1} is obtained from A_i by an elementary row operation. By the above,

$$\mathrm{RowS}(A_0) = \mathrm{RowS}(A_1) = \cdots = \mathrm{RowS}(A_k) = \mathrm{RowS}(B).$$

□

Definition 3.51:

Let $A \in M_{m\times n}(K)$. The ROW-RANK of A is defined as

$$\begin{aligned} \text{row-rank}(A) &:= \dim(\mathrm{RowS}(A)) \\ \text{col-rank}(A) &:= \dim(\mathrm{ColS}(A)). \end{aligned}$$

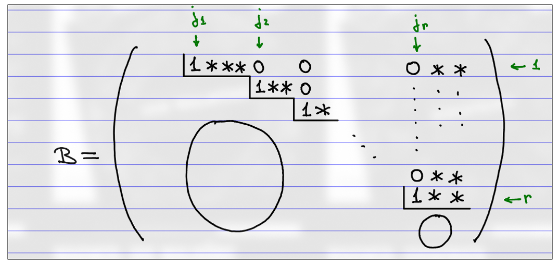
Lemma 3.52:

Let $B \in M_{m\times n}(K)$ which is in row reduced echelon form. Then the rows of B that are not totally 0 form a basis of $\mathrm{RowS}(B)$. In particular, $\text{row-rank}(B)$ is equal to the number of pivots in B.

Also the pivot columns of B form a basis of $\mathrm{ColS}(B)$. In particular, for matrices B that are in row reduced echelon form, we have

$$\text{col-rank}(B) = \text{row-rank}(B).$$

Proof. Consider the following matrix:



Denote by $j_1 < \cdots < j_r$ the column numbers of the pivots. Denote by $u_1, \dots, u_r$ the non-zero rows of B. We claim that $u_1, \dots, u_r$ are linearly independent.

Indeed, assume that $\lambda_1 u_1 + \cdots + \lambda_r u_r = 0$. This is a vector in K^n. Entry number j_k in this vector is precisely λ_k, since u_k has a pivot in column j_k. But if we assume that $\lambda_1 u_1 + \cdots + \lambda_r u_r = 0$, then all entries of this vector are 0. In particular, $\lambda_k = 0$ for all $k = 1, \dots, r$. Thus, $u_1, \dots, u_r$ are linearly independent. This shows our claim.

But $\mathrm{RowS}(B) = \mathrm{Sp}(u_1, \dots, u_r)$ by definition.

By Thm 3.45, $u_1, \dots, u_r$ form a basis of $\mathrm{RowS}(B)$. Also we get that $\text{row-rank}(B) = r$.

For the column space, clearly

$$\mathrm{ColS}(B) \subseteq \{x \in K^m \mid x_{r+1} = \dots x_m = 0\} = K^r \times \{0\}^{m-r}.$$

At the same time, the pivot columns of B are of the form

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ * \end{pmatrix}.$$

So $e_1, \dots, e_r$ form a basis for $K^r \times \{0\}$. Hence we have, $K^r \times \{0\} = \mathrm{Sp}(e_1, \dots, e_r) \subseteq \mathrm{ColS}(B)$.

Thus, $\mathrm{ColS}(B) = K^r \times \{0\}$ and the pivot columns of B form a basis of $\mathrm{ColS}(B)$. In particular, $\text{col-rank}(B) = r = \text{row-rank}(B)$.

□ Lec 16

We thus have found an algorithm to find a basis for the span of a list of vectors $u_1, \dots, u_m \in K^n$, viewed as rows. Just put them as rows of a matrix $A \in M_{m\times n}(K)$, row reduce A to B, and take the non-zero rows of B as a basis of $\mathrm{Sp}(u_1, \dots, u_m)$.

$$A = \begin{pmatrix} - & u_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_m & - \end{pmatrix}.$$

Remark 3.53:

The number of pivots one gets after elimination does NOT depend on the specific elimination process because the number of pivots is equal to the dimension of the row space, which is an invariant of the matrix.

3.6.1 Transpose of a Matrix

Sometimes its useful to switch between rows and columns of a matrix.

Definition 3.54: **Transposed Matrix**

Let $A = (a_{ij}) \in M_{m\times n}(K)$. The TRANSPOSED MATRIX of A, is

$$A^T \coloneqq (b_{ij}) \in M_{n\times m}(K) \text{ with } b_{ij} = a_{ji}.$$

Sometimes also A^t is written.

Example 3.55:

Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Then

$$A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Simple properties of the transpose: $\forall A, B \in M_{m\times n}(K)$ and $\lambda \in K$:

1. $(A+B)^T = A^T + B^T$
2. $(\lambda A)^T = \lambda A^T$
3. $(A^T)^T = A$
4. $\mathrm{RowS}(A^T) = \mathrm{ColS}(A)$ and $\mathrm{ColS}(A^T) = \mathrm{RowS}(A)$

3.7 Sums of Vector Spaces

Let V be a vector space over K.

Definition 3.56: **Sum of Subspaces**

Let $U, W \subseteq V$ be subspaces. We define the SUM as

$$U + W \coloneqq \{u + w \mid u \in U, w \in W\} \subseteq V.$$

In similar fashion, we can also take sums of more subspaces.

Proposition 3.57:

Let $U, W \subseteq V$ be subspaces. Then

1. $U + W = \mathrm{Sp}(U \cup W)$. In particular, $U + W$ is a subspace of V.
2. Suppose U, W are finite-dimensional. Then $U + W$ is also finite-dimensional and one can write a basis for $U + W$ as follows:
   - Choose a basis $p_1, \dots, p_k$ for $U \cap W$ where $k = \dim(U \cap W)$.
   - Extend this to a basis $p_1, \dots, u_1, \dots, u_{l-k}$ of U, where $l = \dim(U)$.
   - Extend $p_1, \dots, p_k$ to a basis of W say $p_1, \dots, p_k, w_1, \dots, w_{m-k}$, where we have $m = \dim(W)$.
   - $p_1, \dots, p_k, u_1, \dots, u_{l-k}, w_1, \dots, w_{m-k}$ are a basis of $U + W$.
3. In particular

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

Proof. 1) Clearly, $U+W \subseteq \mathrm{Sp}(U\cup W)$. We'll show that $\mathrm{Sp}(U\cup W) \subseteq U + W$.

Indeed, let $v \in \mathrm{Sp}(U\cup W)$. Then $v = \sum_{i=1}^s a_i u_i + \sum_{j=1}^r b_j w_j$, where $a_i, b_j \in K$, $u_i \in U$ and $w_j \in W$. But the first sum belongs to U and the second to W because U, W are subspaces. Thus, $v \in U + W$.

Together this shows $U + W = \mathrm{Sp}(U \cup W)$.

2) We first show that

$$S \coloneqq \{p_1, \dots, p_k, u_1, \dots, u_{l-k}, w_1, \dots, w_{m-k}\}$$

are linearly independent. Indeed, if

$$\sum_{i=1}^k a_i p_i + \sum_{j=1}^{l-k} b_j u_j + \sum_{t=1}^{m-k} c_t w_t = 0_V.$$

Write $v \coloneqq c_1 w_1 + \dots c_{m-k} w_k \in W$. Note that

$$v = -(\sum_{i=1}^k a_i p_i + \sum_{j=1}^{l-k} b_j u_j) \in U.$$

So $v \in U \cap W$. Since $p_1 \dots, p_k$ is a basis of $U \cap W$, $\exists$ coefficients $\alpha_1, \dots, \alpha_k \in K$ such that

$$v = \sum_{i=1}^k \alpha_i p_i.$$

Furthermore, since $p_1, \dots, p_k, w_1, \dots, w_{m-k}$ is a basis of W, Thus,

$$\sum_{i=1}^k \alpha_i p_i - \sum_{t=1}^{m-k} c_t w_t = 0_V$$
$$\Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

.

But $p_1, \dots, p_k, u_1, \dots, u_{l-k}$ is a basis of U, so

$$a_1 = \dots = a_k = b_1 = \dots = b_{l-k} = 0.$$

This shows that the p vectors with the u and w vectors are linearly independent. It remains to show that they span $U + W$.

Indeed let $z \in U + W$. Write $z = u + w$. We can write

$$u = \alpha_1 p_1 + \dots + \alpha_k p_k + a_1 u_1 + \dots + a_{l-k} u_{l-k}$$
$$w = \beta_1 p_1 + \dots + \beta_k p_k + b_1 w_1 + \dots + b_{m-k} w_{m-k}.$$

Then,

$$\begin{aligned} z &= (\alpha_1+\beta_1)p_1+\cdots+(\alpha_k+\beta_k)p_k \\ &\quad + a_1u_1+\cdots+a_{l-k}u_{l-k} \\ &\quad + b_1w_1+\cdots+b_{m-k}w_{m-k}. \end{aligned}$$

But this definetly belongs to $\mathrm{Sp}(S)$. Thus S is a basis of $U+W$.

3) We know

$$\begin{aligned} \dim(U+W) &= k+(l-k)+(m-k) \\ &= l+m-k \\ &= \dim(U)+\dim(W)-\dim(U\cap W). \end{aligned}$$

□

Corollary 3.58:

Let $U,W\subseteq V$ be finite-dimensional subspaces. Then the following are equivalent:

1. $\dim(U+W)=\dim(U)+\dim(W)$
2. $\dim(U\cap W)=0$
3. $U\cap W=\{0_V\}$
4. Every $v\in U+W$ can be written uniquely as $v=u+w$ with $u\in U$ and $w\in W$.
5. If $u+w=0_V$, then $u=0_V$ and $w=0_V$.

Proof. $1\Leftrightarrow 2$: Follows from the dimesion formula in Proposition 3.57.

$2\Leftrightarrow 3$: Again using proposition 3.57 and the fact that there is only one space of dimension 0, namely $\{0_V\}$.

$3\Rightarrow 4$: By the definition 3.56, every $v\in U+W$ *can* be written as $v=u+w$ with $u\in U$ and $w\in W$. We need to show now the uniqueness. Assume $v=u+w=u'+w'$ with $u,u'\in U$ and $w,w'\in W$. Then $u-u'=w'-w$. But the left hand side belongs to U and the right hand side to W. Thus, $u-u'\in U\cap W$. By 3), $u-u'=0_V$, so $u=u'$. Consequently, $w=w'$.

$4\Rightarrow 5$: $0=0+0$, so if $0=u+v$ and 4 holds, then $u=0$ and $v=0$.

$5\Rightarrow 3$: Let $v\in U\cap W$. Clearly $0=v+(-v)$ with $v\in U$, $-v\in W$. By 5), $v=0_V$. □

Definition 3.59: Complement of a Subspace

Let $U\subseteq V$ be a subspace. A subspace $W\subseteq V$ is called a COMPLEMENT of U in V if

$$U+W=V \text{ and } U\cap W=\{0_V\}.$$

Lec 17

Notice there are in general many complements of a given subspace. There is not a canonical choice.

Proposition 3.60:

Let V be a finite-dimensional vector space over K and $U\subseteq V$ a subspace. Then there exists a subspace $W\subseteq V$ which is a complement of U in V.

Proof. Choose a basis $u_1,\ldots,u_l$ of U, where $l=\dim(U)$. Extend this list to a basis of V, say

$$u_1,\ldots,u_l,w_1,\ldots,w_m.$$

So $l+m=\dim(V)$. Take $W:=\mathrm{Sp}(w_1,\ldots,w_m)$. Clearly, $U+W=V$ because $U+W=\mathrm{Sp}(u_1,\ldots,u_l,w_1,\ldots,w_m)=V$.

Also $U\cap W=\{0_V\}$ because if $\sum_{i=1}^l a_iu_i=\sum_{j=1}^m b_jw_j$, then

$$\sum_{i=1}^l a_iu_i-\sum_{j=1}^m b_jw_j=0_V.$$

But since $u_1,\ldots,u_l,w_1,\ldots,w_m$ is a basis of V, they are linearly independent, so all coefficients are 0.

Alternatively, by corollary 3.58, we can compute

$$\dim(U\cap W)=\dim(U)+\dim(W)-\dim(U+W)=l+m-(l+m)=0.$$

□

4 Linear Maps

Let V, W be vector spaces over a field K.

A linear map is a function between vector spaces that respects the structure of the vector spaces. Other names include, linear transformations or homomorphisms of vector spaces.

Definition 4.1: Linear Map

A map $T : V \to W$ is called LINEAR if

1. $\forall u, v \in V : T(u+v) = T(u) + T(v)$
2. $\forall v \in V, \forall a \in K : T(a \cdot v) = a \cdot T(v)$

The set of all linear maps $T : V \to W$ is denoted by $\mathrm{Hom}(V, W)$ or $\mathrm{Hom}_K(V, W)$.

Sometimes, we write $Tv = T(v)$. The reason, V and W are over the same field K is that the definition would not make sense otherwise.

Exercise 4.2:

Show that $T : V \to W$ is linear iff

$$T(au + bv) = aT(u) + bT(v) \quad \forall u, v \in V, \forall a, b \in K.$$

Example 4.3:

1) $\mathrm{id} : V \to V$ is a linear map.

2) $0 : V \to W, v \mapsto 0_W$ is a linear map.

3) $D : K[x] \to K[x]$, the (formal) derivative, i.e.

$$D(a_0 + a_1 x + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}$$

is a linear map as

$$\begin{aligned} D(ap(x) + bq(x)) &= ap'(x) + bq'(x) \\ &= aD(p(x)) + bD(q(x)). \end{aligned}$$

4) Let $C([a, b])$ be the vector space of continuous functions $f : [a, b] \to \mathbb{R}$. Then the map

$$T : C([a, b]) \to \mathbb{R}, \quad f \mapsto \int_a^b f(x)\, dx,$$

is linear.

5) $S : K^\infty \to K^\infty, (a_1, a_2, \dots) \mapsto (a_2, a_3, \dots)$ called the SHIFT map is linear.

Example 4.4: Important Example

Let $A \in M_{m \times n}(K)$. Define

$$T_A : K^n_{\mathrm{col}} \to K^m_{\mathrm{col}}, \quad x \mapsto Ax.$$

We showed that $A \cdot (a \cdot x) = a \cdot (Ax)$ and $A \cdot (x + y) = Ax + Ay$. Thus, T_A is linear.

Proposition 4.5:

Let $T : V \to W$ be a linear map. Then:

1. $\forall n \in \mathbb{Z}_{\geq 0}, \forall v_1, \dots, v_n \in V, \forall a_1, \dots, a_n \in K$:

$$T\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i T(v_i).$$

2. $T(0_V) = 0_W$

Proof. 1) We prove this by induction on n.

Base case: $n = 2$: This is just the definition 4.1. Induction step: Assume true for n. We show for $n+1$:

$$\begin{aligned} T\left(\sum_{i=1}^{n+1} a_i v_i\right) &= T\left(\sum_{i=1}^{n} a_i v_i + a_{n+1} v_{n+1}\right) \\ &= T\left(\sum_{i=1}^{n} a_i v_i\right) + T(a_{n+1} v_{n+1}) \\ &= \sum_{i=1}^{n} a_i T(v_i) + a_{n+1} T(v_{n+1}) \\ &= \sum_{i=1}^{n+1} a_i T(v_i). \end{aligned}$$

2) Note that

$$T(0_V) = T(0 \cdot v) = 0 \cdot T(v) = 0_W.$$

□

Theorem 4.6:

Let V be finite-dimensional, and let $v_1, \dots, v_n$ be a basis of V. Then $\forall w_1, \dots, w_n \in W$, $\exists!$ linear map $T : V \to W$ such that $T(v_i) = w_i$.

In other words, we can define a linear map by defining where the basis vectors go.

Proof. Existence of T: Let $v \in V$. We want to define $T(v)$. Since $v_1, \dots, v_n$ is a basis of V, $\exists\ a_1, \dots, a_n \in K$ such that

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n. \tag{4.1}$$

Define $T(v) := a_1 w_1 + a_2 w_2 + \cdots + a_n w_n$. Note that T is well-defined because $\forall v \in V$, the coefficients a_i are unique. We claim that T is linear.

Indeed, let $u, v \in V$. Then $\exists a_i, b_i \in K$ such that

$$u = \sum_{i=1}^n a_i v_i, \quad v = \sum_{i=1}^n b_i v_i.$$

Now,

$$u + v = \sum_{i=1}^n (a_i + b_i) v_i.$$

But also by definition of T by (4.1):

$$\begin{aligned} T(u+v) &= \sum_{i=1}^n (a_i + b_i) w_i \\ &= \sum_{i=1}^n a_i w_i + \sum_{i=1}^n b_i w_i \\ &= T(u) + T(v). \end{aligned}$$

Let now $\alpha \in K, v \in V$ with $v = \sum_{i=1}^n a_i v_i$. Then,

$$\begin{aligned} T(\alpha v) &= T\left(\sum_{i=1}^n \alpha a_i v_i\right) \overset{(4.1)}{=} \sum_{i=1}^n \alpha a_i w_i \\ &= \alpha \sum_{i=1}^n a_i w_i = \alpha T(v). \end{aligned}$$

This proves that T is linear. Note that by definition, $T(v_i) = w_i$. We now have to show uniqueness of T.

Suppose $T, S : V \to W$ are linear maps such that $T(v_i) = S(v_i) = w_i$. We'll show that $T = S$. Let $v \in V$. Write $v = \sum_{i=1}^n a_i v_i$. Then,

$$\begin{aligned} T(v) &= T\left(\sum_{i=1}^n a_i v_i\right) \overset{\text{Prop. 4.5}}{=} \sum_{i=1}^n a_i T(v_i) \\ &= \sum_{i=1}^n a_i w_i \\ &= \sum_{i=1}^n a_i S(v_i) \overset{\text{Prop. 4.5}}{=} S\left(\sum_{i=1}^n a_i v_i\right) = S(v). \end{aligned}$$

□

Example 4.7:

Consider K^n with the standard basis $e_1, \dots, e_n$. Let $w_1, \dots, w_n \in K^m$ be n arbitrary vectors. By Theorem 4.6, $\exists!$ linear map $T : K^n \to K^m$ such that $T(e_i) = w_i$. We can describe T as $T = T_A$ where

$$A = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix}.$$

Recall that $T_A : K^n \to K^m$ is linear. Indeed, for e_i,

$$T_A(e_i) = A \cdot e_i = w_i.$$

So we have constructed a linear map T_A such that $T_A(e_i) = w_i$. By uniqueness in Theorem 4.6, $T = T_A$.

Lemma 4.8:

$\forall$ linear map $T : K^n \to K^m$, $\exists! A \in M_{m\times n}(K)$ such that $T = T_A$.

In fact,

$$A = \begin{pmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{pmatrix}.$$

Lec 18

4.1 Kernel and Image

Let $T : V \to W$ be a linear map.

Definition 4.9: Kernel and Image

The set

$$\operatorname{Ker}(T) := \{v \in V \mid T(v) = 0_W\} \subseteq V$$

is called the KERNEL of T.

The set

$$\operatorname{Im}(T) := \{Tv \mid v \in V\} \subseteq W$$

is called the IMAGE of T.

bWith the notation from set theory, we have

$$\operatorname{Ker}(T) = T^{-1}(\{0_W\}) \quad \text{and} \quad \operatorname{Im}(T) = T(V).$$

Proposition 4.10:

The kernel $\operatorname{Ker}(T)$ is a subspace of V and the image $\operatorname{Im}(T)$ is a subspace of W.

Proof. 1) Let $a, b \in K$ and $u, v \in \operatorname{Ker}(T)$. Then

$$T(av + bu) = aT(v) + bT(u) = a \cdot 0_W + b \cdot 0_W = 0_W.$$

Hence, $av + bu \in \operatorname{Ker}(T)$ and since $0_V \in \operatorname{Ker}(T)$, $\operatorname{Ker}(T)$ is a subspace of V.

2) Let $a, b \in K$ and $w_1, w_2 \in \operatorname{Im}(T)$. Then $\exists v_1, v_2 \in V$ such that

$$T(v_1) = w_1, \quad T(v_2) = w_2.$$

Now $T(av_1 + bv_2) = aT(v_1) + bT(v_2) = aw_1 + bw_2$. Thus $aw_1 + bw_2 \in \operatorname{Im}(T)$. Also $0_W = T(0_V) \in \operatorname{Im}(T)$. Hence, $\operatorname{Im}(T)$ is a subspace of W. □

Example 4.11:

Let $A \in M_{m\times n}(K)$ and $T_A : K^n \to K^m$ be a linear map $T_A(x) = Ax$. Then $\operatorname{Ker}(T_A) = \{x \in K^n \mid Ax = 0\}$ is the solution space of the homogeneous system of linear equations associated to A.

Proposition 4.12:

The following are equivalent for a linear map $T : V \to W$:

1. T is injective $\Leftrightarrow \operatorname{Ker}(t) = \{0_V\}$
2. T is surjective $\Leftrightarrow \operatorname{Im}(T) = W$

Proof. 2) This is true by definition of surjectivity and image. 1) $\Rightarrow$: Clearly $0_V \in \operatorname{Ker}(T)$. So $\{0_V\} \subseteq T^{-1}(\{0_W\})$. Since T is injective, $T^{-1}(\{0_W\})$ contains at most one element. Thus, $\operatorname{Ker}(T) = \{0_V\}$.

$\Leftarrow$: Let $u, v \in V$ such that $T(u) = T(v)$. Then we have

$$T(u) - T(v) = 0_W \Rightarrow T(u - v) = 0_W \Rightarrow u - v \in \operatorname{Ker}(T).$$

But by assumption, $\operatorname{Ker}(T) = \{0_V\}$, so $u - v = 0_V \Rightarrow u = v$ which shows the injectivity. □

Exercise 4.13:

Show that

1. If $V' \subseteq V$ is a subspace, then $T(V') \subseteq W$ is a subspace.
2. If $W' \subseteq W$ is a subspace, then $T^{-1}(W') \subseteq V$ is a subspace.
3. Explain why 1 & 2 are generalisations of Proposition 4.10.

Recall that $f : X \to Y$ is called bijective if it is both injective and surjective. In this case, $\exists!$ inverse map

$$g : Y \to X \quad \text{such that} \quad g \circ f = \operatorname{id}_X, \quad f \circ g = \operatorname{id}_Y.$$

The other direction of this also holds: If $\exists g : Y \to X$ such that $g \circ f = \operatorname{id}_X$ and $f \circ g = \operatorname{id}_Y$, then f is bijective.

Definition 4.14: Isomorphism

A linear map $T : V \to W$ is called an ISOMORPHISM if $\exists$ a linear map $S : W \to V$ such that

$$S \circ T = \mathrm{id}_V, \quad T \circ S = \mathrm{id}_W .$$

We say that V is isomorphic to W if $\exists$ an isomorphism $T : V \to W$ and write $V \cong W$.

Any isomorphism is obviously bijective. The other direction also holds:

Lemma 4.15:

Let $T : V \to W$ be a linear map. Then T is an isomorphism iff T is bijective.

Proof. Let $a, b \in K, w_1, w_2 \in W$. Put $v_1 = T^{-1}(w_1)$ and $v_2 = T^{-1}(w_2)$. Since T is linear we get

$$T(av_1 + bv_2) = aT(v_1) + bT(v_2) = aw_1 + bw_2.$$

But then

$$T^{-1}(aw_1 + bw_2) = av_1 + bv_2 = aT^{-1}(w_1) + bT^{-1}(w_2).$$

Hence, T^{-1} is linear and thus T is an isomorphism. □

In short, for linear maps, isomorphism = bijection.

Lemma 4.16:

Let $T : V \to W, S : W \to U$ be linear maps. Then the composition

$$S \circ T : V \to U, \quad v \mapsto S(T(v))$$

is also a linear map.

Proof. Let $a, b \in K, u, v \in V$. Then

$$\begin{aligned} S \circ T(au + bv) &= S(T(au + bv)) \\ &= S(aT(u) + bT(v)) \\ &= aS(T(u)) + bS(T(v)) \\ &= a(S \circ T)(u) + b(S \circ T)(v). \end{aligned}$$

□

Exercise 4.17:

Show that $\cong$ defines an equivalence relation on the set of vector spaces over K.

Definition 4.18: Endomorphism and Automorphism

A linear map $T : V \to V$ (both domain and target are V) is called an ENDOMORPHISM of V.

We denote the set of all endomorphisms of V by

$$\mathrm{End}(V) := \mathrm{Hom}(V, V).$$

A linear map $T : V \to V$ which is an isomorphism is called an AUTOMORPHISM of V.

Sometimes, if T is injective, we call it a MONOMORPHISM, and if T is surjective, we call it an EPIMORPHISM.

Lemma 4.19:

Let $v_1, \dots, v_n$ be a basis of V. Then $\mathrm{Im}(T) = \mathrm{Sp}(T(v_1), \dots, T(v_n))$.

Proof. We first show $Im(T) \supseteq \mathrm{Sp}(T(v_1), \dots, T(v_n))$. Let $a_1, \dots, a_n \in K$. Then

$$\sum_{i=1}^{n} a_i T(v_i) = T\left(\sum_{i=1}^{n} a_i v_i\right) \in \Im(T).$$

Thus $\mathrm{Sp}(T(v_1), \dots, T(v_n)) \subseteq \Im(T)$.

For the other inclusion, let $w \in \Im(T)$. Then $\exists v \in V$ such that $T(v) = w$. Since $v_1, \dots, v_n$ is a basis of V, $\exists a_1, \dots, a_n \in K$

$$v = \sum_{i=1}^{n} a_i v_i \Rightarrow w = T(v) = T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i T(v_i).$$

Thus $w \in \mathrm{Sp}(T(v_1), \dots, T(v_n))$ and hence this inclusion holds as well. □

Theorem 4.20: Rank Theorem

Let V be finite-dimensional and $T : V \to W$ a linear map. Then

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)).$$

Proof. Let $n := \dim V$ Let $u_1, \dots, u_k$ be a basis of $\mathrm{Ker}(T)$, where $k = \dim(\mathrm{Ker}(T))$. Extend this to a basis of V (Since the kernel is a subspace of V its also finite-dimensional). We extend it to a basis of V:

$$u_1, \dots, u_k, v_1, \dots, v_{n-k}.$$

By Lemma 4.19, we have

$$\begin{aligned} \mathrm{Im}(T) &= \mathrm{Sp}(\underbrace{T(u_1)}_{=0}, \dots, \underbrace{T(u_k)}_{=0}, T(v_1), \dots, T(v_{n-k})) \\ &= \mathrm{Sp}(T(v_1), \dots, T(v_{n-k})). \end{aligned}$$

We claim that $T(v_1), \dots, T(v_{n-k})$ form a basis for $\mathrm{Im}(T)$. Indeed we just proved that theese vectors span $\mathrm{Im}(T)$. We now show linear independence.

Let $a_1, \dots, a_{n-k} \in K$ and assume $\sum_{i=1}^{n-k} a_i T(v_i) = 0$. Since T is linear,

$$T(\sum_{i=1}^{n-k} a_i v_i) = \sum_{i=1}^{n-k} a_i T(v_i) = 0.$$

Thus, $\sum_{i=1}^{n-k} a_i v_i \in \mathrm{Ker}(T)$. But $u_1, \dots, u_k$ is a basis of $\mathrm{Ker}(T)$, so $\exists b_1, \dots, b_k \in K$ such that

$$\begin{aligned} &a_1 v_1 + \dots + a_{n-k} v_{n-k} = b_1 u_1 + \dots + b_k u_k \\ \Rightarrow\ &a_1 v_1 + \dots + a_{n-k} v_{n-k} - b_1 u_1 - \dots - b_k u_k = 0. \end{aligned}$$

But since $u_1, \dots, u_k, v_1, \dots, v_{n-k}$ is a basis of V, they must be linearly independent. Thus, all coefficients are 0, which proves that $T(v_1), \dots, T(v_{n-k})$ are linearly independent.

We can see that $\dim(\mathrm{Im}(T)) = n - k = \dim(V) - \dim(\mathrm{Ker}(T))$. which proves the theorem. □ Lec 19

Corollary 4.21:

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces. Then the following holds

- If $\dim W < \dim V$, then T is not injective.
- If $\dim V < \dim W$, then T is not surjective.
- If $\dim V = \dim W$, then

$$T \text{ injective } \Leftrightarrow T \text{ surjective } \Leftrightarrow T \text{ bijective.}$$

Proof. 1) Since $\mathrm{Im}(T) \subseteq W$, we have

$$\dim(\mathrm{Im}(T)) \leq \dim(W) < \dim(V).$$

By the rank theorem (Theorem 4.20),

$$\dim(\mathrm{Ker}(T)) = \dim(V) - \dim(\mathrm{Im}(T)) > 0.$$

But this means that $\mathrm{Ker}(T) \neq \{0_V\}$, so by Proposition 4.12, T is not injective.

2) Again by the rank theorem,

$$\dim(\mathrm{Im}(T)) = \dim(V) - \dim(\mathrm{Ker}(T)) \leq \dim(V) < \dim(W).$$

Since the image has strictly smaller dimension than W, we have $\mathrm{Im}(T) \subsetneq W$ and thus T is not surjective.

3) T is injective $\Leftrightarrow$ $\dim(\mathrm{Ker}(T)) = 0$. But this is equivalent to $\dim(\mathrm{Ker}(T)) = 0$. By the rank theorem, this is equivalent to $\dim(\mathrm{Im}(T)) = \dim(W)$, which is equivalent to T being surjective.

□

Corollary 4.22:

Two finite-dimensional vector spaces V and W are isomorphic iff $\dim(V) = \dim(W)$.

Proof. $\Rightarrow$: Suppose $T : V \to W$ is an isomorphism. Then T is bijective, so by Proposition 4.12,

$$\mathrm{Ker}(T) = \{0_V\} \Rightarrow \dim(\mathrm{Ker}(T)) = 0,$$

and

$$\mathrm{Im}(T) = W \Rightarrow \dim(\mathrm{Im}(T)) = \dim(W).$$

By the rank theorem, we have

$$\dim(W) = \dim(V) - \dim(\mathrm{Ker}(T)) = \dim(V) - 0 = \dim(V).$$

$\Leftarrow$: Assume $\dim(V) = \dim(W)$. Denote by n their common dimension.

Let $v_1, \ldots, v_n$ be a basis for V and $w_1, \ldots, w_n$ be a basis for W.

Define a linear map $T : V \to W$ by $T(v_i) = w_i$ for $i = 1, \ldots, n$. By Theorem 4.6, such a linear map exists and is unique. By Lemma 4.19, we have

$$\mathrm{Im}(T) = \mathrm{Sp}(T(v_1), \ldots, T(v_n)) = \mathrm{Sp}(w_1, \ldots, w_n) = W.$$

Hence, T is surjective. By Corollary 4.21, we have that T is also injective, and thus bijective, i.e. isomorphic. □

Theorem 4.23:

Let T be an isomorphism. Let S be a set of vectors from V. Write

$$T(s) := \{T(v) \mid v \in S\}.$$

Then

1) S is linearly independent iff $T(S)$ is linearly independent.

2) S spans V iff $T(S)$ spans W.

3) S is a basis of V iff $T(S)$ is a basis of W.

Given $V \cong W$, then V and W have the same algebraic properties. We can think of V and W as the same space represented in two different ways. For all practical purposes, V and W are identical.

Now, unfortunately, there is no *canonical* isomorphism between two vector spaces and none is preferred over another in general. Since this is the case, it is therefore better to look at all of them.

So for example, $K[x]_3 \cong K^4$, but there is no preferred isomorphism between them. A more extreme example is the vector space

$$\left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mid a_{11} + a_{22} = 0 \right\}.$$

This is a 3-dimensional vector space over K, so it is isomorphic to K^3, but there is no obvious isomorphism between them.

Definition 4.24: Rank of a Linear Map

We define the RANK of a linear map $T : V \to W$ as

$$\mathrm{rank}(T) := \dim(\mathrm{Im}(T)).$$

Exercise 4.25:

Let $T : V \to W$, $S : W \to U$ be linear maps where U, V, W are finite-dimensional. Show that

1. $\mathrm{rank}(S \circ T) \leq \min(\mathrm{rank}(S), \mathrm{rank}(T))$.
2. If S is injective, then $\mathrm{rank}(S \circ T) = \mathrm{rank}(T)$.
3. If T is surjective, then $\mathrm{rank}(S \circ T) = \mathrm{rank}(S)$.

4.2 Linear Maps and coordinates

In this section, we will always assume that all vector spaces are finite-dimensional. We will write a basis as an ordered tuple $\mathcal{B} = (v_1, \ldots, v_n)$.

Definition 4.26: Coordinate Vector

Let V be a vector space over K and $\mathcal{B} = (v_1, \ldots, v_n)$ a basis for V. Let $v \in V$. Define the COORDINATE VECTOR of v with respect to the basis $\mathcal{B}$ as

$$[v]_{\mathcal{B}} := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in K^n.$$

Where $a_1, \ldots, a_n \in K$ are the unique scalars such that

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n.$$

Define a map $\Phi_{\mathcal{B}} : V \to K^n$ by

$$\Phi_{\mathcal{B}}(v) = [v]_{\mathcal{B}} \quad \forall v \in V.$$

Proposition 4.27:

The map $\Phi_{\mathcal{B}} : V \to K^n$ is an isomorphism.

Proof. Linearity of $\Phi_{\mathcal{B}}$. Let $a, b \in K$ and $u, v \in V$. Now write v and u as a linear combination of the basis vectors:

$$v = \sum_{i=1}^{n} a_i v_i, \quad a_i \in K \Rightarrow [v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$u = \sum_{i=1}^{n} b_i v_i, \quad b_i \in K.$$

Now it holds that

$$\begin{aligned} av+bu &= a\sum_{i=1}^{n} a_i v_i + b\sum_{i=1}^{n} b_i v_i \\ &= \sum_{i=1}^{n}(aa_i+bb_i)v_i \\ \Rightarrow [av+bu]_{\mathcal{B}} &= \begin{pmatrix} aa_1+bb_1 \\ aa_2+bb_2 \\ \vdots \\ aa_n+bb_n \end{pmatrix} = a\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + b\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \\ &= a[v]_{\mathcal{B}} + b[u]_{\mathcal{B}}. \end{aligned}$$

Now in the $\Phi_{\mathcal{B}}$ definition, we have

$$\Phi_{\mathcal{B}}(av+bu) = [av+bu]_{\mathcal{B}} = a[v]_{\mathcal{B}} + b[u]_{\mathcal{B}} = a\Phi_{\mathcal{B}}(v) + b\Phi_{\mathcal{B}}(u).$$

This shows linearity. We now show that $\Phi_{\mathcal{B}}$ is an isomorphism.

$\Phi_{\mathcal{B}}$ is a map between two vector spaces of the same dimension n. Hence, we will show that $\mathrm{Ker}(\Phi_{\mathcal{B}}) = \{0_V\}$ to prove injectivity.

Indeed, if $v \in \mathrm{Ker}(\Phi_{\mathcal{B}})$, then

$$[v]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

But this means that $v = 0_V$. Thus, $\mathrm{Ker}(\Phi_{\mathcal{B}}) = \{0_V\}$ which implies injectivity. By Corollary 4.21, $\Phi_{\mathcal{B}}$ is also surjective, and thus an isomorphism. ☐

Lec 20

Remark 4.28:

The inverse $\Phi_{\mathcal{B}}^{-1} : K^n \to V$ of $\Phi_{\mathcal{B}}$ can be written as:

$$\Phi_{\mathcal{B}}^{-1}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1v_1 + a_2v_2 + \cdots + a_nv_n.$$

Example 4.29:

Consider $V = \mathbb{R}^2$ and let $\xi = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$ be the standard basis of $\mathbb{R}^2$. Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$ be another basis. Let $v = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. Then

$$[v]_{\xi} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad [v]_{\mathcal{B}} = \begin{pmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{pmatrix}.$$

But how can we do the second equality quickly? We can consider the first the basis vectors of ξ as linear combinations of the basis vectors of $\mathcal{B}$: Write $[v_1]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ Hence, by a system of equations, we have

$$a_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow a_1 = \frac{1}{2}, a_2 = \frac{1}{2}.$$

Similarly, for v_2 we have

$$b_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + b_2\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow b_1 = \frac{1}{2}, b_2 = -\frac{1}{2}.$$

Since the map is linear, we find the desired result.

Let V, W be finite-dimensional vector spaces over K. Let $n := \dim V, m := \dim W$. Let T be a linear map from V to W. Fix a basis $\mathcal{B}$ for V and a basis $\mathcal{C}$ for W.

$$\begin{tikzcd} V \arrow[r, "T"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\ K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"] \arrow[r, "\Phi_{\mathcal{C}}\cdot T\cdot \Phi_{\mathcal{B}}^{-1}"] & K^m \end{tikzcd}$$

Since it does not matter which path we take from K^n to W, we say that the diagram COMMUTES.

Recall that composition of linear maps is linear. Hence, $\Phi_{\mathcal{C}} \cdot T \cdot \Phi_{\mathcal{B}}^{-1} : K^n \to K^m$ is a linear map.

By previous knowledge (4.7), $\exists! A \in M_{m\times n}(K)$ such that

$$\Phi_{\mathcal{C}} \cdot T \cdot \Phi_{\mathcal{B}}^{-1}(x) = T_A.$$

Where $T_A : K^n \to K^m$ is the linear map defined by $T_A x := A \cdot x$.

The matrix A is called the REPRESENTATION of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. It is denoted by $[T]_{\mathcal{C}}^{\mathcal{B}}$.

Now, how do we calculate $[T]_{\mathcal{C}}^{\mathcal{B}}$? Put$A := [T]_{\mathcal{C}}^{\mathcal{B}}$.

$$\begin{tikzcd} V \arrow[r, "T"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\ K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"] \arrow[r, "T_A"] & K^m \end{tikzcd}$$

Let now $e_i \in K^n$ then under the map $\Phi_{\mathcal{B}}^{-1}$, it becomes v_i, the i-th basis vector of $\mathcal{B}$. Applying T to v_i gives $T(v_i) \in W$. Under $\Phi_{\mathcal{B}}$, this becomes $[T(v_i)]_{\mathcal{C}} \in K^m$.

In other words,

$$A = [T]_{\mathcal{C}}^{\mathcal{B}} = \begin{pmatrix} | & & | \\ [T(v_1)]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} \\ | & & | \end{pmatrix}.$$

Another way to describe this $[T]_{\mathcal{C}}^{\mathcal{B}}$ is the following: Write $[T]_{\mathcal{C}}^{\mathcal{B}} = (a_{ij})$. The entries a_{ij} are uniquely defined by:

$$T(v_j) = \sum_{i=1}^{m} a_{ij}w_i \quad \text{for } j = 1, \ldots, n.$$

(Here, $\mathcal{C} = (w_1, \ldots, w_m)$.)

Proposition 4.30:

Let $T : V \to W$ be a linear map and let $\mathcal{B}$, $\mathcal{C}$ be bases for V and W, respectively. Put $A = [T]_{\mathcal{C}}^{\mathcal{B}}$. Then $\forall v \in V$,

$$[T(v)]_{\mathcal{C}} = A \cdot [v]_{\mathcal{B}}.$$

Equivalently: $[Tv]_{\mathcal{C}} = [T]_{\mathcal{C}}^{\mathcal{B}} \cdot [v]_{\mathcal{B}}$.

Proof.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\Phi_{\mathcal{B}}"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\ K^n \arrow[r, "T_A"] & K^m \end{tikzcd}$$

This diagram commutes again. Going around the top right gives us

$$\Phi_{\mathcal{C}}(T(v)) = [T(v)]_{\mathcal{C}}.$$

Which is equal to

$$[T(v)]_{\mathcal{C}} = T_A([v]_{\mathcal{B}}) = A \cdot [v]_{\mathcal{B}}.$$

□

Example 4.31:

Let $V = K^n, W = K^m$, $T : V \to W$ be a linear map with $T = T_Q$. Take $\mathcal{B}$ and $\mathcal{C}$ to be the standard bases of K^n and K^m, then $[T]^{\mathcal{B}}_{\mathcal{C}} = Q$.

Example 4.32:

Let $V = K[x]_3$, $W = K[x]_2$ and let $D : V \to W$ with $D(f) = f'$ be the differentiation map. Take $\mathcal{B} = (1, x, x^2, x^3)$ and $\mathcal{C} = (1, x, x^2)$ as bases of V and W, respectively. We now want to find $[D]^{\mathcal{B}}_{\mathcal{C}}$.

We have

$$\begin{aligned} D(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ D(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2. \end{aligned}$$

Hence we find

$$[D]^{\mathcal{B}}_{\mathcal{C}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

4.3 Matrices

Let $T : V \to W$ and $S : W \to U$ be linear maps between finite-dimensional vector spaces. Let $\mathcal{A} = (v_1, \dots, v_n)$ be a basis for V, $\mathcal{B} = (w_1, \dots, w_m)$ a basis for W and $\mathcal{C} = (u_1, \dots, u_p)$ a basis for U.

Consider $S \circ T : V \to U$. We want to find the relation between $[S \circ T]^{\mathcal{A}}_{\mathcal{C}}$, $[T]^{\mathcal{A}}_{\mathcal{B}}$ and $[S]^{\mathcal{B}}_{\mathcal{C}}$.

Write $[T]^{\mathcal{A}}_{\mathcal{B}} = A = (a_{ij})$ and $[S]^{\mathcal{B}}_{\mathcal{C}} = B = (b_{ij})$. Furthermore, write $[S \circ T]^{\mathcal{A}}_{\mathcal{C}} = C = (c_{ij})$.

We know that

$$T(v_j) = a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m.$$

Furthermore

$$S(w_j) = b_{1j}u_1 + b_{2j}u_2 + \cdots + b_{pj}u_p.$$

Lastly,

$$S \circ T(v_j) = c_{1j}u_1 + c_{2j}u_2 + \cdots + c_{pj}u_p.$$

But $(S \circ T)(v_j) = S(T(v_j))$, so we have

$$\begin{aligned} S(T(v_j)) = S\left(\sum_{i=1}^{m} a_{ij}w_i\right) &= \sum_{i=1}^{m} a_{ij}S(w_i) \\ = \sum_{i=1}^{m} a_{ij}\left(\sum_{k=1}^{p} b_{ki}u_k\right) &= \sum_{k=1}^{p}\left(\sum_{i=1}^{m} b_{ki}a_{ij}\right)u_k. \end{aligned}$$

Lec 21 So the coefficient $c_{kj} = \sum_{i=1}^{m} b_{ki}a_{ij}$.

So

$$[S \circ T]^{\mathcal{A}}_{\mathcal{C}} = \begin{pmatrix} \cdots & c_{1j} & \cdots \\ \cdots & c_{2j} & \cdots \\ & \vdots & \\ \cdots & c_{pj} & \cdots \end{pmatrix}.$$

If we write the matrices A and B we get

$$[S \circ T]^{\mathcal{A}}_{\mathcal{C}} = \begin{pmatrix} \cdots & \cdots & \cdots \\ b_{i1} & \cdots & b_{im} \\ \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} \cdots & a_{1j} & \cdots \\ \cdots & a_{2j} & \cdots \\ & \vdots & \\ \cdots & a_{mj} & \cdots \end{pmatrix}.$$

Definition 4.33: **Matrix Multiplication**

Let $A \in M_{m\times n}(K)$ and $B \in M_{n\times p}(K)$. Write $A = (a_{ij})$ and $B = (b_{ij})$. Define a new matrix $C \in M_{m\times p}(K)$ with entries (c_{ij}), where

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

We denote this new matrix C by $A \cdot B$ and call it the MULTIPLICATION or PRODUCT of A and B.

It is important that the number of columns of A equals the number of rows of B.

Another way to describe $A \cdot B$ is the following:

$$(A) \cdot \begin{pmatrix} | & \cdots & | \\ z_1 & \cdots & z_p \\ | & \cdots & | \end{pmatrix} = \begin{pmatrix} | & \cdots & | \\ Az_1 & \cdots & Az_p \\ | & \cdots & | \end{pmatrix}.$$

Example 4.34:

Compute

$$\begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 4 & 0 \end{pmatrix}.$$

We have

$$A \cdot B = \begin{pmatrix} 7 & 8 \\ 9 & 2 \end{pmatrix}.$$

In this example, we can also compute the other way around and get:

$$B \cdot A = \begin{pmatrix} 5 & 2 & 5 \\ 1 & 0 & 2 \\ 12 & 8 & 4 \end{pmatrix}.$$

This example tells us that in general, matrix multiplication is not commutative, i.e. $A \cdot B \neq B \cdot A$ in general.

Let $A \in M_{n\times n}(K)$. Let $v \in K^n_{\text{col}}$. We define Av previously. Viewing v as a $n \times 1$ matrix, we can see that this is just a special case of matrix multiplication.

Using the same idea, we can define multiplication of a $1 \times n$ matrix with a $n \times n$ matrix. Let $w \in K^{1\times n}$ and $A \in M_{n\times n}(K)$. Then we define

$$w \cdot A = \begin{pmatrix} \sum_{k=1}^{n} w_{1k}a_{k1} & \cdots & \sum_{k=1}^{n} w_{1k}a_{kn} \end{pmatrix}.$$

which is a $1 \times n$ matrix.

So matrix multiplication can also be described like

$$\begin{pmatrix} \cdots & \cdots & \cdots \\ - & v_i & - \\ \cdots & \cdots & \cdots \end{pmatrix} \cdot (B) = \begin{pmatrix} \cdots & \cdots & \cdots \\ - & v_i \cdot B & - \\ \cdots & \cdots & \cdots \end{pmatrix}.$$

Conclusion: $[S \circ T]^{\mathcal{A}}_{\mathcal{C}} = [S]^{\mathcal{B}}_{\mathcal{C}} \cdot [T]^{\mathcal{A}}_{\mathcal{B}}$.

As a notation, let I be an index set. $\forall i, j \in I$, define

$$\delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

This is called the KRONECKER DELTA.

Define the identity matrix $I_n \in M_{n\times n}(K)$ as

$$I_n = (\delta_{ij}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Remark 4.35:

Let v be a finite dimensional vector space with dimension n. Let $\mathcal{B}$ be any basis of V. Consider $\mathrm{id}_V : V \to V$. Then $[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}} = I_n$.

Proposition 4.36:

1) $\forall A \in M_{m\times n}(K)$, $B \in M_{n\times p}(K)$, $C \in M_{p\times q}(K)$, we have

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C.$$

Hence, there is a meaning to write $A \cdot B \cdot C$ without parentheses.

2) $\forall A, B \in M_{m\times n}, C \in M_{n\times p}, C' \in M_{q\times m}$, we have

$$(A + B) \cdot C = A \cdot C + B \cdot C$$
$$C' \cdot (A + B) = C' \cdot A + C' \cdot B.$$

3) $\forall A \in M_{m\times n}(K)$, we have

$$I_m \cdot A = A = A \cdot I_n.$$

4) $\forall \alpha \in K, A \in M_{m\times n}(K), B \in M_{n\times p}(K)$, we have

$$(\alpha A) \cdot B = \alpha(A \cdot B) = A \cdot (\alpha B).$$

Proof. 1) Let $T_A : K^n \to K^m$, $T_B : K^p \to K^n$ and $T_C : K^q \to K^p$ be the linear maps defined by A, B and C, respectively. $\forall l$, denote by E_l the standard basis for K^l.

$$[T_A]^{E_n}_{E_m} = A, \quad [T_B]^{E_p}_{E_n} = B, \quad [T_C]^{E_q}_{E_p} = C.$$

We have

$$(T_A \circ T_B) \circ T_C = T_A \circ (T_B \circ T_C),$$

since composition of functions is associative. Hence

$$[(T_A \circ T_B) \circ T_C]^{E_q}_{E_m} = [T_A \circ (T_B \circ T_C)]^{E_q}_{E_m}.$$

By what we did earlier,

$$[T_A \circ T_B]^{E_q}_{E_m} \cdot [T_C]^{E_q}_{E_p} = [T_A]^{E_n}_{E_m} \cdot [T_B \circ T_C]^{E_q}_{E_n}.$$

But we can expand this further:

$$([T_A]^{E_n}_{E_m} \cdot [T_B]^{E_p}_{E_n}) \cdot [T_C]^{E_q}_{E_p} = [T_A]^{E_n}_{E_m} \cdot ([T_B]^{E_p}_{E_n} \cdot [T_C]^{E_q}_{E_p}).$$

But this is exactly what we wanted to prove.

2) If $T, S : V \to W$ are linear maps, $\mathcal{B}$ is a basis for V and $\mathcal{C}$ is a basis for W, then

$$[T + S]^{\mathcal{B}}_{\mathcal{C}} = [T]^{\mathcal{B}}_{\mathcal{C}} + [S]^{\mathcal{B}}_{\mathcal{C}}.$$

From this, the result follows directly. 3) Look at the definition of I_n in the standard basis. From this, the result follows directly.

4) The outline of the proof is to let $A = (a_ij)$ and $B = (b_ij)$ and compute both sides. □ Lec 22

Remark 4.37:

Let V be a vector space, take $\alpha \in K$. And let $Q(v) = \alpha v$. Then

$$[Q]^{\mathcal{B}}_{\mathcal{B}} = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha \end{pmatrix}.$$

Definition 4.38: Invertible Matrix

Let A_n be a $n \times n$ matrix. We say that A is INVERTIBLE if $\exists B_n$ such that

$$A \cdot B = I_n B \cdot A.$$

Remark 4.39:

If A is invertible, then B st. $A \cdot B = B \cdot A = I_n$ is unique.

Proof. Assume that B, C are both inverses of A. Then

$$B = B \cdot I_n = B \cdot (A \cdot C) = (B \cdot A) \cdot C = I_n \cdot C = C.$$

□

Definition 4.40:

Let A be an invertible matrix. We denote by A^{-1} the unique matrix such that

$$A \cdot A^{-1} = A^{-1} \cdot A = I_n.$$

A^{-1} is called the INVERSE of A.

We denote by $\mathrm{GL}_n(K)$ the set of all invertible $n \times n$ matrices over K. It is called the GENERAL LINEAR GROUP of degree n over K.

Notice that $\mathrm{GL}_n(K) \neq \varnothing$ since $I_n \in \mathrm{GL}_n(K)$. Also, $0_n \notin \mathrm{GL}_n(K)$. Hence, $\mathrm{GL}_n(K) \subsetneq M_{n\times n}(K)$.

Proposition 4.41:

Let $A, B \in \mathrm{GL}_n(K)$. Then $A \cdot B \in \mathrm{GL}_n(K)$ and also $A^{-1} \in \mathrm{GL}_n(K)$. Moreover, $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$ and $(A^{-1})^{-1} = A$.

Proof. It holds that

$$(B^{-1} \cdot A^{-1}) \cdot (A \cdot B) = B^{-1} \cdot (A^{-1} \cdot A) \cdot B = B^{-1} \cdot I_n \cdot B = I_n.$$

Also

$$(A \cdot B) \cdot (B^{-1} \cdot A^{-1}) = A \cdot (B \cdot B^{-1}) \cdot A^{-1} = A \cdot I_n \cdot A^{-1} = I_n.$$

Also

$$A \cdot A^{-1} = I_n = A^{-1} \cdot A.$$

So A is the inverse of A^{-1}. □

Corollary 4.42:
Let $A \in \mathrm{GL}_n(K)$. Then $\forall b \in K^n$, the linear system of equations $Ax = b$ has a unique solution $x \in K^n$, which is given by $x = A^{-1}b$.

Proof. If $Ax = b$, then multiplying both sides by A^{-1} gives

$$A^{-1}Ax = A^{-1}b \Rightarrow I_n x = A^{-1}b \Rightarrow x = A^{-1}b.$$

Also, if $x = A^{-1}b$, then $Ax = A(A^{-1}b) = (AA^{-1})b = I_n b = b$. □

Proposition 4.43:
Let $A \in M_{n\times n}(K)$. A is invertible if and only if $T_A : K^n \to K^n$ is an isomorphism. Moreover, $(T_A)^{-1} = T_{A^{-1}}$.

Proof. Assume first that A is invertible. We claim that $T_{A^{-1}} \circ T_A = \mathrm{id}_{K^n}$ and same for the other way around. This will show that T_A is an isomorphism with inverse $T_{A^{-1}}$.

Indeed, $\forall v \in K^n$, $T_{A^{-1}}(T_A(v)) = T_{A^{-1}}(Av) = A^{-1}(Av) = (A^{-1}A)v = I_n v = v$. Hence, $T_{A^{-1}} \circ T_A = \mathrm{id}_{K^n}$. Similarly, $T_A \circ T_{A^{-1}} = \mathrm{id}_{K^n}$.

Conversely, assume that T_A is an isomorphism. Denote $S := (T_A)^{-1}$. S is a linear map from K^n to K^n. Recall that $\forall F : K^n \to K^n$ linear, $\exists! M \in M_{n\times n}(K)$ such that $F = T_M$. So, $\exists B \in M_{n\times n}(K)$ such that $S = T_B$. Now $\forall v \in K^n$ we have:

$$v = S \circ T_A(v) = T_B(T_A(v)) = T_B(Av) = B(Av) = (BA)v.$$

Applying this to $v = e_1, v = e_2, \ldots, v = e_n$ shows that $BA = I_n$. In a similar way, we show that $AB = I_n$. Hence, A is invertible with inverse B. □

Proposition 4.44:
1) Let $A \in M_{m\times n}(K), B \in M_{n\times p}(K)$. Then

$$(A \cdot B)^T = B^T \cdot A^T.$$

2) If $A \in \mathrm{GL}_n(K)$, then $A^T \in \mathrm{GL}_n(K)$ and

$$(A^T)^{-1} = (A^{-1})^T.$$

Definition 4.45: Triangular and Diagonal Matrices
A matrix $A \in M_{n\times n}(K)$ is called:

- UPPER TRIANGULAR if $a_{ij} = 0$ for all $i > j$.
- LOWER TRIANGULAR if $a_{ij} = 0$ for all $i < j$.
- DIAGONAL if $a_{ij} = 0$ for all $i \neq j$.

Lemma 4.46:
Let $A, B \in M_{n\times n}(K)$ be upper triangular(or lower triangular or diagonal) matrices. Then $A \cdot B$ is also upper triangular(or lower triangular or diagonal, respectively).

We can ask ourselves how $[T]^{\mathcal{B}}_{\mathcal{C}}$ depends on the choice of bases $\mathcal{B}$ and $\mathcal{C}$.

Corollary 4.47:
Let V, W be finite-dimensional vector spaces over K of dimensions n and m, respectively. Let $\mathcal{B}, \mathcal{B}'$ be bases for V and $\mathcal{C}, \mathcal{C}'$ be bases for W. Let $T : V \to W$ be a linear map. Then

$$[T]^{\mathcal{B}'}_{\mathcal{C}'} = [\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}'} \cdot [T]^{\mathcal{B}}_{\mathcal{C}} \cdot [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}.$$

Proof. $T = \mathrm{id}_W \circ T \circ \mathrm{id}_V$. But this we can write as

$$\mathrm{id}_W \circ (T \circ \mathrm{id}_V).$$

This implies that

$$[T]^{\mathcal{B}'}_{\mathcal{C}'} = [\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}'} \cdot [T \circ \mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{C}} = [\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}'} \cdot [T]^{\mathcal{B}}_{\mathcal{C}} \cdot [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}.$$

□

Corollary 4.48:
Furthermore $[id_V]^{\mathcal{B}'}_{\mathcal{B}} \in \mathrm{GL}_n(K)$ and $[\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}'} \in \mathrm{GL}_m(K)$ and

$$([\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}})^{-1} = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'}, \quad ([\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}'})^{-1} = [\mathrm{id}_W]^{\mathcal{C}'}_{\mathcal{C}}.$$

Proof. We know that $\mathrm{id}_V = \mathrm{id}_V \circ \mathrm{id}_V$. Hence,

$$I_n = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}} = [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}} \cdot [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'}.$$

Similarly,

$$I_n = [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}'} = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} \cdot [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}.$$

Hence, $[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'}$ is invertible with inverse $[\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}$. □ Lec 23

Definition 4.49: Change of Basis Matrix
Let V be a finite-dimensional vector space with dimension n. Let $\mathcal{B}, \mathcal{B}'$ be bases for V. The matrix $[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} \in \mathrm{GL}_n(K)$ is called the CHANGE OF BASIS MATRIX between $\mathcal{B}$ and $\mathcal{B}'$, or the TRANSITION MATRIX between $\mathcal{B}$ and $\mathcal{B}'$.

If $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{B}' = (v'_1, \ldots, v'_n)$, then

$$[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} = \begin{pmatrix} | & & | \\ [v_1]_{\mathcal{B}'} & \cdots & [v_n]_{\mathcal{B}'} \\ | & & | \end{pmatrix}.$$

Proposition 4.50:
Let $T : V \to W$ be an isomorphism between finite-dimensional vector spaces. Let $\mathcal{B}, \mathcal{C}$ be bases for V and W, respectively. Then $[T]^{\mathcal{B}}_{\mathcal{C}} \in \mathrm{GL}_n(K)$ and

$$\left([T]^{\mathcal{B}}_{\mathcal{C}}\right)^{-1} = \left[T^{-1}\right]^{\mathcal{C}}_{\mathcal{B}}.$$

Proof. Let $n := \dim V = \dim W$. We have

$$[T^{-1}]^{\mathcal{C}}_{\mathcal{B}} \cdot [T]^{\mathcal{B}}_{\mathcal{C}} = [T^{-1} \circ T]^{\mathcal{B}}_{\mathcal{B}} = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}} = I_n.$$

Similarly,

$$[T]^{\mathcal{B}}_{\mathcal{C}} \cdot [T^{-1}]^{\mathcal{C}}_{\mathcal{B}} = [T \circ T^{-1}]^{\mathcal{C}}_{\mathcal{C}} = [\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{C}} = I_n.$$

□

Consider now $T : V \to V$. Let $\mathcal{B}, \mathcal{B}'$ be bases for V. The question is, what's the relation between $[T]^{\mathcal{B}}_{\mathcal{B}}$ and $[T]^{\mathcal{B}'}_{\mathcal{B}'}$?

Corollary 4.51:
Let $T : V \to V$ be a linear map and let $\mathcal{B}, \mathcal{B}'$ be bases for V. Then

$$[T]^{\mathcal{B}'}_{\mathcal{B}'} = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} \cdot [T]^{\mathcal{B}}_{\mathcal{B}} \cdot [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}.$$

where $[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} = ([\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}})^{-1}$.

Definition 4.52: Matrix Equivalence

1) Let $A, B \in M_{m\times n}(K)$. We say that A and B are EQUIVALENT if $\exists P \in \mathrm{GL}_m(K)$ and $Q \in \mathrm{GL}_n(K)$ such that

$$B = P \cdot A \cdot Q.$$

2) Let $A, B \in M_{n\times n}(K)$. We say that A and B are SIMILAR if $\exists P \in \mathrm{GL}_n(K)$ such that

$$B = P^{-1} \cdot A \cdot P.$$

Equivalence between matrices is an equivalence relation on the set of all $m \times n$ matrices over K.

Furthermore, $A, B \in M_{n\times n}(K)$ are equivalent if and only if $\exists$ two vector spaces V, W of dimension n, m, and bases $\mathcal{B}, \mathcal{B}'$ of V and $\mathcal{C}, \mathcal{C}'$ of W and a linear map $T : V \to W$ such that

$$A = [T]^{\mathcal{B}}_{\mathcal{C}}, \quad B = [T]^{\mathcal{B}'}_{\mathcal{C}'}.$$

Similarity enjoys the same properties, with $V = W$.

As a notation, the zero $p \times q$ matrix is denoted by $0_{p,q}$.

Proposition 4.53:

Let V, W be finite-dimensional vector spaces over K of dimensions n and m, respectively. Let $T : V \to W$ be a linear map with $\mathrm{rank}(T) = r$. Then $\exists$ bases $\mathcal{B}$ of V and $\mathcal{C}$ of W such that

$$[T]^{\mathcal{B}}_{\mathcal{C}} = \left(\begin{array}{c|c} I_r & 0_{r,n-r} \\ \hline 0_{m-r,r} & 0_{m-r,n-r} \end{array}\right) \in M_{m\times n}(K).$$

Proof. Put $l = \dim(\ker(T))$. By the rank theorem (4.20), we have $r = n - l$. Let $v_1, \dots, v_l$ be a basis for $\ker(T)$. We can extend it to a basis $(v_1, \dots, v_n)$ of V. We'll take

$$\mathcal{B} = (v_{l+1}, \dots, v_n, v_1, \dots, v_l).$$

In the proof of the rank theorem, we saw that

$$(T(v_{l+1}), \dots, T(v_n))$$

is a basis for $\mathrm{Im}(T)$. Define $w_i := T(v_{l+i})$ for $i = 1, \dots, r$. (note that $l + r = n$). We can extend $(w_1, \dots, w_r)$ to a basis

$$\mathcal{C} = (w_1, \dots, w_r, w_{r+1}, \dots, w_m)$$

of W.

So now, we can construct the matrix $[T]^{\mathcal{B}}_{\mathcal{C}}$.

$$[T]^{\mathcal{B}}_{\mathcal{C}} = \begin{pmatrix} | & & | & | & & | \\ [T(v_{l+1})]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} & [T(v_1)]_{\mathcal{C}} & \cdots & [T(v_l)]_{\mathcal{C}} \\ | & & | & | & & | \end{pmatrix}.$$

Now $[T(v_i)]_{\mathcal{C}} = 0$ for $i = 1, \dots, l$ since $v_i \in \ker(T)$. Also, for $j = 1, \dots, r$, we have

$$T(v_{l+j}) = w_j = 1 \cdot w_j + 0 \cdot w_{r+1} + \dots + 0 \cdot w_m.$$

So we can write $[T(v_{l+j})]_{\mathcal{C}}$ as desired. $\square$

Corollary 4.54:

Let $A \in M_{m\times n}(K)$. Then $\exists P \in \mathrm{GL}_m(K)$ and $Q \in \mathrm{GL}_n(K)$ such that

$$P \cdot A \cdot Q = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

where $r = \text{col-rank}(A)$.

Proof. Consier $T_A : K^n \to K^m$. Then $r := \text{col-rank}(A)$. From the definition of column rank, we have

$$\mathrm{ColS}(A) = \mathrm{Im}(T_A).$$

So also $r = \mathrm{rank}(T_A)$. By 4.53, $\exists$ bases $\mathcal{B}$ of K^n and $\mathcal{C}$ of K^m such that

$$[T_A]^{\mathcal{B}}_{\mathcal{C}} = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

We also know that

$$[T_A]^{\mathcal{B}}_{\mathcal{C}} = [\mathrm{id}_{K^m}]^{E_m}_{\mathcal{C}} \cdot A \cdot [\mathrm{id}_{K^n}]^{\mathcal{B}}_{E_n}.$$

If we now call $P := [\mathrm{id}_{K^m}]^{E_m}_{\mathcal{C}} \in \mathrm{GL}_m(K)$ and $Q := [\mathrm{id}_{K^n}]^{\mathcal{B}}_{E_n} \in \mathrm{GL}_n(K)$, we are done. $\square$

A consequence for this corollary is that $\text{col-rank}(A) \leq \min(m, n)$. This is because $r = \text{col-rank}(A)$ and A is $m \times n$.

Corollary 4.55:

Let $A, B \in M_{m\times n}(K)$. Then A and B are equivalent if and only if $\text{col-rank}(A) = \text{col-rank}(B)$.

Proof. Assume $\text{col-rank}(A) = \text{col-rank}(B) = r$. By 4.54,

$$A \sim \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right) \sim B.$$

By transitivity of equivalence, we have $A \sim B$.

Assume now that $A \sim B$. Then $\exists P \in \mathrm{GL}_m(K)$ and $Q \in \mathrm{GL}_n(K)$ such that $B = P \cdot A \cdot Q$. Hence $T_B = T_P \circ T_A \circ T_Q$. Then

$$\begin{aligned} \mathrm{Im}(T_B) &= T_B(K^n) = T_P(T_A(T_Q(K^n))) \\ &= T_P(T_A(K^n)) = T_P(\mathrm{Im}(T_A)) \\ &= \mathrm{Im}(T_A). \end{aligned}$$

Hence, $\mathrm{rank}(T_B) = \mathrm{rank}(T_A)$, so $\text{col-rank}(B) = \text{col-rank}(A)$. $\square$

Theorem 4.56: Row Rank = Column Rank

Let $A \in M_{m\times n}(K)$. Then

$$\text{row-rank}(A) = \text{col-rank}(A).$$

This common number is denoted by $\mathrm{rank}(A)$ and is called the RANK of A. Lec 24

To show this, we need some preparation.

Lemma 4.57:

Let $T : V \to W$ be a linear map and let $S : W \to U$ be an isomorphism and $L : P \to V$ also an isomorphism. Then

1) $\mathrm{rank}(S \circ T) = \mathrm{rank}(T)$.

2) $\mathrm{rank}(T \circ L) = \mathrm{rank}(T)$.

Proof. 1) We have

$$\operatorname{rank}(S \circ T) = \dim(\operatorname{Im}(S \circ T)) = \dim(S(\operatorname{Im}(T))).$$

Since S is an isomorphism, it is bijective. Hence,

$$\dim(S(\operatorname{Im}(T))) = \dim(\operatorname{Im}(T)) = \operatorname{rank}(T).$$

2) We have

$$\operatorname{rank}(T \circ L) = \dim(\operatorname{Im}(T \circ L)) = \dim(T(L(P))).$$

Since L is an isomorphism, it is bijective. Hence,

$$\dim(T(L(P))) = \dim(T(V)) = \operatorname{rank}(T).$$

□

Lemma 4.58:

Let $A \in M_{n\times n}(K)$. Then A is invertible iff $T_A : K^n \to K^n$ is an isomorphism. Moreover, $T_{A^{-1}} = T_A^{-1}$.

Proof. This is exactly Proposition 4.43. □

Lemma 4.59:

Let $A, B \in M_{m\times n}(K)$. Let $P \in \mathrm{GL}_m(K)$ and $Q \in \mathrm{GL}_n(K)$. Then:

1) $\text{col-rank}(A) = \text{col-rank}(PAQ)$.

2) $\text{row-rank}(A) = \text{row-rank}(PAQ)$.

In other words, if $B \sim A$, then $\text{col-rank}(B) = \text{col-rank}(A)$ and $\text{row-rank}(B) = \text{row-rank}(A)$.

Proof. Put $B := PAQ$. Then $T_B = T_P \circ T_A \circ T_Q$. Since P and Q are invertible, by lemma 4.58, T_P and T_Q are isomorphisms.

By lemma 4.57, $\operatorname{rank}(T_B) = \operatorname{rank}(T_A)$. Hence, $\text{col-rank}(B) = \text{col-rank}(A)$.

For the row rank, we have

$$B^T = (PAQ)^T = Q^T A^T P^T.$$

Since P, Q are invertible, so are P^T, Q^T. By the same argument as before, $\text{col-rank}(B^T) = \text{col-rank}(A^T)$. Hence, $\text{row-rank}(B) = \text{row-rank}(A)$. □

Proof. [Theorem 4.56] By Corollary 4.54, $\exists P \in \mathrm{GL}_m(K)$ and $Q \in \mathrm{GL}_n(K)$ such that

$$PAQ = \left(\begin{array}{c:c} I_r & 0 \\ \hdashline 0 & 0 \end{array}\right),$$

for some $r \geq 0$. Denote by B the matrix on the right-hand side.

Clearly, $\text{col-rank}(B) = r$ and $\text{row-rank}(B) = r$. By lemma 4.59,

$$\text{col-rank}(A) = \text{col-rank}(B) = r.$$

Similarly,

$$\text{row-rank}(A) = \text{row-rank}(B) = r.$$

Thus, $\text{col-rank}(A) = \text{row-rank}(A)$. □

Notice that $\operatorname{Im}(T_A) = \operatorname{ColS}(A)$. Hence, also

$$\operatorname{rank}(A) = \dim(\operatorname{ColS}(A)) = \operatorname{rank}(T_A).$$

Remark 4.60:

Suppose A' is a row reduced echelon matrix which is row-equivalent to A. We know that $RowS(A) = RowS(A')$. Hence, $\operatorname{rank}(A) = \operatorname{rank}(A')$ which is exactly the number of pivots in A'.

Let $A \in M_{m\times n}(K)$. Consider the system of equations $Ax = b$, where $b \in K^m$. Denote by $\operatorname{Sol}(A, b) = x \in K^n | Ax = b$ the set of solutions.

We know that for $b = 0$, $\operatorname{Sol}(A, 0)$ is a subspace of K^n.

Claim 4.61:

It holds that

$$\dim \operatorname{Sol}(A, 0) = n - \operatorname{rank}(A).$$

Proof. This is because $\operatorname{Sol}(A, 0) = \ker(T_A)$. By the rank theorem,

$$\dim(\ker(T_A)) = n - \dim(\operatorname{Im}(T_A)) = n - \operatorname{rank}(A).$$

□

Let A' be a row reduced echelon matrix which is row-equivalent to A. The number $n - \operatorname{rank}(A)$ is called the NUMBER OF FREE VARIABLES of the system $Ax = b$.

Denote by $x_{i1}, \dots, x_{il}$ the free variables of the system $Ax = b$ where $1 \leq i_1 < \dots < i_l \leq n$ and $l = n - \operatorname{rank}(A)$. Denote by $x_{j1}, \dots, x_{jr}$ the pivot variables where $1 \leq j_1 < \dots < j_r \leq n$ and $r = \operatorname{rank}(A)$.

For each choice of values for the free variables, the pivot variables are uniquely determined. In other words, we have a linear map

$$\Phi : K^l \to \operatorname{Sol}(A, 0).$$

To be more particular, $\Phi(\lambda_1, \dots, \lambda_l)$ is the unique solution on $Ax = 0$ such that $x_{i1} = \lambda_1, \dots, x_{il} = \lambda_l$.

Proposition 4.62:

The map Φ is linear. Moreover, it is an isomorphism.

Proof. Consider row number k of A' and of $A'x = 0$. If $1 \leq k \leq r$, then it looks like $x_{j_k} +$ (linear combination of free variables that come after j_k) $= 0$.

So, $x_{j_k} = -\sum_{\{q|1\leq q\leq l, i_q > j_k\}} a'_{k,i_q} x_{i_q}$.

Hence, the pivots depend linearly on the free variables. Thus, Φ is linear.

To show that Φ is an isomorphism, we show that $\operatorname{Ker}(\Phi) = \{0\}$ and that Φ is surjective.

If $\Phi(\lambda_1, \dots, \lambda_l) = 0$, then by definition, $\lambda_1 = \dots = \lambda_l = 0$.

But $\Phi : K^l \to \operatorname{Sol}(A, 0)$ and $\dim \operatorname{Sol}(A, 0) = n - \operatorname{rank}(A) = l$. Hence, Φ is surjective, which concludes the proof. □

Notice that from this proposition, we also have

$$\Phi^{-1} : \operatorname{Sol}(A, 0) \to K^l,$$

which maps each solution $x \in \operatorname{Sol}(A, 0)$ to the corresponding values of the free variables.

Corollary 4.63:

Let $b \in K^m, A \in M_{m\times n}(K)$. Then

1) $\operatorname{Sol}(A, b) \neq \varnothing$ iff $b \in \operatorname{ColS}(A)$. 2) If $\operatorname{Sol}(A, b) \neq \varnothing$ and $y \in \operatorname{Sol}(A, b)$, then

$$\operatorname{Sol}(A, b) = y + \operatorname{Sol}(A, 0) = \{y + x | x \in \operatorname{Sol}(A, 0)\}.$$

Proof. Exc with hints.

1) If $A = (v_1 \cdots v_n)\,, x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, then $Ax = x_1v_1 + \cdots + x_nv_n$.

2) If $x \in \mathrm{Sol}(A, 0)$, then $A(y + x) = Ay + Ax = b + 0 = b$.

If $z \in \mathrm{Sol}(A, b)$, then $A(z - y) = Az - Ay = b - b = 0$. □

Proposition 4.64:
Let $A \in M_{m\times n}(K)$ and $b \in K^m$. Then the following statements are equivalent:

1) $\mathrm{rank}(A|b) = \mathrm{rank}(A)$,

2) The system of equations $Ax = b$ has a solution,

Proof. Write $A = (v_1 \quad \cdots \quad v_n)$. Then statement 1) is equivalent to

$$\text{col-rank}(A|b) = \text{col-rank}(A).$$

This is the same as saying that $\mathrm{Sp}(v_1, \ldots, v_n, b) = \mathrm{Sp}(v_1, \ldots, v_n)$, which is equivalent to $b \in \mathrm{Sp}(v_1, \ldots, v_n) = \mathrm{ColS}(A)$. But this is equivalent to statement 2) by Corollary 4.63. □

Proposition 4.65:
Let $A \in M_{m\times n}(K)$ and $b \in K^m$. Then the following statements are equivalent:

1) $\mathrm{rank}(A) = \mathrm{rank}(A|b) = n$,

2) The system of equations $Ax = b$ has a unique solution.

Proof. [outline] $\mathrm{rank}(A) = n \Leftrightarrow \dim \mathrm{Sp}(v_1, \ldots, v_n) = n \Leftrightarrow (v_1, \ldots, v_n)$ form a basis of K^m.

□

Lec 25

4.4 Elementary Row Operations

The main point of this section is to show that elementary row operations can be written by matrix multiplication. As notation, let $n \in \mathbb{Z}_{\geq 1}$ and let $1 \leq i, j \leq n$. Denote by $E_{ij} \in M_{n\times n}(K)$ the matrix with 1 at position (i, j) and 0 elsewhere.

$$E_{ij} = \begin{array}{c} \begin{array}{ccccc} & & \text{column j} & & \end{array} \\ \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{array}{c} \\ \\ \Leftarrow \text{row i} \\ \\ \\ \end{array} \end{array}.$$

We'll define now three other types of matrices.

Type 1: Let $i \neq j$ and let $\alpha \in K$. Define $Q_{ij}(\alpha)$ to be the matrix obtained from I_n by adding α times row j to row i. We have

$$Q_{ij}(\alpha) = I_n + \alpha E_{ij}.$$

Type 2: Define P_{ij} to be the matrix obtained from I_n by swapping rows i and j. We have

$$P_{ij} = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}.$$

Type 3: Let $\alpha \in K, \alpha \neq 0$. Define $S_i(\alpha)$ to be the matrix obtained from I_n by multiplying row i by α. We have

$$S_i(\alpha) = I_n + (\alpha - 1)E_{ii}.$$

Example 4.66:
In $M_{3x3}(K)$, we have

$$Q_{1,3}(\alpha) = \begin{pmatrix} 1 & 0 & \alpha \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$P_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$S_2(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Lemma 4.67:
Let $A \in M_{n\times p}(K)$. Then:

1) Multiplying A from the left by $Q_{ij}(\alpha)$ results in applying the row-operation $R_i + \alpha R_j \to R_i$ to A.

2) Multiplying A from the left by P_{ij} results in applying the row-operation $R_i \leftrightarrow R_j$ to A.

3) Multiplying A from the left by $S_i(\alpha)$ results in applying the row-operation $\alpha R_i \to R_i$ to A.

Proof. [Outline] Can be done by explicit straightforward calculation. For example, for 1):

$$Q_{ij}(\alpha)A = (I_n + \alpha E_{ij})A = A + \alpha E_{ij}A.$$

□

Similarly to elementary row operations, we can define elementary column operations. They are defined as follows:

Type 1': Multiply A from the rigth by $Q_{ij}(\alpha)$ results in the operation $C_j + \alpha C_i \to C_j$.

Type 2': Multiplying A from the right by P_{ij} results in the operation $C_i \leftrightarrow C_j$.

Type 3': Multiplying A from the right by $S_i(\alpha)$ results in the operation $\alpha C_i \to C_i$.

To prove that these operations work as intended, one can use the fact that

$$(AB)^T = B^T A^T.$$

Lemma 4.68:
Every elementary matirx is invertible and the inverse of each of them is also an elementary matrix. In fact

$$Q_{ij}(\alpha)^{-1} = Q_{ij}(-\alpha)$$
$$P_{ij}^{-1} = P_{ji}$$
$$S_i(\alpha)^{-1} = S_i\left(\alpha^{-1}\right).$$

Proof. [Outline] We can check this by using the definition. For example, $Q_{ij}(\alpha)Q_{ij}(-\alpha)$ is the same as applying the row operation $R_i + \alpha R_j \to R_i$ followed by $R_i - \alpha R_j \to R_i$, which results in the identity operation. Similarly one shows $Q_{ij}(-\alpha)Q_{ij}(\alpha) = I_n$.

The other proofs are similar. □

Theorem 4.69: Lec 26

For every $A \in \mathrm{GL}_n(K)$, there exist $k \geq 1$ and elementary matrices $T_1, \ldots, T_k$ such that

$$T_k T_{k-1} \cdots T_1 A = I_n.$$

Furthermore, $A = T_1^{-1} T_2^{-1} \cdots T_k^{-1}$ and

$$A^{-1} = T_k T_{k-1} \cdots T_1.$$

Proof. We proved that every matrix $A \in M_{n\times n}(K)$ can be brought to a row reduced echelon form by a finite number of elementary row operations.

Since A is invertible, $\mathrm{rank}(A) = n$. Hence, the row reduced echelon form of A is I_n.

But A' is obtained from A by a sequence of elementary row operations, hence $\exists$ elementary matrices $T_1, \ldots, T_k$ such that

$$T_k T_{k-1} \cdots T_1 A = I_n.$$

The rest follows by straightforward manipulation using the invertibility of the elementary matrices. □

Theorem 4.70:

Let $A \in M_{n\times n}(K)$. Place A together with the identity matrix on the right to A as augmented matrix $(A|I_n)$.

Then, A is invertible iff the matrix $(A|I_n)$ can be brought to the form $(I_n|B)$ by a finite number of elementary row operations. In this case, $B = A^{-1}$.

To prove this, we need the following lemma.

Lemma 4.71:

Let B, C, D be $n \times n$ matrices. Then

$$B(C|D) = (BC|BD).$$

Proof. Exercise. (by direct calculation) □

Proof. [of Theorem 4.70] Assume A is invertible. By theorem 4.69, there exist elementary matrices $T_1, \ldots, T_k$ such that

$$T_k T_{k-1} \cdots T_1 A = I_n.$$

By lemma 4.71, we have

$$T_k T_{k-1} \cdots T_1 (A|I_n) = (I_n | T_k T_{k-1} \cdots T_1).$$

By theorem 4.69, $T_k T_{k-1} \cdots T_1 = A^{-1}$.

Conversely, assume that $(A|I_n)$ can be brought to $(I_n|B)$ by elementary row operations. Then, A itself can be brought to I_n by the same row operations. Hence, A is row-equivalent to I_n. By theorem 4.69,A is invertible. □

Example 4.72: Matrix inversion

Invert the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ over Q.

Solution.

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right)$$

4.5 The Dual Space

Given V, W as vector spaces, $\mathrm{Hom}(V, W)$ is also a vector space. If V and W are finite-dimensional, then $\mathrm{Hom}(V, W)$ is also finite-dimensional and

$$\dim(\mathrm{Hom}(V, W)) = \dim(V) \cdot \dim(W).$$

Proposition 4.73: The Hom Space is a Vector Space

$\mathrm{Hom}(V, W)$ has the structure of a vector space over K with the following operations

$$(T_1 + T_2)(v) := T_1(v) + T_2(v)$$
$$(\alpha T)(v) := \alpha(T(v)).$$

Proof. We know that the set of all functions from V to W is a vector space with the above operations. We need to show that $\mathrm{Hom}(V, W)$ is a subspace of this vector space.

First off all, the zero map $0 : V \to W$ defined by $0(v) = 0_W$ for all $v \in V$ is linear. Hence, $0 \in \mathrm{Hom}(V, W)$.

We claim that $T_1 + T_2$ is linear. Indeed, for $v_1, v_2 \in V$ and $\alpha, \beta \in K$, we have

$$\begin{aligned}(T_1 + T_2)(\alpha v_1 + \beta v_2) &= T_1(\alpha v_1 + \beta v_2) + T_2(\alpha v_1 + \beta v_2) \\ &= \alpha T_1(v_1) + \beta T_1(v_2) + \alpha T_2(v_1) + \beta T_2(v_2) \\ &= \alpha(T_1(v_1) + T_2(v_1)) + \beta(T_1(v_2) + T_2(v_2)) \\ &= \alpha(T_1 + T_2)(v_1) + \beta(T_1 + T_2)(v_2).\end{aligned}$$

So indeed $T_1 + T_2 \in \mathrm{Hom}(V, W)$.

Next, one shows that αT is linear for $\alpha \in K$ and $T \in \mathrm{Hom}(V, W)$. The proof is similar to the one above.

Hence, $\mathrm{Hom}(V, W)$ is closed under addition and scalar multiplication. Therefore, $\mathrm{Hom}(V, W)$ is a subspace of the vector space of all functions from V to W. Hence, $\mathrm{Hom}(V, W)$ is a vector space. □

The neutral element of $\mathrm{Hom}(V, W)$ is the zero map $0 : V \to W$ defined by $0(v) = 0_W$ for all $v \in V$.

Theorem 4.74:

Let V, W be finite-dimensional vector spaces and $\mathcal{B}$ a basis for V and $\mathcal{C}$ a basis for W. Define

$$\begin{aligned}\Psi^{\mathcal{B}}_{\mathcal{C}} : \mathrm{Hom}(V, W) &\to M_{m\times n}(K) \\ T &\mapsto [T]^{\mathcal{B}}_{\mathcal{C}}.\end{aligned}$$

Then, $\Psi^{\mathcal{B}}_{\mathcal{C}}$ is an isomorphism.

Proof. Let $T \in \mathrm{Hom}(V, W)$ and $M := [T]^{\mathcal{B}}_{\mathcal{C}} \in M_{m\times n}(K)$. Recall the following
$$\begin{tikzcd} V \arrow[r, "T"] & W \arrow[d, "\Phi_{\mathcal{C}}"] \\ K^n \arrow[u, "\Phi_{\mathcal{B}}^{-1}"] \arrow[r, "T_M"] & K^m \end{tikzcd}$$
where

$$\Phi_{\mathcal{B}}(v) := [v]_{\mathcal{B}}$$
$$\Phi_{\mathcal{C}}(w) := [w]_{\mathcal{C}}.$$

We have seen that then $\Phi_{\mathcal{B}}^{-1} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \sum_{i=1}^n a_i v_i$ where we used the basis $\mathcal{B} = (v_1, \ldots, v_n)$.

We first claim, that $\Psi^{\mathcal{B}}_{\mathcal{C}}$ is bijective. To do this, define

$$\begin{aligned}\Theta : M_{m\times n}(K) &\longrightarrow \mathrm{Hom}(V, W) \\ A &\longmapsto \Theta(A) = \Phi_{\mathcal{C}}^{-1} \circ T_A \circ \Phi_{\mathcal{B}}.\end{aligned}$$

The corresponding commutative diagram is
$$\begin{tikzcd} V \arrow[r, "\Theta(A)"] \arrow[d, "\Phi_{\mathcal{B}}"'] & W \\ K^n \arrow[r, "T_A"] & K^m \arrow[u, "\Phi_{\mathcal{C}}^{-1}"'] \end{tikzcd}$$
. Consider now,
$$\Psi_{\mathcal{C}}^{\mathcal{B}} \circ \Theta : M_{m\times n}(K) \to M_{m\times n}(K).$$
Let $A \in M_{m\times n}(K)$. Then
$$\Psi_{\mathcal{C}}^{\mathcal{B}} \circ \Theta(A) = \Psi_{\mathcal{C}}^{\mathcal{B}}\left(\Phi_{\mathcal{C}}^{-1} \circ T_A \circ \Phi_{\mathcal{B}}\right) = [\Phi_{\mathcal{C}}^{-1} \circ T_A \circ \Phi_{\mathcal{B}}]_{\mathcal{C}}^{\mathcal{B}} = T_A \Phi_{\mathcal{B}} = A.$$
The last step follows from the definition of T_A and the fact that
$$[\Phi_{\mathcal{B}}(v_j)]_{\mathcal{B}} = [v_j]_{\mathcal{B}} = e_j.$$
Hence, $\Psi_{\mathcal{C}}^{\mathcal{B}} \circ \Theta = \mathrm{id}_{M_{m\times n}(K)}$. Consider now $\Theta \circ \Psi_{\mathcal{C}}^{\mathcal{B}} : \mathrm{Hom}(V,W) \to \mathrm{Hom}(V,W)$ applied to $T \in \mathrm{Hom}(V,W)$.
$$\Theta \circ \Psi_{\mathcal{C}}^{\mathcal{B}}(T) = \Theta([T]_{\mathcal{C}}^{\mathcal{B}}) = \Phi_{\mathcal{C}}^{-1} \circ T_{[T]_{\mathcal{C}}^{\mathcal{B}}} \circ \Phi_{\mathcal{B}} = T.$$
Hence, $\Theta \circ \Psi_{\mathcal{C}}^{\mathcal{B}} = \mathrm{id}_{\mathrm{Hom}(V,W)}$. Therefore, $\Psi_{\mathcal{C}}^{\mathcal{B}}$ is bijective with inverse Θ.

It remains to show that $\Psi_{\mathcal{C}}^{\mathcal{B}}$ is linear. Indeed, let $\alpha \in K$, and $T \in \mathrm{Hom}(V,W)$. Then
$$\Psi_{\mathcal{C}}^{\mathcal{B}}(\alpha T) = \begin{pmatrix} | & \cdots & | \\ [\alpha T(v_1)]_{\mathcal{C}} & \cdots & [\alpha T(v_n)]_{\mathcal{C}} \\ | & \cdots & | \end{pmatrix}$$
$$= \alpha \begin{pmatrix} | & \cdots & | \\ [T(v_1)]_{\mathcal{C}} & \cdots & [T(v_n)]_{\mathcal{C}} \\ | & \cdots & | \end{pmatrix} = \alpha \Psi_{\mathcal{C}}^{\mathcal{B}}(T).$$
Similarly, one shows that
$$\Psi_{\mathcal{C}}^{\mathcal{B}}(T_1 + T_2) = \Psi_{\mathcal{C}}^{\mathcal{B}}(T_1) + \Psi_{\mathcal{C}}^{\mathcal{B}}(T_2).$$
□

Corollary 4.75: Dimension of the Hom Space

If V, W are finite-dimensional vector spaces, then so is $\mathrm{Hom}(V,W)$ and
$$\dim(\mathrm{Hom}(V,W)) = \dim(V) \cdot \dim(W).$$

Proof. Let $\dim(V) = n$ and $\dim(W) = m$. By theorem 4.74, $\mathrm{Hom}(V,W) \cong M_{m\times n}(K)$. But $\dim(M_{m\times n}(K)) = mn$. Hence,
$$\dim(\mathrm{Hom}(V,W)) = \dim(M_{m\times n}(K)) = mn = \dim(V)\cdot\dim(W).$$
□

Definition 4.76: Ring

A RING has the same axioms as a field, except that multiplication need not be commutative and that multiplicative inverses need not exist for every nonzero element.

In a ring, the element 1 is called the UNITY of R. If a ring is commutative, we say that it is a COMMUTATIVE RING.

Example 4.77:

1) Every field is a commutative ring.

2) $R = \mathbb{Z}$ is a commutative ring.

3) The set $M_{n\times n}(K)$ is a ring with the usual matrix addition and multiplication. It is not commutative for $n \geq 2$.

Let R, S be rings. A homomorphism of rings is a map $\varphi : R \to S$ such that
$$\varphi(a+b) = \varphi(a)+\varphi(b), \quad \varphi(ab) = \varphi(a)\varphi(b), \quad \varphi(1_R) = 1_S.$$

Corollary 4.78:

Let V be a vector space over K. Then, $\mathrm{End}(V) := \mathrm{Hom}(V,V)$ is a ring with the usual addition and composition of linear maps. The unity is the identity map id_V. Moreover, if V is finite-dimensional, and $\mathcal{B}$ is a basis of V, then the map
$$\Psi_{\mathcal{B}}^{\mathcal{B}} : \mathrm{End}(V) \to M_{n\times n}(K)$$
$$T \mapsto [T]_{\mathcal{B}}.$$
is an isomorphism of rings.

Proof. $\forall T_1, T_2, T_3 \in \mathrm{End}(V)$ and $\alpha \in K$, we have
$$T_1 \circ (T_2 \circ T_3) = (T_1 \circ T_2) \circ T_3.$$
Thus multiplication is associative. The identity map id_V satisfies
$$\mathrm{id}_V \circ T = T \circ \mathrm{id}_V = T.$$
So id_V is the unity of $\mathrm{End}(V)$. Also,
$$T_1 \circ (T_2 + T_3) = T_1 \circ T_2 + T_1 \circ T_3,$$
because T is linear.

One can check all the other ring axioms similarly.

Assume now that V is finite-dimensional etc. We already know that $\Psi_{\mathcal{B}}^{\mathcal{B}}(T_1 + T_2) = \Psi_{\mathcal{B}}^{\mathcal{B}}(T_1) + \Psi_{\mathcal{B}}^{\mathcal{B}}(T_2)$ and that $\Psi_{\mathcal{B}}^{\mathcal{B}}$ is bijective. It remains to show that
$$\Psi_{\mathcal{B}}^{\mathcal{B}}(T_1 \circ T_2) = \Psi_{\mathcal{B}}^{\mathcal{B}}(T_1) \cdot \Psi_{\mathcal{B}}^{\mathcal{B}}(T_2).$$
But this is exactly Proposition 4.41.
$$[T_1 \circ T_2]_{\mathcal{B}}^{\mathcal{B}} = [T_1]_{\mathcal{B}}^{\mathcal{B}} \cdot [T_2]_{\mathcal{B}}^{\mathcal{B}} = \Psi_{\mathcal{B}}^{\mathcal{B}}(T_1) \cdot \Psi_{\mathcal{B}}^{\mathcal{B}}(T_2).$$
Finally,
$$\Psi_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = [\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}} = I_n.$$
□ Lec 27

A special case of the Hom space is the dual space.

Definition 4.79: Dual Space

Let V be a vector space over K. The DUAL SPACE of V is defined as
$$V^* := \mathrm{Hom}(V, K).$$
The elements of V^* are linear maps $l : V \to K$ and are called FUNCTIONALS on V.

Corollary 4.80:

Let V be a finite-dimensional vector space over K. Then, V^* is also finite-dimensional and
$$\dim(V^*) = \dim(V).$$

Proof. This is a special case of corollary 4.75 with $W = K$. □

Example 4.81:

Take $V = K^n_{\mathrm{col}}$. We will identify V^* with K^n_{row}. If $l \in K^n_{\mathrm{row}}$, the l defines a functional $f_l : K^n_{\mathrm{col}} \to K$ by $f_l(v) = l \cdot v$ where we view l as a $1\times n$ matrix and v as an $n \times 1$ matrix. The map $D : K^n_{\mathrm{row}} \to (K^n_{\mathrm{col}})^*$ defined by $D(l) = f_l$ is an isomorphism of vector spaces. As an exercise, check that D is linear and an isomorphism.

Let V be a finite-dimensional vector space over K and let $\mathcal{B} = (v_1, \ldots, v_n)$ be a basis of V.

Define n functionals $v_1^*, \dots, v_n^* \in V^*$ by

$$v_i^*(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Since $\mathcal{B}$ is a basis for V, this determines v_i^* uniquely.

Lemma 4.82:

The elements $v_1^*, \dots, v_n^*$ form a basis of V^*. This basis is called the DUAL BASIS of $\mathcal{B}$ and is denoted by $\mathcal{B}^* = (v_1^*, \dots, v_n^*)$.

Moreover, $\forall l \in V^*$,

$$l = \sum_{i=1}^{n} l(v_i) v_i^*.$$

In other words, $l(v) = \sum_{i=1}^{n} l(v_i) v_i^*(v)$ for all $v \in V$.

Proof. Let $l \in V^*$. Let $l \in V^*$. We claim that

$$l = \sum_{i=1}^{n} l(v_i) v_i^*. \tag{4.2}$$

In other words, $\forall v \in V$,

$$l(v) = \sum_{i=1}^{n} l(v_i) v_i^*(v). \tag{4.3}$$

The lefthand side of (4.2) is a linear map $V \to K$ and so must the righthand side. Hence, to show LHS = RHS, it suffices to check (4.3) for $v \in \mathcal{B}$.

So we will check (4.3) for $v = v_1, \dots, v_n$. Let $1 \leq k \leq n$, and substitute $v = v_k$ in (4.3). We get

$$\text{RHS}(2) = \sum_{i=1}^{n} l(v_i) \underbrace{v_i^*(v_k)}_{=\delta_{ij}} = l(v_k).$$

The formula we've just proven shows also that $v_1^*, \dots, v_n^*$ span V^*. But $\dim(V^*) = n$. Hence, $v_1^*, \dots, v_n^*$ form a basis of V^*. □

Let V be a finite-dimensional vector space over K. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for V. Then we also get two dual bases $\mathcal{B}^*$ and $\mathcal{C}^*$ for V^*. What is the relationship between

$$[\text{id}_V]_{\mathcal{C}}^{\mathcal{B}} \quad \text{and} \quad [\text{id}_{V^*}]_{\mathcal{B}^*}^{\mathcal{C}^*}?.$$

Proposition 4.83:

It holds that

$$[\text{id}_{V^*}]_{\mathcal{B}}^{\mathcal{C}} = \left([\text{id}_V]_{\mathcal{C}}^{\mathcal{B}}\right)^T.$$

Furthermore,

$$[id_V^*]_{\mathcal{C}}^{\mathcal{B}} = \left(\left([\text{id}_V]_{\mathcal{C}}^{\mathcal{B}}\right)^T\right)^{-1}.$$

Proof. Write $\mathcal{B} = (v_1, \dots, v_n), \mathcal{C} = (w_1, \dots, w_n)$. Then

$$A := [\text{id}_V]_{\mathcal{C}}^{\mathcal{B}} = \begin{pmatrix} | & \cdots & | \\ [v_1]_{\mathcal{C}} & \cdots & [v_n]_{\mathcal{C}} \\ | & \cdots & | \end{pmatrix}.$$

We also write $A = (a_{ij})$. We have

$$v_i = \sum_{k=1}^{n} a_{ki} w_k. \tag{4.4}$$

Let $\mathcal{C}^*$ be the dual basis of $\mathcal{C}$. By the previous lemma 4.82, if we take $l = w_j^*$ then,

$$w_j^* = \sum_{i=1}^{n} w_j^*(v_i) v_i^*.$$

By Equation (4.4) we find

$$w_j^* = \sum_{i=1}^{n} w_j^* \left(\sum_{k=1}^{n} a_{ki} w_k\right) v_i^*.$$

By definition of the dual basis the only nonzero term is when $k = j$.

$$w_j^* = \sum_{i=1}^{n} a_{ji} * v_i^*.$$

But this means that

$$[\text{id}_V]_{\mathcal{B}^*}^{\mathcal{C}^*} = (a_{ji}) = A^T.$$

□

Lemma 4.84:

Let V, W be vector spaces over K. Let $T : V \to W$ be a linear map. Define a new map

$$T^* : W^* \to V^*,$$

defined by $T^*(l) = l \circ T \in V^*$. This is called the DUAL MAP to T.

Note that $l \circ T$ is indeed in V^* because $l \circ T$ is linear as T and l are lienar themselves.

Then T^* is a linear map

The operation becomes reasonable if we look at a diagram

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[dr, "l\circ T"] & W \arrow[d, "l"] \\ & K \end{tikzcd}$$

Proof. Let l be a functional on W and $\alpha \in K$. We have that $T^*(\alpha l)$ is the map $(\alpha l) \circ T$. So for every $v \in V$,

$$T^*(\alpha l)(v) = (\alpha l)(T(v)) = \alpha(l \circ T)(v) = \alpha T^*(l)(v).$$

Hence, $T^*(\alpha l) = \alpha T^*(l)$.

Let now l_1, l_2 be functionals on W. Then,

$$\begin{aligned} (T^*(l_1 + l_2))(v) &= (l_1 + l_2)(T(v)) \\ &= l_1(T(v)) + l_2(T(v)) \\ &= T^*(l_1)(v) + T^*(l_2)(v). \end{aligned}$$

Since this is true for all $v \in V$, we have

$$T^*(l_1 + l_2) = T^*(l_1) + T^*(l_2).$$

□

Proposition 4.85:

Let U, V, W be vector spaces over K. Let $T : V \to W$, $S : W \to U$ be linear maps. Consider $T^* : W^* \to V^*$ and $S^* : U^* \to W^*$ the dual maps. Then,

$$(S \circ T)^* = T^* \circ S^*.$$

Proof. Exercise. □

Exercise 4.86:

Let V, W be vector spaces.

1) Show that $(0 : V \to W)^* = 0 : W^* \to V^*$.

2) Show that $(\text{id}_V)^* = \text{id}_{V^*}$.

Lemma 4.87:

Let $S : V \to W$ be a linear map between two finite-dimensional vector spaces. Let $\mathcal{B}, \mathcal{C}$ be bases for V and W respectively. Consider $S^* : W^* \to V^*$ the dual map and the dual bases $\mathcal{B}^*, \mathcal{C}^*$. Then,

$$[S^*]_{\mathcal{B}^*}^{\mathcal{C}^*} = \left([S]_{\mathcal{C}}^{\mathcal{B}}\right)^T.$$

Proof. Write $\mathcal{B} = (v_1, \dots, v_n), \mathcal{C} = (w_1, \dots, w_m)$. Let $A = [S]_{\mathcal{C}}^{\mathcal{B}} = (a_{ij})$. Then,

$$S(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Now, for $w_j^* \in C^*$, we have

$$S^*(w_j^*) = w_j^* \circ S = \sum_{i=1}^{n} (w_j^* \circ S)(v_i) v_i^*.$$

But the term $(w_j^* \circ S)(v_i)$ is nonzero only when $S(v_i)$ contains w_j with a nonzero coefficient. From the equation above, this coefficient is a_{ji}. Hence,

$$S^*(w_j^*) = \sum_{i=1}^{n} a_{ji} v_i^*.$$

This means that the coordinates of $S^*(w_j^*)$ in the basis $\mathcal{B}^*$ are given by the j-th row of A. Hence,

$$[S^*]_{\mathcal{B}^*}^{\mathcal{C}^*} = A^T = \left([S]_{\mathcal{C}}^{\mathcal{B}}\right)^T.$$

Lec 28

□

Definition 4.88: Reflexivity

Let V be a Vector space. Define $V^{**} = (V^*)^*$ which is sometimes called the BIDUAL of V.

Theorem 4.89:

Let V be a finite-dimensional vector space over K. Then, there exists a canonical isomorphism $\tau : V \to V^{**}$, given by:

$$\forall v \in V, \tau(v) : V^* \to K, .$$

is the map such that $\tau(v)(l) = l(v)$. In other words,

$$V \ni v \overset{\tau}{\mapsto} (V^* \ni l \mapsto l(v) \in K).$$

Proof. We claim, τ is linear. Indeed, if $v \in V, \alpha \in K$ we have:

$$\begin{aligned}\tau(\alpha v) &= (V^* \ni l \mapsto l(\alpha v) \in K)\\ &= (V^* \ni l \mapsto \alpha l(v) \in K)\\ &= \alpha(V^* \ni l \mapsto l(v) \in K) = \alpha\tau(v).\end{aligned}$$

As an exercise, check that $\tau(v_1 + v_2) = \tau(v_1) + \tau(v_2)$ for all $v_1, v_2 \in V$.

We now claim, that τ is injective which is equivalent to showing that $\ker(\tau) = \{0\}$.

Indeed, assume that $\tau(v) = 0$ meaning that $l(v) = 0 \forall l \in V^*$. Choose a basis $\mathcal{B} = (v_1, \dots, v_n)$ of V such that $v_1 = v$. Write $v = c_1 v_1 + \dots + c_n v_n$, with $c_1, \dots, c_n \in K$. Apply $l(v) = 0$ for $l = v_k^*$. We get $0 = l(v) = c_k$ for all $1 \le k \le n$. Hence, $v = 0$.

Finally, since $\dim(V) = \dim(V^*)$, also $\dim(V^{**}) = \dim(V^*) = \dim(V)$. Hence, τ is an injective linear map between two vector spaces of the same dimension. Hence, τ is an isomorphism. □

The above theorem is NOT true for infinite-dimensional vector spaces.

Definition 4.90: Naturality

The map $\tau : V \to V^{**}$ can be defined for every vector space V. To emphasize dependence on V, we will write here τ_V instead of τ.

Let V, W be two vector spaces over K. Then $\forall$ linear map $T : V \to W$, we have the following commutative diagram:

$$\begin{tikzcd} V \arrow[r, "\tau_V"] \arrow[d, "T"] & V^{**} \arrow[d, "T^{**}"] \\ W \arrow[r, "\tau_W"] & W^{**} \end{tikzcd}$$

Since this works for any linear map T, we call τ a NATURAL ISOMORPHISM between the identity functor and the double dual functor.

Exercise 4.91:

Let V be a finite dimensional vector space over K. Let $\mathcal{B}$ be a basis for V and $\mathcal{B}^*$ be the dual basis for V^* and $\mathcal{B}^{**}$ be the dual basis for V^*. Show that $\tau_V(v_i) = v_i^{**}$ for all $v_i \in \mathcal{B}$.

4.6 Direct Sums

Let us talk about another way to construct new vector spaces from old ones.

Proposition 4.92:

Let V, W be vector spaces over K. Then, the Cartesian product $V \times W$ becomes a vector space over K if we endow it with the following operations:

$$\begin{aligned}(v_1, w_1) + (v_2, w_2) &:= (v_1 + v_2, w_1 + w_2)\\ \alpha(v, w) &:= (\alpha v, \alpha w)\\ 0_{V \times W} &:= (0_V, 0_W).\end{aligned}$$

Proof. Just go over the axioms of a vector space. □

$V \times W$ is denoted usually by $V \oplus W$ and is called the DIRECT SUM of V and W. Its elements are denoted by (v, w) or sometimes by $v \oplus w$.

Remark 4.93:

$V \oplus W$ comes with the following canonical linear maps:

1. $i_V : V \to V \oplus W$ defined by $i_V(v) = (v, 0)$.
2. $i_W : W \to V \oplus W$ defined by $i_W(w) = (0, w)$.
3. $p_V : V \oplus W \to V$ defined by $p_V(v, w) = v$.
4. $p_W : V \oplus W \to W$ defined by $p_W(v, w) = w$.

1 and 2 are called embeddings and 3 and 4 are called projections.

Using i_V and i_W, we can view V as a subspace of $V \oplus W$ and W as a subspace of $V \oplus W$.

Proposition 4.94:

Let V be a vector space and $U \subseteq V$ be a subspace. Let $U' \subseteq V$ be another subspace which is a complement of U. Then, the map $\varphi : U \oplus U' \to V$ defined by $\varphi(u, u') = u + u'$ is an isomorphism.

Proof. Linearity of φ follows immediately from the definition.

Injectivity: Suppose $\varphi(u, u') = 0$. Then, $u' = -u \in U \cap U'$. Since $U \cap U' = \{0\}$, we have $u = u' = 0$. Hence, $\mathrm{Ker}(\varphi) = \{0\}$ and φ is injective.

Surejctivity: By assumption, $U + U' = V$. Let $v \in V$. Write $v = u + u'$ with $u \in U, u' \in U'$. Then, $\varphi(u, u') = v$. Hence, φ is surjective. □

The idea of the direct sum can be generalized to arbitrary many spaces. Let $U_1, \dots, U_m$ be finitely many vector spaces over K. We can define $U_1 \oplus \cdots \oplus U_m$ as the set of all tuples similarly to what we've done before.

Let $\{U_i\}_{i \in I}$ be a family of vector spaces, parametrized by $i \in I$.

$$\prod_{i \in I} U_i = \left\{ t : I \to \bigcup_{i \in I} U_i | t(i) \in U_i \forall i \in I \right\}.$$

One can endow this with the structure of a vector space over K.

Define also

$$\bigoplus_{i \in I} U_i = \left\{ t \in \prod_{i \in I} U_i |\ t(i) \neq 0 \text{ for only finitely many } i \right\}.$$

If I is finite, then $\prod_{i \in I} U_i = \bigoplus_{i \in I} U_i$. But, if for example $I = \mathbb{Z}_{\geq 1}, U_i = K \forall i$, then $\prod_{i \in I} U_i$ is the set of all sequences of elements of K while $\bigoplus_{i \in I} U_i$ is the set of all sequences of elements of K which are zero for all but finitely many i.

Lec 29

4.7 Quotient Space

Let V be a vector space over K and let $U \subset V$ be a subspace. We'll define a new vector space V/U together with a linear map $\pi : V \to V/U$ such that π is surjective and $\mathrm{Ker}(\pi) = U$.

We begin by defining an equivalence relation on the set V.

Definition 4.95: Equivalence relation associated to U

Let $v_1, v_2 \in V$. We declare $v_1 \sim v_2$ if $v_1 - v_2 \in U$.

Claim 4.96:

$\sim$ is an equivalence relation on V.

Proof. Reflexivity: $v \sim v$ since $v - v = 0 \in U$.

Symmetry: If $v_1 \sim v_2$, then $v_1 - v_2 \in U$. Hence, $v_2 - v_1 = -(v_1 - v_2) \in U$. Hence, $v_2 \sim v_1$.

Transitivity: If $v_1 \sim v_2$ and $v_2 \sim v_3$, then $v_1 - v_2 \in U$ and $v_2 - v_3 \in U$. Hence, $v_1 - v_3 = (v_1 - v_2) + (v_2 - v_3) \in U$. Hence, $v_1 \sim v_3$. □

Denote by $[v]$ the equivalence class of $v \in V$. One can also think of $[v]$ as $[v] = v + U = \{v + u | u \in U\}$.

Definition 4.97:

Let V and U as before. Define $V/U := V/\sim$ as the set of all equivalence classes of $\sim$. Elements of V/U are denoted by $[v]$, where $v \in V$. We call V/U the QUOTIENT SPACE of V by U.

We'll turn the set V/U into a vector space over K by defining the following operations:

Addition: Let $x, y \in V/U$. Pick representatives $v, w \in V$ such that $x = [v]$ and $y = [w]$. Define

$$x + y = [v] + [w] := [v + w].$$

Multiplication: Let $x \in V/U$ and $\alpha \in K$. Pick a representative $v \in V$ such that $x = [v]$. Define

$$\alpha x = \alpha[v] := [\alpha v].$$

Zero: Define $0_{V/U} = [0_V]$.

Proposition 4.98:

The operations above are well-defined and turn V/U into a vector space over K.

Proof. Multiplication: Suppose v, v' give the same equivalence class, namely $[v] = [v'] = x$. Let $\alpha \in K$. We need to check that $[\alpha v] = [\alpha v']$. Indeed, $v - v' \in U$. Hence, $\alpha v - \alpha v' = \alpha(v - v') \in U$. Hence, $[\alpha v] = [\alpha v']$.

Addition: Suppose $x_1 = [v_1] = [v_1']$ and $x_2 = [v_2] = [v_2']$. Now,

$$(v_1 + v_2) - (v_1' + v_2') = (v_1 - v_1') + (v_2 - v_2') \in U.$$

Hence, $[v_1 + v_2] = [v_1' + v_2']$.

The axioms of the vector space can be checked one by one. □

We now define our map $\pi : V \to V/U$ by $\pi(v) = [v]$.

Proposition 4.99:

The map $\pi : V \to V/U$ is linear, surjective and $\mathrm{Ker}(\pi) = U$.

Proof. Linearity: Let $v_1, v_2 \in V$ and $\alpha \in K$. We have:

$$\pi(v_1 + v_2) = [v_1 + v_2] = [v_1] + [v_2] = \pi(v_1) + \pi(v_2)$$
$$\pi(\alpha v) = [\alpha v] = \alpha[v] = \alpha\pi(v).$$

Surjectivity: Let $x \in V/U$. Pick a representative $v \in V$ such that $x = [v]$. Then, $\pi(v) = [v] = x$.

Kernel: If $v \in \mathrm{Ker}(\pi)$, then $\pi(v) = [v] = 0_{V/U} = [0_V]$. Hence, $v \sim 0_V$ meaning that $v - 0_V = v \in U$. Hence, $\mathrm{Ker}(\pi) \subseteq U$.

Let now $u \in U$. Then, $\pi(u) = [u] = [0_V] = 0_{V/U}$. Hence, $u \in \mathrm{Ker}(\pi)$ implying that $U \subseteq \mathrm{Ker}(\pi)$. □

Geometrically, we can think of V/U as the space obtained by collapsing U to a single point. For example, if $V = \mathbb{R}^2$ and U is the x-axis, then V/U can be thought of as the set of all lines parallel to the x-axis.

Proposition 4.100:

Let V be a finite dimensional vector space and $U \subseteq V$ be a subspace. Then, $\dim(V/U) = \dim(V) - \dim(U)$.

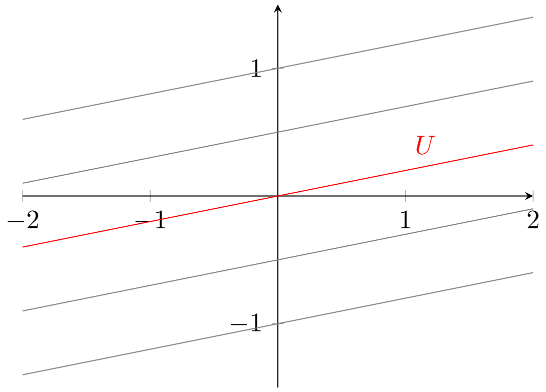


Figure 3: Quotient space in $\mathbb{R}^2$

Proof. Since $\pi : V \to V/U$ is surjective, then V/U is finite dimensional.[2] We apply now the rank theorem to π. Hence,

$$\dim(V) = \dim(\operatorname{Ker}(\pi)) + \dim(\operatorname{Im}(\pi)) = \dim(U) + \dim(V/U).$$

□

Theorem 4.101: Isomorphism theorem

Let V, W be vector spaces over K and $T : V \to W$ linear. Define a new map $\overline{T} : V/\operatorname{Ker}(T) \to \Im(T)$ as follows:

$$V/Ker(T) \ni [v] \overset{\overline{T}}{\longmapsto} T(v) \in \operatorname{Im}(T).$$

Then $\overline{T}$ is a well-defined isomorphism and the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\pi"] & W \\ V/\operatorname{Ker}(T) \arrow[ur, "\overline{T}"] & \end{tikzcd}$$

Proof. Well-Defined: Suppose $x = [v] = [v']$. Then, $v - v' \in \operatorname{Ker}(T)$. Hence, $T(v - v') = 0$. Hence, $T(v) = T(v')$.

Linearity: Let $x_1 = [v_1], x_2 = [v_2]$ and $\alpha \in K$. We have:

$$\begin{aligned} \overline{T}(\alpha[v]) &= \overline{T}([\alpha v]) = T(\alpha v) = \alpha T(v) = \alpha\overline{T}([v]) \\ \overline{T}([v_1] + [v_2]) &= \overline{T}([v_1 + v_2]) = T(v_1 + v_2) = T(v_1) + T(v_2) \\ &= \overline{T}([v_1]) + \overline{T}([v_2]). \end{aligned}$$

Injectivity: Suppose $\overline{T}([v]) = 0$. Then, $T(v) = 0$. Hence, $v \in \operatorname{Ker}(T)$. Hence, $[v] = 0_{V/\operatorname{Ker}(T)}$.

Surjectivity: Let $w \in \operatorname{Im}(T)$. Then, there exists $v \in V$ such that $T(v) = w$. Hence, $\overline{T}([v]) = w$.

Since $\overline{T}$ is surjective and bijective, $\overline{T}$ is an isomorphism.

The commutativity of the diagram follows immediately from the definition of $\overline{T}$. □

[2] This holds in general.

Theorem 4.102:

Let V be a vector space over K and $U \subset V$ be a subspace. The quotient V/U has the following universal property: For every vector space W and every linear map $T : V \to W$ with $T(U) = 0$, there exists a unique linear map $T' : V/U \to W$ such that the following diagram commutes ($T = T' \circ \pi$):

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\pi"] & W \\ V/U \arrow[ur, "T'"] & \end{tikzcd}.$$

Moreover, $\operatorname{Ker}(T') = \operatorname{Ker}(T)/U$.

Jargon: Every linear map $T : V \to W$ that send U to zero FACTORS through V/U. I.e $T = T' \circ \pi$. Lec 30

Proof. Define $T' : V/U \to W$ by $T'([v]) = T(v)$. We claim that the diagram commutes. Indeed, for every $v \in V$, we have

$$T'(\pi(v)) = T'([v]) = T(v).$$

Uniqueness: Suppose that $T = T' \circ \pi = T'' \circ \pi$. Let $x \in V/U$ and choose a representative $v \in V$ such that $x = [v]$. Then,

$$T(v) = T'(\pi(v)) = T''(\pi(v)) \Rightarrow T'([v]) = T''([v]) \Rightarrow T'(x) = T''(x).$$

Since x was arbitrary, we conclude that $T' = T''$.

□

Exercise 4.103:

1. Show that T' is well-defined.
2. Show that T' is linear.

Let V be a vector space and $U \subseteq V$ be a subspace. Let $W \subseteq V$ be a complement of U[3].

Proposition 4.104:

There exists a canonical isomorphism $Q : W \to V/U$, defined by

$$Q(w) := [w] \in V/U.$$

Proof. Consider the inclusion $i : W \to V$ ($i(w) = w$) and the projection $\pi : V \to V/U$. Recall that

$$\pi(v) = [v] \Rightarrow Q = \pi \circ i.$$

So Q is linear. We now claim that the kernel of Q is zero. Indeed, if $w \in \operatorname{Ker}(Q)$, then

$$[w] = 0_{V/U} = [0_V] \Rightarrow w \sim 0_V \Rightarrow w \in U.$$

But $w \in W$ and $W \cap U = \{0\}$. Hence, $w = 0$. So $\operatorname{Ker}(Q) = \{0\}$.

We claim now that Q us surjective. Let $x \in V/U$. Choose $v \in V$ with $x = [v]$. Since $V = W + U$, there exists $w \in W$ and $u \in U$ such that $v = w + u$. Hence, by definition of $\sim$, we have $[v] = [w + u] = [w]$. Hence, $Q(w) = [w] = [v] = x$.

Alternatively we can see that the dimensions of W and V/U are the same. Indeed, $\dim(W) = \dim(V) - \dim(U)$ and $\dim(V/U) = \dim(V) - \dim(U)$. □

Exercise 4.105:

Describe the inverse of Q.

We now want to relate the quotient space to the dual space.

[3] $W \subseteq V$ is a subspace such that $W \cap U = \{0\}$ and $W + U = V$.

Proposition 4.106:

Let V be a vector space and $U \subseteq V$ be a subspace. Define

$$U^\perp := \{l \in V^* \mid l|_U \equiv 0\}.$$

So $U^\perp$ is the set of all linear functionals on V that vanish on U.

1. $U^\perp \subset V^*$ is a subspace.

2. There exists a canonical isomorphism $(V/U)^* \cong U^\perp$, defined by $s \mapsto s \circ \pi$. If we assume that V is finite dimensional, then $\exists$ also a canonical isomorphism $V^*/U^\perp \cong U^*$.

Proof. 1. Clearly, $0 \in U^\perp$. If $l_1, l_2 \in U^\perp$ and $\alpha, \beta \in K$, then

$$\alpha l_1 + \beta l_2 \in U^\perp.$$

Because, for every $u \in U$, we have

$$(\alpha l_1 + \beta l_2)(u) = \alpha l_1(u) + \beta l_2(u) = 0.$$

2. Define a map $L : (V/U)^* \to U^\perp$ as follows:

Let $s \in (V/U)^*$, i.e. $s : V/U \to K$ is linear. Define $L(s) \in V^*$ to be $L(s) := s \circ \pi$. So L is: $s \mapsto s \circ \pi$.

$$\begin{tikzcd} V \arrow[r, "s\circ\pi"] \arrow[d, "\pi"] & K \\ V/U \arrow[ur, "s"] & \end{tikzcd}.$$

We claim, that L is an isomorphism. To see this, we will define another map $P : U^\perp \to (V/U)^*$ and show that P is the inverse of L.

Let $t \in U^\perp$, i.e. $t : V \to K$ such that $t|_U \equiv 0$. By the universal property of the quotient spaces, there exists a unique $t' : V/U \to K$ such that $t = t' \circ \pi$.

$$\begin{tikzcd} V \arrow[r, "t"] \arrow[d, "\pi"] & K \\ V/U \arrow[ur, "t'"] & \end{tikzcd}.$$

Define $P(t) = t'$. We claim now that $L \circ P = \mathrm{id}_{U^\perp}$ and $P \circ L = \mathrm{id}_{(V/U)^*}$.

Indeed, if $s : V/U \to K$ then,

$$P \circ L(s) = P(s \circ \pi) = s.$$

Also, $\forall t \in U^\perp$, we have

$$L \circ P(t) = L(t') = t' \circ \pi = t.$$

Since L is linear, P is also a linear map and both of them are isomorphisms.

Assume now that V is finite dimensional and consider the map $R : V^* \to U^*$ defined by $\forall \varphi \in V^*$, $R(\varphi) = \varphi|_U$. So R is the restriction map. We claim that R is surjective. Indeed, let $\psi \in U^*$. We need to show that there exists $\varphi \in V^*$ such that $\varphi|_U = \psi$. Pick a basis $\{u_1, \ldots, u_k\}$ of U and extend it to a basis $\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\}$ of V.

Define $\varphi : V \to K$ by $\varphi(u_i) = \psi(u_i)$ for $i = 1, \ldots, k$ and $\varphi(v_j) = 0$ for $j = k+1, \ldots, n$. Since $\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\}$ is a basis of V, the map φ is well-defined and linear. Moreover, $\varphi|_U = \psi$. Hence, R is surjective.

Also, $\ker(R) = U^\perp$ by definition. By theorem 4.101, we get that

$$V^*/\ker(R) \cong \mathrm{Im}(R) = U^*.$$

So $V^*/U^\perp \cong U^*$. □

Exercise 4.107:

1. Check that $s \circ \pi$ is indeed in $U^\perp$.

2. Show that L is linear.

3. Check that if $i : U \to V$ is the inclusion, then $R = i^*$.

Exercise 4.108:

Let $T : V \to W$ be a linear map. Then $(\mathrm{Im}\, T)^\perp = \mathrm{Ker}(T^*)$.

Lec 31

5 Determinants

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(K)$.

Proposition 5.1:

A is invertible if and only if $ad - bc \neq 0$. Moreover, if A is invertible, then

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

The scalar $ad - bc$ is called the DETERMINANT of A and is denoted by $\det(A)$ or $|A|$.

Proof. Homework.

Outline: A is not invertible if and only if $\operatorname{rank}(A) < 2$ if and only if the columns of A are linearly dependent if and only if there exists $\lambda \in K$ such that $a = \lambda c$ and $b = \lambda d$ or that $c = \lambda a$ and $d = \lambda b$. In both cases, $ad - bc = 0$. □

Definition 5.2: n-linear function

Let $D : M_{n\times n}(K) \to K$ be a function. We say that D is n-LINEAR if $\forall 1 \le i \le n$, D is a linear function of the i-th row when the other rows are held fixed.

We can think of $M_{n\times n}(K)$ as the coordinate space $K^n_{\text{row}} \times \cdots \times K^n_{\text{row}}$ (n times) and if

$$A = \begin{pmatrix} \cdots & \alpha_1 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \alpha_n & \cdots \end{pmatrix}.$$

write $D(A) = D(\alpha_1, \ldots, \alpha_n)$ where α_i is the i-th row of A. Then, D is n-linear if $\forall 1 \le i \le n$, D is linear in α_i when the other rows are held fixed so

1. $D(\alpha_1, \ldots, \alpha_i + \beta_i, \ldots, \alpha_n) = D(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + D(\alpha_1, \ldots, \beta_i, \ldots, \alpha_n)$.
2. $D(\alpha_1, \ldots, \lambda\alpha_i, \ldots, \alpha_n) = \lambda D(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n)$.

Notice that $D(A + B) \neq D(A) + D(B)$ in general and $D(\lambda A) \neq \lambda D(A)$ in general. So D is not linear as a function from $M_{n\times n}(K)$ to K.

In the following, for a matrix A, write $A(i,j)$ or A_{ij} or a_{ij} for the entry in the i-th row and j-th column of A.

Example 5.3:

Fix integers $k_1, \ldots, k_n$ such that $1 \le k_i \le n$ and $a \in K$. Define $D : M_{n\times n}(K) \to K$ by $D(A) = a \cdot A(1, k_1) \cdot \cdots \cdot A(n, k_n)$.

We claim that D is n-linear.

Proof. Consider

$$\begin{pmatrix} \cdots & \alpha_1 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \alpha_n & \cdots \end{pmatrix} = \begin{pmatrix} A(1,1) & \ldots & A(1,n) \\ \vdots & \vdots & \vdots \\ A(i,1) & \ldots & A(i,n) \\ \vdots & \vdots & \vdots \\ A(n,1) & \ldots & A(n,n) \end{pmatrix}.$$

Now, we can write

$$D(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) = b \cdot A(i, k_i).$$

where $b \in K$ does not depend on row i.

If we now change row i to $\lambda\alpha_i + \beta_i$, then we see that D is linear in row i when the other rows are held fixed. □

Let us try to find all 2-linear functions $D : M_{2\times 2}(K) \to K$.

Write $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$. If D is 2-linear, then $D\begin{pmatrix} A(1,1) & A(1,2) \\ A(2,1) & A(2,2) \end{pmatrix}$ is determined by

$$\begin{aligned} D(A) &= D\left((A(1,1), A(1,2)), (A(2,1), A(2,2))\right) \\ &= D((A(1,1)e_1 + A(1,2)e_2), A(2,1)e_1 + A(2,2)e_2) \\ &= A(1,1)D(e_1, A(2,1)e_1 + A(2,2)e_2) + \\ &+ A(1,2)D(e_2, A(2,1)e_1 + A(2,2)e_2) \\ &= A(1,1)A(2,1)D(e_1,e_1) + A(1,1)A(2,2)D(e_1,e_2) \\ &+ A(1,2)A(2,1)D(e_2,e_1) + A(1,2)A(2,2)D(e_2,e_2). \end{aligned}$$

By the four scalars, D is completely determined. Hence, the space of 2-linear functions from $M_{2\times 2}(K)$ to K is 4-dimensional.

Lemma 5.4:

Any linear combination of n-linear functions is again an n-linear function.

Proof. It is enough to prove for the case of two n-linear functions.

Let $D, E : M_{n\times n}(K) \to K$ be n-linear functions and $a, b \in K$. Consider the function $aD + bE : M_{n\times n}(K) \to K$ defined by $(aD + bE)(A) = aD(A) + bE(A)$.

Let $1 \le i \le n$. Then

$$\begin{aligned} &(aD + bE)(\alpha_1, \ldots, \alpha_i + \beta_i, \ldots, \alpha_n) \\ &= aD(\alpha_1, \ldots, \alpha_i + \beta_i, \ldots, \alpha_n) + bE(\alpha_1, \ldots, \alpha_i + \beta_i, \ldots, \alpha_n) \\ &= a(D(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + D(\alpha_1, \ldots, \beta_i, \ldots, \alpha_n)) + \\ &+ b(E(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + E(\alpha_1, \ldots, \beta_i, \ldots, \alpha_n)) \\ &= (aD + bE)(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + (aD + bE)(\alpha_1, \ldots, \beta_i, \ldots, \alpha_n). \end{aligned}$$

Similarly, we can check that

$$(aD + bE)(\alpha_1, \ldots, \lambda\alpha_i, \ldots, \alpha_n) = \lambda(aD + bE)(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n).$$

□

Example 5.5:

Consider $D : M_{2\times 2}(K) \to K$ defined by

$$D(A) = A_{11} \cdot A_{22} - A_{12} \cdot A_{21}.$$

This is 2-linear which can be seen as $A_{11}A_{22}$ and $-A_{12}A_{21}$ are 2-linear (Example 5.3) and D is a linear combination of them.

Remark

$D = det(A)$ as a function $M_{2\times 2}(K) \to K$ has two additional properties:

1. $D(I) = 1$ where I is the identity matrix.
2. If A' is obtained from A by interchanging its rows, then $D(A') = -D(A)$.

Definition 5.6: Alternating function

Let $D : M_{n\times n}(K) \to K$ be an n-linear function. We say that D is ALTERNATING if for every matrix A in which some two rows are equal, we have $D(A) = 0$.

Proposition 5.7:

Let D be an n-linear function. Assume that D has the property that whenever $A \in M_{n\times n}(K)$ has two equal adjacent rows, then $D(A) = 0$. Then, D is alternating and $\forall$ matrix B, if B' is obtained from B by interchanging two rows, then $D(B') = -D(B)$.

Proof. We begin with proving the second statement. Assume first, that B' is obtained from B by interchanging two adjacent rows.

Consider

$$D(\beta_1, \dots, \beta_k + \beta_{k+1}, \beta_k + \beta_{k+1}, \dots, \beta_n).$$

By assumption this must be zero. Opening the brackets, we find that

$$D(\beta_1, \dots, \beta_k, \beta_{k+1}, \dots, \beta_n) + D(\beta_1, \dots, \beta_{k+1}, \beta_k, \dots, \beta_n) = 0.$$

Lec 32 So interchanging two adjacent rows indeed changes the sign of D.

Suppose now that B' is obtained from B by interchanging rows k and l where $k < l$ (but not nessecairly adjacent). We can obtain B' from B by doing a sequence of interchanges of adjacent rows.

Indeed, we can begin with interchanging rows k and $k+1$, and continue in this way until we get

$$\beta_1, \dots, \beta_{k-1}, \beta_{k+1}, \dots, \beta_{l-1}, \beta_l, \beta_k, \dots, \beta_n.$$

This requires $r := l - k$ interchanges of adjacent rows. Now, we can continue with interchanging rows $l-1$ and $l-2$, and continue in this way until we get

$$\beta_1, \dots, \beta_{k-1}, \beta_l, \dots, \beta_{l-1}, \beta_k, \dots, \beta_n.$$

This requires $(l-1) - k = r - 1$ interchanges of adjacent rows. Hence, in total, we need $r + (r-1) = 2r - 1$ interchanges of adjacent rows to get B' from B. Since $2r-1$ is odd, we conclude that $D(B') = -D(B)$.

1. Let A be a matrix in which row i and row j are equal. We can obtain a matrix A' from A by interchanging rows i and j. Since row i and row j are equal, we have $A = A'$. On the other hand, by the second statement, we have $D(A') = -D(A)$. Hence, $D(A) = -D(A)$ which implies that $D(A) = 0$. So D is alternating. □

Definition 5.8: Determinant

A function $D : M_{n\times n}(K) \to K$ is called a DETERMINANT FUNCTION if it is n-linear, alternating and $D(I) = 1$ where I is the identity matrix.

As a warm-up let us find all determinant functions $D : M_{2\times 2}(K) \to K$. Write $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$ and $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. We saw that every 2-linear function satisfies

$$\begin{aligned} D(A) = {} & A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) \\ & + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2). \end{aligned}$$

Since D is alternating, we have $D(\epsilon_1, \epsilon_1) = D(\epsilon_2, \epsilon_2) = 0$. Also $D(\epsilon_1, \epsilon_2) = -D(\epsilon_2, \epsilon_1)$.

Thus, we can write

$$D(A) = D(\epsilon_1, \epsilon_2)(A_{11}A_{22} - A_{12}A_{21}).$$

Since $D(I) = 1$, we have $D(\epsilon_1, \epsilon_2) = 1$. Hence, the only determinant function from $M_{2\times 2}(K)$ to K is the function defined by

$$D(A) = A_{11}A_{22} - A_{12}A_{21} = \det(A).$$

Definition 5.9:

Let $A \in M_{n\times n}(K)$, $n \geq 2$ and let $1 \leq i, j \leq n$. Denote by $A(i|j) \in M_{(n-1)\times(n-1)}(K)$ the matrix obtained from A by deleting row i and column j.

If D is a $(n-1)$-linear function, denote by

$$D_{ij}(A) := D(A(i|j)).$$

Theorem 5.10:

Let $n \geq 2$ and D an alternating $(n-1)$-linear function. Let $1 \leq j \leq n$. Define $E_j : M_{n\times n}(K) \to K$ by

$$E_j(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot D_{ij}(A).$$

Then, E_j is an n-linear function and it is alternating. Moreover, if D is a determinant function, then so is E_j.

Proof. Let $A \in M_{n\times n}(K)$ and $1 \leq j \leq n$. $D_{ij}(A)$ is independent of row i of A since it gets deleted. Since D is $(n-1)$-linear, then also $A \mapsto D_{ij}(A)$ is "linear" when viewed as a function of each of its rows of A except of row i.

So it follows that the function $A \mapsto A_{ij} \cdot D_{ij}(A)$ is n-linear. Since E_j is a linear combination of such functions, it is also n-linear.

Assume now that D is alternating. Indeed, by proposition 5.7, it is enough to show that $E_j(A) = 0$ whenever A has two equal adjacent rows.

So let A be a matrix in which row k and row $k+1$ are equal. Let $1 \leq i \leq n$, such that $i \neq k, k+1$. The matrix $A(i|j)$ has two equal adjacent rows, hence $D_{ij}(A) = D(A(i|j)) = 0$. Thus $E_j(A) = (-1)^{k+j}A_{kj} \cdot D_{kj}(A) + (-1)^{k+1+j}A_{k+1,j} \cdot D_{k+1,j}(A)$. But since row k and row $k+1$ are equal, we have $A_{kj} = A_{k+1,j}$ and $D_{kj}(A) = D_{k+1,j}(A)$. Hence, $E_j(A) = 0$.

Finally, if D is a determinant function, then $E_j(I) = D(I) = 1$. So E_j is also a determinant function. □

Corollary 5.11:

For every $n \geq 1$, $\exists$ at least one determinant function from $M_{n\times n}(K)$ to K.

Proof. By induction on n. For $n = 1$, the function defined by $D((a)) = a$ is a determinant function.

Assume now that $n \geq 2$ and there exists a determinant function $D : (n-1) \times (n-1)(K) \to K$. By Theorem 5.10, the function $E_1 : M_{n\times n}(K) \to K$ defined by

$$E_1(A) = \sum_{i=1}^{n} (-1)^{i+1} A_{i1} \cdot D_{i1}(A)$$

is a determinant function. □

Example 5.12:

For a 2×2 matrix B we have

$$\det : B \mapsto \det(B) = B_{11}B_{22} - B_{12}B_{21}.$$

For a 3×3 matrix A, we have

$$\det_1 : M_{3\times 3}(K) \to K, A \mapsto \det_1(A) =$$
$$A_{11} \cdot \det \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix} - A_{21} \cdot \det \begin{pmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{pmatrix} +$$
$$A_{31} \cdot \det \begin{pmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{pmatrix}$$
$$\det_2 : M_{3\times 3}(K) \to K, A \mapsto \det_2(A) =$$
$$A_{12} \cdot \det \begin{pmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{pmatrix} - A_{22} \cdot \det \begin{pmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{pmatrix} +$$
$$A_{32} \cdot \det \begin{pmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{pmatrix}$$
$$\det_3 : M_{3\times 3}(K) \to K, A \mapsto \det_3(A) =$$
$$A_{13} \cdot \det \begin{pmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix} - A_{23} \cdot \det \begin{pmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{pmatrix} +$$
$$A_{33} \cdot \det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Lec 33

5.1 Uniqueness of determinant functions

Let $D : M_{n\times n}(K) \to K$ be an n-linear alternating function. Let

$$A = \begin{pmatrix} - & \alpha_1 & - \\ & \vdots & \\ - & \alpha_n & - \end{pmatrix}, I = \begin{pmatrix} - & \varepsilon_1 & - \\ & \vdots & \\ - & \varepsilon_n & - \end{pmatrix} \in M_{n\times n}(K).$$

We have, that $\alpha_i = \sum_{j=1}^n A_{ij}\varepsilon_j$. Since D is n-linear, we have

$$D(A) = D(\alpha_1, \dots, \alpha_n) = D\left(\sum_{j=1}^n A_{1j}\varepsilon_j, \alpha_2, \dots, \alpha_n\right).$$

Since D is linear in the first argument, we have

$$D(A) = \sum_{j=1}^n A_{1j} D(\varepsilon_j, \alpha_2, \dots, \alpha_n).$$

Repeating the same argument for the second, third, ..., n-th argument, we have

$$\begin{aligned} D(A) &= \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n A_{1j_1}A_{2j_2}\dots A_{nj_n} D(\varepsilon_{j_1}, \dots, \varepsilon_{j_n}) \\ &= \sum_{1\le k_1,k_2,\dots,k_n\le n} A_{1k_1}A_{2k_2}\dots A_{nk_n} D(\varepsilon_{k_1}, \dots, \varepsilon_{k_n}). \end{aligned}$$

Since D is alternating, if $k_i = k_j$ for some $i \neq j$, then $D(\varepsilon_{k_1}, \dots, \varepsilon_{k_n}) = 0$. Hence, we can restrict the sum to the case where $k_1, k_2, \dots, k_n$ are distinct. Such a sequence is called a PERMUTATION of degree n.

We can think of permutations as functions

$$\sigma : \{1, \dots, n\} \to \{1, \dots, n\}, \text{ bijective }.$$

For example, $(k_1, \dots, k_n) = (\sigma(1), \dots, \sigma(n))$ for some permutation σ.

So

$$D(A) = \sum_\sigma A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} D(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(n)}).$$

Where the sum runs over all permutations σ of degree n.

The number of permutations of degree n is $n!$. For example, for $n = 3$, there are 6 permutations.

Every permutation σ can be written as a composition

$$\sigma = \theta_1 \circ \dots \circ \theta_k,$$

of transpositions θ_i where a TRANSPOSITION is a permutation such that $\exists i, j$ such that $\theta(i) = j$, $\theta(j) = i$ and $\theta(k) = k$ for all $k \neq i, j$.

Why? Start with $(1, \dots, n)$. If $\sigma(1) \neq 1$, then interchange 1 and $\sigma(1)$ to get a new permutation θ_1 such that $\theta_1(1) = 1$. If $\theta_1 \circ \sigma(2) \neq 2$, then interchange 2 and $\theta_1 \circ \sigma(2)$ to get a new permutation θ_2 such that $\theta_2 \circ \theta_1 \circ \sigma(2) = 2$. Continuing in this way, we get a permutation $\theta_k \circ \dots \circ \theta_1 \circ \sigma$ such that $\theta_k \circ \dots \circ \theta_1 \circ \sigma(i) = i$ for all $1 \le i \le n$.

In general there is not a unique way to write a permutation as a composition of transpositions.

$$\begin{aligned} &(1, 2, 3) &&\to (3, 2, 1) \\ &(1, 2, 3) \to (2, 1, 3) \to (2, 3, 1) &&\to (3, 2, 1). \end{aligned}$$

Lemma 5.13:

If $\sigma = \theta_1 \circ \dots \circ \theta_k = \theta'_1 \circ \dots \circ \theta'_{k'}$ are two ways to write σ as a composition of transpositions, then $k \equiv k' \mod 2$.

Define the SIGN of σ by $\operatorname{sgn}(\sigma) = (-1)^k$ where k is the number of transpositions in any decomposition of σ.

Proof. We know that determinant functions exist $\forall n \ge 1$. Fix one such function E. Consider

$$E(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(n)}).$$

If $(\sigma_1, \dots, \sigma_n)$ can be obtained from $(1, \dots, n)$ by m transpositions, then the matrix

$$\begin{pmatrix} - & \varepsilon_{\sigma(1)} & - \\ & \vdots & \\ - & \varepsilon_{\sigma(n)} & - \end{pmatrix}$$

can be obtained from the identity matrix by m interchanges of adjacent rows. Since E is alternating, we have

$$E(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(n)}) = (-1)^m E(I) = (-1)^m.$$

Equivalently, we can find

$$E(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(n)}) = (-1)^{k'} E(I) = (-1)^{k'}.$$

Hence, $(-1)^k = (-1)^{k'}$ and $k \equiv k' \mod 2$. □

If we return to $D(A)$, we have

$$\begin{aligned} D(A) &= \sum_\sigma A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} D(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(n)}) \\ &= D(I) \sum_\sigma A_{1\sigma(1)}A_{2\sigma(2)}\dots A_{n\sigma(n)} \operatorname{sgn}(\sigma). \end{aligned}$$

Theorem 5.14: Uniqueness of the Determinant

$\forall$ n-linear alternating function D we have

$$D(A) = D(I) \sum_{\sigma} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \dots A_{n\sigma(n)}.$$

In particular, if D is a determinant function, then

$$D(A) = \sum_{\sigma} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \dots A_{n\sigma(n)}.$$

So D is unique.

Corollary 5.15:

For every n-linear alternating function D, we have that $D(A) = \det(A) \cdot D(I)$.

5.2 Permutations

Denote by S_n the set of all permutations on n elements. S_n is a group under composition of functions. We can also invert elements of S_n by inverting the corresponding functions. For example, if $\sigma = (3, 1, 2)$, then $\sigma^{-1} = (2, 3, 1)$. The identity element of S_n is the permutation $\mathrm{id} = (1, 2, \dots, n)$.

Lemma 5.16:

Given two permutations σ, τ, we have

$$\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\tau).$$

Proof. Let $\sigma = \theta_1 \circ \dots \circ \theta_k$ and $\tau = \theta'_1 \circ \dots \circ \theta'_{k'}$ be two decompositions of σ and τ into transpositions. Then, $\sigma \circ \tau = \theta_1 \circ \dots \circ \theta_k \circ \theta'_1 \circ \dots \circ \theta'_{k'}$ is a decomposition of $\sigma \circ \tau$ into transpositions. Hence, $\operatorname{sgn}(\sigma \circ \tau) = (-1)^{k+k'} = (-1)^k \cdot (-1)^{k'} = \operatorname{sgn}(\sigma) \cdot \operatorname{sgn}(\tau)$. □

Theorem 5.17:

$\forall A, B \in M_{n\times n}(K)$, we have

$$\det(AB) = \det(A) \cdot \det(B).$$

Proof. Fix B and consider the function $D : M_{n\times n}(K) \to K$ defined by

$$D(A) \coloneqq \det(AB).$$

We claim that D is n-linear and alternating.

Indeed, write $A = \begin{pmatrix} - & \alpha_1 & - \\ & \vdots & \\ - & \alpha_n & - \end{pmatrix}$. Then,

$$A \cdot B = \begin{pmatrix} - & \alpha_1 \cdot B & - \\ & \vdots & \\ - & \alpha_n \cdot B & - \end{pmatrix}.$$

So, $D(A) = \det(\alpha_1 \cdot B, \dots, \alpha_n \cdot B)$. Fix $1 \le i \le n$. If α_i and α'_i are two row vectors, then

$$(\alpha_i + \alpha'_i) \cdot B = \alpha_i \cdot B + \alpha'_i \cdot B.$$

Therefore,

$$\begin{aligned} &D(\alpha_1, \dots, \alpha_i + \alpha'_i, \dots, \alpha_n) \\ &= D(\alpha_1, \dots, \alpha_i, \dots, \alpha_n) + D(\alpha_1, \dots, \alpha'_i, \dots, \alpha_n). \end{aligned}$$

since $\det$ is linear in each of its rows. Similarly, if $\lambda \in K$, then

$$D(\alpha_1, \dots, \lambda\alpha_i, \dots, \alpha_n) = \lambda D(\alpha_1, \dots, \alpha_i, \dots, \alpha_n).$$

If $\alpha_i = \alpha_j$ then $\alpha_i \cdot B = \alpha_j \cdot B$. Hence, $D(A) = 0$ since $\det$ is alternating. So D is alternating, proving the claim.

By corollary 5.15, we have $D(A) = D(I) \cdot \det(A)$. Since $D(I) = \det(IB) = \det(B)$, we have $\det(AB) = \det(A) \cdot \det(B)$. □

Corollary 5.18:

If A is invertible, then $\det(A) \neq 0$ and

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

Proof. If A is invertible, then

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

□ Lec 34

Recall that $\forall M \in M_{n\times n}(K)$, we have the transposed matrix

$$M^T \text{ with } M^T_{ij} = M_{ji}.$$

Theorem 5.19:

For every $A \in M_{n\times n}(K)$, we have $\det(A^T) = \det(A)$.

Proof. If σ is a permutation of $1, \dots, n$, then $(A^T)_{1\sigma(1)} (A^T) = A_{\sigma(1)1}$, so

$$\begin{aligned} \det(A^T) &= \sum_{\sigma} \operatorname{sgn}(\sigma) (A^T)_{1\sigma(1)} (A^T)_{2\sigma(2)} \dots (A^T)_{n\sigma(n)} \\ &= \sum_{\sigma} \operatorname{sgn}(\sigma) A_{\sigma(1)1} A_{\sigma(2)2} \dots A_{\sigma(n)n}. \end{aligned}$$

If $\sigma i = j$ then $A_{\sigma(i)i} = A_{j\sigma^{-1}(j)}$.

Claim: $A_{\sigma(1)1} A_{\sigma(2)2} \dots A_{\sigma(n)n} = A_{1\sigma^{-1}(1)} A_{2\sigma^{-1}(2)} \dots A_{n\sigma^{-1}(n)}$.

This is tree because every factor on the left hand side appears exactly once on the right hand side.

Furthermore, $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ since σ and σ^{-1} can be written as the same number of transpositions. Hence,

$$\det(A^T) = \sum_{\sigma} \operatorname{sgn}(\sigma^{-1}) A_{1\sigma^{-1}(1)} A_{2\sigma^{-1}(2)} \dots A_{n\sigma^{-1}(n)} = \det(A).$$

Since σ runs over all permutations, so does σ^{-1}, proving the theorem. □

Corollary 5.20:

If $D : M_{n\times n}(K) \to K$ is an alternating n-linear function of the rows, then D has the same properties when viewed as a function of the columns.

In fact, $D(A) = D(A^T)$.

Proof. $\forall A$, $D(A) = \det(A) \cdot D(I) = \det(A^T) \cdot D(I) = D(A^T)$. □

Theorem 5.21:

If B is obtained from A is obtained by elementary row operations, then

$$\begin{aligned} R_i + cR_j \to R_i &\Rightarrow \det(B) = \det(A) \\ R_i \leftrightarrow R_j &\Rightarrow \det(B) = -\det(A) \\ cR_i \to R_i &\Rightarrow \det(B) = c \cdot \det(A). \end{aligned}$$

Proof. 2 follows from the determinant being alternating. 3 follows from n-linearity. For 1, we have

$$\begin{aligned} \det(B) &= D(R_1, \dots, R_i + cR_j, \dots, R_n) \\ &= D(R_1, \dots, R_i, \dots, R_n) + c \cdot D(R_1, \dots, R_j, \dots, R_n) \\ &= \det(A) + c \cdot 0 = \det(A). \end{aligned}$$

we want to compute Determinants of block matrices

$$M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

Proposition 5.22:

If $A \in M_{k\times k}(K)$, $B \in M_{k\times(n-k)}(K)$ and $C \in M_{(n-k)\times(n-k)}(K)$, then

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det(A)\cdot\det(C).$$

Proof. Define the function D as

$$D(A,B,C) = \det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

If we fix A, B, then the function $C \mapsto D(A,B,C)$ is an alternating $(n-k)$-linear function. Hence, $D(A,B,C) = D(A,B,I)\cdot\det(C)$.

But using ERO's we can get rid of all the entries of B without changing the determinant. Hence,

$$D(A,B,I) = D(A,0,I) = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} = \det(A)\cdot\det(I) = \det(A).$$

□

Example 5.23:

Compute

$$\det \begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & 1 & -1 \\ 1 & 2 & 3 & 0 \end{pmatrix} \in M_{4\times 4}(\mathbb{R}).$$

Solution. We perform gaussian elimination to get an upper triangular matrix. We have

$$\begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & 1 & -1 \\ 1 & 2 & 3 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{pmatrix}.$$

This is exactly the form of the matrix in proposition 5.22 so

$$\det(A) = 128.$$

Let D be the determinant function on $M_{(n-1)\times(n-1)}(K)$. Recall that $\forall 1 \le j \le n$, the function $E_j : M_{n\times n}(K) \to K$ defined by

$$E_j(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij}\cdot D_{ij}(A).$$

We know now, that $\forall 1 \le j \le n$, E_j is a determinant function. The scalars $c_{ij} := (-1)^{i+j} D_{ij}(A)$ are called the i,j COFACTORS of A. We have that

$$\det(A) = \sum_{i=1}^{n} A_{ij}\cdot c_{ij}.$$

Claim 5.24:

If we replace A_{ij} by A_{ik} where $k \ne j$ is fixed, then $\sum_{i=1}^{n} A_{ik}\cdot c_{ij} = 0$.

□

Proof. Let B be the $n\times n$ matrix obtained from A by replacing column j of A by column k of A. Then, $\det(B) = 0$ since B has two equal columns. But $\det(B) = \sum_{i=1}^{n} A_{ik}\cdot c_{ij}$, proving the claim. □

In summary, $\forall A \in M_{n\times n}(K)$, we have $\forall j,k$

$$\sum_{i=1}^{n} A_{ik}\cdot c_{ij} = \delta_{jk}\cdot\det(A).$$

The matrix $(c_{ij})^T$ is called the CLASSICAL ADJOINT of A and is denoted by $\operatorname{adj}(A)$. We have

$$(\operatorname{adj}(A))_{ij} = c_{ji} = (-1)^{i+j}\det(A(j|i)).$$

With this, our previous notation becomes

$$(\operatorname{adj}(A))\cdot A = (\det(A))\cdot I.$$

Claim 5.25:

$A\cdot(\operatorname{adj}(A)) = (\det(A))\cdot I.$

Proof. Apply the previous claim to A^T to get

$$(\operatorname{adj}(A^T))\cdot A^T = (\det(A^T))\cdot I.$$

Also

$$(\operatorname{adj}(A))^T\cdot A^T = (\det(A))\cdot I.$$

But $\det(A) = \det(A^T)$. Transposing both sides, we get

$$A\cdot\operatorname{adj}(A) = (\det(A))\cdot I.$$

□

Theorem 5.26:

Let $A \in M_{n\times n}(K)$. A is invertible if and only if $\det(A) \ne 0$. In this case,

$$A^{-1} = \frac{1}{\det(A)}\cdot(\operatorname{adj}(A)).$$

5.3 Cramer's Rule

Let $A \in M_{n\times n}(K)$ and $b \in K^n$. Consider the system of equations

$$A\cdot x = b.$$

If $A\cdot x = b$ then $\operatorname{adj}(A)\cdot A\cdot x = \operatorname{adj}(A)\cdot b$. Hence

$$\det(A)\cdot x = \operatorname{adj}(A)\cdot b.$$

Thus, if $\det(A) \ne 0$, then

$$x = \frac{1}{\det(A)}\cdot\operatorname{adj}(A)\cdot b.$$

If we write this in coordinates, we have $\forall 1 \le j \le n$,

$$\det(A)\cdot x_j = \sum_{i=1}^{n} c_{ji}\cdot b_i = \sum_{i=1}^{n} (-1)^{i+j}\cdot b_i\cdot\det(A(i|j)).$$

Consider the matrix

$$A(j,b) = \begin{pmatrix} \cdots & | & \cdots \\ A & b & A \\ \cdots & | & \cdots \end{pmatrix}.$$

The matrix obtained from A by replacing column j of A by the column vector b.

So by previous formulas for E_j, but applied to $A(j,b)$ instead of A, we have

$$\det(A(j,b)) = \sum_{i=1}^{n} (-1)^{i+j} \cdot b_i \cdot \det(A(i|j)).$$

This is Cramer's Rule.

$$x_j = \frac{\det(A(j,b))}{\det(A)}.$$

5.4 Determinants and Endomorphisms

Let $A \in M_{n\times n}(K)$ and $P \in \mathrm{GL}_n(K)$. Then, what happens if we take the determinant of the conjugation of A by P?

$$\det(P^{-1}AP) = \det(P^{-1}) \cdot \det(A) \cdot \det(P) = \det(A),$$

since $\det(P^{-1}) = \det(P)^{-1}$.

So similar matrices have the same determinant.

Let V be a finite n-dimensional vector space over K and $T : V \to V$ be a linear map. Pick a basis $\mathcal{B}$ of V. This gives us a matrix $[T]_{\mathcal{B}}^{\mathcal{B}}$ of T with respect to $\mathcal{B}$.

Consider the determinant of this matrix.

$$\det([T]_{\mathcal{B}}^{\mathcal{B}}) \in K.$$

> **Question 5.27:**
> Does $\det([T]_{\mathcal{B}}^{\mathcal{B}})$ depend on the choice of $\mathcal{B}$?

No, let $\mathcal{B}'$ be another basis of V. Then, recall that

$$[T]_{\mathcal{B}'}^{\mathcal{B}'} = [\mathrm{id}_V]_{\mathcal{B}'}^{\mathcal{B}} \cdot [T]_{\mathcal{B}}^{\mathcal{B}} \cdot [\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}'}.$$

As we have seen, the transition matrices $[\mathrm{id}_V]_{\mathcal{B}'}^{\mathcal{B}}$ and $[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}'}$ are inverses of each other. So the two matrices are conjugates of each other. Hence, they have the same determinant.

Notice that with this notion,

$$\det(\mathrm{id}_V) = 1.$$

Recall also, that if $T, S : V \to V$ are linear maps, then

$$[S \circ T]_{\mathcal{B}}^{\mathcal{B}} = [S]_{\mathcal{B}}^{\mathcal{B}} \cdot [T]_{\mathcal{B}}^{\mathcal{B}}.$$

And also, if T is an isomorphism, then

$$[T^{-1}]_{\mathcal{B}}^{\mathcal{B}} = ([T]_{\mathcal{B}}^{\mathcal{B}})^{-1}.$$

Applying the determinant to these formulas, we get

$$\det(S \circ T) = \det(S) \cdot \det(T)$$
$$\det(T^{-1}) = \frac{1}{\det(T)}.$$

Of course, $\det(0_{V\to V}) = 0$.

Important: This entire story fails if we take two different bases $\mathcal{B}$ and $\mathcal{C}$ and consider $\det([T]_{\mathcal{B}}^{\mathcal{C}})$. This is not an invariant of T.

Following this, we cannot define $\det(T)$ for $T : V \to W$.

> **Note**
> If we have $\dim V = \dim W$, then if $\det(T) = 0$ in any basis then $\det(T) = 0$ in every basis. In this case, T is not an isomorphism.

6 Eigenvectors and Eigenvalues

Consider the following definition.

Definition 6.1: Eigenvalue and Eigenvector

Let V be a vector space over K and $T : V \to V$ be a linear map. We say, that $\lambda \in K$ is an EIGENVALUE of T if there exists a nonzero vector $v \in V$ such that

$$T(v) = \lambda v.$$

Any vector with this property is called an EIGENVECTOR of T for the eigenvalue λ.

Definition 6.2:

Let V be a finite-dimensional vector space over K and $T : V \to V$ be a linear map. We say that T is DIAGONALIZABLE if there exists a basis $\mathcal{B}$ of V such that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}.$$

Lemma 6.3:

T is diagonalizable if and only if there exists a basis of V consisting of eigenvectors of T.

Moreover, if for some basis $\mathcal{B}$ the matrix has the shape as in definition 6.2, then $\forall i$, λ_i is an eigenvalue of T and the i-th vector of $\mathcal{B}$ is an eigenvector of T for the eigenvalue λ_i.

Proof. Recall that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} | & & | \\ [Tv_1]_{\mathcal{B}} & \dots & [Tv_n]_{\mathcal{B}} \\ | & & | \end{pmatrix}.$$

Where

$$[Tv_j]_{\mathcal{B}} = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}.$$

$[T]_{\mathcal{B}}^{\mathcal{B}}$ has the shape as in definition 6.2 if and only if $\alpha_{ij} = 0$ for all $i \neq j$ and $\alpha_{jj} = \lambda_j$ for some $\lambda_j \in K$. This is equivalent to $Tv_j = \lambda_j v_j$ for all j, which is equivalent to $\mathcal{B}$ being a basis of eigenvectors of T. $\square$

Definition 6.4:

A matrix $A \in M_{n\times n}(K)$ is called DIAGONALIZABLE if the linear map $T_A : K^n \to K^n$ is diagonalizable.

Example 6.5:

Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, with $K = \mathbb{R}$. We have

$$T_A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } T_A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

So 3 is an eigenvalue of T_A and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of T_A for the eigenvalue 3. Similarly, $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector of T_A for the eigenvalue 1.

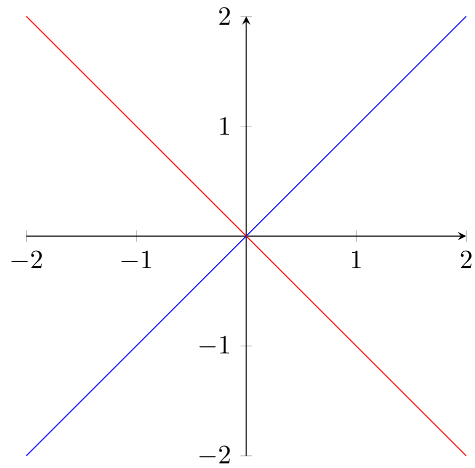


Figure 4: Eigenvalues and eigenvectors of A

In this case,

$$[T_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

So how do we find eigenvalues and eigenvectors? We want to have

$$T_A \cdot v = \lambda v.$$

So also

$$(T_A - \lambda \operatorname{id}_V) \cdot v = 0.$$

We need that $T_A - \lambda I$ is not invertible, so it has determinant 0. Lec 35

Proposition 6.6:

Let V be a vector space over K and $T \in \operatorname{End}(V)$. Suppose $\lambda_1, \dots, \lambda_k$ are pairwise distinct eigenvalues of T. Let $v_1, \dots, v_n$ be eigenvectors for $\lambda_1, \dots, \lambda_k$ respectively. Then, $v_1, \dots, v_k$ are linearly independent.

Proof. We prove this by induction on k.

Base Case ($k = 1$): v_1 is nonzero, so it is linearly independent.

Induction Step: Suppose the proposition holds for every list of k pairwise distinct eigenvalues and corresponding eigenvectors.

Let $\lambda_1, \dots, \lambda_{k+1}$ be pairwise distinct eigenvalues of T and $v_1, \dots, v_{k+1}$ be eigenvectors for $\lambda_1, \dots, \lambda_{k+1}$.

Suppose $v_1, \dots, v_{k+1}$ are linearly dependent. Then, there exists $1 \leq i \leq k+1$ such that v_i is a linear combination of the other v_j's.

$$v_{k+1} = \alpha_1 v_1 + \dots + \alpha_k v_k.$$

Applying T to both sides, we get

$$\lambda_{k+1} v_{k+1} = \alpha_1 \lambda_1 v_1 + \dots + \alpha_k \lambda_k v_k.$$

Substituting the value of v_{k+1} from the previous equation,

$$\lambda_{k+1}(\alpha_1 v_1 + \dots + \alpha_k v_k) = \alpha_1 \lambda_1 v_1 + \dots + \alpha_k \lambda_k v_k.$$

Rearranging, we get

$$\alpha_1(\lambda_{k+1} - \lambda_1) v_1 + \dots + \alpha_k(\lambda_{k+1} - \lambda_k) v_k = 0.$$

By the induction hypothesis, $v_1, \dots, v_k$ are linearly independent. Hence,

$$\alpha_1(\lambda_{k+1} - \lambda_1) = \dots = \alpha_k(\lambda_{k+1} - \lambda_k) = 0.$$

Since $\lambda_1, \dots, \lambda_{k+1}$ are pairwise distinct, $\lambda_{k+1} - \lambda_i \neq 0$ for all i. Hence, $\alpha_1 = \dots = \alpha_k = 0$. Substituting back, we get $v_{k+1} = 0$, which is a contradiction since v_{k+1} is an eigenvector. Hence, $v_1, \dots, v_{k+1}$ are linearly independent. $\square$

Proposition 6.7:

Let V be a vector space over K and $T \in \mathrm{End}(V)$. Then, $\lambda \in K$ is an eigenvalue of T if and only if $T - \lambda \,\mathrm{id}_V$ is not injective.

Proof. λ is an eigenvalue of T is by definition the same as

$$\exists v \neq 0 \in V \text{ st } Tv = \lambda v.$$

This is the same as

$$Tv - \lambda v = 0 \text{ for some } v \neq 0.$$

This is the same as

$$(T - \lambda \,\mathrm{id}_V)v = 0 \text{ for some } v \neq 0.$$

But this exactly is that

$$\mathrm{Ker}(T - \lambda \,\mathrm{id}_V) \neq \{0\}.$$

□

Since we work with endomorphisms, beeing injective is the same as beeing bijective. So we can translate the previous proposition to say that λ is an eigenvalue of T if and only if $T - \lambda \,\mathrm{id}_V$ is not an isomorphism. This is the same as saying that $T - \lambda \,\mathrm{id}_V$ is not invertible, which is the same as saying that $\det(T - \lambda \,\mathrm{id}_V) = 0$.

Definition 6.8: Characteristic Polynomial

Let V be a finite-dimensional vector space over K and $T \in \mathrm{End}(V)$. The CHARACTERISTIC POLYNOMIAL of T is the polynomial

$$p_T(x) = \det(T - x \cdot \mathrm{id}_V) \in K[x].$$

If we instead work with matrices, we can define the characteristic polynomial of a matrix $A \in M_{n\times n}(K)$ as

$$p_A(x) = \det(A - x \cdot I) \in K[x].$$

Definition 6.9:

Let V be a finite-dimensional vector space over K and $T \in \mathrm{End}(V)$. Let λ be an eigenvalue of T. The EIGENSPACE of T corresponding to λ is the subspace

$$\mathrm{Eig}_T(\lambda) = \{v \in V : T(v) = \lambda v\} = \mathrm{Ker}(T - \lambda \,\mathrm{id}_V).$$

Notice, that also $0 \in \mathrm{Eig}_T(0)$ since $T(0) = 0$.

Lemma 6.10:

Let $A \in M_{n\times n}(K)$. Then,

1) $p_A(x)$ is a polynomial of degree n with leading coefficient $(-1)^n$.

2) $p_{A^T}(x) = p_A(x)$.

3) If A is upper triangular with diagonal entries $a_{11}, \dots, a_{nn}$, then

$$p_A(x) = (\lambda_1 - x)\dots(\lambda_n - x).$$

4) If A has a block upper triangular form, where $B \in M_{r\times r}(K)$, $C \in M_{s\times s}(K)$, then

$$p_A(x) = p_B(x) \cdot p_C(x).$$

5) If A and B are similar, then $p_A(x) = p_B(x)$.

Proof. 1) $p_A(x) = \det(A - xI)$. The leading term of $\det(A - xI)$ is the product of the leading terms of the diagonal entries of $A - xI$, which is $(-x)(-x)\dots(-x) = (-1)^n x^n$.

2) $p_{A^T}(x) = \det(A^T - xI) = \det((A - xI)^T) = \det(A - xI) = p_A(x)$.

3) Exercise

4) Exercise

5) $P_B(x) = \det(B - xI) = \det(P^{-1}AP - xI) = \det(P^{-1}(A - xI)P) = \det(A - xI) = p_A(x)$. □

Corollary 6.11:

If V is an n-dimensional vector space over K and $T \in \mathrm{End}(V)$, then $p_T(x)$ is a polynomial in x of degree n with leading coefficient $(-1)^n$.

Proof. Let $\mathcal{B}$ be a basis of V. Then, $p_T(x) = p_{[T]^{\mathcal{B}}_{\mathcal{B}}}(x)$, which is a polynomial of degree n with leading coefficient $(-1)^n$. Since $p_T(x)$ does not depend on the choice of $\mathcal{B}$, the result follows. □

Definition 6.12: Trace

Let $A \in M_{n\times n}(K)$. Define the TRACE of A as

$$\mathrm{tr}(A) := A_{11} + A_{22} + \dots + A_{nn}.$$

If V is finite-dimensional and $T \in \mathrm{End}(V)$, define the TRACE of T as

$$\mathrm{tr}(T) := \mathrm{tr}([T]^{\mathcal{B}}_{\mathcal{B}}).$$

Lemma 6.13:

1) $\forall A, B \in M_{n\times n}(K)$, $\mathrm{tr}(A \cdot B) = \mathrm{tr}(B \cdot A)$.

2) If A and B are similar, then $\mathrm{tr}(A) = \mathrm{tr}(B)$. In particular, if $\mathcal{B}$ and $\mathcal{B}'$ are two bases of V, then $\mathrm{tr}([T]^{\mathcal{B}}_{\mathcal{B}}) = \mathrm{tr}([T]^{\mathcal{B}'}_{\mathcal{B}'})$.

Proof. 1) We compute

$$\begin{aligned}\operatorname{tr}(A \cdot B) &= \sum_{i=1}^{n} (A \cdot B)_{ii} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} \cdot B_{ji} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} B_{ji} \cdot A_{ij} \\ &= \operatorname{tr}(B \cdot A).\end{aligned}$$

2) If $B = P^{-1}AP$, then

$$\operatorname{tr}(B) = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(AP \cdot P^{-1}) = \operatorname{tr}(A).$$

If, $\mathcal{B}'$ is another basis for V, then

$$\operatorname{tr}([T]^{\mathcal{B}'}_{\mathcal{B}'}) = \operatorname{tr}([\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}'} \cdot [T]^{\mathcal{B}}_{\mathcal{B}} \cdot [\mathrm{id}_V]^{\mathcal{B}'}_{\mathcal{B}}) = \operatorname{tr}([T]^{\mathcal{B}}_{\mathcal{B}}).$$

Lec 36

□

Lemma 6.14:

Let $A \in M_{n\times n}(K)$. Then, the coefficient of x^{n-1} in $p_A(x)$ is $(-1)^{n-1} \cdot \operatorname{tr}(A)$. Furthermore, the coefficient of x^0 is $\det(A)$.

Proof. Recall that $p_A(x) = \det(A - xI)$. Denote this matrix by B. The determinant of B is given as

$$(A_{11} - x)(A_{22} - x)\dots(A_{nn} - x) + \sum_{\sigma\neq \mathrm{id}} \operatorname{sgn}(\sigma) B_{1\sigma(1)} \dots B_{n\sigma(n)}$$

.

If σ is not the identity, there must be at least two indices $i \neq j$ such that $\sigma(i) \neq i$ and $\sigma(j) \neq j$. So at least two factors do not contain x, hence the degree of the second term is at most $n - 2$.

So $p_A(x) = (A_{11} - x)(A_{22} - x)\dots(A_{nn} - x) +$ lower degree terms. Opening the brackets, we get

$p_A(x) = (-1)^n x^n + (-1)^{n-1}(A_{11} + \dots + A_{nn})x^{n-1} +$ lower deg terms.

But $A_{11} + \cdots + A_{nn} = \operatorname{tr}(A)$. Hence, the coefficient of x^{n-1} is $(-1)^{n-1} \cdot \operatorname{tr}(A)$.

The last coefficient is very easy to determine. Plug in $x = 0$ to get $p_A(0) = \det(A)$, which is the coefficient of x^0. □

Definition 6.15:

Let V be a vector space over K and $U_1, \dots, U_r \subseteq V$ be linear subspaces. Denote

$$W := U_1 + \cdots + U_r = \{u_1 + \cdots + u_r : u_i \in U_i\}.$$

We say that W is a direct sum of $U_1, \dots, U_r$ if $\forall w \in W$, there exists a unique representation as a sum $w = u_1 + \cdots + u_r$ with $u_i \in U_i$.

Exercise 6.16:

Show that $U_1 + \cdots + U_r$ is a direct sum if and only if any of the following equivalent conditions hold:

1) The only way to write 0 as a sum $u_1 + \cdots + u_r$ with $u_i \in U_i$ is by taking all u_i's to be 0.

2) For every $2 \leq j \leq r$ we have $U_j \cap (U_1 + \cdots + U_{j-1}) = \{0\}$.

Under the additional assumption that the U_i's are finite dimensional, also

3) $\dim(U_1 + \cdots + U_r) = \dim(U_1) + \cdots + \dim(U_r)$.

4) Whenever B_i is a basis of U_i for all i, then $\cup_{i=1}^r B_i$ is a basis of $U_1 + \cdots + U_r$.

Exercise 6.17:

If $U_1 + \cdots + U_r$ is a direct sum, then there exists a canonical isomorphism between $U_1 + \cdots + U_r$ and $U_1 \oplus \cdots \oplus U_r$, given by

$$u_1 + \cdots + u_r \mapsto (u_1, \dots, u_r).$$

Proposition 6.18:

Let V be a vector space over K, and $T \in \operatorname{End}(V)$. Let $\lambda_1, \dots, \lambda_r \in K$ be pairwise distinct eigenvalues of T. Then, $\operatorname{Eig}_T(\lambda_1) + \cdots + \operatorname{Eig}_T(\lambda_r)$ is a direct sum.

Proof. Assume $u_1 + \cdots + u_r = u_1' + \cdots + u_r'$ where $u_i, u_i' \in \operatorname{Eig}_T(\lambda_i)$ for all i. Rewriting this, we get

$$(u_1 - u_1') + \cdots + (u_r - u_r') = 0.$$

Ignore in this sum all the summands which are 0.

$$(u_{i_1} - u_{i_1}') + \cdots + (u_{i_l} - u_{i_l}') = 0.$$

Notice that $u_{i_j} - u_{i_j}' \in \operatorname{Eig}_T(\lambda_{i_j}) \setminus 0$ for all j. So $u_{i_j} - u_{i_j}'$ is an eigenvector of T. By proposition 6.6, we have that $u_{i_1} - u_{i_1}', \dots, u_{i_l} - u_{i_l}'$ are linearly independent. This is a contradiction since they sum to 0. Hence, $u_i = u_i'$ for all i, proving the proposition. □

Corollary 6.19:

Assume V is finite-dimensional. Let $T \in \operatorname{End}(V)$ and $\lambda_1, \dots, \lambda_r$ be all the eigenvalues of T (we assume they are pairwise distinct). Then, T is diagonalizable if and only if

$$\sum_{i=1}^{r} \dim(\operatorname{Eig}_T(\lambda_i)) = \dim(V).$$

Proof. $\Leftarrow$) Choose a basis $\mathcal{B}_i$ of $\operatorname{Eig}_T(\lambda_i)$. We've seen that $W := \operatorname{Eig}_T(\lambda_1) + \cdots + \operatorname{Eig}_T(\lambda_r)$ is a direct sum.

But then, $(B_1, B_2, \dots, B_r)$ is a basis of W. Hence,

$$\dim(W) = \sum_{i=1}^{r} \dim(\operatorname{Eig}_T(\lambda_i)) = \dim(V).$$

Since $W \subseteq V$, we get $W = V$. So V has a basis consisting of eigenvectors of T. Hence, T is diagonalizable.

$\Rightarrow$) Let $\mathcal{B}$ be a basis of V such that $[T]^{\mathcal{B}}_{\mathcal{B}}$ is diagonal. Then, the vectors of $\mathcal{B}$ are eigenvectors of T. Hence, $\mathcal{B}$ contains a basis of $\operatorname{Eig}_T(\lambda_i)$ for all i. So $\sum_{i=1}^{r} \dim(\operatorname{Eig}_T(\lambda_i)) \leq \dim(V)$. But $\sum_{i=1}^{r} \dim(\operatorname{Eig}_T(\lambda_i)) = \dim(V)$, so we get $\sum_{i=1}^{r} \dim(\operatorname{Eig}_T(\lambda_i)) = \dim(V)$. □

Example 6.20:

Let $A = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then

$$p_A(x) = (x-1)^2 = p_B(x).$$

Example 6.21:

Pick $\lambda \in K$. Consider the matrix

$$J := J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix},$$

which is called the Jordan block. Then,

$$p_J(x) = (\lambda - x)^n.$$

The only Eigenvalue of J is λ. The eigenvectors of J are

$$J - \lambda I = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The columns 2 to n are linearly independent, so $\operatorname{rank}(J - \lambda I) = n - 1$. Hence, $\dim(\operatorname{Eig}_J(\lambda)) = \dim(\operatorname{Ker}(J - \lambda I)) = 1$. It's eigenvector is $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

Lemma 6.22:

Let $A \in M_{n\times n}(K)$. Assume A is diagonalizable. Then, $p_A(x)$ splits as

$$p_A(x) = (\lambda_1 - x) \dots (\lambda_r - x).$$

Where $\lambda_1, \dots, \lambda_r$ are the eigenvalues of A, where repetitions are allowed.

Notice that the other direction is not true. For example, consider the above example on Jordan blocks.

Proof. By assumption, $\exists$ an invertible matrix P such that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} =: \Lambda.$$

Hence,

$$p_A(x) = p_\Lambda(x) = (\lambda_1 - x) \dots (\lambda_n - x).$$

□

Example 6.23:

Consider $K = \mathbb{R}, A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $T_A : \mathbb{R}^2 \to \mathbb{R}^2$. Then

$$p_A(x) = x^2 + 1.$$

This does not have any zeros in $K = \mathbb{R}$ so no eigenvalues exist.

Lec 37

6.1 Geometric and Algebraic Multiplicity

Let's do a quick digression on polynomials. Let $0 \neq p(x) \in K[x]$ be a nonzero polynomial. By dividing $p(x)$ by $x - \lambda$ we can write $p(x) = (x - \lambda)q(x) + r$ where $q(x) \in K[x]$ with $\deg(q) = \deg(p) - 1$ and $r \in K$.

Substituting $x = \lambda$, we get that $p(\lambda) = r$. So

$$p(x) = (x - \lambda)q(x) + p(\lambda).$$

So $p(\lambda) = 0$ if and only if $p(x) = (x - \lambda)q(x)$. We say that λ is a ZERO or ROOT of $p(x)$.

Sometimes, λ is also a zero of $q(x)$. This leads to the following definition.

Definition 6.24:

Let $p(x) \in K[x]$ be a nonzero polynomial and $\lambda \in K$. We say that λ is a ZERO of $p(x)$ of MULTIPLICITY $m \in \mathbb{Z}_{\geq 1}$ if

$$p(x) = (x - \lambda)^m g(x),$$

Where $g(x) \in K[x]$ is a polynomial such that $g(\lambda) \neq 0$.

If $m = 1$, we say that λ is a SIMPLE ZERO of $p(x)$. If $m \geq 2$, we say that λ is a MULTIPLE ZERO of $p(x)$.

Definition 6.25:

Let V be a vector space over K and $T \in \operatorname{End}(V)$ and $\lambda \in K$ be an eigenvalue of T. We define the GEOMETRIC MULTIPLICITY of λ as

$$m_g(T, \lambda) := \dim(\operatorname{Eig}_T(\lambda)).$$

Assume in addition that V is finite-dimensional. We define the ALGEBRAIC MULTIPLICITY of λ as the multiplicity of λ as a zero of $p_T(x)$, which we denote by $m_a(T, \lambda)$.

Proposition 6.26:

The geometric multiplicity is smaller than or equal to the algebraic multiplicity, i.e. $m_g(T, \lambda) \leq m_a(T, \lambda)$.

Proof. Let $v_1, \dots, v_k$ be a basis for $\operatorname{Eig}_T(\lambda)$, so $k = m_g(T, \lambda)$. We can extend this basis to a basis $\mathcal{B}$ of V given by

$$\mathcal{B} = (v_1, \dots, v_k, w_{k+1}, \dots, w_n).$$

We have

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{pmatrix} \lambda & 0 & \dots & 0 & * & \dots & * \\ 0 & \lambda & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda & * & \dots & * \\ 0 & 0 & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & * & \dots & * \end{pmatrix}.$$

The characteristic polynomial of T is given by

$$p_T(x) = (-1)^k(\lambda - x)^k \cdot \det \begin{pmatrix} * & \dots & * \\ \vdots & \ddots & \vdots \\ * & \dots & * \end{pmatrix}.$$

From this, we see that the algebraic multiplicity of λ is at least k, which is the geometric multiplicity of λ. Hence, $m_g(T, \lambda) \leq m_a(T, \lambda)$. □

Example 6.27:

Consider $J = J_{\lambda,n}$. We saw that $p_J(x) = (\lambda - x)^n$, so $m_a(J,\lambda) = n$. We also saw that $\dim(\mathrm{Eig}_J(\lambda)) = 1$, so $m_g(J,\lambda) = 1$. Hence, $m_g(J,\lambda) \le m_a(J,\lambda)$.

Theorem 6.28:

Let V be an n-dimensional vector space over K and $T \in \mathrm{End}(V)$. The following 5 statements are all equivalent:

1) T is diagonalizable.

2) $\exists$ a basis for V consisting of eigenvectors of T.

3) The characteristic polynomial of T splits as a product of linear factors and for every eigenvalue λ of T, we have $m_g(T,\lambda) = m_a(T,\lambda)$.

4) $\sum_{i=1}^k \dim \mathrm{Eig}_T(\lambda_i) = n$, where $\lambda_1, \dots, \lambda_k$ are all the eigenvalues of T.

5) $V \cong \bigoplus_{i=1}^k \mathrm{Eig}_T(\lambda_i)$, where $\lambda_1, \dots, \lambda_k$ are all the eigenvalues of T.

Proof. 1) $\Leftrightarrow$ 2) we have already seen.

1) $\Rightarrow$ 3) Let $\mathcal{B}$ be a basis of V of which $[T]_\mathcal{B}^\mathcal{B}$ is diagonal. By changing the order of the elements of $\mathcal{B}$, we can arrange that

$$[T]_\mathcal{B}^\mathcal{B} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{pmatrix}.$$

Where λ_i appears $m_g(T,\lambda_i)$ times. We can write

$$p_T(x) = (\lambda_1 - x)^{m_g(T,\lambda_1)} \dots (\lambda_k - x)^{m_g(T,\lambda_k)}.$$

So $p_T(x)$ splits as a product of linear factors.

3) $\Rightarrow$ 4) Let $\lambda_1, \dots, \lambda_k$ be all the eigenvalues of T. By assumption,

$$p_T(x) = (\lambda_1 - x)^{m_a(T,\lambda_1)} \dots (\lambda_k - x)^{m_a(T,\lambda_k)}.$$

And $m_g(T,\lambda_i) = m_a(T,\lambda_i)$ for all i. Notice, that $\lambda_1, \dots, \lambda_k$ are all eigenvalues of T. Hence,

$$\sum_{i=1}^k \dim \mathrm{Eig}_T(\lambda_i) = \sum_{i=1}^k m_g(T,\lambda_i) = \sum_{i=1}^k m_a(T,\lambda_i) = n.$$

4) $\Rightarrow$ 5) Recall that $\mathrm{Eig}_T(\lambda_1) + \dots + \mathrm{Eig}_T(\lambda_k)$ is a direct sum. So

$$\dim V = \sum_{i=1}^k \dim \mathrm{Eig}_T(\lambda_i) = \dim(\mathrm{Eig}_T(\lambda_1) + \dots + \mathrm{Eig}_T(\lambda_k)).$$

Since the dimensions are equal, and $\mathrm{Eig}_T(\lambda_1) + \dots + \mathrm{Eig}_T(\lambda_k) \subseteq V$, we get that $V \cong \mathrm{Eig}_T(\lambda_1) + \dots + \mathrm{Eig}_T(\lambda_k) = \bigoplus_{i=1}^k \mathrm{Eig}_T(\lambda_i)$.

5) $\Rightarrow$ 1) If $V \cong \mathrm{Eig}_T(\lambda_1) \oplus \dots \oplus \mathrm{Eig}_T(\lambda_k)$, then since $\mathrm{Eig}_T(\lambda_1) + \dots + \mathrm{Eig}_T(\lambda_k) \subset V$ is a direct sum, we must have that

$$\mathrm{Eig}_T(\lambda_1), \dots, \mathrm{Eig}_T(\lambda_k) = V.$$

So if we choose a basis $\mathcal{B}_i$ of $\mathrm{Eig}_T(\lambda_i)$ for all i, then $\cup_{i=1}^k \mathcal{B}_i$ is a basis of V consisting of eigenvectors of T. Hence, T is diagonalizable. ☐

Corollary 6.29:

If $p_T(x)$ splits as a product of linear factors and each eigenvalue λ has $m_a(T,\lambda) = 1$, then T is diagonalizable.

Proof. We have that $1 \le m_g(T,\lambda) \le m_a(T,\lambda) = 1$. Hence, $m_g(T,\lambda) = m_a(T,\lambda)$ for all eigenvalues λ of T. By the previous theorem, T is diagonalizable. ☐

We call the process of bringing a matrix / endomorphism to a upper triangular form a TRIANGULARIZATION.

Definition 6.30:

We say that $T \in \mathrm{End}(V)$ is TRIANGULARIZABLE if there exists a basis $\mathcal{B}$ of V such that $[T]_\mathcal{B}^\mathcal{B}$ is upper triangular.

Theorem 6.31:

Let V be a vector space over K and $T \in \mathrm{End}(V)$. Then T is triangularizable if and only if the characteristic polynomial of T splits as a product of linear factors.

As a short digression, the fundamental theorem of algebra.

Theorem 6.32: Fundamental Theorem of Algebra

Let $p(x) \in \mathbb{C}[x]$ be a nonzero polynomial. Then, $p(x)$ splits as a product of linear factors and a constant factor, namely

$$p(x) = c \cdot (x - \lambda_1) \dots (x - \lambda_n).$$

Lec 38

Corollary 6.33:

Let V be a finite dimensional vector space over $\mathbb{C}$ and $T \in \mathrm{End}(V)$. Then, T is triangularizable.

Proof. [Theorem 6.31] $\Rightarrow$) Let $\mathcal{B}$ be a basis for V such that $[T]_\mathcal{B}^\mathcal{B}$ is upper triangular. Then, by lemma 6.10, the characteristic polynomial of T splits as a product of linear factors.

$$p_T(x) = \det([T]_\mathcal{B}^\mathcal{B} - xI) = (\lambda_1 - x) \dots (\lambda_n - x).$$

$\Leftarrow$) We will prove this by induction on $n := \dim(V)$. For $n = 1$, the result is trivial. Assume $n \ge 1$ and the result holds for all vector spaces of dimension not more than n. Let V be an $n+1$ dimensional vector space and $T \in \mathrm{End}(V)$ such that $p_T(x)$ splits as a product of linear factors.

Since $p_T(x)$ splits as a product of linear factors, T has an eigenvalue $\lambda_1 \in K$. Let $u_1 \in \mathrm{Ker}(T - \lambda_1 \,\mathrm{id}_V)$ be a nonzero eigenvector of T corresponding to λ_1. Extend u_1 to a basis $\mathcal{B}' = (u_1, u_2, \dots, u_{n+1})$ of V. Then

$$[T]_{\mathcal{B}'}^{\mathcal{B}'} \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \dots & * \end{pmatrix}.$$

Let us call the $n \times n$ matrix in the bottom right corner A. Let $T_A : K^n \to K^n$ be the endomorphism corresponding to A. We have that $p_T(x) = (\lambda_1 - x) \cdot p_A(x)$. Since $p_T(x)$ splits as a product of linear factors, $p_A(x)$ also splits as a product of linear factors. By the induction hypothesis, there exists a basis $\mathcal{B}$ of K^n such that $[T_A]_\mathcal{C}^\mathcal{C}$ is upper triangular. Denote by $\mathcal{E}$ the standard basis of K^n. Then,

$$[S_A]_\mathcal{C}^\mathcal{C} = [\mathrm{id}_{K^n}]_\mathcal{C}^\mathcal{E} \cdot [T_A]_\mathcal{E}^\mathcal{E} \cdot [\mathrm{id}_{K^n}]_\mathcal{E}^\mathcal{C}.$$

Define

$$P := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & [\mathrm{id}_{K^n}]_\mathcal{E}^\mathcal{C} & \\ 0 & & & \end{pmatrix}.$$

Claim

We claim P is invertible and

$$P^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & [\mathrm{id}_{K^n}]^{\mathcal{E}}_{\mathcal{C}} & \\ 0 & & & \end{pmatrix}.$$

aswell as that $P^{-1}MP$ is upper triangular.

We compute $P \cdot Q$ we get the identity matrix, so P is invertible and P^{-1} is as claimed. We compute

$$P^{-1}MP = \begin{pmatrix} \lambda_1 & \tilde{*} & \dots & \tilde{*} \\ 0 & & & \\ \vdots & & [\mathrm{id}_{K^n}]^{\mathcal{E}}_{\mathcal{C}} \cdot [T_A]^{\mathcal{E}}_{\mathcal{E}} \cdot [\mathrm{id}_{K^n}]^{\mathcal{C}}_{\mathcal{E}} & \\ 0 & & & \end{pmatrix}.$$

Claim

We claim that $\exists$ a basis $\mathcal{B}$ of K^n such that $[\mathrm{id}]^{\mathcal{B}}_{\mathcal{B}'}$ is P.

Write

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{pmatrix}.$$

Where $m = n+1$. Write B' as $B' = (u_1, u_2, \dots, u_m)$. Define

$$\begin{aligned} v_1 &:= p_{11}u_1 + p_{21}u_2 + \dots + p_{m1}u_m \\ v_2 &:= p_{12}u_1 + p_{22}u_2 + \dots + p_{m2}u_m \\ &\vdots \\ v_m &:= p_{1m}u_1 + p_{2m}u_2 + \dots + p_{mm}u_m. \end{aligned}$$

This is a basis of V (exercise). We now have $[T]^{\mathcal{B}}_{\mathcal{B}} = P^{-1}MP$, which is upper triangular. Hence, T is triangularizable. □

6.2 Minimal Polynomials and Cayley-Hamilton

Let $A \in M_{n\times n}(K)$. $\forall k \geq 1$, let $A^k := \underbrace{A \cdot A \cdot \dots \cdot A}_{k \text{ times}}$. We also define $A^0 = I$.

Let $g(x) = a_d x^d + a_{d-1}x^{d-1} + \dots + a_0 \in K[x]$ be a polynomial. We define

$$g(A) := a_d A^d + a_{d-1}A^{d-1} + \dots + a_0 I.$$

Claim 6.34:

For every $A \in M_{n\times n}(K)$, there exists a nonzero polynomial $g(x) \in K[x]$ such that $g(A) = 0$.

Proof. Consider $I_n, A, A^2, \dots, A^{n^2}$. This is a collection of n^2+1 matrices in the n^2-dimensional vector space $M_{n\times n}(K)$. Hence, they are linearly dependent. So $\exists a_0, a_1, \dots, a_{n^2} \in K$, not all zero, such that

$$a_0 I_n + a_1 A + \dots + a_{n^2}A^{n^2} = 0.$$

Let $g(x) = a_{n^2}x^{n^2} + \dots + a_1 x + a_0$. Then, $g(A) = 0$. □

Everything here, also applies to $T \in \mathrm{End}(V)$, where $T^k := \underbrace{T \circ T \circ \dots \circ T}_{k \text{ times}}$ and $T^0 = \mathrm{id}_V$. We can also define $g(T)$ for a polynomial $g(x)$ as above, by

$$g(T) = a_d T^d + a_{d-1}T^{d-1} + \dots + a_0 \,\mathrm{id}_V.$$

Also if V is finite-dimensional, then $\mathrm{End}(V)$ has dimension $\dim(V)^2$, so we can find a nonzero polynomial $g(x)$ such that $g(T) = 0$ by the same argument as above.

Definition 6.35:

Let $T \in \mathrm{End}(V)$. A MINIMAL POLYNOMIAL for T is a nonzero polynomial $g(x) \in K[x]$ of the minimal possible degree, such that $g(T) = 0$.

Theorem 6.36: Cayley-Hamilton

1) Let $A \in M_{n\times n}(K)$. Then, $p_A(A) = 0$.

2) Let V be a finite-dimensional vector space over K and $T \in \mathrm{End}(V)$. Then, $p_T(T) = 0$.

Exercise 6.37:

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$p_A(x) = x^2 - (a+d)x + (ad - bc).$$

Compute $p_A(A)$ and verify that it is the zero matrix.

We first do an outline of an alternative analytic proof

Proof. Assume first, that $K = \mathbb{C}$. Proof first for $A \in M_{n\times n}(\mathbb{C})$, which is diagonalizable. Indeed, if there is an invertible matrix Q, such that $Q^{-1}AQ = \Lambda$ is diagonal, with diagonal entries $\lambda_1, \dots, \lambda_n$, then

$$p_A(x) = c_n x^n + c_{n-1}x^{n-1} + \dots + c_0.$$

And

$$Q^{-1}p_A(A)Q = p_A(\Lambda) = \begin{pmatrix} p_A(\lambda_1) & 0 & \dots & 0 \\ 0 & p_A(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_A(\lambda_n) \end{pmatrix}.$$

But this is the zero matrix since $p_A(\lambda_i) = 0$ for all i. Hence, $p_A(A) = 0$, since Q is invertible. □

We need some preparations for the algebraic proof.

Definition 6.38:

Let V be a vector space over K and $T \in \mathrm{End}(V)$. Let $v \in V$. Define $\langle v \rangle_T := \mathrm{Sp}\{v, T(v), T^2(v), \dots\}$.

$\langle v \rangle_T$ is called the T-CYCLIC SUBSPACE generated by v.

Exercise 6.39:

1) Show that $\langle v \rangle_T$ is invariant under T. i.e.

$$T(\langle v \rangle_T) \subseteq \langle v \rangle_T.$$

2) $\langle v \rangle_T$ is the smallest subspace of V which both contains v and is invariant under T.

Lemma 6.40:

Suppose that $v, Tv, \dots, T^{d-1}v$ for $d \geq 1$ are linearly independent, and

$$T^d v = a_{d-1}T^{d-1}v + \dots + a_1 Tv + a_0 v.$$

Then,

1) $\mathcal{B} = (v, Tv, \dots, T^{d-1}v)$ is a basis of $\langle v \rangle_T$.

2) Denote $S := T|_{\langle v \rangle_T}$. Then,

$$[S]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{d-1} \end{pmatrix}.$$

Proof. 1) Denote by $\mathcal{U} := \mathrm{Sp}\{v, Tv, \dots, T^{d-1}v\}$. Note that for $w \in \mathcal{B}$ then $T(w) \in \mathcal{U}$. So U is T-invariant. Furthermore, $v \in U$, By the previous exercise, $\langle v \rangle_T \subset U$. But clearly, also $U \subset \langle v \rangle_T$. Hence, $\langle v \rangle_T = U$.

Since by assumption, the elements of $\mathcal{B}$ are linearly independent, and $\mathcal{B}$ spans $\langle v \rangle_T$, we get that $\mathcal{B}$ is a basis of $\langle v \rangle_T$.

2) Direct calculation. $\square$

Lemma 6.41:

Suppose that $v, Tv, \dots, T^{d-1}v$ for $d \geq 1$ are linearly independent, and

$$T^d v = a_{d-1}T^{d-1}v + \dots + a_1 Tv + a_0 v.$$

Then,

$$p_S(x) = (-1)^d (x^d - a_{d-1}x^{d-1} - \dots - a_1 x - a_0).$$

Proof. We compute

$$[S]_{\mathcal{B}}^{\mathcal{B}} - xI = \begin{pmatrix} -x & 0 & \dots & 0 & a_0 \\ 1 & -x & \dots & 0 & a_1 \\ 0 & 1 & \dots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & a_{d-1} - x \end{pmatrix}.$$

We will argue by induction on d. For $d = 1, 2$ the result is easy to verify. Assume $d \geq 3$ and the result holds for all smaller values of d. We compute

$$\det([S]_{\mathcal{B}}^{\mathcal{B}} - xI) =$$
$$-x \cdot \det \begin{pmatrix} -x & \dots & 0 & a_1 \\ 1 & \dots & 0 & a_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & a_{d-1} - x \end{pmatrix} + (-1)^{d+1} a_0 \cdot 1.$$

But this first matrix is the same as $[S]_{\mathcal{B}}^{\mathcal{B}} - xI$ for $d-1$, so by the induction hypothesis, we get that

$$\det([S]_{\mathcal{B}}^{\mathcal{B}} - xI) = (-1)^d (x^d - a_{d-1}x^{d-1} - \dots - a_1 x - a_0).$$

$\square$

Proof. [of Theorem 6.36] Let V be finite dimensional and $T \in \mathrm{End}(V)$. Then $p_T(x) := \det(T - x\,\mathrm{id}_V)$. We want to show that $p_T(T) = 0$. This is equivalent to showing that $p_T(T)(v) = 0$ for all $v \in V$.

Let $v \in V$. If $v = 0$, then $p_T(T)(v) = 0$. Assume $v \neq 0$. Consider the cyclic subspace $\langle v \rangle_T$. Take the maximal $d \geq 1$ such that $v, Tv, \dots, T^{d-1}v$ are linearly independent. Since V is finite-dimensional, such a d is well defined. Since d is maximal, we have that $T^d v$ is a linear combination of $v, Tv, \dots, T^{d-1}v$. So

$$T^d v = a_{d-1}T^{d-1}v + \dots + a_1 Tv + a_0 v.$$

Put $n = \dim(V)$. If $d < n$ extend $v, Tv, \dots, T^{d-1}v$ to a basis $\mathcal{E}$ of V. Then, $[T]_{\mathcal{E}}^{\mathcal{E}}$ has the form of the following block matrix.

$$[T]_{\mathcal{E}}^{\mathcal{E}} = \begin{pmatrix} [S]_{\mathcal{B}}^{\mathcal{B}} & * \\ 0 & C \end{pmatrix}.$$

The characteristic polynomial of T is given by

$$p_T(x) = p_S(x) \cdot p_C(x).$$

Put $R := p_T(T) \in \mathrm{End}(V)$. So $R = p_S(T) \cdot p_C(T)$. So it follows, that $R(v) = p_C(T) \circ p_S(T)(v)$. But by lemma 6.41,

$$p_S(T)(v) = (-1)^d (T^d v - a_{d-1}T^{d-1}v - \dots - a_1 Tv - a_0 v) = 0.$$

So $R(v) = 0$ for all $v \in V$. Hence, $p_T(T) = 0$. $\square$

Lec 39

7 Euclidean and Hermitian Spaces

In the following we will work over $K = \mathbb{R}$ or $K = \mathbb{C}$.

Definition 7.1: Scalar Product

Let V be a vector space over $\mathbb{R}$. A SCALAR PRODUCT (or inner product) is a function $V \times V \to \mathbb{R}$, denoted by $(v, w) \mapsto \langle v, w\rangle$, that satisfies

1) Linearity in the first variable: $\langle av + bw, u\rangle = a\langle v, u\rangle + b\langle w, u\rangle$.

2) Linearity in the second variable: $\langle v, au + bw\rangle = a\langle v, u\rangle + b\langle v, w\rangle$.

3) Symmetry: $\langle v, w\rangle = \langle w, v\rangle$.

4) Positivity: $\forall v \neq 0$, $\langle v, v\rangle > 0$.

We call $(V, \langle\cdot,\cdot\rangle)$ an EUCLIDEAN SPACE or a space with a scalar product.

Definition 7.2: Hermitian Product

Let V be a vector space over $\mathbb{C}$. A HERMITIAN PRODUCT (or inner product) is a function $V \times V \to \mathbb{C}$, denoted by $(v, w) \mapsto \langle v, w\rangle$, that satisfies

1) Linearity in the first variable: $\langle av + bw, u\rangle = a\langle v, u\rangle + b\langle w, u\rangle$.

2) Complex antilinearity or Sesquilinearity in the second variable: $\langle v, au + bw\rangle = \overline{a}\langle v, u\rangle + \overline{b}\langle v, w\rangle$.

3) Hermitian property: $\langle v, w\rangle = \overline{\langle w, v\rangle}$.

4) Positivity: $\forall v \neq 0$, $\langle v, v\rangle > 0$.

We call $(V, \langle\cdot,\cdot\rangle)$ a HERMITIAN SPACE or a space with a Hermitian product.

Notice that for positivity, $\langle v, v\rangle \in \mathbb{R}$.

Two properties we will use often about complex numbers are the following:

1) $\alpha \in \mathbb{C}$ is real iff $\alpha = \overline{\alpha}$.

2) $\forall \beta \in \mathbb{C}$, $\beta \cdot \overline{\beta} = |\beta|^2$.

Lemma 7.3:

Let V be an Euclidean or Hermitian space and $v, w \in V$. Then,

1) $\langle v, 0\rangle = \langle 0, v\rangle = 0$.

2) If $\langle v, w\rangle = 0$ for all $v \in V$, then $w = 0$.

3) If $\langle v, w\rangle = \langle v, w'\rangle$ for all $v \in V$, then $w = w'$.

Proof. Exercise. □

Definition 7.4: Norm

Let V be an Euclidean or Hermitian space and $v \in V$. We define the NORM of v as $\|v\| := \sqrt{\langle v, v\rangle} \in \mathbb{R}_{\geq 0}$ induced by $\langle\cdot,\cdot\rangle$.

Definition 7.5: Unit Vector

A vector $u \in V$ is called a UNIT VECTOR if $\|u\| = 1$.

Note that $\forall v \neq 0$, $\frac{1}{\|v\|}v$ is a unit vector. So every nonzero vector can be rescaled to a unit vector. We call $\frac{v}{\|v\|}$ the NORMALIZATION of v.

Definition 7.6: Distance

$\forall v, w \in V$, we define the DISTANCE between v and w as $d(v, w) = \|v - w\|$.

Definition 7.7: Orthogonality

Two vectors $u, v \in V$ are called ORTHOGONAL if $\langle u, v\rangle = 0$. In this case, we write $u \perp v$.

Definition 7.8: Orthogonal System

A subset $S \subset V$ is called an ORTHOGONAL SYSTEM if $0 \notin S$ and $\forall u, v \in S$ with $u \neq v$, we have $u \perp v$.

Definition 7.9: Orthonormal System

An orthogonal system $S \subset V$ is called an vocaborthonormal system if $\forall u \in S$, $\|u\| = 1$.

Theorem 7.10: Pythagorean Theorem

Let V be an Euclidean or Hermitian space and $u, v \in V$ orthogonal. Then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Proof. We compute

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \|u\|^2 + 0 + 0 + \|v\|^2 \\ &= \|u\|^2 + \|v\|^2.\end{aligned}$$

□

Exercise 7.11:

1) Let V be Euclidean. Show that $\forall u, v \in V$, we have

$$\langle u, v\rangle = \frac{1}{2}(\|u+v\|^2 - \|u\|^2 - \|v\|^2).$$

2) What happens with the above, if V is Hermitian?

Important about the first exercise is, that we can find the scalar product after having defined the norm.

Example 7.12: Standard Scalar Product

The standard scalar product on $\mathbb{R}^n$ is given by

$$\langle u, v\rangle := u^T \cdot v = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Example 7.13: Standard Hermitian Product

The standard Hermitian product on $\mathbb{C}^n$ is given by

$$\langle u, v\rangle := u^\dagger \cdot v = u^t\overline{v} = \overline{u_1}v_1 + \overline{u_2}v_2 + \cdots + \overline{u_n}v_n.$$

Example 7.14:

If V is an Euclidean or Hermitian space and $W \subset V$ is a subspace, then the restriction of $\langle \cdot, \cdot \rangle$ to $W \times W$ is a scalar product on W. Hence, W is an Euclidean or Hermitian space aswell.

Definition 7.15: Symmetric Matrices

A matrix $A \in M_{n\times n}(K)$ is called SYMMETRIC if $A^T = A$.

Definition 7.16: Hermitian Matrices

Let $A \in M_{m\times n}(\mathbb{C})$. Write $\overline{A} = (\overline{a_{ij}})$.

Let $A \in M_{n\times n}(\mathbb{C})$. We define the ADJOINT MATRIX of A as $A^\dagger = \overline{A}^T$.

$A \in M_{n\times n}(\mathbb{C})$ is called HERMITIAN if $A^\dagger = A$.

Example 7.17:

Let $A \in M_{n\times n}(\mathbb{R})$. Define $\langle u, v\rangle_A = u^T A v$. This is always bilinear. Assume that A is symmetric. Then,

$$\langle u, v\rangle_A = u^T A v = v^T A^T u = v^T A u = \langle v, u\rangle_A.$$

The question now is, if $\langle \cdot, \cdot\rangle_A$ is a scalar product. In general, no. For example, the 0 matrix.

Definition 7.18: Positive Definite Matrices

A symmetric matrix $A \in M_{n\times n}(\mathbb{R})$ is called POSITIVE DEFINITE if $\forall u \in \mathbb{R}^n \setminus \{0\}$, $u^T A u > 0$.

Claim 7.19:

1) If A is positive definite, then $\langle u, v\rangle_A = u^T A v$ is a scalar product on $\mathbb{R}^n$.

2) If $\langle \cdot, \cdot\rangle_A$ is a scalar product on $\mathbb{R}^n$, then A is positive definite.

Proof. On friday □

Example 7.20:

Let $a_1, \dots, a_n > 0$, then

$$D = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix},$$

is positive definite, since for $u \in \mathbb{R}^n \setminus \{0\}$, we have

$$u^T D u = a_1 u_1^2 + a_2 u_2^2 + \dots + a_n u_n^2 > 0.$$

Lec 40

Definition 7.21: Positive Definite Hermitian Matrices

Let $A \in M_{n\times n}(\mathbb{C})$ be Hermitian. We say that A is POSITIVE DEFINITE if $\forall u \in \mathbb{C}^n \setminus \{0\}$, $u^T A \overline{u} > 0$.

Claim 7.22:

IF A is Hermitian, then $v^T A \overline{v} \in \mathbb{R}$ for all $v \in \mathbb{C}^n$.

Proof.

$$\begin{aligned} \overline{v^T A \overline{v}} &= \overline{v}^T \overline{A} v = (\overline{v}^T \overline{A} v)^T \\ &= v^T \overline{A}^T \overline{v} \\ &= v^T A \overline{v}. \end{aligned}$$

□

Given a positive definite $A \in M_{n\times n}(\mathbb{C})$, we can define a Hermitian product on $\mathbb{C}^n$ by $\langle u, v\rangle_A = u^T A \overline{v}$.

Claim 7.23:

$\langle \cdot, \cdot\rangle_A$ is a Hermitian product on $\mathbb{C}^n$ iff A is Hermitian and positive definite.

Proof. $\Rightarrow$) Assume $\langle \cdot, \cdot\rangle_A$ is a Hermitian product on $\mathbb{C}^n$. Then, $\forall v, w \in \mathbb{C}^n$, we have

$$\langle v, w\rangle_A = \overline{\langle w, v\rangle_A}.$$

Thus,

$$v^T A \overline{w} = \overline{w^T A \overline{v}} = \overline{w}^T \overline{A} v.$$

Since this is a scalar, we can transpose it and get

$$v^T A \overline{w} = v^T \overline{A}^T \overline{w}.$$

Take $v = e_i$ and $w = e_j$, where e_i is the standard basis of $\mathbb{C}^n$. We get that

$$\begin{aligned} e_i^T A \overline{e_j} &= e_i^T \overline{A}^T \overline{e_j} \\ e_i^T A e_j &= e_i^T \overline{A}^T e_j \\ a_{ij} &= \overline{a_{ji}}. \end{aligned}$$

Thus A is Hermitian

A is also positive definite since for Hermitian products, we must have

$$\langle v, v\rangle_A > 0 \, \forall v \neq 0.$$

$\Leftarrow$) Assume A is positive definite. Then, $\langle \cdot, \cdot\rangle_A$ is linear in the first variable and sesquilinear in the second variable by definition. Since A is Hermitian, we have

$$\langle v, w\rangle_A = v^T A \overline{w} = \overline{w^T A \overline{v}} = \overline{\langle w, v\rangle_A}.$$

Finally, since A is positive definite, we have $\langle v, v\rangle_A > 0$ for all $v \neq 0$. Hence, $\langle \cdot, \cdot\rangle_A$ is a Hermitian product on $\mathbb{C}^n$. □

Exercise 7.24:

Let $A = \begin{pmatrix} a & b \\ \overline{b} & d \end{pmatrix}$. Show that A is positive definite iff $\mathbb{R} \ni a > 0$ and $\det(A) \in \mathbb{R}_{>0}$.

Example 7.25:

Given $V = \mathbb{R}^2$, define two scalar products on V by

$$\langle u, v\rangle_1 = u_1 v_1 + u_2 v_2, \quad \langle u, v\rangle_2 = u^T A v,$$

where $A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$, with $a_1, a_2 > 0$.

In the first case, the unit sphere(circle) is given by

$$\mathcal{S} = \{v \in \mathbb{R}^2 \mid \|v\| = 1\}.$$

In this case, $\mathcal{S}$ looks like a unit circle. In the second case, $\mathcal{S}'$ looks like an ellipse, given by

$$\mathcal{S}' = \{v \in \mathbb{R}^2 \mid a_1 v_1^2 + a_2 v_2^2 = 1\}.$$

Exercise 7.26:

Let $V = C[a,b]$ be the vector space of continuous functions on $[a,b]$. Define

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

Show, that $\langle \cdot, \cdot\rangle$ is a scalar product on V.

Interestingly, according to this scalar product, say on $C[-\pi,\pi]$, $\sin(x)$ and 1 are orthogonal.

7.1 Norms and Angles

We want to define geometric properties on Euclidean and Hermitian spaces, such as angles.

Definition 7.27: Norm

Let V be a vector space over $K = \mathbb{R}$ or $K = \mathbb{C}$. A norm $\|\cdot\|$ on V is a function $V \to \mathbb{R}_{\geq 0}$, denoted by $v \mapsto \|v\|$, that satisfies

1. Triangle inequality: $\|u+v\| \leq \|u\| + \|v\|$.
2. Absolute homogeneity: $\|av\| = |a| \cdot \|v\|$.
3. Non-degeneracy: If $\|v\| = 0$, then $v = 0$.

Claim 7.28:

If $\langle \cdot, \cdot\rangle$ is an inner product on V, then

$$\|v\| = \sqrt{\langle v, v\rangle},$$

is a norm on V.

Proof. Soon ™ □

Exercise 7.29:

Define

$$\| \begin{pmatrix} a \\ b \end{pmatrix} \| := \max\{|a|, |b|\}.$$

And

$$\| \begin{pmatrix} a \\ b \end{pmatrix} \| := |a| + |b|.$$

Show that these are norms and draw the unit sphere in these norms.

Show that none of them does from a scalar product on $\mathbb{R}^2$.

Hint: If $\|\cdot\|$ is induced from a scalar product, then

$$\langle u, v\rangle = \frac{1}{2}(\|u+v\|^2 - \|u\|^2 - \|v\|^2).$$

Theorem 7.30: Cauchy-Schwarz Inequality

Let $(V, \langle \cdot, \cdot\rangle)$ be a Euclidean or Hermitian space. Then, $\forall u, v \in V$, we have

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|.$$

Moreover, inequality holds iff u and v are linearly dependent.

Proof. If $u = 0$, the inequality becomes an obvious equality. Assume thus, $u \neq 0$.

Put $w := v - \lambda u$, where $\lambda = \frac{\langle v, u\rangle}{\|u\|^2}$. Then,

$$\langle w, u\rangle = \langle v, u\rangle - \lambda\langle u, u\rangle = 0.$$

Now,

$$\begin{aligned} 0 \leq \|w\|^2 &= \langle w, w\rangle \\ &= \langle v - \lambda u, v - \lambda u\rangle \\ &= \langle v, v\rangle - \overline{\lambda}\langle v, u\rangle - \lambda\langle u, v\rangle + |\lambda|^2\langle u, u\rangle \\ &= \|v\|^2 - |\lambda|^2\|u\|^2 \\ &= \|v\|^2 - \frac{|\langle v, u\rangle|^2}{\|u\|^2}. \end{aligned}$$

From this, we get that $|\langle v, u\rangle| \leq \|u\| \cdot \|v\|$.

Inequality holds, iff $\|w\|^2 = 0$, iff $w = 0$, iff $v = \lambda u$, iff u and v are linearly dependent. □

Corollary 7.31:

Let $(V, \langle \cdot, \cdot\rangle)$ be a Euclidean or Hermitian space. Then $V \ni v \mapsto \sqrt{\langle v, v\rangle} \in \mathbb{R}_{\geq 0}$ is a norm on V.

Proof. We consider the proof over $\mathbb{C}$. The second and third properties of norms are easy to verify. We will verify the triangle inequality. Let $u, v \in V$. Then,

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + 2\Re(\langle u, v\rangle) + \|v\|^2 \end{aligned}$$

.

Note that $\forall z \in \mathbb{C}$, $\Re(z) \leq |z|$. So,

$$\begin{aligned} \|u+v\|^2 &\leq \|u\|^2 + 2|\langle u, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\| \cdot \|v\| + \|v\|^2 \qquad \text{Cauchy-Schwarz} \\ &= (\|u\| + \|v\|)^2. \end{aligned}$$

Hence, $\|u+v\| \leq \|u\| + \|v\|$. □

Definition 7.32: Angle

Let $(V, \langle \cdot, \cdot\rangle)$ be a Euclidean space. Let $u, v \in V$ nonzero. We define the ANGLE between u and v to be the unique number $\alpha \in [0, \pi]$ such that

$$\cos(\alpha) = \frac{\langle u, v\rangle}{\|u\| \cdot \|v\|} = \langle \frac{u}{\|u\|}, \frac{v}{\|v\|}\rangle.$$

Lec 41

We want to find an orthonormal basis of a Euclidean or Hermitian space. To this extent we use the Gram-Schmidt orthogonalization process.

Theorem 7.33:

Let $(V, \langle \cdot, \cdot\rangle)$ be an inner product space.

1. If $S \subset V$ is an orthogonal system, then S is linearly independent.
2. If $v_1, \ldots, v_n$ is an orthogonal system and $v \in \mathrm{Sp}\{v_1, \ldots, v_n\}$ and we write $v = a_1v_1 + \cdots + a_nv_n$, then
$$a_j = \frac{\langle v, v_j\rangle}{\|v_j\|^2}.$$

Proof. 2. Let $1 \leq j \leq n$. We compute

$$\langle v, v_j\rangle = \sum_{k=1}^n a_k\langle v_k, v_j\rangle = a_j\|v_j\|^2.$$

Thus, $a_j = \frac{\langle v, v_j \rangle}{\|v_j\|^2}$.

1. Suppose $\sum_{j=1}^n a_j v_j = 0$. Then, for $1 \leq j \leq n$, we have by statement 2, that

$$a_j = \frac{\langle 0, v_j \rangle}{\|v_j\|^2} = 0.$$

□

Proposition 7.34:

Let V, W be vector spaces over $\mathbb{R}$ or $\mathbb{C}$ and $\langle \cdot, \cdot \rangle$ an inner product on W. Let $\mathcal{B}$ be a basis for V and $\mathcal{C} = (w_1, \ldots, w_n)$ an orthonormal basis for W. Let $T \in \mathrm{Hom}(V, W)$. Write $[T]_{\mathcal{C}}^{\mathcal{B}} = (a_{ij})$. Then,

$$a_{ij} = \langle T(v_j), w_i \rangle.$$

In case $\mathcal{C}$ is only an orthogonal basis, then

$$a_{ij} = \frac{\langle T(v_j), w_i \rangle}{\|w_i\|^2}.$$

Proof. Exercise. □

Theorem 7.35: Gram-Schmidt Process

Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$. Let $(v_1, \ldots, v_n)$ be a basis of V. Define new vectors $w_1, \ldots, w_n$ by induction as follows:

$$w_1 := v_1$$
$$w_j := v_j - \sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\|w_i\|^2} w_i, \quad 2 \leq j \leq n.$$

Then,

1. $(w_1, \ldots, w_n)$ is an orthogonal basis of V.
2. $(w_1/\|w_1\|, \ldots, w_n/\|w_n\|)$ is an orthonormal basis of V.
3. $\forall 1 \leq j \leq n$, $\mathrm{Sp}\{v_1, \ldots, v_j\} = \mathrm{Sp}\{w_1, \ldots, w_j\}$.

What this process essentially does is filter out the components of v_j in the direction of $w_1, \ldots, w_{j-1}$, so that w_j is orthogonal to $w_1, \ldots, w_{j-1}$.

Proof. We claim that for $1 \leq j \leq n$, w_j is well defined (i.e. $w_i \neq 0$ for $1 \leq i \leq j-1$). Moreover, $w_1, \ldots, w_j$ is an orthogonal system with the same span as $v_1, \ldots, v_j$. We will verify this by induction on j.

Base case ($j = 1$): $w_1 = v_1 \neq 0$. Moreover, w_1 is an orthogonal system and $\mathrm{Sp}\{w_1\} = \mathrm{Sp}\{v_1\}$.

Let $2 \leq j \leq n$ and assume our claim holds for $w_1, \ldots, w_{j-1}$. Then,

$$w_j = v_j - \sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\|w_i\|^2} w_i.$$

By induction hypothesis, $\|w_i\|^2 \neq 0$ for $1 \leq i \leq j-1$. So w_j is well defined. We now show $w_j \neq 0$. Assume $w_j = 0$. Then,

$$v_j = \sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\|w_i\|^2} w_i.$$

So $v_j \in \mathrm{Sp}\{w_1, \ldots, w_{j-1}\}$. By induction hypothesis, we get that $\mathrm{Sp}\{w_1, \ldots, w_{j-1}\} = \mathrm{Sp}\{v_1, \ldots, v_{j-1}\}$. So $v_j \in \mathrm{Sp}\{v_1, \ldots, v_{j-1}\}$, contradicting the fact that $v_1, \ldots, v_n$ is a basis. Hence, $w_j \neq 0$.

We now show orthogonality. Indeed, let $1 \leq k \leq j-1$. We have

$$\langle w_j, w_k \rangle = \langle v_j, w_k \rangle - \sum_{i=1}^{j-1} \frac{\langle v_j, w_i \rangle}{\|w_i\|^2} \langle w_i, w_k \rangle.$$

The only term surviving in the sum is the $i = k$ term, which is $\langle v_j, w_k \rangle$. So

$$\langle w_j, w_k \rangle = \langle v_j, w_k \rangle - \frac{\langle v_j, w_k \rangle}{\cancel{\|w_k\|^2}} \cancel{\langle w_k, w_k \rangle} = 0.$$

Thus, $w_1, \ldots, w_j$ is an orthogonal system.

It remains to show that $\mathrm{Sp}\{w_1, \ldots, w_j\} = \mathrm{Sp}\{v_1, \ldots, v_j\}$. Clearly, $w_j \in \mathrm{Sp}\{w_1, \ldots, w_{j-1}, v_j\}$. By induction hypothesis, we thus get $w_j \in \mathrm{Sp}\{v_1, \ldots, v_j\}$. Hence, $\mathrm{Sp}\{w_1, \ldots, w_j\} \subset \mathrm{Sp}\{v_1, \ldots, v_j\}$.

Since $(w_1, \ldots, w_j)$ is orthogonal, and $(v_1, \ldots, v_j)$ is a j dimensional space, we must have $\dim(\mathrm{Sp}\{w_1, \ldots, w_j\}) = j$. Hence, $\mathrm{Sp}\{w_1, \ldots, w_j\} = \mathrm{Sp}\{v_1, \ldots, v_j\}$. □

Corollary 7.36:

Every finite dimensional inner product space has an orthonormal basis.

7.2 The orthogonal complement

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{R}$ or $\mathbb{C}$.

Definition 7.37:

Let $S \subset V$ be a nonempty subset. Define the ORTHOGONAL COMPLEMENT of S as

$$S^\perp := \{v \in V \mid \langle v, s \rangle = 0 \, \forall s \in S\}.$$

If $S = \{w\}$, we write $w^\perp$ instead of $\{w\}^\perp$.

Lemma 7.38:

Let $S, T \subset V$ be nonempty subsets. Then,

1. $S^\perp$ is a subspace of V.
2. $0^\perp = V$ and $V^\perp = \{0\}$.
3. $S \cap S^\perp = \varnothing$ or $S \cap S^\perp = \{0\}$.
4. If $S \subset T$, then $T^\perp \subset S^\perp$.
5. $(\mathrm{Sp}(S))^\perp = S^\perp$.
6. $S \subset (S^\perp)^\perp$.

Proof. 1. $0 \in S^\perp$ since $\langle 0, s \rangle = 0$ for all $s \in S$. Furthermore, if $v, w \in S^\perp$ and $a, b \in K$, then $\langle av + bw, s \rangle = a\langle v, s \rangle + b\langle w, s \rangle = 0$ for all $s \in S$. Hence, $av + bw \in S^\perp$. Thus, $S^\perp$ is a subspace of V.

2. $0^\perp = \{v \in V \mid \langle v, 0 \rangle = 0\} = V$ since every vector is perpendicular to 0. Let $v \in V^\perp$. Thus $\langle v, v \rangle = 0$. By positivity, we get $v = 0$. Hence, $V^\perp = \{0\}$.

3. If $S \cap S^\perp = \varnothing$, we are done. So assume $S \cap S^\perp \neq \varnothing$. Let $v \in S \cap S^\perp$. Then, $\langle v, v \rangle = 0$. By positivity, we get $v = 0$. Hence, $S \cap S^\perp = \{0\}$.

4. Exercise.

5. We have $S \subset \mathrm{Sp}(s)$, so by 4, we get $(\mathrm{Sp}(S))^\perp \subset S^\perp$. Let $v \in S^\perp$. Let $u \in \mathrm{Sp}(S)$. Write $u = \sum_{i=1}^k a_i u_i$, with $a_i \in K$ and $u_i \in S$. Then,

$$\langle v, u \rangle = \sum_{i=1}^k \overline{a_i} \langle v, u_i \rangle = 0.$$

6. Let $s \in S$. Then $\forall v \in S^\perp$, $\langle v, s \rangle = 0$. Hence, $\langle s, v \rangle = 0$. So $s \perp v$, implying $s \in (S^\perp)^\perp$. □

Theorem 7.39:

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space (not necessarily finite dimensional) and $U \subset V$ a finite dimensional subspace. Then, $U^\perp$ is a complement of U in V, i.e. $V = U + U^\perp$ and $U \cap U^\perp = \{0\}$. Hence, $V \cong U \oplus U^\perp$ canonical.

Proof. By Lemma 7.38, we have $U \cap U^\perp = \{0\}$. Let $r := \dim U$. Pick an orthonormal basis $(e_1, \dots, e_r)$ of U. Let $v \in V$ and define

$$\tilde{P}_U(v) := \sum_{i=1}^{r} \langle v, e_i\rangle e_i.$$

We have $v = \underbrace{\tilde{P}_U(v)}_{\in U} + (v - \tilde{P}_U(v))$. We claim, that $v - \tilde{P}_U(v) \in U^\perp$. Indeed,

$$\langle v - \tilde{P}_U(v), e_i\rangle = \langle v, e_i\rangle - \sum_{k=1}^{r} \langle v, e_k\rangle\langle e_k, e_i\rangle = \langle v, e_i\rangle - \langle v, e_i\rangle = 0.$$

This shows that $v - \tilde{P}_U(v) \in U^\perp$. Hence, $V = U + U^\perp$. □

Definition 7.40:

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and $U \subset V$ a finite dimensional subspace. We have seen that $U^\perp$ is a complement of U in V, so $V \cong U \oplus U^\perp$. Define a map $P_U : V \to U$ by $P_U(v) = u$, where $v = u + w$ with $u \in U$ and $w \in U^\perp$. We call P_U the ORTHOGONAL PROJECTION to U.

Lec 42

Let us do a short digression. Let W be a vector space over any field K and let $W_1 \subset V$ be a subspace. Let $W_2 \subset V$ be a complement. Given $v = w_1 + w_2$ with $w_1 \in W_1$ and $w_2 \in W_2$, we can define a map $P_{W_1} : V \to W_1$ by $P_{W_1}(v) = w_1$.

Claim 7.41:

P_{W_1}, P_{W_2} are linear, $\operatorname{Im} P_{W_1} = W_1$, and $\operatorname{Ker} P_{W_1} = W_2$.

Proof. $\forall v \in W, v - P_{W_1}(v) \in W_2$. So

$$v = P_{W_1}(v) + P_{W_2}(v).$$

□

Proposition 7.42:

P_U satisfies the following properties:

1. P_U is linear.
2. $\operatorname{Im} P_U = U$ and $\operatorname{Ker} P_U = U^\perp$.
3. $\forall v \in V$, $v - P_U(v) \in U^\perp$.
4. $\forall u \in U$, $P_U(u) = u$.
5. If $e_1, \dots, e_r$ is an orthonormal basis of U, then $\forall v \in V$,

$$P_U(v) = \sum_{i=1}^{r} \langle v, e_i\rangle e_i = \tilde{P}_U(v).$$

Proof. 1-4 follow from 7.41. We now verify 5. Let $v \in V$. Write

$$v = \sum_{i=1}^{r} \langle v, e_i\rangle e_i + (v - \sum_{i=1}^{r} \langle v, e_i\rangle e_i).$$

Clearly $\sum_{i=1}^{r} \langle v, e_i\rangle e_i \in U$. We further saw, that $v - \sum_{i=1}^{r} \langle v, e_i\rangle e_i \perp U$. Hence, $P_U(v) = \sum_{i=1}^{r} \langle v, e_i\rangle e_i$.

So by uniqueness of the decomposition $v = u + w$ with $u \in U$ and $w \in U^\perp$, we get that $P_U(v) = \tilde{P}_U(v)$. □

Corollary 7.43:

Let $(V, \langle\cdot,\cdot\rangle)$ be a inner product space. Let $U \subset V$ be a finite dimensional subspace. Then

$$\left(U^\perp\right)^\perp = U.$$

Proof. By Lemma 7.38, we have $U \subset (U^\perp)^\perp$. We thus have to show the other inclusion. Let $v \in (U^\perp)^\perp$. We have that $V = U + U^\perp$ and $U \cap U^\perp = \{0\}$. So we can write $v = u + w$ with $u \in U$ and $w \in U^\perp$.

Since $u \in U \subset (U^\perp)^\perp$, and $v \in (U^\perp)^\perp$, we get that $w = v - u \in (U^\perp)^\perp$. But $w \in U^\perp$ and we know that $U^\perp \cap (U^\perp)^\perp = \{0\}$. Hence, $w = 0$ and $v = u \in U$. □

Corollary 7.44:

If V is a finite dimensional inner product space and $U \subset V$ is a subspace, then $\dim U + \dim U^\perp = \dim V$.

We want to return to matrices, in this case orthogonal (for $\mathbb{R}$) and unitary (for $\mathbb{C}$) matrices.

Definition 7.45: Orthogonal and Unitary Matrices

1. A matrix $A \in M_{n\times n}(\mathbb{R})$ is ORTHOGONAL if it's columns form an orthonormal basis of $\mathbb{R}^n$ with respect to the standard scalar product.

   The set of all orthogonal matrices is denoted by $O(n)$.

2. A matrix $A \in M_{n\times n}(\mathbb{C})$ is UNITARY if it's columns form an orthonormal basis of $\mathbb{C}^n$ with respect to the standard hermitian product.

   The set of all unitary matrices is denoted by $U(n)$.

Lemma 7.46:

$A \in O(n)$ iff $A^TA = I_n$ iff $AA^T = I_n$ iff $A^{-1} = A^T$.

Lemma 7.47:

$A \in U(n)$ iff $A^\dagger A = I_n$ iff $AA^\dagger = I_n$ iff $A^{-1} = A^\dagger$.

Proof. The proof for 7.46 and 7.47 are similar, so we will only verify the second one. Write

$$A = (v_1 \dots v_n).$$

Where $v_k \in C^n_{\text{cols}}, k = 1, \dots, n$. Then,

$$A^T\overline{A} = \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix} \begin{pmatrix} \overline{v_1} & \dots & \overline{v_n} \end{pmatrix} = \begin{pmatrix} v_1^T\overline{v_1} & \dots & v_1^T\overline{v_n} \\ \vdots & \ddots & \vdots \\ v_n^T\overline{v_1} & \dots & v_n^T\overline{v_n} \end{pmatrix}.$$

So the entry i, j is $\langle v_i, v_j\rangle$.

We know that A is unitary iff $v_1, \dots, v_n$ is an orthonormal basis of $\mathbb{C}^n$. This is only iff $\langle v_i, v_j\rangle = \delta_{ij}$, which is only iff $A^T\overline{A} = I_n$. From this, we get the rest from simple algebraic manipulations. □

Corollary 7.48:

1. Let $A \in M_{n\times n}(\mathbb{R})$. Then, $A \in O(n)$ iff $A^T \in O(n)$.
2. Let $A \in M_{n\times n}(\mathbb{C})$. Then, $A \in U(n)$ iff $A^\dagger \in U(n)$.

Thus, we can also see if the rows of A form an orthonormal basis.

7.3 QR Decomposition

We want to decompose matrices as a product of an orthogonal/unitary matrix and an upper triangular matrix. This is called the QR decomposition.

Theorem 7.49: QR Decomposition

1. Let $A \in \mathrm{GL}_n(\mathbb{R})$. Then, there exists $Q \in O(n)$ and an upper triangular matrix $R \in M_{n\times n}(\mathbb{R})$ such that
$$A = QR.$$
2. Let $A \in \mathrm{GL}_n(\mathbb{C})$. Then, there exists $Q \in U(n)$ and an upper triangular matrix $R \in M_{n\times n}(\mathbb{C})$ such that
$$A = QR.$$

Proof. We'll prove both statements together. Write
$$A = (v_1 \dots v_n).$$
Here, $v_k \in K^n_{\text{cols}}$, $k = 1, \dots, n$. Since A is invertible, $v_1, \dots, v_n$ form a basis of K^n. Applying the Gram-Schmidt process to $v_1, \dots, v_n$, we get an orthonormal basis $e_1, \dots, e_n$ of K^n.

Recall that $\forall 1 \le j \le n$,
$$\mathrm{Sp}\{e_1, \dots, e_j\} = \mathrm{Sp}\{v_1, \dots, v_j\}.$$
Furthermore, we have
$$\begin{aligned} v_1 &= \langle v_1, e_1\rangle e_1 \\ v_2 &= \langle v_2, e_1\rangle e_1 + \langle v_2, e_2\rangle e_2 \\ v_j &= \langle v_j, e_1\rangle e_1 + \dots + \langle v_j, e_j\rangle e_j. \end{aligned}$$
We can conveniently write this as
$$(v_1 \dots v_n) = \underbrace{(e_1 \dots e_n)}_{Q} \underbrace{\begin{pmatrix} \langle v_1, e_1\rangle & \langle v_2, e_1\rangle & \dots & \langle v_n, e_1\rangle \\ 0 & \langle v_2, e_2\rangle & \dots & \langle v_n, e_2\rangle \\ 0 & 0 & \dots & \langle v_n, e_3\rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \langle v_n, e_n\rangle \end{pmatrix}}_{R}.$$
Q is orthogonal/unitary since $e_1, \dots, e_n$ is an orthonormal basis. R is upper triangular by construction. Hence, $A = QR$. ☐

Theorem 7.50: Extension of QR Decomposition

Let $A \in M_{m\times n}(K)$ with with $r = \operatorname{rank} A$. Then, $\exists Q \in O(m)$ (if $K = \mathbb{R}$) or $Q \in U(m)$ (if $K = \mathbb{C}$) and
$$R = \begin{pmatrix} C & * \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(K).$$
Where $C \in M_{r\times r}(K)$ is an invertible upper triangular matrix, such that
$$A = QR.$$

Proof. See lecture notes. ☐

Lec 43

7.4 Dual Spaces Reloaded

Let V be a vector space over K. Recall that $V^* := \mathrm{Hom}(V, K)$ is the dual space of V. Assume $V = K^n_{\text{cols}}$ and let $\varphi : K^n \to K$ be a functional. Then, $\exists!$ $A = (a_1, \dots, a_n) \in M_{1\times n}(K)$ such that
$$\varphi(v) = Av.$$
Indeed, if we denote by $e_1, \dots, e_n$ the standard basis of K^n, then we can write $\varphi(e_i) = a_i$ for $1 \le i \le n$. Then, $\varphi(v) = Av$ for all $v \in K^n$.

Define $\phi_u : V \to K$, by $\phi_u(v) = \langle v, u\rangle$. $\phi_u(\cdot) = \langle \cdot, u\rangle$ is called the FUNCTIONAL ASSOCIATED TO u.

Claim 7.51:

ϕ_u is linear. So $\phi_u \in V^*$.

Theorem 7.52:

Let $(V, \langle \cdot,\cdot\rangle)$ be a finite dimensional inner product space. Let $\varphi \in V^*$. Then $\exists! u \in V$ such that $\varphi = \varphi_u$.

In fact, we get a map $\Phi : V \to V^*$, defined by $\Phi(u) = \varphi_u$. For $K = \mathbb{R}$, Φ is an isomorphism. For $K = \mathbb{C}$, Φ is injective and surjective but only complex anti-linear, i.e.
$$\Phi(au + bv) = \bar{a}\Phi(u) + \bar{b}\Phi(v).$$

So at least, if $K = \mathbb{R}$, we get a canonical isomorphism between V and V^*.

Proof. Fix an orthonormal basis $e_1, \dots, e_n$ of V. Let $\varphi \in V^*$ and $v \in V$. We have:
$$v = \sum_{i=1}^n \langle v, e_i\rangle e_i.$$
So applying ϕ to both sides, we get
$$\varphi(v) = \varphi\left(\sum_{i=1}^n \langle v, e_i\rangle e_i\right) = \sum_{i=1}^n \langle v, e_i\rangle \varphi(e_i) = \left\langle v, \sum_{i=1}^n \overline{\varphi(e_i)} e_i \right\rangle.$$
Define $u := \sum_{i=1}^n \overline{\varphi(e_i)} e_i$. Then, $\varphi(v) = \langle v, u\rangle$ for all $v \in V$. Hence, $\varphi = \varphi_u$.

For uniqueness, suppose $\varphi_{u_1} = \varphi_{u_2}$. Then, $\forall v \in V$,
$$\langle v, u_1\rangle = \langle v, u_2\rangle.$$
So $\langle v, u_1 - u_2\rangle = 0$ for all $v \in V$. In particular, $\langle u_1 - u_2, u_1 - u_2\rangle = 0$. By positivity, we get $u_1 - u_2 = 0$. Hence, $u_1 = u_2$. (Lemma 7.3). ☐

Exercise 7.53:

Φ is linear/ complex anti-linear as claimed.

Let V, W be vector spaces over K and $T : V \to W$ linear. We have seen, that T induces a new map $T^* : W^* \to V^*$, called the dual map, defined by
$$T^*(\varphi) = \varphi \circ T.$$
$$\begin{tikzcd} V \arrow[r, "T"] \arrow[dr, "T^*"'] & W \arrow[d, "\varphi"] \\ & W^* \end{tikzcd}.$$

Assume now, that $(V, \langle \cdot,\cdot\rangle_V)$ and $(W, \langle \cdot,\cdot\rangle_W)$ are inner product spaces and $\dim V < \infty$. Let $T : V \to W$ be a linear map. Define $T' : W \to V$ using our "canonical" identifications $W \cong W^*$ and $V \cong V^*$. The corresponding diagram for $T' = \Phi_V^{-1} \circ T^* \circ \Phi_W$ is
$$\begin{tikzcd} W^* \arrow[r, "T^*"] & V^* \arrow[d, "\Phi_V^{-1}"] \\ W \arrow[u, "\Phi_w"] \arrow[r, "T'"] & V \end{tikzcd}.$$

Notice, that since $\dim V < \infty$, Φ_V is bijective and thus Φ_V^{-1} is well defined.

Let us look at some properties of T'. Note that

$$\Phi_V \circ T' = T^* \circ \Phi_W.$$

Thus $\forall w \in W$, $\Phi_V(T'(w)) = T^*(\Phi_W(w))$. So we can write

$$\varphi_{T'(w)} = T^*(\varphi_w) \in V^*.$$

This is the same as saying that $\forall v \in V$,

$$\langle v, T'(w)\rangle_V = \varphi_w(T(v)) = \langle T(v), w\rangle_W.$$

The map T' is characterized by:

$$\langle Tv, w\rangle_W = \langle v, T'(w)\rangle_V \quad \forall v \in V, w \in W.$$

Claim 7.54:
T' is a linear map.

Proof. If $K = \mathbb{R}$, then Φ_V and Φ_W are linear, so $T' = \Phi_V^{-1} \circ T^* \circ \Phi_W$ is linear as well.

If $K = \mathbb{C}$, then Φ_V and Φ_W are both additive and complex anti-linear. As an exercise, verify that if $R : X \to Y$ is a bijective complex anti-linear map, then $R^{-1} : Y \to X$ is also additive and complex anti-linear.

So Φ_V^{-1} and Φ_W are both additive and complex anti-linear. T^* is linear. So

$$\begin{aligned} T'(w_1 + w_2) &= \Phi_V^{-1} \circ T^* \circ \Phi_W(w_1 + w_2) \\ &= \Phi_V^{-1} \circ T^* \circ \Phi_W(w_1) + \Phi_V^{-1} \circ T^* \circ \Phi_W(w_2) \end{aligned}$$

.

So T' is additive. Moreover, for $a \in \mathbb{C}$,

$$\begin{aligned} T'(aw) &= \Phi_V^{-1} \circ T^* \circ \Phi_W(aw) \\ &= \Phi_V^{-1} \circ T^* \circ (\overline{a}\Phi_W(w)) \\ &= \Phi_V^{-1} \circ (\overline{a}T^* \circ \Phi_W(w)) \\ &= a\Phi_V^{-1} \circ T^* \circ \Phi_W(w) = aT'(w). \end{aligned}$$

□

As a notation, we will write T^* instead of T' and call it the ADJOINT of T.

Lemma 7.55:
Let U, V, W be inner product spaces over K. Assume they are all finite dimensional. Let $S, T : V \to W$ and $R : W \to U$ be linear maps. Then,

1. T^* is linear,
2. $(S + T)^* = S^* + T^*$,
3. $(\lambda T)^* = \overline{\lambda} T^*$ for $\lambda \in K$,
4. $(T^*)^* = T$,
5. $(\mathrm{id}_V)^* = \mathrm{id}_V$,
6. $(R \circ T)^* = T^* \circ R^*$.

Proof. 1. Has been proven already.

2. Let $w \in W, v \in V$. Then

$$\begin{aligned} \langle v, (S+T)^*(w)\rangle_V &= \langle (S+T)(v), w\rangle_W \\ &= \langle Sv, w\rangle_W + \langle Tv, w\rangle_W \\ &= \langle v, S^*(w)\rangle_V + \langle v, T^*(w)\rangle_V \\ &= \langle v, S^*(w) + T^*(w)\rangle_V \\ &= \langle v, (S^* + T^*)(w)\rangle_V. \end{aligned}$$

Since this holds $\forall v \in V$, by lemma 7.3, we get that $(S + T)^*(w) = (S^* + T^*)(w)$ for all $w \in W$. Hence, $(S + T)^* = S^* + T^*$.

3. Exercise.

4. Let $v \in V, w \in W$. Then,

$$\begin{aligned} \langle w, (T^*)^*(v)\rangle_W &= \langle T^*(w), v\rangle_V \\ &= \overline{\langle v, T^*(w)\rangle_V} \\ &= \overline{\langle Tv, w\rangle_W} \\ &= \langle w, Tv\rangle_W. \end{aligned}$$

Again, by lemma 7.3, we get that $(T^*)^*(v) = Tv$ for all $v \in V$. Hence, $(T^*)^* = T$.

5. Exercise

6. Let $v \in V, u \in U$. Then,

$$\begin{aligned} \langle v, (R \circ T)^* u\rangle &= \langle (R \circ T)(v), u\rangle_U \\ &= \langle R(T(v)), u\rangle_U \\ &= \langle T(v), R^*(u)\rangle_W \\ &= \langle v, T^*(R^*(u))\rangle_V. \end{aligned}$$

Once again, by lemma 7.3, we get that $(R \circ T)^* u = T^*(R^*(u))$ for all $u \in U$. Hence, $(R \circ T)^* = T^* \circ R^*$. □

Lemma 7.56:
Let V, W be finite dimensional inner product spaces over K. Let $T : V \to W$ be linear. Consider $T^* : W \to V$ the adjoint of T. Then,

1. $\mathrm{Ker}(T^*) = (\mathrm{Im}\, T)^\perp$,
2. $\mathrm{Im}(T^*) = (\mathrm{Ker}\, T)^\perp$.
3. $\mathrm{Ker}(T) = (\mathrm{Im}\, T^*)^\perp$,
4. $\mathrm{Im}(T) = (\mathrm{Ker}\, T^*)^\perp$.

Proof. Let $w \in W$. Then, $w \in \mathrm{Ker}\, T^*$ iff $T^* w = 0$ iff $\langle v, T^* w\rangle_V = 0$ for all $v \in V$ iff $\langle Tv, w\rangle_W = 0$ for all $v \in V$ iff $w \perp Tv$ for all $v \in V$ iff $w \perp \mathrm{Im}\, T$ iff $w \in (\mathrm{Im}\, T)^\perp$. Hence, $\mathrm{Ker}(T^*) = (\mathrm{Im}\, T)^\perp$. This proves 1. The rest of the statements follow from 1 and the fact that $(T^*)^* = T$ and $(S^\perp)^\perp = S$ for any subspace S of a finite dimensional inner product space. □ Lec 44

We want to find the matrix representation of adjoint maps.

Proposition 7.57:
Let V, W be finite dimensional inner product spaces and $T : V \to W$ linear. Let $\mathcal{B} = (e_1, \ldots, e_n)$ be an orthonormal basis from V, and $\mathcal{C} = (f_1, \ldots, f_m)$ be an orthonormal basis from W. Then

$$[T^*]_{\mathcal{B}}^{\mathcal{C}} = \overline{[T]_{\mathcal{C}}^{\mathcal{B}}}^T = \left([T]_{\mathcal{C}}^{\mathcal{B}}\right)^\dagger.$$

Proof. Write $A = [T]_{\mathcal{C}}^{\mathcal{B}} = (a_{ij})$ and $B = [T^*]_{\mathcal{B}}^{\mathcal{C}} = (b_{ij})$.

We have

$$b_{ij} = \langle T^* f_j, e_i\rangle_V = \overline{\langle e_i, T^* f_j\rangle_V} = \overline{\langle Te_i, f_j\rangle_W} = \overline{a_{ji}}.$$

□

Corollary 7.58:
Let $A \in M_{m\times n}(K)$ and $T_A : K^n \to K^m$ be the linear map defined by $T_A(v) = Av$. Then, if we endow K^n and K^m with the standard inner product, we have $(T_A)^* = T_{A^\dagger}$.

For the record, some properties of the adjoint matrix:

1. $(A + B)^\dagger = A^\dagger + B^\dagger$,
2. $(\lambda A)^\dagger = \overline{\lambda} A^\dagger$ for $\lambda \in K$,

3. $(A^\dagger)^\dagger = A$,

4. $(I_n)^\dagger = I_n$,

5. $(AB)^\dagger = B^\dagger A^\dagger$.

7.5 The spectral theorem

We want to see when we can diagonalize matrices.

Definition 7.59: Orthogonal Diagonalizability

Let V be an inner product space and $T \in \mathrm{End}(V)$. We say that T is ORTHOGONALLY DIAGONALIZABLE if there exists an orthonormal basis $\mathcal{B}$ of V consisting of eigenvectors of T.

Lemma 7.60:

If T is orthogonally diagonalizable, then $T \cdot T^* = T^* \cdot T$.

Proof. Let B be an orthonormal basis for V such that all vectors in B are eigenvectors of T. Then, $[T]_B^B$ is a diagonal matrix. By proposition 7.57, $[T^*]_B^B$ is the conjugate transpose of $[T]_B^B$, which is also a diagonal matrix. Hence, $[T]_B^B$ and $[T^*]_B^B$ commute, so $T \cdot T^* = T^* \cdot T$. □

Definition 7.61: Normal Endomorphisms / Matrices

Let V be a finite dimensional inner product space. $T \in \mathrm{End}(V)$ is called NORMAL if $T \cdot T^* = T^* \cdot T$.

A matrix $A \in M_{n\times n}(K)$ is called NORMAL if

$$A \cdot A^\dagger = A^\dagger \cdot A.$$

So if T is orthogonally diagonalizable, then T is normal. Is the converse true? NO! Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In that case $A \cdot A^\dagger = A^\dagger \cdot A$, but A has no eigenvalues at all.

Theorem 7.62: Spectral Theorem

Let V be a finite dimensional inner product space over $\mathbb{C}$ and $T \in \mathrm{End}(V)$. Then, T is normal iff T is orthogonally diagonalizable.

Lemma 7.63:

Let V be a finite dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$ and $T \in \mathrm{End}(V)$ be normal. Then

1. $\forall v \in V$, $\|Tv\| = \|T^*v\|$.
2. $\forall \lambda \in K$, the endomorphism $T - \lambda\,\mathrm{id}_V$ is normal.
3. If $v \in V$ is an eigenvector of T with eigenvalue λ, then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.
4. If v_1, v_2 are eigenvectors of T with distinct eigenvalues, then $v_1 \perp v_2$.

Proof. 1. We compute $\|Tv\|^2$ and $\|T^*v\|^2$:

$$\begin{aligned}\|Tv\|^2 &= \langle Tv, Tv\rangle = \langle v, T^* \circ Tv\rangle \\ &= \langle v, T \circ T^* v\rangle = \langle T^*v, T^*v\rangle = \|T^*v\|^2.\end{aligned}$$

By positivity, we get $\|Tv\| = \|T^*v\|$.

2. We compute $(T - \lambda\,\mathrm{id}_V) \cdot (T - \lambda\,\mathrm{id}_V)^*$ and find

$$\begin{aligned}(T - \lambda\,\mathrm{id}_V) \cdot (T - \lambda\,\mathrm{id}_V)^* &= (T - \lambda\,\mathrm{id}_V) \cdot (T^* - \overline{\lambda}\,\mathrm{id}_V) \\ &= T \cdot T^* - \overline{\lambda}T - \lambda T^* + |\lambda|^2\,\mathrm{id}_V \\ &= T^* \cdot T - \lambda T^* - \overline{\lambda}T + |\lambda|^2\,\mathrm{id}_V \\ &= (T^* - \overline{\lambda}\,\mathrm{id}_V) \cdot (T - \lambda\,\mathrm{id}_V) \\ &= (T - \lambda\,\mathrm{id}_V)^* \cdot (T - \lambda\,\mathrm{id}_V).\end{aligned}$$

3. If v is an eigenvector of T with eigenvalue λ, then

$$(T - \lambda\,\mathrm{id}_V)(v) = 0.$$

By 2. $T - \lambda\,\mathrm{id}_V$ is normal. By 1. $\|(T - \lambda\,\mathrm{id}_V)(v)\| = \|(T - \lambda\,\mathrm{id}_V)^*(v)\|$. So

$$\|(T - \lambda\,\mathrm{id}_V)^*(v)\| = 0.$$

By positivity, we get $(T - \lambda\,\mathrm{id}_V)^*(v) = 0$. So $T^*v = \overline{\lambda}v$.

4. We compute

$$\begin{aligned}\lambda_1\langle v_1, v_2\rangle &= \langle \lambda_1 v_1, v_2\rangle = \langle Tv_1, v_2\rangle \\ &= \langle v_1, T^*v_2\rangle = \langle v_1, \overline{\lambda}_2 v_2\rangle = \lambda_2\langle v_1, v_2\rangle.\end{aligned}$$

It follows, that $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0$. Since $\lambda_1 \neq \lambda_2$, we get $\langle v_1, v_2\rangle = 0$. Hence, $v_1 \perp v_2$. □

Proof. [of theorem 7.62] $\Rightarrow$: Follows from lemma 7.60.

$\Leftarrow$: Since T is complex, by theorem 6.31 T is trigonalizable. Thus, $\exists$ a basis $\mathcal{C}$ for V such that $[T]_\mathcal{C}^\mathcal{C}$ is upper triangular. Applying Gram-Schmidt to $\mathcal{C}$, we get an orthonormal basis $\mathcal{B} = (v_1, \ldots, v_n)$ for V which is orthonormal.

Note that because of the Gram-Schmidt process, $[T]_\mathcal{B}^\mathcal{B}$ is also upper triangular.[4] We'll show now, that because of normality, $[T]_\mathcal{B}^\mathcal{B}$ is actually diagonal. Write

$$[T]_\mathcal{B}^\mathcal{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

Since B is orthonormal, we know that

$$\|Tv_1\|^2 = |a_{11}|^2. \tag{7.1}$$

Since $[T^*]_\mathcal{B}^\mathcal{B}$ is the conjugate transpose of $[T]_\mathcal{B}^\mathcal{B}$, we have

$$[T^*]_\mathcal{B}^\mathcal{B} = \begin{pmatrix} \overline{a_{11}} & 0 & \cdots & 0 \\ \overline{a_{12}} & \overline{a_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & \overline{a_{nn}} \end{pmatrix}.$$

So we get

$$\|T^*v_1\|^2 = |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2.$$

Combining this with Equation (7.1), we get $|a_{12}|^2 + \cdots + |a_{1n}|^2 = 0$. By positivity, we get $a_{12} = \cdots = a_{1n} = 0$. So the first row of $[T]_\mathcal{B}^\mathcal{B}$ has only one non-zero entry. Inductively applying this argument, we get the result. □ Lec 45

Corollary 7.64:

Let V be a finite dimensional inner product space over $\mathbb{R}$. Let $T \in \mathrm{End}(V)$ be normal. If T is trigonalizable, then T is orthogonally diagonalizable.

As a reminder, trigonalizabilty holds iff $p_T(x)$ splits into linear factors over $\mathbb{R}$, which is not always the case. So the spectral theorem does not hold for all normal endomorphisms over $\mathbb{R}$.

[4] Since the span of the first j vectors of $\mathcal{B}$ is the same as the span of the first j vectors of $\mathcal{C}$.

Lemma 7.65:

Let V be a finite dimensional inner product space over $\mathbb{R}$ and $T \in \mathrm{End}(V)$. If T is orthogonally diagonalizable, then T is self-adjoint, i.e. $T = T^*$.

Proof. Let B an orthonormal basis for V such that $[T]_B^B$ is diagonal. Then, $[T^*]_B^B$ is the conjugate transpose of $[T]_B^B$. But since the entries are real, the conjugate transpose is just the transpose, which is the same as the original matrix. Hence, $[T^*]_B^B = [T]_B^B$. □

Definition 7.66:

Let V be an inner product space over $\mathbb{R}$ or $\mathbb{C}$ and $T \in \mathrm{End}(V)$. T is called SELF-ADJOINT if $T = T^*$.

A matrix $A \in M_{n\times n}(K)$ is called SELF-ADJOINT if $A = A^\dagger$. (So over $\mathbb{R}$, A is self-adjoint iff A is symmetric).

Exercise 7.67:

Show that $T : V \to V$ is self adjoint iff $\forall$ orthonormal basis $\mathcal{B}$ of V, $[T]_\mathcal{B}^\mathcal{B}$ is self-adjoint.

Theorem 7.68:

Let V be a finite dimensional inner product space over $\mathbb{R}$. Then T is orthogonally diagonalizable iff T is self-adjoint.

Lemma 7.69:

Let V be an inner product space over $\mathbb{R}$ or $\mathbb{C}$ and let $T \in \mathrm{End}(V)$ be self-adjoint. Then:

1) All eigenvalues of T are real.

2) $p_T(x)$ factors into linear factors over K.

Proof. 1. Obvious for $K = \mathbb{R}$. Assume $K = \mathbb{C}$ and let $\lambda \in \mathbb{C}$ be an eigenvalue of T with eigenvector v. Then, by Lemma 7.63, v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$. Since $T = T^*$, we get $\lambda = \overline{\lambda}$. Hence, $\lambda \in \mathbb{R}$.

2. In $\mathbb{C}$, this follows from the fundamental theorem of algebra. So assume $K = \mathbb{R}$. Let $\mathcal{B}$ be an orthonormal basis for V. Cosnider $A := [T]_\mathcal{B}^\mathcal{B} \in M_{n\times n}(\mathbb{R})$. We know that $p_A(x) = p_T(x)$. We'll show that $p_A(x)$ factors into linear factors over K.

We have $T^* = T$ so $[T^*]_\mathcal{B}^\mathcal{B} = [T]_\mathcal{B}^\mathcal{B}$. By proposition 7.57, $[T^*]_\mathcal{B}^\mathcal{B}$ is the conjugate transpose of $[T]_\mathcal{B}^\mathcal{B}$. So $A = A^T$. Consider $\mathbb{C}^n$ with its standard inner product (Hermitian product) and consider A now as a complex matrix.

Let $T_A : \mathbb{C}^n \to \mathbb{C}^n$ be the linear map defined by $T_A(v) = Av$, and let $\mathcal{E} = (e_1, \ldots, e_n)$ be the standard basis for $\mathbb{C}^n$. Then, $\mathcal{E}$ is orthonormal. We also have

$$[T_A]_\mathcal{E}^\mathcal{E} = A.$$

Now, $p_{T_A}(x) = p_A(x)$, also when we view it as a polynomial over $\mathbb{C}$. Now, T_A is self-adjoint when considered as an endomorphism of $\mathbb{C}^n$. Thus, by the fundamental theorem of algebra, $p_{T_A}(x)$ factors into linear factors over $\mathbb{C}$,

$$p_A(x) = \prod_{k=1}^{n} (x - \lambda_k) \in \mathbb{C}[x].$$

By 1. $\lambda_k \in \mathbb{R}$ for all k. Hence, $p_A(x)$ factors into linear factors over $\mathbb{R}$ as well. □

Proof. [Theorem 7.68] $\Rightarrow$: Follows from lemma 7.65.

$\Leftarrow$: Follows from lemma 7.69 and corollary 7.64. □

Let's extend this to matrices. A matrix $A \in M_{n\times n}(K)$ is called normal, if $A \cdot A^\dagger = A^\dagger \cdot A$.

Example 7.70:

1. For $K = \mathbb{R}$, symmetric matrices are normal. 2. For $K = \mathbb{R}$, orthogonal matrices are normal. 3. For $K = \mathbb{C}$, self-adjoint matrices are normal. 4. For $K = \mathbb{C}$, unitary matrices are normal

Theorem 7.71:

1. If $A \in M_{n\times n}(\mathbb{C})$ is normal, then $\exists$ a unitary matrix U such that $U^{-1}AU$ is diagonal.

2. If $A \in M_{n\times n}(\mathbb{R})$ is symmetric, $\exists$ an orthogonal matrix O such that $O^{-1}AO$ is diagonal.

Proof. Follows from the spectral theorems. □

7.6 Isometries

We want to focus a bit more on the geometry of inner product spaces.

Definition 7.72:

Let V, W be inner product spaces over $K = \mathbb{R}$ or $\mathbb{C}$. A linear map $T : V \to W$ is called an LINEAR ISOMETRY if $\forall v_1, v_2 \in V$,

$$\langle Tv_1, Tv_2\rangle_W = \langle v_1, v_2\rangle_V.$$

Remark, that if T is a linear isometry, then

1. $\forall v \in V$, $\|Tv\| = \|v\|$.

2. $\forall p, q \in V$, $d_w(Tp, Tq) = d_V(p, q)$, where d_V and d_W are the distance functions induced by the inner products on V and W respectively.

3. T is injective. Indeed, if $v \in \mathrm{Ker}\, T$, then $\|v\| = \|Tv\| = 0$. By positivity, we get $v = 0$.

Theorem 7.73:

Let V be a finite dimensional inner product space over $K = \mathbb{R}$ or $\mathbb{C}$ and $T \in \mathrm{End}(V)$. Then, the following are equivalent:

1. T is an isometry.

2. $\forall v \in V$, $\|Tv\| = \|v\|$.

3. $\forall$ orthonormal system of vectors $(e_1, \ldots, e_k)$ in V, $(Te_1, \ldots, Te_k)$ is also an orthonormal system of vectors in V.

4. $\exists$ an orthonormal basis $e_1, \ldots, e_n$ of V such that $(Te_1, \ldots, Te_n)$ is also an orthonormal basis of V.

5. $T \circ T^* = \mathrm{id}_V$.

6. $T^* \circ T = \mathrm{id}_V$.

7. $T^* = T^{-1}$.

8. T^* is an isometry.

Proof. We only proof some.

$2 \Rightarrow 1$: Recall that
$$\langle v, w\rangle = \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right).$$
So,
$$\begin{aligned}\langle Tv, Tw\rangle &= \frac{1}{2}\left(\|Tv+Tw\|^2 - \|Tv\|^2 - \|Tw\|^2\right)\\ &= \frac{1}{2}\left(\|T(v+w)\|^2 - \|Tv\|^2 - \|Tw\|^2\right)\\ &= \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right)\\ &= \langle v, w\rangle.\end{aligned}$$

$1 \Rightarrow 3$: Let $(e_1, \dots, e_k)$ be an orthonormal system of vectors in V. Then, $\forall 1 \le i, j \le k$,
$$\langle Te_i, Te_j\rangle = \langle e_i, e_j\rangle = \delta_{ij}.$$
So $(Te_1, \dots, Te_k)$ is also an orthonormal system of vectors in V.

$4 \Rightarrow 5$: Let $e_1, \dots, e_n$ be an orthonormal basis of V such that $(Te_1, \dots, Te_n)$ is also an orthonormal basis of V. Let $1 \le j \le n$. Then $\forall 1 \le i \le n$,
$$\langle e_i, e_j\rangle = \delta_{ij} = \langle Te_i, Te_j\rangle = \langle e_i, T^* \circ Te_j\rangle.$$
Thus,
$$\langle e_i, e_j - T^* \circ Te_j\rangle = 0 \quad \forall 1 \le i \le n.$$
So $T^* \circ Te_j = e_j$ for all $1 \le j \le n$. Hence, $T^* \circ T = \mathrm{id}_V$.

□

Lets go back to matrices. Recall that A is orthogonal if $A \cdot A^T = I_n$, and A is unitary if $A \cdot A^\dagger = I_n$. So We denote orthogonal matrices by $O(n)$ and unitary matrices by $U(n)$.

Proposition 7.74:

Let V be a finite dimensional inner product space over $\mathbb{R}$ (or $\mathbb{C}$). Let $\mathcal{B}$ be an orthonormal basis for V and $T \in \mathrm{End}(V)$. Then, T is an isometry iff $[T]^{\mathcal{B}}_{\mathcal{B}}$ is an orthogonal (or unitary) matrix.

Proof. Exercise. □

Remark 7.75:

If $A \in O(n)$, then $\det A = \pm 1$. If $A \in U(n)$, then $|\det A| = 1$.

Lec 46

Note, that $\forall A \in O(n), \det(A) = \pm 1$. Indeed,
$$1 = \det(I_n) = \det(A^T A) = \det(A^T)\det(A) = (\det A)^2.$$
Similarly, if $A \in U(n)$, then $|\det A| = 1$. Indeed,
$$1 = \det(I_n) = \det(A^\dagger)\det(A) = \overline{\det A}\det A = |\det A|^2.$$

Definition 7.76: Special Groups

We define the SPECIAL ORTHOGONAL GROUP as
$$SO(n) := \{A \in O(n) \mid \det A = 1\}.$$
Further, we define the SPECIAL UNITARY GROUP as
$$SU(n) := \{A \in U(n) \mid \det A = 1\}.$$

Note that theese groups are not closed under addition, so they are not vector spaces. Nonetheless we have the following result.

Proposition 7.77:

Let $G = O(n), U(n), SO(n)$ or $SU(n)$. Then, $I_n \in G$, for all $A, B \in G$, $AB \in G$, and for all $A \in G$, $A^{-1} \in G$.

We now want to classify the elements of $O(n)$ in the case $n = 2$ and $n = 3$.

Lemma 7.78: Eigenvalues of Orthogonal Matrices

Let $A \in O(n)$. If $\lambda \in \mathbb{R}$ is an eigenvalue of A, then $\lambda = \pm 1$.

Proof. Suppose $Av = \lambda v$ where $0 \neq v \in \mathbb{R}^n$. Since A is orthogonal, we have
$$\langle v, v\rangle = \langle Av, Av\rangle = \langle \lambda v, \lambda v\rangle = \lambda^2\langle v, v\rangle.$$
But $\langle v, v\rangle > 0$ since $v \neq 0$. Hence, $\lambda^2 = 1$ and thus the result follows. □

Proposition 7.79: Geometric Description of $O(2)$

Let $A \in O(2)$. If $\det A = 1$, then $\exists \theta \in [0, 2\pi)$ such that
$$A = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} =: R_\theta.$$
The map T_A is an anticlockwise rotation of $\mathbb{R}^2$ by θ around the origin.

If $\det A = -1$, then $\exists$ an orthonormal basis $B = (v_1, v_2)$ for $\mathbb{R}^2$ such that $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ has the following matrix representation with respect to B:
$$[T_A]^B_B = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$
In other words, T_A is a reflection about the line spanned by v_1.

Moreover, $\exists \theta \in [0, 2\pi)$ such that $A = R_\theta \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$.

The basis v_1, v_2 is given by
$$v_1 = \begin{pmatrix}\cos\frac{\theta}{2}\\ \sin\frac{\theta}{2}\end{pmatrix}, v_2 = \begin{pmatrix}\cos(\frac{\theta}{2} + \frac{\pi}{2})\\ \sin(\frac{\theta}{2} + \frac{\pi}{2})\end{pmatrix}.$$

Proof. Note, that if $v \in \mathbb{R}^2$ has norm 1, then $\exists \theta \in [0, 2\pi)$ such that $v = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}$.

Let $A = \begin{pmatrix}u_1 & u_2\end{pmatrix} \in O(2)$. Since $\|u_1\| = 1$, we can write $u_1 = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}$ for some $\theta \in [0, 2\pi)$. Since u_2 is orthogonal to u_1 and has norm 1, we have $u_2 = \begin{pmatrix}\cos(\theta + \frac{\pi}{2})\\ \sin(\theta + \frac{\pi}{2})\end{pmatrix}$.

But inside $\mathrm{Sp}(u_2)$ there exist precisely two unit vectors, one of them is u_2 and the other is $-u_2$. So either $u_2 = \begin{pmatrix}\cos(\theta + \frac{\pi}{2})\\ \sin(\theta + \frac{\pi}{2})\end{pmatrix}$ or $u_2 = \begin{pmatrix}\cos(\theta - \frac{\pi}{2})\\ \sin(\theta - \frac{\pi}{2})\end{pmatrix}$.

In the first case, we have $\det A = 1$ and $A = R_\theta$. In the second case, by direct calculation, we have $\det A = -1$ and $A = R_\theta\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$. □

The summary is, that $SO(2)$ is just rotations and $O(2) \setminus SO(2)$ is just reflections.

Corollary 7.80:

Let V be a 2-dimensional euclidean space and $T : V \to V$ a linear isometry. Then, $\det T = \pm 1$. Moreover, $\exists$ an orthonormal basis $\mathcal{B}$ of V such that if $\det T = 1$, then
$$[T]_{\mathcal{B}}^{\mathcal{B}} = R_\theta.$$
If $\det T = -1$, then
$$[T]_{\mathcal{B}}^{\mathcal{B}} = R_\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Proposition 7.81: Geometric Description of $SO(3)$

Let $A \in SO(3)$. Then 1 is an eigenvalue of A. Moreover, $\exists$ an orthonormal basis $\mathcal{B} = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ and an angle $\theta \in [0, 2\pi)$ such that
$$[T_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$
So T_A is a rotation with angle θ about the axis spanned by v_1.

Moreover, if -1 is an eigenvalue of A, then $\theta = \pi$. In this case
$$[T_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$
Also, if -1 is not an eigenvalue of A, then $\theta \neq \pi$.

Proof. The characteristic polynomial $p_A(x)$ of A has degree 3 and real coefficients. So $p_A(x)$ has a real root λ. By lemma 7.78, $\lambda = \pm 1$. Take $v_1 \in \mathbb{R}^3$ an eigenvector of λ. We normalize v_1 to get a unit eigenvector.

Consider the subspace
$$L := v_1^\perp = \{v \in \mathbb{R}^3 \mid \langle v, v_1 \rangle = 0\}.$$
Since A is orthogonal, and $A(\mathrm{Sp}(v_1)) = \mathrm{Sp}(v_1)$, we get that $A(L) = L$.

Indeed, if $v \in L$, then
$$0 = \langle v, v_1 \rangle = \langle Av, Av_1 \rangle = \langle Av, \lambda v_1 \rangle = \pm \langle Av, v_1 \rangle = 0.$$
Since this is an orthogonal complement, we also have that
$$\mathbb{R}^3 = \mathrm{Sp}(v_1) + L, \quad \mathrm{Sp}(v_1) \cap L = \{0\}.$$
Consider now $T_A|_L : L \to L$. Then L is a 2-dimensional euclidean space and $T_A|_L$ is a linear isometry. By corollary 7.80, $\exists$ an orthonormal basis $\mathcal{B}' = (v_2, v_3)$ of L and $\theta \in [0, 2\pi)$ such that
$$[T_A|_L]_{\mathcal{B}'}^{\mathcal{B}'} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$
Notice, that if $\lambda = 1$, then we need that $\det Q = 1$ so $Q \in SO(2)$.

If $\lambda = -1$, then we need that $\det Q = -1$ so $Q \in O(2) \setminus SO(2)$. But then $\theta = \pi$ and
$$[T_A|_L]_{\mathcal{B}'}^{\mathcal{B}'} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
So choosing $\mathcal{B} = (v_2, v_1, v_3)$, we get the desired result. □

Proposition 7.82:

Let $A \in O(3) \setminus SO(3)$. Then, $\exists$ an orthonormal basis $\mathcal{B} = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ such that
$$[T_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$
So T_A does a rotation by θ around $\mathrm{Sp}(v_1)$ and then a reflection about the plane $v_1^\perp$.

Proof. Let $A' = -A$. Then, $A' \in SO(3)$. By proposition 7.81, $\exists$ an orthonormal basis $\mathcal{B} = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ and $\theta \in [0, 2\pi)$ such that
$$[T_{A'}]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$
But $A = -A'$. So
$$[T_A]_{\mathcal{B}}^{\mathcal{B}} = -[T_{A'}]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -\cos\theta & \sin\theta \\ 0 & -\sin\theta & -\cos\theta \end{pmatrix}.$$
So replacing θ by $\theta + \pi$, we get the desired result. □

7.7 Bilinear Forms

Lec 47

We want to generalize the notion of inner products by dropping the positivity condition.

Definition 7.83: Bilinear Forms

Let V be a vector space over K. A function $B : V \times V \to K$ is called a BILINEAR FORM if the following conditions are satisfied:

1. $\forall w \in V$, the map $B(\cdot, w) : V \to K$ is linear.
2. $\forall v \in V$, the map $B(v, \cdot) : V \to K$ is linear.

Lemma 7.84:

Let V be a vector space over K and $\mathcal{C} = (e_1, \ldots, e_n)$ a basis of V and B a bilinear form on V. Then, $\exists$ a matrix $[B]_{\mathcal{C}} \in M_{n\times n}(K)$ such that
$$B\left(\sum_{i=1}^n c_i e_i, \sum_{j=1}^n d_j e_j\right) = \begin{pmatrix} c_1 & \ldots & c_n \end{pmatrix} [B]_{\mathcal{C}} \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}.$$

Proof. Define $[B]_{\mathcal{C}} := (a_{ij})$ where $a_{ij} = B(e_i, e_j)$. Then,
$$\begin{aligned} B\left(\sum_{i=1}^n c_i e_i, \sum_{j=1}^n d_j e_j\right) &= \sum_{i=1}^n c_i B(e_i, \sum_{j=1}^n d_j e_j) \\ &= \sum_{i=1}^n c_i \sum_{j=1}^n d_j B(e_i, e_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i d_j a_{ij} \\ &= \begin{pmatrix} c_1 & \ldots & c_n \end{pmatrix} [B]_{\mathcal{C}} \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}. \end{aligned}$$
□

Recall that if $T : V \to V$ is linear, and $\mathcal{C}$ is a basis of V, then we have
$$[T]_{\mathcal{C}}^{\mathcal{C}} \in M_{n\times n}(K).$$

If $\mathcal{D}$ is another basis of V, then we have

$$[T]_{\mathcal{D}}^{\mathcal{D}} = [\mathrm{id}_V]_{\mathcal{D}}^{\mathcal{C}}[T]_{\mathcal{C}}^{\mathcal{C}}[\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}.$$

For Bilinear forms, this story is a bit different.

Lemma 7.85:

Let V be a finite dimensional vector space over K. Let B be a bilinear form and $\mathcal{C}, \mathcal{D}$ be two bases of V. Then,

$$[B]_{\mathcal{D}} = \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)^T [B]_{\mathcal{C}}[\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}.$$

Proof. Let $\mathcal{C} = (v_1, \dots, v_n)$ and $\mathcal{D} = (w_1, \dots, w_n)$. Then,

$$([B]_{\mathcal{C}})_{j,k} = B(v_j, v_k), \quad ([B]_{\mathcal{D}})_{j,k} = B(w_j, w_k).$$

Recall also, that column j of $[\mathrm{id}_V]_{\mathcal{D}}^{\mathcal{C}}$ is the coordinate vector of w_j with respect to $\mathcal{C}$.

$$w_j = \sum_{p=1}^{n} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{p,j} v_p.$$

Thus,

$$\begin{aligned}
([B]_{\mathcal{D}})_{j,k} &= B(w_j, w_k) \\
&= B\left(\sum_{p=1}^{n} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{p,j} v_p, \sum_{q=1}^{n} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{q,k} v_q\right) \\
&= \sum_{p=1}^{n}\sum_{q=1}^{n} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{p,j} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{q,k} B(v_p, v_q) \\
&= \sum_{p=1}^{n}\sum_{q=1}^{n} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{p,j} \left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{q,k} ([B]_{\mathcal{C}})_{p,q} \\
&= \left(\left([\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)^T [B]_{\mathcal{C}}[\mathrm{id}_V]_{\mathcal{C}}^{\mathcal{D}}\right)_{j,k}.
\end{aligned}$$

□

Definition 7.86: Quadratic Forms

1. A bilinear form $B : V \times V \to K$ is called SYMMETRIC if $\forall v, w \in V$, $B(v, w) = B(w, v)$.

2. Assume $K = \mathbb{R}$. A bilinear form $B : V \times V \to \mathbb{R}$ is called POSITIVE DEFINITE if $\forall v \in V \setminus \{0\}$, $B(v, v) > 0$.

3. Define $q_B : V \to K$ by $q_B(v) = B(v, v)$. We call q_B the QUADRATIC FORM associated with B.

Example 7.87:

1. If $(V, \langle\cdot,\cdot\rangle)$ is an Euclidean space, then $B(v, w) = \langle v, w\rangle$ is a symmetric positive definite bilinear form. It's associated quadratic form is $q(v) = \|v\|^2$.

2. Let $B \in M_{n\times n}(K)$. Then $B : K^n \times K^n \to K$ defined by

$$B(v, w) = v^T Bw.$$

If $B^T = B$, then B is symmetric. If B is positive definite, then B is positive definite as a bilinear form.

Theorem 7.88: Sylvester's inertia theorem

Let $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form on a vector space V over $\mathbb{R}$ with dimension n. Then, $\exists$ a basis $\mathcal{C}$ for V such that

$$[B]_{\mathcal{C}} = \begin{bmatrix} I_k & 0 & 0 \\ 0 & -I_l & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

for some $k, l \geq 0$ such that $k + l \leq n$. Moreover, k and l do not depend on the choice of $\mathcal{C}$. In fact,

$$k = \max\{\dim W | W \leq V \text{s.t.} B|_{W\times W} \text{ is pos. def. }\}.$$

$$l = \max\{\dim W | W \leq V \text{s.t.} - B|_{W\times W} \text{ is pos. def. }\}.$$

$$n - k - l = \max\{\dim Z | Z \leq V \text{s.t.} B|_{Z\times V} \equiv 0\}.$$

k is called the POSITIVITY INDEX of B, l is called the NEGATIVITY INDEX of B, and $n - k - l$ is called the NULLITY of B.

We further define the SIGNATURE of B as $\sigma(B) := k - l$.

Proof. Choose any basis $\mathcal{C}'$ for V. Let

$$A := [B]_{\mathcal{C}'} \in M_{n\times n}(\mathbb{R}).$$

Note, that A is symmetric since B is symmetric. By the spectral theorem, A is an orthogonally diagonalizable matrix. So $\exists P \in O(n)$ such that

$$P^{-1}AP = \text{ diagonal.}$$

But $P^{-1} = P^T$. Thus, P^TAP is diagonal. Let $\mathcal{C}''$ be the (unique) basis of V for which $[\mathrm{id}_V]_{\mathcal{C}'}^{\mathcal{C}''} = P$. So we have

$$[B]_{\mathcal{C}''} = P^TAP = \begin{pmatrix} \eta_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \eta_n \end{pmatrix},$$

where $\eta_j \in \mathbb{R}$. By changing the order of the elements of $\mathcal{C}''$, we can assume that $\eta_1, \dots, \eta_k > 0$, $\eta_{k+1}, \dots, \eta_{k+l} < 0$, and $\eta_{k+l+1} = \dots = \eta_n = 0$.

Write $\lambda_i = +\sqrt{\eta_i}$ for $i = 1, \dots, k$ and $\lambda_j = +\sqrt{-\eta_j}$ for $j = k + 1, \dots, k + l$. Note, that if

$$\mathcal{C}'' = (v_1'', \dots, v_n''),$$

then,

$$B(v_i'', v_i'') = \lambda_i^2 \quad B(v_j'', v_j'') = -\lambda_j^2.$$

Define $\mathcal{C} = (v_1, \dots, v_n)$ where $v_i = \frac{1}{\lambda_i} v_i''$ for $i = 1, \dots, k$ and $v_j = \frac{1}{\lambda_j} v_j''$ for $j = k+1, \dots, k+l$ and $v_m = v_m''$ for $m = k+l+1, \dots, n$. Then, $\mathcal{C}$ is a basis of V and

$$[B]_{\mathcal{C}} = \begin{bmatrix} I_k & 0 & 0 \\ 0 & -I_l & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let's show that k and l do not depend on $\mathcal{C}$. To that extent, put $k' = \max\{\dim W | W \leq V \text{s.t.} B|_{W\times W} \text{ is pos. def. }\}$. We will show that $k = k'$. Clearly, $k \leq k'$ since B is positive definite on $U \times U$, where $U =^{\langle} v_1, \dots, v_k\rangle$.

Suppose by contradiction, that $k' \geq k + 1$. Then $\exists W \subset V$ subspace such that $\dim W \geq k + 1$ and $B|_{W\times W}$ is positive definite. Take $X := \mathrm{Sp}(v_k + 1, \dots, v_n)$ with $\dim X = n - k$. Then,

$$\dim(W \cap X) = \dim W + \dim X - \dim(W + X) \geq (k+1) + (n-k) - n.$$

Using our assumptions, we get

$$\dim(W \cap X) \geq k + 1 + n - k - n = 1.$$

So $W \cap X \neq \{0\}$. Let $0 \neq w \in W \cap X$. Write

$$w = a_{k+1}v_{k+1} + \dots + a_nv_n.$$

Then,
$$B(w,w) = -a_{k+1}^2 - \cdots - a_{k+l}^2 \leq 0.$$
But this contradicts the fact that $B|_{W\times W}$ is positive definite. Hence, $k = k'$. Similarly, we can show that l does not depend on $\mathcal{C}$. □

Theorem 7.89:

Let V be a finite dimensional euclidean space of dimension n and $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form. Then, $\exists$ an orthonormal basis $\mathcal{C}$ of V such that
$$[B]_{\mathcal{C}} = \begin{pmatrix} \eta_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \eta_n \end{pmatrix},$$
where $\eta_i > 0$ for $i = 1, \dots, k$, $\eta_j < 0$ for $j = k+1, \dots, k+l$, and $\eta_m = 0$ for $m = k+l+1, \dots, n$.

The scalars $\eta_1, \dots, \eta_n$only depend on B and the inner product.

Lec 48

Proof. Analogous to the proof of Sylvester's inertia theorem (theorem 7.88). □

Recall that a symmetric bilinear form B is called positive definite if $B(v,v) > 0$ for all $v \neq 0$.

Theorem 7.90:

Let V be a finite dimensional vector space over $\mathbb{R}$ and B a symmetric bilinear form on V. Then, the following conditions are equivalent:

1. B is positive definite.
2. $\exists$ a basis $\mathbb{C} = (c_1, \dots, c_n)$ such that all principal minors of $[B]_{\mathcal{C}} = \det(B(c_j, c_k))_{j,k=1,\dots,l}$ for $l = 1, \dots, n$ are positive.
3. $\forall$ basis $\mathcal{C}$, all principal minors of $[B]_{\mathcal{C}}$ are positive.

Proof. $1 \Rightarrow 2$: Follows immediately from Sylvester's inertia theorem (theorem 7.88).

$2 \Rightarrow 1$: We do induction on $n = \dim V$. If $n = 1$, then $B(c_1, c_1) > 0$ and thus B is positive definite.

Suppose now, $2 \Rightarrow 1$ holds for all vector spaces of dimension $\leq n$ and any bilinear form on them. Let V be an $n+1$-dimensional vector space over $\mathbb{R}$ and $B : V \times V \to \mathbb{R}$ a symmetric bilinear form and $\mathcal{C} = (e_1, \dots, e_{n+1})$ such that all principal minors of $[B]_{\mathcal{C}}$ are positive.

Let $W := \mathrm{Sp}(e_1, \dots, e_n)$, and let $\mathcal{C}_W = (e_1, \dots, e_n)$ be a basis of W. Then, the principal minors of $[B]_{\mathcal{C}_W}$ are positive. By the induction hypothesis, $B|_{W\times W}$ is positive definite. But the principal minors are the same as the first n principal minors of $[B]_{\mathcal{C}}$. If we define $\langle w_1, w_2 \rangle_B := B(w_1, w_2)$, we get an inner product on W. Consider the linear functional $L : W \to \mathbb{R}$ given $L(w) := B(w, e_{n+1})$. By the Riesz representation theorem $\exists w' \in W$ such that $L(\cdot) = \langle \cdot, w' \rangle_B$. In other words, $\exists w'$ such that
$$B(w, e_{n+1}) = B(w, w') \Rightarrow B(w, e_{n+1} - w') = 0 \forall w \in W.$$
Define a new basis $\mathcal{C}' = (e_1, \dots, e_n, e_{n+1} - w')$ of V. Then, we have $B(e_j, e_{n+1} - w') = 0$ for $j = 1, \dots, n$. So the matrix representation of B with respect to $\mathcal{C}'$ is
$$[B]_{\mathcal{C}'} = \begin{pmatrix} & & & 0 \\ & [B]_{\mathcal{C}_W} & & \vdots \\ & & & 0 \\ 0 & \dots & 0 & B(e_{n+1} - w', e_{n+1} - w') \end{pmatrix}.$$
So $[B]_{\mathcal{C}'} = S^T [B]_{\mathcal{C}} S$ where S is the change of basis matrix from $\mathcal{C}$ to $\mathcal{C}'$.

Notice that $B(e_{n+1} - w', e_{n+1} - w')$ must be positive, since
$$\det([B]_{\mathcal{C}'}) = \det(S)^2 \det([B]_{\mathcal{C}}) > 0.$$
So also
$$\det([B]_{\mathcal{C}'}) = \det([B]_{\mathcal{C}_W}) B(e_{n+1} - w', e_{n+1} - w') > 0.$$
Let now $v \in V$ and write $v = w + \alpha(e_{n+1} - w')$ with $w \in W$ and $\alpha \in \mathbb{R}$. We have, that
$$B(v,v) = B(w,w) + \alpha^2 B(e_{n+1} - w', e_{n+1} - w') > 0.$$
Hence, B is positive definite.

$2 \Rightarrow 3$: Exercise □

Let's quickly return to quadratic forms. Assume $B \in M_{n\times n}(K)$. Then, the quadratic form $q : K^n \to K$ associated with B is the same as the one associated with B^T beacuse
$$v^T B v = v^T B^T v.$$

Assume, in our field K, we have $2 \neq 0$. Then also the matrix
$$\frac{1}{2}(B + B^T),$$
gives the same quadratic form as B. Note, that regardless of B, the matrix $\frac{1}{2}(B+B^T)$ is symmetric. So if $2 \neq 0$ in K, then every quadratic form is associated with a symmetric bilinear form.

If $2 \neq 0$ in K, and $B : V \times V \to K$ is a symmetric bilinear form, with associated quadratic form q, then:

1. $B(v,w) = \frac{1}{2}(q(v+w) - q(v) - q(w))$ for all $v, w \in V$.
2. $B(v,w) = \frac{1}{4}(q(v+w) - q(v-w))$ for all $v, w \in V$.

If $q : \mathbb{R}^n \to \mathbb{R}$ is a quadratic form, we can write it as
$$q(x_1, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

In geometry, a QUADRIC is the set
$$Q = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid q(x_1, \dots, x_n) = C\}.$$

Example 7.91: Quadrics

In $n = 2$, we have for $a, b > 0$,
$$Q = \{(x_1, x_2) \mid ax_1^2 - bx_2^2 = 1\}.$$
This set is a hyperbola. If
$$Q = \{(x_1, x_2) \mid ax_1^2 + bx_2^2 = 1\},$$
we get an ellipse.

Theorem 7.92:

Let V be a finite dimensional vector space over K where $2 \neq 0$. Let B be a symmetric bilinear form on V. Then, $\exists$ a basis $\mathcal{C}$ for V such that $[B]_{\mathcal{C}}$ is a diagonal matrix.

Proof. We argue by induction on $n = \dim V$. If $n = 1$ or $B \equiv 0$, then any basis $\mathcal{C}$ of V will do.

Let $n \geq 2$ and assume the statement holds for $n-1$. Let V be an n-dimensional space and B a symmetric bilinear form on V. If $B(v,v) = 0$

for all $v \in V$, then by the polarization formula, $B \equiv 0$. So we can assume that $\exists v_1 \in V$ such that $B(v_1, v_1) \neq 0$.

Put $W := \mathrm{Sp}(v_1)$. Then W is 1-dimensional. Define

$$W^{\perp_B} := \{v \in V \mid B(v, w) = 0 \forall w \in W\}.$$

As an exercise, show that $W^{\perp_B}$ is a subspace of V. We claim that $W \cap W^{\perp_B} = \{0\}$. Indeed, if $v \in W \cap W^{\perp_B}$, then $v = \alpha v_1$ for some $\alpha \in K$. So

$$0 = B(v, v_1) = \alpha B(v_1, v_1).$$

But since $B(v_1, v_1) \neq 0$, we get $\alpha = 0$. So $v = 0$.

We now claim that $W^{\perp_B}$ is a complement of W in V. It remains to show that $W + W^{\perp_B} = V$. Let $v \in V$. Consider

$$u = v - \frac{B(v, v_1)}{B(v_1, v_1)} v_1.$$

Note, that $u \in W^{\perp_B}$ since

$$B(v_1, u) = B(v_1, v) - \frac{B(v, v_1)}{B(v_1, v_1)} B(v_1, v_1) = 0.$$

So $v = \frac{B(v_1, v)}{B(v_1, v_1)} v_1 + u \in W + W^{\perp_B}$. Hence, $V = W + W^{\perp_B}$. We conclude, that $\dim W^{\perp_B} = n - 1$. By the induction hypothesis, $\exists$ a basis $\mathcal{C}' = (v_2, \dots, v_n)$ of $W^{\perp_B}$ such that $[B]_{\mathcal{C}'}$ is diagonal. Then, $\mathcal{C} = (v_1, v_2, \dots, v_n)$ is a basis of V and $[B]_{\mathcal{C}}$ is also diagonal. □

Lec 49

We want to look at symmetric bilinear forms over $\mathbb{C}$.

Theorem 7.93:

Let $B : V \times V \to \mathbb{C}$ be a symmetric bilinear form on a finite dimensional vector space V over $\mathbb{C}$ of dimension n. Then, $\exists$ a basis $\mathcal{C}$ of V such that

$$[B]_{\mathcal{C}} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

In fact, $n - r = \max\{\dim W | W \leq V \text{s.t.} B|_{W \times V} \equiv 0\}$. r is called the RANK of B.

Proof. By theorem 7.92, $\exists$ a basis

$\mathcal{C}' = (v_1', \dots, v_n')$ of V such that $[B]_{\mathcal{C}'}$ is diagonal. Let Choose now $\lambda_j \in \mathbb{C}$ such that $\lambda_j^2 = \eta_j$ and take the basis

$$\mathcal{C} = (\lambda_1^{-1} v_1', \dots, \lambda_r^{-1} v_r', v_{r+1}', \dots, v_n').$$

In this basis, $B(v_i, v_j) = 0 \forall i \neq j$ and for $1 \leq j \leq r$, we have

$$B(v_j, v_j) = \lambda_j^{-2} B(v_j', v_j') = \lambda_j^{-2} \eta_j = 1.$$

□

8 Jordan Normal Form

Let V be a finite dimensional vector space over K and $T : V \to V$ a linear map. The simplest case is when T is diagonalizable. In this case, $\exists$ a basis $\mathcal{B}$ such that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Unfortunately, we don't always have this luxury. The general case, is where the characteristic polynomial of T does not split as a product of linear factors. In this case, T is not even trigonalizable.

If T is trigonalizable, then $p_T(x)$ is a product of linear factors. The reason why T is not diagonalizable is because $p_T(x)$ for some eigenvalue λ, the geometric multiplicity of λ is less than its algebraic multiplicity.

Definition 8.1: Jordan Block

Let $\lambda \in K$, $n \in \mathbb{Z}_{\geq 1}$. Define the JORDAN BLOCK of size n with eigenvalue λ to be the matrix

$$J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix},$$

which is an $n \times n$ matrix.

Lemma 8.2:

1. λ is the only eigenvalue of $J_{\lambda,n}$.
2. The characteristic polynomial of $J_{\lambda,n}$ is $(x - \lambda)^n$.
3. The geometric multiplicity of λ is 1. In fact,

$$\mathrm{Eig}_\lambda(J_{\lambda,n}) = \mathrm{Sp}(e_1).$$

Proof. 1 + 2: Exercise.

3. To find $\mathrm{Eig}_\lambda(J_{\lambda,n})$, we need to solve

$$(J_{\lambda,n} - \lambda I)x = 0.$$

So we need to find the dimension of the kernel of the matrix

$$J_{\lambda,n} - \lambda I = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Writing down the equations, we get

$$x_2 = x_3 = \dots = x_n = 0.$$

So this leaves x_1 free and thus $\dim \mathrm{Eig}_\lambda(J_{\lambda,n}) = 1$ and $\mathrm{Eig}_\lambda(J_{\lambda,n}) = \mathrm{Sp}(e_1)$. □

Theorem 8.3: Jordan Normal Form

Let V be an n-dimensional vector space over K and $T : V \to V$ a linear map. Assume that $p_T(x)$ splits as a product of linear factors in $K[x]$. Then, $\exists$ a basis $\mathcal{B}$ for V such that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} J_{\lambda_1,n_1} & 0 & \dots & 0 \\ 0 & J_{\lambda_2,n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{\lambda_k,n_k} \end{bmatrix},$$

where $k \geq 1$, $\lambda_1, \dots, \lambda_k \in K$ are not necessarily distinct , and $n_1, \dots, n_k \in \mathbb{Z}_{\geq 1}$ are such that

$$n_1 + \dots + n_k = n.$$

Moreover, up to permutation, the blocks J_{λ_i,n_i} are uniquely determined by T.

Example 8.4:

The following 8 by 8 matrix is in Jordan normal form:

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

The Eigenvalues of this matrix are 2 and -1. The characteristic polynomial is

$$(x-2)^3(x+1)^5.$$

the algebraic multiplicity of 2 is 3 and the algebraic multiplicity of -1 is 5. The geometric multiplicity of 2 is 1 and the geometric multiplicity of -1 is 2.

Definition 8.5: Minimal Polynomial

A minimal polynomial of $T : V \to V$ is a polynomial $\mu_T(x) \in K[x]$ which is not 0, has minimal degree and satisfies $\mu_T(T) = 0$.

Similarly for matrices.

Claim 8.6:

1. Let $f \in K[x]$ be a polynomial such that $f(T) = 0$. Then $\forall$ eigenvalue λ of T, $f(\lambda) = 0$.

2. $\mu_A(x) | p_A(x)$

Proof. 1. Write $f(x) = a_0 + a_1 x + \dots + a_m x^m$. Then

$$0 = f(T) = a_0 I + a_1 T + \dots + a_m T^m.$$

If λ is an eigenvalue of T and v is a corresponding eigenvector, then

$$0 = f(T)v = a_0 v + a_1 \lambda v + \dots + a_m \lambda^m v = f(\lambda)v.$$

Since $v \neq 0$, we get $f(\lambda) = 0$.

2. We can always write $p_A(x) = q(x)\mu_A(x) + r(x)$ where $q, r \in K[x]$ and $\deg r < \deg \mu_A(x)$. Then, plugging in A, we get by Cayley-Hamilton

$$0 = p_A(A) = q(A)\mu_A(A) + r(A) = r(A).$$

Hence $r(A) = 0$. But $\deg r < \deg \mu_A(x)$ and $\mu_A(x)$ is the minimal polynomial. So $r(x) = 0$ and thus $\mu_A(x) | p_A(x)$. □

Let us focus a bit more on Jordan Blocks. Write $J_{\lambda,n} = \lambda I + N$, where

$$N = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Lemma 8.7:

1. The matrix N is nilpotent, i.e. $\exists m \in \mathbb{Z}_{\geq 1}$ such that $N^m = 0$.

$$N^2 = \begin{pmatrix} 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \dots, N^n = 0.$$

2. $\forall k \geq 1$, $J_{\lambda,n}^k$ is given by

$$\begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots & \binom{k}{n-1}\lambda^{k-n+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \binom{k}{n-2}\lambda^{k-n+2} \\ 0 & 0 & \lambda^k & \dots & \binom{k}{n-3}\lambda^{k-n+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda^k \end{pmatrix}.$$

In other words,

$$J_{\lambda,n}^k = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} N^j.$$

Proof. 1. Calculation (Exercise).

2. We have

$$J_{\lambda,n}^k = (\lambda I + N)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} N^j.$$

□

In the JNF, we can read the geometric multiplicity of an eigenvalue λ as the number of Jordan blocks with eigenvalue λ. The minimal polynomial is

$$\mu_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{s(\lambda_i)}.$$

where $s(\lambda_i)$ is the size of the largest Jordan block with eigenvalue λ_i. Lec 50

Proof. The proof is based on: Given $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, where $B \in M_{l \times l}(K)$ and $C \in M_{m \times m}(K)$ and $0 \neq u = \begin{pmatrix} v \\ w \end{pmatrix} \in K^{l+m}$, then u is an eigenvector of A with eigenvalue λ if and only if

$$Bv = \lambda v \text{ and } Cw = \lambda w.$$

Moreover, $\operatorname{Eig}_\lambda(A) \cong \operatorname{Eig}_\lambda(B) \oplus \operatorname{Eig}_\lambda(C)$. In particular, $m_g(A,\lambda) = m_g(B,\lambda) + m_g(C,\lambda)$. This is an exercise. From this, it follows that

$$m_g([T]^{\mathcal{B}}_{\mathcal{B}}, \lambda) = \# \text{ of Jordan blocks with eigenvalue } \lambda.$$

This is because $m_g(J_{\eta,m}, \eta') = \delta_{\eta,\eta'}$.

Let's look at the statement about the minimal polynomial. If

$$A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \Rightarrow A^k = \begin{bmatrix} B^k & 0 \\ 0 & C^k \end{bmatrix}.$$

So if $q(x) \in K[x]$, then $q(A) = \begin{bmatrix} q(B) & 0 \\ 0 & q(C) \end{bmatrix}$. Furthermore,

$$q(J_{\lambda,n}) = \begin{pmatrix} q(\lambda) & * & \dots & * \\ 0 & q(\lambda) & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q(\lambda) \end{pmatrix}.$$

If $q(\lambda) \neq 0$, then $q(J_{\lambda,n})$ is invertible. Consider now $q(x) = (x-\lambda)^s$. Then,

$$q(J_{\lambda,n}) = (J_{\lambda,n} - \lambda I)^s = N^s.$$

If $s < n$, this is a nonzero matrix and if $s \geq n$, this is the zero matrix. □

We now want to outline of the proof of the existence of the JNF.

Lemma 8.8:

Let V be a n dimensional vector space over K. Let $T \in \operatorname{End}(V)$ and $\lambda \in K$ an eigenvalue of T. Define the GENERALIZED EIGENSPACE of λ as

$$\widetilde{\operatorname{Eig}}_\lambda(T) = \bigcup_{j=1}^{\infty} \operatorname{Ker}((T-\lambda I)^j).$$

Then,

1. $\widetilde{\operatorname{Eig}}_\lambda(T)$ is a subspace of V. In fact, $\widetilde{\operatorname{Eig}}_\lambda(T) = \operatorname{Ker}((T-\lambda I)^n)$.

2. If $v \in \widetilde{\operatorname{Eig}}_\lambda(T)$ is an eigenvector of T corresponding to some eigenvalue μ of T, then $\mu = \lambda$. In other words, the only eigenvectors of T that belong to $\widetilde{\operatorname{Eig}}_\lambda(T)$ are those corresponding to λ.

Proof. 1. We have

$$\{0\} \subsetneq \operatorname{Ker}(T - \lambda \operatorname{id}_V) \subset \dots \subset \operatorname{Ker}((T-\lambda \operatorname{id}_V)^{n+1}) \subset V.$$

We have a sequence of $n+$ nested subspaces of V, where $\dim V = n$ and the 1st subspace is at least 1-dimensional. Thus,

$$\exists 1 \leq k \leq n : \operatorname{Ker}((T-\lambda \operatorname{id}_V)^k) = \operatorname{Ker}((T-\lambda \operatorname{id}_V)^{k+1}).$$

We claim that $\operatorname{Ker}((T-\lambda \operatorname{id}_V)^k) = \operatorname{Ker}((T-\lambda \operatorname{id}_V)^l)$ for all $l \geq k$. We do induction on $l \geq k$. The base case $l = k+1$ is by definition.

Let now $l > k+1$. We have seen that $\operatorname{Ker}((T-\lambda \operatorname{id}_V)^l) \supset \operatorname{Ker}((T-\lambda \operatorname{id}_V)^{l-1})$. Let $v \in \operatorname{Ker}((T-\lambda \operatorname{id}_V)^l)$. Then,

$$(T-\lambda \operatorname{id}_V)^l v = 0..$$

Put $w := (T-\lambda \operatorname{id}_V)v$. Thus, we have

$$(T-\lambda \operatorname{id}_V)^{l-1} w = 0.$$

By induction hypothesis, $w \in \operatorname{Ker}((T-\lambda \operatorname{id}_V)^k)$. But $\operatorname{Ker}((T-\lambda \operatorname{id}_V)^k) = \operatorname{Ker}((T-\lambda \operatorname{id}_V)^{k+1})$. So $w \in \operatorname{Ker}((T-\lambda \operatorname{id}_V)^{k+1})$. Hence,

$$(T-\lambda \operatorname{id}_V)^{k+1} w = 0 \Rightarrow (T-\lambda \operatorname{id}_V)^{k+2} v = 0.$$

2. Let $v \in V$ be an eigenvector of T with eigenvalue μ and assume $v \in \widetilde{\operatorname{Eig}}_\lambda(T)$. Then, $Tv = \mu v$ implying

$$T^j v = \mu^j v \forall j \geq 0.$$

Hence $\forall q(x) \in K[x]$, $q(T)v = q(\mu)v$. In particular, if $q(x) = (x-\lambda)^n$, then

$$0 = q(T)v = q(\mu)v = (\mu - \lambda)^n v.$$

□

Definition 8.9: Invariant Subspace

Let W be a vector space over K. Let $S \in \operatorname{End}(W)$. A subspace $U \subset W$ is called S-INVARIANT if $S(U) \subset U$. In this case, we get a restricted endomorphism

$$S|_U : U \to U.$$

Exercise 8.10:

Let $S \in \operatorname{End}(W)$ and $U \subset W$ an S-invariant subspace. Then,

$$p_{S|_U}(x) | p_S(x).$$

Hint: Choose a basis $\mathcal{B}$ of U and extend it to a basis $\mathcal{C}$ of W.

Lemma 8.11:

Let V be a finite dimensional vector space over K and let $T \in \operatorname{End}(V)$ and $f(x) \in K[x]$. Then $\operatorname{Ker}(f(T))$ is T-invariant. In particular, if λ is an eigenvalue of T, then $\widetilde{\operatorname{Eig}}_\lambda(T)$ is T-invariant.

Proof. If $v \in \operatorname{Ker}(f(T))$, then $f(T)Tv = Tf(T)v = 0$. Hence, $Tv \in \operatorname{Ker}(f(T))$. So the kernel of $f(T)$ is T-invariant. If we take $f(x) = (x-\lambda)^n$, then $\operatorname{Ker}(f(T)) = \widetilde{\operatorname{Eig}}_\lambda(T)$ is T-invariant. □

Proposition 8.12:

Let V be a finite dimensional vector space over K and $T \in \operatorname{End}(V)$. Assume that $p_T(x)$ splits as a product of linear factors in $K[x]$. Then,

$$V = \bigoplus_{\lambda \text{ eigenvalue of } T} \widetilde{\operatorname{Eig}}_\lambda(T).$$

Moreover, $m_a(T,\lambda) = \dim \widetilde{\operatorname{Eig}}_\lambda(T)$.

Proof. Postponed □

Definition 8.13: Nilpotent

1. Let V be a finite dimensional vector space over K and $N \in \operatorname{End}(V)$. N is called NILPOTENT if $\exists m \in \mathbb{Z}_{\geq 1}$ such that $N^m = 0$. The minimal such m is called the NILPOTENCY INDEX of N.

2. $A \in M_{n\times n}(K)$ is called NILPOTENT if $\exists m \in \mathbb{Z}_{\geq 1}$ such that $A^m = 0$. The minimal such m is called the NILPOTENCY INDEX of A.

Example 8.14:

$J_{0,n}$ is nilpotent with nilpotency index n.

Exercise 8.15:

1. If $N \in \operatorname{End}(V)$ is nilpotent, then 0 is an eigenvalue of N. In fact, it is the only eigenvalue of N.

2. If $N \in \operatorname{End}(V)$ is nilpotent, then $N^{\dim V} = 0$.

Corollary 8.16:
Let $T \in \mathrm{End}(V)$ and λ an eigenvalue of T. Then,

$$(T - \lambda \,\mathrm{id}_V)|_{\widetilde{\mathrm{Eig}}_\lambda(T)} = 0,$$

where $m = \dim \widetilde{\mathrm{Eig}}_\lambda(T)$. Moreover, if $p_T(x)$ splits as a product of linear factors, then

$$\dim \widetilde{\mathrm{Eig}}_\lambda(T) = m_a(T, \lambda).$$

Lec 51

Proof. [JNF] It's enough to proof JNF for $N \in \mathrm{End}(U)$ nilpotent. We can do this, because if $T \in \mathrm{End}(V)$ and $p_T(x)$ splits as a product of linear factors, then

$$V = \bigoplus_\lambda \widetilde{\mathrm{Eig}}_\lambda(T).$$

Each $\widetilde{\mathrm{Eig}}_\lambda(T)$ is T-invariant so we can consider

$$N_\lambda := (T - \lambda \,\mathrm{id}_V)|_{\widetilde{\mathrm{Eig}}_\lambda(T)}.$$

Where $N_\lambda \in \mathrm{End}(\widetilde{\mathrm{Eig}}_\lambda(T))$. Also $p_{N_\lambda}(x)$ also splits as a product of linear factors. Also $N_\lambda^{m_\lambda}$ is zero, where $m_\lambda = \dim \widetilde{\mathrm{Eig}}_\lambda(T)$. So if we know the JNF for N_λ, then we get the JNF for $T|_{\widetilde{\mathrm{Eig}}_\lambda(T)}$ because

$$[T|_{\widetilde{\mathrm{Eig}}_\lambda(T)}]_{\mathcal{B}}^{\mathcal{B}} = \lambda I + [N_\lambda]_{\mathcal{B}}^{\mathcal{B}}.$$

To get the JNF for T, do the same on each generalized eigenspace and combine the bases together.

Let us show JNF for nilpotent endomorphisms. We write $N \in \mathrm{End}(V)$ nilpotent, with $n = \dim(V)$ and $p_N(x)$ splits as a product of linear factors. We will argue by induction on n. If $n = 1$, then $N = 0$ and then

$$[N]_{\mathcal{B}}^{\mathcal{B}} = 0 = J_{0,1}.$$

Let $n > 1$ and assume the JNF exists for all nilpotent N and every vector space V of dimension $\leq n-1$. Let V be n-dimensional and $N \in \mathrm{End}(V)$ nilpotent with $p_N(x)$ splitting as a product of linear factors. Define

$$k := \min\{j \in \mathbb{Z}_{\geq 1} : N^j = 0\} \leq n.$$

If $k = 1$, then $N = 0$ and we are done. So assume $k \geq 2$. Then, $N^{k-1} \neq 0$. Hence $\exists e \in V$ such that $N^{k-1}e \neq 0$. Denote $e_k := e$. Denote $e_j := N^{k-j}e_k$ for $1 \leq j \leq k-1$. So the vectors are

$$N^{k-1}e_k, \dots, Ne_k, e_k.$$

We claim that these vectors are linearly independent. Suppose that

$$c_1e_1 + \cdots + c_ke_k = 0.$$

Applying N^{k-1} to both sides, we get

$$c_1N^{k-1}e_1 + \cdots + c_kN^{k-1}e_k = c_1N^{k-1}e_1 = 0.$$

So $c_1 = 0$. Applying N^{k-2} to both sides, we get $c_2 = 0$ and so on. Hence, $e_1, \dots, e_k$ are linearly independent. Define $W := \mathrm{Sp}(e_1, \dots, e_k)$. Then, W is N-invariant.

Take $\mathcal{B}_1 := (e_1, \dots, e_k)$. We have

$$[N|_W]_{\mathcal{B}_1}^{\mathcal{B}_1} = J_{0,k}.$$

Consider V/W. Since N sends W to W, N descends to a linear map. Letting $N' : V/W \to V/W$ be the induced linear map, N' is also nilpotent. Also $\dim V/W = n-k < n$. So by the induction hypothesis, $\exists$ a basis $B' = (u_1, \dots, u_l)$ where $l = n-k$ for V/W such that

$$[N']_{B'}^{B'} = \text{JNF}.$$

Chosse $f_1, \dots, f_l \in V$ such that $[f_i] = u_i \forall i$. In the basis

$$B_2 = (e_1, \dots, e_k, f_1, \dots, f_l)$$

we have

$$[N]_{B_2}^{B_2} = \begin{bmatrix} J_{0,k} & A \\ 0 & \text{JNF} \end{bmatrix}.$$

To get rid of the A, we will work on every block in the JNF of N'. We need to modify $f_1, \dots, f_m$ by adding suitable elements from W. We will not go into detail here...

For uniqueness, if $N \in \mathrm{End}(V)$ is nilpotent, then: If $A = [N]_{\mathcal{B}}^{\mathcal{B}}$ is in JNF, then

$$\dim \mathrm{Ker}(N) = \#\text{ of Jordan blocks in } A.$$

Furthermore,

$$\dim \mathrm{Ker}(N^2) = \#\text{ of Jordan blocks} + \#\text{ Blocks of size} \geq 2.$$

This goes on recursively.

$\square$