Linear Algebra 2

Lecture notes, University (of Technology) Graz
based on the lecture by Franz Lehner

Lukas Prokop

June 14, 2019

Contents

6 Preface

↓ *This lecture took place on 2018/03/05.*

6.1 Lecture

- Mon, 08:15–09:45, lecture
- Wed, 08:15–09:45, lecture
- Mon, 16:00–18:00, tutorial, AE01
- Mon, 13:15–14:00, conversatorium (BE01)

6.2 Linear algebra 1

Person. *Gottfried Wilhelm von Leibniz (1646–1716)*

Results from 1693:

- Vector spaces (first definition in 1880)
- Matrices and linear maps

From now, it will be more specific (→ matrices). In general, we discuss the question "when is a matrix invertible"?

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

We need to invert the matrix.

Assuming $a \neq 0$. We multiply the first row with $\cdot \frac{1}{a} \cdot (-c)$.

$$\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \\ \hline 0 & d - \frac{c}{a} \cdot b & -\frac{c}{a} & 1 \end{array}$$

We then divide by $d - \frac{c}{a}b$ if $\neq 0$.

If $a = 0$ and $c = 0$, rank is certainly not 2.

If $a = 0$ and $c \neq 0$, we multiply with $\frac{1}{c}(-a)$.

$$\begin{array}{cc} a & b \\ c & d \\ \hline 0 & b - \frac{ad}{c} \end{array}$$

we divide $b - \frac{ad}{c}$ if $\neq 0$.

When does such a system have a non-trivial solution? There is a non-trivial solution iff $ad - bc \neq 0$.

$ad - bc \neq 0$ iff $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible.

Leibniz was not the first discovering it. The result was found before 1685 by Sehi Takahazu.

7 Determinants

7.1 Definition

Definition 7.1.

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} =: ad - bc =: \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

is called determinant of matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

7.2 Properties

Remark.

- *The determinant is linear in every row and every column. For fixed b and d, it is*

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \det \begin{pmatrix} x & b \\ y & d \end{pmatrix} = dx - by \qquad \textit{is linear}$$

$$\mathbb{K}^2 \to \mathbb{K}$$

$$\begin{aligned} \det \begin{pmatrix} \lambda x + \mu x' & b \\ \lambda y + \mu y' & d \end{pmatrix} &= d \cdot (\lambda x + \mu x') - b \cdot (\lambda y + \mu y') \\ &= \lambda(dx - by) + \mu(dx' - by') \\ &= \lambda \det \begin{pmatrix} x & b \\ y & d \end{pmatrix} + \mu \det \begin{pmatrix} x' & b \\ y' & d \end{pmatrix} \end{aligned}$$

  The determinant is bilinear in rows and columns.

$$\det(\lambda v + \mu v', w) = \lambda \det(v, w) + \mu \det(v', w)$$

Let $v = \begin{pmatrix} a \\ c \end{pmatrix}$.

$$\det(v, \lambda w + \mu w') = \lambda \det(v, w) + \mu \det(v, w')$$

Let $w = \begin{pmatrix} b \\ d \end{pmatrix}$. *Follows analogously.*

- *If two rows are the same, then* $\det(M) = 0$.

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ab - ba = 0$$

$$\det \begin{pmatrix} a & a \\ c & c \end{pmatrix} = ac - ca = 0$$

- *The determinant of the unit matrix is one.*

$$\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$$

Theorem 7.2 (properties 1–3 characterize the determinant)**.** *If* $\varphi : \mathbb{K}^2 \times \mathbb{K}^2 \to \mathbb{K}$ *and* φ *satisfies the properties 1–3,* φ *is the determinant.*

bilinear[1]

$$\varphi(\lambda v + \mu v', w) = \lambda \varphi(v, w) + \mu \varphi(v', w)$$
$$\forall v, w, v', w' : \mu(v, \lambda w + \mu w') = \lambda \varphi(v, w) + \mu \varphi(v, w')$$

$$\forall v : \varphi(v, v) = 0$$
$$\implies \varphi = \det$$

$\varphi(e_1, e_2) = 1$

Proof.

$$v = \begin{pmatrix} a \\ c \end{pmatrix} = a \cdot e_1 + c \cdot e_2$$

$$w = \begin{pmatrix} d \\ b \end{pmatrix} = b \cdot e_1 + d \cdot e_2$$

$$\begin{aligned} \varphi(v, w) &= \varphi(a \cdot e_1 + c \cdot e_2, b \cdot e_1 + d \cdot e_2) \\ &= a \cdot \varphi(e_1, b \cdot e_1 + d \cdot e_2) + c \cdot \varphi(e_2, b \cdot e_1 + d \cdot e_2) \\ &= ab \cdot \underbrace{\varphi(e_1, e_1)}_{=0} + ad \cdot \varphi(e_1, e_2) + cb \cdot \varphi(e_2, e_1) + cd \cdot \underbrace{\varphi(e_2, e_2)}_{=0} \end{aligned}$$

Is zero, because of property 3.

$$= ad \cdot \underbrace{\varphi(e_1, e_2)}_{=1} + cb \cdot \varphi(e_2, e_1)$$

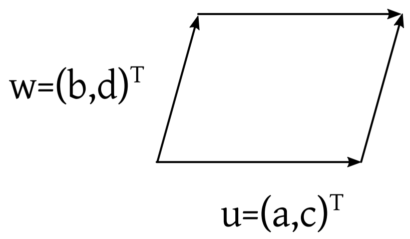


Figure 1: Geometric interpretation of determinants

$$
\begin{aligned}
0 = \varphi(e_1 + e_2, e_1 + e_2) = \underbrace{\varphi(e_1, e_1)}_{=0} + \underbrace{\varphi(e_1, e_2)}_{=1} + \varphi(e_2, e_1) + \underbrace{\varphi(e_2, e_2)}_{=0} \\
\implies \varphi(e_2, e_2) = -1
\end{aligned}
$$

□

Corollary 7.3.

$$\forall v, w : \varphi(v, w) = -\varphi(w, v)$$

Corollary 7.4 (Geometrical interpretation). *See Figure 1. The determinant* $\det(v, w)$ *is the area of the spanned parallelogram. We denote F as the function returning the area of a geometric object.*

Proof. $\text{area}(v, w)$ satisfies properties $(i) - (iii)$.
Consider orthogonal e_1 and e_2. $F = 1 = \det(e_1, e_2)$. $\det(e_2, e_1) = -1$.
The sign indicates the orientation of the area. □

By property 2, if $v = w$, then $F = 0$. By property 1,

1. If v and w are *linear dependent*[2], then

   $$\lambda v + \mu w = 0 \qquad (\lambda, \mu) \neq (0, 0)$$

   Without loss of generality, $\mu \neq 0 \implies w = -\frac{\lambda}{\mu} \cdot v$.

2. To show:

   $$F(\lambda v, w) = \lambda \cdot F(v, w)$$
   $$F(v + v', w) = F(v, w) + F(v', w)$$

   Let $\lambda \in \mathbb{N}$. We multiply the area n times.

   $$F(n \cdot v, w) = n \cdot F(v, w)$$

[2]Hence, one vector is a multiple of the other

3. $$F\left(\frac{1}{n}\cdot v, w\right) = \frac{1}{n}F(v, w)$$

   follows from $F(\lambda v, w) = \lambda \cdot F(v, w)$, because $v = n \cdot \left(\frac{1}{n}v\right)$:

   $$F\left(n\left(\frac{1}{n}v\right), w\right) = n \cdot F\left(\frac{1}{n}v, w\right)$$

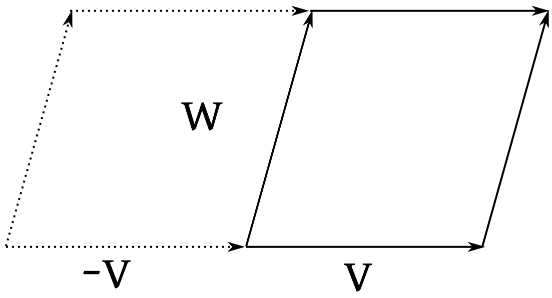
4. 

   Figure 2: The sign changes if the orientation changes

   If we combine (2) and (3),

   $$F\left(\frac{m}{n}v, w\right) = \frac{m}{n}F(v, w)$$

   See Figure 2.

5. By continuity, $F(\lambda v, w) = \lambda F(v, w) \forall \lambda \in \mathbb{R}_+$[3]. If the orientation changes, the sign changes. By this property, this actually holds for $\mathbb{R}$, not only $\mathbb{R}_+$.

   Analogously:

   $$F(v, \lambda w) = \lambda F(v, w) \forall \lambda \in \mathbb{R} \forall v, w \in \mathbb{R}^2$$

6. To show: $F(v + v', w) = F(v, w) + F(v', w)$

   If v and w are linear independent, then $F(v + w, w) = F(v, w)$. In general, for a parallelogram of height h and vector w, it holds that

   $$F = |w| \cdot h$$

   The height of the parallelogram stays the same.

   $$F(v, w) = F(v + w, w)$$

7. $$F(\lambda v + \mu w, w) = \lambda F(v, w)$$

   Case $\mu = 0$ Already shown, $F(\lambda v, w) = \lambda F(v, w) \forall \lambda \in \mathbb{R}$.

[3] By the way, how are real numbers defined?

Case $\mu \neq 0$ $F(\lambda v + \mu w, w) = \frac{1}{\mu} F(\lambda v + \mu w, \mu w) = \frac{1}{\mu} F(\lambda v, \mu w) = F(\lambda v, w) = \lambda F(v, w)$

8. Let v and w be linear independent, then they define a basis of $\mathbb{R}^2$.

$$v_1 = \lambda_1 v + \mu_1 w$$
$$v_2 = \lambda_2 v + \mu_2 w$$

$$\begin{aligned}
\to F(v_1 + v_2, w) &= F(\lambda_1 v + \mu, w + \lambda_2 v + \mu_2 w, w) \\
&= F((\lambda_1 + \lambda_2) v + (\mu_1 + \mu_2) w, w) \\
&= F((\lambda_1 + \lambda_2) v, w) \\
&= (\lambda_1 + \lambda_2) F(v, w) \\
&= \lambda_1 F(v, w) + \lambda_2 F(v, w) \\
&= F(\lambda_1 v, w) + F(\lambda_2 v, w) \\
&= F(\lambda_1 v + \mu_1 w, w) + F(\lambda_2 v + \mu_2 w, w) \\
&= F(v_1, w) + F(v_2, w)
\end{aligned}$$

This shows that additivity is given.

7.3 Determinant form

Definition 7.5. *Let V be an n-dimensional vector space over $\mathbb{K}$. A* determinant form *is a map*

$$\triangle : V^n \to \mathbb{K}$$
$$(a_1, \dots, a_n) \mapsto \triangle(a_1, \dots, a_n)$$

Let $n = 2$.

$$\triangle : \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) \mapsto \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

It satisfies the properties of *multilinearity*:

1. $\triangle(a_1, \dots, \lambda a_k, \dots, a_n) = \lambda \triangle(a_1, \dots, a_n)$
2. $\triangle(a_1, \dots, a_k + v, \dots, a_n) = \triangle(a_1, \dots, a_k, \dots, a_n) + \triangle(a_1, \dots, a_{k-1}, v, a_{k+1}, \dots, a_n)$

Multilinearity is given, if linearity is given in every component. Hence, if $a_1, \dots, a_{k-1}, a_{k+1}, \dots, a_n$ are fixed, then

$$V \to \mathbb{K}$$
$$v \mapsto \triangle(a_1, \dots, a_{k-1}, v, a_{k+1}, \dots, a_n) \text{ linear}$$

Furthermore, it satisfies the following property:

3. $\triangle(a_1, \dots, a_n) = 0$
   if $\exists k \neq l : a_k = a_l$. If $\triangle \not\equiv 0$, then $\triangle$ is called *non-trivial*.

Corollary 7.6. 4. $\triangle(a_1, \dots, a_k + \lambda a_i, \dots, a_n) = \triangle(a_1, \dots, a_k, \dots, a_n) \forall \lambda \in \mathbb{K}, \forall i \neq k$

5. $\triangle(a_1,\dots,a_i,\dots,a_j,\dots,a_n) = -\triangle(a_1,\dots,a_j,\dots,a_i,\dots,a_n)$

Proof. 4.

$$\begin{aligned}\triangle(a_1,\dots,a_k+\lambda a_i,\dots,a_n) &= \triangle(a_1,\dots,a_k,\dots,a_n) + \triangle(a_1,\dots,a_{k-1},\lambda a_i,a_{k+1},\dots,a_n) \\ &= \triangle(a_1,\dots,a_n) + \underbrace{\lambda\triangle(a_1,\dots,a_{k-1},a_i,a_{k+1},\dots,a_n)}_{=0 \qquad \text{because } a_i \text{ occurs twice}}\end{aligned}$$

5.

$$\begin{aligned}0 &= \triangle(a_1,\dots,a_i+a_j,\dots,a_i+a_j,\dots,a_n) \\ &= \triangle(a_1,\dots,a_i,\dots,a_i,\dots,a_n) \\ &+ \triangle(a_1,\dots,a_i,\dots,a_j,\dots,a_n) \\ &+ \triangle(a_1,\dots,a_j,\dots,a_i,\dots,a_n) \\ &+ \triangle(a_1,\dots,a_j,\dots,a_j,\dots,a_n)\end{aligned}$$

The first and last summands are zero. Multilinearity is given:

$$\lambda(a_1,\dots,\lambda a_k,\dots,a_n) = \lambda\triangle(a_1,\dots,a_n)$$

$$\lambda(a_1,\dots,\lambda a_k+v,\dots,a_n) = \lambda\triangle(a_1,\dots,a_n) + \triangle(a_1,\dots,a_{k-1},v,a_{k+1},\dots,a_n)$$

□

↓ *This lecture took place on 2018/03/07.*

Determinant form: $\dim V = n$

$$\triangle : V^n \to \mathbb{K}$$

1. $\triangle(a_1,\dots,a_{k-1},\lambda a_k,a_{k+1},\dots,a_n) = \lambda\triangle(a_1,\dots,a_n)$
2. $\triangle(a_1,\dots,a_{k-1},a_k+v,a_{k+1},\dots,a_n = \triangle(a_1,\dots,a_k,\dots,a_n)+\triangle(a_1,\dots,v,\dots,a_n)$
3. $\triangle(a_1,\dots,a_n) = 0$ if $\exists i \neq j : a_i = a_j$

Multilinearity is given by the first two properties.

$\triangle \not\equiv 0$

Then the fourth property follows:

4. $\triangle(a_1,\dots,a_k+\lambda a_i,\dots,a_n) = \triangle(a_1,\dots,a_n) \forall i \neq k \forall \lambda \in \mathbb{K}$
5. $\triangle(a_1,\dots,a_i,\dots,a_j,\dots,a_n) = -\triangle(a_1,\dots,a_j,\dots,a_i,\dots,a_n)$

Example. *Let* $n = 2$, $V = \mathbb{K}^2$.

$$\triangle\left(\begin{pmatrix}a\\c\end{pmatrix},\begin{pmatrix}b\\d\end{pmatrix}\right) = ad - bc = \det\begin{pmatrix}a & b\\ c & d\end{pmatrix}$$

7.4 Permutations and transpositions

Definition 7.7. *A* permutation *is a bijective map* $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$*.* σ_n *is the set of all permutations.*

$$|\sigma_n| = n!$$

Remark 7.8. σ_n *in regards of composition defines a group with neutral element* id *and is called* symmetric group.

Remark 7.9. *For* $n \geq 3$*, it is non-commutative.*

Example 7.10. *Permutations:*

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

So, e.g. 2 is mapped to 3 (right side of $\circ$*) and 3 is mapped to 3 (left side of* $\circ$*). Hence 2 is mapped to 3 (right-hand side of* $=$*).*

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

Definition 7.11. *A* transposition *is a permutation exchanging exactly 2 elements.*

$$\tau_{ij} : \begin{cases} i \mapsto j \\ j \mapsto i \\ k \mapsto k \forall k \notin \{i, j\} \end{cases}$$

$$\tau_{ij}^{-1} = \tau_{ij}$$

Lemma. *Every permutation* $\sigma \in \sigma_n$ *with* $\sigma \neq \mathrm{id}$ *can be denoted as product of transpositions.*

Example.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 3 & 5 & 4 & 7 & 6 & 2 \end{pmatrix}$$

Proof.

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

Find transpositions $\tau_1, \dots, \tau_k$ such that $\sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_k$.

If $\sigma = \mathrm{id}$, then $k = 0$.

If $\sigma \neq \mathrm{id}$,

$$k_1 := \min\{i \mid \sigma(i) \neq i\} \neq \emptyset$$

$$\tau_1 := \tau_{k_1 \sigma(k_1)}$$

$$\sigma_1 := \tau_i \circ \sigma$$

if $\sigma_i = \mathrm{id}$, then $\tau_1 \circ \sigma = \mathrm{id}$. Then $\sigma = \tau_1^{-1} = \tau_i$.

$$k_2 := \min\{i \mid \sigma_i(i) \neq i\}$$

$$\tau_2 \coloneqq \tau_{k_2\sigma_1(k_2)}$$
$$\sigma_2 \coloneqq \tau_2 \circ \sigma_1$$

□

Example. *Let* $k_1 = 2$.

$$\tau_1 = \tau_{23}$$

$$\sigma_1 = \tau_{23} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 3 & 5 & 4 & 7 & 6 & 2 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 5 & 4 & 7 & 6 & 3 \end{pmatrix}$$

$k_2 = 3$.

$$\tau_2 = \tau_{35}$$

$$\sigma_2 = \tau_2 \circ \sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 4 & 7 & 6 & 5 \end{pmatrix}$$

$k_3 = 5$.

$$T_3 = T_{57}$$

$$\sigma_3 = \tau_3 \circ \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{pmatrix}$$
$$= \mathrm{id}$$

$$\tau_3 \circ \tau_2 \circ \tau_1 \circ \sigma = \mathrm{id}$$
$$\implies \tau_2 \circ \tau_1 \circ \sigma = T_3^{-1} \circ \mathrm{id} = \tau_3$$
$$\tau_1 \circ \sigma = \tau_2^{-1} \circ T_3 = \tau_2 \circ \tau_3$$
$$\sigma = \tau_1 \circ \tau_2 \circ \tau_3$$

and so on and so forth.

$$\tau_k$$
$$\sigma_k = \tau_k \circ \tau_{k-1} \circ \cdots \circ \tau_i \circ \sigma = \mathrm{id}$$
$$\implies \sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$$

Remark. *This decomposition is not unique.*

Definition 7.12. *Let* $\pi \in \sigma_n$ *be a permutation. A* malposition *(dt. Fehlstand) of* π *is a pair* (i, j) *such that* $i < j$ *and* $\pi(i) > \pi(j)$.

$$f_\pi \coloneqq \left| \left\{ (i, j) \,\middle|\, (i, j) \text{ is malposition of } \pi \right\} \right|$$

$$\mathrm{sign}(\pi) \coloneqq (-1)^{f_\pi} \eqqcolon (-1)^\pi$$

is called signature of π

Example 7.13.

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 3 & 5 & 4 & 7 & 6 & 2 \end{pmatrix}$$

Malpositions:

$$\{(2,7),(3,4),(3,7),(5,6),(5,7),(4,7),(6,7)\}$$

$$2 < 7$$

$$\pi(2) - 3 > 2 = \pi(7)$$

$$f_\pi = 7$$

Theorem 7.14.

$$\operatorname{sign}(\pi) = \prod_{\substack{i,j \\ i<j}} \frac{\pi(j) - \pi(i)}{j - i}$$

1. $\binom{n}{2}$ *factors*
2. *for transposition,* $\operatorname{sign} \tau = -1$.

Proof.

$$\prod_{i<j} \frac{\pi(j) - \pi(i)}{j - i} = \frac{\prod_{i<j}(\pi(j) - \pi(i))}{\prod_{i<j}(j - i)}$$

π is bijective in $\{1,\dots,n\}$. Hence, every difference (the expression $j-i$) occurs exactly one time in the enumerator and the denomiator with sign ± 1 depending on whether (i,j) is a malposition or not (statement 1). In other words, the term in the enumerator and denominator cancel out if they are not malpositions.

$$\operatorname{sign}(\pi(j) - \pi(i)) = \begin{cases} +1 & \pi(j) > \pi(i) \\ -1 & \pi(j) < \pi(i) \text{ hence malposition} \end{cases}$$

Consider any transposition, let k be the first index with $k \neq \pi(k)$ and let l be the last index with $l \neq \pi(l)$. $(k,k+1),(k,k+2),\dots,(k,l-1)$ are $(l-1-k+1+1)$ malpositions. $(k+1,l),(k+2,l),\dots,(l-1,l)$ are $(l-1-k+1+1)$ malpositions. The sum gives an even number. Additional, we have malposition (k,l), thus an odd number of malpositions is given. Thus $\operatorname{sign} \tau = -1$ (statement 2). ◻

Example.

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 3 & 5 & 4 & 7 & 6 & 2 \end{pmatrix}$$

Malposition:

$$\{(2,7),(3,4),(3,7),(5,6),(5,7),(4,7),(6,7)\}$$

$$2 < 7$$

$$\pi(2) - 3 > 2 = \pi(7)$$

$$f_\pi = 7$$

$$\frac{\prod_{i<j}(\pi(j)-\pi(i))}{\pi_{i<j}(j-i)} = \frac{\prod_{i<j}(j-i)\cdot(-1)^{f_\pi}}{\prod_{i<j}(j-i)} = \operatorname{sign}\pi$$

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

$$\begin{aligned}
\prod_{i<j}\frac{\pi(j)-\pi(i)}{j-i} &= \frac{\pi(2)-\pi(1)}{2-1}\cdot\frac{\pi(3)-\pi(1)}{3-1}\cdot\frac{\pi(3)-\pi(2)}{3-2} \\
&= \frac{(2-3)\cdot(1-3)\cdot(1-2)}{(2-1)(3-1)(3-2)} \\
&= (-1)^3 = -1
\end{aligned}$$

Malpositions:

1. $(1,2)$
2. $(1,3)$
3. $(2,3)$

Transposition: Let $k < \tau(k)$.

$$\tau = \left\{ \begin{array}{cccccccccc} 1 & 2 & \dots & k-1 & k & k+1 & \dots & \tau(k) & \tau(k+1) & \dots & n \\ 1 & 2 & \dots & k-1 & \tau(k) & k+1 & \dots & k & \tau(k+1) & \dots & n \end{array} \right\}$$

Malpositions (denoted F_τ*):*

$$F_\tau = \begin{cases} (k,k+1),\dots,(k,\tau(k)) \\ (k+1,\tau(k)),(k+2,\tau(k)),\dots,(\tau(k)-1,\tau(k)) \end{cases}$$

Let us count on a specific example:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 6 & 4 & 5 & 3 & 7 \end{pmatrix}$$

$$\begin{cases} (3,4),(3,5),(3,6) \\ (4,6),(5,6) \end{cases}$$

$$|F_\tau| = (\tau(k)-k)+((\tau(k)-1)-k) = 2\tau(k)-2k-1 = 2(\tau(k)-k)-1 \textit{ even}$$

Theorem 7.15.

1. $\operatorname{sign}(\mathrm{id}) = 1$
2. $\operatorname{sign}(\pi\circ\sigma) = \operatorname{sign}(\pi)\cdot\operatorname{sign}(\sigma)$
   Hence, $\operatorname{sign}\sigma_n \to \{\pm 1\}$ *is a homomorphism.*
   $(\{+1,-1\},\cdot)$ *is a group* $\tilde{=}$ $(\mathbb{Z}_2,+)$

$$+1 \to [0]_2 \qquad -1 \to [1]_2$$

3. $\operatorname{sign}(\pi^{-1}) = \operatorname{sign}(\pi)$

Proof. 1. obvious, because there are no malpositions

2.
$$\operatorname{sign}(\pi \circ \sigma) = \prod_{i<j} \frac{(\pi \circ \sigma(j) - \pi \circ \sigma(i))}{j-i} \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{\sigma(j)-\sigma(i)}$$

because of bijectivity

$$= \underbrace{\prod_{i<j} \frac{\pi(\sigma(j)) - \pi(\sigma(i))}{\sigma(j)-\sigma(i)}}_{\operatorname{sign}\pi} \cdot \underbrace{\prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}}_{\operatorname{sign}\sigma}$$

3. Homomorphism

$\operatorname{sign}(\pi^{-1} \circ \pi) = 1 \iff \operatorname{sign}(\pi^{-1}) \cdot \operatorname{sign}(\pi) = 1 \iff \operatorname{sign}(\pi^{-1}) = \operatorname{sign}(\pi)^{-1}$

□

Remark. *Recall that the kernel of a homomorphism defines a subgroup.*

Corollary 7.16. 1. *If* $\pi = \tau_1 \circ \cdots \circ \tau_k$ *is a product of transpositions, then* $\operatorname{sign}(\pi) = (-1)^k$

2. $\mathfrak{a}_n = \{\pi \in \sigma_n \mid \operatorname{sign}(\pi) = +1\} = \ker(\{\operatorname{sign} : \sigma_n \to \{\pm 1\}\})$ *is a subgroup of* σ_n*, the so-called* alternating group
$$|\mathfrak{a}_n| = \frac{n!}{2}$$

Corollary 7.17.

$$\dim V = n \qquad \triangle : V^n \to \mathbb{K} \qquad \textit{determinant form}$$

then it holds that $\forall \sigma \in \sigma_n : \triangle(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) = \operatorname{sign}(\sigma) \cdot \triangle(a_1, \ldots, a_n)$

Proof. If $\sigma = \tau$ is a transposition, by the fourth property we have:

$$\triangle(a_{\tau(1)}, \ldots, a_{\tau(n)}) = -\triangle(a_1, \ldots, a_n) \implies \operatorname{sign}(\tau) = -1$$

The general case: $\sigma = \tau_1 \circ \cdots \circ \tau_k$ and let $\sigma = \tau_1 \circ \sigma_1$ and $\sigma_1 = \tau_2 \circ \sigma_2$.

$$\begin{aligned}
\triangle(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) &= \triangle(a_{\tau_1(\sigma_1(1))}, \ldots, a_{\tau_1(\sigma_1(n))}) \\
&= -\triangle(a_{\sigma_1(1)}, \ldots, a_{\sigma_1(n)}) && \text{[one transposition applied]} \\
&= (-1)\triangle(a_{\sigma_1(1)}, \ldots, a_{\sigma_1(n)}) \\
&= (-1)^2\triangle(a_{\sigma_2(1)}, \ldots, a_{\sigma_2(n)}) && \text{[two transpositions applied]} \\
&= \ldots \\
&= (-1)^k\triangle(a_1, \ldots, a_n) \\
&= \operatorname{sign}(\sigma)\triangle(a_1, \ldots, a_n)
\end{aligned}$$

□

7.5 Leibniz formula for determinants

Definition 7.18 (Definition with theorem)**.** *Define the determinant of matrix A.*

$$\triangle(a_1, \dots, a_n) = \triangle(b_1, \dots, b_n) \cdot \det A$$

if $a_j = \sum_{i=1}^n a_{ij} b_i$. Hence,

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = \Phi_B(a_j)$$

Let $\dim V = n$*. Let $B = (b_1, \dots, b_n)$ be a basis of V. $a_1, \dots, a_n \in V$ with coordinates*

$$\Phi_B(a_j) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} \qquad A := \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

Then $\triangle(a_1, \dots, a_n) = \det(A) \cdot \triangle(b_1, \dots, b_n)$ where

$$\det(A) := \sum_{\pi \in \sigma_n} \operatorname{sign}(\pi) a_{1\pi(1)} a_{2\pi(2)} \dots a_{n\pi(n)}$$

is called determinant of A

This formula was discovered by Leibniz.

Example. *Consider $n = 2$.*

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \underbrace{a_{11}a_{22}}_{\pi = \mathrm{id}} - \underbrace{a_{12}a_{21}}_{\pi = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}}$$

Proof.

$$a_j = \sum_{i=1}^n a_{ij} b_i$$

$$\triangle(a_1, \dots, a_n) = \triangle\left(\sum_{i_1=1}^n a_{i_1,1} b_{i_1}, \sum_{i_2=1}^n a_{i_2,2} b_{i_2}, \dots, \sum_{i_n=1}^n a_{i_n,n} b_{i_n} \right)$$

because it is multilinear

$$= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_n=1}^n a_{i_1,1} a_{i_2,2} \dots a_{i_n,n} \cdot \triangle(b_{i_1}, b_{i_2}, \dots, b_{i_n})$$

where $\triangle$ is zero if $b_i = b_j$.

$$\implies i_1, \dots, i_n \text{ are all different elements in } \{1, \dots, n\}$$

$$
\begin{gathered}
\implies \text{every element occurs exactly once} \\
i_1, \ldots, i_n \text{ is permutation of } 1, \ldots, n \\
\exists \sigma \in \sigma_n : i_1 = \sigma(1), \ldots, i_n = \sigma(n)
\end{gathered}
$$

$$
\begin{aligned}
&= \sum_{\sigma \in \sigma_n} a_{\sigma(1)1} a_{\sigma(2)2} \ldots a_{\sigma(n)n} \qquad \underbrace{\triangle(b_{\sigma(1)} \ldots b_{\sigma(n)})}_{\text{sign } \sigma \cdot \triangle(b_1, \ldots, b_n) \text{ because of Corollary 7.17}} \\
&= \sum_{\pi \in \sigma_n} a_{\pi(1)1} \ldots a_{\pi(n)n} \cdot \operatorname{sign}(\pi) \triangle(b_1, \ldots, b_n)
\end{aligned}
$$

□

Corollary 7.19. *A determinant form is uniquely defined by the value $\triangle(b_1, \ldots, b_n)$ on a basis. Especially, $\triangle(b_1, \ldots, b_n) \not\equiv 0 \iff \triangle(b_1, \ldots, b_n) \neq 0$ [for any basis] $\iff \triangle(b_1, \ldots, b_n) \neq 0$ [for every basis].*

Proof. Assume $\triangle(b_1, \ldots, b_n) = 0$ for any basis. Every other basis can be expressed by $b_1, \ldots, b_n$ and the formula gives $\triangle(a_1, \ldots, a_n) = 0 \quad \forall a_1, \ldots, a_n$. □

> ↓ *This lecture took place on 2018/03/12.*

Theorem 7.20.

$$\triangle \text{ non-trivial} \iff \triangle(b_1, \ldots, b_n) \neq 0 \text{ for every basis}$$

Theorem 7.21. *Inverse of Definition 7.18. Given basis $B = (b_1, \ldots, b_n)$.*

$$\triangle(a_1, \ldots, a_n) := \det\left[\Phi_B(a_1), \ldots, \Phi_B(a_n)\right]$$

defines a non-trivial determinant form such that $\triangle(b_1, \ldots, b_n) = 1$.

Example. *Let $a_1 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$ and $a_2 = \begin{pmatrix} 12 \\ 10 \end{pmatrix}$ with $A = (a_1, a_2)$. Let $b_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $b_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ with $B = (b_1, b_2)$. So $\Phi_B(A) = \begin{pmatrix} 2 & 6 \\ 0 & 2 \end{pmatrix}$.*

Let $\triangle$ such that

$$\triangle(\begin{pmatrix} 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 12 \\ 10 \end{pmatrix}) \stackrel{!}{=} \det(\Phi_B(A)) = -8$$

where -8 is given by Leibniz' formula.

$$\implies \triangle(B) = 1$$

Namely,

$$\triangle(M) = \det(M) \cdot \frac{1}{4}$$

All determinants distinguish each other by some factor.

Remark. *Let* $\begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}$ *and* $\begin{pmatrix}0\\1\\0\end{pmatrix}$ *be some basis B. Let* $v = \begin{pmatrix}7\\3\\-5\end{pmatrix}$*. Then* $\Phi_B(v) = \begin{pmatrix}-12\\7\\15\end{pmatrix}$ *is the representation of v with basis B.*

Corollary 7.22. *Let* $\triangle$ *be a non-trivial determinant form* $\triangle(v_1, \dots, v_n) \neq 0 \iff$ *Then* $v_1, \dots, v_n$ *is linearly independent.*

Proof. $\Rightarrow$ Immediate, because $v_1, \dots, v_n$ is a basis.

$\Leftarrow$ Assume $v_1, \dots, v_n$ is linearly independent. Without loss of generality, $v_n = \sum_{k=1}^{n-1} \lambda_k v_k$.

$$\begin{aligned}
\triangle(v_1, \dots, v_n) &= \triangle(v_1, \dots, v_{n-1}, \sum_{k=1}^{n-1} \lambda_n v_k) \\
&= \sum_{k=1}^{n-1} \lambda_k \triangle \underbrace{(v_1, \dots, v_{n-1}, v_k)}_{=0 \text{ because } v_k \text{ occurs twice}} \\
&= 0
\end{aligned}$$

□

Remark (Summary).

1. *The determinant form defines a 1-dimensional vector space.*
2. *There exists a non-trivial determinant form. Given a basis* $b_1, \dots, b_n$

$$\triangle(b_1, \dots, b_n) = \mathbf{1}$$

By Theorem 7.21, $\triangle(a_1, \dots, a_n) = \det(\Phi_B(a_1), \dots, \Phi_B(a_n))$.

Proof of Theorem 7.21.

1.

$$\begin{aligned}
\triangle(a_1, \dots, \lambda a_k, \dots, a_n) &= \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots \lambda a_{\pi(k)k} \dots a_{\pi(n)n} \\
&= \lambda \cdot \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots a_{\pi(n)n} \\
&= \lambda \cdot \triangle(a_1, \dots, a_n)
\end{aligned}$$

2.

$$\begin{aligned}
\triangle(a_1, \dots, a_k + v, \dots, a_n) &= \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots (a_{\pi(k)k} + v_{\pi(k)}) \dots a_{\pi(n)n} \\
&= \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots a_{\pi(k)k} \dots a_{\pi(n)n} \\
&+ \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots v_{\pi(k)k} \dots a_{\pi(n)n} \\
&= \triangle(a_1, \dots, a_k, \dots, a_n) + \triangle(a_1, \dots, v, \dots, a_n)
\end{aligned}$$

This proves multilinearity.

3. Let $a_k = a_l, a_{ik} = a_{il} \forall i = 1, \dots, n$. Without loss of generality, $k < l$.

$$\triangle(a_1, \dots, a_k) = \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots a_{\pi(k)k} \dots a_{\pi(l)l} \dots a_{\pi(n)n} = (\text{ref}^*)$$

Let $\tau = \tau_{kl}$, exchange of k and l.

Claim.

$$\sigma_n = \underbrace{\mathcal{A}_n}_{\substack{\text{alternating group} \\ =\{\pi \mid \text{sign}(\pi)=+1\}}} \cup \underbrace{\mathcal{A}_n \cdot \tau}_{=\{\pi\circ\tau \mid \pi\in\mathcal{A}_n\}}$$

The set of all permutations is the set of even permutations unified with the set of even permutations with one transposition applied. Thus, one transposition suffices to turn even permutations into odd permutations.

Proof. Direction $\Leftarrow$. Let $\text{sign}(\pi) = -1$.

$$\Rightarrow \pi = (\pi \circ \underbrace{\tau) \circ \tau}_{=\text{id}}$$

$\sigma = \pi \circ \tau$ has $\text{sign}(\sigma) = \text{sign}(\pi \circ \tau) = \text{sign}(\pi) \cdot \text{sign}(\tau) = (-1) \cdot (-1) = 1$.

$$\sigma \in \mathcal{A}_n \text{ and } \pi = \sigma \circ \tau$$

□

$$\begin{aligned}
(\text{ref}^*) &= \sum_{\pi \in \mathcal{A}_n} \underbrace{(-1)^\pi}_{=+1} a_{\pi(1)1} \dots a_{\pi(n)n} \\
&+ \sum_{\substack{\pi \in \mathcal{A}_n \tau \\ \pi = \sigma\circ\tau}} \underbrace{(-1)^{\text{sign}(\pi)}}_{=-1} a_{\pi(1)1} \dots a_{\pi(n)n} \\
&= \sum_{\pi \in \mathcal{A}_n} a_{\pi(1)1} \dots a_{\pi(n)n} - \sum_{\sigma \in A_n} \underbrace{a_{\sigma\circ\tau(1)1} \dots a_{\sigma\circ\tau(k)2} \dots a_{\sigma\circ\tau(l)l} \dots a_{\sigma\circ\tau(n)n}}_{a_{\sigma(1)1} \dots \underbrace{a_{\sigma(l)k}}_{=a_{\sigma(l)l}} \dots \underbrace{a_{\sigma(k)l}}_{=a_{\sigma(k)k}} \dots a_{\sigma(n)n}} = 0
\end{aligned}$$

□

This previous part, beginning with the reference from 2018/03/12, was actually added on 2018/03/14, because we skipped it by accident.

$$\triangle(a_1, \dots, a_n)$$

Determinant form $\iff$

multilinear $\triangle(a_1, \dots, \lambda a_k + \mu a'_k, \dots, a_n) = \lambda \triangle(a_1, \dots, a_k, \dots, a_n) + \mu \triangle(a_1, \dots, a'_k, \dots, a_n)$

anti-symmetrical $\triangle(a_1,\dots,a_k,\dots,a_l,\dots,a_n) = -\triangle(a_1,\dots,a_l,\dots,a_k,\dots,a_n)$

$$\triangle(a_{\pi(1)},\dots,a_{\pi(n)}) = (-1)^{\pi}\triangle(a_1,\dots,a_n)$$

where $(-1)^{\pi} := \operatorname{sign}(\pi) = (-1)^{(F(\pi))}$

$$F(\pi) = \left\{(i,j) \,\middle|\, i < j \wedge \pi(i) > \pi(j)\right\}$$

$$\operatorname{sign}(\pi \circ \sigma) = \operatorname{sign}(\pi) \cdot \operatorname{sign}(\pi) \cdot \operatorname{sign}(\sigma)$$

Basis $b_1,\dots,b_n$.

$$\triangle(\sum_{i=1}^{n} a_{i1}b_i,\dots,\sum_{i=1}^{n} a_{in}b_i) = \det A \cdot \triangle(b_1,\dots,b_n)$$

$$\det(A) = \sum_{\pi\in\sigma_n}(-1)^{\pi}a_{1\pi(1)}\dots a_{n\pi(n)} = \sum_{\pi\in\sigma_n}(-1)^{\pi}a_{\pi(1)1}\dots a_{\pi(n)n}$$

Lemma 7.23. *Let V, W be vector spaces over $\mathbb{K}$ with $\dim V = \dim W = n$. Let $\triangle : W^n \to \mathbb{K}$ be a determinant form and $f : V \to W$ linear.*

$$V \xrightarrow{f} W$$

$$V^n \xrightarrow{f^{(n)}} W^n \xrightarrow{\triangle} \mathbb{K}$$

$$(v_1,\dots,v_n) \mapsto (f(v_1),\dots,f(v_n))$$

$$\implies \triangle^f : V^n \to \mathbb{K}$$

$$\triangle^f(v_1,\dots,v_n) = \triangle(f(v_1),\dots,f(v_n))$$

is a determinant form on V.

Proof. We only need to prove multilinearity.

$$\begin{aligned}
&\triangle^f(v_1,\dots,\lambda v_k + \mu v_k',\dots,v_n)\\
&= \triangle(f(v_1),\dots,f(\lambda v_k + \mu v_k'),\dots,f(v_n))\\
&= \triangle(f(k),\dots,\lambda f(v_k) + \mu f(v_k'),\dots,f(v_k))\\
&= \lambda\triangle(f(v_1),\dots,f(v_k),\dots,f(v_n)) + \mu\triangle(f(v_1),\dots,f(v_k'),\dots,f(v_n))\\
&= \lambda\triangle^f(v_1,\dots,v_k,\dots,v_n) + \mu\triangle^f(v_1,\dots,v_k',\dots,v_n)
\end{aligned}$$

□

7.6 About determinants

Corollary 7.24. *Let $V = W$, $\triangle : V^n \to \mathbb{K}$ determinant form.*

$$f : V \to V \text{ linear } \implies \triangle^f \text{ is determinant form (Lemma 7.23)}$$

Because there is (except for one factor) only one determinant form:

$$\exists c_f \in \mathbb{K} : \triangle^f(v_1, \dots, v_n) = c_f \cdot \triangle(v_1, \dots, v_n) \forall v_1, \dots, v_n \in V$$

$$\det(f) := c_f \text{ is called determinant on } f$$

Proof. Let $\triangle_1, \triangle_2$ be two determinant forms. Let $b_1, \dots, b_n$ be a basis.

$$\triangle_1(v_1, \dots, v_n) = \det A \cdot \triangle_1(b_1, \dots, b_n)$$

$$\triangle_2(v_1, \dots, v_n) = \det A \cdot \triangle_2(b_1, \dots, b_n)$$

$$v_j = \sum_{i=1}^n a_{ij} b_i$$

$$\implies \triangle_2(v_1, \dots, v_n) = \frac{\triangle_2(b_1, \dots, b_n)}{\triangle_1(b_1, \dots, b_n)} \cdot \triangle_1(v_1, \dots, v_n)$$

$$\implies c_f = \frac{\triangle^f(b_1, \dots, b_n)}{\triangle(b_1, \dots, b_n)} =: \det(f)$$

□

Corollary 7.25. *$B = (b_1, \dots, b_n)$ is basis of V. $\phi_B^B(f)$ is matrix representation of linear f and $\det(f) = \det \phi_B^B(f)$ (LHS by Corollary 7.24, RHS by Definition 7.18 $\sum_\pi (-1)^\pi \dots$).*

Proof.

$$\det(f) = \frac{\triangle(f(b_1)), \dots, \triangle(f(b_n)))}{\triangle(b_1, \dots, b_n)}$$

$$f(b_j) = \sum_{i=1}^n \phi_B(f(b_i))_i \cdot b_i = \sum_{i=1}^n \left(\phi_B^B(f)\right)_{ij} b_i$$

with $\phi_B^B(f)_{ij} = \phi_B(f(b_j))_i$.

$$\det f = \frac{\det \phi_B^B(f) \cdot \triangle(b_1, \dots, b_n)}{\triangle(b_1, \dots, b_n)}$$

□

Theorem 7.26. *$f : V \to V$ is invertible $\iff$ $\det(f) \neq 0$.*

Proof. $\implies$ Let $\triangle$ be a non-trivial determinant form.

$$B = (b_1, \dots, b_n) \text{ is a basis } \implies \triangle(b_1, \dots, b_n) \neq 0$$

$$\det(f) = \frac{\triangle(f(b_1), \dots, f(b_n))}{\triangle(b_1, \dots, b_n)}$$

$\Longleftarrow$ $\det(f) \neq 0 \implies (f(b_1), \dots, f(b_n))$ is basis $\iff$ f is invertible.

If f is invertible then $(f(b_1), \dots, f(b_n))$ is basis.

$$\implies \triangle(f(b_1), \dots, f(b_n)) \neq 0 \implies \det(f) \neq 0$$

If f is not invertible then $f(b_1) \dots f(b_n)$ is linear dependent

$$\exists k : f(b_k) = \sum_{i \neq k} \lambda_i f(b_i)$$

Without loss of generality: $k = n$

$$\begin{aligned}
\triangle(f(b_1), \dots, f(b_n)) &= \triangle(f(b_1), \dots, f(b_{n-1}), \sum_{i=1}^{n-1} \lambda_i f(b_i)) \\
&= \sum_{i=1}^{n} \lambda_i \triangle(\underbrace{f(b_1), \dots, f(b_{n-1})}_{=0 \forall i \in \{1, \dots, n-1\}}, f(b_i)) \\
&= 0
\end{aligned}$$

□

Corollary 7.27. *For a matrix* $A \in \mathbb{K}^{n \times n}$ *it holds that* $\det A \neq 0 \iff$ *A has full rank.*

Proof. $\implies$ If A is invertible $\ker(A) = \{0\}$, so A has full rank.

$\Longleftarrow$ If A does not have full rank, consider $x \neq 0$ with $x \in \ker(A)$ then $Ax = 0$ and $A(2x) = 0$. Thus it is not injective and therefore not invertible.

Remark. *If A has full rank it is surjective (column space spans all n dimensions) and injective* ($x \neq y \implies Ax \neq Ay$)*. Thus invertible.*

□

Theorem 7.28. $f, g : V \to V$ *linear.*

$$\implies \det(f \circ g) = \det(f) \cdot \det(g)$$

for a matrix: $\det(A \cdot B) = \det(A) \cdot \det(B)$

Proof. **If f and g are invertible**

$$\det(f) = \frac{\triangle(f(b_1), \dots, f(b_n))}{\triangle(b_1, \dots, b_n)}$$

for arbitrary bases $(b_1, \dots, b_n)$ of V.

$$\begin{aligned}
\det(f \circ g) &= \frac{\triangle(f(g(b_1)), \dots, f(g(b_n)))}{\triangle(b_1, \dots, b_n)} \cdot \frac{\triangle(g(b_1), \dots, g(b_n))}{\triangle(g(b_1), \dots, g(b_n))} \\
&= \underbrace{\frac{\triangle(f(g(b_1)), \dots, f(g(b_n)))}{\triangle(g(b_1), \dots, g(b_n))}}_{\det(f) \neq 0} \cdot \underbrace{\frac{\triangle(g(b_1), \dots, g(b_n))}{\triangle(b_1, \dots, b_n)}}_{\det(g) \neq 0}
\end{aligned}$$

If f or g is not invertible

$$f \text{ is not invertible} \implies \det(f) = 0$$

Same for g.

Claim. *$f \circ g$ invertible $\iff$ f invertible and g invertible.*

$f \circ g$ invertible $\implies$ $f \circ g$ surjective $\implies$ f surjective $\implies$ ($\dim V < \infty$) f is bijective.

$f \circ g$ invertible $\implies$ $f \circ g$ injective $\implies$ g injective $\implies$ g bijective.

$$\implies f \circ g \text{ is not invertible}$$
$$\det(f \circ g) = 0 = \det(f) \cdot \det(g)$$

□

7.7 Laws of determinants

Corollary 7.29. *For $A, B \in \mathbb{K}^{n \times n}$ it holds that*

1. $\det(A \cdot B) = \det(A) \cdot \det(B)$
2. $\det(A^{-1}) = \frac{1}{\det(A)}$ *if invertible*
3. $\det(A) = 0 \iff \operatorname{rank}(A) < n$
4. $\det(A^t) = \det(A)$

Proof of Corollary 7.29. 1. $\det(A \cdot B) = \det(f_A \circ f_B) = \det(f_A) \cdot \det(f_B) = \det(A) \cdot \det(B)$ (compare with Corollary 7.25)

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

Remark (From the practicals)**.**

$$\det(A) = \det(f_A)$$

Shown so far:

$$\det f = \det\left(\phi_B^B(f)\right)$$
$$A = \phi_B^B\left(f_A\right)$$

for $B = (e_1, \dots, e_n)$

□

Direct proof of Corollary 7.29 (1).

$$A = \begin{bmatrix} s_1 & \dots & s_n \\ \vdots & & \vdots \end{bmatrix}$$

s_i are column vectors of A. Let $\triangle$ be uniquely defined det. form by $\triangle(e_1, \dots, e_n) = 1$.

$$A \cdot B = \begin{bmatrix} s_1 & \dots & s_n \\ \vdots & & \vdots \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ \vdots & & & \vdots \\ b_{n1} & & & b_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} s_1 b_{11} + s_2 b_{21} + \dots + s_n b_{n1} & s_1 b_{12} + s_2 b_{22} + \dots + s_n b_{n2} & \dots & s_1 b_{1n} + s_2 b_{2n} + \dots + s_n b_{nn} \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

$$\det(A \cdot B) = \frac{\triangle(s_1(A \cdot B), \dots, s_n(A \cdot B))}{\triangle(e_1, \dots, e_n)} = \triangle\left(\sum_{i_1=1}^{n} s_{i_1} b_{i_1 1}, \sum_{i_2=1}^{n} s_{i_2} b_{i_2 2}, \dots, \sum_{i_n=1}^{n} s_{i_n} b_{i_n n}\right)$$

$$= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} b_{i_1 1} b_{i_2 2} \dots b_{i_n n} \underbrace{\triangle(s_{i_1}, \dots, s_{i_n})}_{=0}$$

if one index occurs twice. It suffices to consider $\sum_{i_1, \dots, i_n}$ such that all ij are different. If all are different, then all occur exactly once. Hence, $i_1, \dots, i_n$ is permutation of $1, \dots, n$.

$$= \sum_{\pi \in \sigma_n} b_{\pi(1)1} \dots b_{\pi(n)n} \triangle(s_{\pi(1)} \dots s_{\pi(n)})$$
$$= \sum_{\pi \in \sigma_n} \underbrace{(-1)^{\pi} b_{\pi(1)1} \dots b_{\pi(n)n}}_{=\det(B)} \underbrace{\triangle(s_1, \dots, s_n)}_{=\det(A)} = \det(B) \cdot \det(A)$$

□

Proof of Corollary 7.29 (3). A is not invertible $\iff$ f_A is not invertible. $\implies$ $\det(A) = 0 \iff \det(f_A) = 0 \iff f_A$ is not bijective $\iff \operatorname{rank}(A) < n$. □

Proof of Corollary 7.29 (4).

$$\det(A^t) = \sum_{\pi \in \sigma_n} (-1)^{\pi} (A^t)_{\pi(1)1} \dots (A^t)_{\pi(n)n}$$
$$= \sum_{\pi \in \sigma_n} (-1)^{\pi} a_{1\pi(1)} \dots a_{n\pi(n)}$$

Remark.

$$\sigma_n \to \sigma_n$$
$$\pi \mapsto \pi^{-1} \textit{ is bijective}$$
$$\textit{injective: } \pi^{-1} = \sigma^{-1} \implies \pi = \sigma$$
$$\textit{surjective: } \pi = (\pi^{-1})^{-1}$$

$$= \sum_{\pi \in \sigma_n} (-1)^{\pi^{-1}} a_{1\pi^{-1}(1)} \dots a_{n\pi^{-1}(n)}$$

Every index i occurs once on the left side and once on the right side. i occurs right

$$\pi^{-1}(j) = i \iff j = \pi(i)$$

$$= \sum_{\pi \in \sigma_n} (-1)^{\pi} a_{\pi(1)1} \dots a_{\pi(n)n}$$

$$\operatorname{sign}(\pi \circ \pi^{-1}) = 1$$
$$= \operatorname{sign}(\pi) \cdot \operatorname{sign}(\pi^{-1})$$

Remark (A small exercise)**.**

$$\det(A) = \det(f_A)$$

$$\prod_{j=1}^{n} a_{j,\pi^{-1}(j)} = \prod_{i=1}^{n} a_{\pi(i),\pi^{-1}(\pi(i))} = \prod_{i=1}^{n} a_{\pi(i),i} \qquad j := \pi(i)$$

□

Definition 7.30.

$$\operatorname{perm}(A) := \sum_{\pi \in \sigma_n} a_{\pi(1)1} \dots a_{\pi(n)n}$$

is called permanent of A.

Open problem: for which matrix does $\operatorname{perm}(A) = 0$ *hold?*

Example 7.31 (Computation of the determinant)**.**

$$\dim \leq 3$$

$$n = 2 : \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$n = 3 : \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{2}3 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \sum_{\sigma \in \sigma_n} (-1)^{\pi} a_{\pi(1)1} a_{\pi(2)2} a_{\pi(3)3}$$

By the Cayley-Graph of group σ_3 *we can see that* $\sigma_3 = \left\langle (\underline{12}), (\underline{\underline{23}}) \right\rangle = -1$.

$$\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}$$

Rule of Sarrus *holds only for* $n = 2$ *or* $n = 3$.

↓ *This lecture took place on 2018/03/14.*

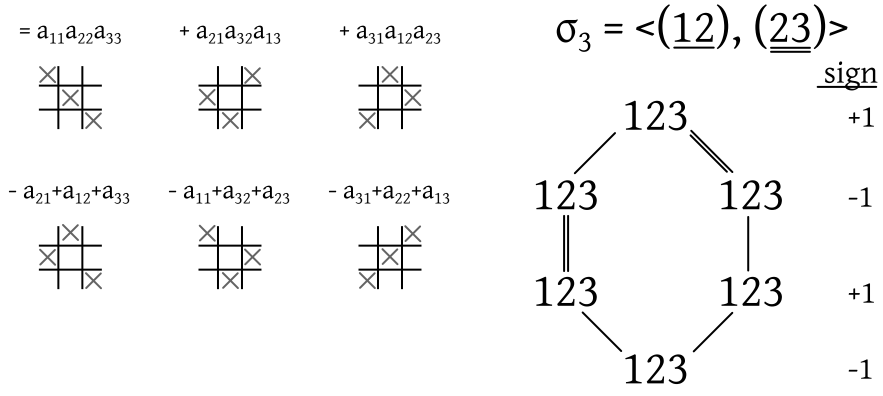


Figure 3: Cayley graph (right) and permutation factors in Sarrus' Rule (left) in the 3D case

Example (Rule by Sarrus). *Let $n = 2$:*

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Let $n = 3$:

$$\begin{vmatrix} 1 & 2 & 5 & 1 & 2 \\ 2 & 5 & 14 & 2 & 5 \\ 5 & 14 & 42 & 5 & 14 \end{vmatrix} = 1$$

$$\begin{aligned} &1 \cdot 5 \cdot 42 + 2 \cdot 14 \cdot 5 + 5 \cdot 2 \cdot 14 - 5 \cdot 5 \cdot 5 - 1 \cdot 14 \cdot 14 - 2 \cdot 2 \cdot 42 \\ &= 14 \cdot (1 \cdot 5 \cdot 3 + 2 \cdot 5 + 5 \cdot 2) - 125 - 14 \cdot (14 + 2 \cdot 2 \cdot 3) \\ &= 14 \cdot 35 - 125 - 14 \cdot 26 \\ &= 14 \cdot 9 - 125 = 1 \end{aligned}$$

An error in the computation will be enhanced.

Let $n = 4$. $|\sigma_n| = 24$ makes consideration of all permutations impractical.

Lemma 7.32. *Let A be an upper triangular matrix, hence $a_{ij} = 0$ if $i > j$.*

$$\implies \det(A) = a_{11}a_{22}\dots a_{nn}$$

Proof.

$$\det(A) = \sum_{\pi \in \sigma_n} (-1)^\pi a_{\pi(1)1} \dots a_{\pi(n)n}$$

such that $\pi(j) \leq j \forall j$.

$$\implies \text{id}$$

$$
\begin{aligned}
\pi(j) \le j \forall j \implies \pi(1) &\le 1 \implies \pi(1) = 1 \\
\pi(2) &\le 2 \implies \pi(2) = 2 \\
\pi(3) &\le 3 \implies \pi(3) = 3 \\
&\dots \\
\pi(n) &\le n \implies \pi(n) = n
\end{aligned}
$$

□

Theorem 7.33. *Let $A = (a_{ij})$ be a $n \times n$ matrix.*

1. *Let $z_1, \dots, z_n$ be row vectors of A. Then*

$$\det \begin{bmatrix} z_1 & \dots \\ \vdots & \\ z_n & \dots \end{bmatrix} = \det \begin{bmatrix} z_1 & \dots \\ z_i + \lambda z_j & \dots \\ \vdots & \\ z_n & \dots \end{bmatrix} \forall i \neq j, \lambda \in \mathbb{K}$$

2. *Let $S_1, \dots, S_n$ be columns of A. Then,*

$$\det \begin{pmatrix} S_1 & \dots & S_n \\ \vdots & & \vdots \end{pmatrix} = \det \begin{pmatrix} S_1 & \dots & S_i + \lambda S_j & \dots & S_j & \dots & S_n \\ \vdots & & \vdots & & \vdots & & \vdots \end{pmatrix}$$

Proof. 2. Proof for column i:

$$\triangle(s_1, \dots, s_n) = \triangle(s_1, \dots, s_i + \lambda s_j, \dots, s_n)$$

$$\left(= \triangle(s_1, \dots, s_i, \dots, s_n) + \lambda \underbrace{\triangle(s_1, \dots, s_j, \dots, s_j, \dots, s_n)}_{=0} \right)$$

1. Second proof Row form is multiplication from left with matrix of structure

$$I + \lambda E_{ij} \qquad i \neq j$$

$$\det((I + \lambda E_{ij})A) = \underbrace{\det(I + \lambda E_{ij})}_{\text{triangular matrix}=1} \cdot \det(A)$$

□

Example 7.34.

$$\begin{vmatrix} 1 & 2 & 5 \\ 2 & 5 & 14 \\ 5 & 14 & 42 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 5 \\ 0 & 1 & 4 \\ 0 & 4 & 17 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

Example.

$$\begin{vmatrix} 1 & 0 & 3 & -2 \\ 2 & 6 & 4 & 1 \\ 3 & 3 & -1 & -1 \\ -1 & 2 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 3 & -2 \\ 0 & 6 & -2 & 5 \\ 0 & 3 & -10 & 5 \\ 0 & 2 & 7 & -1 \end{vmatrix}$$

$$= \frac{1}{3}\frac{1}{2}\begin{vmatrix} 1 & 0 & 3 & -2 \\ 0 & 6 & -2 & 5 \\ 0 & 6 & -20 & 10 \\ 0 & 6 & 21 & -3 \end{vmatrix} = \frac{1}{6}\begin{vmatrix} 1 & 0 & 3 & -2 \\ 0 & 6 & -2 & 5 \\ 0 & 0 & -18 & 5 \\ 0 & 0 & 23 & -8 \end{vmatrix} = \frac{1}{6}\cdot 6\begin{vmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & -2 & 5 \\ 0 & 0 & -18 & 5 \\ 0 & 0 & 23 & -8 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 8 & 5 \\ 0 & 0 & -8 & 5 \\ 0 & 0 & 7 & -8 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 8 & 5 \\ 0 & 0 & -8 & 5 \\ 0 & 0 & -1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 8 & 5 \\ 0 & 0 & 0 & 29 \\ 0 & 0 & -1 & -3 \end{vmatrix}$$

$$-\begin{vmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 8 & 5 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 29 \end{vmatrix} = 29$$

Remark (Laws, discussed so far)**.**

$$\begin{vmatrix} z_1 & \dots \\ \lambda \cdot z_1 & \dots \\ z_n & \dots \end{vmatrix} = \lambda \begin{vmatrix} z_1 & \dots \\ \vdots & \\ z_k & \dots \\ \vdots & \\ z_n & \dots \end{vmatrix}$$

$$\begin{vmatrix} z_1 & \dots \\ z_1 + \lambda z_j & \dots \\ z_n & \dots \end{vmatrix} = \begin{vmatrix} z_1 & \dots \\ \vdots & \\ z_i & \dots \\ \vdots & \\ z_n & \dots \end{vmatrix} \qquad (i \neq j)$$

$$\begin{vmatrix} z_1 & \dots \\ \vdots & \\ z_i & \dots \\ \vdots & \\ z_j & \dots \\ \vdots & \\ z_n & \dots \end{vmatrix} = -\begin{vmatrix} z_1 & \dots \\ \vdots & \\ z_j & \dots \\ \vdots & \\ z_i & \dots \\ \vdots & \\ z_n & \dots \end{vmatrix}$$

$$\begin{vmatrix} a_{11} & \dots & & & \\ & a_{22} & \dots & & \\ & & a_{33} & \dots & \\ & & & \ddots & \\ 0 & & & & a_{nn} \end{vmatrix} = a_{11} \cdot a_{nn}$$

Lemma 7.35. *1.*

$$\begin{vmatrix} a_{11} & * & * & * \\ 0 & & & \\ 0 & & B & \\ 0 & & & \end{vmatrix} = a_{11} \cdot \det(B)$$

2.

$$\begin{vmatrix} & & & 0 \\ & B & & 0 \\ & & & 0 \\ * & * & * & a_{nn} \end{vmatrix} = \det(B) \cdot a_{nn}$$

3. *If there are individual square matrices ($A_1, A_2, \ldots, A_k$) along the diagonal of a matrix, the determinant of the matrix is the product of the determinant of the submatrices.*

$$\det(A) = \det(A_1) \cdot \det(A_2) \cdot \ldots \cdot \det(A_k)$$

Proof. We only prove the second property.

All permutations which do not map index 1 to 1, introduce a factor zero making the product zero. If index 1 is mapped to 1, the product in Leibniz' formula is multiplied with a_{11} in all permutations. We can extract factor a_{11} and get the determinant of B multiplied with a_{11}.

$$\begin{aligned} \begin{vmatrix} & & & 0 \\ B & & & \vdots \\ & & & 0 \\ a_{n,1} & \ldots & a_{n,n-1} & a_{n,n} \end{vmatrix} &= \sum_{\pi \in \sigma_n} (-1)^{\pi} a_{\pi(1)1} \ldots a_{\pi(n)n} \\ &= \sum_{\pi' \in \sigma_{n-1}} (-1)^{\pi'} a_{\pi'(1)1} \ldots a_{\pi'(n-1)n-1} \cdot a_{nn} \\ &= \det(B) \cdot a_{nn} \end{aligned}$$

$$\{\pi \in \sigma_n \mid \pi(n) = n\}$$

$$\pi(n) = n$$

$$B = \begin{pmatrix} a_{11} & \ldots & a_{1,n-1} \\ \vdots & & \\ a_{n-1,1} & \ldots & a_{n,n-1} \end{pmatrix}$$

Same idea: If

$$A = \begin{bmatrix} & 0 & \\ & \vdots & \\ & 0 & \\ \vdots & a_{ij} & \vdots \\ & 0 & \\ & \vdots & \\ & 0 & \end{bmatrix}$$

Exchange the i-th row with the last row.

$$= \pm 1 \begin{vmatrix} & 0 \\ & \vdots \\ & 0 \\ \cdots & 0 \\ & 0 \\ & \vdots \\ & a_{ij} \end{vmatrix}$$

□

Definition 7.36.

$$A \in \mathbb{K}^{n \times n}$$

$A_{k,l}$ is an $(n-1) \times (n-1)$ matrix, that is created by omitting the k-th row and l-th column.

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,l-1} & a_{1,l+1} & \dots & a_{1,n} \\ \vdots & & & & & \vdots \\ a_{k-1,1} & \dots & a_{k-1,l-1} & a_{k-1,l+1} & \dots & a_{k-1,n} \\ a_{k+1,1} & \dots & a_{k+1,l-1} & a_{k+1,l+1} & \dots & a_{k+1,n} \\ \vdots & & & & & \vdots \\ a_{n,1} & \dots & a_{n,l-1} & a_{n,l+1} & \dots & a_{n,n} \end{bmatrix}$$

Person. *Pierre-Simon Laplace (1749–1827)*

Definition 7.37 (Laplace expansion)**.** *In German, this theorem is called Entwicklungssatz von Laplace.*

Let l be fixed.

$$\det(A) = \sum_{k=1}^{n} a_{kl}(-1)^{k+l} \det(A_{kl}) \qquad \text{“Expansion along column } l\text{”}$$

Let k be fixed.

$$\det(A) = \sum_{l=1}^{n} a_{kl}(-1)^{k+l} \det(A_{kl}) \qquad \text{“Expansion along row } k\text{”}$$

Example 7.38.

$$\begin{vmatrix} 1 & 2 & 5 \\ 2 & 5 & 14 \\ 5 & 14 & 42 \end{vmatrix} = \sum_{l=1}^{3} (-1)^{1+l} \det(A_{1l}) \qquad \text{for } k = 1 \text{ fixed}$$

$$= 1 \begin{vmatrix} 5 & 14 \\ 14 & 42 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 14 \\ 5 & 42 \end{vmatrix} + 5 \cdot \begin{vmatrix} 2 & 5 \\ 5 & 14 \end{vmatrix}$$

$$= 1 \cdot (5 \cdot 42 - 14 \cdot 14) - 2(2 \cdot 42 - 5 \cdot 14) + 5 \cdot (2 \cdot 14 - 5 \cdot 9)$$

$$= 1 \cdot (5 \cdot 3 \cdot 14 - 14 \cdot 14) - 2 \cdot (2 \cdot 3 \cdot 13 - 5 \cdot 14)$$

$$= 14 - 2 \cdot 14 + 5 \cdot 15 = 1$$

Consider $k = 2$.

$$-2 \cdot \begin{vmatrix} 2 & 5 \\ 14 & 42 \end{vmatrix} + 5 \cdot \begin{vmatrix} 1 & 5 \\ 5 & 42 \end{vmatrix} - 14 \cdot \begin{vmatrix} 1 & 2 \\ 5 & 14 \end{vmatrix}$$
$$= -2(3 \cdot 14 \cdot 2 - 14 \cdot 5) + 5 \cdot (42 - 25) - 14 \cdot (14 - 10)$$
$$= -2 \cdot 14 + 5 \cdot 17 - 4 \cdot 14 = -28 + 85 - 56 = 85 - 84 = 1$$

↓ *This lecture took place on 2018/03/19.*

Review:

- Determinants are multilinear (in rows and columns)
- Determinants switches its sign if two rows or row columns are exchanged
- $\triangle(s_1, \dots, s_n) = (-1)^\pi \triangle(s_{\pi(1)}, \dots, s_{\pi(n)})$ where s_i are column vectors
- $$\begin{vmatrix} a_{11} & 0 & \dots & 0 \\ * & & & \\ \vdots & & B & \\ * & & & \end{vmatrix} = a_{11} \cdot \det B \qquad B = A_{11}$$

  where A_{kl} is the $(n-1) \times (n-1)$ matrix created by removal of the k-th row and l-th column. This is a special case of Laplace expansion.

7.8 Laplace expansion

$$\begin{aligned} \det A &= \sum_{k=1}^n (-1)^{k+l} a_{kl} \cdot \det A_{kl} \qquad && \text{for fixed } l \in \{1, \dots, n\} \\ &= \sum_{l=1}^n (-1)^{k+l} a_{kl} \cdot \det A_{kl} && \text{for fixed } k \in \{1, \dots, n\} \end{aligned}$$

So in the case of (a very classic example)

$$\begin{vmatrix} a_{11} & 0 & \dots & 0 \\ * & & & \\ \vdots & & B & \\ * & & & \end{vmatrix} = a_{11} \cdot (-1)^{1+1} \cdot \det A_{11}$$

for fixed $k = 1$:

$$\sum_{l=1}^n (-1)^{1+l} \underbrace{a_{1l}}_{=0 \text{ for } l>1} \det A_{1l}$$

Proof. Let $l \in \{1, \ldots, n\}$ be fixed. Let e_k be a unit vector. For the l-th column,

$$s_l = \sum_{k=1}^{n} a_{kl} e_k = \begin{pmatrix} a_{1l} \\ a_{2l} \\ \vdots \\ a_{nl} \end{pmatrix}$$

$$\begin{aligned}
\det(A) &= \triangle(s_1, s_2, \ldots, s_{l-1}, \sum_{k=1}^{n} a_{kl} \cdot e_k, s_{l+1}, \ldots, s_n) \\
&= \sum_{k=1}^{n} a_{kl} \triangle(s_1, \ldots, s_{l-1}, e_k, s_{l+1}, \ldots, s_n) \\
&= \sum_{k=1}^{n} a_{kl} \begin{vmatrix}
a_{11} & a_{12} & \vdots & a_{1,l-1} & 0 & a_{1,l+1} & \ldots & a_{1n} \\
a_{21} & a_{22} & \vdots & a_{2,l-1} & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & \vdots & \vdots & 0 & \vdots & & \vdots \\
\vdots & \vdots & \vdots & \vdots & 1 & \vdots & & \vdots \\
\vdots & \vdots & \vdots & \vdots & 0 & \vdots & & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \vdots & a_{n,l-1} & 0 & a_{n,l+1} & \ldots & a_{nn}
\end{vmatrix}
\end{aligned}$$

Recognize the one in row k. We consecutively exchange row k with the row above until it becomes row 1. This gives $k-1$ exchanges. Hence a cycle $(1 \ldots k)$. This gives sign $= (-1)^{k-1}$.

$$= \sum_{k=1}^{n} a_{kl} (-1)^{k-1} \begin{vmatrix}
a_{k1} & a_{k2} & \ldots & a_{k,l-1} & 1 & a_{k,l+1} & \ldots & a_{kn} \\
a_{11} & a_{12} & \ldots & & 0 & & & a_{1n} \\
\vdots & \vdots & \ldots & & 0 & & & \vdots \\
a_{k-1,1} & a_{k-1,2} & \ldots & & 0 & & & a_{k-1,n} \\
a_{k+1,1} & a_{k+1,2} & \ldots & & 0 & & & a_{k+1,n} \\
\vdots & \vdots & \ldots & & 0 & & & \vdots \\
a_{n1} & a_{n2} & \ldots & & 0 & & & a_{nn}
\end{vmatrix}$$

Now we can do $l-1$ column exchanges to move the one into the first column. This gives a cycle $(1, 2, \ldots, l)$ and sign $= (-1)^{l-1}$

$$= \sum_{k=1}^{n} a_{kl} (-1)^{k-1} (-1)^{l-1} \begin{vmatrix}
1 & a_{k1} & a_{k2} & \ldots & a_{k,l-1} & a_{k,l+1} & \ldots & a_{k,n} \\
0 & a_{11} & a_{12} & \ldots & a_{1,l-1} & a_{1,l+1} & \ldots & a_{1,n} \\
0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & a_{2,n} \\
0 & a_{k-1,1} & a_{k-1,2} & \ldots & a_{k-1,l-1} & a_{k-1,l+1} & \ldots & a_{k-1,n} \\
0 & a_{k+1,1} & a_{k+1,2} & \ldots & a_{k+1,l-1} & a_{k+1,l+1} & \ldots & a_{k+1,n} \\
0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & a_{2,n} \\
0 & a_{n1} & a_{n2} & \ldots & a_{nl-1} & a_{nl+1} & \ldots & a_{nn}
\end{vmatrix}$$

where the k-th row and l-th column is removed

$$= \sum_{k=1}^{n} (-1)^{k+l} a_{kl} \det A_{kl}$$

□

Example 7.39. *Let* $A = (a_{kl})_{\substack{1 \le k \le n \\ 1 \le l \le n}} = (-1)^{k+l}$.

$$A = \begin{pmatrix} +1 & -1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & +1 \end{pmatrix}$$

Theorem 7.40. $\hat{a}_{kl} := (-1)^{k+l} \det A_{lk}$ *is called* cofactor.

$$\hat{A} = \left[\hat{a}_{kl}\right]_{k,l=1}^{n}$$

is called complementary matrix *or* adjugate matrix *of* A.

$$\begin{aligned} \hat{a}_{kl} &= (-1)^{k+l} \det \text{ (the matrix without row } l \text{ and column } k\text{)} \\ &= (-1)^{k+l} \det A_{lk} = \frac{\partial}{\partial a_{lk}} \det A \end{aligned}$$

Then it holds that

$$A^{-1} = \frac{1}{\det A} \hat{A}$$

Proof. Show that $\hat{A} \cdot A = I \cdot \det(A)$. Let $B = \hat{A} \cdot A$.

$$b_{kl} = \sum_{i=1}^{n} \hat{a}_{ki} \cdot a_{il} = \sum_{i=1}^{n} (-1)^{k+i} \det A_{ik} \cdot a_{il}$$

Case $\mathbf{k = l}$

$$b_{ll} = \sum_{i=1}^{n} (-1)^{l+i} \det A_{il} \cdot a_{il} \underbrace{=}_{\substack{\text{Laplace expansion} \\ \text{with } l\text{-th column}}} \det A$$

Case $\mathbf{k \neq l}$ Without loss of generality: $k < l$.

$$b_{kl} = \sum_{i=1}^{n} \det(A_{ik})(-1)^{k+i} a_{il}$$

unlike Laplace expansion along k-th column, this expression uses column l. So column k is omitted and column l is used twice.

$$= \det \begin{bmatrix} a_{11} & \dots & a_{1l} & \dots & a_{1l} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nl} & \dots & a_{nl} & \dots & a_{nn} \end{bmatrix}$$

where the left l-column is column k replaced with values of column l and the right l-column is the original column l

$$\underbrace{=}_{\text{two equal columns}} 0$$

Thus, the sum of both cases is $b_{kl} = \det(A) + 0$. Thus $B = \hat{A} \cdot A = I \cdot \det(A)$.

□

Example 7.41 (Small inverse matrices). *Let $n = 2$.*

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

$$\hat{a}_{11} = (-1)^{1+1} \cdot \det A_{11} \qquad \hat{a}_{21} = (-1)^{2+1} \cdot \det A_{12}$$
$$\hat{a}_{12} = (-1)^{1+2} \cdot \det A_{21} \qquad \hat{a}_{22} = (-1)^{2+2} \cdot \det A_{22}$$

Remark (Cayley 1855).

$$A^{-1} = \frac{1}{\nabla} \begin{bmatrix} \partial_a \nabla & \partial_c \nabla \\ \partial_b \nabla & \partial_d \nabla \end{bmatrix} \qquad \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Example. *Let $n = 3$.*

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}$$

Corollary 7.42. *Let $A \in \mathbb{Z}^{n \times n}$. If $\det A = 1 \implies A^{-1} \in \mathbb{Z}^{n \times n}$.*
Let $A \in \mathbb{Z}^{n \times n}$ and $\det A = 1$. Let $B \in \mathbb{Z}^{n \times n}$ and $\det B = 1$.

$$\implies \det(A \cdot B) = 1 \qquad \implies \det(A^{-1}) = 1$$

Definition 7.43. *Integer matrices with* $\det = 1$ *define a group called* special linear group.

$$\mathrm{SL}(n, \mathbb{Z}) = \left\{ A \in \mathbb{Z}^{n \times n} \,\middle|\, \det A = 1 \right\}$$

Or in general for a ring R:

$$\mathrm{SL}(n, R) = \left\{ A \in R^{n \times n} \,\middle|\, \det A = 1 \right\}$$

7.9 Cramer's Rule

Person. *Gabriel Cramer (1704–1752)*

Theorem 7.44 (Cramer's Rule)**.** *Show by Cramer in 1750, by McLaurin 1748 for $n \leq 3$.*

Let A be a invertible matrix with column vectors $a_1, \dots, a_n$. Then the solution $Ax = b$ ($\implies$ $x = A^{-1}b$ has a unique solution) is given by

$$x_i = \frac{\triangle(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n)}{\triangle(a_1, \dots, a_n)}$$
$$= \frac{\det\left(\begin{bmatrix} a_1 & \dots & a_{i-1} & b & a_{i+1} & \dots & a_n \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}\right)}{\det A}$$

$n+1$ determinants of form $n \times n$. In practice infeasible except for small matrices.

Geometrical proof for $n = 2$.

$$A = \begin{pmatrix} a_1 & a_2 \\ \vdots & \vdots \end{pmatrix}$$
$$Ax = b \qquad a_1 \cdot x + a_2 \cdot x_2 = b$$
$$\triangle(a_1, a_2) = A(a_1, a_2)$$

where A is the area function. Compare with Figure 4.

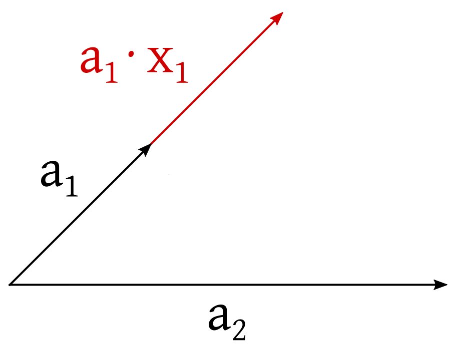


Figure 4: Geometrical proof for $n = 2$. $\triangle(a_1, a_2)$ is the area of the parallelogram

$$\triangle(b, a_2) = A(b, a_2) = \triangle(x_1 \cdot a_1, a_2) = x_1 \cdot \triangle(a_1, a_2)$$
$$\implies x_1 = \frac{\triangle(b, a_2)}{\triangle(a_1, a_2)}$$

□

Generic proof. Let $x = A^{-1} \cdot b = \frac{1}{\det A} \cdot \hat{A} \cdot b$.

$$\begin{aligned}
x_i &= \frac{1}{\det A} \cdot \sum_{k=1}^{n} \hat{a}_{ik} b_k \\
&= \frac{1}{\det A} \sum_{k=1}^{n} (-1)^{i+k} \det A_{ki} \cdot b_k \\
&\underbrace{=}_{\substack{\text{see proof of} \\ \text{Laplace expansion}}} \frac{1}{\det A} \sum_{k=1}^{n} \triangle(a_1, \dots, a_{i-1}, e_k, a_{i+1}, \dots, a_n) b_k \\
&= \frac{\triangle(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n)}{\det A}
\end{aligned}$$

□

Example.

$$\begin{aligned}
2x_1 + x_2 &= 7 \\
x_1 - 3x_2 &= 0
\end{aligned}$$

$$\begin{aligned}
A &= \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} \\
\det(A) &= 2 \cdot (-3) - 1 = -7
\end{aligned}$$

$$\begin{aligned}
x_1 &= -\frac{1}{7} \begin{vmatrix} 7 & 1 \\ 0 & -3 \end{vmatrix} = 3 \\
x_2 &= -\frac{1}{7} \begin{vmatrix} 2 & 7 \\ 1 & 0 \end{vmatrix} = 1
\end{aligned}$$

Remark. *For large n (hence $n \geq 4$), Cramer's Rule is impractical (tiresome and unstable). But it helps with theoretical considerations.*

1. *The map $A \mapsto \det A$ is continuous and differentiable.*
2. *if $\det A \neq 0 \implies$ the set of invertible matrices is open*[4]
3. *The solution of system $Ax = b$ depends continuously on a_{ij} and b_i* [5]

---

[4] Hence for all invertible A, there exists some neighborhood such that all matrices in this neighborhood are invertible.

$$\text{e.g. } d(A, B) = \max_{i,j} \left| a_{ij} - b_{ij} \right|$$

[5] This justifies why Computational Mathematics (dt. Numerik) is practical and interesting

$$\forall \varepsilon \exists \delta : d(b, b') < \delta \implies d(x, x') < \varepsilon$$

8 Inner products

8.1 Definition

Definition 8.1.

$$\mathbb{R}^3 : \left\| \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \right\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

By Pythagorem Theorem

Geometrical proof of the Pythagorem Theorem. Claim: $a^2 + b^2 = c^2$

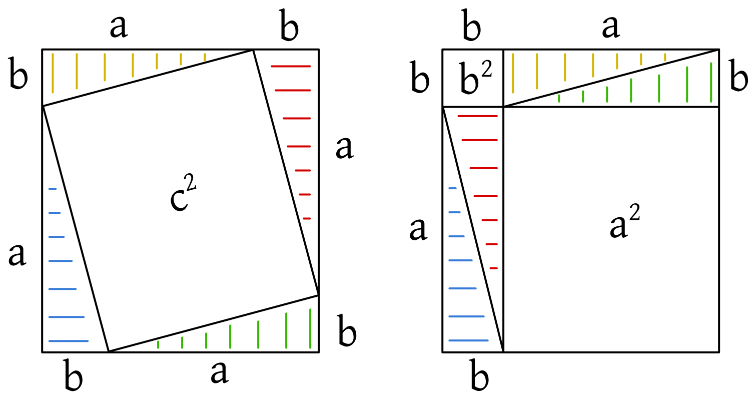


Figure 5: Proof construction of the Pythagorem Theorem

$\square$

↓ *This lecture took place on 2018/03/21.*

The norm is given by

$$\left\| \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \right\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

Definition (Scalar product in $\mathbb{R}^2/\mathbb{R}^3$)**.**

$$\langle a, b \rangle := ||a|| \cdot ||b|| \cdot \cos\theta$$

where θ is the angle between vector a and b.

Theorem. $\langle a, a \rangle = ||a||^2$

Remark. *Recall that*

$$\cos 0 = 1 \qquad \cos\frac{\pi}{2} = 0 \qquad \cos\pi = -1 \qquad \cos\frac{3}{2}\pi = 0$$

$$\sin 0 = 0 \qquad \sin \frac{\pi}{2} = 1 \qquad \sin \pi = 0 \qquad \sin \frac{3}{2}\pi = -1$$

$$\sin \theta = \cos(\theta - \frac{\pi}{2})$$
$$\cos(\pi - \theta) = -\cos(\theta)$$
$$\sin(-\theta) = \cos(\theta)$$
$$\sin(\pi - \theta) = \sin(\theta)$$
$$\sin(-\theta) = -\sin(\theta)$$

Theorem 8.2. *1.* $\langle a, a \rangle = ||a||^2$

2. $\langle a, a \rangle = 0 \iff a = 0$
3. $\langle a, b \rangle = 0 \iff a = 0 \vee b = 0 \vee \theta = \frac{\pi}{2} \vee \theta = \frac{3}{2}\pi$, *hence orthogonal*
4. $\langle a, b \rangle > 0 \iff$ *acute angle*
5. $\langle a, b \rangle < 0 \iff$ *obtuse angle (dt. stumpf)*

Theorem 8.3. *1.* $\langle a, b \rangle = \langle b, a \rangle$

2. $\langle \lambda a, b \rangle = \lambda \cdot \langle a, b \rangle = \langle a, \lambda \cdot b \rangle$
3. $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$

Thus, linear in a and b. Thus, bilinear.

Proof. 1. $\langle a, b \rangle = ||a|| \, ||b|| \cdot \cos \theta = ||b|| \, ||a|| \cdot \cos \theta = \langle b, a \rangle$

2. Assume $\lambda > 0$. Angle stays the same.

$$\langle \lambda a, b \rangle = ||\lambda a|| \cdot ||b|| \cdot \cos \theta = \lambda \cdot ||a|| \cdot ||b|| \cdot \cos \theta$$

Assume $\lambda < 0$. θ becomes $\pi - \theta$.

$$\langle \lambda a, b \rangle = ||\lambda a|| \cdot ||b|| \cdot \cos(\pi - \theta) = |\lambda| \cdot ||a|| \cdot ||b|| \cdot (-\cos(\theta)) = \lambda \cdot ||a|| \cdot ||b||$$

3. Let $||c|| = 1$. $\langle a, c \rangle = ||a|| \cdot \cos \theta$.

$$\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$$

Projections will add up.

In the generic case:

$$\langle a + b, c \rangle = \left\langle a + b, ||c|| \cdot \frac{c}{||c||} \right\rangle \underbrace{=}_{\text{by (2.)}} ||c|| \left\langle a + b, \frac{c}{||c||} \right\rangle$$

∎

Theorem 8.4.

$$\left\langle \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\rangle = a_1 b_1 + a_2 b_2 + a_3 b_3$$

Proof.

$$\begin{aligned}
\langle a, b\rangle &= \langle a_1 e_1 + a_2 e_2 + a_3 e_3, b\rangle \\
&= a_1 \langle e_1, b\rangle + a_2 \langle e_2, b\rangle + a_3 \langle e_3, b\rangle \\
&= a_1 b_1 + a_2 b_2 + a_3 b_3 \\
\langle e_i, b\rangle &= \langle e_i, b_1 e_1 + b_2 e_2 + b_3 e_3\rangle \\
&= b_1 \langle e_i, e_1\rangle + b_2 \langle e_i, e_2\rangle + b_3 \langle e_i, e_3\rangle \\
&= b_1 \delta_{i1} + b_2 \delta_{i2} + b_3 \cdot \delta_{i3} \qquad \text{with } \delta \text{ as Kronecker delta} \\
&= b_i
\end{aligned}$$

□

In this chapter, we will talk about vector spaces in which we will discuss scalar products with properties 1–3 from Theorem 8.3.

$$\begin{aligned}
\text{in } \mathbb{R}^n : \langle x, y\rangle &= \sum_{i=1}^{n} x_i y_i \\
\text{in } V \subseteq \mathbb{R}^\infty : \langle x, y\rangle &= \sum_{i=1}^{\infty} x_i y_i \qquad \text{if convergent!}
\end{aligned}$$

For this space, $(e_i)_{i\in\mathbb{N}}$ is a basis.

$$\text{in } C[a, b] \qquad \langle f, g\rangle = \int f(x) g(x)\, dx$$

is the Delta function.

Or better: $(\sin nx)_{n\in\mathbb{N}} \cup (\cos nx)_{n\in\mathbb{N}}$.

$$\begin{aligned}
\int_0^{2\pi} \sin(nx)\cos(mx)\, dx &= 0 \forall m, n \\
\int_0^{2\pi} \sin(nx)\sin(mx)\, dx &= 0 \text{ if } m \neq n
\end{aligned}$$

Person. *Jean-Baptiste Joseph Fourier (1768/03/21–1830/05/16)*

Theorem 8.5 (1822 Fourier)**.** *Every function f in $[0, 2\pi]$ can be denoted as*

$$\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \\
a_n &= \langle f, \cos(nx)\rangle = \int_0^{2\pi} f(x)\cos(nx)\, dx \\
b_n &= \langle f, \sin(nx)\rangle = \int_0^{2\pi} f(x)\sin(nx)\, dx
\end{aligned}$$

This theorem cannot be proven, because it depends on the definition of "function". The answer to the question, which functions satisfy this theorem, is an open research topic.

8.2 Law of cosines

Theorem 8.6 (Law of cosines)**.** *In German, "Kosinussatz".*

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

$$\begin{aligned}\left\|\vec{c}\right\|^2 &= \left\|\vec{b}-\vec{a}\right\|^2\\ &= \left\langle \vec{b}-\vec{a}, \vec{b}-\vec{a}\right\rangle\\ &= \left\langle \vec{b}, \vec{b}\right\rangle - \left\langle \vec{a}, \vec{b}\right\rangle - \left\langle \vec{b}, \vec{a}\right\rangle + \langle \vec{a}, \vec{a}\rangle\\ &= \|b\|^2 - 2\,\|a\|\,\|b\|\cos\gamma + \|a\|^2\end{aligned}$$

$$\|a\| \cdot \|b\| \cdot \sin\theta = \text{ area of the spanned parallelogram}$$

How to find an orthogonal vector?

Remark (Orthogonal vector in $\mathbb{R}^2$)**.** *Find $\vec{b}$ such that $\left\langle \vec{a}, \vec{b}\right\rangle = 0$, $a_1b_1 + a_2b_2 = 0$. For example, $b_1 = a_2$ and $b_2 = -a_1$.*

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} a_2 \\ -a_1 \end{pmatrix}$$

8.3 Outer product

Definition 8.7. *Called* outer product *(only in $\mathbb{R}^3$) or* cross product*.*
Let $a, b \in \mathbb{R}^3$ and let $a \times b$ be the vector which

1. $\|a \times b\| = \|a\| \cdot \|b\| \cdot \sin\theta$ *is the area of the spanned parallelogram.*
2. $a \times b \perp a$ *and* $a \times b \perp b \iff \langle a \times b, a\rangle = 0$ *and* $\langle a \times b, b\rangle = 0$
3. $(a, b, a \times b)$ *is clockwise.*

$$\textit{When does } a \times b = 0 \textit{ hold? } a = 0, b = 0, \sin\theta = 0, \textit{ hence } \theta = 0 \vee \theta = \pi$$

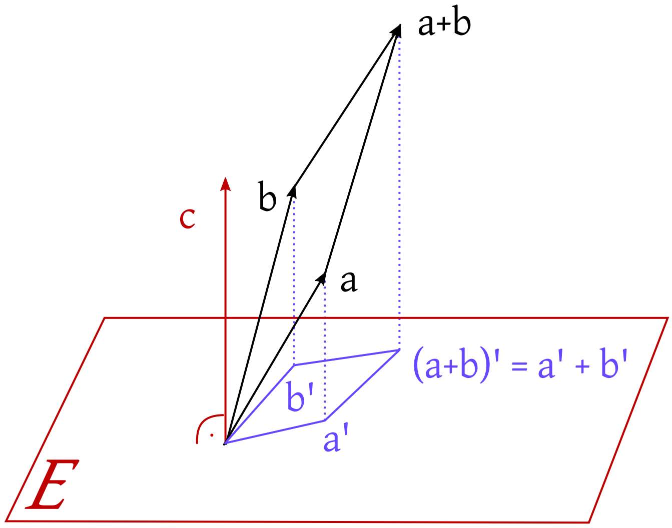
$$\iff a, b \textit{ are linear independent}$$

Theorem 8.8.

- $b \times a = -a \times b$
- $(\lambda a) \times b = \lambda(a \times b) = a \times (\lambda b)$
- $(a + b) \times c = a \times c + b \times c$

Proof.

- Orientation swaps. Consider the right-hand rule. If you assign b to your index finger, a to your middle finger, you retrieve direction $b \times a$ with the thumb. Now assign a (index finger) and b (middle finger) and you retrieve the opposite direction, namely $-(b \times a)$.



- If $\lambda > 0$, it follows immediate. If $\lambda < 0$, lengths stay the same, but orientation swaps.
- If $c = 0$, it is trivial. If $c \neq 0$, E is the plane orthogonal to c. a' and b' are projections of a and b to E.

  1. $(a+b)' = a' + b'$
  2. $a \times c = a' \times c$.

  $$\begin{aligned} ||a \times c|| &= ||a||\,||c|| \cdot \sin\theta \\ &= ||a'|| \cdot ||c|| \\ &= ||a' \times c|| \end{aligned}$$

     - Orientation of $a \times c$ and $a' \times c$ is the same
     - The plane, spanned by c and a, is also spanned by c and a'

  $$||a'|| = ||a|| \cdot \underbrace{\cos\left(\frac{\pi}{2}\right)}_{=\sin\theta} - \theta)$$

  Hence,

  $$(a+b) \times c = (a+b)' \times c = (a'+b') \times c \overset{!}{=} a' \times c + b' \times c = a \times c + b \times c$$

  $$(a'+b') \times c = a' + b'$$

  rotated by $90°$ multiplied by $||c||$

  $$a' \times c = a'$$

rotated by $90°$ multiplied by $||c||$

$$a' \times c + b' \times c = (a' + b') \times c$$

The relation $u + v = w$ will be preserved under rotation by $90°$ and multiplication with λ.

□

Corollary 8.9. *The cross product is a map of* $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ *such that*

- *bilinear*
- *antisymmetrical,* $a \times b = -b \times a$
- $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$

$$e_i \times e_j = e_k \cdot \operatorname{sign} \pi \qquad \pi = \begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix}$$

Corollary 8.10.

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix} = \begin{bmatrix} \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ -\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} \\ \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \end{bmatrix} \underbrace{=}_{\substack{\text{by Laplace expansion along} \\ \text{the third column} \\ \text{[not an actual equality]}}} \begin{vmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \\ a_3 & b_3 & e_3 \end{vmatrix}$$

Proof.

$$\begin{aligned} &(a_1 e_1 + a_2 e_2 + a_3 e_3) \times (b_1 e_1 + b_2 e_2 + b_3 e_3) \\ &= a_1 b_1 e_1 \times e_1 + a_1 b_2 e_1 \times e_2 + a_1 b_3 e_1 \times e_3 \\ &+ a_2 b_1 e_2 \times e_1 + a_2 b_2 e_2 \times e_2 + a_2 b_3 e_2 \times e_3 \\ &= a_3 b_1 e_3 \times e_1 + a_3 b_2 e_3 \times e_2 + a_3 b_3 e_3 \times e_3 \\ &= a_1 b_2 e_3 - a_1 b_3 e_2 - a_2 b_1 e_3 + a_2 b_3 e_1 + a_3 b_1 e_2 - a_3 b_2 e_1 \\ &= (a_2 b_3 - a_3 b_2) e_1 + (a_3 b_1 - a_1 b_3) e_2 + (a_1 b_2 - a_2 b_1) e_3 \end{aligned}$$

□

Theorem 8.11.

$$\langle a \times b, c \rangle = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

This corresponds to the volume of the spanned parallelepiped (dt. "Spat"). $||a \times b||$ *is the area of the parallelogram and* $||c||$ *its height.*

Equivalently, $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ *is the area of the parallelogram.*

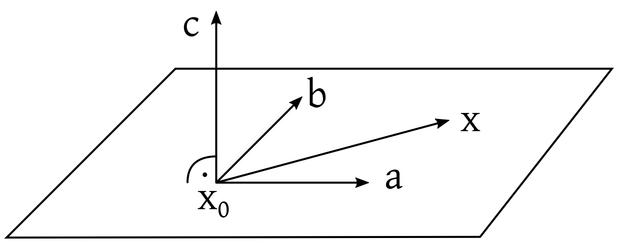
Proof. Laplace expansion in third column □

Example 8.12. *Let planes in $\mathbb{R}^3$ be given. Let x_0 be any point in E.*

$$E := \left\{x_0 + \lambda a + \mu b \,\middle|\, \lambda, \mu \in \mathbb{R}\right\}$$

$$c = a \times b$$

$$\left\{x \in \mathbb{R}^3 \,\middle|\, x - x_0 \perp c\right\} = \left\{x \in \mathbb{R}^3 \,\middle|\, \langle x - x_0, c\rangle = 0\right\}$$



8.4 Inner products and positive definiteness

From now on $\mathbb{K}$ will be $\mathbb{R}$ or $\mathbb{C}$.

Definition 8.13. *An inner product on a vector space V is a map*

$$V \times V \to \mathbb{K} \qquad (x, y) \mapsto \langle x, y\rangle$$

1. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle \quad \forall x, y, z \in V$
2. $\langle \lambda x, y\rangle = \lambda \langle x, y\rangle \quad \forall \lambda \in \mathbb{K} \forall x, y \in V$
3. $\langle y, x\rangle = \overline{\langle x, y\rangle} \quad \forall x, y \in V$

where $\overline{\langle x, y\rangle}$ denotes the complex conjugate.

$$\langle x, \lambda y\rangle \underbrace{=}_{\text{by (3)}} \overline{\langle \lambda y, x\rangle} \underbrace{=}_{\text{by (2)}} \overline{\lambda \langle y, x\rangle} = \overline{\lambda} \langle x, y\rangle$$

The inner product is linear in x and semi-linear in y, thus sesquilinear[6].
In physics, the notation is different:

$$\langle x|y\rangle \qquad \langle \lambda x|y\rangle = \overline{\lambda} \langle x|y\rangle \qquad \langle x|\lambda y\rangle = \lambda \langle x|y\rangle$$

$$\langle x| \ldots \textit{bra} \qquad |y\rangle \ldots \textit{ket} \qquad \langle x|y\rangle \qquad \textit{bracket}$$

The inner product is called positive-semidefinite, *if*

$$\langle x, x\rangle \geq 0 \qquad \forall x \in X$$

If $\langle x, x\rangle = 0 \iff x = 0$, then $\langle\cdot,\cdot\rangle$ is called positive definite.

[6] In Latin, sesqui means 1.5

↓ *This lecture took place on 2018/04/09. Easter holidays finished.*

Lemma 8.14. *1.* $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$

2. $\langle x, \lambda y \rangle = \overline{\lambda} \cdot \langle x, y \rangle$

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

Definition 8.15. *An inner product is* positive semidefinite, *if* $\langle x, x \rangle \geq 0$. *Is* positive definite, *if* $\langle x, x \rangle > 0$ *for all* $x \neq 0$. *Is* negative definite, *if* $\langle x, x \rangle < 0$ *for all* $x \neq 0$. *Is* indefinite, *if neither positive nor negative semidefinite.*

A positive definite inner product is called scalar product. *A positive definite inner product is in* Hermitian form, *if* $\mathbb{K} = \mathbb{C}$. *A positive definite inner product is also called* unitary product, *if* $\mathbb{K} = \mathbb{C}$.

So quadratic form over $\mathbb{R}$ *and Hermitian form over* $\mathbb{C}$.

Remark. *For example, the expression* $\langle x, x \rangle > 0$ *requires that* $x \in \mathbb{R}$, *but we defined the inner product over* $\mathbb{C}$ *as well. In Euclidean spaces,* $\langle x, x \rangle \in \mathbb{R}$ *anyhow, but more generally, the condition should require the absolute value:* $|\langle x, x \rangle| > 0$.

Example 8.16. • *The dot product over* $\mathbb{R}^n$ *or* $\mathbb{C}^n$ *is given by,*

$$V = \mathbb{R}^n : \quad \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i=1}^n x_i y_i \qquad V = \mathbb{C}^n : \quad \left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i=1}^n x_i \overline{y_i}$$

$$V = \mathbb{C}^n \implies \langle x, x \rangle = \sum_{i=1}^n x_i \overline{x_i} = \sum_{i=1}^n |x_i|^2 \geq 0 \rightarrow \textit{ positive semi-definite}$$

• *Let* $A \in \mathbb{R}^{n \times n}$. *Let* $x, y \in \mathbb{R}^n$.

$$\begin{aligned} \langle x, y \rangle_A &= x^t \cdot A \cdot y \qquad \textit{is bilinear} \\ &= \sum_{i=1}^n x_i \sum_{j=1}^n a_{ij} y_j = \sum_{i,j=1}^n a_{ij} x_i y_j \end{aligned}$$

hence $\langle x, y \rangle_A = \langle y, x \rangle_A$. *It must hold that*

$$\sum_{i,j=1}^n a_{ij} x_i y_j = \sum_{i,j=1}^n a_{ij} y_i x_j \forall x, y$$

We let $x = e_k$ *and* $y = e_l$.

$$\implies a_{kl} = a_{lk} \forall k, l$$

Hence $A = A^T$. *A is symmetrical.*

Let $A \in \mathbb{C}^{n \times n}$. *Let* $x, y \in \mathbb{C}^n$.

$$\langle x, y \rangle_A = \sum_{i=1}^n \sum_{j=1}^n x_i a - i j \overline{y_j}$$

$$\langle x,y\rangle_A = \langle y,x\rangle_A \,\forall x,y$$
$$\iff A^T = \overline{A} \qquad \textit{is in Hermitian form}$$
$$a_{ji} = \overline{a_{ij}} \forall i,j$$

- $$V = C[a,b] = \{f : [a,b] \to \mathbb{K} \textit{ continuous}\}$$
$$\langle f,g\rangle = \int_a^b f(t)\overline{g(t)}\,dt \qquad \textit{is a scalar product}$$
$$\langle f,f\rangle = \int_a^b \left|f(t)\right|^2\,dt \geq 0$$

- *Consider* $V = l_2$ *(*$\mathbb{R}^\infty$ *would be too large) where* $l_2 = \left\{(x_n)_{n\in\mathbb{N}} \,\middle|\, x_n \in \mathbb{R}, \sum_{n=1}^{\infty} x_n^2 < \infty\right\}$.
$$\langle x,y\rangle = \sum_{n=1}^{\infty} x_n y_n \qquad \textit{is a scalar product}$$
Does it converge? This is not obvious.

  Fourier claimed that this example (4) and example (3) are the same. He claimed every function can be written as $f(x) = \sum_{n=0}^{\infty} a_n e^{inx}$.

$$x \cdot x = \langle x,x\rangle = \sum_{i=1}^{n} x_i^2 = \|x\|^2$$

Definition 8.17. *Let V be a vector space. A* norm *on V is a map* $\|\cdot\| : V \to [0,\infty[$ *such that*

1. $\|x\| \geq 0$ *and* $\|x\| = 0 \iff x = 0$
2. $\|\lambda \cdot x\| = |\lambda| \cdot \|x\| \qquad \forall \lambda \in K, \forall x \in V$
3. $\left\|x+y\right\| \leq \|x\| + \left\|y\right\|$ *is the triangle inequality*

Remark. *Every norm is a metric with* $d(x,y) = \left\|x-y\right\|$.

d is translation invariant. $d(x+x_0, y+x_0) = d(x,y)$. *This is compatible to a vector space.*

In a black hole (→ physics), you have a different metric in every point (Riemannian geometry): $\langle x,y\rangle_{A(x,y)}$.

Example 8.18. *Let* $V = \mathbb{R}^n$.

- $\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)$ *is called* Euclidean norm.
- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ *is called* L^1 norm *or* Manhattan norm.
- $\|x\|_\infty = \max\{|x_i| \mid i = 1,\dots,n\}$

Let $V = C[a,b]$.

$$\left\|f\right\|_1 := \int_a^b \left|f(t)\right| \, dt$$

$$\left\|f\right\|_\infty := \max_{t\in[\overline{a},b]} \left|f(t)\right| \qquad \textit{is a } L^\infty\textit{-norm}$$

$$\left\|f\right\|_2 := \left(\int \left|f(t)\right|^2 \, dt\right)^{\frac{1}{2}}$$

The L^1-norm, gives rise to the Lebesgue integral.

Theorem 8.19. *Let $\langle\cdot,\cdot\rangle$ be a scalar product in V (hence, positive-definite inner product). Then $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm on V.*

Proof.

- $\|x\| \geq 0, \|x\| = 0 \iff \langle x, x\rangle = 0 \iff x = 0$
- $\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda \cdot \overline{\lambda} \cdot \langle x, x\rangle} = \sqrt{\lambda^2 \cdot \langle x, x\rangle} = |\lambda| \cdot \sqrt{\langle x, x\rangle}$
- Triangle inequality. We will prove this after proving the CBS inequality.

□

8.5 Cauchy-Bunyakovsky-Schwarz inequality

Lemma 8.20 (Cauchy-Bunyakovsky-Schwarz inequality)**.** *Cauchy (1789–1857) for $\mathbb{R}^n$, Bunyakovsky (1804–1889) for infinite dimensions, Schwarz (1843–1921) generically.*

$$\left|\langle x, y\rangle\right| \leq \|x\| \cdot \left\|y\right\|$$

Hence, l^2 if $\sum_{n=1}^\infty x_n^2 < \infty$ and $\sum_{n=1}^\infty y_n^2 < \infty$. $\langle x, x\rangle < \infty$ and $\langle y, y\rangle < \infty$.

$$\implies \sum x_n y_n \leq \sqrt{\sum x_n^2}\sqrt{\sum y_n^2}$$

If $\left|\langle x, y\rangle\right| = \|x\| \cdot \left\|y\right\| \iff x, y$ are linear dependent.

Proof. Now we can continue with part 3 of the proof of Theorem 8.19. Triangle inequality:

$$\begin{aligned}
\left\|x+y\right\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\leq \|x\|^2 + 2\left|\langle x, y\rangle\right| + \left\|y\right\|^2 \\
&\leq \|x\|^2 + 2\|x\|\left\|y\right\| + \left\|y\right\|^2 \\
&= \left(\|x\| + \left\|y\right\|\right)^2
\end{aligned}$$

□

Proof of CBS inequality, Lemma 8.20.

Case 1: $y = 0$ trivial

Case 2: $y \neq 0$ Let $\lambda \in \mathbb{K}$ be arbitrary.

$$
\begin{aligned}
0 &\leq \langle x - \lambda y, x - \lambda y \rangle \\
&= \langle x, x \rangle - \langle x, \lambda y \rangle - \langle \lambda y, x \rangle + \langle \lambda y, \lambda y \rangle \\
&= \langle x, x \rangle - \overline{\lambda} \langle x, y \rangle - \lambda \langle y, x \rangle + |\lambda|^2 \langle y, y \rangle
\end{aligned}
$$

This holds for all λ, hence also for $\lambda = \frac{\langle x,y \rangle}{\langle y,y \rangle}$. Because $y \neq 0 \implies \langle y, y \rangle > 0$, we can divide.

$$
\begin{aligned}
&= \langle x, x \rangle - \frac{\overline{\langle x, y \rangle}}{\langle y, y \rangle} \cdot \langle x, y \rangle - \frac{\langle x, y \rangle}{\langle y, y \rangle} \cdot \langle y, x \rangle + \frac{\left|\langle x, y \rangle\right|^2}{\langle y, y \rangle^2} \cdot \langle y, y \rangle \\
&= \langle x, x \rangle - \frac{\left|\langle x, y \rangle\right|^2}{\langle y, y \rangle} - \frac{\left|\langle x, y \rangle\right|^2}{\langle y, y \rangle} + \frac{\left|\langle x, y \rangle\right|^2}{\langle y, y \rangle} \\
&= \|x\|^2 - \frac{\left|\langle x, y \rangle\right|^2}{\left\|y\right\|^2} \\
&\implies \|x\|^2 \cdot \left\|y\right\|^2 - \left|\langle x, y \rangle\right|^2 \geq 0
\end{aligned}
$$

□

Alternative proof of CBS inequality in $\mathbb{R}^n$.

$$
\begin{aligned}
0 &\leq \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i y_j - x_j y_i)^2 \\
&= \sum_{i,j=1}^{n} \left(x_i^2 y_j^2 - 2 x_i y_j x_j y_i + x_j^2 y_i^2 \right) \\
&= \sum_{i,j}^{n} x_i^2 y_j^2 - 2 \sum_{i,j} x_i x_j y_i y_j + \sum_{i,j} x_j^2 y_i^2 \\
&= 2 \sum_i x_i^2 \sum_j y_j^2 - 2 \sum_i x_i y_i \sum_j x_j y_j \\
&= 2 \|x\|^2 \left\|y\right\|^2 - 2 \langle x, y \rangle^2 \\
&\rightsquigarrow \|x\|^2 \left\|y\right\|^2 = \langle x, y \rangle^2 + \frac{1}{2} \sum_i \sum_j (x_i y_j - x_j y_i)^2
\end{aligned}
$$

So for $n = 3$, $\|x\|^2 \left\|y\right\|^2 = \langle x, y \rangle^2 + \left\|x \times y\right\|^2$. Hence, equality is given iff x and y are linear dependent.

In the general case: If $\left|\langle x, y \rangle\right| = \|x\| \cdot \left\|y\right\|$. From the proof, it follows that $\exists \lambda : \langle x - \lambda y, x - \lambda y \rangle = 0$

$$
\implies x - \lambda y = 0 \implies x, y \text{ are linear independent}
$$

□

Theorem 8.21. *Let V be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Let $B = \{b_1, \dots, b_n\}$ be a basis. $\langle \cdot, \cdot \rangle$ is an inner product. What does $\langle \cdot, \cdot \rangle$ look like in regards of the coordinate?*

There exists a unique matrix A in Hermitian form (hence, $a_{ij} = \overline{a_{ji}}, A = \overline{A^T}$) such that $\forall x, y \in V : \langle x, y \rangle = \Phi_B(x)^T \cdot A \cdot \overline{\Phi_B(y)}$. If $\langle \cdot, \cdot \rangle$ is positive definite, A is invertible.

Remark.

$$\langle x, y \rangle = \sum x_i \overline{y_i}$$

corresponds to $A = I$.

$$x^T \cdot I \cdot \overline{y} = x^T \cdot \overline{y}$$

How about $A = -I$.

$$\langle x, y \rangle_A = -\sum x_i \overline{y}_i$$

This is not a scalar product (because of negative definiteness).

Proof. Let $x = \sum_{i=1}^n \xi_i b_i, y = \sum_{j=1}^n \eta_j b_j$.

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n \xi_i b_i, \sum_{j=1}^n \eta_j b_j \right\rangle = \sum_{i=1}^n \xi_i \sum_{j=1}^n \overline{\eta_j} \qquad \underbrace{\left\langle b_i, b_j \right\rangle}_{=:a_{ij} \text{ is unique } a_{ij} = \left\langle b_i, b_j \right\rangle}$$

$$= \sum_{i=1}^n \sum_{j=1}^n \xi_i a_{ij} \overline{\eta}_j = \xi^T \cdot A \cdot \overline{\eta} = \Phi_B(x)^T \cdot A \cdot \Phi_B(y)$$

$$a_{ji} = \left\langle b_j, b_i \right\rangle = \overline{\left\langle b_i, b_j \right\rangle} = \overline{a_{ij}}$$

Show: If $\langle \cdot, \cdot \rangle$ is positive definite, then A is invertible. It suffices to show that $\ker A = \{0\}$.

Assume: $A \cdot \xi = 0 \implies \xi^T \cdot A \cdot \xi = 0$. Let $x = \sum_{i=1}^n \xi_i b_i$. $\langle x, x \rangle = 0 \implies x = 0 \implies \xi = \Phi_B(x) = 0$ □

Definition 8.22.

- *Let $A \in \mathbb{C}^{n \times n}$. The matrix $A^* := \overline{A^T}$ ($(A^*)_{ij} = \overline{a_{ji}}$) is called* conjugate transpose.
- *A is called* self-adjoint *if $A = A^*$ (dt. selbst-adjungiert).*
- *If $\mathbb{K} = \mathbb{R}$ and $A = \overline{A}$, then A is called* symmetric.
  If $\mathbb{K} = \mathbb{C}$ and $A = A^*$, then A is called Hermitian.
- *$A = A^*$ is called (positive/negative) (semidefinite/definite) if the corresponding sesquilinear form satisfies*

  $$\langle \xi, \eta \rangle_A = \xi^T \cdot A \cdot \overline{\eta} \circ 0 \textit{ where } \circ \textit{ is one of } \geq, >, \leq \textit{ or } < \textit{ respectively}$$

  Thus A is positive definite iff $\forall \xi \in \mathbb{K}^n : \langle \xi, \eta \rangle_A = \xi^T \cdot A \cdot \overline{\eta} \geq 0$
- *If $\exists \xi : \xi^T A \overline{\xi} > 0$ and $\exists \eta : \eta^T A \overline{\eta} < 0$, then A is called indefinite.*

↓ *This lecture took place on 2018/04/11.*

Inner product: $\langle x, y\rangle$

- $\forall x : \langle x, x\rangle \geq 0$ positive semi-definite
- $\forall x \neq 0 : \langle x, x\rangle > 0$ positive definite

with respect to basis $b_1, \dots, b_n$.

$$\langle x, y\rangle = \sum a_{ij}\xi_i\overline{\eta_j} \qquad a_{ij} = \left\langle b_i, b_j\right\rangle$$

Remark. $A = A^*$ *is called* positive semidefinite *if* $A \geq 0$ *if* $\forall \xi : \xi^T A\overline{\xi} \geq 0$.

$A = A^*$ *is called* positive definite $\iff A > 0 \iff \forall \xi \in \mathbb{K}^n \setminus \{0\} : \xi^T A\overline{\xi} > 0$ *with* $\xi^T A\overline{\xi} = \sum_{i=1}\sum_{j=1} a_{ij}\xi_i\overline{\xi_j}$.

Example.

$$A = I > 0$$

$$\xi^T I\overline{\xi} = \sum_{i=1}^n \xi_i\overline{\xi_i} = \sum |\xi_i|^2 > 0 \qquad \text{if } \xi \neq 0$$

$$A = -I < 0 \text{ is negative definite}$$

$$A = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{bmatrix}$$

is indefinite:

$$e_1^T A e_1 > 0 \qquad e_n^T A e_n < 0$$

Remark. *For a diagonal matrix*

$$A = \begin{bmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{bmatrix}$$

$A = A^* \iff a_i = \overline{a}_i$, *hence for all* $a_i \in \mathbb{R}$.

For a diagonal matrix it holds that

$$\begin{aligned} A > 0 \text{ if all } a_i > 0 : \xi^T A\overline{\xi} = \sum_{i=1}^n a_i\,|\xi_i|^2 \geq 0 \\ A \leq 0 \text{ if all } a_i \geq 0 \text{ if } \xi^T A\overline{\xi} = 0 \implies \text{all } a_i \cdot |\xi_i|^2 = 0 \\ A < 0 \text{ if all } a_i < 0 \\ A \leq 0 \text{ if all } a_i \leq 0 \\ \text{indefinite if } \exists i : a_i > 0 \exists j : a_j < 0 \end{aligned}$$

Remark. *Remember, that the rank of matrix satisfies:*

$$\exists P, Q \in \mathrm{GL}(n) : PAQ = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$$

$$A \sim PAQ \textit{ is equivalent}$$

8.6 Congruence of matrices

Definition 8.23 (Congruence)**.** *Consider two self-adjoint matrices $A, B \in \mathbb{K}^{n \times n}$ are called congruent (denoted $A \hat{=} B$) if $\exists C \in \mathrm{GL}(n, \mathbb{K})$ such that $C^* AC = B$.*

Remark. *C is invertible, hence C^T is invertible.*

$$(C^T)^{-1} = (C^{-1})^T \qquad (C^{-1})^T \cdot C^T = (C \cdot C^{-1})^T = I^T = I$$

$$(\overline{A}^{-1}) = \overline{A^{-1}}$$

$$(AB)^* = \overline{(AB)^T} = \overline{B^T A^T} = \overline{B^T} \cdot \overline{A^T} = B^* \cdot A^*$$

$C^* AC$ is self-adjoint.

$$(C^* AC)^* = C^* \cdot A^* \cdot (C^*)^* = C^* \cdot A \cdot C$$

Theorem 8.24. *Every Hermitian matrix is congruent to a diagonal matrix D of structure:*

$$\mathrm{diag}(D) = (1, 1, \dots, 1, -1, \dots, -1, 0, \dots, 0)$$

Proof. The proof is given by an algorithm.

We construct matrix C inductively such that

$$C^* AC = \mathrm{diag}(\pm 1, \dots, 0)$$

Consider $n = 1$.

$$A = [a_{11}]$$

If $a_{11} = 0$ where $a_{11} \in \mathbb{R}$, we don't have to do anything. If $a_{11} \neq 0$ and $a_{11} \in \mathbb{R}$ (because A is self-adjoint),

$$C = \left[\frac{1}{\sqrt{|a_{11}|}} \right]$$

$$C^* AC = \left[\frac{1}{\sqrt{|a_{11}|}} \cdot a_{11} \cdot \frac{1}{\sqrt{|a_{11}|}} \right] = [\mathrm{sign}(a_{11})]$$

Example 8.25.

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

Remark. *It seems we need to take the absolute value in the complex numbers: Let $a = 3 + 4i$. $|a| = 5$.*

$$C^*AC = \left[\frac{1}{\sqrt{|a_{11}|}} \cdot |a_{11}| \cdot \frac{1}{\sqrt{|a_{11}|}}\right] = \left[\frac{1}{\sqrt{5}} \cdot 5 \cdot \frac{1}{\sqrt{5}}\right] = [1]$$

Then $n - 1 \to n$:

Case 1: $A = 0$ nothing to do.

Case 2: $a_{11} = 0$ **Case 2a:**

$$\exists j : a_{jj} \neq 0 : \begin{bmatrix} 0 & & \\ & a_{jj} & \\ & & \end{bmatrix}$$

$$T_{(1,j)} = \begin{bmatrix} 0 & & & & & & & 1 \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & 0 & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ 1 & & & & & & & 1 \end{bmatrix} = T^*_{(ij)}$$

Permutation matrix that swaps 1 with j.

$$T^*_{(1j)} A T_{(1j)} = \begin{bmatrix} a_{ji} & \dots & \dots \\ \vdots & \ddots & \\ \vdots & & 0 \end{bmatrix}$$

where $T^*_{(1j)}$ exchanges j-th and first row and $T_{(1j)}$ exchanges j-th and first column.

Case 2b : all $a_{jj} = 0$. Choose i, j such that $a_{ij} \neq 0$. Let E_{ij} be a zero matrix with 1 at row i and column j.

$$C = I + E_{ij} e^{i\theta}$$

where θ such that $a_{ij} = e^{i\theta} \left|a_{ij}\right|$.

Example. $a_{12} \neq 0$

$$C_1 = \begin{bmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{bmatrix}$$

$$C_1^* A C_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & i \\ 1 & 0 & 1 \\ -i & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & i \\ 1 & 2 & 1+i \\ -i & 1-i & 0 \end{bmatrix}$$

In the general case:

$$C^*AC = (I + E_{ji} e^{-i\theta}) A (I + E_{ij} e^{i\theta})$$

$$
\begin{aligned}
(C^*AC)_{jj} &= \left(A + E_{ji}e^{-i\theta}A + AE_{ij}e^{+i\theta} + E_{ji}AE_{ij}\right)_{jj} \\
&= \underbrace{a_{jj}}_{=0} + \underbrace{(E_{ji}e^{-i\theta}A)_{jj}}_{e^-i\theta a_{jj}=\left|a_{ij}\right|} + \underbrace{(AE_{ij}e^{+i\theta})_{jj}}_{a_{ji}e^{+i\theta}=\overline{a_{ij}}e^{i\theta}=\left|a_{ij}\right|} + \underbrace{a_{ii}}_{=0} \\
&= 2\left|a_{ij}\right|
\end{aligned}
$$

Case 2a is shown.

Example.

$$
C_2 = \begin{bmatrix} 0 & 1 & \\ 1 & 0 & \\ & & 1 \end{bmatrix} = T_{(12)}
$$

$$
A_2 = C_2^* A_1 C_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & i \\ 1 & 2 & i+1 \\ -i & 1-i & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1+i \\ 1 & 0 & i \\ 1+i & -i & 0 \end{bmatrix}
$$

Case 3 $a_{11} \neq 0$

$$
C = \begin{bmatrix} 1 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \dots & -\frac{a_{in}}{a_{11}} \\ & 1 & \dots & 0 & 0 \\ & \vdots & 1 & & 0 \\ & 0 & & \ddots & \\ & 0 & 0 & \dots & 1 \end{bmatrix}
$$

Example.

$$
C_3 = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1+i}{2} \\ & 1 & \\ & & 1 \end{bmatrix}
$$

$$
A_3 = C_3^* A_2 C_3 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1-i}{2} & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 & 1+i \\ 1 & 0 & i \\ 1-i & -i & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1+i}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

$$
\begin{bmatrix} 2 & 1 & 1+i \\ 0 & -\frac{1}{2} & \frac{1}{2}(-i+i) \\ 0 & \frac{1}{2}(-1-i) & -1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{-1+i}{2} \\ 0 & \frac{-1-i}{2} & -1 \end{bmatrix}
$$

$$
C^*AC = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \tilde{A} & \\ 0 & & & \end{bmatrix}
$$

$$
\tilde{A} \in \mathbb{K}^{(n-1)\times(n-1)}
$$

$$
\tilde{A} = \tilde{A}^*
$$

$$
C' = \begin{bmatrix} \frac{1}{\sqrt{|a_{11}|}} & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}
$$

$$(C')^*(C^*AC)C' = \begin{bmatrix} \frac{a_{11}}{|a_{11}|} & 0 & 0 \\ 0 & & \\ \vdots & & \\ 0 & & \tilde{A} \end{bmatrix} \text{where } \frac{a_{11}}{|a_{11}|} = \pm 1$$

Apply this algorithm to $\tilde{A}$.

Example (Part 4).

$$C_4 = \begin{bmatrix} \frac{1}{\sqrt{2}} & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

$$A_4 = C_4^* A_3 C_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{-1+i}{2} \\ 0 & \frac{-1-i}{2} & -1 \end{bmatrix}$$

$$\tilde{A} = \begin{bmatrix} -\frac{1}{2} & \frac{-1+i}{2} \\ \frac{-1-i}{2} & -1 \end{bmatrix}$$

$$C_5 = \begin{bmatrix} 1 & & \\ & 1 & -1+i \\ & 0 & 1 \end{bmatrix}$$

$$A_5 = C_5^* A_4 C_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1-i & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{-1+i}{2} \\ 0 & \frac{-1-i}{2} & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1+i \\ 0 & 0 & 1 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{-1+i}{2} \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1+i \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$C_6 = \begin{bmatrix} 1 & & \\ & \sqrt{2} & \\ & & 1 \end{bmatrix}$$

$$\sqrt{2} = \frac{1}{\sqrt{\frac{1}{2}}}$$

$$C_6^* A_5 C_6 = \begin{bmatrix} 1 & & \\ & -1 & \\ & & 0 \end{bmatrix}$$

$$C_6^* \dots C_2^* C_1^* A C_1 C_2 \dots C_6 = \begin{bmatrix} 1 & & \\ & -1 & \\ & & 0 \end{bmatrix} \implies \textit{indefinite}$$

$$C = C_1 C_2 \dots C_6$$

$$C^* = C_6^* C_5^* \dots C_1^*$$

□

Remark 8.26. 1. *If $A \geq 0$, C arbitrary $\implies C^*AC \geq 0$.*

$$\xi^T(C^*AC)\overline{\xi} = \underbrace{(\xi^T C^*)}_{\xi^T\overline{C^T}=\overline{\overline{\xi^T}C^T}=\overline{(C\cdot\overline{\xi})^T}=\overline{\eta^T}} A \underbrace{(C\overline{\xi})}_{\eta} = \overline{\eta}^T A \overline{\overline{\eta}} \geq 0$$

2. *If $A > 0$, C invertible*

$$\begin{aligned} &\implies C^*AC > 0 \\ \text{if } \xi^T C^*AC\overline{\xi} = 0 &\implies \eta = C\overline{\xi} = 0 \text{ because } A > 0 \\ &\implies \overline{\xi} = 0 \text{ because } C \text{ is invertible} \end{aligned}$$

Corollary 8.27. *If we apply the example 8.25 to $A > 0$,*

$$C^*AC = \operatorname{diag}(\pm 1, \dots, \pm 1, \dots, 0, \dots) \text{ is still positive definite} \implies C^*AC = I$$

Person. *James Joseph Sylvester (1814–1897)*

Theorem 8.28 (Sylvester's law of inertia)**.** *Let $A \in \mathbb{C}^{n \times n}$ be Hermitian. $C \in \operatorname{GL}(n, \mathbb{C})$ by the algorithm such that*

$$C^*AC = \operatorname{diag}(\pm 1, \dots, \pm 1, -1, \dots, -1, 0, \dots, 0)$$

Then the number of $+1$, -1 and zeros is uniquely determined (it does not depend on the order to the operands).

Proof. C is invertible, hence

$$\operatorname{rank}(A) = \operatorname{rank} \operatorname{diag}(+1, \dots, +1, -1, \dots, -1, 0, \dots, 0)$$

Let r be the number of $+1$ and s be the number of -1. The number of $+1$ and -1 is uniquely determined.

Hence, it suffices to show that the number r of $+1$ is uniquely defined.

Let $\tilde{C}$ be another matrix such that

$$\tilde{C}^*A\tilde{C} = \operatorname{diag}(\pm 1, \dots, \pm 1, -1, \dots, -1, 0, \dots, 0)$$

with $\tilde{r}$ ones and $\tilde{s}$ minus ones.

It suffices to show that $r \leq \tilde{r}$. We know $r + s = \tilde{r} + \tilde{s}$.

C is an invertible matrix, hence a change of basis. In this new basis $B' = \{b_1, \dots, b_n\}$, it holds that

$$x^*Ax = \overline{x^T}Ax = \overline{\Phi_B(x)^T} \cdot D \cdot \Phi_B(x)$$

$$\begin{aligned} A &= (C^*)^{-1}DC^{-1} \\ \overline{x^T}Ax &= \overline{x^T}(C^*)^{-1}D \underbrace{C^{-1}x}_{\overline{C}^{-1}x} \end{aligned}$$

Equivalently, $\tilde{C}$ is a change of basis to basis $\tilde{B}$ such that $x^* A x = \Phi_{\tilde{B}}(x)^* \tilde{D} \tilde{\Phi}_{\tilde{B}}(x)$. For $x \in \mathcal{L}(\{b_1, \ldots, b_r\}) \setminus \{0\}$,

$$\Phi_B(x) = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\implies x^* A x = \Phi_B(x)^* D \Phi_B(x)$$

$$= (\overline{\xi}_1, \ldots, \overline{\xi}_r, 0, \ldots, 0) \operatorname{diag}(+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0) \cdot \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{i=1}^{r} |\xi_i|^2 > 0$$

On the other hand, $\forall x \in \mathcal{L}(\tilde{b}_{\tilde{r}+1}, \ldots, \tilde{b}_n)$.

$$\Phi_{\tilde{B}}(x) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{\xi}_{\tilde{r}+1} \\ \vdots \\ \tilde{\xi}_n \end{pmatrix}$$

$$x^* A x = \Phi_{\tilde{B}}(x)^* \tilde{D} \Phi_{\tilde{B}}(x)$$

$$= (0, \ldots, 0, \tilde{\xi}_{\tilde{r}+1}, \ldots, \tilde{\xi}_n) \operatorname{diag}(+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0) \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tilde{\xi}_{\tilde{r}+1} \\ \vdots \\ \tilde{\xi}_n \end{bmatrix} \leq 0$$

$$\implies \mathcal{L}(b_1, \ldots, b_r) \cap \mathcal{L}(\tilde{b}_{\tilde{r}+1}, \ldots, \tilde{b}_n) = \{0\}$$

$$\text{dimension } r + (n - \tilde{r}) \leq n \implies r \leq \tilde{r}$$

□

↓ *This lecture took place on 2018/04/16.*

$$A = A*$$

Conjugate complex. The important question: When does it hold that

$$A > 0$$

Hence

$$\forall x \in \mathbb{C}^n : x^* A x \geq 0$$

$$A > 0 \text{ if } x^* A x > 0 \forall x \neq 0$$

$$(x^*)_i = \overline{x}_i$$

$$\exists C \in \mathrm{GL}(n, \mathbb{C}) \text{ such that}$$

$$C^* A C \underbrace{=}_{\text{congruence}} \mathrm{diag}(+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0)$$

where the number of $+1$ is r (see Sylvester's Law of inertia).

Definition 8.29. *If* $A = A*$ *is congruent to* $\mathrm{diag}(+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0)$ *with* $+1$ *occuring r times and* -1 *occuring s times.*

Then $\mathrm{ind}(A) := r$ *is called* index of A. $\mathrm{sign}(A) := r - s$ *is called* signature of A.

Corollary 8.30.

1. $A > 0 \iff A \hat{=} I \iff \mathrm{ind}(A) = n$
2. $A \geq 0 \iff \mathrm{ind}(A) = \mathrm{sign}(A) = \mathrm{rank}(A)$
3. $A \hat{=} B \iff \mathrm{ind}(A) = \mathrm{ind}(B) \wedge \mathrm{sign}(A) = \mathrm{sign}(B)$

It is left as an exercise to the reader that congruence is an equivalence relation.

1. $I \cdot A \cdot I = A$
2. $A \hat{=} B \implies C^* A C = B \implies A = (C^*)^{-1} B C^{-1} = (C^{-1}) * B C^{-1} \implies B \hat{=} A$
3. $C_1^* A_1 C_1 = A_2 \wedge C_2^* A_2 C_2 = A_3 \implies \underbrace{C_2^* C_1^* A_1 C_1 C_2}_{=(C_1 C_2)^* A_1 (C_1 C_2) \implies A_1 \hat{=} A_3} = A_3$

Furthermore it will be shown in the practicals that $A > 0 \iff \exists C A = C^* C$

Remark (Idea)**.**

$$\det(C^* A C) = \det \mathrm{diag}(+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0)$$

$$\det(C^*) \det(A) \det(C) = \begin{cases} 0 & \textit{if } \mathrm{rank}(A) < n \\ (-1)^{\textit{number of} -1} & \end{cases}$$

$$\overline{\det(C)} \det(A) \det(C)$$

If $A > 0$,

$$\left|\det(C)\right|^2 \cdot \det(A) = 1 \implies \det(A) > 0$$

Lemma 8.31.

1. $\det(C^*) = \overline{\det(C)}$

2. $A = A^* \implies \det(A) \in \mathbb{R}$

3. $A = A^*, B = B^*, A \hat{=} B \implies \operatorname{sign}\det(A) = \operatorname{sign}\det(B)$

4. $A > 0 \implies \det(A) > 0$ *but not the other way around:*

$$\det \begin{bmatrix} -1 & \\ & -1 \end{bmatrix} = 1$$

Proof. 1.

$$\det(C^*) = \underbrace{\sum_{\sigma \in \Sigma_n} (-1)^\sigma \underbrace{(C^*)_{1\sigma(1)}}_{\overline{C_{\sigma(1)1}}} \dots \underbrace{(C^*)_{n\sigma(n)}}_{\overline{C_{\sigma(n)n}}}}_{=\overline{\sum_\sigma (-1)^\sigma C_{\sigma(1)1} \dots C_{\sigma(n)n} = \overline{\det(C)}}}$$

2. immediate

3. $A\hat{B} \implies C^*AC = B$

$$\det(C^*AC) = \det(B)$$

$$\underbrace{\left|\det(C)\right|^2}_{>0} \cdot \det(A) = \det(B)$$

4. $A \hat{=} I \implies \operatorname{sign}\det(A) = \operatorname{sign}\det(I) = 1$

□

Definition 8.32. *Let* $A \in \mathbb{K}^{m \times n}, r \leq \min\{m, n\}$.

$$I = \underbrace{\{i_1 < \ldots < i_r\}}_{\subseteq\{1,\ldots,m\}} \qquad J = \underbrace{\{j_1 < \ldots < j_r\}}_{\subseteq\{1,\ldots,n\}}$$

Then

$$[A]_{I,J} = \begin{vmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \dots & a_{i_1 j_r} \\ a_{i_2 j_1} & a_{i_2 j_2} & \dots & a_{i_2 j_r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_r j_1} & a_{i_r j_2} & \dots & a_{i_r j_r} \end{vmatrix}$$

is called minor of A.

Example. *Let* $r = 1, I = \{i_1\}, J = \{j_1\}, [A]_{\{i_1\},\{j_1\}} = a_{i_1 j_1}$.

Definition 8.33. *If* $m = n$ *with* $I = \{1, \ldots, r\}$ *and* $J = \{1, \ldots, r\}$, *then*

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rr} \end{vmatrix}$$

the first minor of A (Hauptminoren).

$$A < 0 \iff (-A) > 0$$

$$\det(\lambda A) = \lambda^* \det(A)$$

Theorem 8.34 (dt. Hauptminorenkriterium)**.** *Let $A = A^*$, then it holds that*

1. $A > 0 \iff$ *all first minors satisfy* $\det(A_r) > 0$
2. $A < 0 \iff (-1)^r \det(A_r) > 0 \forall r \in \{1, \ldots, n\}$

Proof.

Direction $\implies$: For $r = n$: $\det(A_r) = \det(A) > 0$. It suffices to show: submatrices

$$A_r = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1r} \\ \vdots & \ddots & & \\ a_{r1} & \ldots & & a_{rr} \end{bmatrix}$$

are positive definite. Hence, $\forall x \in \mathbb{C}^r$ with $x \neq 0 : x^* A_r x > 0$.

$$x \in \mathbb{C}^r \setminus \{0\} : x^* A_r x = \begin{bmatrix} x^* & \underbrace{0}_{n-r} \end{bmatrix} \cdot A \cdot \begin{bmatrix} x \\ 0 \end{bmatrix} > 0$$

$$= \begin{bmatrix} x^* & 0 \end{bmatrix} \begin{bmatrix} A_r & & & * \\ & & & \vdots \\ & & & * \\ * & \ldots & * & * \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix}$$

Remark: *every submatrix* $\begin{bmatrix} a_{i_1 i_1} & \ldots & a_{i_1 i_r} \\ \vdots & \ddots & \vdots \\ a_{i_r i_1} & \ldots & a_{i_r i_r} \end{bmatrix}$ of a pos. def. matrix is pos. definite.

Direction $\impliedby$: Assume all first minors $\det(A_r) > 0$.

We use complete induction:

Let $n = 1$ and $r = 1$: $A = [a_{11}]$ and $\det(A_1) = a_{11}$. $A > 0 \iff a_{11} > 0$.

Consider $n - 1 \to n$: Assume all first minors are greater 0. Then all first minors of matrix A_{n-1} are greater 0 by induction hypothesis.

$$\text{Theorem 8.24} \implies \exists C : C^* A_{n-1} C = I_{n-1}$$

$$A' = \begin{bmatrix} & & & 0 \\ & C^* & & \vdots \\ & & & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} A \begin{bmatrix} & & & 0 \\ & C & & \vdots \\ & & & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} & & & 0 \\ & C^* & & \vdots \\ & & & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} & & & a_{1,n} \\ & A_{n-1} & & \vdots \\ & & & a_{n-1,n} \\ \overline{a_{1,n}} & \dots & \overline{a_{n-1,n}} & a_{n,n} \end{bmatrix} \begin{bmatrix} & & & 0 \\ & C & & \vdots \\ & & & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} & & & a_{1,n} \\ & I & & \vdots \\ & & & a_{n-1,n} \\ \overline{a_{1,n}} & \dots & \overline{a_{n-1,n}} & a_{n,n} \end{bmatrix}$$

$$C' = \begin{bmatrix} 1 & & 0 & -a_{1,n} \\ & \ddots & & \vdots \\ 0 & & 1 & -a_{n-1,n} \\ 0 & \dots & 0 & 1 \end{bmatrix} = \left[\begin{array}{c|c} I & -b \\ \hline 0 & 1 \end{array}\right] \text{ with } b := \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n-1,n} \end{bmatrix}$$

$$\begin{aligned}
(C')^* A' C' &= \left[\begin{array}{c|c} I & 0 \\ \hline -b^* & 1 \end{array}\right] \left[\begin{array}{c|c} I & b \\ \hline b^* & a_{n,n} \end{array}\right] \left[\begin{array}{c|c} I & -b \\ \hline 0 & 1 \end{array}\right] \\
&= \left[\begin{array}{c|c} I & b \\ \hline 0 & -b^*b + a_{nn} \end{array}\right] \left[\begin{array}{c|c} I & -b \\ \hline 0 & 1 \end{array}\right] \\
&= \left[\begin{array}{c|c} I & 0 \\ \hline 0 & -b^*b + a_{nn} \end{array}\right]
\end{aligned}$$

Hence

- $A \hat{=} A' \hat{=} \begin{bmatrix} I & 0 \\ 0 & -b^*b + a_n \end{bmatrix}$
- $\exists C'' = \begin{bmatrix} C & 0 \\ 0 & 1 \end{bmatrix} \cdot C'$ such that $(C'')^* A C'' = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & a_{n,n} - b^*b \end{array}\right]$
- $\det(A) \cdot \left|\det(C'')\right|^2 = \det \begin{bmatrix} I & 0 \\ 0 & a_{n,n} - b^*b \end{bmatrix} = a_{n,n} - b^*b$

$$\implies a_{n,n} - b^*b > 0 \qquad \implies \begin{bmatrix} I & 0 \\ 0 & -b^*b + a_{nn} \end{bmatrix} \text{ is positive definite}$$

□

Back to the scalar product:

Person. *David Hilbert (1862–1943)*

Definition 8.35. 1. (a) *A vector space with a positive definite inner product is called* Euclidean space *($K = \mathbb{R}$, $\dim < \infty$) or* unitary space *($K = \mathbb{C}$)*

(b) *Hilbert space if* $\dim = \infty$.

$$\|v\| = \sqrt{\langle v, v\rangle}$$
$$\|\lambda v\| = |\lambda| \cdot \|v\|$$

in $\mathbb{R}^2$: $\langle a, b\rangle = \|a\|\,\|b\| \cos\varphi$

2. *An element* $v \in V$ *is called* normed *if* $\|v\| = 1$ *(if not, then* $\frac{v}{\|v\|}$ *is normed)*

3. *Let* $v, w \in V \setminus \{0\}$. *Then the angle spanned between* v *and* w *is the angled* $\varphi \in [0, \pi]$ *such that* $\cos\varphi = \frac{\Re\langle v,w\rangle}{\|v\|\|w\|}$

4. *Two vectors* $v, w \in V$ *are orthogonal* ($v \perp w$) *if* $\langle v, w\rangle = 0$ *(hence* $\varphi = \frac{\pi}{2}$*)*

Theorem 8.36. 1. $\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\,\|v\|\,\|w\| \cos\varphi$ *(Law of cosines)*

2. *if* $v \perp w$: $\|v + w\|^2 = \|v\|^2 + \|w\|^2$ *(Pythagorean Theorem)*

3. $\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2)$ *(Parallelogram Law)*

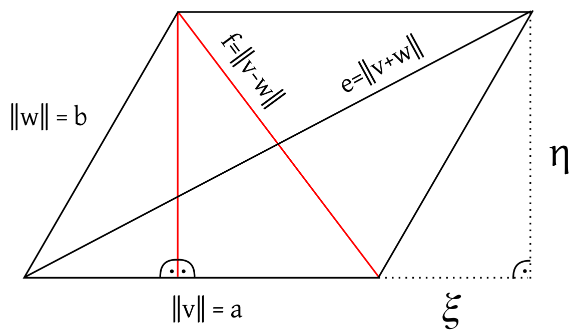


Figure 6: Geometrical proof of Theorem 8.36

$$e^2 + f^2 = 2(a^2 + b^2)$$
$$e^2 = (a + \xi)^2 + \eta^2$$
$$f^2 = (a - \xi)^2 + \eta^2$$
$$e^2 + f^2 = (a + \xi)^2 + (a - \xi)^2 + 2\eta^2$$
$$= a^2 + \xi^2 + a^2 + \xi^2 + 2\eta^2 = 2a^2 + 2b^2$$

Proof. 1. Show the Law of Cosines:

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\ &= \|v\|^2 + \langle v, w\rangle + \overline{\langle v, w\rangle} + \|w\|^2 \\ &= \|v\|^2 + 2\underbrace{\Re\langle v, w\rangle}_{\cos\varphi \cdot \|v\| \cdot \|w\|} + \|w\|^2\end{aligned}$$

2. Show the Pythagorean theorem: immediate, $\langle v, w\rangle = 0$
3. Show the Parallelogram law:

$$\begin{aligned}\|v+w\|^2 + \|v-w\|^2 &= \|v\|^2 + \|w\|^2 + 2\Re\langle v, w\rangle + \|v\|^2 + \|-w\|^2 + 2\Re\langle v, -w\rangle \\ &= 2\|v\|^2 + 2\|w\|^2 + 0\end{aligned}$$

Other norms:

$$\left\|\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}\right\|_1 = \sum_1^n |x_i| \qquad \left\|\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}\right\|_\infty = \max|x_i|$$

□

Remark 8.37. *You can show (von Neumann did): A norm on $\mathbb{R}^n$ satisfies the Parallelogram Law iff $\exists$ a scalar product on $\mathbb{R}^n$ such that $\|v\| = \sqrt{\langle v, v\rangle}$*

Definition 8.38. *Let $(v, \langle\cdot,\cdot\rangle)$ be a vector space with scalar product. A family $(v_i)_{i\in I} \subseteq V$ is called*

orthogonal *if $\forall i \neq j : \left\langle v_i, v_j\right\rangle = 0$*

orthonormal *if additionally $\|v_i\| = 1 \forall i$*
hence $\forall i, j : \left\langle v_i, v_j\right\rangle = \delta_{ij}$

orthonormal basis *if they are orthonormal and give a basis of V.*

Example 8.39. 1. *Canonical basis in $\mathbb{R}^n$ in regards of the standard scalar product*

$$\left\langle e_i, e_j\right\rangle = \delta_{ij}$$

2. *Fourier $\left\{\sqrt{2}\sin 2\pi x,\ \sqrt{2}\sin 4\pi x, \dots,\ \sqrt{2}\sin(2k\pi x), \dots\right\}$ with $k \in \mathbb{N}$ union with $\left\{\sqrt{2}\cos 2\pi x,\ \sqrt{2}\cos 4\pi x, \dots\right\} \cup \{g\}$ on $C[0,1]$.*

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$$

And this is wrong unless we redefine the term basis (not every function is built using the sine/cosine). A basis here is every function:

$$f(x) = \sum_{k=0}^{\infty} a_k \cos(2k\pi x) + \sum_{k=1}^{\infty} b_k \sin(2k\pi 2)$$

And this is wrong as well unless we define equality more precisely (in the usual sense, it is wrong). Lebesgue did this later.

Remark. *For JPEG compression, Fourier transformation is applied. Hence, we consider the music (amplitudes) as f and*

$$f(x) = \sum_{k=0}^{n} a_k \cos 2k\pi x + \sum_{k=1}^{n} b_k \sin 2k\pi 2$$

with n finite.

Theorem 8.40. *Let $(v_i)_{i\in I} \subseteq V$, $v_i \neq 0 \forall i$*

1. $(v_i)_{i\in I}$ *orthogonal* $\iff \left(\frac{v_i}{\|v_i\|}\right)_{i\in I}$ *is orthonormal*
2. $(v_i)_{i\in I}$ *is orthogonal, then* $(v_i)_{i\in I}$ *is linear independent.*

↓ *This lecture took place on 2018/04/18.*

Proof of Theorem 8.40. Let $\sum_{k=1}^n \lambda_k v_{i_k} = 0$.

$$\implies 0 = \left\langle \sum_{k=1}^n \lambda \cdot v_{i_k}, v_i \right\rangle = \sum_{k=1}^n \lambda_k \langle v_{i_k}, v_i \rangle$$

$\forall l \in \{1, \dots, n\}$: Let $i = i_l$.

$$i_l = \sum_{k=1}^n \lambda_k \underbrace{\langle v_{i_k}, v_{i_l} \rangle}_{=\begin{cases} 0 & i_k \neq i_l \\ \|v_{i_l}\|^2 & i_k = i_l \end{cases}} = \lambda_l \cdot \left\|v_{i_l}\right\|^2 \implies \lambda_l = 0$$

□

Theorem 8.41. *Let $B = (b_1, \dots, b_n)$ be an orthonormal basis of a finite dimensional vector space over $\mathbb{K}$. For $v \in V$, let $\Phi_B(v) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$. For $w \in V$, let $\Phi_B(w) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$.*

1. $\lambda_i = \langle v, b_i \rangle$
2. $\langle v, w \rangle = \sum_{i=1}^n \lambda_i \overline{\mu_i}$

Proof. 1.

$$\langle v, b_i \rangle = \left\langle \sum_{j=1}^n \lambda_j b_j, b_i \right\rangle = \sum_{j=1}^n \lambda_j \cdot \underbrace{\langle b_j, b_i \rangle}_{=\delta_{ji}} = \lambda_i$$

2.

$$\langle v, w \rangle = \left\langle \sum_{i=1}^n \lambda_i b_i, \sum_{j=1}^n \mu_j b_j \right\rangle = \sum_{i=1}^n \lambda_i \sum_{j=1}^n \overline{\mu_j} \underbrace{\langle b_i, b_j \rangle}_{\delta_{ij}} = \sum_{i=1}^n \lambda_i \cdot \overline{\mu_i}$$

Compare: B is an arbitrary basis:

$$\langle v, w \rangle = \Phi_B(v)^T \cdot A \cdot \overline{\Phi_B(w)}$$

$$a_{ij} = \left\langle b_i, b_j \right\rangle = \delta_{ij}$$
$$A = I$$
$$\rightarrow \langle v, w \rangle = \Phi_B(v)^T \cdot \overline{\Phi_B(w)}$$

□

Definition 8.42. *Let V be a vector space with a scalar product. Let $v \in V$, then*

$$v^\perp = \{w \in V \mid \langle v, w \rangle = 0\}$$

For $M \subseteq V : M^\perp = \{w \in V \mid \forall u \in M : \langle u, w \rangle = 0\}$ is called orthogonal complement *of v or* orthogonal complement of *M.*

Compare with Figure 7.

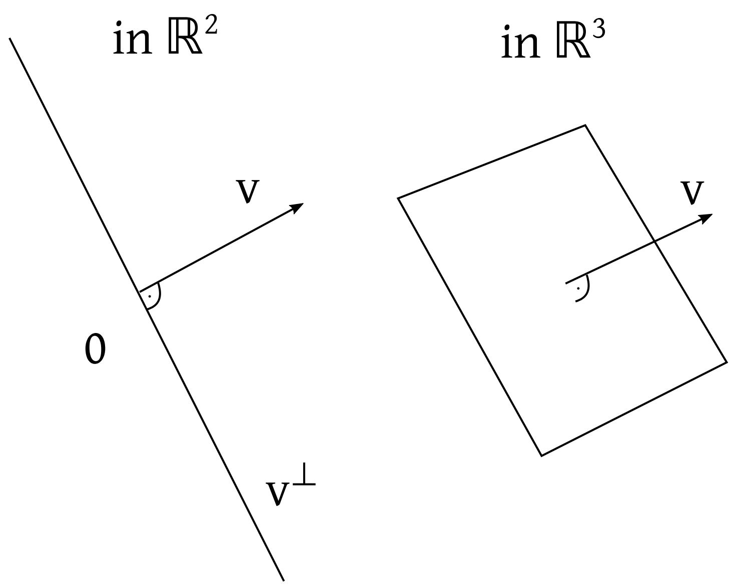


Figure 7: Orthogonal complement

In $\mathbb{R}^n$:

$$\{w \mid \langle v, w \rangle = 0\} = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \middle| \sum_1^n a_i x_i = 0 \right\} \quad \text{if } v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Theorem 8.43. *Let V be a vector space with scalar product. $M, N \subseteq V$ are partitions.*

1. *$M^\perp$ is a subspace.*
2. $M \subseteq N \implies N^\perp \subseteq M^\perp$
   $(M_1 \cup M_2)^\perp = M_1^\perp \cap M_2^\perp$

3. $\{0\}^\perp = V$
4. $V^\perp = \{0\}$
5. $M \cap M^\perp \subseteq \{0\}$
6. $M^\perp = \mathcal{L}(M)^\perp$
7. $M \subseteq (M^\perp)^\perp$

Proof. 1.

$$\begin{aligned} v^\perp &= \{w \in V \mid \langle v, w\rangle = 0\} \\ T_v &: V \to \mathbb{K} \text{ (linear functional)} \\ &\quad w \mapsto \langle w, v\rangle \end{aligned}$$

$$\begin{aligned} v^\perp &= \{w \mid T_v(w) = 0\} = \ker(T_v) \text{ is a subspace} \\ M^\perp &= \bigcap_{v\in M} v^\perp = \bigcap_{v\in M} \ker(T_v) \text{ is a subspace} \end{aligned}$$

2. $M \subseteq N \implies N^\perp \subseteq M^\perp$

$$\begin{aligned} (M_1 \cup M_2)^\perp &= \{w \mid \forall v \in M_1 : \langle w, v\rangle = 0 \land \forall v \in M_2 : \langle w, v\rangle = 0\} \\ &= M_1^\perp \cap M_2^\perp \end{aligned}$$

3. trivial: $\forall v \in V : \langle v, 0\rangle = 0$
4. Let $w \in V$ such that $\langle w, v\rangle = 0 \forall v \in V$. Especially for $v = w$.

$$\implies \underbrace{\langle w, w\rangle}_{\|w\|^2} = 0 \implies w = 0$$

$$\implies V^\perp = \{0\}$$

5. Let $w \in M \cap M^\perp$, hence

$$\begin{aligned} \forall v \in M &: \langle w, v\rangle = 0 \\ w \in M &\implies \langle w, w\rangle = 0 \\ &\implies w = 0 \end{aligned}$$

If $0 \notin M$, then $M \cap M^\perp = \emptyset$.

6. It is immediate that $\mathcal{L}(M)^\perp \subseteq M^\perp$ because

$$M \subseteq \mathcal{L}(M) \underbrace{\implies}_{\text{by point (2.)}} \mathcal{L}(M)^\perp \subseteq M^\perp$$

Show that: $M^\perp \subseteq \mathcal{L}(M)^\perp$. Hence, $\forall v \in M^\perp \implies v \in \mathcal{L}(M)^\perp$.
Let $v \in M^\perp$, $w \in \mathcal{L}(M)$.

$$\exists w_1, \ldots, w_n \in M \, \exists \lambda_1, \ldots, \lambda_n \in \mathbb{K} : w = \sum_{i=1}^n \lambda_i w_i$$

$$\langle w, v\rangle = \left\langle \sum_{i=1}^{n} \lambda_i w_i, v \right\rangle \underbrace{=}_{\substack{\text{by linearity in}\\ \text{1st argument}}} \sum_{i=1}^{n} \lambda_i \underbrace{\left\langle \underbrace{w_i}_{\in M}, \underbrace{v}_{\in M^\perp} \right\rangle}_{=0} = 0$$

$$\implies v \perp w \quad \forall w \in \mathcal{L}(M)$$

7. Show that $\forall v \in M : v \in (M^\perp)^\perp$. Hence, $\forall w \in M^\perp : v \perp w$

$$M^\perp = \{w \mid \forall v \in M : v \perp w\}$$
$$\implies \forall v \in M \forall w \in M^\perp : v \perp w$$
$$\implies \forall v \in M : v \in \bigcap_{w \in M^\perp} w^\perp = (M^\perp)^\perp$$

□

Corollary 8.44. *Let $U \subseteq V$ be a subspace. By Theorem 8.43 (1), $U^\perp$ is a subspace and $U \cap U^\perp = \{0\}$ because of Theorem 8.43 (5), $U + U^\perp$ is a direct sum in $\mathbb{R}^n$ such that $U + U^\perp = \mathbb{R}^n$.*

Remark 8.45. *If $\dim(V) = \infty$, it must not hold that $U + U^\perp = V$.*

Example.

$$V = l^2 = \left\{ (x_n)_{n \in \mathbb{N}} \,\middle|\, \sum |x_n|^2 < \infty \right\}$$

$$\begin{aligned} U &= \mathcal{L}((e_i)_{i \in \mathbb{N}}) \\ &= \left\{ (x_n)_{n \in \mathbb{N}} \,\middle|\, x_n = 0 \textit{ except for finite many } n \right\} \end{aligned}$$

$$U^\perp = \{e_i \mid i \in \mathbb{N}\}^\perp = \left\{ (x_n)_{n \in \mathbb{N}} \,\middle|\, \underbrace{\langle (x_n)_{n \in \mathbb{N}}, e_i \rangle}_{=\left\{ (x_n)_{n \in \mathbb{N}} \,\middle|\, \forall i \in \mathbb{N} : x_i = 0 \right\} = \{0\}} = 0 \forall i \in \mathbb{N} \right\}$$

$$\begin{aligned} \langle (x_n)_n, (y_n)_n \rangle &= \sum_{n=1}^{\infty} x_n \overline{y_n} \\ &\implies U^\perp = \{0\} \\ &\textit{but } U + U^\perp \neq l_2 \end{aligned}$$

$U \dot{+} U^\perp$ is a direct sum.

$$v \in U \dot{+} U^\perp$$
$$U \xrightarrow{\pi_U} U$$
$$U^\perp \xrightarrow{\pi_{U^\perp}} U^\perp$$

Every $v \in U \dot{+} U^\perp$ has a unique decomposition:

$$v = u + w \qquad u \in U, w \in U^\perp$$

Definition 8.46. *Let V be a vector space. A subset $K \subseteq V$ is called convex*[7] *if*

$$\forall \lambda \in [0, 1] : \forall x, y \in K : \lambda x + (1 - \lambda) y \in K$$

[7] Wide-sighted people with glasses use a glass with convex curvature.

Example 8.47. *Subspaces are convex.*

1. $$U \subseteq V : \forall x, y \in U \forall \lambda, \mu : \lambda x + \mu y \in U$$
   Especially: $\lambda \in [0,1], \mu = 1 - \lambda$

2. *Let* $(V, \|\cdot\|)$ *be a normed space.*
   $$B_{\|\cdot\|}(0,1) = \left\{ x \in V \mid \underbrace{\|x\| < 1}_{\text{unit circle}} \right\}$$
   We discussed three different norms so far. In $\mathbb{R}^2$ *with* $\|\cdot\|_2$ *(Euclidean norm), the unit circle is a circle of radius* 1*. In* $\mathbb{R}^2$ *with* $\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\|_\infty = \max(|x|, |y|)$ *(infinity norm), the unit circle is a square from* $(-1,-1)$ *to* $(1,1)$*. This square contains the circle of radius* 1*. In* $\mathbb{R}^2$ *with* $\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\|_1 = |x| + |y|$ *(Manhattan norm), the unit circle is a square rotated by 45 degrees from* $(-1,0)$ *to* $(1,0)$*. It also contains the circle of radius* 1*.*
   Let $x, y \in B(0,1)$*, hence* $\|x\| < 1$, $\|y\| < 1$.
   $$\begin{aligned} \|\lambda x + (1-\lambda)y\| &\underbrace{\leq}_{\text{by triangle ineq.}} \lambda \|x\| + (1-\lambda)\|y\| \\ &< \lambda + (1-\lambda) \\ &= 1 \\ &\implies \lambda x + (1-\lambda)y \in \mathcal{B}(0,1) \end{aligned}$$

3. *Translation in a convex set gives a convex set. Let* K *be convex.* $K' = x_0 + K = \{x_0 + z \mid z \in K\}$ *Let* $x', y' \in K' \implies x' = x_0 + x$ *and* $y' = x_0 + y$*.*
   $$\begin{aligned} \implies \lambda x' + (1-\lambda)y' &= \lambda \cdot (x_0 + x) + (1-\lambda)(x_0 + y) \\ &= x_0 + \underbrace{\lambda x + (1-\lambda)y}_{\in K} \end{aligned}$$
   Especially: linear manifolds are convex. $B(x_0, 1)$ *is convex.*

4. $K \subseteq V$ *convex.* $f : V \to W$ *is linear.* $\implies f(K)$ *is convex.*

Optimization: Given a set M and a function $f : M \to \mathbb{R}$. Find $y \in M$ such that $f(y)$ is minimal.

Find $y \in M$ such that $d(x_0, y)$ is minimal. Compare with Figure 8.

Now if M is convex (consider M convex in $(\mathbb{R}^n, \|\cdot\|_2)$), there exists a unique element $y \in M$ such that $\|x_0 - y\|$ is minimal.

Finite elements (in computational mathematics) is the same idea.

Theorem 8.48. $(V, \langle \cdot, \cdot \rangle)$ *is a vector space with scalar product.* $K \subseteq V$ *is convex. Let* $x \in V$ *be given. Let* $y_0 \in K$*. Then the following statements are equivalent:*

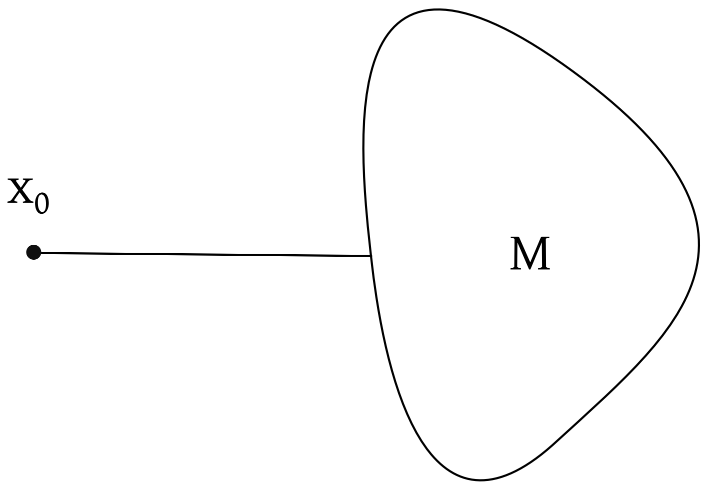


Figure 8: A generic optimization problem

1. $\forall y \in K : \left\|x - y_0\right\| \leq \left\|x - y\right\|$
2. $\forall y \in K : \Re \langle x - y_0, y - y_0 \rangle \leq 0$
3. $\forall y \in K \setminus \{y_0\} : \left\|x - y_0\right\| < \left\|x - y\right\|$

Compare with Figure 9. In the special case if $K = U$ is a subspace, then the following statement is given (equivalent to statement 2)

2'. $\forall y \in U : \langle x - y_0, y - y_0 \rangle = 0$

Proof.

1 → 2. Let $y \in K : 1 > \varepsilon > 0$.

$$y_\varepsilon = \underbrace{y_0 + \varepsilon(y - y_0)}_{\varepsilon y + (1-\varepsilon) y_0 \text{ because of convexity}} \in K$$

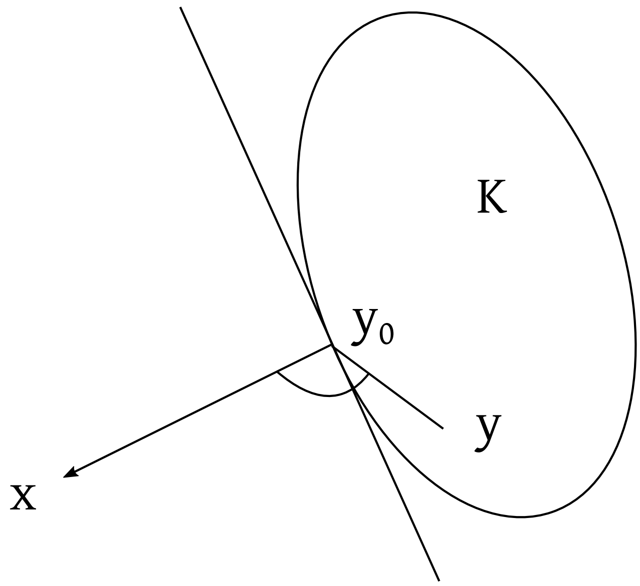


Figure 9: Optimization on a convex set

$$
\begin{aligned}
\forall \varepsilon \in (0,1): \left\|x - y_0\right\|^2 &\leq \left\|x - y_\varepsilon\right\|^2 \\
&= \left\|x - (y_0 + \varepsilon(y - y_0))\right\|^2 \\
&= \left\|(x - y_0) - \varepsilon(y - y_0)\right\|^2 \\
&= \left\|x - y_0\right\|^2 - 2\varepsilon \Re \left\langle x - y_0, y - y_0 \right\rangle + \varepsilon^2 \left\|y - y_0\right\|^2 \\
\implies \forall 0 < \varepsilon < 1 : 0 &\leq -2\varepsilon \Re \left\langle x - y_0, y - y_0 \right\rangle + \varepsilon^2 \left\|y - y_0\right\|^2 \\
&= \varepsilon \cdot \left(-2\Re \left\langle x - y_0, y - y_0 \right\rangle + \varepsilon \left\|y - y_0\right\|^2\right) \\
&\underbrace{\implies}_{\varepsilon \to 0} 0 \leq -2\Re \left\langle x - y_0, y - y_0 \right\rangle
\end{aligned}
$$

$2 \to 3$.

$$
\begin{aligned}
\left\|x - y\right\|^2 &= \left\|(x - y_0) + (y_0 - y)\right\|^2 \\
&= \left\|(x - y_0) - (y - y_0)\right\|^2 \\
&= \left\|x - y_0\right\|^2 + \left\|y - y_0\right\|^2 \underbrace{-2\Re \left\langle x - y_0, y - y_0 \right\rangle}_{\geq 0} \\
&\geq \left\|x - y_0\right\|^2 + \left\|y - y_0\right\|^2 \\
&> \left\|x - y_0\right\|^2 \\
&y \neq y_0
\end{aligned}
$$

$3 \to 1$. trivial.

$2 \to 2'$. Consider $K = U$ is subspace.

$$\forall y \in Y : \Re \langle x - y_0, y - y_0 \rangle \leq 0$$

U is a subspace.

$$\left\{ y - y_0 \,\middle|\, y \in U \right\} = \{ z \mid z \in U \} = U - y_0$$

$$\left.\begin{array}{ll} \forall z \in U : & \Re \langle x - y_0, z \rangle \leq 0 \\ \forall z \in U : & \Re \langle x - y_0, -z \rangle \leq 0 \end{array}\right\} \implies \forall z \in U : \Re \langle x - y_0, z \rangle = 0$$

Case $K = \mathbb{C}$:

$$i \cdot U = U$$

$$\implies z \in U : \Re \langle x - y_0, iz \rangle = 0$$

$$\Re \bar{i} \langle x - y_0, z \rangle = \Im \langle x - y_0, z \rangle$$

□

Corollary 8.49. *Let $(V, \langle \cdot, \cdot \rangle)$ be a vector space.*

1. *$K \subseteq V$ is convex, $x \in V$. Then the optimization problem*

$$\begin{cases} \left\| x - y \right\| = \min! \\ y \in K \end{cases}$$

has at most one solution.

2. *If $K = U$ subspace, then there exists at most one $y_0 \in U$ such that $x - y_0 \in U^\perp$.*

↓ *This lecture took place on 2018/04/23.*

Orthonormal basis:

$$\left\langle b_i, b_j \right\rangle = \delta_{ij}$$

$$v = \sum \lambda_i b_i \rightsquigarrow \langle v, b_i \rangle = \lambda_i$$

Given: an arbitrary basis of a subspace
Find: orthonormal basis of the subspace

$$K \subseteq V \text{ convex}$$

V with scalar product.

Then the optimization problem

$$\left\| x - y \right\| = \min \qquad Y \in K$$

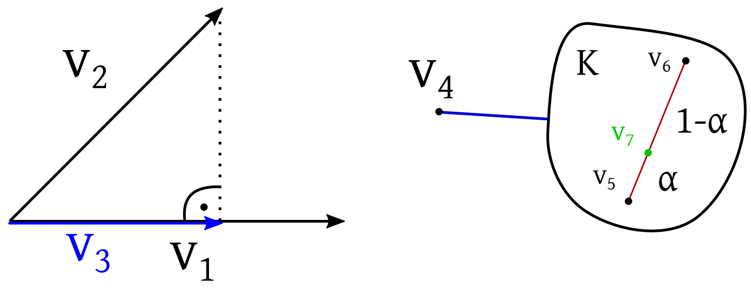


Figure 10: Projection of vector v_1 onto vector v_2 results in projected vector v_3 (left). K is a convex set (right). Due to convexity, there is a unique shortest path of v_4 to K. And a convex set satisfies for any two $v_5, v_6 \in K$ that $v_7 \in K$ with $v_7 = (v_6 - v_5)\lambda$ where $\lambda \in [0, 1]$

has at most one solution.

y is the solution.

$$\iff \Re \langle x - y_0, y - y_0 \rangle \leq 0 \forall y \in K$$

If K is the subspace U ($x - y_0 \perp U$), then

$$\Re \langle x - y_0, y \rangle = 0 \forall y \in K$$

$$U^\perp = \left\{ y \middle| \, y \perp U \right\}$$

is subspace.

$$U \cap U^\perp = \{0\}$$

If $x \in U \cap U^\perp$, then $x \perp x = \langle x, x \rangle = ||x||^2 = 0$.

Orthogonal complement: $U + U^\perp$ is direct sum.
Every $x \in U + U^\perp$ has a unique decomposition.

$$x = u + v \qquad u \in U, v \in U^\perp$$

The maps $x \mapsto u$ and $x \mapsto v$ are linear.

Definition 8.50. *Assume $U \dot{+} U^\perp = V$. Then the projection maps*

$$\pi_U : V \to V \qquad \pi_U^\perp : V \to V$$

such that $\pi_U(x) \in U$ and $\pi_{U^\perp}(x) \in U^\perp$ and $x = \pi_U(x) + \pi_{U^\perp}(x)$ are orthogonal projections.

Remark. 1. $x \in U \iff \pi_U(x) = x \iff \pi_{U^\perp}(x) = 0$

2. $x \in U^\perp \iff \pi_U(x) = 0 \iff \pi_{U^\perp}(x) = x$

3. $\pi_{U^\perp} = \mathrm{id} - \pi_U$

$$\pi_U(x) \in U$$
$$\implies \text{remark (4): } \pi_U(\pi_U(x)) = \pi_U(x)$$
$$(\sim)\text{: } \pi_U \circ \pi_U = \pi_U \text{ idempotent}$$
$$\pi_U \text{ is linear: } \pi_U \circ \pi_{U^\perp} = 0$$

Theorem 8.51. *Let $V = U \dot{+} U^\perp$.*

1. $\forall x, y \in V : \langle x, \pi_U(y)\rangle = \langle \pi_U(x), y\rangle = \langle \pi_U(x), \pi_U(y)\rangle$
2. *Compare with Figure 11.*

$$\left\|\pi_U(x)\right\| \leq \|x\| \wedge \left\|\pi_U(x)\right\| = \|x\| \iff x \in U$$

Proof:

(a)
$$x = \pi_U(x) + \pi_{U^\perp}(x) \qquad y = \pi_U(y) + \pi_{U^\perp}(y)$$
$$\langle x, \pi_U(y)\rangle = \langle \pi_U(x) + \pi_{U^\perp}(x), \pi_U(y)\rangle = \langle \pi_U(x), \pi_U(y)\rangle + \underbrace{\Big\langle \underbrace{\pi_U(x)}_{\in U^\perp}, \underbrace{\pi_U(y)}_{\in U} \Big\rangle}_{=0}$$
$$\langle \pi_U(x), y\rangle = \langle \pi_U(x), \pi_U(y)\rangle + \langle \pi_U(x), \pi_{U^\perp}(y)\rangle$$

(b)
$$x = \pi_U(x) + \pi_{U^\perp}(x)$$
$$\implies \|x\|^2 = \left\|\pi_U(x)\right\|^2 + \left\|\pi_{U^\perp}(x)\right\|^2 \geq \left\|\pi_U(x)\right\|^2$$
$$\textit{By equality} \iff \left\|\pi_{U^\perp}(x)\right\| = 0 \iff x = \pi_U(x) \iff x \in U$$

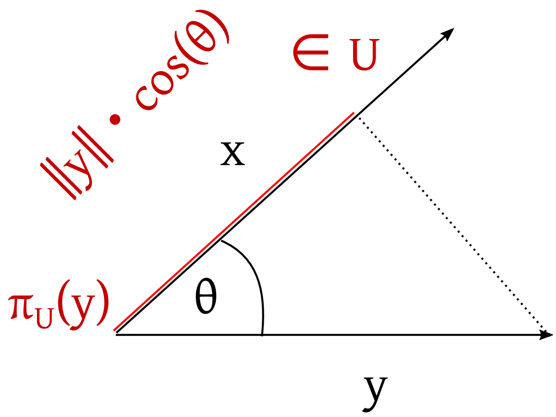


Figure 11: Projection

Person. *Jørgen Pedison Gram (1850–1916)*

Definition 8.52. *Let* $v_1, v_2, \ldots \in V$.

$$\operatorname{Gram}(v_1, \ldots, v_m) = \begin{bmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \ldots & \langle v_1, v_m \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & \ldots & \langle v_2, v_m \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle v_m, v_1 \rangle & \langle v_m, v_2 \rangle & \ldots & \langle v_m, v_m \rangle \end{bmatrix}$$

is called Gram matrix of tuple $v_1, v_2, \ldots, v_m$

Remark 8.53. *In case* $V = \mathbb{C}^n$.

$$\langle v, w \rangle = \overline{w}^T \cdot v = \sum_{i=1}^n \lambda_i \overline{\mu_i} = (\overline{\mu}_1, \ldots, \overline{\mu}_n) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

$$v = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \qquad w = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

Hence, if

$$v_i = \begin{pmatrix} \beta_{1i} \\ \vdots \\ \beta_{ni} \end{pmatrix} \qquad i = 1, \ldots, m$$

$$\begin{aligned} V &= \begin{pmatrix} v_1 & v_2 & \ldots & v_m \\ \vdots & \vdots & & \vdots \end{pmatrix} \in \mathbb{C}^{n \times m} \\ &= \begin{pmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1m} \\ \vdots & \vdots & & \vdots \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{nm} \end{pmatrix} \\ (V^* V)_{ij} &= \sum_{k=1}^n (v^*)_{ik} v_{kj} = \sum_{k=1}^n \overline{\beta_{ki}} \beta_{kj} = \overline{\langle v_i, v_j \rangle} \\ &= \begin{pmatrix} v_1^* & \ldots \\ \vdots & \\ v_m^* & \ldots \end{pmatrix} \begin{pmatrix} v_1 & \ldots & v_m \\ \vdots & & \vdots \end{pmatrix} \\ V^* V &= \overline{\operatorname{Gram}(v_1, \ldots, v_m)} \end{aligned}$$

Theorem 8.54. *Let* $v_1, \ldots, v_m \in V$. $G = \operatorname{Gram}(v_1, \ldots, v_m)$.

1. $G = G^*$ *is Hermitian, positive* semi*definite.*

$$\xi^T \cdot G \cdot \overline{\xi} = \left\| \sum_{i=1}^m \xi_i v_i \right\|^2 \geq 0$$

2. $\xi \in \ker G \iff \sum_{i=1}^m \overline{\xi_i} v_i = 0$

3. *G is positive definite iff $(v_1, \ldots, v_m)$ are linear independent.*

Proof. 1. $g_{ij} = \left\langle v_i, v_j \right\rangle = \overline{\left\langle v_j, v_i \right\rangle} = \overline{g_{ji}}$

$$\xi^T \cdot G \cdot \overline{\xi} = \sum_{i=1}^n \sum_{j=1}^n \xi_i g_{ij} \overline{\xi_j} = \sum_{i=1}^n \sum_{j=1}^n \xi_i \overline{\xi_j} \left\langle v_i, v_j \right\rangle = \left\langle \sum_{i=1}^n \xi_i v_i, \sum_{j=1}^n \xi_j v_j \right\rangle = \left\| \sum_{i=1}^n \xi_i v_i \right\|^2$$

2. Direction $\implies$. $\xi \in \ker G \implies G\xi = 0 \implies \xi^T \cdot G \cdot \xi = 0$

$$\xi^T \cdot G \cdot \xi = \xi^T \cdot G \cdot \overline{\overline{\xi}} \underbrace{=}_{(1)} \left\| \sum_{i=1}^m \overline{\xi_i} v_i \right\|^2$$

Direction $\impliedby$. If $\left\| \sum_1^m \xi_i v_i \right\| = 0$

$$(G \cdot \xi)_i = \sum_{j=1}^n \left\langle v_i, v_j \right\rangle \xi_j = \sum_{j=1}^n \left\langle v_i, \overline{\xi_j} v_j \right\rangle = \left\langle v_i, \underbrace{\sum_{j=1}^n \overline{\xi_j}\, v_j}_{=0} \right\rangle = 0$$

$$\implies G \cdot \xi = 0$$

3. G is positive definite

$$\begin{aligned}
&\iff \forall \xi \neq 0 : \xi^T \cdot G \cdot \overline{\xi} > 0 \\
&\iff \forall \xi \neq 0 : \left\| \sum_1^m \xi_i \cdot v_i \right\|^2 > 0 \\
&\iff \forall \xi \neq 0 : \sum_{i=1}^m \xi_i v_i \neq 0 \\
&\iff (v_1, \ldots, v_m) \text{ is linear independent} \\
&\iff \ker G = \{0\} \\
&\iff G \text{ is invertible}
\end{aligned}$$

□

Theorem 8.55. *Let $U \subseteq V$ be a subspace. V is a vector space with scalar product.*

$$(u_1, \ldots, u_m) \text{ is basis of } U \qquad G = \mathrm{Gram}(u_1, \ldots, u_m) = \left[\left\langle u_i, u_j \right\rangle\right]_{i,j=1,\ldots,m}$$

Then the projection is $\pi_U(x) = \sum_{j=1}^m \eta_j u_j$ where $\eta = \overline{G}^{-1} \cdot \begin{pmatrix} \langle x, u_1 \rangle & \ldots & \langle x, u_m \rangle \end{pmatrix}^T$. Additionally, if $u_1, \ldots, u_m$ would be an orthonormal basis, then the coordinate of x is given by $\begin{pmatrix} \langle x, u_1 \rangle & \ldots & \langle x, u_m \rangle \end{pmatrix}^T$.

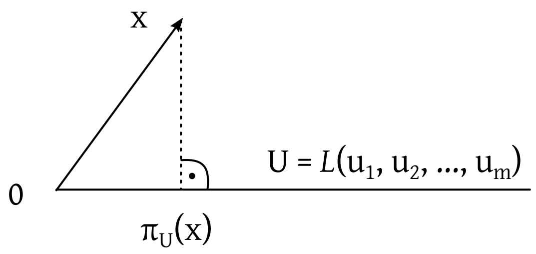


Figure 12: Projection $\pi_U(x)$ of x onto U

Proof. Let $u = \sum_{j=1}^{m} \eta_j u_j$. Compare with Figure 12. Show that $x - u \in U^\perp = \mathcal{L}(u_1, \dots, u_m)^\perp = \{u_1, \dots, u_m\}^\perp = \bigcap_{i=1}^{m} u_i^\perp$. Hence, show that $x - u \perp u_i \forall i \in \{1, \dots, m\}$.

$$\begin{aligned}\langle u_i, u\rangle &= \left\langle u_i, \sum_{j=1}^{m} \eta_j u_j \right\rangle = \sum_{j=1}^{m} \left\langle u_i, u_j \right\rangle \cdot \overline{\eta_j} = \sum_{j=1}^{m} g_{ij}\overline{\eta_j} \\ &= (G\overline{\eta})_i = \langle u_i, x\rangle\end{aligned}$$

because

$$\overline{G} \cdot \eta = \begin{pmatrix} \langle x, u_1 \rangle \\ \vdots \\ \langle x, u_m \rangle \end{pmatrix} \qquad G \cdot \overline{\eta} = \begin{pmatrix} \overline{\langle x, u_1 \rangle} \\ \vdots \\ \overline{\langle x, u_m \rangle} \end{pmatrix} = \begin{pmatrix} \langle u_1, x \rangle \\ \vdots \\ \langle u_m, x \rangle \end{pmatrix}$$

Hence, $\forall i \in \{1, \dots, m\}$:

$$\langle u_i, u\rangle = \langle u_i, x\rangle \implies \forall i \in \{1, \dots, m\} : \langle u_i, x - u\rangle = 0 \implies x - u \in \{u_1, \dots, u_m\}^\perp$$

□

Example 8.56. *Find polynomial $p(t)$ of degree 2 such that*

$$\int_0^1 \left|t^3 - p(t)\right|^2 \, dt \overset{!}{=} \min$$

$V = C[0, 1]$, *scalar product*

$$\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}\, dt$$

$$\begin{aligned} U &= \text{ polynomial function of degree } \leq 2 \\ x &= t \mapsto t^3 \notin U \end{aligned}$$

Find $p \in U$ such that $\left\|x - p\right\|^2 \overset{!}{=} \min$

$$\left\|x - p\right\|^2 = \int \left|x(t) - p(t)\right|^2 \, dt$$

Basis of $U = \mathcal{L}(\{1, t, t^2\})$

$$u_i(t) = t^{i-1} \qquad i = 1, 2, 3$$

Gram matrix:

$$g_{ij} = \left\langle u_i, u_j \right\rangle = \int_0^1 t^{i-1} t^{j-1} \, dt = \int_0^1 t^{i+j-2} \, dt = \left. \frac{t^{i+j-1}}{i+j-1} \right|_0^1 = \frac{1}{i+j-1}$$

$$G = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

Hilbert matrix:

$$\left[\frac{1}{i+j-1} \right]_{i,j=1,\dots,k}$$

This matrix is very unstable (in the equation system $Gx = b$*) and therefore a very important test matrix in computational mathematics (ie. Numerics).*

$$u = \sum_{j=1}^{3} \eta_j u_j \qquad \eta = \overline{G}^{-1} \cdot \begin{pmatrix} \langle x, u_1 \rangle \\ \langle x, u_2 \rangle \\ \langle x, u_3 \rangle \end{pmatrix}$$

$$\left\langle x, u_j \right\rangle = \int_0^1 x(t) u_j(t) \, dt = \int_0^1 t^3 \cdot t^{j-1} \, dt = \int_0^1 t^{2+j} \, dt = \frac{1}{3+j}$$

$$\eta = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}^{-1} \begin{pmatrix} \frac{1}{4} \\ \frac{1}{5} \\ \frac{1}{6} \end{pmatrix} = \begin{bmatrix} 9 & -36 & 30 & & & \\ -36 & 192 & -180 & 30 & -180 & 180 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{5} \\ \frac{1}{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{20} \\ -\frac{3}{5} \\ \frac{3}{2} \end{bmatrix}$$

(Assume that we don't know 180 *in the bottom-right corner precisely. Consider* $180 + \varepsilon$*, then this error* ε *explodes tremendously in the solution).*

Corollary 8.57. *Special case* u_i *is orthonormal basis of* U *(*$\rightarrow G = I$*) Then it holds that*

1. $\forall v \in V : \pi_U(v) = \sum_{i=1}^{m} \langle v, v_i \rangle \cdot u_i$

2. $$\|v\|^2 \geq \sum_{i=1}^{m} |\langle v, v_i \rangle|^2 \qquad \textit{(Bessel's inequality)}$$

$$\|v\|^2 = \sum_{i=1}^{m} |\langle v, u_i \rangle|^2 \iff v \in U \qquad \textit{(Parseval's identity)}$$

$$\eta_j = \overline{G}^{-1} \begin{pmatrix} \langle v, u_1 \rangle \\ \vdots \\ \langle v, u_m \rangle \end{pmatrix}$$

Person. *Friedrich Bessel (1784–1846)*

Person. *Marc-Antoine Parseval (1755–1836)*

Proof. Gram's matrix = I.

$$\eta_j = \left\langle v, u_j \right\rangle$$

□

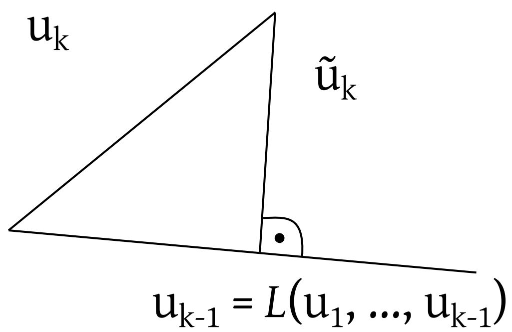


Figure 13: Projection used in the Gram-Schmidt process

8.7 Gram-Schmidt process

Given: finite, linear independent vectors $v_1, \dots, v_m$
Find: orthonormal basis of $(v_1, \dots, v_m)$.

Theorem 8.58 (Gram–Schmidt process for orthogonalization)**.** *Let $(v_1, \dots, v_m) \subseteq V$ be linear independent. Let $U := \mathcal{L}(v_1, \dots, v_m)$. Then $\exists u_1, \dots, u_m$ as orthonormal basis of U, specifically inductive*

$$u_1 = \frac{v_1}{||v_1||}$$

and for $k = 2, \dots, m$:

$$\tilde{u}_k = v_k - \sum_{j=1}^{k-1} \left\langle v_k, u_j \right\rangle \cdot u_j \qquad u_k = \frac{\tilde{u}_k}{||u_k||}$$

Proof. **Induction base** $k = 1$ is trivial

Induction step $k - 1 \to k$. Assume

$$\mathcal{L}(u_1, \dots, u_{k-1}) = \mathcal{L}(v_1, \dots, v_{k-1}) =: U_{k-1}$$

$$\tilde{u}_k = v_k - \pi_{U_{k-1}}(v_k) \in U_{k-1}^{\perp} \text{ because of Theorem 8.57}$$

$$\implies \tilde{u}_k \perp u_1, \dots, u_{k-1} \implies (u_1, \dots, u_{k-1}, \frac{\tilde{u}_k}{||\tilde{u}_k||})$$

is an orthonormal basis.

$$\mathcal{L}(u_1, \dots, u_{n-1}, \frac{\tilde{u}_k}{||\tilde{u}_k||}) = \mathcal{L}(u_1, \dots, u_{k-1}, v_k)$$

then $\tilde{u}_k - v_k \in \mathcal{L}(u_1, \dots, u_{k-1})$

$\square$

↓ *This lecture took place on 2018/04/25.*

Gram-Schmidt process:

$$\mathcal{L}(v_1, v_2) = \mathcal{L}(v_2 - p(v_2), v_1) \qquad v_2 - p(v_2) \perp v_1$$

Given: $v_1, \dots, v_m$

$$u_i = \frac{v_i}{\|v_i\|}$$

$$\tilde{u}_k = v_k - \sum_{i=1}^{k-1} \langle v_k, u_i \rangle \cdot u_i$$

$$u_k = \frac{\tilde{u}_k}{\|\tilde{u}_k\|} \qquad \frac{\langle v_k, \tilde{u}_i \rangle \, \tilde{u}_i}{\|\tilde{u}_i\|^2}$$

Example. *Let $V = \mathbb{R}^3$.*

$$\langle x, y \rangle = x^t A y$$

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

$$v_i = \text{ standard basis } e_i$$

$$\|v_1\|^2 = \langle v_1, v_1 \rangle = v_1^T A v_1 = a_{11} = 1$$

$$\|v_2\|^2 = \langle v_2, v_2 \rangle = a_{12} = 3$$

$$u_1 = \frac{v_1}{\|v_1\|} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\tilde{u}_2 = v_2 - u_1 \langle v_2, u_1 \rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot (0\,1\,0) A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} \qquad \|\tilde{u}_2\|^2 = \langle \tilde{u}_2, \tilde{u}_2 \rangle = (1\,1\,0) \cdot A \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 2 \qquad u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\tilde{u}_3 = v_3 - u_1 \langle v_3, u_1 \rangle - u_2 \langle v_3, u_2 \rangle$$

$$= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \underbrace{(0\,0\,1) \cdot A \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{a_{31}=1} - \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \cdot \underbrace{(0\,0\,1) \cdot A \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}_{a_{31}+a_{32}=0} \cdot \frac{1}{\sqrt{2}} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$\|\tilde{u}_3\|^2 = (-1\,0\,1) \cdot A \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 1 - 1 - 1 + 2 = 1 \qquad u_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Remark 8.59. *This is an alternative method to build orthogonal projection on subspace $U \subseteq \mathbb{C}^n$ with standard scalar product.*

1. *Determine an orthonormal basis of U:* $u_1, \dots, u_m \in \mathbb{C}^{n\times 1}$
2. $P = \sum_{i=1}^{m} u_1 \cdot u_i^*$

$$P \cdot v = \sum_{i=1}^{m} u_i \underbrace{u_i^* \cdot v}_{\langle v, v_i \rangle} = \sum_{i=1}^{m} u_i \langle v, v_i \rangle$$

Gram matrix = I

Example 8.60 (Example 8.56 again)**.**

$$V = C[0,1] \qquad U = \mathcal{L}(1, x, x^2) =: \mathcal{L}(v_1, v_2, v_3)$$

$$\langle f, g \rangle = \int_0^1 f(t)\overline{g(t)}\, dt$$

Orthonormal basis:

$$\|v_i\|^2 = \int_0^1 1^2\, dt = 1$$

$$u_1 = 1$$

$$\tilde{u}_2 = v_2 - u_1 \cdot \langle v_2, u_1 \rangle = x - 1 \cdot \underbrace{\int_0^1 t \cdot 1\, dt}_{=\frac{1}{2}} = x - \frac{1}{2}$$

$$\|\tilde{u}_2\|^2 = \int_0^1 (t - \frac{1}{2})^2\, dt = \left.\frac{(t-\frac{1}{2})^3}{3}\right|_0^1 = \frac{(\frac{1}{2})^3 - (-\frac{1}{2})^2}{3} = \frac{1}{12}$$

$$u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} = \sqrt{12} \cdot (x - \frac{1}{2})$$

$$\begin{aligned}
\tilde{u}_3 &= v_3 - u_1 \langle v_3, u_1 \rangle - u_2 \cdot \langle v_3, u_2 \rangle \\
&= x^2 - 1 \cdot \underbrace{\int_0^1 t^2 \cdot 1\, dt}_{=\frac{1}{3}} - \sqrt{12}(x - \frac{1}{2}) \int_0^1 t^2 \sqrt{12}(t - \frac{1}{2})\, dt \\
&= x^2 - \frac{1}{3} - 12(x - \frac{1}{2}) \cdot \frac{1}{12} \\
&= x^2 - x + \frac{1}{6}
\end{aligned}$$

Side note:

$$\int_0^1 t^2 (t - \frac{1}{2})\, dt = \int_0^1 (t^3 - \frac{1}{2}t^2)\, dt = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$$

$$\|\tilde{u}_3\|^2 = \int_0^1 (t^2 - t + \frac{1}{6})^2\, dt = \frac{1}{180}$$

$$\implies u_3 = \sqrt{180} \cdot (x^2 - x + \frac{1}{6})$$

Projection:

$$\int_0^1 (t^3 - p(t))^2 \, dt = \min!$$

Solution: $\pi_U(x^3) \qquad U = \mathcal{L}(1, x, x^2)$

$$\begin{aligned}\pi_U(x^3) &= u_1 \left\langle x^3, u_1 \right\rangle + u_2 \left\langle x^3, u_2 \right\rangle + u_3 \left\langle x^3, u_3 \right\rangle \\ &= 1 \cdot \int_0^1 t^3 \cdot 1 \, dt + \sqrt{12}\left(x - \frac{1}{2}\right) \int_0^1 t^3 \sqrt{12}\left(t - \frac{1}{2}\right) dt \\ &+ \sqrt{180}\left(x^2 - x + \frac{1}{6}\right) \int_0^1 t^3 \sqrt{180}\left(t^2 - t + \frac{1}{6}\right) dt\end{aligned}$$

Consider $\langle f, g \rangle := \int_{-1}^1 \sqrt{1 - t^2} f(t) \overline{g(t)} \, dt$. Take $1, x, x^2, \dots$ and apply Gram-Schmidt process to retrieve the Chebyshev polynomials.

$$\int_0^1 f(t) g(t) \, dt \qquad \text{Laguerre polynomials}$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} f(t) g(t) \, dt \qquad \text{Hermite polynomials}$$

8.8 Riesz representation theorem

Person. *Frigyes Riesz (1880–1956)*

Theorem 8.61. *Let $(V, \langle \cdot, \cdot \rangle)$ be a vector space with scalar product* $\dim V < \infty$.

V^* is the dual space, namely $\mathrm{Hom}(V, \mathbb{K})$ *which is the space of linear functionals. For some fixed $y \in V$, the map $T_y(x) = \langle x, y \rangle$ is linear in x (and therefore $T_y \in V^*$).*

Then the map $V \to V^*$ (where an element of V^* is a map $V \to \mathbb{K}$ with $x \mapsto \langle x, y \rangle$) is an antilinear isomorphism (antiisomorphism).

This is trivial in $\mathbb{R}$, but in $\mathbb{C}$ it is much more complex (pun intended).

Hence,

1. For every y it holds that $Ty \in V^*$
2. For every linear functional $f \in V^* : \exists! y \in V : f = T_y$
3. Let $y \mapsto Ty$ be an antilinear map.

$$T_{\lambda y_1 + \mu y_2} = \overline{\lambda} T y_1 + \overline{\mu} T y_2$$

Example (For point 2)**.**

$$V = C[0, 1]$$

Scalar product: $\langle f, g \rangle = \int f(t) g(t) \, dt$. *Let $F : C[0, 1] \to \mathbb{R}$ linear. Then by the Riesz representation theorem, there exists $g \in C[0, 1] : F(f) = \int f(t) g(t) \, dt$.*

For example $f \to f(1)$

$$\exists g(t) : f(1) = \int_0^1 f(t) g(t) \, dt$$

In physics, e.g. the Dirac delta function.

Proof of point 3. We show linearity.

$$Ty(x) = \langle x, y \rangle \text{ is linear in } X \implies T_y \in V^*$$

$$\begin{aligned}\forall x \in V : T_{\lambda y_1 + \mu y_2}(x) &= \langle x, \lambda y_1 + \mu y_2 \rangle = \overline{\lambda} \langle x, y_1 \rangle + \overline{\mu} \langle x, y_2 \rangle \\ &= \overline{\lambda} T y_1(x) + \overline{\mu} T y_2(x) = (\overline{\lambda} T y_1 + \overline{\mu} T y_2)(x) \\ &\implies T_{\lambda y_1 + \mu y_2} = \overline{\lambda} T y_1 + \overline{\mu} T y_2\end{aligned}$$

We show injectivity: the map $y \mapsto Ty$ is injective.

Assume: $Ty = 0$ (zero functional). Show $y = 0$. $Ty = 0$ means $\forall x \in V : Ty(x) = 0$, especially for $x = y$, $T_y(y) = \langle y, y \rangle = 0 \implies y = 0$.

We show surjectivity: the map $y \mapsto Ty$ is surjective.

Let $u_1, \ldots, u_n$ be an orthonormal basis (exists because of Gram-Schmidt).

Given: $f \in V^*$. Find: y such that $f = Ty$.

$$\text{Hence, } \forall x \in V : f(x) = \langle x, y \rangle \quad \underbrace{\iff}_{\text{by Fortsetzungssatz}} \quad f(u_i) = \langle u_i, y \rangle$$

Let $y = \sum_{j=1}^{n} \overline{f(u_j)} \cdot u_j$.

$$\implies \langle u_i, y \rangle = \left\langle u_1, \sum_{j=1}^{n} \overline{f(u_j)} u_j \right\rangle = \sum_{j=1}^{n} f(u_j) \underbrace{\langle u_i, u_j \rangle}_{\delta_{ij}} = f(u_i)$$

Hence, y satisfies the condition. ☐

Remark. *The Riesz representation theorem also holds in infinite dimensions (thus, in Hilbert spaces, a generalization of Euclidean spaces). In those spaces, there exists some Hilbert base:*

$$(u_i)_{i \in I} : x = \sum_{i \in I} \langle x, u_i \rangle \cdot u_i \forall x$$

So every x has such a representation and in infinite dimensions, this representation is a series.

Corollary 8.62.

1. $v = 0 \iff \forall w \in V : \langle v, w \rangle = 0$
2. $\|v\| = \sup \{ |\langle v, w \rangle| \mid \|w\| \leq 1 \}$

Equivalently in the dual space:

1. $v = 0 \iff \forall f \in V^* : f(v) = 0$
2. $\|v\| = \sup \left\{ \left| f(v) \right| \,\middle|\, f \in V^*, \left\| f \right\| \leq 1 \right\}$

holds in general in a normed space.

Remark. *We make a small revision: dual space* $V^* = \mathrm{Hom}(V, \mathbb{K})$

$$W \xrightarrow{T} V \xrightarrow{f} \mathbb{K}$$

$$\implies f \circ T : W \to \mathbb{K} \in W^*$$

is a linear functional on W*. Hence, the map* $\mathrm{Hom}(V, \mathbb{K}) \to \mathrm{Hom}(W, \mathbb{K})$ *and* $f \mapsto f \cdot T$ *is linear.*

$$(\lambda f + \mu g) \circ T = \lambda \cdot f_0 T + \mu g \circ T \qquad \text{“transposed map”}$$

Linear map: $T^* : V^* \to W^*$.

Let V, W *be spaces with a scalar product. Then* $V \simeq V^*$ *and* $W \simeq W^*$ *where* $\simeq$ *means anti-isomorphic.* $T : W \to V \implies T^* : V \to W$.

8.9 Adjoint maps

Definition 8.63 (Theorem and definition)**.** *Let* $(V, \langle \cdot, \cdot \rangle_V)$ *and* $(W, \langle \cdot, \cdot \rangle_W)$ *be spaces with a scalar product.* $\dim V, \dim W < \infty$.

$$T \in \mathrm{Hom}(W, V) \text{ hence, } T : W \to V \text{ linear}$$

1. *For every* $v \in V$ *the map* $w \mapsto \langle T(w), v \rangle_V$ *is linear.*
2. $\forall v \in V \exists! u \in W \forall w \in W : \langle T(w), v \rangle_V = \langle w, u \rangle_W$ *and* $T^*(v) = u$.
   Hence,
   $$\langle T(w), v \rangle_V = \langle w, T^*(v) \rangle_W \qquad \forall w \in W \quad \forall v \in V$$
3. *The map* $T^* : V \to W$ *with* $v \mapsto u$ *is linear, hence* $T^* \in \mathrm{Hom}(V, W)$ *and is called* adjoint map.
4. *The map* $\mathrm{Hom}(W, V) \mapsto \mathrm{Hom}(V, W)$ *with* $T \mapsto T^*$ *is antilinear and* $T^{**} = T$.

Proof. 1. $\langle T(w), v \rangle = T_V(T(w)) = T_v \circ T(w)$
   Composition of linear maps is linear.

2. $T_V \circ T \in W^*$. By Riesz representation theorem, $\exists! u \in W : T_V \circ T(w) = \langle w, u \rangle \, \forall w \in W \implies \langle T(w), v \rangle = \langle w, u \rangle$

3. Show that,
   $$\forall v_1, v_2 \in V \forall \lambda, \mu : T^*(\lambda v_1 + \mu v_2) = \lambda T^*(v_1) + \mu T^*(v_2)$$
   It suffices to show that
   $$\langle w, T^*(\lambda v_1 + \mu v_2) \rangle = \langle w, \lambda T^*(v_1) + \mu T^*(v_2) \rangle \, \forall w \in W$$
   Compare with corollary: $w_1 = w_2$ in $W \iff \forall w : \langle w, w_1 \rangle = \langle w, w_2 \rangle$.
   $$\begin{aligned} \langle w, T^*(\lambda v_1 + \mu v_2) \rangle &= \langle T(w), \lambda v_1 + \mu v_2 \rangle \\ &= \overline{\lambda} \langle T(w), v_1 \rangle + \overline{\mu} \langle T(w), v_2 \rangle \\ &= \overline{\lambda} \langle w, T^*(v_1) \rangle + \overline{\mu} \langle w, T^*(v_2) \rangle \\ &= \langle w, \lambda T^*(v_1) \rangle + \langle w, \mu T^*(v_2) \rangle \\ &= \langle w, \lambda T^*(v_1) + \mu T^*(v_2) \rangle \end{aligned}$$

4. Show $(\lambda T_1 + \mu T_2)^* = \overline{\lambda} T_1^* + \overline{\mu} T_2^*$.

$$\iff \forall v \in V : (\lambda T_1 + \mu T_2)^* v = (\overline{\lambda} T_1^* + \overline{\mu} T_2^*)(v)$$
$$\forall v \in V \forall w \in W : \langle w, (\lambda T_1 + \mu T_2)^*(v) \rangle = \left\langle w, (\overline{\lambda} T_1^* + \overline{\mu} T_2^*)(v) \right\rangle$$

Hence,

$$\begin{aligned}
\langle w, (\lambda T_1 + \mu T_2)^*(v) \rangle &= \langle (\lambda T_1 + \mu T_2)(w), v \rangle \\
&= \lambda \langle T_1(w), v \rangle + \mu \langle T_2(w), v \rangle \\
&= \lambda \langle w, T_1^*(v) \rangle + \mu \langle w, T_2^*(v) \rangle \\
&= \left\langle w, \overline{\lambda} T_1^*(v) \right\rangle + \langle w, \overline{\mu} T_2^*(v) \rangle \\
&= \left\langle w, \overline{\lambda} T_1^*(v) + \overline{\mu} T_2^*(v) \right\rangle \\
&= \left\langle w, (\overline{\lambda} T_1^* + \overline{\mu} T_2^*)(v) \right\rangle
\end{aligned}$$

$$T : W \to V \qquad T^* : V \to W \qquad T^{**} : W \to V$$

Show that $\forall w \in W : T^{**}(w) = T(w)$. Hence $\forall w \in W \forall v \in V : \langle T^{**}(w), v \rangle_V = \langle T(w), v \rangle_V$

$$\begin{aligned}
\langle T^{**}(w), v \rangle_V &= \overline{\langle v, T^{**}(w) \rangle} = \overline{\langle T^*(v), v \rangle} = \langle w, T^*(v) \rangle \\
&= \langle T(w), v \rangle \\
\langle Tw, v \rangle &= \langle w, T^* v \rangle
\end{aligned}$$

If $V = W$, then $T = T^*$.

$\square$

Assume $u = D^*(x)$ exists $\in \mathbb{R}[x]$

$$\implies M := \max_{t \in [0,1]} \left| u(t) \right| < \infty$$
$$\left| \|x^n| D^*(x) \right| = \left| \int_0^1 t^n \cdot u(t)\, dt \right| \le \int_0^1 t^n \cdot M\, dt = \frac{M}{n+1}$$
$$\implies \frac{n}{n+1} \le \frac{M}{n+1} \forall n \in \mathbb{N}$$
$$\implies u(x) \notin \mathbb{R}[x]$$

Example 8.64 (For Definition 8.63, point 3)**.** *If* $\dim V = \infty$*, then not every linear map has an adjoint map!*

$$V = \mathbb{R}[x] \qquad \langle f, g \rangle = \int_0^1 f(t) g'(t)\, dt$$
$$D : V \to V \qquad p(x) \mapsto p'(x)$$

Recall: The derivative of a linear combination is the linear combination of derivatives. Assume D *has an adjoint* D^**.*

$$\implies \langle x^n, D^*(x) \rangle = \langle D(x^n), x \rangle = \int_0^1 n t^{n-1} t\, dt = \frac{n}{n+1}$$

↓ *This lecture took place on 2018/05/02.*

Riesz representation theorem
V with scalar product
$\operatorname{Hom}(V, \mathbb{K}) \simeq V$ where $\simeq$ is antilinear
$\forall f \in \operatorname{Hom}(V, \mathbb{K}) : \exists! y \in V : f = T_y$

$$T_y(x) = \langle x, y \rangle$$

$$T_{\lambda x + \mu y} = \overline{\lambda} T_x + \overline{\mu} T_y$$

For $f \in \operatorname{Hom}(V, W)$, the map $x \mapsto \langle f(x), y \rangle \in \operatorname{Hom}(V, \mathbb{K})$

$$\implies \exists! z \in V : \forall x \in V : \langle f(x), y \rangle = \langle x, z \rangle$$

$$z =: f^*(y) \text{ ...adjoint map}$$

$$f^* : W \to V \text{ is linear}$$

$$\operatorname{Hom}(V, W) \to \operatorname{Hom}(W, V)$$

$$f \mapsto f^*$$

is an antilinear *involution*.

$$f^{**} = f$$

8.10 The linear adjoint map is the complex transpose

Theorem 8.65. *Let $B \subseteq V, C \subseteq W$ be orthonormal bases. $f \in \operatorname{Hom}(V, W)$.*

$$\Phi_B^C(f^*) = \Phi_C^B(f)^* = \overline{\Phi_C^B(f)^T}$$

Proof.

$$A = \Phi_C^B(f)$$

Column $s_j(A)$ are the coordinates of $b_j \in B$ in regards of basis C.

$$\begin{aligned} a_{ij} &= \text{ i-th coordinate of } f(b_j) \\ &= \Phi_C(f(b_j))_i = \left\langle f(b_j), c_i \right\rangle \\ &= \left\langle b_j, f^*(c_i) \right\rangle = \overline{\left\langle f^*(c_i), b_j \right\rangle} \\ &= \text{ j-th coordinate of } f^*(c_i) \\ &= \overline{\Phi_B^C(f^*)_{ji}} = \overline{\tilde{a}_{ji}} \text{ if } \tilde{A} = \Phi_B^C(f^*) \end{aligned}$$

□

Theorem 8.66. *Let U, V, W be finite-dimensional.*

$$U \xrightarrow{f} V \xrightarrow{g} W$$

1. $(g \circ f)^* = f^* \circ g^*$
2. $f^{**} = f$
3. $\ker f = (\operatorname{im} f^*)^\perp$
4. $\operatorname{im} f = (\operatorname{kern} f^*)^\perp$
5. *f injective $\iff$ f^* surjective*
6. *f surjective $\iff$ f^* injective*

Proof. 1. Let $u \in V, w \in W$

$$\begin{aligned}\langle (g \circ f)(u), w\rangle_W &= \langle g(f(u)), w\rangle_W \\ &= \langle f(u), g^*(w)\rangle_V \\ &= \langle u, f^*(g^*(w))\rangle_U\end{aligned}$$

holds $\forall u \in U \forall w \in W$. By definition

$$\langle (g \circ f)(u), w\rangle_W = \langle u, (g \circ f)^*(w)\rangle$$

Hence,

$$\implies (g \circ f)^* = f^* \circ g^*$$

3. Show that
   - $\operatorname{kern} f \subseteq (\operatorname{im} f^*)^\perp$
   - $(\operatorname{im} f^*)^\perp \subseteq \operatorname{kern} f$

   Proof:

   - Let $u \in \operatorname{kern} f$. Show that $\forall y \in \operatorname{im} f^* : \langle u, y\rangle = 0$

   $$y \in \operatorname{im} f^* \implies \exists v \in V : y = f^*(v)$$

   $$\langle u, y\rangle_U = \langle u, f^*(v)\rangle_U = \left\langle \underbrace{f(u)}_{=0}, v \right\rangle_V = 0$$

   - Let $u \in (\operatorname{im} f^*)^\perp \implies \forall v \in V : u \perp f^*(v) \implies \forall v \in V : \langle u, f^*(v)\rangle_U = 0.$

   $$\begin{aligned}\forall v \in V : \langle f(u), v\rangle_V &= 0 \\ \implies f(u) &\in V^\perp = \{0\} \\ \implies u &\in \operatorname{kern} f\end{aligned}$$

4. Apply (3) to f^*.

$$\begin{aligned}\operatorname{kern} f^* &= (\operatorname{im} f^{**})^\perp = (\operatorname{im} f)^\perp \\ \implies (\operatorname{kern} f^*)^\perp &= (\operatorname{im} f)^{\perp\perp} \underbrace{=}_{\dim < \infty} \operatorname{im} f\end{aligned}$$

□

Remark (Addition to Theorem 8.43)**.** *So, if subspace $U \subseteq V$. Then $U^{\perp\perp} = U$.*

Proof: It holds that $U \dot{+} U^\perp = V$ and $U^\perp \dot{+} U^{\perp\perp} = V$. $U \subseteq U^{\perp\perp}$ and $\dim U = \dim U^{\perp\perp} \implies U = U^{\perp\perp}$.

8.11 Unitary transformations and self-adjoint matrices

Definition 8.67. *Let V be a vector space with scalar product.*

1. $f : V \to V$ *is called* self-adjoint, *if* $f = f^*$. *Hence* $\forall x, y \in V : \langle f(x), y \rangle = \langle x, f(y) \rangle \iff \Phi_B^B(f) = \Phi_B^B(f)^*$ *if B is orthonormal basis of V.*
2. $f \in \operatorname{Hom}(V, W)$ *is called* unitary transformation *or* linear isometry *if*

$$\forall x, y \in V : \langle f(x), f(y) \rangle = \langle x, y \rangle$$

esp. $\left\| f(x) \right\| = \|x\|$, *hence lengths (and also angles) are preserved.*
mostly *it is additionally required that f is invertible.*

Remark 8.68. 1. *Unitary transformations are injective.*

2. *If* $\dim V = \dim W < \infty$ *and* $f : V \to W$ *is linear and unitary, then f is invertible and* $f^{-1} = f^*$.
3. *If* $\dim V = \infty$, $f : V \to V$ *is isometry, it does not imply that f is invertible.*

Proof. 1. Recognize that $\langle f(x), f(x) \rangle = \langle x, x \rangle$ implies $\left\| f(x) \right\|^2 = \|x\|^2$. Then $f(v) = 0 \implies \left\| f(v) \right\| = \|v\| = 0 \implies v = 0$

$$\operatorname{kern} f = \{0\}$$

2. f unitary $\overset{(1.)}{\implies} f$ injective $\implies f$ surjective.

$$\begin{aligned} \forall x, y \in V : \langle x, y \rangle &= \langle f(x), f(y) \rangle \\ &= \langle x, f^* \circ f(y) \rangle \end{aligned}$$

hence for fixed y, it holds that

$$\begin{aligned} &\forall x \in V : \langle x, y \rangle = \langle x, f^* \circ f(y) \rangle \\ &\qquad \implies y = f^* \circ f(y) \text{ for all } y \implies f^* \circ f = \mathrm{id} \end{aligned}$$

3. $V = l^2 = \left\{ (x_n)_n \,\middle|\, \sum |x_n|^2 < \infty \right\}$

$$S : l^2 \to l^2$$

$$\begin{aligned} (x_1, x_2, \dots) &= (0, x_1, x_2, \dots) \\ \left\| S(x) \right\| &= \|x\| \end{aligned}$$

$$
\begin{aligned}
\langle S(x), S(y)\rangle &= \langle (0, x_1, x_2, \ldots), ()\rangle \\
&= 0 + \sum_{i=1}^{\infty} x_i \overline{y_i} \\
&= \langle x, y\rangle \\
\langle x, S^* y\rangle &= \langle Sx, y\rangle \\
&= \langle (0, x_1, x_2, \ldots), (y_1, y_2, \ldots)\rangle \\
&= 0 \cdot \overline{y_1} + x_1 \cdot \overline{y_2} + x_2 \cdot \overline{y_3} + \ldots \\
&= \langle (x_1, x_2, \ldots), (y_1, y_2, \ldots)\rangle \\
S^*(y_1, y_2, \ldots) &= (y_2, y_3, \ldots) \\
\left\langle S_x, S_y \right\rangle &= \langle x, S^* S y\rangle \, \forall x, y \\
&\implies S^* \circ S = \text{id} \\
\text{but } S \circ S^*(x_1, x_2, \ldots) &= S(x_2, x_3, \ldots) \\
&= (0, x_2, x_3, \ldots) \\
&\implies S \circ S^* \neq \text{id} \\
& S \text{ is not invertible}
\end{aligned}
$$

This shifting of indices works in a finite number of dimensions, but does not work in infinity (in this case, you miss one dimension).

□

8.12 Unitary matrices and orthogonal matrices

Definition 8.69. 1. *A matrix U is called* unitary *if $U^* U = I$*

2. *A matrix $U \in \mathbb{R}^{n \times n}$ is called* orthogonal *if $U^T U = I$*

Theorem 8.70. *For a matrix $T \in \mathbb{C}^{n \times n}$ it holds equivalently:*

1. *T is unitary ($T^* \cdot T = I$)*
2. *$\forall x \in \mathbb{C}^n : ||Tx|| = ||x||$ (isometry)*
3. *$\forall x, y \in \mathbb{C}^n : \Re\langle Tx, Ty\rangle = \Re\langle x, y\rangle$*
4. *$\forall x, y \in \mathbb{C}^n : \langle Tx, Ty\rangle = \langle x, y\rangle$*
5. *The columns of T define an orthonormal basis of $\mathbb{C}^n$*

Proof. **1. → 2.**

$$||Tx||^2 = \langle Tx, Tx\rangle = \langle x, T^* Tx\rangle = \langle x, Ix\rangle = ||x||^2$$

2. → 3.

$$
\begin{aligned}
\left\|T(x+y)\right\|^2 &= \left\|x+y\right\|^2 \\
\left\|T(x-y)\right\|^2 &= \left\|x-y\right\|^2 \\
\left\|Tx+Ty\right\|^2 &= \|Tx\|^2 + 2\Re\langle Tx, Ty\rangle + \left\|Ty\right\|^2 \\
\left\|Tx-Ty\right\|^2 &= \|Tx\|^2 - 2\Re\langle Tx, Ty\rangle + \left\|Ty\right\|^2 \\
\hline
\left\|Tx+Ty\right\|^2 - \left\|Tx-Ty\right\|^2 &= 4\Re\langle Tx, Ty\rangle \\
\text{analogously, } \left\|x+y\right\|^2 - \left\|x-y\right\|^2 &= 4\Re\langle x, y\rangle \\
\hline
\implies \Re\langle Tx, Ty\rangle &= \Re\langle x, y\rangle
\end{aligned}
$$

3. → 4.

$$\Re\langle Tx, Ty\rangle = \Re\langle x, y\rangle \qquad \forall x, y \in \mathbb{C}^n$$

also holds for $i \cdot y$ instead of y

$$\Re\langle Tx, iTy\rangle = \Re\langle x, iy\rangle \qquad \forall x, y \in \mathbb{C}^n$$

$$\Re\left(-i\left\langle Tx, Ty\right\rangle\right) = \Re(-i\left\langle x, y\right\rangle)$$

$$
\begin{aligned}
\Re(-i(a+ib)) &= \Re(-ia+b) = b \\
\Re(-i \cdot z) &= \Im(z)
\end{aligned}
$$

$$\Im\langle Tx, Ty\rangle = \Im\langle x, y\rangle \qquad \forall x, y \in \mathbb{C}^n$$

$\Re$ and $\Im$ are equivalent.

$$\implies \langle Tx, Ty\rangle = \langle x, y\rangle \qquad \forall x, y$$

(this is a common proof pattern, that you only show it for $\Re$ and $\Im$ follows immediately)

4. → 5. $e_1, \dots, e_n$ define some orthonormal basis.

$$\implies (Te_1, \dots, Te_n) \text{ is orthonormal basis}$$

$$u_i = T_{e_i} = \text{ i-th column of } T$$

$$\left\langle u_i, u_j\right\rangle = \left\langle Te_i, Te_j\right\rangle = \left\langle e_i, e_j\right\rangle = \delta_{ij}$$

5. → 4. $(T^*T)_{ij}$ is the i-th column vector of T^* times the j-th column vector of T.

$$u_j^* \cdot u_j = \left\langle u_j, u_i\right\rangle = \delta_{ji}$$

$$\implies T^*T = \begin{bmatrix} 1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1 \end{bmatrix} = I$$

□

What do isometries of $\mathbb{R}^n$ or $\mathbb{C}^n$ look like?

Definition 8.71. *An isometry between two metric spaces (M_1, d_1) and (M_2, d_2). Metric d:*

$$\begin{aligned} d(x,y) &\geq 0 \\ d(x,y) = 0 &\iff x = y \\ d(x,y) &\leq d(x,z) + d(z,y) \end{aligned}$$

is a map $f : M_1 \to M_2$ such that

$$d_2(f(x), f(y)) = d_1(x,y)$$

Every normed space has metric $d(x,y) = \|x - y\|$. An isometry between two spaces is a (not necessarily linear) map $f : V \to W$ such that $\|f(x) - f(y)\| = \|x - y\|$.

Example (Translation)**.**

$$x_0 \in V \qquad T_{x_0} : V \to V \qquad x \mapsto x + x_0$$

Translation T_{x_0} is an isometry because $\|T_{x_0}(x) - T_{x_0}(y)\| = \|x + x_0 - (y + x_0)\| = \|x - y\|$. But translation is not unitary because of non-linearity: $T_{x_0}(0) = 0 + x_0 \neq 0$.

Other examples in $\mathbb{R}^n$:

1. rotation
2. reflection
3. unitary/orthogonal map

Example (Rotation in $\mathbb{R}^2$)**.**

$$U(e_1) = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} \qquad U(e_2) = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}$$

Compare with Figure 14.

$$U_\alpha = \begin{bmatrix} \cos\alpha & \dots & -\sin\alpha \\ & \ddots & \\ \sin\alpha & \dots & \cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \cdot \cos\alpha + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \sin\alpha$$

Tangent a:

$$\begin{gathered} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} \\ \vec{x}(t) \perp \vec{x}(t) \\ \dot{\vec{x}}(t) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \vec{x}(t) \\ \vec{x}(t) = e^{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} t} \cdot \vec{x_0} \end{gathered}$$

Compare with Figure 15.

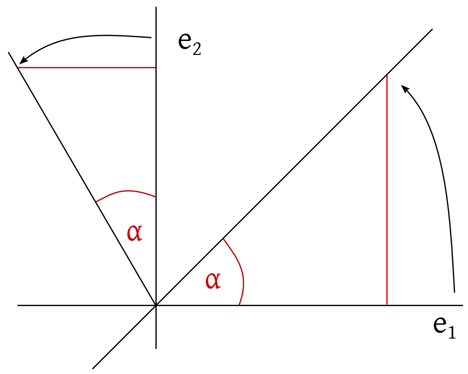


Figure 14: Rotation in $\mathbb{R}^2$

Example (Rotation considered as motion)**.**

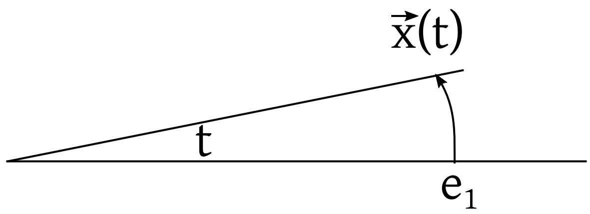


Figure 15: Rotation in $\mathbb{R}^2$ considered as motion. Commonly done by physicists.

$$x'(t) = a \cdot x(t) \implies x(t) = c \cdot e^{at}$$
$$\frac{dx}{dt} = ax$$
$$dx = ax \cdot dt$$
$$\int \frac{dx}{x} = \int a \cdot dt$$
$$\log x = at + C$$
$$x = C_1 \cdot e^{at}$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
$$e^{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} t} = \sum_{n=0}^{\infty} \frac{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^n}{n!} t^n$$
$$e^{it} = \cos t + i \cdot \sin t$$

insert $\sum_{n=0}^{\infty} \frac{(it)^n}{n!}$ *and split* $\Re$ *and* $\Im$.

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} -1 & \\ & -1 \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^4 = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$$

$$i^2 = -1 \qquad i^3 = -i \qquad i^4 = 1$$

$$e^{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} t} = \cos(t) \cdot \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} + \sin(t) \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$
$$U_{\alpha+\beta} = U_\alpha \cdot U_\beta$$
$$\begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\alpha \end{bmatrix}$$
$$= \begin{bmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\cos\beta - \sin\alpha\sin\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & \sin\alpha\cos\beta + \cos\alpha\sin\beta \end{bmatrix}$$

Example (Reflection in $\mathbb{R}^2$).

$$S(e_1) = \begin{bmatrix} \cos(2\varphi) \\ \sin(2\varphi) \end{bmatrix}$$
$$S(e_2) = \begin{bmatrix} \cos(2\varphi - \frac{\pi}{2}) \\ \sin(2\varphi - \frac{\pi}{2}) \end{bmatrix} = \begin{bmatrix} \sin(2\varphi) \\ -\cos(2\varphi) \end{bmatrix}$$
$$\frac{\pi}{2} - 2\psi = \frac{\pi}{2} - 2\left(\frac{\pi}{2} - \varphi\right) = 2\varphi - \frac{\pi}{2}$$
$$S = \begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix}$$

↓ *This lecture took place on 2018/05/07.*

Linear isometries:

Theorem 8.72.

$$\mathcal{O}(n) = \left\{ U \in \mathbb{R}^{n\times n} \,\middle|\, U^T U = I \right\} \qquad \textit{orthogonal group}$$
$$\mathcal{U}(n) = \left\{ U \in \mathbb{C}^{n\times n} \,\middle|\, U^* U = I \right\} \qquad \textit{unitary group}$$
$$\mathcal{SO}(n) = \{ U \in \mathbb{O} \mid \det(U) = 1 \} \subseteq \mathcal{O}(n) \qquad \textit{subgroup, special orthogonal group}$$
$$\mathcal{SU}(n) = \{ U \in \mathbb{U} \mid \det(U) = 1 \} \subseteq \mathcal{U}(n) \qquad \textit{subgroup, special unitary group}$$
$$\mathcal{GL}(n, \mathbb{K}) = \left\{ A \in \mathbb{K}^{n\times n} \,\middle|\, \textit{invertible} \right\} \qquad \textit{general linear group}$$
$$\mathcal{SL}(n, \mathbb{K}) = \left\{ A \in GL(n) \,\middle|\, \det(A) = 1 \right\} \qquad \textit{special linear group}$$

Then, e.g. $\mathcal{O}(2)$ is the group of rotations and reflections.

Remark. *For $U \in \mathcal{U}(n)$ it holds that $\left|\det(U)\right| = 1$. Why?*

We know: $U^*U = I \implies \det(U^*U) = I = \det(U^*)\det(U) = \det(\overline{U}^T)\det(U) = \overline{\det(U^T)}\det(U) = \overline{\det(U)}\det(U) = \left|\det(U)\right|^2 = 1.$

Example (Rotation)**.**

$$U = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

$$\det U_\varphi = \cos^2(\varphi) + \sin^2(\varphi) = 1 \qquad \implies U_\varphi \in \mathcal{SO}(2)$$

$$S_\varphi = \begin{bmatrix} \cos(2\varphi) & \sin(2\varphi) \\ \sin(2\varphi) & -\cos(2\varphi) \end{bmatrix}$$

$$\det(S_\varphi) = -\cos^2(2\varphi) - \sin^2(2\varphi) = -1$$

General orthogonal matrix in $\mathcal{O}(2)$.

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \textit{ with } \overline{U}U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Resulting constraints:

$$a^2 + c^2 = 1 \tag{1}$$

$$b^2 + d^2 = 1 \tag{2}$$

$$ab + cd = 0 \tag{3}$$

$$a = \cos\varphi \qquad c = \sin\varphi \qquad b = \cos\psi \qquad d = \sin\psi$$

$$\cos\varphi \cdot \cos\psi + \sin\varphi \cdot \sin\psi = 0$$

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} = \begin{bmatrix} \cos(\alpha+\beta) \end{bmatrix}$$

$$\cos\alpha\cos\beta - \sin\alpha\sin\beta = \cos(\alpha+\beta)$$

$$\cos\varphi \cdot \cos\psi = \cos(\varphi - \psi)$$

$$\cos\alpha = 0 \textit{ for } \alpha = \frac{\pi}{2} + k\cdot\pi = (k+\frac{1}{2})\pi \qquad (k \in \mathbb{Z})$$

$$\implies \varphi - \psi = (k+\frac{1}{2})\pi$$

$$\varphi = \psi + (k+\frac{1}{2})\pi$$

$$\cos\varphi = \cos(\psi + (k+\frac{1}{2})\pi) = \cos\psi\cos(k+\frac{1}{2})\pi - \sin\psi\underbrace{\sin(k+\frac{1}{2})\pi}_{\varepsilon\in\{\pm 1\}}$$

$$= -\varepsilon\cdot\sin\psi \implies \sin\psi = -\varepsilon\cos\varphi$$

$$\sin\alpha\cos\beta + \cos\alpha\sin\beta = \sin(\alpha+\beta)$$

$$\sin(\varphi) = \sin(\psi + (k+\frac{1}{2})\pi) = \underbrace{\sin\psi\cos\left(k+\frac{1}{2}\right)\pi}_{=\varepsilon\cdot\cos(\psi)} + \underbrace{\cos\psi\sin\left(k+\frac{1}{2}\right)\pi}_{=0}$$

$$\cos\psi = \varepsilon\sin\varphi$$

$$U = \begin{bmatrix}\cos\varphi & \varepsilon\cdot\sin(\psi)\\ \sin\varphi & -\varepsilon\cos\varphi\end{bmatrix} = \underbrace{\begin{bmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{bmatrix}}_{\substack{\textit{rotation}\\ \det=1}}\cdot \underbrace{\begin{bmatrix}1 & \\ & -\varepsilon\end{bmatrix}}_{\substack{\varepsilon=1:\\ \textit{reflection on x-axis}\\ \varepsilon=-1:\mathrm{id}}}$$

$$U_\varphi = \cos\varphi\begin{bmatrix}1 & \\ & 1\end{bmatrix} + \sin\varphi\begin{bmatrix} & -1\\ 1 & \end{bmatrix}$$

Hence, every orthogonal matrix is either a rotation ($\det = 1$) *or a reflection* ($\det = -1$)*.*

$$\mathcal{SO}(2) : \left\{U_\varphi = \cos\varphi + i\cdot\sin\varphi \qquad 1 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, i = \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\right\}$$

$$\mathcal{SU}(2) : \left\{a_0 + ia_1 + ja_2 + ka_3 \,\middle|\, \sum a_i^2 = 1\right\}$$

8.13 Quaternions

Person. *William Rowan Hamilton (1805–1865).*

Remark (Quaternions)**.** *Hamilton defined the complex numbers in the modern sense in 1833.*

$$C = \left\{(a,b) \,\middle|\, a,b\in\mathbb{R}\right\}$$

$$(a,b)\cdot(c,d) = (ac-bd, ad+bc)$$

He tried to invent them over 10 years for the third dimension. He failed. On 1843/10/16, he invented the quaternions (a memorial next to Brougham Bridge in Dublin reminds of the moment of discovery). It works on four dimensions, but it is non-commutative. It is a screw field (dt. Schiefkörper).

$$ij = k \qquad jk = i \qquad ki = j \qquad ji = -k \qquad kj = -i \qquad ik = -j$$

anti-commutative.

$$i^2 = j^2 = k^2 = -1$$

$$(a_0 + a_1 i + a_2 j + a_3 k)(b_0 + b_1 i + b_2 j + b_3 k) \qquad \textit{linear}$$

$$(a_0 + \vec{a})(b_0 + \vec{b}) = a_0 b_0 + a_0\vec{b} + b_1\vec{a} + \vec{a}\times\vec{b}$$

$$\mathcal{SO}(2) \approx \left\{\cos\varphi + i\cdot\sin\varphi \,\middle|\, \varphi\in[0,2\pi]\right\} = \{z\in\mathbb{C} \mid |z| = 1\} = \mathcal{T}\text{ Torus}$$

$$\mathcal{SU}(2) = \left\{a_0 + ia_1 + ja_2 + ka_3 \,\middle|\, \sum a_i^2 = 1\right\}$$

$$\mathcal{SO}(2) \approx \left\{\cos\varphi + i\sin\varphi \,\middle|\, q\in[0,2\pi]\right\}$$

9 Polynomials and algebras

Definition 9.1 (Algebra)**.** *Let $\mathbb{K}$ be a field, a $\mathbb{K}$ algebra is a vector space $\mathcal{A}$ over $\mathbb{K}$ with a multiplication operator $* : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ with $(a, b) \to a * b$ such that*

1. $a * (b + c) = a * b + a * c$ *(distributive law, $a, b, c \in \mathcal{A}$)*
2. $(a + b) * c = a * c + b * c$
3. $\lambda \cdot (a * b) = (\lambda \cdot a) * b = a * (\lambda \cdot b)$ *($a, b \in \mathcal{A}, \lambda \in \mathbb{K}$, associativity)*

Remark 9.2. Associativity *is not generally required.*

$$a * (b * c) = (a * b) * c$$

If satisfied, it is called associative algebra.

Commutativity *is not generally required.*

$$a * b = b * a$$

If satisfied, it is called commutative algebra.

Example 9.3.

1. $(\mathbb{K}, +, * = \cdot)$ *is a one-dimensional $\mathbb{K}$ algebra.*
2. $(\mathbb{K}^{n\times n}, +, * =$ *matrix multiplication) is an associative non-commutative algebra where* $\mathbb{K}^{n\times n} \simeq \mathrm{Hom}(V, V)$ *and* $f * g = f \circ g$.
3. $\mathbb{K}^{\times} = \{f : X \to \mathbb{K}\}$. *Let X be an arbitrary set.*

$$(\lambda f + \mu g)(x) = \lambda \cdot f(x) + \mu \cdot g(x)$$
$$(f * g)(x) = f(x) \cdot g(x)$$

   $(\mathbb{K}^{\times}, +, *)$ *is an associative, commutative algebra.*

4. $\mathbb{R}^3$ *with $a \times b$ is an algebra.*

$$a \times b = -b \times a$$

   is non-commutative and also non-associative:

$$a \times (b \times c) \neq (a \times b) \times c$$

   Jacobian identity:

$$a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0$$

5. $\mathcal{A} = \mathbb{K}^{n\times n}$

$$A * B = [A, B] = A \cdot B - B \cdot A \qquad \text{“commutator”}$$

   is an algebra with Jacobian identity. Lie algebra:

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$$
$$[A, B] = -[B, A]$$

   The so-called Lie groups (like $\mathcal{O}(n), \mathcal{U}(n), \mathcal{SO}(n), \mathcal{SU}(n)$).

6. $\mathcal{A} = \mathbb{K}^{n \times n}$
$$A * B = A \cdot B + B \cdot A$$
is associative. It is an Jordan algebra.

Person. *Pascual Jordan (1902–1980)*[8].

Person. *Oskar Perron (1880/05/07–1975)*

Definition 9.4.
$$\mathbb{K}^\infty = \left\{ (a_0, a_1, a_2, \dots) \,\middle|\, a_i \in \mathbb{K} \right\}$$
$$P_{\mathbb{K}} = \left\{ (a_0, a_1, \dots, a_n, 0, \dots) \,\middle|\, n \in \mathbb{N}, a_i \in \mathbb{K} \right\}$$
We define $*$ as the Cauchy product:
$$(a_n)_{n \geq 0} * (b_n)_{n \geq 0} = (c_n)_{n \geq 0} \qquad c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

Lemma 9.5. 1. $(P_{\mathbb{K}}, *)$ *is a commutative, associative algebra with one-element* $(1, 0, \dots)$. *The basis is given with* $1, x, x^2, \dots$. *The algebra is called* polynomial algebra
$$\mathbb{K}[x] = \left\{ \sum_{k=0}^{n} a_k x^k \,\middle|\, a_k \in \mathbb{K}, n \in \mathbb{N} \right\}$$

2. $(\mathbb{K}^\infty, *)$ *is a commutative algebra with one-element* $(1, 0, \dots)$ *and is called* algebra of formal power series[9]
$$\mathbb{K}[[x]] = \left\{ \sum_{k=0}^{\infty} a_k x^k \,\middle|\, a_k \in \mathbb{K} \right\}$$

Proof. We show algebra properties:

1. Show that $\forall a, b \in P_{\mathbb{K}} : a * b \in P_{\mathbb{K}}$, hence only finitely many c_n are $\neq 0$.
   Remark: $a_k = 0 \forall k > m$ and $b_k = 0 \forall k > n$.

   Claim.
$$c_k = 0 \qquad \forall k > m + n$$

$$\begin{aligned} c_k &= \sum_{l=0}^{k} a_l b_{k-l} \\ &= \sum_{l=0}^{m-1} a_l b_{k-l} \qquad \text{equality if } l > m \implies a_l = 0 \\ &= 0 \end{aligned}$$
$$k > m + n, l < m \implies -l > -m \implies k - l \underbrace{>}_{\implies b_{k-l} = 0} m + n - m = n$$

[8] Different Jordan than in Gauss-Jordan and different than C. Jordan (19th century) about to come
[9] We don't need to consider convergence. This is purely formal object.

2. About the Cauchy product:

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} = \sum_{k'=0}^{n} a_{n-k'} b_{k'} = (b * a)_n \qquad (k' = n - k)$$

3. Law of distributivity:

$$\begin{aligned} [(a+b) * c]_n &= \sum_{k=0}^{n} (a+b)_k \cdot c_{n-k} \\ &= \sum_{l=0}^{n} (a_k c_{n-k}) + (b_k c_{n-k}) \\ &= (a * c)_n + (b * c)_n \end{aligned}$$

□

Definition 9.6. *Let $x^0 = (1, 0, \dots)$ and $x^k = (0, \dots, 1, 0, \dots)$ create a basis. The elements of $p(x) = \mathbb{K}[x]$ are called polynomials in the formal variable x*

$$\deg p(x) = \max\{k \mid a_k \neq 0\} \text{ *is called* degree of the polynomial}$$

$$\deg(0) := -\infty$$

Lemma 9.7.

1. $\deg(p(x) \cdot q(x)) = \deg(p(x)) + \deg(q(x))$
2. $\mathbb{K}[x]$ *is zero-divisor-free, hence* $p(x) \cdot q(x) = 0 \implies p(x) = 0 \vee q(x) = 0$

9.1 The difference of polynomials and polynomial functions

Definition 9.8. *Every polynomial $p(x) \in \mathbb{K}[x]$ induces a polynomial function $p : \mathbb{K} \to \mathbb{K}$ with $\alpha \mapsto p(\alpha)$ with $p \in \mathbb{K}^{\mathbb{K}}$.*

$$\implies (\lambda p + \mu q)(\alpha) = \lambda \cdot p(\alpha) + \mu \cdot q(\alpha)$$

$$(p \cdot q)(\alpha) = p(\alpha) \cdot q(\alpha)$$

The map $\mathbb{K}[x] \to \mathbb{K}^{\mathbb{K}}$ with $p(x) \mapsto$ polynomial function p is linear and multiplicative (called algebra homomorphism*).*

Remark 9.9. *A polynomial and a polynomial function are not the same. If $|\mathbb{K}| < \infty$, a difference occurs. For example, consider $\mathbb{Z}_5$:*

$$\left|\mathbb{Z}_5^{\mathbb{Z}_5}\right| = 5^5$$

$$\left|\mathbb{K}[x]\right| = \infty$$

where $\mathbb{Z}_5^{\mathbb{Z}_5}$ is a set of polynomial functions and $\mathbb{K}[x]$ is a set of polynomials. For example, $\prod_{\alpha \in \mathbb{K}}(x - \alpha)$ corresponds to the polynomial function 0. Consider $\mathbb{K} = \mathbb{Z}_3$. If $\alpha \in \mathbb{Z}_3$ and $x \in \mathbb{Z}_3$, we get $(x-0)(x-1)(x-2)$ and choosing any $x \in \mathbb{Z}_3$ makes at least one factor *zero. Thus, the polynomial function 0 is given. Hence the map $\mathbb{K}[x] \to \mathbb{K}^{\mathbb{K}}$ is surjective but not injective.*

On finite fields, every function is a polynomial function.

$$\eta_i = f(\xi_i) \qquad \{\xi_1, \dots, \xi_n\} = \mathbb{K}$$

From the practicals, it will follow that there exists a polynomial of degree n such that $p(\xi_i) = \eta_i$.

Definition 9.10. *An algebra homomorphism is a linear map between ψ and two $\mathbb{K}$-algebras $\mathcal{A}$ and $\mathcal{B}$ such that $\forall a, b \in \mathcal{A} : \psi(a *_A b) = \psi(a) *_B \psi(b)$.*

Example 9.11.

1. *$\mathbb{K}[x] \to \mathbb{K}^{\mathbb{K}}$ with $p(x) \mapsto$ polynomial function*
2. *Let $\alpha \in \mathbb{K}$ be fixed. $\psi_\alpha : \mathbb{K}[x] \to \mathbb{K}$ with $p(x) \mapsto p(\alpha)$ is an algebra homomorphism of $\mathbb{K}[x] \to \mathbb{K}$.*

$$\psi_\alpha(\lambda p + \mu q) = (\lambda p + \mu q)(\alpha) = \lambda p(\alpha) + \mu q(\alpha) = \lambda \psi_\alpha(p) + \mu \psi_\alpha(q)$$

3. *Consider $\iota : \mathbb{K} \to \mathbb{K}[x]$ with $\iota : \alpha \mapsto \alpha \cdot x^0$.*

$$(\alpha \cdot x^0) \cdot (\beta \cdot x^0) = (\alpha \cdot \beta) \cdot x^0$$

Theorem 9.12 (Insertion theorem, dt. Einsetzungssatz)**.** *Let $\mathcal{A}$ be an associative algebra with one-element $\mathbf{1}_A$ and $\iota : \mathbb{K} \to \mathcal{A}$ with $\alpha \mapsto \alpha \cdot \mathbf{1}_A$ is the insertion of $\mathbb{K}$.*

Then for every $a \in \mathcal{A}$ the map

$$\psi_a : \mathbb{K}[x] \to \mathcal{A}$$
$$\sum_{k=0}^{n} c_k x^k \mapsto \sum_{k=0}^{n} c_k a^k$$

is the unique algebra homomorphism of $\mathbb{K}[x] \to \mathcal{A}$ with the property $\psi_a(x) = a$. We say, $\mathbb{K}[x]$ is a free, associative algebra over $\mathbb{K}$. *Every algebra homomorphism $\mathbb{K}[x] \to \mathcal{A}$ has this structure.*

↓ *This lecture took place on 2018/05/09.*

We consider algebras as vector spaces with associative multiplication. For example, matrices and polynomials. An algebra homomorphism is linear and multiplicative.

$$\Phi(a + b) = \Phi(a) * \Phi(b)$$

A is an associative algebra with $\mathbf{1}_{\mathcal{A}}$.

$$l : \begin{array}{c} \mathbb{K} \to \mathcal{A} \\ \alpha \to \alpha \cdot \mathbf{1}_{\mathcal{A}} \end{array}$$

$a \in \mathcal{A} \implies \mathcal{L}(a^0, a^1, a^2, a^3, \dots) \subseteq \mathcal{A}$ subalgebra.

1. Show linear and multiplicative property.

$$\exists!\Phi_a : \mathbb{K}[a] \to \mathcal{A} \text{ algebra homomorphism}$$

such that $\Phi_a(x) = a$, namely $\Phi_a\left(\sum_{k=0}^n c_k x^k\right) = \sum_{k=0}^n c_k a^k$.

2. Every homomorphism $\Psi : \mathbb{K}[x] \to \mathcal{A}$ has this structure.

Proof. Let $a := \Psi(x) \implies \Psi(x^n) = \Psi(x)^n = a^n$ by homomorphism.

$$\Psi \text{ linear} \implies \Psi\left(\sum_{k=0}^n c_k x^k\right) = \sum_{k=0}^n c_k \Psi(x^k) = \sum_{k=0}^n c_k a^k$$

$x^0, x^1, \dots$ give a basis of $\mathbb{K}[x]$. Hence $\Psi = \Phi_a$ with $a = \Psi(x)$. On the opposite (1.): Obviously Φ_a is linear. Multiplicative: Show that

$$\underbrace{\Psi_a(p(x)\cdot q(x))}_{=p(a)\cdot q(a)} \stackrel{!}{=} \underbrace{\Phi_a(p(x)) \cdot_{\mathcal{A}} \Phi_a(q(x))}_{=p(a)\cdot q(a)}$$

□

Example 9.13. *1.* $\mathcal{A} = \mathbb{K}$.

$$\Psi_\alpha : \substack{\mathbb{K}[x]\to\mathbb{K} \\ p(x)\mapsto p(\alpha)}$$

2. $\mathcal{A} = \mathbb{K}^{n\times n} \approx \mathrm{Hom}(V, V)$

$$A^0 = I \qquad A^n = A\cdot A^{n-1}$$

$$l : \substack{\mathbb{K}\to\mathbb{K}^{n\times n} \\ \alpha\mapsto\alpha\cdot I}$$

$$\Psi_\alpha : \substack{\mathbb{K}[x]\to\mathbb{K}^{n\times n} \\ p(x)\mapsto p(A) \\ \sum_{k=0}^n c_k x^k \mapsto \sum_{k=0}^n c_k\cdot A^k}$$

Remark 9.14. *Let $\mathbb{K}[x]$ be a free, associative algebra over $\mathbb{K}$ with a generator. Hence, for all associative algebras $\mathcal{A}$, given some element $a \in \mathcal{A}$. There exists exactly one homomorphism $\varphi : \mathbb{K}[x] \to \mathcal{A}$ such that $\varphi(x) = a$.*

Compare it with a free group with one generator. Is a group G generated by x such that $\forall$ groups H, if $h \in H$ given, there exists exactly one group homomorphism $\varphi : G \to H$ such that $\varphi(x) = h$. Namely, $G = (\mathbb{Z}, +)$ is generated by **1***. Given $h \in H \to \varphi_h : \mathbb{Z} \to H_k$ and $k \mapsto h$.*

Definition 9.15. *A root of a polynomial $p(x) \in \mathbb{K}[x]$ is a $\xi \in \mathbb{K}$ such that $p(\xi) = \Psi_\xi(p) = 0$, hence $p(x) \in \ker \Psi_\xi$.*

Remark. *1.* $p(x) = c_0$ *is a root* $\iff c_0 = 0$

2. $p(x) = c_0 + c_1 x$ *is the only root,* $\xi = -\frac{c_0}{c_1}$.

3. $p(x) = c_0 + c_1 x + c_2 x^2$ *has two roots over* $\mathbb{C}$

4. $p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$ *has three roots*

To find roots, formulas up to fourth degree exist. For degree ≥ 5, there is no equation.

Person. *Paolo Ruffini (1765–1822)*

Person. *Niels Henrik Abel (1802–1829)*

Person. *Gerolamo Cardano (1501–1576)*

Remark. *Cardano was a polymath.*

- *founder of probability theory*
- Liber de ludo aleae*: important book on probability*
- *Cardan joint (dt. Kardanische Welle)*
- *Gimbal (dt. Kardanische Aufhängung)*
- *used* $\sqrt{-1}$ *as a valid expression for the first time*
- *published a solution for roots of cubic polynomials (Ars Magna, 1545)*

Person. *Scipione del Ferro (1465–1526)*

- *used a solution for roots of cubic polynomials in competitions, kept it secret*
- *came up with the same solution like Tartaglia*
- *lost competitions on cubic polynomials to Antonio Fiore, because Ferro's solution was not generic enough*

Person. *Niccolò Fontana Tartaglia (1500–1557)*

1. *Cardano cajoled Tartaglia into revealing his solution to the cubic equations by promising not to publish them.*

Person. *Ludovico Ferrari (1522–1565)*

Person. *Viète, Francois (1540–1603)*

Theorem 9.16 (Method by Cardano/del Ferro)**.**

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 = 0$$

$$x \to x + a \qquad \text{such that } a_2 = 0$$

$$x^3 + px + q = 0$$

Cubus p.6 rebus aeq 20
$x^3 + 6x = 20$

$x = $ *res*, $x^2 = $ *census*, $x^3 = $ *cubus.*

Approach: $x = u + v$.

$$u^3 + 3u^2 v + 3uv^2 + v^3 + p(u+v) + q = 0$$

$$u^3 + v^3 + (3uv + p)(u + v) + q = 0$$

Requirement: u and v such that $3uv + p = 0$.

$$\begin{cases} u^3 + v^3 + q = 0 & \implies v^3 = -(q + u^3) \\ 3uv + p = 0 & \implies uv = -\frac{p}{3u} \end{cases}$$

$$u^3 \cdot v^3 = -\frac{p^3}{27}$$

$$-u^3(q + u^3) = -\frac{p^3}{27}$$

$$u^6 + qu^3 - \frac{p^3}{27} = 0$$

$$u^3 = ?$$

Equation for degree 2 by Viète Francois:

$$(y - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta$$

$$x^2 + px + q$$

$$p = -(\alpha + \beta)$$

$$q = \alpha \cdot \beta$$

$$\alpha = \frac{1}{2}\left[(\alpha + \beta) + \sqrt{(\alpha - \beta)^2}\right]$$

$$\beta = \frac{1}{2}\left[(\alpha + \beta) - \sqrt{(\alpha - \beta)^2}\right]$$

$$\frac{\alpha}{\beta} = \frac{1}{2}\left(\alpha + \beta \pm \sqrt{(\alpha - \beta)^2}\right) = \frac{1}{2}\left(\alpha + \beta \pm \sqrt{\underbrace{\alpha^2 + \beta^2 - 2\alpha\beta}_{(\alpha+\beta)^2 - 4\alpha\beta}}\right) = \frac{1}{2}\left(-p \pm \sqrt{p^2 - 4q}\right)$$

Hence,

$$u^3 = \frac{1}{2}\left(-q \mp \sqrt{q^2 + \frac{4p^3}{27}}\right)$$

$$u^3 = \frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

$$u = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

$$v^3 = -q - u^3 = -\frac{q}{2} \mp \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

$$x = u + v = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q^2}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

Theorem 9.17 (Division with remainder). $p(x), q(x) \in \mathbb{K}[x]$, $q(x) \neq 0$. *Then there exists exactly one polynomial* $s(x), r(x) \in \mathbb{K}[x]$,

$$p(x) = s(x) \cdot q(x) + r(x)$$

with $\deg r(x) < \deg q(x)$.

Proof. Induction over $\deg p(x)$.

Induction base

$$\deg p(x) < \deg q(x) \rightsquigarrow p(x) = 0 \cdot q(x) + p(x)$$

If $\deg p(x) \geq \deg q(x)$,

$$p(x) = \sum_{k=0}^{n} a_k x^k \qquad q(x) = \sum_{k=0}^{m} b_k x^k$$

$$a_n \neq 0 \qquad m \leq n \qquad b_m \neq 0$$

$$p_1(x) = p(x) - \frac{a_n}{b_m} \cdot q(x) \cdot x^{n-m}$$

cancels the largest term $a_n x^n$ in $p(x)$.

$$\begin{aligned}
&= \sum_{k=0}^{n} a_k x^k - \frac{a_n}{b_m} \sum_{k=0}^{m} b_k x^{k+n-m} \\
&= a_n x^n + \sum_{k=0}^{n-1} a_k x^k - \frac{a_n}{b_m} b_m \cdot x^{m+n-m} - \frac{a_n}{b_m} \sum_{k=0}^{m-1} b_k x^{k+n-m}
\end{aligned}$$

what remains is a polynomial of degree $\deg p_1(x) \leq n - 1$.

$$\implies p(x) = \frac{a_n}{b_m} x^{n-m} \cdot q(x) + p_1(x)$$

By induction hypothesis,

$$p_1(x) = s_1(x) \cdot q(x) + r_1(x)$$

Hence,

$$p(x) = \left(\frac{a_n}{b_m} x^{n-m} + s_1(x) \right) q(x) + r_1(x)$$

□

Example 9.18.

$$\begin{aligned}
p(x) &= 3x^5 - x^4 + 2x^3 + x^2 + 1 \\
q(x) &= x^2 - 3x + 1
\end{aligned}$$

$$
\begin{array}{rrrrrrl}
3x^5 & -x^4 & +2x^3 & +x^2 & & +1 & : x^2-3x+1 = 3x^2+8x^2+23x+62 \\
-3x^5 & +9x^4 & -3x^3 & & & & \\
\cline{1-3}
0 & 8x^4 & -x^3 & +x^2 & & +1 & \\
 & 8x^4 & -24x^3 & +8x^2 & & & \\
\cline{2-6}
 & 0 & 23x^3 & -7x^2 & & +1 & \\
 & & 23x^3 & -69x^2 & +23x & & \\
\cline{3-6}
 & & 0 & 62x^2 & -23x & +1 & \\
 & & & 62x^2 & -186x & +62 & \\
\cline{4-6}
 & & & 0 & 163x & -61 &
\end{array}
$$

Hence, $s(x) = 3x^3 + 8x^2 + 23x + 62$ *and* $r(x) = 163x - 61$.

Definition 9.19. $q(x)$ divides $p(x)$ $\iff$ *the remainder is zero* $\iff$ *there exists* $s(x)$ *such that* $p(x) = s(x) \cdot q(x)$.

Theorem 9.20. 1. *If* $p(x) = s(x) \cdot (x - \xi) + r$, *then* $q(x) = x - \xi$ *with* $p(\xi) = r$.

2. ξ *is root of* $p(x)$ $\iff$ $x - \xi$ *divides* $p(x)$

Theorem 9.21 (Ruffini-Horner's method)**.** *Given* $p(x) \in \mathbb{K}[x]$, $\lambda \in \mathbb{K}$.
Find $p(\lambda)$.

$$
\begin{aligned}
p(x) &= a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \\
&= a_n \lambda^n + a_{n-1} \lambda^{n-1} + \ldots + a_1 \lambda + a_0 \\
&= \left(a_n \lambda^{n-1} + \cdots + a_n\right) \lambda + a_0 \\
&= \left((a_n \lambda^{n-2} + \cdots + a_n)\lambda + a_1\right) \lambda + a_0 \\
&= \vdots
\end{aligned}
$$

Algorithm:

$$
\xi_n = a_n \text{ for } k = n-1, \ldots, 0 \qquad \xi_k = \lambda \xi_{k+1} + a_k
$$
$$
p(\lambda) = \xi_0
$$

If $p(x) = s(x)(x - \lambda) + r$, $p(\lambda) = r$.

Horner's method provides a more convenient method to evaluate a polynomial for given x than exponentiation by a high degree.

Example 9.22.

$$
3x^5 - x^4 + 2x^3 + x^2 + 1
$$
$$
p(5) = ? \qquad \xi_5 = 3
$$

$$
\left|
\begin{array}{rrrrrrrl}
3x^5 & -x^4 & +2x^3 & +x^2 & +1 & : (x-5) & = & 3x^4 + 14x^3 + 72x^2 + 361x + 1805 \\
3x^5 & -15x^4 & & & & & & \\
0 & 14x^4 & +2x^3 & +x^2 & +1 & & & \\
 & 14x^4 & -70x^3 & & & & & \\
 & 0 & +72x^3 & +x^2 & +1 & & & \\
 & & 72x^3 & -360x^2 & & & & \\
 & & 0 & +361x^2 & +1 & & & \\
 & & & 361x^2 & -1805x & & & \\
 & & & & 1805x & +1 & & \\
 & & & & 1805x & -5 \cdot 1805 & & \\
 & & & & & 5 \cdot 1805 + 1 & &
\end{array}
\right|
$$

$$\xi_5 = 3$$
$$\xi_4 = 5 \cdot \xi_5 + (-1) = 5 \cdot 3 - 1 = 14$$
$$\xi_3 = 5 \cdot 14 + 2 = 72$$
$$\xi_2 = 5 \cdot 72 + 1 = 361$$
$$\xi_1 = 5 \cdot 361 + 0 = 1805$$
$$\xi_0 = 5 \cdot 1805 + 1 = 9026$$

Definition 9.23. *A polynomial $p(x) \in \mathbb{K}[x]$ is called* reducible, *if $\exists p_1(x), p_2(x) : \deg p_1(x) < \deg p(x)$ and $p(x) = p_1(x) \cdot p_2(x)$ (is the factorization). $\deg p_2(x) < \deg p(x)$ (proper divisor). Otherwise the polynomial is called* irreducible.

Remark 9.24. *An irreducible polynomial of degree > 1 has no roots.*

Example 9.25.

- *Consider $x^2 = -2$ irreducible over $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{R}$. Its roots are $\pm\sqrt{2}$.*
  It is reducible over $\mathbb{R}$: $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.
  It is reducible over $\mathbb{Q}(\sqrt{2}) = \left\{a + b\sqrt{2} \,\middle|\, a, b \in \mathbb{Q}\right\}$.
- *Consider $x^2 + 1$ irreducible over $\mathbb{Q}$, $\mathbb{R}$ and reducible over $\mathbb{C}$. Its roots are $\pm i$.*
  $\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$. $x^2 + 1 = (x + i)(x - i)$.
- *Consider $\mathbb{K} = \mathbb{Z}_2$ and $p(x) = x^2 + x + 1$. This polynomial has no roots and is irreducible.*
- *$x^5 + x + 1$ has no roots, is reducible.*
  $$x^5 + x + 1 = (x^2 + x + 1)(x^3 + x^2 + 1)$$
  Is there some field $\mathbb{K} \supseteq \mathbb{Z}_2$ such that $x^3 + x^2 + 1$ has roots?
  <u>Yes</u>. Let α be a number such that $\alpha^3 + \alpha^2 + 1 = 0 \implies \alpha^3 = -\alpha^2 - 1 = \alpha^2 + 1$.
  $$\mathbb{K} = \mathbb{Z}_2(\alpha) = \left\{a + b\alpha + c\alpha^2 \,\middle|\, a, b, c \in \mathbb{Z}_2\right\}$$
  with $\alpha^3 = \alpha^2 + 1$ is a field.
  Let i be a number such that $i^2 + 1 = 0$, thus $i^2 = -1$
  $$\mathbb{C} = \mathbb{R}(i) = \{a + bi \mid a, b \in \mathbb{R}\}$$

Hence, irreducible is not equivalent to some root exists. The implication works only in one direction. There always exists some field such that roots exist.

Theorem 9.26 (Fundamental theorem of Algebra). *$\mathbb{C}$ is algebraically closed, hence every polynomial has a root over $\mathbb{C}$.*

Corollary 9.27. *Every polynomial over $\mathbb{C}$ ...*

1. *has a factorization $p(x) = (x - \xi_1)(x - \xi_2) \dots (x - \xi_n)$.*
2. *$p(x)$ is irreducible $\iff \deg p(x) \leq 1$.*

Remark. *No algebraic proof exists. It is more like a Fundamental Theorem of Calculus over complex numbers. The proof is given by the Lionville theorem (not done here).*

Theorem 9.28. *For arbitrary fields, it holds that every polynomial has exactly one factorization (except for its order) in irreducible factors.*

↓ *This lecture took place on 2018/05/14.*

9.2 The greatest common divisor of polynomials

The Euclidean algorithm determines the greatest common divisor.

Consider $n = q \cdot m + r$. For the Euclidean algorithm, it holds that $\gcd(n, m) = \gcd(m, r)$ The analogous solution holds for polynomials. Consider $p(x) = s(x) \cdot q(x) + r(x)$. Then the $\gcd(p(x), q(x))$ returns the polynomial of maximum degree that divides the polynomial with leading coefficient 1.

Corollary 9.29. *The Euclidean algorithm also works for polynomials.*

An application: Find all multiple roots (i.e. roots with multiplicity greater 1).

$$(x - \xi)^k \,\Big|\, p(x) \implies (x - \xi)^{k-1} \,\Big|\, p'(x)$$

$$\begin{aligned} p(x) &= s(x) \cdot (x - \xi)^k \\ p'(x) &= s'(x) \cdot (x - \xi)^k + s(x) \cdot k \cdot (x - \xi)^{k-1} = (s'(x)(x - \xi) + s(x) \cdot k)(x - \xi)^{k-1} \\ &\implies (x - \xi)^{k-1} | \gcd(p(x), p'(x)) \end{aligned}$$

10 Eigenvectors and eigenvalues

Given $f : V \to V$. Find a basis of V such that $\Phi_B^B(f)$ has the simplest possible representation. Hence,

$$\Phi_B^B(f) = A = \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix} \qquad A \cdot e_i = \lambda_i \cdot e_i$$

Find vector $v \in V$ such that $f(v) = \lambda \cdot v$.
0 can be an eigenvalue, but not an eigenvector.
Not every A has λ satisfying $\forall v : A \cdot v = \lambda \cdot v$.

Definition 10.1. $f \in \operatorname{Hom}(V, V) = \operatorname{End}(V)$. $\lambda \in \mathbb{K}$ *is called* eigenvalue *if* $\exists v \in V \setminus \{0\} : f(v) = \lambda \cdot v$. *Then v is called* eigenvector *of eigenvalue* λ. $\operatorname{spec}(f) = \{$*eigenvalues of* $f\}$ *is called* spectrum of f.

In 1925 in quantum mechanisms, it was discovered that the spectrum of light is given as a linear map (spectrum in the mathematical sense).

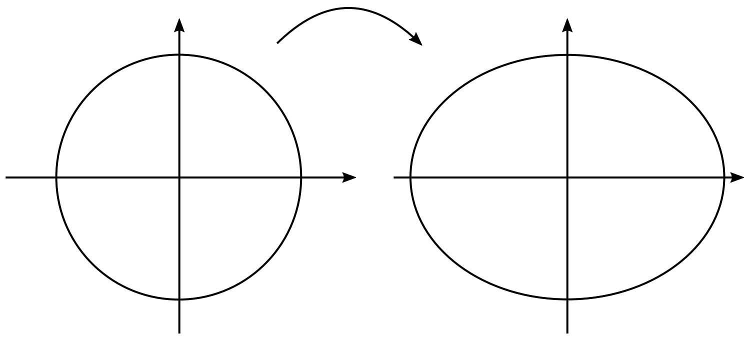


Figure 16: How map f might transform a circle

10.1 Eigenspace

Lemma 10.2. *For $\lambda \in \mathbb{K}$, $f \in \text{End}(V)$.*

$$\eta_\lambda = \{v \mid f(v) = \lambda \cdot v\}$$

is a subspace and is called eigenspace of f for eigenvalue λ.

Proof.

$$f(v) = \lambda \cdot v \iff f(v) - \lambda \cdot v = 0 \iff (f - \lambda \cdot \text{id})(v) = 0 \iff v \in \underbrace{\ker(f - \lambda \cdot \text{id})}_{\text{subspace}}$$

□

Example 10.3. 1. $f = \mu \cdot \text{id}$. $\text{spec}(f) = \{\mu\}$. $f(v) = \mu \cdot v \forall v \in V$. $\eta_\mu = V$.

2. *Let $b_1, \dots, b_n$ be a basis of V. Let $\lambda_1, \dots, \lambda_n \in \mathbb{K}$. Then there exists a unique, linear map f such that $f(b_i) = \lambda_i \cdot b_i$. Every b_i is an eigenvector to eigenvalue λ_i.*

$$\eta_\lambda = \mathcal{L}(\{b_i \mid \lambda_i = \lambda\})$$

Assume $f(v) = \lambda \cdot v$.

$$\begin{aligned} v &= \alpha_1 \cdot b_1 + \dots + \alpha_n b_n \\ f(v) &= \alpha_1 f(b_1) + \dots + \alpha_n f(b_n) \\ &= \alpha_1 \lambda_1 b_1 + \dots + \alpha_n \lambda_n b_n \\ &= \lambda(\alpha_1 b_1 + \dots + \alpha_n b_n) \\ \hline \implies 0 &= \alpha_1(\lambda_1 - \lambda) b_1 + \dots + \alpha_n(\lambda_n - \lambda) \cdot b_n \end{aligned}$$

$$\textit{linear indep.} \implies \forall i : \alpha_i(\lambda_i - \lambda) = 0$$

$$\textit{hence either } \alpha_i = 0 \textit{ or } \lambda_i = \lambda$$

$$\implies \text{spec}(f) = \{\lambda_1, \dots, \lambda_n\}$$

$$\Phi_B^B(f) = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

3. *Let* $V = C^\infty(\mathbb{R})$.

$$\frac{d}{dx}y(x) = \lambda \cdot y(x) \qquad \frac{dy}{dx} = \lambda \cdot y$$

$$\int \frac{dy}{y} = \int \lambda \cdot dx$$

$$\log(y) = \lambda \cdot x + C \qquad \text{Eigen function } \textit{(compare with Fourier analysis)}$$

$$y = C \cdot e^{\lambda x}$$

$$\frac{d}{dx}e^{\lambda x} = \lambda \cdot e^{\lambda x}$$

4. *Let* $V = C^\infty[0, a]$.

$$\frac{d^2}{dx^2}y(x) = \lambda \cdot y(x)$$

$$\frac{d^2}{dx^2}e^{\lambda x} = \frac{d}{dx}\lambda e^{\lambda x} = \lambda^2 e^{\lambda x}$$

$$\frac{d^2}{dx^2}e^{i\omega x} = -\omega^2 e^{i\omega x}$$

$$\frac{d^2}{dx^2}\sin \omega x = \frac{d}{dx}\omega \cdot \cos(\omega x) = -\omega^2 \cdot \sin(\omega x)$$

$$\frac{d^2}{dx^2}\cos \omega \alpha = \frac{d}{dx}(-\omega)\sin(\omega x) = -\omega^2 \cos(\omega x)$$

$$y(0) = y_0 \to y(x) = y_0 \cdot e^{\lambda x}$$

$$y(0) = y(a) = 0$$

$$y(x) = \sin(\omega x)$$

$$\omega a = k \cdot \pi \implies y(0) = y(a) = \pi$$

$$\omega = \frac{k \cdot \pi}{a}$$

Eigenvalues of $H = P^2 + Q$ *and* $PQ - QP = \frac{\hbar}{i}I$. *Heisenberg: Quantum mechanics is not commutative (impulses are matrices, not values).*

Definition 10.4. *Let* A *be a* $n \times n$ *matrix.* λ *is called right-sided eigenvalue if* $\exists x \in \mathbb{K}^n \setminus \{0\} : Ax = \lambda \cdot x$. λ *is called left-sided eigenvalue if* $\exists x \in \mathbb{K}^n \setminus \{0\} : x^T A = \lambda \cdot x^T$ *But this definition is satisfied* $\iff$ $A^T x = \lambda \cdot x$, *hence right-sided eigenvalue of* A^T. *Thus, these definitions collapse.*

Lemma 10.5. *Left-sided eigenvalue* $\iff$ *right-sided eigenvalue.*

Proof. Let λ be a right-sided eigenvalue.

$$\begin{aligned} Ax = \lambda x &\iff (A - \lambda \cdot I) \cdot x = 0 \\ &\iff \exists x \neq 0 : x \in \ker(A - \lambda I) \\ &\iff \ker(A - \lambda I) \neq \{0\} \\ &\iff \operatorname{rank}(A - \lambda I) < n \\ &\iff \operatorname{rank}(A^T - \lambda I) < n \\ &\iff \ker(A^T - \lambda I) \neq \{0\} \\ &\iff \exists x \neq 0 : A^T x = \lambda \cdot x \\ &\iff \lambda \text{ is a left-sided } \mathit{eigenvalue} \end{aligned}$$

□

Example 10.6. *For* $\dim = \infty$*, this must not hold.*

$$S : \begin{array}{c} \mathbb{K}^\infty \to \mathbb{K}^\infty \\ (x_1, x_2, \dots) \mapsto (x_2, x_3, \dots) \end{array}$$

$$S(1, 0, \dots) = (0, 0, \dots)$$

$$\implies (1, 0, \dots) \textit{ is eigenvector for eigenvalue } 0$$

hence, element of $\ker(S)$.

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \vdots & 0 & 1 & 0 \\ \vdots & \vdots & 0 & 1 \\ \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & & & \end{bmatrix}$$

$$S^T = \begin{bmatrix} 0 & & & \\ 1 & 0 & & 0 \\ & 1 & \ddots & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{bmatrix}$$

$S^T(x_1, x_2, \dots) \mapsto (0, x_1, x_2)$ *is injective.* $\ker(S^T) = \{0\}$*. Hence* 0 *is no eigenvalue.* 0 *is right-sided eigenvalue of S, but not left-sided eigenvalue.*

Remark. *The theory of eigenvalues in infinite-dimensional spaces is more complex than the finite-dimensional case.*

Definition 10.7. *For* $A \in \mathbb{K}^{n \times n}$.

$$\begin{aligned} \operatorname{spec}(A) &= \{\textit{right-sided eigenvalue of } A\} \\ &= \{\textit{left-sided eigenvalue of } A\} \end{aligned}$$

is called spectrum of A.

Remark (Proof exercise). $\dim V = n, f \in \mathrm{End}(V)$, *$B$ is basis of V.*

$$\implies \mathrm{spec}(f) = \mathrm{spec}(\Phi_B^B(f))$$

Remark. *Recall that any basis exchange can be done using some T:*

$$\Phi_B^B(A) = T^{-1}AT$$

Corollary 10.8. *The spectrum does not depend on the choice of the basis. Hence,*

$$\mathrm{spec}(T^{-1}AT) = \mathrm{spec}(A)$$

Direct proof. Let x be an eigenvector of A.

$$\begin{aligned} Ax = \lambda x &\iff A \cdot I \cdot x = \lambda x \iff A \cdot T \cdot T^{-1} \cdot x = \lambda x \\ &\iff T^{-1} \cdot A \cdot T \cdot T^{-1} \cdot x = T^{-1} \cdot \lambda x = \lambda \cdot T^{-1} \cdot x \\ &\iff y := T^{-1}x \text{ is eigenvector of } T^{-1}AT \\ &\iff T^{-1}ATy = \lambda y \\ &\iff \lambda \text{ is eigenvalue of } T^{-1}AT \end{aligned}$$

□

Remark. *λ is eigenvalue of A.*

$$\begin{aligned} &\iff \ker(\lambda \cdot I - A) \neq \{0\} \\ &\iff \mathrm{rank}(\lambda \cdot I - A) < n \\ &\iff \det(\lambda \cdot I - A) = 0 \end{aligned}$$

10.2 Characteristic polynomial

Theorem 10.9 (Theorem and definition).

Definition $\chi_A(\lambda) := \det(\lambda \cdot I - A)$ *is a polynomial function and is called* characteristic polynomial of A.

Theorem *λ is eigenvalue* $\iff \chi_A(\lambda) = 0$

Example 10.10.

$$A = \begin{bmatrix} -1 & 1 & 2 \\ -1 & -5 & 2 \\ 2 & -2 & -4 \end{bmatrix}$$

$$\chi_A(\lambda) = \det(\lambda I - A) = \begin{vmatrix} \lambda+1 & -1 & -2 \\ 1 & \lambda+5 & -2 \\ -2 & 2 & \lambda+4 \end{vmatrix} = \begin{vmatrix} \lambda+1 & -1 & -2 \\ 1 & \lambda+5 & -2 \\ 0 & 2\lambda+12 & \lambda \end{vmatrix}$$

$$= \begin{vmatrix} \lambda & -\lambda-6 & 0 \\ 1 & \lambda+5 & -2 \\ 0 & 2\lambda+12 & \lambda \end{vmatrix} = \lambda \cdot \begin{vmatrix} \lambda+5 & -2 \\ 2\lambda+12 & \lambda \end{vmatrix} - \begin{vmatrix} -\lambda-6 & 0 \\ 2\lambda+12 & \lambda \end{vmatrix}$$

$$= \lambda \cdot [\lambda^2 + 5\lambda + 4\lambda + 24] - \lambda(-\lambda - 6)$$

$$= \lambda(\lambda^2 + 5\lambda + 4\lambda + 24 + \lambda + 6)$$
$$= \lambda(\lambda^2 + 10\lambda + 30)$$
$$x_1 = 0 \qquad \lambda_{2,3} = \frac{-10 \pm \sqrt{10^2 - 120}}{2} = \frac{-10 \pm 2\sqrt{-5}}{2} = -5 \pm i\sqrt{5}$$

Thus, the existence of eigenvalues depends on the field.

Remark. *A square matrix A is invertible if and only if 0 is not an eigenvalue of A.*

10.3 Symmetrical minor

Theorem 10.11. *Let $A \in \mathbb{K}^{n\times n}$. Then $\chi_A(x) = \det(x \cdot I - A)$ is a polynomial of degree n. Specifically, $\chi_A(x) = \sum_{k=0}^{n}(-1)^{n-k}c_k(A) \cdot x^k$ with $c_k(A) := \sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=n-k}} \det(A_{JJ})$ with $A_{JJ} := (a_{ij})_{\substack{i\in J\\ j\in J}}$ are called* symmetrical minors.

Remark. *What are special values of c_i?*

$$c_0 = \det(A) \qquad c_n = 1 \qquad c_{n-1} = \sum a_{ii} = \mathrm{Tr}(A)$$

Proof. The proof is given using the Leibniz formula for determinants.

$$\det(x \cdot I - A) = \sum_{\pi\in\sigma_n}(-1)^\pi \prod_{i=1}^{n} \underbrace{(x \cdot I - A)_{\pi(i),i}}_{x\cdot\delta_{\pi(i),i} - a_{\pi(i),i}}$$
$$= (x - a_{11})(x - a_{22})\dots(x - a_{nn}) + \underbrace{\sum_{\substack{\pi\in\sigma_n\\ \pi\neq\mathrm{id}}}(-1)^\pi \prod_{i=1}^{n}\left(x\delta_{\pi(i),i} - a_{\pi(i),i}\right)}_{\text{for at least 2 } i,\ \delta_{\pi(i),i} = 0}$$
$$= \text{ expression of degree } n + \text{ expression of degree } n-2$$

Hence $\det(x \cdot I - A)$ retains as expression of polynomial degree n. Hence the degree of $\chi_A(x)$ is n.

$$\det \prod_{i=1}^{n}(x\delta_{\pi(i),i} - a_{\pi(i),i}) = \#\{i \mid \pi(i) = i\}$$
$$= \#\,\text{fixedpoints}(\pi)$$

Let $s_1, \dots, s_n$ be the columns of A.

$$\det(xI - A_i) = \triangle(x \cdot e_1 - s_1, x \cdot e_2 - s_2, \dots, x \cdot e_n - s_n) = \sum_{I\subseteq\{1,\dots,n\}} \triangle(y_1, \dots, y_n)$$

$$y_i = \begin{cases} x \cdot e_i & i \in I \\ -s_{i_k} & i \in I^C \end{cases}$$

Let $k \in I$.

$$\triangle(y_1, \dots, y_{k-1}, x \cdot e_k, y_{k+1}, \dots, y_n) = \begin{vmatrix} y_1 & y_2 & \dots & y_{k-1} & 0 & y_{k+1} & \dots & y_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \end{vmatrix}$$

Permute the k-th column into the first column: $(-1)^{k-1}$.
Permute the k-th row into the first row: $(-1)^{k-1}$.

$$= \begin{vmatrix} x & \tilde{y}_1 & \tilde{y}_2 & \dots & k \tilde{-} 1 & k \tilde{+} 1 & \dots & \tilde{n} \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix} = x \cdot \begin{vmatrix} \tilde{\tilde{y}}_1 & \dots & \tilde{\tilde{y}}_{k-1} & \tilde{\tilde{y}}_{k+1} & \vdots & \tilde{\tilde{y}}_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

where $\tilde{y}$ is the permutation of y_i such that the k-th row moved to the first.

Every time, one x is eliminated, the corresponding row and column of A is removed. In the end,

$$x^{|I|} \cdot \underbrace{\det A_{I^C I^C}}_{\text{minor of the complement } |I^C| = n-k} \cdot (-1)^{|I^C|}$$

$$\implies \chi_A(x) = \sum_{I \subseteq \{1,\dots,n\}} x^{|I|} \cdot \det[A_{I^C I^C}](-1)^{|I^C|} \qquad = \sum_{k=0}^{n} x^k (-1)^{n-k} c_k(A)$$

with $c_k(A) = \sum_{|J|=n-k} \det[A_{jj}]$. □

↓ *This lecture took place on 2018/05/16.*

Lemma 10.12.

$$\chi_{T^{-1}AT}(x) = \chi_A(x)$$

Proof.

$$\begin{aligned} \chi_{T^{-1}AT}(x) &= \det(xI - T^{-1}AT) \\ &= \det(xT^{-1}T - T^{-1}AT) \\ &= \det(T^{-1}(x \cdot I)T - T^{-1}(A)T) \\ &= \det(T^{-1}(x \cdot I - A) \cdot T) \\ &= \det(T^{-1}) \cdot \det(xI - A) \cdot \det(T) \\ &= \frac{1}{\det T} \cdot \chi_A(x) \cdot \det T = \chi_A(x) \end{aligned}$$

$\square$

$$A = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \rightsquigarrow \operatorname{spec}(A) = \{a_{11}, \dots, a_{nn}\}$$

Eigenvector: $e_1, \dots, e_n$.

Remark (Question)**.** *Does a change of basis exist, hence $T \in \mathrm{GL}(n)$, such that $T^{-1}AT = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$? Then the eigenvalues are necessarily on the diagonal.*

10.4 Diagonalizable matrix

Definition 10.13. *A is called* diagonalizable *if $\exists T \in \mathrm{GL}(n)$ such that $T^{-1} \cdot AT$ is a diagonal matrix, i.e. A is* similar *to a diagonal matrix.*

Remark (Recall)**.**

Equivalence *$A = PBQ$ with invertible P, Q $\iff$ $\operatorname{rank}(A) = \operatorname{rank}(B)$.*

Congruence *$\exists$ invertible $C : A = C^*BC$*
For $A = A^*, B = B^*, \rightsquigarrow$ index and Sylvester's law of inertia.

Similarity *$A = TBT^{-1}$ with invertible T. This is related to eigenvalues.*

Later on *$\exists T$ such that $T^* = T^{-1}$ unitary. $T^*T = I$.*

Lemma 10.14. *A is diagonalizable $\iff$ $\exists$ basis of eigenvectors.*

Proof. B is invertible such that

$$B^{-1}AB = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \iff \begin{cases} \exists \text{ columns } b_1, \dots, b_n \text{ define a basis} \\ AB = B \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \\ A \cdot \begin{bmatrix} b_1 & b_2 & \dots & b_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \\ \begin{bmatrix} Ab_1 & Ab_2 & \dots & Ab_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} b_1\lambda_1 & b_2\lambda_2 & \dots & b_n\lambda_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{cases}$$

$$\iff \begin{cases} \exists \text{ basis } b_1, \dots, b_n \\ A \cdot b_i = \lambda \cdot b_i \qquad i = 1, \dots, n \end{cases}$$

$\square$

Example 10.15.

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 4 & -3 & -8 \\ -2 & 2 & 5 \end{bmatrix}$$

$$\begin{aligned}
\chi_A(\lambda) = \det(\lambda I - A) &= \begin{vmatrix} \lambda+1 & -2 & -4 \\ -4 & \lambda+3 & 8 \\ 2 & -2 & \lambda-5 \end{vmatrix} = \begin{vmatrix} \lambda-1 & -2 & -4 \\ \lambda-1 & \lambda+3 & 8 \\ 0 & -2 & \lambda-5 \end{vmatrix} \\
&= (\lambda-1)\begin{vmatrix} 1 & -2 & -4 \\ 1 & \lambda+3 & 8 \\ 0 & -2 & \lambda-5 \end{vmatrix} = (\lambda-1)\begin{vmatrix} 1 & -2 & -4 \\ 0 & \lambda+5 & 12 \\ 0 & -2 & \lambda-5 \end{vmatrix} \\
&= (\lambda-1)(\lambda^2-25+24) = (\lambda-1)(\lambda^2-1) = (\lambda-1)^2(\lambda+1)
\end{aligned}$$

Eigenvalue $(\lambda - 1)$ *has multiplicity* 2.

Eigenvector: $\ker(\lambda \cdot I - A)$
Eigenvalue: $\lambda = \pm 1$

Consider $\lambda = +1$*:* $\ker(I - A)$

Homogeneous equation system:

$$\begin{array}{ccc|c} 2 & -2 & -4 & 0 \\ -4 & 4 & 8 & 0 \\ 2 & -2 & -4 & 0 \\ \hline 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{array}$$

$\dim\ker(I - A) = 2$. $2x_1 = 2x_2 + 4x_3$.

Basis:

$$\begin{pmatrix}1\\1\\0\end{pmatrix} \quad \begin{pmatrix}2\\0\\1\end{pmatrix}$$

Consider $\lambda = -1$*:* $\ker(-I - A)$

$$\begin{array}{ccc|c} 0 & -2 & -4 & 0 \\ -4 & 2 & 8 & 0 \\ 2 & -2 & -6 & 0 \\ \hline 0 & -2 & -4 & \\ 0 & -2 & -4 & \\ \hline 0 & 0 & 0 & \end{array}$$

$\dim\ker(-I - A) = 1$.

Basis:

$$\begin{aligned} x_3 &= 1 \\ x_2 &= -2x_3 = -2 \\ x_1 &= \frac{2x_2 + 6x_3}{2} = 1 \end{aligned}$$

$$b_3 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

with $B = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix}$ *it holds that* $B^{-1}AB = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}$.

Example (Application)**.**

$$A = B^{-1} \cdot \underbrace{\begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & \Lambda_n \end{bmatrix}}_{\Lambda} \cdot B$$

$$\begin{aligned} A^2 &= B^{-1}\Lambda B \cdot B^{-1}\Lambda B = B^{-1}\Lambda^2 B \\ A^3 &= B^{-1}\Lambda^3 B \\ &\vdots \\ A^k &= B^{-1}\Lambda^k B \end{aligned}$$

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{B^{-1}\Lambda^k B}{k!} = B^{-1} \sum_{k=0}^{\infty} \frac{\Lambda^k}{k!} B = B^{-1} \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n} \end{bmatrix}$$

Remark. *Leondaro Pisano (1170–1250) wrote his book "Liber Abbaci" (1202) to introduce the Arabic numbers (and zero) in Europe. He also introduced the Fibonacci sequence using the growth of a rabbit population.*

10.5 Fibonacci sequence and golden ratio

Remark (Fibonacci sequence)**.**

$$F_0 = F_1 = 1 \qquad F_n = F_{n-1} + F_{n-2}$$

Can we find a formula for F_n*?*

Person. *Pingala (200 BC)*

Remark. *How many ways are there for the equation* $x_1 + \cdots + x_k = n$ *for given* n *and* x_i *in* $\{1, 2\}$*? The answer is the Fibonacci sequence.*

His application was the number of long syllables (2) or short syllables (1) in a sentence of given length in Sanskrit.

Remark (Growth of Fibonacci sequence)**.**

$$F_{n+1} = F_n + F_{n-1} \qquad F_n = F_n$$

$$\begin{aligned} \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \begin{pmatrix} F_n + F_{n-1} \\ F_n \end{pmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2 \begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^3 \begin{bmatrix} F_{n-2} \\ F_{n-3} \end{bmatrix} \\ &= \cdots = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{aligned}$$

and where $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ *is diagonalizable.*

$$\chi_A(\lambda) = \begin{vmatrix} \lambda - 1 & -1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 - \lambda - 1 \qquad \lambda_{1,2} = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1 \pm \sqrt{5}}{2}$$

Eigenvector: *Consider* $\lambda_1 := \frac{1+\sqrt{5}}{2}$.

$$\left(\begin{array}{cc|c} \frac{1+\sqrt{5}}{2} - 1 & -1 & 0 \\ -1 & \frac{1+\sqrt{5}}{2} & 0 \end{array}\right) \implies x_1 = \frac{1+\sqrt{5}}{2} x_2 \qquad b_1 = \begin{bmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{bmatrix}$$

Consider $\lambda_2 = \frac{1-\sqrt{5}}{2}$.

$$\left(\begin{array}{cc|c} \frac{1-\sqrt{5}}{2} & -1 & 0 \\ -1 & \frac{1-\sqrt{5}}{2} & 0 \end{array}\right) \implies x_1 = \frac{1-\sqrt{5}}{2} x_2 \qquad b_2 = \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix} \qquad \det B = \frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2} = \sqrt{5} \qquad B^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ -1 & \frac{1+\sqrt{5}}{2} \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad B^{-1}AB = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = A^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} = B \begin{bmatrix} (\frac{1+\sqrt{5}}{2})^n & \\ & (\frac{1-\sqrt{5}}{2})^n \end{bmatrix} \cdot B^{-1} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

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

$$\frac{F_{n+1}}{F_n} = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n+2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}} \xrightarrow{n \to \infty} \frac{1+\sqrt{5}}{2} \approx 1.6180$$

This limit is called Golden ratio. *It is the ratio:*

$$\frac{a}{a+b} = \frac{b}{a} \qquad \frac{F_n}{F_{n-1}} = \frac{1}{1 + \frac{1}{1 + \frac{1}{\dots}}}$$

Theorem 10.16. *Eigenvectors corresponding to different eigenvalues are linear independent.*

Proof. Let $\lambda_1, \dots, \lambda_r$ be different eigenvalues of A. Let $v_1, \dots, v_r$ be their respective eigenvectors. Proof by induction over r.

Case $r = 1$**:** immediate, $v_1 \neq 0$.

Case $r - 1 \to r$**:** Let $\alpha_1 v_1 + \dots + \alpha_r v_r = 0$.

$$\implies A \sum_{i=1}^{r} \alpha_i v_i = 0 \qquad \sum_{i=1}^{r} \alpha_i A v_i = 0 \qquad \sum_{i=1}^{r} \alpha_i \lambda_i v_i = 0$$

$$\begin{array}{rrl} (1): & \sum_{i=1}^{r} \alpha_i v_i & = 0 \\ (2): & \sum_{i=1}^{r} \lambda_i \alpha_i v_i & = 0 \\ \hline (2) - \lambda_r(1) \implies & \sum_{i=1}^{r} (\lambda_i - \lambda_r) \alpha_i v_i & = 0 \end{array}$$

By induction hypothesis: $v_1, \dots, v_{r-1}$ are linear independent.

$$\begin{aligned} \implies (\lambda_1 - \lambda_r)\alpha_1 &= 0 \\ (\lambda_2 - \lambda_r)\alpha_2 &= 0 \\ &\vdots \\ (\lambda_{r-1} - \lambda_r)\alpha_{r-1} &= 0 \end{aligned}$$

As all eigenvalues are different, $\lambda_i - \lambda_r \neq 0 \forall i < r$.

$$\implies \alpha_1 = \alpha_2 = \dots = \alpha_{r-1} = 0$$

$$(1) \implies \alpha_r \cdot v_r = 0 \implies \alpha_r = 0 \text{ because } v_r \neq 0$$

□

Corollary 10.17. *An $n \times n$ matrix with n different eigenvalues is diagonalizable.*

Hence, for every eigenvalue there exists some eigenvector. They are linear independent and n elements. Hence they define a basis.

Example 10.18.

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

$$\chi_A(\lambda) = \begin{vmatrix} \lambda & -1 \\ 0 & \lambda \end{vmatrix} = \lambda^2$$

$$\mathrm{spec}(A) = \{0\}$$

$$\dim \ker(A) = 1$$

is not a basis of eigenvectors.

$$A^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

A is nilpotent, hence a square matrix M such that $\exists k \in \mathbb{N}_{\geq 1} : M^k = 0$.

10.6 Multiplicities of eigenvalues

Definition 10.19. *Let λ be the eigenvalue of a matrix $A \implies \chi_A(\lambda) = 0$. The*

geometric multiplicity of λ *is defined as* $d(\lambda) := \dim \ker(\lambda I - A) > 0$

algebraic multiplicity of λ *is defined as, $k(\lambda)$, the multiplicity of λ as root of $\chi_A(\lambda)$*

$$d(\lambda) \leq k(\lambda)$$

Lemma 10.20. *A matrix is diagonalizable iff for different eigenvalues $\lambda_1, \ldots, \lambda_r$ it holds that*

$$d(\lambda_1) + d(\lambda_2) + \cdots + d(\lambda_r) = n$$

Proof.

Direction $\implies$: There exists a basis of eigenvectors $b_1, \ldots, b_n$.

$$V = \eta_{\lambda_1} + \cdots + \eta_{\lambda_r} \qquad \eta_{\lambda_i} = \ker(\lambda_i I - A)$$

is a direct sum (because eigenvectors for different eigenvalues are linear independent). Let $v_1 \in \eta_{\lambda_1}, \ldots, v_r \in \eta_{\lambda_r}$ such that $v_1 + \cdots + v_r = 0$.

$$Av_i = \lambda_i v_i \implies v_1, \ldots, v_r \text{ are linear independent } \implies \text{ all } v_i = 0$$

$$\implies n = \dim V = \dim(\eta_{\lambda_1}) + \cdots + \dim(\eta_{\lambda_r}) = d(\lambda_1) + \cdots + d(\lambda_r)$$

Direction $\impliedby$. Let B_j be the basis of η_{λ_j}, hence $\left|B_j\right| = d(\lambda_j)$. The sum $\eta_{\lambda_1} + \cdots + \eta_{\lambda_r}$ is direct. $\implies$ $B_i \cup \cdots \cup B_r$ is linear independent.

$$|B_1 \cup \cdots \cup B_r| = \sum_{j=1}^{r} d(\lambda_j) \underbrace{=}_{\text{by induction}} \eta$$

$B_i \cup \cdots \cup B_n$ is basis of $\mathbb{K}^n$ of eigenvectors.

□

Remark. *Why are there at most n eigenvalues?*

We know the characteristic polynomial has degree n.

For every eigenvalue, we have an algebraic multiplicity. Every algebraic multiplicity is at least 1. If there are more than n eigenvalues, namely $n+1$, the sum of algebraic multiplicities of eigenvalues is at least $n+1$. Thus the degree of the characteristic polynomial (which is the sum of algebraic multiplicities of eigenvalues), is at least $n+1$ which contradicts our previous proof.

Theorem 10.21. *For every eigenvalue, it holds that*

$$d(\lambda) \leq k(\lambda)$$

Hence, the geometric multiplicity is smaller than the algebraic multiplicity.

Proof. Let $\lambda \in \operatorname{spec}(A)$. Let $d = d(\lambda)$. Let $(b_1, \ldots, b_d)$ be a basis of $\ker(\lambda I - A)$. We extend this vector to a basis of $\mathbb{K}^n : (b_1, \ldots, b_d, \ldots, b_n)$.

$$B = \begin{pmatrix} b_1 & b_2 & \ldots & b_n \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$
\begin{aligned}
AB &= \begin{bmatrix} Ab_1 & Ab_2 & \dots & Ab_d & Ab_{d+1} & \dots & Ab_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \\
&= \begin{bmatrix} \lambda b_1 & \lambda b_2 & \dots & \lambda b_d & Ab_{d+1} & \dots & Ab_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \\
&= \begin{bmatrix} b_1 & \dots & b_d & b & \dots & b_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \lambda & & & \dots \\ & \lambda & & \dots \\ & & \lambda & \dots \\ 0 & 0 & 0 & \dots \\ & \dots 0 \dots & & \dots \\ 0 & 0 & 0 & \dots \end{bmatrix} \text{where } \lambda \text{ occurs in } d \text{ different columns}
\end{aligned}
$$

$$
B^{-1}AB = \begin{bmatrix} \lambda & & & \dots \\ & \ddots & & \dots \\ & & \lambda & \dots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \tilde{A} \\ 0 & 0 & 0 & \end{bmatrix} =: M
$$

$$
\begin{aligned}
\chi_A(x) = \chi_{B^{-1}AB}(x) &= \det(x \cdot I - M) \\
&= \begin{bmatrix} x-\lambda & & & \dots \\ & \ddots & & \dots \\ & & x-\lambda & \dots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & xI_{n-d} - \tilde{A} \\ 0 & 0 & 0 & \end{bmatrix} \\
&= (x-\lambda)^d \det(xI - \tilde{A})
\end{aligned}
$$

$\implies$ $x - \lambda$ is d-multiple factor of $\chi_A(x)$ $\implies$ $k(\lambda) \geq d(\lambda)$. □

↓ *This lecture took place on 2018/05/23.*

Corollary 10.22. *A is diagonalizable iff $\exists T$:*

$$
T^{-1}AT = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}
$$

where T is basis of eigenvectors.

$$
\iff d(\lambda) = \textit{geometric multiplicity} = \dim \eta_\lambda
$$

$$
\overset{!}{=} k(\lambda) = \textit{algebraic multiplicity}
$$

Example.

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

Eigenvalues:

$$\chi_A(x) = \begin{vmatrix} x-\lambda & -1 & 0 \\ 0 & x-\lambda & -1 \\ 0 & 0 & x-\lambda \end{vmatrix} = (x-\lambda)^3$$

The only eigenvalue: λ. $k(\lambda) = 3$.

$$\ker \eta_\lambda = \ker \begin{bmatrix} 0 & -1 & 0 \\ & 0 & -1 \\ & & 0 \end{bmatrix}$$

$$d(\lambda) = \dim \ker \eta_\lambda = 1 \implies \textit{not diagonalizable}$$

Person. *Camille Jordan (1838–1922): Jordan curve theorem.*

11 Jordan Normal Form (JNF)

Person. *Causs-Wilhelm-Jordan (1842–1899)*

Nilpotent matrix:

$$\begin{bmatrix} 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

11.1 Invariant subspaces

Definition 11.1. *Let $f : V \to V$ (or matrix $A \in \mathbb{K}^{n \times n}$) be linear. A subspace $U \subseteq V$ is called* invariant *under f, if $f(U) \subseteq U$ (and accordingly $\forall x \in U : Ax \in U$)*

Example 11.2.

1. $\{0\}$. $f(0) = 0 \in \{0\}$. *V is trivially invariant.*
2. $\ker f$. *Let $x \in \ker f \implies f(x) = 0 \implies f(f(x)) = 0 \implies f(x) \in \ker f$.*
   $\operatorname{im} f$ *is invariant. $y \in \operatorname{im} f \implies f(y) \in \operatorname{im} f$.*
3. *Eigenspaces are invariant.*
   $$f(x) = \lambda \cdot x \implies f(f(x)) = f(\lambda \cdot x) = \lambda \cdot f(x) \implies f(x) \in \eta_\lambda$$
4. *If U are invariant with $\dim U = 1$, then $U = \mathcal{L}(x)$ with x as eigenvector.*
   If $x \in U \setminus \{0\}$ ($\implies U = \mathcal{L}(x)$), $f(x) \in U$. $\exists \lambda \cdot f(x) = \lambda \cdot x \implies x$ is eigenvector.
5. $A = \begin{bmatrix} a_{11} & \dots & \dots & \dots \\ & a_{22} & \dots & \dots \\ & & \ddots & \dots \\ & & & a_{nn} \end{bmatrix}$.
   $$A \cdot e_1 = a_{11} \cdot e_1 \implies e_1 \textit{ is eigenvalue} \implies \mathcal{L}(e_1) \textit{ is invariant}$$

$$A(\lambda_1 e_1+\lambda_2 e_2) = \lambda_1 a_{11} e_1 + a_{12}\lambda_2 e_2 + a_{22}\lambda_2 e_2 \in \mathcal{L}(e_1, e_2) \implies \mathcal{L}(e_1, e_2) \text{ is invariant}$$

$$A \cdot e_k \in \mathcal{L}(e_1, \dots, e_k) \implies \forall k : \mathcal{L}(e_1, \dots, e_k) \text{ is invariant}$$

Numerically unstable:

$$\begin{bmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{bmatrix} \text{ is diagonalizable.} \qquad \begin{bmatrix} \lambda & & \varepsilon \\ & \ddots & \\ 0 & & \lambda \end{bmatrix} \text{ is not diagonalizable}$$

Theorem 11.3. *Let $A \in \mathbb{K}^{n \times n}$, $V = \mathbb{K}^n$.*

1. *If $U \subseteq V$ is invariant under A and $p(x) \in \mathbb{K}[x]$, then U is invariant under $p(A)$.*

$$p(x) := \sum_{k=0}^{n} a_k x^k \qquad p(A) = \sum_{k=0}^{n} a_k A^k \qquad \psi_A : {}_{x \mapsto A}^{\mathbb{K}[x] \to \mathbb{K}^{n \times n}}$$

2. *$U_1, \dots, U_k$ are invariant subspaces.*

$$\implies U_1 \cap \dots \cap U_k, U_1 + \dots + U_k \text{ are invariant with respect to } A$$

Proof. 1. Let $x \in U$.

$$\implies Ax \in U \qquad A^2 \cdot x = A \cdot \underbrace{(Ax)}_{\in U} \in U \quad \dots \quad A^k x \in U \implies \sum a_k A^k x \in U$$

because it is a linear combination of elements of U where U is the subspace with $A^k x \in U$.

2. Let $x \in \bigcap_{i=1}^{k} U_i$. $\implies \forall i : x \in U_i \implies \forall i : Ax \in U_i \implies Ax \in \bigcap_{i=1}^{k} U_i$. Let $x \in U_1 + \dots + U_k \implies x = u_1 + u_2 + \dots + u_k$ for $u_i \in U_i$.

$$\implies Ax = (\underbrace{Au_1}_{\in U_1} + \underbrace{Au_2}_{\in U_2} + \dots + \underbrace{Au_k}_{\in U_k}) \in U_1 + \dots + U_k$$

$$\implies U_1 + \dots + U_k \text{ is invariant}$$

□

Lemma 11.4. *Let $f : V \to V$ and $U \subseteq V$ be an invariant subspace. $\implies f|_U : U \to U$ is a homomorphism. (If U is not invariant, $\varphi|_U : U \to V$ must not map $U \to U$.)*

Theorem 11.5. *Let $f : V \to V$. Let $U, W \subseteq V$ be invariant with $V = U \dot{+} W$. Let $B = \{b_1, \dots, b_m\}$ be a basis of U, $B' = \left\{b'_1, \dots, b'_n\right\}$ is basis of W. $\implies B \cup B'$ is basis of V.*

$$\Phi_{B \cup B'}^{B \cup B'}(f) = \left[\begin{array}{c|c} \Phi_B^B(f|_U) & 0 \\ \hline 0 & \Phi_{B'}^{B'}(f|_W) \end{array}\right]$$

Proof of Theorem 11.3. In the first m columns, we have the images of b_i (basis of U)

$$U \text{ invariant} \implies f(b_i) \in U$$

$$\implies \text{coordinates in regards of } b'_1 \dots b'_n \text{ are } 0$$

$$\begin{bmatrix} f(b_1) & \dots & f(b_m) & f(b'_1) & \dots & f(b'_m) \\ & \ddots & & 0 & & 0 \\ & & & 0 & & 0 \\ 0 & \dots & 0 & \ddots & & \\ 0 & \dots & 0 & & \ddots & \\ 0 & \dots & 0 & & & \ddots \end{bmatrix}$$

In the last n columns, we can find the images of b'_j. W is invariant $\implies f(b'_j) \in W \implies$ coordinate in regards of $b_1, \dots, b_m$ are 0. □

Corollary 11.6. *Let $f : V \to V$. $U_1, \dots, U_k \subseteq V$ is invariant with $V = U_1 \dot{+} U_2 \dot{+} \dots \dot{+} U_k$. Let B_i be basis of $U_i \implies B = B_1 \cup \dots \cup B_k$ is basis of V and*

$$\Phi_B^B(f) = \left[\begin{array}{c|ccc} \Phi_{B_1}^{B_1}(f|_{U_1}) & 0 & 0 & 0 \\ \hline 0 & \Phi_{B_2}^{B_2}(f|_{U_2}) & 0 & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & & & \Phi_{B_k}^{B_k}(f|_{U_k}) \end{array}\right]$$

Hence, if V can be decomposed into a direct sum of invariant subspaces, then A can be transformed into block diagonal form. (A is diagonalizable $\iff$ V can be decomposed into direct sum of one-dimensional subspaces)

Corollary 11.7. *Corollary related to Corollary 11.6.*

$$\chi_f(x) = \prod_{i=1}^{k} \chi_{f|_{U_i}}(x)$$

11.2 Fitting lemma

Person. *Hans Fitting (1906–1938)*

Lemma 11.8 (Fitting lemma)**.** *Let* $\dim V = n, f \in \operatorname{End}(V)$.

1. $\{0\} \subseteq \ker f \subseteq \ker f^2 \subseteq \ker f^3 \subseteq \dots$
   $\operatorname{im} f \supseteq \operatorname{im} f^2 \supseteq \operatorname{im} f^3 \supseteq \dots$
2. $\exists m \le n : \ker f^m = \ker f^{m+1}$
3. *The following statements are equivalent:*
   (a) $\ker f^m = \ker f^{m+1}$
   (b) $\operatorname{im} f^m = \operatorname{im} f^{m+1}$

(c) $\ker f^m = \ker f^{m+k} \forall k \geq 1$

(d) $\operatorname{im} f^m = \operatorname{im} f^{m+k} \forall k \geq 1$

(e) $\ker f^m \cap \operatorname{im} f^m = \{0\}$

(f) $V = \ker f^m \dot{+} \operatorname{im} f^m$

Proof. 1. Let $x \in \ker f$. Then $f^2(x) = f(f(x)) = f(0) = 0$.
Let $y \in \operatorname{im} f^2$. Then $\exists x : y = f(f(x))$. Thus $f(x) \in \operatorname{im} f$.

2. If $\{0\} \subsetneq \ker f \subsetneq \ker f^2 \subsetneq \cdots \subsetneq \ker(f^m)$

$$\implies 0 < \dim \ker f < \dim \ker f^2 < \cdots < \dim \ker f^m \qquad \implies m \leq n$$

3. We prove a set of equivalences.

- We prove (a) $\iff$ (b). Because of (1.), we know

$$\begin{aligned} &\ker(f^m) \subseteq \ker(f^{m+1}) \\ \implies &\ker(f^m) = \ker(f^{m+1}) \\ \iff &\dim \ker f^m = \dim \ker f^{m+1} \\ \iff &n - \dim \operatorname{im}(f^m) = n - \dim \operatorname{im}(f^{m+1}) \\ \iff &\dim \operatorname{im}(f^m) = \dim \operatorname{im}(f^{m+1}) \end{aligned}$$

  Because of (1.), $\operatorname{im} f^m \supseteq f^{m+1}$

$$\iff \operatorname{im}(f^m) = \operatorname{im}(f^{m+1})$$

- The proof of (c) $\iff$ (d) follows analogously. The proofs (a) $\iff$ (c) and (d) $\iff$ (b) are trivial.
- We prove (a) $\iff$ (c):

$$0 \subseteq \ker f \subseteq \ker f^2 \subseteq \ker f^3 \subseteq \dots$$

$$m_0 = \min\left\{ m \mid \ker(f^m) = \ker(f^{m+1}) \right\}$$

  Claim.
$$\ker f^{m_0+k} = \ker f^{m_0+k+1} \forall k \geq 0$$

  Direction $\subseteq$: Immediate.
  Direction $\supseteq$: Let $x \in \ker f^{m_0+k+1} \implies f^{m_0+k+1}(x) = f^{m_0+1}(f^k(x)) = 0$.

$$\implies f^k(x) \in \ker f^{m_0+1} = \ker f^{m_0} \implies f^{m_0+k}(x) = 0 \implies x \in \ker f^{m_0+k}$$

  with $f^k(x) \in \ker f^{m_0+1} = \ker f^{m_0}$ by the definition of m_0.
- We prove (b) $\iff$ (d). Let $m_0 = \min\{m \mid \operatorname{im} f^m = \operatorname{im} f^{m+1}\}$.
  Claim. $\operatorname{im} f^{m_0+k} = \operatorname{im} f^{m_0+k+1} \forall k \geq 0$.

  Direction $\supseteq$: Trivial.

Direction $\subseteq$**:** Let $y \in \operatorname{im} f^{m_0+k}$.

$$\implies \exists x : y = f^{m_0+k}(x) = f^k(f^{m_0}(x)) \text{ with } f^{m_0}(x) \in \operatorname{im} f^{m_0} = \operatorname{im} f^{m_0+1}$$

hence $\exists z : f^{m_0}(x) = f^{m_0+1}(z)$.

$$\implies y = f^k(f^{m_0+1}(z)) = f^{m_0+k+1}(z) \in \operatorname{im} f^{m_0+k+1}$$

- We prove $\{(a), (b), (c), (d)\} \iff$ (e). Let $W = \operatorname{im} f^m$ be invariant under f^m.

$$g \coloneqq f^m|_W \in \operatorname{Hom}(W, W)$$
$$\ker g = \ker(f^m) \cap W = \ker f^m \cap \operatorname{im} f^m$$
$$\ker f^m \cap \operatorname{im} f^m = \{0\} \iff \ker g = \{0\}$$
$$\iff g \text{ linear} \wedge g \text{ injective } \wedge g \text{ surjective } \iff \operatorname{im} g = W$$
$$\iff f^m(f^m(V)) = f^m(V) \iff f^{2m}(W) = f^m(W)$$
$$f^{m+m}(v) = f^m(v)$$
$$\operatorname{im} f^{m+m} = \operatorname{im} f^m$$

This is equivalent to statements 1 and 3, thus rendering our assumption $\ker f^m \cap \operatorname{im} f^m = \{0\}$ valid.

- We prove (d) $\iff$ (b).

$$\operatorname{im} f^{m+m} = \operatorname{im} f^m \iff \operatorname{im} f^{m+1} = \operatorname{im} f^m$$

- (f) $\implies$ (e) is trivial.
- We prove (e) $\impliedby$ (f).

$$\left.\begin{array}{c} \dim \operatorname{im} f^m + \dim \ker f^m = n \\ \operatorname{im} f^m \cap \ker f^m = \{0\} \end{array}\right\} \implies \operatorname{im} f^m \dot{+} \ker f^m = V$$

because of dimensionality reasons.

□

Remark. *The intersection of image and kernel of a linear map is only* $\{0\}$.

↓ *This lecture took place on 2018/05/28.*

Fitting Lemma:

$$\ker A \subseteq \ker A^2 \subseteq \cdots \subseteq \ker A^r$$
$$\operatorname{im} A \supseteq \operatorname{im} A^2 \supseteq \cdots \supseteq \operatorname{im} A^r = \operatorname{im} A^{r+1}$$
$$\ker A^r \oplus \operatorname{im} A^r = V$$
$$\ker A^r \cap \operatorname{im} A^r = \{0\}$$

Example 11.9.

$$A = \begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix} \qquad \ker A = \mathcal{L}\begin{pmatrix}1\\0\\0\end{pmatrix} \qquad \operatorname{im} A = \mathcal{L}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right)$$

$$A^2 = \begin{bmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix} \qquad \ker A^2 = \mathcal{L}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right) \qquad \operatorname{im} A^2 = \mathcal{L}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right)$$

$$A^3 = \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix} \qquad \ker A^3 = \mathbb{K}^3 \qquad \operatorname{im} A^3 = \{0\}$$

$$\boxed{x \in \ker A^k \implies A \cdot x \in \ker A^{k-1}}$$

Example.

$$A = \begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix} \qquad \textit{eigenvalue: } \lambda \qquad \lambda I - A = \begin{bmatrix} 0 & -1 & \\ & 0 & -1 \\ & & 0 \end{bmatrix}$$

$$\ker(\lambda I - A) = \mathcal{L}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) \qquad \ker((\lambda I - A)^2) = \mathcal{L}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right)$$

$$\begin{aligned} x \in \ker((\lambda I - A)^2) &\implies (\lambda I - A)x \in \ker(\lambda I - A) \\ &\implies Ax = \lambda x + b \qquad b \in \ker(\lambda I - A) \end{aligned}$$

11.3 Generalized space

Definition 11.10. *Let* $A \in \mathbb{K}^{n \times n}$, $\lambda \in \operatorname{spec}(A)$. *Then* $\ker(\lambda I - A)^n$ *is called* generalized space *(dt. Hauptraum) of* A *for eigenvalue* λ. *The elements are called* generalized eigenvectors.

Actually, $\ker(\lambda I - A)^n = \ker(\lambda I - A)^r$ *and a generalized eigenvector satisfies*

$$A \cdot x = \lambda \cdot x + y \qquad y \in \ker(\lambda I - A)^{r-1}$$

with r *as first index such that* $\ker(\lambda I - A)^r = \ker(\lambda I - A)^{r+1}$.

Fitting:

$$\mathbb{K}^n = \ker(\lambda I - A)^r \oplus \operatorname{im}(\lambda I - A)^r$$

Next step: decomposition for *different* eigenvalues.

Lemma 11.11. *Let* $\lambda_1, \dots, \lambda_k$ *be different eigenvalues of* A *and* $\ker(\lambda I - A)^{r_i}$ *the corresponding generalized spaces where*

$$\ker(\lambda_i I - A)^{r_{i-1}} \subsetneq \ker(\lambda_i I - A)^{r_i} = \ker(\lambda_i \cdot I - A)^{r_i + 1}$$

$$\implies \bigcap_{i=1}^{k} \operatorname{im}(\lambda_i - A)^{r_i} \cap \ker((\lambda_1 I - A)^{r_1} (\lambda_2 I - A)^{r_2} \dots (\lambda_k I - A)^{r_k}) = \{0\}$$

Remark.

$$\ker(\lambda_1 I - A)^{r_1} \dots (\lambda_k I - A)^{r_k} \supseteq \ker(\lambda_i I - A)^{r_i} \forall i$$
$$\supseteq \ker(\lambda_1 I - A)^{r_1} + \dots + \ker(\lambda_k I - A)^{r_k}$$

if $(\lambda_i I - A)^{r_i} \cdot x = 0$,

$$(\lambda_1 I - A)^{r_1} \dots (\lambda_k I - A)^{r_k} \cdot x$$
$$= (\lambda_1 I - A)^{r_1} \dots (\lambda_{i-1} I - A)^{r_{i-1}} (\lambda_{i+1} I - A)^{r+1} \dots (\lambda_k I - A)^{r_k} \cdot \underbrace{(\lambda_i \cdot I - A)^{r_i}}_{=0} x = 0$$

If one of the factors is zero, the product is zero.

$$p(A) \cdot q(A) = q(A) \cdot p(A) \textit{ for arbitrary polynomials } p(x) \textit{ and } q(x)$$

especially: $p_i(x) = (\lambda_i - x)^{r_i}$. *[I think this can be derived from Lemma 9.5 (1)]*

Example 11.12.

$$A = \begin{bmatrix} 1 & 1 & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 2 & 1 & 0 & & \\ & & & & 2 & 1 & & \\ & & & & & 2 & & \\ & & & & & & 0 & -1 \\ & & & & & & 1 & 0 \end{bmatrix} \textit{ over } \mathbb{R}$$

$$\operatorname{spec}(A) = \{1, 2\} \cup \underbrace{(\{\pm i\})}_{\notin \mathbb{R}}$$

Consider $\lambda = 1$.

$$\ker(I - A) = \ker \begin{bmatrix} 0 & -1 & & & & & & \\ & 0 & & & & & & \\ & & 0 & & & & & \\ & & & -1 & -1 & & & \\ & & & & -1 & -1 & & \\ & & & & & -1 & & \\ & & & & & & 1 & 1 \\ & & & & & & -1 & 1 \end{bmatrix}$$

$$\ker(I - A) = \mathcal{L}(e_1, e_2) \qquad \textit{the two zero columns}$$

$$\ker(I - A)^2 = \ker \begin{bmatrix} 0 & 0 & & & & & & \\ 0 & 0 & & & & & & \\ & & 0 & & & & & \\ & & & -1 & -1 & & & \\ & & & & -1 & -1 & & \\ & & & & & -1 & & \\ & & & & & & 1 & 1 \\ & & & & & & -1 & 1 \end{bmatrix}$$

$$\ker(I - A)^2 = \mathcal{L}(e_1, e_2, e_3)$$

$$\operatorname{im}(I-A)^2 = \mathcal{L}(e_4, e_5, e_6, e_7, e_8)$$

Consider $\lambda = 2$.

$$\ker(2I - A) = \ker \begin{bmatrix} 1 & -1 & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 0 & -1 & & & \\ & & & & 0 & -1 & & \\ & & & & & 0 & & \\ & & & & & & 2 & 1 \\ & & & & & & -1 & 2 \end{bmatrix}$$

$$\ker(2I - A) = \mathcal{L}(e_4)$$

$$\ker(2I - A)^2 = \ker \begin{bmatrix} 1 & 1 & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 0 & 0 & 1 & \\ & & & & 0 & 0 & \\ & & & & & 0 & \\ & & & & & & [\textit{invertible}] \end{bmatrix}$$

$$\ker(2I - A)^2 = \mathcal{L}(e_4, e_5)$$

$$\ker(2I - A)^3 = \mathcal{L}(e_4, e_5, e_6)$$

$$\ker(2I - A)^3 = \mathcal{L}(e_1, e_2, e_3, e_7, e_8)$$

$$\bigcap_{i=1}^{2} \operatorname{im}(\lambda_i I - A)^{r_i} = \mathcal{L}(e_1, e_2, e_3, e_7, e_8) \cap \mathcal{L}(e_4, e_5, e_6, e_7, e_8) = \mathcal{L}(e_7, e_8)$$

$$(I - A)^2 \cdot (2I - A)^3 = \begin{bmatrix} 0 & 0 & & & \\ 0 & 0 & & & \\ & & 0 & & \\ & & & [\textit{invertible } 3 \times 3] & \\ & & & & [\textit{invertible } 2 \times 2] \end{bmatrix}$$

$$\cdot \begin{bmatrix} [\textit{invertible } 2 \times 2] & & & & & \\ & 1 & & & & \\ & & 0 & 0 & 0 & \\ & & 0 & 0 & 0 & \\ & & 0 & 0 & 0 & \\ & & & & & [\textit{invertible } 2 \times 2] \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & & & & & \\ 0 & 0 & & & & & \\ & & 0 & & & & \\ & & & 0 & 0 & 0 & \\ & & & 0 & 0 & 0 & \\ & & & 0 & 0 & 0 & \\ & & & & & & [\textit{invertible } 2 \times 2] \end{bmatrix}$$

$$\ker\,(I - A)^2 \cdot (2I - A)^3 = \mathcal{L}(e_1, \ldots, e_6) = \left(\bigcap (\lambda_i I - A)^{r_i}\right)^c$$

Remark. *Let f and g be matrices. Let A and B be corresponding linear maps. Then $A \cdot B$ is the matrix corresponding to $f \circ g$.*

Proof of Lemma 11.11. Show: If $x \in \bigcap i = 1^k \operatorname{im}(\lambda_i I - A)^{r_i}$ and $(\lambda_1 I - A)^{r_1} \dots (\lambda_k I - A)^{r_k} \cdot x = 0$, then $x = 0$.

Proof by induction over k:

Case $k = 1$

$$x \in \operatorname{im}(\lambda_1 I - A)^{r_1} \wedge (\lambda_1 I - A)^{r_1} x = 0$$
$$\xRightarrow{\text{Fitting}} x = 0$$

Case $k \to k+1$ Let $x \in \bigcap_{i=1}^{k} \operatorname{im}(\lambda_i I - A)^{r_i}$ and $(\lambda_1 I - A)^{r_1} \dots (\lambda_{k+1} I - A)^{r_{k+1}} x = 0$. Let $y = (\lambda_{k+1} - A)^{r_{k+1}} x \implies y \in \ker(\lambda_i I - A)^{r_i} \dots (\lambda_k I - A)^{r_k}$.

$$\begin{aligned}
\forall i \in \{1, \dots, k+1\} \, \exists u_i : x &= (\lambda_i I - A)^{r_i} \cdot u_i \\
y = (\lambda_{k+1} - A)^{r_{k+1}} x &= (\lambda_{k+1} - A)^{r_{k+1}} (\lambda_i I - A)^{r_i} \cdot u_i \\
&= (\lambda_i I - A)^{r_i} (\lambda_{k+1} I - A)^{r_{k+1}} u_i \\
&\in \operatorname{im}(\lambda_i I - A)^{r_i}
\end{aligned}$$

$$p(A)q(A) = q(A)p(A)$$
$$p(x) = (\lambda_{k+1} - x)^{r_{k+1}} \qquad q(x) = (\lambda_i - x)^{r_i}$$
$$\implies y \in \bigcap_{i=1}^{k} \operatorname{im}(\lambda_i I - A)^{r_i}$$

By the induction hypothesis, $y = 0$.

$$\implies x \in \ker(\lambda_{k+1} I - A)^{r_{k+1}} \wedge x \in \operatorname{im}(\lambda_{k+1} I - A)^{r_{k+1}}$$
$$\xRightarrow{\text{Fitting}} x = 0$$

□

Lemma 11.13.

1. $\forall \lambda \neq \mu \in \operatorname{spec}(A) \forall k, l \geq 1 : \ker(\lambda I - A)^k \cap \ker(\mu I - A)^l = \{0\}$
2. *The sum $\sum_{i=1}^{k} \ker(\lambda_i I - A)^{r_i}$ is direct for arbitrary pairwise different $\lambda_1, \dots, \lambda_k$.*

Proof. Proof of the first statement. Induction over $m = k + l$.

Induction base Consider $m = 2, k = l = 1$.

$$\ker(\lambda I - A) \cap \ker(\mu I - A) = \{0\}$$

The eigenvectors for different eigenvalues are linear independent.

Induction step $m-1 \to m$**:** Consider $m \geq 3$. Without loss of generality: $k \geq 2$. Let $x \in \ker(\lambda I - A)^k \cap \ker(\mu I - A)^l$. Let $y = (\lambda I - A)x \in \ker(\lambda I - A)^{k-1} \cap \ker(\mu I - A)^l$. Then,

$$(\mu I - A)^l \cdot y = (\mu I - A)^l (\lambda I - A) \cdot x = (\lambda I - A) \underbrace{(\mu I - A)^l \cdot x}_{=0} = 0$$

Let $k - 1 + l = m - 1$. By induction hypothesis, $y = 0$.

$$\implies x \in \ker(\lambda I - A)$$

$$\implies x \in \ker(\lambda I - A) \cap \ker(\mu I - A)^l \xRightarrow[1 + l \leq m - 1]{\text{induction hypothesis}} x = 0$$

Proof of the second statement. Induction over k.

Induction base $k = 1$**:** trivial

Induction step $k \to k+1$

Show: if $v_i \in \ker(\lambda_i I - A)^{r_i} \quad i = 1, \ldots, k+1$ and $v_1 + \cdots + v_{k+1} = 0 \implies$ all $v_i = 0$.

Let $w_i = (\lambda_{k+1} I - A)^{r_{k+1}} v_i \implies w_{k+1} = 0$.

$$\sum_{i=1}^{k} w_i = \sum_{i=1}^{k+1} w_i = (\lambda_{k+1} I - A)^{r_{k+1}} \underbrace{\sum_{i=1}^{k+1} v_i}_{=0} = 0$$

$$(\lambda_i - A)^{r_i} w_i = (\lambda_i I - A)^{r_i} (\lambda_{k+1} I - A)^{r_{k+1}} v_i = (\lambda_{k+1} I - A)^{r_{k+1}} \underbrace{(\lambda_i I - A)^{r_i} \cdot v_i}_{=0} = 0$$

$$p(x) = (\lambda_i - x)^{r_i} \qquad q(x) = (\lambda_{k+1} - x)^{r_{k+1}}$$

$$\implies w_i \in \ker(\lambda_i I - A)^{r_i}$$

$$\xRightarrow{\text{induction hypothesis}} w_i = 0 \forall i$$

$$\implies v_i \in \ker(\lambda_{k+1} I - A)^{r_{k+1}}$$

$$v_i \in \ker(\lambda_i I - A)^{r_i}$$

$$\implies v_i \in \ker(\lambda_{k+1} I - A)^{r_{k+1}} \cap \ker(\lambda_i I - A)^{r_i} = \{0\}$$

$$\implies v_i = 0 \qquad \forall i = 1, \ldots, k$$

$$\implies 0 + \cdots + 0 + v_{k+1} = 0 \implies v_{k+1} = 0$$

□

Theorem 11.14. *Let $\lambda_1, \ldots, \lambda_k$ be pairwise different eigenvalues of $A \in \mathbb{K}^{n \times n}$. Let $W := \bigcap_{i=1}^{k} \mathrm{im}(\lambda_i I - A)^n$.*

1. $V = \ker(\lambda_1 I - A)^n \oplus \cdots \oplus \ker(\lambda_k I - A)^n \oplus \bigcap_{i=1}^k \operatorname{im}(\lambda_i I - A)^n$
2. *W is invariant under A and $\lambda_i \notin \operatorname{spec}(A|_W) \forall i \in \{1, \ldots, k\}$*

Proof. 1. Induction over k

Induction base $k = 1$:

$$V = \ker(\lambda_i I - A)^n \oplus \operatorname{im}(\lambda_i I - A)^n \qquad \text{(Fitting)}$$

Induction step $k \to k+1$: We assume:

$$V = \ker(\lambda_i I - A)^n \oplus \cdots \oplus \ker(\lambda_k I - A)^k \oplus W_k$$

$$W_k = \bigcap_{i=1}^{k} \operatorname{im}(\lambda_i I - A)^n$$

W_k is invariant: $y \in W_k \overset{!}{\implies} A_y \in W_k$. Let $y \in W_k$. $\implies \forall i = 1, \ldots, k$: $\exists x_i : y = (\lambda i - A)^n x_i$.

$$\implies Ay = A \cdot (\lambda_i I - A)^n x_i = (\lambda_i I - A)^n \cdot A x_i \in \operatorname{im}(\lambda_i I - A)^n$$

For all $i = 1, \ldots, k$ it holds that

$$\implies Ay \in \bigcap_{i=1}^{k} \operatorname{im}(\lambda_i I - A)^n$$

$$p(x) = x \qquad q(x) = (\lambda_i - x)^n$$

Consider $g : W_k \to W_k$ with $x \mapsto Ax$ with $\dim(W_k) \leq n$.

$$\text{Fitting} \implies \ker(\lambda_{k+1} - g)^n \oplus \operatorname{im}(\lambda_{k+1} - g)^n = W_k$$

where $\operatorname{im}(\lambda_{k+1} - g)^n \subseteq \operatorname{im}(\lambda_{k+1} - A)^n$.

$$\begin{aligned}
&\subseteq \ker(\lambda_{k+1} I - A)^n + (\operatorname{im}(\lambda_{k+1} - A)^n) \cap W_k \\
&= \ker(\lambda_{k+1} I - A)^n + \bigcap_{i=1}^{k+1} \operatorname{im}(\lambda_i I - A)^n
\end{aligned}$$

$$\implies V = \ker(\lambda_1 I - A)^n + \cdots + \ker(\lambda_k I - A)^n + \ker(\lambda_{k+1} I - A)^n + W_{k+1}$$

Claim: This sum is direct.
Let $x_i \in \ker(\lambda_i I - A)^n$ and $i = 1, \ldots, k+1$. Let $y \in W_{k+1} = \bigcap_{i=1}^{k+1} \operatorname{im}(\lambda_i I - A)^{r_i}$. Show that all $x_i = 0$ and $y = 0$. Thus $\sum_{i=1}^{k+1} x_i + y = 0$.

$$0 = \prod_{i=1}^{k+1} (\lambda_i I - A)^n \left(\sum_{i=1}^{k+1} x_i + y \right) = \sum_{i=1}^{k+1} 0 + \prod_{i=1}^{k+1} (\lambda_i I - A)^n \cdot y$$

$$\implies y \in \ker \prod_{i=1}^{k+1} (\lambda_i I - A)^n \cap \bigcap_{i=1}^{k+1} \operatorname{im}(\lambda_i I - A)^n \xRightarrow{\text{Lemma 11.11}} y = 0$$

$$\implies \sum_{i=1}^{k+1} x_i = 0 \implies x_i \text{ are linear independent}$$
$$\implies x_i = 0 \forall i \qquad \implies \text{sum is direct}$$

2. W_k is invariant, see proof of part (1)

$$\ker(\lambda_i - A) \cap W_k \subseteq \ker(\lambda_i - A)^n \cap \{0\}$$

$\implies$ no eigenvector for λ_1 in W_k.

□

↓ *This lecture took place on 2018/05/30.*

The sum of generalized spaces $\ker(\lambda_1 I - A)^n + \dots + \ker(\lambda_k I - A)^n + W$ is direct. The generalized spaces are invariant and also $W = \bigcap_{i=1}^k \operatorname{im}(\lambda_i I - A)^n$ and the restriction $A|_W$ has no λ_i as eigenvalue.

Let B_0 be a basis of W. Let B_i be a basis of $\ker(\lambda_i I - A)^n$. Then $B := B_1 \cup B_2 \cup \dots \cup B_k \cup B_0$ is a basis and in this basis,

$$\Phi_B^B(A) = \begin{bmatrix} [B_1] & & & & \\ & [B_2] & & & \\ & & \ddots & & \\ & & & [B_k] & \\ & & & & [B_0] \end{bmatrix}$$

If $x \in \mathcal{L}(B_i) \implies Ax \in \mathcal{L}(B_i)$. By invariance, $\Phi_B^B(A)$ has block diagonal form.

$$\ker(\lambda_i I - A)^n = \mathcal{L}(B_i)$$
$$\implies (\lambda_i I - A)^n \big|_{\mathcal{L}(B_i)} = 0 \implies \text{nilpotent}$$

Theorem 11.15. *Let $\mathbb{K}$ be algebraically closed (hence, every matrix has eigenvalues) and let $\lambda_1, \dots, \lambda_k$ be all eigenvalues of a matrix $A \in \mathbb{K}^{n \times n}$.*

$$\implies \mathbb{K}^n = \ker(\lambda_1 I - A)^n \oplus \dots \oplus \ker(\lambda_k I - A)^n$$

Proof.

$$\mathbb{K}^n = \ker(\lambda_1 I - A)^n \oplus \dots \oplus \ker(\lambda_n I - A)^n \oplus W$$
$$W = \bigcap_i \operatorname{im}(\lambda_i I - A)^n$$

$A|_W$ has no eigenvalue (because eigenvalue of $A|_W$ are also eigenvalues of A, but none of λ_i is an eigenvalue of $A|_W$), otherwise the sum is not direct. $\implies$ W is trivial ($W = \{0\}$). □

11.4 Nilpotent matrix

Definition 11.16. *A matrix/linear map* $f : V \to V$ *is called* nilpotent*, if* $\exists k \in \mathbb{N} : f^k = 0$. *The smallest k is called* index *of* f.

$$(\lambda_i I - A)|_{i\text{-th generalized space}} \text{ is nilpotent}$$

Goal: Structure of nilpotent matrices:

$$\begin{bmatrix} 0 & * & \ddots & 0 \\ & \ddots & \ddots & \\ & \ddots & \ddots & * \\ 0 & & & 0 \end{bmatrix}$$

Lemma 11.17. *Let* $\ker(f^m) \subseteq \ker(f^{m+1}) \subseteq \ker(f^{m+2})$

$$\begin{aligned} u_1 \dots u_p \dots &\ \textit{basis of } \ker f^n \\ u_1 \dots u_p v_1 \dots v_k \dots &\ \textit{basis of } \ker f^{m+1} \\ u_1 \dots u_p v_1 \dots v_k w_1 \dots w_r \dots &\ \textit{basis of } \ker f^{m+2} \end{aligned}$$

Then $(u_1 \dots u_p, f(w_1), \dots, f(w_r))$ *is linear independent.*

Proof. Immediate: $f(w_i) \in \ker f^{m+1}$, thus $f(\ker f^{m+2}) \subseteq \ker f^{m+1}$.

Show that: $\sum_{i=1}^p \lambda_i u_i + \sum_{j=1}^r \mu_j f(w_j) = 0 \overset{!}{\implies}$ all $\lambda_i = 0, \mu_j = 0$.

$$\implies \underbrace{f^m(\dots)}_{=\sum_{j=1}^p \lambda_i \underbrace{f^m(u_i)}_{=0} + \sum_{i=1}^r \mu_j f^{m+1}(w_j) = 0} = 0$$

$$\implies \sum_{i=1}^r \mu_j w_j \in \ker f^{m+1}$$

but $\ker f^{m+2} = \ker f^{m+1} \oplus \underbrace{\mathcal{L}(w_1, \dots, w_r)}_{w_j \in \mathcal{L}(w_1, \dots, w_r)}$. Hence, $\ker f^{m+1} \cap \mathcal{L}(w_1, \dots, w_r) = \{0\}$.

$$\implies \sum_{i=1}^r \mu_j w_j = 0 \xRightarrow{w_j \text{ linear indep.}} \text{all } \mu_j = 0$$

$$\implies \sum_{i=1}^p \mu_i u_i = 0 \xRightarrow{u_i \text{ linear indep.}} \text{all } \lambda_i = 0$$

□

11.5 Jordan's normal form

Theorem 11.18. Jordan's normal form *is a nilpotent matrix. Let* $\dim V = n$. $f : V \to V$ *is nilpotent of index p* ($f^p = 0$). $d = \dim\ker f$. *Then there exists a basis B of V such that*

$$\Phi_B^B(f) = \begin{bmatrix} [N_1] & & & \\ & [N_2] & & \\ & & \ddots & \\ & & & [N_d] \end{bmatrix}$$

where

$$N_i = \begin{bmatrix} 0 & 1 & \ddots & 0 \\ & 0 & 1 & \\ & & \ddots & 1 \\ 0 & & \ddots & 0 \end{bmatrix}_{n_i \times n_i}$$

$$p = n_1 \geq n_2 \geq \cdots \geq n_d \geq 1$$

$$n_1 + \cdots + n_d = n$$

Proof. Let $U_k = \ker f^k$, $\dim U_k = m_k$. $U_1 \subseteq U_2 \subseteq \cdots \subseteq U_p = V$. $d = m_1 \leq m_2 \leq m_3 \leq \cdots \leq m_p = n$.

$$f(U_i) \subseteq U_{i-1}$$

$$\underbrace{[\underbrace{[\underbrace{[u_1 \dots u_{m_1}]}_{U_1} u_{m_1+1} \dots u_{m_2}]}_{U_2} \dots u_{m_{p-1}+1} \dots U_{m_p}]}_{U_p}$$

$u_1 \dots u_{m_k}$ is basis of U_k.

We start from behind:

$$\ker f^{p-2} \leq \ker f^{p-1} \leq \ker f^p$$

We apply Lemma 11.17.

$$u_1 \dots u_{m_{p-2}} | u_{m_{p-2}+1} \dots u_{m_{p-1}} | u_{m_{p-1}+1} \dots u_{m_p}$$

$$v_1^{(p)} := u_{m_{p-1}+1} \qquad v_2^{(p)} = u_{m_{p-1}+2} \dots v_{m_p - m_{p-1}}^{(p)} := u_{m_p}$$

is basis of $U_p \ominus U_{p-1}$.

$$v_1^{(p+1)} = f(v_1^{(p)}) \qquad v_2^{(p-1)} = f(v_2^{(r)}) \ddots v_{m_p-m_{p-1}}^{(p-1)} \in U_{p-1} \underbrace{\ominus}_{(*)} U_{p-2}$$

(*) by Lemma 11.17 $f(v_j^{(p)})$ linear independent of $u_1 \dots u_{m_{p-2}}$.

And these $v_i^{(p-1)}$ are linear independent of $u_1 \dots u_{m_{p-2}}$. Extend $u_1 \dots u_{m_{p-2}} v_1^{(p-1)} \dots v_{m_p-m_{p-1}}^{(p-1)}$ to basis of U_{p-1}: $v_{m_p-m_{p-1}+1}^{(p+1)} \dots v_{m_{p-1}-m_{p-2}}$ chosen from $u_{m_{p-2}+1} \dots u_{m_{p-1}}$.

$$m_{p-2} + \cdots + m_{p-1} - m_{p-2} = m_{p-1}$$

$$u_1 \dots u_{m_{p-2}} | u_{m_{p-2}+1} \dots u_{m_{p-1}} | u_{m_{p-1}+1} \dots u_{m_p}$$

$$\underbrace{\underbrace{\underbrace{u_1 \dots u_{m_{p-2}}}_{U_{p-2}} v_1^{(p-1)} \dots v_{m_{p-1}-m_{p-2}}^{(p-1)}}_{f(n_{m_{p-1}+1}) \dots f(u_{m_p}) U_{p-1}} u_{m_{p-1}+1} \dots u_{m_p}}_{U_p}$$

where $u_{m_{p-1}+1} = v_1^{(p)} \dots u_{m_p} = v_{m_p - m_p - 1}^{(p)}$.

Iteration:

$$\begin{aligned} v_1^{(p-2)} &= f(v_1^{(p-1)}) \in U_{p-2} \ominus U_{p-3} \\ v_2^{(p-2)} &= f(v_2^{(p-1)}) \\ &\vdots \\ v_{m_{p-1}-m_{p-2}}^{(p-2)} &= f\left(v_{m_{p-1}-m_{p-2}}^{(p-1)}\right) \\ \rightsquigarrow u_1 \dots u_{m_{p-3}} v_1^{(p-2)} &\dots v_{m_{p-1}-m_{p-2}}^{(p-2)} \subseteq U_{p-2} \end{aligned}$$

are linear independent. $\to$ extend to basis of U_{p-2}:

$$u_1 \dots u_{m_{p-3}} v_1^{(p-2)} \dots v_{m_{p-2}-m_{p-3}}^{(p-2)}$$

and so on and so forth.

In the end, we get a basis:

$$\begin{array}{cccccccc} v_1^{(p)} & v_2^{(p)} & \dots & v_{m_p-m_{p-1}}^{(p)} & & & & \\ v_1^{(p-1)} & v_2^{(p-1)} & \dots & v_{m_p-m_{p-1}}^{(p-1)} & \dots & v_{m_{p-1}-m_{p-2}}^{(p-1)} & & \\ v_1^{(p-2)} & v_2^{(p-2)} & \dots & v_{m_p-m_{p-1}}^{(p-2)} & \dots & v_{m_{p-1}-m_{p-2}}^{(p-2)} & \dots & v_{m_{p-2}-m_{p-3}}^{(p-2)} \\ \vdots & & & & & & & \\ v_1^{(1)} & v_2^{(1)} & & & & & & v_{m_1}^{(1)} \end{array}$$

where each successive row can be reached by applying f. The last row represents the basis of U_1, all rows give the basis of U_p.

1. The last row is basis of U_1
2. f maps k-th row to $k-1$-th column.

$$B = \begin{bmatrix} \vdots & & \\ \vdots & \vdots & \\ \vdots & \vdots & \vdots \end{bmatrix}$$

$$B = V_1^{(1)} v_n^{(2)} \dots v_1^{(p)} v_2^{(2)} v_2^{(2)} \dots v_2^{(p)} \dots v_k^{(1)} v_k^{(2)} \dots v_k^{(p-1)} \dots v_{M_i}$$

$$B = v_1^{(i)} \dots v_1^{(p)} v_2^{(i)} \dots v_2^{(n_2)} v_3^{(1)} \dots v_3^{(n_3)} \dots v_d^{(1)} \dots v_d^{(n_d)}$$

$$n_3 \leq n_2 \leq n_1$$

$$f(v_i^{(i)}) = 0 \forall i = 1, \dots, d$$

$$f(v_i^{(2)}) = v_i^{(1)} \qquad f(v_i^{(3)}) = v_1^{(2)}$$

$$\Phi_B^B(f) = \begin{bmatrix} 0 & 1 & & & & \vdots & & \\ & 0 & 1 & & & \vdots & & \\ & & \ddots & \ddots & & \vdots & & \\ & & & & 0 & \vdots & & \\ \vdots & & & & 0 & 1 & & \\ \vdots & & & & & 0 & 1 & \\ \vdots & & & & & & \ddots & \\ \vdots & & & & & & & 1 \\ \vdots & & & & & & & 0 \\ 0 & 0 & \dots & 0 & & & & \end{bmatrix}$$

where this matrix goes on with these block matrices in the diagonal from n_1 to n_d. $\square$

Example 11.19.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & -2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 & 0 & -4 & 0 & 0 \end{bmatrix} \rightsquigarrow \left[\begin{array}{cccccccc} 0 & 0 & 3 & 0 & -2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 3 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$x_2 = -x_6 \qquad x_5 = 0 \qquad x_7 = 2x_3 = 0 \qquad x_3 = 0$$

Bases of $\ker(A) = e_1, e_4, e_8, -e_2 + e_6 =: \{u_1, u_2, u_3, u_4\}$.

$$\ker A = \ker N_1, N_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \textit{ are pivot rows}$$

$$\ker A^2 = \ker N_1 \cdot A$$

$$\textit{because}: A^2 x = 0 \iff Ax \in \ker A \iff Ax \in \ker N_1 \iff N_1 \cdot Ax = 0$$

$$\ker(N_1 \cdot AU) = \ker \begin{bmatrix} 0 & 0 & 1 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\ker A^2 : x_3 = 2x_5$$

$$\textit{Basis of } \ker A^2 : e_1, e_2, e_4, e_6, e_7, e_8, 2e_3 + e_5$$

$$\begin{array}{cccc} u_1 u_2 & u_3 u_4 & u_5 u_6 & u_7 \\ e_1 e_4 & e_8 - e_2 + e_6 & e_2 e_7 & 2e_3 + e_5 \end{array}$$

$$\textit{Basis of } U_2 \qquad m_2 = 7$$

$$A^3 = 0$$
$$U_3 = \ker A^3 = \mathbb{R}^8$$

Basis of U_3.

$$\begin{array}{cccc} u_1 u_2 & u_3 u_4 & u_5 u_6 & u_7 u_8 \\ e_1 e_4 & e_8 - e_2 + e_6 & e_2 e_7 & 2e_3 + e_5 e_3 \end{array}$$

$$p = 3 \qquad d = 4$$

$\rightarrow$ *4 blocks,* $n_i \leq 3$.

$$A \cdot v_1^{(3)} = A \cdot e_3 = 3e_2 - 2e_6$$

$$\begin{array}{cccc} u_1 u_2 & u_3 u_4 & u_1^{(2)} u_2^{(2)} & v_3^{(3)} v_1^{(3)} \\ e_1 e_4 & e_8 - e_2 + e_6 & 3e_2 - 2e_6 e_7 & 2e_3 + e_5 e_3 \end{array}$$

$$v_1^{(3)} = v_8 = e_3$$

$$v_1^{(2)} = A \cdot v_1^{(3)} = 3e_2 - 2e_6 \qquad v_2^{(2)} = e_7 \qquad v_3^{(2)} = 2e_3 + e_5$$

$$v_1^{(1)} = Av_1^{(2)} = e_1 + 3e_4 - 4e_8 \qquad v_2^{(1)} = -e_2 + e_6 \qquad v_3^{(1)} = 4e_2 + e_4 - 4e_6 \qquad v_4^{(1)} = e_1$$

$$v_1^{(1)} = A \cdot v_1^{(2)} = A \cdot \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \\ -4 \end{bmatrix}$$

$$v_2^{(1)} = A \cdot v_2^{(2)} = Ae_7 = -e_2 + e_6$$

$$v_3^{(2)} = A \cdot \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 0 \\ 1 \\ 0 \\ -4 \\ 0 \\ 0 \end{bmatrix}$$

	u_1	u_2	u_3	u_4
step 2	e_1	e_4	e_8	$-e_2+e_6$
step 3	$e_1+3e_4-4e_8$	$-e_2+e_6$	$4e_2+e_4-4e_6$	e_1

	u_5	u_6	u_7	u_8
step 2	$3e_2-2e_6$	e_7	$2e_3+e_5$	e_3
step 3	$3e_2-2e_6$	e_7	$2e_3+e_5$	e_3

$$B = \underbrace{e_1+3e_4-4e_8, 3e_2-2e_6, e_3}_{n_1=3} \mid \underbrace{-e_2+e_6, e_7}_{n_2=2} \mid \underbrace{4e_2+e_4-4e_6, 2e_3+e_5}_{n_3=2} \mid \underbrace{e_1}_{n_4=1}$$

$$\Phi_B^B(A) = \begin{bmatrix} 0 & 1 & & & & & & \\ & 0 & 1 & & & & & \\ & & & 0 & 1 & & & \\ & & & 0 & 0 & & & \\ & & & & & 0 & 1 & \\ & & & & & 0 & 0 & \\ & & & & & & & 0 \end{bmatrix} \textit{ is the Jordan norm form of } A$$

> ↓ *This lecture took place on 2018/06/04.*

$$A \text{ nilpotent} \implies A^k = 0$$

$\exists$ basis B such that

$$B^{-1}AB = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \end{bmatrix}$$

In general: Decomposition in generalized spaces:

$$U_i = \ker(\lambda_i I - A)^n \qquad V = \bigoplus_{i=1}^{l} U_i$$

$$(\lambda_i I - A)|_{U_i} \text{ is nilpotent}$$

$\rightsquigarrow$ basis B_i such that

$$(\lambda_i I - A)|_{U_i} = \begin{bmatrix} 0 & 1 & & & & \\ & & & & & \\ & 0 & \ddots & & & \\ & & 0 & 0 & & \\ & & & 0 & 1 & \\ & & & & \ddots & \ddots \end{bmatrix}$$

$$A|_{U_i} \sim \begin{bmatrix} \lambda_i & 0 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i & 1 \end{bmatrix}$$

11.6 Jordan block

Definition 11.20. *A matrix of form* $J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 0 & & & \\ & & \lambda & 1 & & \\ & & & \lambda & 1 & \\ & & & & \ddots & \ddots \end{bmatrix} \in \mathbb{K}^{n+m}$ *is called* Jordan block *of length k to eigenvalue* λ.

Remark 11.21. *1.* $\chi_{J_k(\lambda)}(x) = (x-\lambda)^k$

2. $J_k(\lambda) - \lambda \cdot I$ *is nilpotent with index k.*

Theorem 11.22. *Let* $\mathbb{K}$ *be an algebraically closed field (hence, every polynomial has roots). Then every matrix* $A \in \mathbb{K}^{n+m}$ *is similar to a matrix of Jordan normal form.*

$$\implies \exists B \in \mathrm{GL}(\mathbb{K}, n) : B^{-1}AB = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_n \end{bmatrix}$$

where every J_i *is a Jordan block to eigenvalue of* A.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the different eigenvalues of A. Let U_i be the generalized spaces.

$$U_i = \ker(\lambda_i I - A)^n$$

$$V = U_1 \oplus \cdots \oplus U_m$$

By Theorem 11.15 and the field is algebraically closed,

$$\implies U_i \text{ are invariant } \wedge (\lambda_i I - A)|_{U_i} \text{ is nilpotent}$$

By Theorem 11.18, $\exists$ basis B_i of U_i,

$$\Phi_{B_i}^{B_i}((\lambda_i I - A)|_{U_i}) = \begin{bmatrix} N_{n_{i_1}} & & & \\ & N_{n_{i_2}} & & \\ & & \ddots & \\ & & & N_{n_{i d_i}} \end{bmatrix}$$

with

$$N_k = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} = J_k(0)$$

where d_i is the geometric multiplicity of eigenvalue $\lambda_i = \dim\ker(\lambda_i I - A)$.

$$B = B_1 \cup B_2 \cup \cdots \cup B_n$$

$$\implies \Phi_B^B(A) = B^{-1}AB = \begin{bmatrix} J_{n_1 1}(\lambda_1) & & & & & & & \\ & J_{n_1 2}(\lambda_1) & & & & & & \\ & & \ddots & & & & & \\ & & & J_{n_1 d_1}(\lambda_1) & & & & \\ & & & & \ddots & & & \\ & & & & & J_{n_n 1}(\lambda_n) & & \\ & & & & & & \ddots & \\ & & & & & & & J_{n_n d_n}(\lambda_n) \end{bmatrix}$$

where columns of n_1 give U_1 and columns of n_n give U_n. □

Remark. *In essence, this can be proven because Φ_B^B is linear.*

Theorem 11.23. *Let* $B^{-1}AB = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix} \in \mathbb{K}^{n+m}$ *be a Jordan normal form with* $J_i = J_{k_i}(\lambda_i)$.

1. $\sum_{i=1}^q k_i = n$
2. $\operatorname{spec}(A) = \{\lambda_i\}$, *potentially with repetitions.*

$$\forall \lambda \in \operatorname{spec}(A): \qquad d(\lambda) = \#\{i : \lambda_i = \lambda\} \qquad k(\lambda) = \sum_{\lambda_i = \lambda} k_i$$

Geometric multiplicity of λ equals the number of corresponding Jordan blocks. Algebraic multiplicity of λ equals the num of sizes of corresponding Jordan blocks.

3. *The smallest exponent r such that*

$$\ker((\lambda I - A)^r) = \ker((\lambda I - A)^{r+1})$$

is the largest length of a corresponding Jordan block.

$$\min\left\{r : \ker((\lambda I - A)^r) = \ker(C, \dots)^{r+1}\right\} = \max\{k_i : \lambda_i = \lambda\}$$

The reason is given in an example:

Example.

$$A = \begin{bmatrix} J_{k_1}(\lambda) & \\ & J_{k_2}(\lambda) \end{bmatrix} \qquad A - \lambda I = \begin{bmatrix} J_{k_1}(0) & \\ & J_{k_2}(0) \end{bmatrix}$$

$$(A - \lambda I)^r = \begin{bmatrix} J_{k_1}(0)^r & \\ & J_{k_2}(0)^r \end{bmatrix} \stackrel{!}{=} 0$$

$$\implies J_{k_1}(0)^r = 0 \land J_{k_2}(0)^r = 0$$

$$\implies r \geq k_1 \land r \geq r_k \implies r = \max\{k_1, k_2\}$$

4. *Let $k \in \mathbb{N}$. $\# \{i : \lambda_i = \lambda \wedge k_i \geq k+1\} = \operatorname{rank}(\lambda I - A)^k - \operatorname{rank}(\lambda I - A)^{k+1}$*
5. *The Jordan blocks are uniquely determined (except for the order)*

Proof. 1. Immediate.

2. For every Jordan block, there exists exactly one eigenvector.

$$(\operatorname{rank}(J_k(\lambda) - \lambda I_k)) = k = 1$$

$$\chi_{J_{k_i}(\lambda_i)}(x) = (x-\lambda)^k$$

$$\implies \chi_A(x) = \prod_{i=1}^{q} \chi_{J_{k_i}(\lambda_i)}(x) = \prod_{i=1}^{q} (x-\lambda_i)^{k_i}$$

3. Let $A = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix}$.

$$(\lambda I - A)^k = \begin{bmatrix} (\lambda - J_1)^k & & \\ & \ddots & \\ & & (\lambda - J_q)^k \end{bmatrix}$$

If $\lambda \neq \lambda_i$, then $\lambda - J_k(\lambda_i)$ is invertible (triangular matrix with all entries on the main diagonal $\neq 0$). If $\lambda = \lambda_i$, then $\lambda - J_{k_i}(\lambda_i)$ is nilpotent.

$$\operatorname{rank}((\lambda - J_{k_i}(\lambda_i))^k) = \begin{cases} k_i - k & k_i > k \\ 0 & \text{else} \end{cases}$$

4.

$$\operatorname{rank}(\lambda_i - J_{k_i}(\lambda_i))^k - \operatorname{rank}(\lambda_i - J_{k_i}(\lambda_i))^{k+1} = \begin{cases} 1 & k_i \geq k+1 \\ 0 & k_i \leq k \end{cases}$$

$$\operatorname{rank}(\lambda - J_{k_i}(\lambda_i))^k = \begin{cases} k_i & \text{if } \lambda \neq \lambda_i \\ k_i - k & \text{if } \lambda = \lambda_i \text{ and } k_i \geq k \\ 0 & \text{if } \lambda = \lambda_i \wedge k_i < k \end{cases}$$

$$\operatorname{rank}(\lambda - A)^k = \sum_{\lambda \neq \lambda_j} k_j + \sum_{\lambda = \lambda_i} \begin{cases} k_i - k & \text{if } k_i \geq k \\ 0 & \text{if } k_i < k \end{cases}$$

$$\operatorname{rank}(\lambda - A)^{k+1} = \sum_{\lambda \neq \lambda_j} k_j + \sum_{\lambda = \lambda_i} \begin{cases} k_i - (k+1) & \text{if } k_i \geq k \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned} \operatorname{rank}(\lambda - A)^k - \operatorname{rank}(\lambda - A)^{k+1} &= \overbrace{0}^{\substack{\text{first} \\ \text{sum}}} + \sum_{\lambda = \lambda_i} (k_i - k) - (k_i - (k+1)) \\ &= \sum_{\lambda = \lambda_i} 1 \text{ if } k_i > k+1 \\ &= \{i : k_i > k+1\} \end{aligned}$$

5. Given some eigenvalues, is the Jordan normal form unique?

   The number of Jordan blocks is uniquely determined by $d(\lambda)$.

   By the previous item, $\forall k \in \mathbb{N}_0$ we know the number of Jordan blocks of size $k \geq 1$. This way, $\forall k \in \mathbb{N}_0$ we can determine the number of Jordan blocks of size k. See Wikipedia.

$\square$

Lemma 11.24. *Let $A \in \mathbb{K}^{n+n}$ matrix.*

$$\Psi_A : {}^{\mathbb{K}[x] \to \mathbb{K}^{n+n}}_{p(x) \mapsto p(A)}$$

$$a_0 + a_1 x + \dots + a_k x^k \mapsto a_0 \cdot I + a_1 A + \dots + a_k A^k$$

1. $p\left(\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}\right) = \begin{bmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{bmatrix}$
2. $p\left(\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{bmatrix}\right) = \begin{bmatrix} p(A_1) & & \\ & \ddots & \\ & & p(A_n) \end{bmatrix}$
3. $A = T^{-1}BT \implies p(A) = T^{-1}p(B)T$

Proof. Because of linearity, it suffices to show that this holds for basis polynomials x^k.

1. Immediate.
2. $\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_m \end{bmatrix}^k = \begin{bmatrix} A_1^k & & \\ & \ddots & \\ & & A_m^k \end{bmatrix}$
3. $$\begin{aligned} A^k &= (T^{-1}BT)^k \\ &= (T^{-1}BT)(T^{-1}BT)\dots(T^{-1}BT) \\ &= T^{-1}B^kT \end{aligned}$$

   $\implies p(A)$ can be reduced to $p(\text{JNF})$.

$\square$

Remark. *Recognize that a polynomial is a linear combination of basis elements. Some basis element might be non-linear like x^5, but this does not invalidate our proof.*

Lemma 11.25. *For some Jordan block $J_k(\lambda)$ it holds that*

$$p(J_k(\lambda))_{i,j} = \begin{cases} \frac{p^{(j-i)}(\lambda)}{(j-i)} & j > i \\ p(\lambda) & j = i \\ 0 & j < i \text{ (below the diagonal)} \end{cases}$$

Proof.

$$(A+B)^M = \sum_{k=0}^{M} \binom{M}{k} A^k B^{M-k}$$

In general, $(A+B)^2 = AA + AB + BA + BB$, but here $A = I$ and therefore $AB = BA$.

$$\begin{aligned}
(\lambda I + N)^M &= \sum_{k=0}^{M} \binom{M}{k} \lambda^{M-k} \cdot I \cdot N^k \\
&= \begin{bmatrix} \lambda^M & \binom{n}{1}\lambda^{n-1} & \\ & \ddots & \\ & & \lambda^M \end{bmatrix} \\
&[k = 0 \implies \operatorname{diag}(\lambda^M), N^1 = I] \\
&= \sum_{k=0}^{n} \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \cdot \lambda^{n-k} \cdot N^k = \sum \frac{(\lambda^M)^{(k)}}{k!}
\end{aligned}$$

□

Example 11.26 (Application).

1. *Fibonacci → Tribonacci:* $T_n = T_{n-1} + T_{n-2} + T_{n-3}$ *(see practicals)*
2. *Discrete dynamic systems: Predator-prey equations:*

   $$F_n := \text{ number of foxes} \qquad H_n := \text{ number of rabbits}$$

   When is F_n *and* H_n *stationary?*

   $$F_{n+1} = p \cdot F_n + q \cdot H_n$$

   $$H_{n+1} = -t \cdot F_n + q \cdot H_n$$

   $$\rightarrow \begin{pmatrix} F_{n+1} \\ H_{n+1} \end{pmatrix} = \begin{bmatrix} p & q \\ -t & g \end{bmatrix} \cdot \begin{pmatrix} F_n \\ H_n \end{pmatrix}$$

   When it is balanced? This depends on the matrix ...

   $$\begin{bmatrix} p & q \\ -t & g \end{bmatrix} \sim \begin{bmatrix} \lambda_1 & 1_v.0 \\ & \lambda_2 \end{bmatrix}^n \xrightarrow{n\to\infty} \begin{cases} 0 & |\lambda_1|, |\lambda_2| < 1 \\ \infty & |\lambda_1|, |\lambda_2| > 1 \\ \textit{balanced} & \textit{if } |\lambda_1| = 1 \\ \textit{dynamic balance} & \textit{if } |\lambda_1| = |\lambda_2| = 1 \end{cases}$$

   $$\vec{x} = A \cdot \vec{x}$$

   $$\frac{d}{dx} = A \cdot x \rightarrow \textit{ solution } x = e^{A \cdot t} \cdot x_0$$

11.7 Matrix exponentiation

Theorem 11.27 (Matrix exponential function)**.** *In general: A is diagonalizable.*

$$\implies A = T^{-1}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} T \qquad \to f(A) = T^{-1}\begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} T$$

If not diagonalizable: Jordan norm form.

$$\rightsquigarrow f(J_k(\lambda)) = \text{ ?}$$

- *For polynomials, immediate.*
- *Otherwise, only works for analytical functions. Hence,*

$$f(x) = \text{ Taylor series } = \sum_{k=0}^{\infty} a_k (x-\lambda)^k$$

$$f(J_k(\lambda)) = \sum_{k=0}^{\infty} a_k \underbrace{(J_k(\lambda) - \lambda)^k}_{\text{nilpotent}}$$

nilpotent $\implies$ *series escapes* $\implies$ *convergent.*

$$f(x) = e^x \implies e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

$$JNF \implies A = B \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_n \end{bmatrix} B^{-1}$$

We have to determine e^{J_i}, because $e^A = B \cdot \begin{bmatrix} e^{J_1} & & \\ & \ddots & \\ & & e^{J_k} \end{bmatrix} B^{-1}$.

$$\begin{aligned} J &= \lambda \cdot I + N \\ \to e^J &= e^{\lambda I + N} \\ &= e^{\lambda I} \cdot e^N && \text{because } \lambda I \text{ and } N \text{ commute} \\ &= e^{\lambda} \cdot \sum_{k=0}^{\infty} \frac{N^k}{k!} \\ &= e^{\lambda} \cdot \sum_{k=0}^{M-1} \frac{N^k}{k!} && \text{if } N \text{ is nilpotent} \\ &= e^{\lambda} \cdot \begin{bmatrix} 1 & \frac{1}{2} & \cdots \\ & \ddots & \ddots \\ & & \frac{1}{2} \\ & & 1 \end{bmatrix} \end{aligned}$$

↓ *This lecture took place on 2018/06/06.*

Example 11.28.

$$e^A \begin{cases} \frac{dx}{dt} = Ax \rightsquigarrow x(t) = e^{A \cdot t} x_0 \\ x(0) = x_0 \end{cases}$$

for $t \to \infty$: $e^{At}x_0$. Asymptotically, depends on the eigenvalue.

$$A = B^{-1} \underbrace{J}_{JNF} B \qquad e^{At} = B^{-1} e^{Jt} B = B^{-1} \begin{bmatrix} e^{J_1 t} & & & \\ & e^{J_2 t} & & \\ & & \ddots & \\ & & & e^{J_k t} \end{bmatrix} \cdot B$$

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix} \text{Jordan blocks} \qquad J_k = \begin{bmatrix} \lambda & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix}$$

$$e^{J_k t} = e^{\lambda t} \cdot e^{\overbrace{\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}}^{\text{polynomial}} t}$$

$$e^{\lambda t} = e^{(\xi + i\eta)\beta} \xrightarrow{t \to \infty} \begin{cases} \infty & \text{if } \xi > 0 \\ 0 & \text{if } \xi < 0 \end{cases}$$

$$e^{(\xi + i\eta)\beta} = e^{\xi t} \cdot e^{i\eta t}$$

with $\left|e^{i\eta t}\right| = 1$ and $\xi = \Re\lambda$. → if $\Re\lambda < 0 \forall$ eigenvalue λ_i.

$$e^{At}x_0 \xrightarrow{t \to \infty} 0$$

for arbitrary x_0. If $\Re\lambda_i > 0 \forall$ eigenvalue $\implies$ $e^{At}x_0 \xrightarrow{t \to \infty} \infty$.

If $\Re\lambda_i < 0$ for some specific λ_i and $\Re\lambda_i > 0$ for other λ_i. Asymptotically depends on the initial value x_0.

Example (Pendulum)**.**

$$m \cdot l \cdot \dot{\varphi} = -m \cdot g \cdot \sin\varphi \approx -g \cdot \varphi$$

$$l \cdot \ddot{\varphi} = -g \cdot \varphi = -\omega^2 \cdot \varphi$$

$$\psi = \frac{\dot{\varphi}}{\omega} \qquad \dot{\varphi} = \omega \cdot \psi$$

$$\dot{\psi} = \frac{\ddot{\varphi}}{\omega} = -\frac{\omega^2 \cdot \varphi}{\omega} = -\omega \cdot \varphi$$

$$\frac{d}{dt}\begin{pmatrix}\varphi \\ \psi\end{pmatrix} = \begin{pmatrix}\omega\psi \\ -\omega\varphi\end{pmatrix} = \begin{bmatrix}0 & \omega \\ -\omega & 0\end{bmatrix}\begin{pmatrix}\varphi \\ \psi\end{pmatrix}$$

$$\begin{pmatrix}\varphi(t) \\ \psi(t)\end{pmatrix} = e^{\begin{bmatrix}0 & \omega \\ -\omega & 0\end{bmatrix}t}\begin{bmatrix}\varphi_0 \\ \psi_0\end{bmatrix}$$

eigenvalue: $\lambda^2 + \omega^2 = 0, \omega = \pm i\omega$.

$$\varphi(t) \sim \Re(c \cdot e^{i\omega t} \sim a\cos\omega t + b \cdot \sin\omega t)$$

$$\omega = \sqrt{\frac{g}{l}}$$

Example.

$$\frac{\partial T(x,t)}{\partial t} = \triangle T(x,t)$$

$$\triangle T(x,t) = \frac{\partial^2}{\partial x^2} T(x,t)$$

$$\triangle = \sum \frac{\partial^2}{\partial x_i^2} \qquad \textit{Laplace operator}$$

$$T(x,t) = e^{t\triangle} T(x,0)$$

Example (Schrödinger's equation)**.**

$$-\frac{\hbar}{i}\frac{\partial\psi}{\partial t} = H\psi \rightsquigarrow \psi(t) = e^{-\frac{i}{\hbar}\triangle t}\psi_0$$

Hamilton:

$$H = \triangle + V(x)$$

11.8 Minimal polynomial and annihilator

Definition 11.29 (Theorem and definition)**.** *Let* $A \in \mathbb{K}^{N\times N}$.

1. $\exists p(x) \in \mathbb{K}[x] : p(A) = 0$
2. $\exists$ *a unique polynomial* $m_A(x) \in \mathbb{K}[x]$ *with minimal degree and leading coefficients* 1 *such that* $m_A(A) = 0$. $m_A(x)$ *is called* minimal polynomial of A.
3. $\mathrm{Ann}(A) = \left\{p(x) \in \mathbb{K}[x] \,\middle|\, p(A) = 0\right\}$ *is called* annihilator of A *and it holds that* $p(x) \in \mathrm{Ann}(A) \iff m_A(x) \,\middle|\, p(x)$
4. $m_A(\lambda) = 0 \forall \lambda \in \mathrm{spec}(A)$

Proof. 1. $A^0, A^1, A^2, A^3, \dots \in \mathbb{K}^{N\times N}$. Infinitely many elements of a finite dimensional vector space are linear dependent.

$$\implies \exists n \exists a_0, a_1, \dots, a_n : a_0 A^0 + \underbrace{aA}_{=p(A)} + \dots + a_n A^n = 0$$

$$p(x) = a_0 + a_1 x + \dots + a_n x^n$$

2. + 3. Let n be minimal such that $A^0, \dots, A^n$ are linear dependent ($\implies n \leq N^2$) and $a_0 I + a_1 A + \dots + a_n A^n = 0$ with $a_n \neq 0$ (this will be shown in the practicals).

$$\implies m_A(x) = \frac{a_0}{a_n} + \frac{a_1}{a_n} x^1 + \dots + \frac{a_{n-1}}{a_n} x^{n-1} + x^n$$

is the unique minimal polynomial.

Assume $p(x) \in \mathbb{K}[x]$ with $p(A) = 0 \implies \deg(p(x)) \geq n$.

By the division algorithm: $\exists q(x) \in \mathbb{K}[x] \exists r(x) \in \mathbb{K}[x]$ with $\deg(r(x)) < \deg(m_A(x))$ and $p(x) = q(x) m_A(x) + r(x)$. Insert A:

$$\begin{aligned} p(A) &= q(A) \cdot m_A(A) + r(A) \\ 0 &= 0 + r(A) \\ &\implies r(A) = 0 \\ \deg(r(x)) < n &\overset{\text{minimality } n}{\implies} r(x) \equiv 0 \\ &\implies p(x) = q(x) \cdot m_A(x) \\ &\implies m_A(x) \,\big|\, p(x) \implies (c) \end{aligned}$$

Especially, if $\deg(p(x)) = n = \deg(m_A(x)) \implies \deg(q(x)) = 0 \implies p(x) = c \cdot m_A(x)$.

4. Will be shown in the practicals.

□

11.9 Cayley-Hamilton Theorem

Theorem 11.30 (Cayley-Hamilton Theorem)**.**

$$\chi_A(A) = 0 \qquad (\iff \chi_A(x) \in \operatorname{Ann}(A))$$

Corollary 11.31.

$$m_A(x) \,\big|\, \chi_A(x)$$

and therefore the roots of $m_A(x)$ are the eigenvalues of A.

Proof. Three different proofs will be given:

1. $\chi_A(x) = \det(xI - A)$ and $\chi_A(A) = \det(AI - A) = 0$ (incorrect, used among physicists)

2. Using Jordan's normal form (if $\mathbb{K}$ is algebraically closed[10]):

$$A = B \cdot J \cdot B^{-1}$$

$$\chi_A(A) = B \cdot \chi_A(J) \cdot B^{-1}$$

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix} \rightsquigarrow \chi_A(J) = \begin{bmatrix} \chi_A(J_1) & & \\ & \ddots & \\ & & \chi_A(J_q) \end{bmatrix}$$

$$J_i = \begin{bmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_{n_i} \end{bmatrix}$$

We know that $\sum_{i,\lambda_i=\lambda} n_i = k(\lambda)$ algebraic multiplicity of eigenvalue λ. $\chi_A(J_i)$.

$$\chi_A(x) = \prod_j (\lambda - \lambda_j)^{kj}$$

$$\chi_A(J_i) = \prod_{j \neq i} (J_i - \lambda_j I)^{kj} \cdot \underbrace{(J_i - \lambda_i I)^{ki}}_{\underbrace{\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}^{k_i}}_{n_i \leq k_i} = 0}$$

$\chi_A(j) = 0$ for all Jordan blocks of A.

$$\chi_A(A) = B \begin{bmatrix} \chi_A(J_i) & & \\ & \ddots & \\ & & \chi_A(J_q) \end{bmatrix} B^{-1} = 0$$

3. Complementary matrix:

$$A \cdot \widehat{A} = \det(A) \cdot I$$

$$\widehat{A}_{ij} = -(-1)^{i+j} \det(A_{ji})$$

where A_{ji} denotes removing the j-th row and i-th column.

$$\widehat{xI - A} \in \mathbb{K}[x]^{n \times n} = [b_{ij}(x)]_{i,j=1,\ldots,n}$$

$$b_{ij}(x) = (-1)^{i+j} \det(xI - A)_{ji} \in \mathbb{K}[x] \text{ with degree } \leq n-1$$

$$b_{ij}(x) = \sum_{k=0}^{n-1} b_{ijk} x^k$$

$$\widehat{xI - A} = \left[\sum_{k=0}^{n-1} b_{ijk} x^k \right]_{i,j=1,\ldots,n} = \sum_{k=0}^{n-1} [b_{ijk}]_{i,j=1,\ldots,n} \cdot x^k = \sum_{k=0}^{n-1} B_k \cdot x^k$$

[10] And therefore the third proof is required.

Example.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\begin{aligned}
\widehat{xI - A} &= \widehat{\begin{bmatrix} x-a & -b \\ -c & x-d \end{bmatrix}} \\
&= \begin{bmatrix} x-d & b \\ c & x-a \end{bmatrix} \in \mathbb{K}[x]^{2\times 2} \qquad \textit{matrix with polynomial entries} \\
&= \underbrace{\begin{bmatrix} -d & b \\ c & -a \end{bmatrix}}_{B_0} + \underbrace{\begin{bmatrix} 1 & \\ & 1 \end{bmatrix} x}_{B_1 x} \in \mathbb{K}^{2\times 2}[x]
\end{aligned}$$

$$\begin{aligned}
(xI - A)(\widehat{xI - A}) &= \det(xI - A) \cdot I \\
&= \chi_A(x) \cdot I
\end{aligned}$$

Let $\chi_A(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1} + x^n$.

$\implies (xI-A)(B_0+B_1x+B_2x^2+\cdots+B_{n-1}x^{n-1}) = (c_0+c_1x+\cdots+c_{n-1}x^{n-1}+x^n)I$

$-AB_0+(B_0-AB_1)x+(B_1-AB_2)x^2+\cdots+(B_{n-2}-AB_{n-1})x^{n-1}+B_{n-1}x^n = c_0I+c_1Ix+\cdots+I\cdot x^n$

We apply coefficient comparison (multiply with A^k *from left):*

$$\begin{aligned}
c_0 \cdot I &= -AB_0 & c_0 I &= -AB_0 \\
c_1 \cdot I &= B_0 - AB_0 & +c_1 A &= AB_0 - A^2 B_1 \\
c_2 \cdot I &= B_1 - AB_2 & +c_2 A^2 &= A^2 B_1 - A^3 B_2 \\
\vdots &= \vdots \\
c_{n-1} \cdot I &= B_{n-2} - AB_{n-1} & +c_{n-1} A^{n-1} &= A^{n-1} B_{n-2} - A^n B_{n-1} \\
I &= B_{n-1} & +A^n &= A^n B_{n-1}
\end{aligned}$$

$$\chi_A(A) = \sum_{k=0}^{n-1} c_k A^k + A^n = -AB_0 + \sum_{k=1}^{n-1} (A^k B_{k-1} - A^{k+1} B_k) + A^k B_{n-1} = 0$$

Thus this proof has proven it for every zero-divisor-free field.

□

Corollary 11.32 (Corollary for second proof)**.**

1. *The minimal polynomial has the structure* $m_A(x) = \prod(\lambda - \lambda_i)^{m_i}$ *where* m_i *is the smallest exponent for* $\ker(\lambda_i - A)^m = \ker(\lambda_i - A)^{m+1}$*, hence this equals the largest length of a Jordan block for eigenvalue* λ_i*.*

2. *A is diagonalizable* $\iff$ *all* $m_i = 1 \iff m_A(x) = \prod_{i=1}^{k}(\lambda - \lambda_i) \iff m_A(x)$ *has only simple roots.*

Example 11.33 (Application). *Let* $A \in \mathbb{K}^{2\times 2}$. *We consider* $A \in \mathbb{C}^{2\times 2}$.

$$e^{\alpha I + A} = e^{\alpha} \cdot e^{A}$$

Without loss of generality: $\mathrm{Tr}(A) = 0$. *Otherwise consider* $\mathring{A} = A - \frac{\mathrm{Tr}(A)}{2} \cdot I$.

$$\implies \mathrm{Tr}(\mathring{A}) = \mathrm{Tr}(A) - \frac{\mathrm{Tr}(A)}{2} \cdot \mathrm{Tr}(I) = 0$$

$$e^{A} = e^{\frac{\mathrm{Tr}(A)}{2}} \cdot e^{\mathring{A}}$$

Without loss of generality: $A = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}$.

$$\begin{aligned} \chi_A(A) &= (X-a)(X+a) - bc \\ &= x^2 - a^2 - bc \\ &= x^2 - \delta \end{aligned}$$

$$\delta = a^2 + bc = -\det(A)$$

$$e^{A} = \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

Cayley-Hamilton Theorem:

$$\chi_A(A) = 0$$

$$A^2 - \delta I = 0 \implies A^2 = \delta I, A^3 = \delta A, A^4 = (A^2)^2 = \delta^2 \cdot I$$

$$\begin{aligned} A^{2n} &= \delta^n \cdot I \\ A^{2n+1} &= \delta^n \cdot A \end{aligned}$$

$$\begin{aligned} e^{A} = \sum_{n=0}^{\infty} \frac{A^n}{n!} &= \sum_{k=0}^{\infty} \frac{A^{2k}}{2k!} + \sum_{k=0}^{\infty} \frac{A^{2k+1}}{(2k+1)!} \\ &= \sum_{k=0}^{\infty} \frac{\delta^k}{2k!} I + \sum_{k=0}^{\infty} \frac{\delta^k}{(2k+1)!} \cdot A \\ &= \cosh\left(\sqrt{\delta}\right) + \sinh(\sqrt{\delta}) \cdot A \end{aligned}$$

$$Ax = \lambda x \qquad A^2 x = \lambda^2 x \qquad \dots \qquad A^k x = \lambda^k x$$

$$p(A) \cdot x = p(\lambda) \cdot x \qquad \text{if } \lambda \in \mathrm{spec}(A) \implies p(\lambda) \in \mathrm{spec}(p(A))$$

11.10 Spectral mapping theorem

Theorem 11.34 (Spectral mapping theorem). *For* $A \in \mathbb{K}^{n\times n}$ *and* $\mathbb{K}$ *be algebraically closed and* $p(x) \in \mathbb{K}[x]$ *is* $\mathrm{spec}(p(A)) = p(\mathrm{spec}(A)) =: \left\{p(\lambda) \,\middle|\, \lambda \in \mathrm{spec}(A)\right\}$.

Proof. **Direction $\supseteq$:**

$$\forall \lambda \in \mathrm{spec}(A) : p(\lambda) \in \mathrm{spec}(p(A))$$

Direction $\subseteq$:

$$\forall \mu \in \operatorname{spec}(p(A)) : \exists \lambda \in \operatorname{spec}(A) : p(\lambda) = \mu$$

$\square$

Example 11.35. *If $\mathbb{K}$ is not algebraically closed, then $\subseteq$ does not hold! For example, consider* $A = \begin{bmatrix} & 1 \\ -1 & \end{bmatrix} \in \mathbb{R}^{2\times 2}$.

$$\operatorname{spec}(A) = \{\pm i\} \implies \operatorname{spec}_{\mathbb{R}}(A) = \emptyset$$

$$p(x) = x^2 \implies A^2 = -I$$

has eigenvalue $\{-1\}$ *but there exists no* $\lambda \in \operatorname{spec}_{\mathbb{R}}(A)$ *such that* $\lambda^2 = -1$.

Let $\mu \in \operatorname{spec}(p(A))$.

$$q(x) = p(x) - \mu = (x - \mu_1)\dots(x - \mu_m)$$

where μ_i are the roots of $q(x)$.

$$\implies q(A) = p(A) - \mu I \text{ is not invertible}$$

$$\begin{aligned} q(A) = (A - \mu_i I)(A - \mu_2 I)\dots(A - \mu_m I) &\text{ not invertible} \\ \implies \exists i : (A - \mu_i I) &\text{ not invertible} \\ \implies \mu_i \in \operatorname{spec}(A) \\ \implies q(\mu_i) = 0 \\ q(\mu_i) = p(\mu_i) - \mu \\ \implies \mu = p(\mu_i) &\text{ and } \mu_i \in \operatorname{spec}(A) \end{aligned}$$

↓ *This lecture took place on 2018/06/11.*

12 Normal matrices

Definition 12.1 (Revision: Inner product)**.** *$\langle x, y\rangle$ is sesquilinear where the first argument is linear and the second argument is antilinear.*

$$\begin{aligned} \langle \lambda x + \mu y, z\rangle &= \lambda \langle x, z\rangle + \mu \langle y, z\rangle \\ \langle x, y\rangle &= \overline{\langle y, x\rangle} \\ \langle x, x\rangle &\geq 0 \text{ or equivalently, } \langle x, x\rangle = 0 \iff x = 0 \end{aligned}$$

Adjoint map: $f : V \to W$ where B is the basis of V and C is the basis of W.

$$\langle f(x), y\rangle = \langle x, f^*(y)\rangle$$

f^* is linear: $W \to V$. $f \mapsto f^*$ is antilinear.

$$\Phi_C^B(f) \text{ is its matrix representation}$$

$$\Phi_B^C(f^*) = \Phi_C^B(f)^*$$
$$(A^*)_{ij} = \overline{a_{ji}}$$

Definition 12.2. *A linear map $f : V \to V$ (and accordingly a matrix A) is called* normal *if*

$$f \circ f^* = f^* \circ f \qquad AA^* = A^*A$$

Special case: $A = A^*$ is self-adjoint.

A real-valued self-adjoint matrix is called symmetrical*: $A = A^T$.*

Example 12.3. *Unitary matrices are normal: $U^*U = I \implies UU^* = I$. A, B normal $\implies A \cdot B$.*

$$(AB)^* \cdot AB \stackrel{?}{=} AB(AB)^* = ABB^*A^*$$
$$= (AB)^* \cdot AB = B^*A^* \cdot AB \text{ only if } AB = BA$$

Example. *$A + B$ is normal if and only if $AB = -BA$.*

$$A, B \text{ self-adjoint} \implies A + B \text{ self-adjoint}$$

$$(A + B)^* = A^* + B^* = A + B$$

Lemma 12.4. *$A \in \mathbb{C}^{n\times n}$ is normal.*

1. $\ker A = \ker A^*$
2. $\ker A = \ker A^2$

Corollary 12.5. *$A \in \mathbb{C}^{n\times n}$ is normal. (So $\lambda I + A$ is normal)*

1. $\ker(\lambda I - A) = \ker(\overline{\lambda} I - A^*)$
2. $\ker(\lambda I - A)^2 = \ker(\lambda I - A)$

$$\implies \text{generalized spaces} = \text{eigenspaces} \qquad \implies \text{diagonalizable}$$

Proof. 1.

$$\begin{aligned}
x \in \ker A &\iff Ax = 0 \\
&\iff \|Ax\|^2 = 0 \\
&\iff \langle Ax, Ax\rangle = 0 \\
&\iff \langle x, A^*Ax\rangle = 0 \\
&\iff \langle x, AA^*x\rangle = 0 \\
&\iff \langle A^*x, A^*x\rangle = 0 \\
&\iff \|A^*x\|^2 = 0 \\
&\iff \|A^*x\| = 0 \\
&\iff A^*x = 0 \\
&\iff x \in \ker A^*
\end{aligned}$$

2. $\ker A \subseteq \ker A^2$ immediate (Fitting)

$$\begin{aligned}
\text{Let } x \in \ker A^2 &\implies A^2x = 0 \\
&\implies Ax \in \ker A \\
&\implies Ax \in \ker A^* \\
&\implies A^*Ax = 0 \\
&\implies \langle A^*Ax, x\rangle = 0 \\
&\implies \langle Ax, Ax\rangle = 0 \\
&\implies \|Ax\|^2 = 0 \\
&\implies Ax = 0 \\
&\implies x \in \ker A
\end{aligned}$$

□

Lemma 12.6. *Let A be normal.*

$$\lambda \neq \mu \in \operatorname{spec} A \implies \ker(\lambda I - A) \perp \ker(\mu I - A)$$

Proof. Let $Ax = \lambda x$ and $Ay = \mu y$

$$\xRightarrow{\textit{Corollary } 12.5} A^*y = \overline{\mu}y$$

$$\begin{aligned}
\langle Ax, y\rangle &= \langle \lambda x, y\rangle = \lambda \langle x, y\rangle \\
\langle Ax, y\rangle &= \langle x, A^*y\rangle = \langle x, \overline{\mu}y\rangle = \mu \langle x, y\rangle \\
\implies \lambda \langle x, y\rangle &= \mu \langle x, y\rangle \\
\implies \underbrace{(\lambda - \mu)}_{\neq 0} \langle x, y\rangle &= 0 \implies \langle x, y\rangle = 0
\end{aligned}$$

□

Remark (Summary). *Let A be normal. Then generalized spaces are eigenspaces. So they are diagonalizable. Then there exists some basis of eigenvectors. Eigenspaces are always orthogonal, so there exists an orthogonal basis of eigenvectors.*

Theorem 12.7. *For $A \in \mathbb{C}^{n\times n}$, the following statements are equivalent:*

1. *A is normal.*
2. *$\exists$ orthonormal basis of eigenvectors*
3. *$\exists$ unitary matrix U such that $U^*AU = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, hence A is unitarily diagonalizable.*

Proof. 1. $\to$ 2. A is diagonalizable, so there exists a basis of eigenvectors and accordingly, $V = \mathbb{C}^n$ is the orthogonal sum of eigenspaces. Construct an orthonormal basis in every eigenspace (Gram-Schmidt process). Thus, the union of these bases gives an orthonormal basis of V.

2. $\to$ 3. Let B be an orthonormal basis of eigenvectors.

$$\implies U = \begin{pmatrix} b_1 & \dots & b_n \\ \vdots & & \vdots \end{pmatrix} \implies AU = U \cdot \Lambda$$

$$U^{-1} = U^* \implies AU = U\Lambda$$

where $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.

3. $\to$ 1.

$$U^*AU = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\implies A = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^*$$

$$A^* = U \begin{bmatrix} \overline{\lambda}_1 & & \\ & \ddots & \\ & & \overline{\lambda}_n \end{bmatrix} U^*$$

$$
\begin{aligned}
AA^* &= U\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\cdot \underbrace{U^*\cdot U}_{I}\begin{bmatrix}\overline{\lambda}_1 & & \\ & \ddots & \\ & & \overline{\lambda}_n\end{bmatrix}U^* \\
&= U\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix}\overline{\lambda}_1 & & \\ & \ddots & \\ & & \overline{\lambda}_n\end{bmatrix}U^* \\
&= U\begin{bmatrix}\lambda_1\overline{\lambda}_1 & & \\ & \ddots & \\ & & \lambda_n\overline{\lambda}_n\end{bmatrix}U^* \\
&= U\begin{bmatrix}\overline{\lambda}_1 & & \\ & \ddots & \\ & & \overline{\lambda}_n\end{bmatrix}U^*\cdot U\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}U^* \\
&= A^*A
\end{aligned}
$$

□

Person. *Issai Schur (1875–1941)*

12.1 Schur's decomposition, QR decomposition

Theorem 12.8 (Schur's decomposition)**.** *Let $A \in \mathbb{C}^{n\times n}$.*

1. $\implies$ *$\exists U \in \mathcal{U}(n) : U^*AU = R$ is an upper triangular matrix.*

$$U^*AU = \begin{bmatrix}\lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix} \qquad \lambda_i = \text{ eigenvalue of } A$$

   is called Schur decomposition, Schur normal form *or* QR decomposition.

2. *If $A \in \mathbb{R}^{n\times n}$ and $\chi_A(\lambda)$ decomposes into linear factors.*

$$\implies U \text{ has real-valued entries, thus } U \in \mathcal{O}(n)$$

Proof. Proof by induction over n.

Induction base $n = 1$ immediate.

Induction step Let $\lambda \in \operatorname{spec}(A)$. Let u be an eigenvector with $\|u\| = 1$.

$$Au = \lambda u$$

Extend u to an orthonormal basis of $\mathbb{C}^n$

$$(u, w_1, \dots, w_{n-1})$$

$$U_1 = \begin{bmatrix} u & & \\ \vdots & W & \\ \vdots & & \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 & \dots & w_{n-1} \\ \vdots & & \vdots \end{bmatrix} \in \mathbb{C}^{n\times(n-1)}$$

$$U_1^* = \begin{bmatrix} u^* & \dots & \dots \\ & W^* & \\ & & \end{bmatrix} \text{ where } W^* \in \mathbb{C}^{(n-1)\times n}$$

$$U_1^* A U_1 = \begin{bmatrix} u^* & \dots & \dots \\ & W^* & \\ & & \end{bmatrix} A \begin{bmatrix} u & & \\ \vdots & & \\ \vdots & W & \\ \vdots & & \end{bmatrix}$$

$$= \begin{bmatrix} u^* & \dots & \dots & \dots \\ & W^* & & \\ & & & \end{bmatrix} \begin{bmatrix} \lambda u & & \\ \vdots & AW & \\ \vdots & & \end{bmatrix} \text{ where } AW \in \mathbb{C}^{n\times(n-1)} = \begin{bmatrix} \lambda & u^*AW \\ 0 & \\ \vdots & [W^*AW] \\ 0 & \end{bmatrix}$$

$$\text{where } W^*AW \in \mathbb{C}^{(n-1)\times(n-1)}$$

$$I = U_1^* U_1 = \begin{bmatrix} u^* & \dots & \dots & \dots \\ & W^* & & \\ & & & \end{bmatrix} \begin{bmatrix} u & & \\ \vdots & & \\ \vdots & W & \\ \vdots & & \end{bmatrix} = \begin{bmatrix} 1 & u^*W \\ 0 & \\ \vdots & W^*w \\ 0 & \end{bmatrix}$$

By induction hypothesis, there exists unitary $U_2 \in \mathcal{U}(n-1) : U_2^* W^* A W = R_2$.

$$U_2^* W^* A W U_2 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & U_2^* & \\ 0 & & & \end{bmatrix} \begin{bmatrix} \lambda & & * & \\ 0 & & & \\ \vdots & & W^*AW & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & U & \\ 0 & & & \end{bmatrix}$$

$$= \begin{bmatrix} \lambda & * & * & * \\ 0 & & & \\ \vdots & & R_2 & \\ 0 & & & \end{bmatrix}$$

is a triangular matrix. For real values, it is trivial.

□

Corollary 12.9. *A matrix is normal* $\iff$ *Schur normal form = diagonal matrix.*

$$U^*AU = R \implies A = URU^*$$

$$A^* = UR^*U^*$$

$$A^*A = AA^* \iff R^*R = RR^* \iff R \text{ is diagonal matrix}$$

The proof is left as an exercise to the reader.

Theorem 12.10. *Let* $A \in \mathbb{C}^{n\times n}, A = A^* \implies \operatorname{spec}(A) \subseteq \mathbb{R}$.

Compare with $z \in \mathbb{C} : z \in \mathbb{R} \iff z = \overline{z}$.

Proof. A is normal.

$$\ker(\lambda I - A) = \ker(\lambda I - A)^* = \ker(\overline{\lambda} I - A^*) = \ker(\overline{\lambda} I - A)$$

hence if $x \neq 0$ and $Ax = \lambda x \implies Ax = \overline{\lambda} x \implies \lambda = \overline{\lambda}$. □

Corollary 12.11. *If* $A \in \mathbb{R}^{n\times n}$ *is symmetrical.*

$$\implies \exists Q \in O(n) : Q^T A Q = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \text{ with } \lambda_i \in \mathbb{R}$$

Remark (Revision)**.** *A matrix is called positive definite, if* $A = A^*$ *and* $\langle Ax, x\rangle > 0 \forall x \neq 0$. *Let x be an eigenvector.* $Ax = \lambda x$ *and* $\langle Ax, x\rangle = \lambda \langle x, x\rangle = \lambda \cdot ||x||^2 > 0$

Theorem 12.12. *Let* $A \in \mathbb{C}^{n\times n}$ *self-adjoint.*

1. $A > 0 \iff \operatorname{spec}(A) \subseteq (0, \infty)$, *hence eigenvalues are* > 0.
2. $A \geq 0 \iff \operatorname{spec}(A) \subseteq [0, \infty)$, *hence eigenvalues* ≥ 0.
3. *Analogously for negatively (semi)definite.*
4. *A is indefinite* $\iff$ $\exists$ *at least one positive and one negative eigenvalue.*

$$\operatorname{spec}(A) \cap (-\infty, 0) \neq \emptyset \text{ and } \operatorname{spec}(A) \cap (0, \infty) \neq \emptyset$$

Proof. 1. **Direction** $\implies$ $A > 0 \implies A$ is self-adjoint $\implies \operatorname{spec}(A) \subseteq \mathbb{R}$

$$\text{and } Ax = \lambda x \implies \lambda > 0 \quad \text{(self-adjoint)}$$

Direction $\impliedby$ A is self-adjoint, $\operatorname{spec}(A) \subseteq (0, \infty)$. Show that $\langle Ax, x\rangle > 0 \forall x \neq 0$.

A is self-adjoint $\implies$ there exists an orthonormal basis of eigenvectors $(u_1, \dots, u_n)$. Let $x \in \mathbb{C}^n \setminus \{0\} \implies x = \sum_{i=1}^n \alpha_i u_i$ and $A \cdot u_i = \lambda_i \cdot u_i$.

$$\begin{aligned}
\implies \langle Ax, x\rangle &= \left\langle A \sum_{i=1}^n \alpha_i u_i, \sum_{j=1}^n \alpha_j u_j \right\rangle \\
&= \left\langle \sum_{i=1}^n \alpha_i \lambda_i u_i, \sum_{j=1}^n \alpha_j u_j \right\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \lambda_i \overline{\alpha_j} \underbrace{\langle u_i, u_j \rangle}_{\delta_{ij}} \\
&= \sum_{i=1}^n |\alpha_i|^2 \lambda_i > 0 \text{ (if all } \lambda_i > 0) \\
&= \sum_{i=1}^n |\alpha_i|^2 \lambda_i \geq 0 \text{ (if all } \lambda_i \geq 0)
\end{aligned}$$

This proves (1.) and (2.)

□

Remark (Application: Taylor expansion)**.**

$$\begin{aligned}
f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 - \dots \\
f(x, y) &= f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \\
&+ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0 y_0)(x - x_0)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(x_0, y_0)(y - y_0)^2 \\
&+ \left(\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) + \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)\right)(x - x_0)(y - y_0)
\end{aligned}$$

Gradient:

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

$$H_f(x_0, y_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \qquad \text{“Hesse matrix”}$$

$$f(x, y) = f(x_0, y_0) + (\nabla f)^T \cdot \underbrace{\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}}_{\vec{\triangle x}} + \frac{1}{2}(\vec{\triangle x})^T H_f \vec{\triangle x} + O((\triangle x)^2)$$

$H_f^T = H_f$ has real eigenvalues. Extrema are found if $\nabla f = 0$. Minimum at $(x_0, y_0) \iff H_f(x_0, y_0) > 0$. Maximum at (x_0, y_0) if $H_f(x_0, y_0) < 0$.

↓ *This lecture took place on 2018/06/13.*

$$A = A^*$$
$$\mathbb{K} = \mathbb{R} : A = A^T \qquad \text{Taylor series}$$

Example (Bicycling)**.** *Consider the wheel when bicycling. It has radius r, ω is the angle. $\vec{v}$ is the tangent.*

Angular momentum:

$$\vec{L} = m \cdot \vec{r} \times \vec{v}$$

is maintained.

$$\left|\vec{L}\right| = m \cdot r \cdot v = m \cdot r^2 \cdot v$$

$\omega = \dot{\omega}$ *is angular speed.* $v = r \cdot \omega$.

$$\vec{a} \times (\vec{b} \times \vec{c}) = (\vec{a}^t \cdot \vec{c}) \cdot \vec{b} - (\vec{a}^t \cdot \vec{b}) \cdot \vec{c}$$

$$\begin{aligned}
\vec{v} &= \vec{w} \times \vec{r} \\
\vec{L} &= m \cdot \vec{r} \times \vec{v} \\
&= m \cdot \vec{r} \times (\vec{\omega} \times \vec{v}) \\
&= m \cdot \left\|\vec{r}\right\|^2 \times \vec{\omega} - (\vec{r}^t \cdot \vec{\omega}) \cdot \vec{r} \\
&= m \cdot \left(\left\|\vec{r}\right\|^2 I - \vec{r}\vec{r}^t\right) \cdot \vec{\omega} \\
&= \underbrace{H}_{\text{inertia tensor}} \cdot \vec{\omega}
\end{aligned}$$

$$H = r^2 I - \vec{r} \cdot \vec{r}^t$$

$\implies$ *at least 3 main axes.*

12.2 Application: Conic section

Example 12.13 (Conic section)**.** *Radius:* $r = z$. *Cone:* $\left\{(x, y, z)\middle| x^2 + y^2 = z^2\right\}$.

Plane: $E = \left\{u + \xi v + \mu w \middle| \xi, \mu \in \mathbb{R}\right\}$. *Without loss of generality* $v \perp w$. $\|v\| = \|w\| = 1$.

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \qquad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \qquad w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

Intersection:

$$x^2 = (u_1 + \xi v_1 + \eta w_1)^2 + y^2 = (u_2 + \xi v_2 + \eta w_2)^2 = z^2 = (u_3 + \xi v_3 + \eta w_3)^2$$

$\implies$ *quadratic equation.*

$$a\xi^2 + 2b\xi\eta + c\eta^2 + d\xi + e\eta = f$$

$$\begin{pmatrix} \xi & \eta \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \begin{pmatrix} d & e \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = f$$

Which curve?

Step 1: Move the center to the origin *Hence, apply translation such that* $d = e = 0$. *Let* $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ *be the center. By translation, let* $\xi = x_0 + x$ *and* $\eta = y_0 + y$. *Choose* x_0, y_0 *such that d and e are different.*

$$a(x+x_0)^2 + 2b(x+x_0)(y+y_0) + c(y+y_0)^2 + d(x+x_0)^2 + e(y+y_0) + f = 0$$

$$ax^2 + 2ax_0x + ax_0^2 + 2bxy + 2by_0x + 2bx_0y + 2bx_0y_0 + xy^2 + 2cy_0y + cy_0^2 + dx + dx_0 + ey + ey_0 - f = 0$$

$$ax^2 + 2bxy + cy^2 + (2ax_0 + 2by_0 + d)x + (2bx_0 + 2cy_0 + e)y + ax_0^2 + 2bx_0y_0 + cy_0^2 + dx_0 + ey_0 - f = 0$$

x_0, y_0 *such that*

$$2ax_0 + 2by_0 = -d \qquad 2bx_0 + 2cy_0 = -e$$

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} -\frac{d}{2} \\ -\frac{e}{2} \end{pmatrix}$$

is solvable?

Case 1: determinant is zero

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 0$$

Existence of such a solution is not guaranteed. So, 0 *is an eigenvalue. There exists an orthogonal matrix* Q *(rotation/reflection).*

$$Q^t \begin{pmatrix} a & b \\ b & c \end{pmatrix} Q = \begin{bmatrix} \lambda_1 & \\ & 0 \end{bmatrix}$$

Rotation of planes:

$$\begin{pmatrix} x \\ y \end{pmatrix} = Q^t \begin{pmatrix} \xi \\ \eta \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} \xi \\ \eta \end{pmatrix} = Q \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{pmatrix} \tilde{d} \\ \tilde{e} \end{pmatrix} = Q^t \cdot \begin{pmatrix} d \\ e \end{pmatrix}$$

$$\begin{pmatrix} x & y \end{pmatrix} Q^t A Q \begin{pmatrix} x \\ y \end{pmatrix} + \underbrace{\begin{pmatrix} d & e \end{pmatrix} Q}_{\begin{pmatrix} \tilde{d} \\ \tilde{e} \end{pmatrix}} \begin{pmatrix} x \\ y \end{pmatrix} = f$$

$$\lambda_1 x^2 + \tilde{d}x + \tilde{e}y = f$$

If $\tilde{e} = 0$*:*

$$\lambda_1 x^2 + \tilde{x} = f \text{ and } y \text{ arbitrary}$$

$$x = \frac{-\tilde{d} \pm \sqrt{\tilde{d}^2 + 4\lambda_1 f}}{2\lambda_1}$$

if $\tilde{d}^2 + 4\lambda_1 f < 0$*, no solution.* 1 *or* 2 *lines.*
If $\tilde{e} \neq 0$*:*

$$y = \frac{f - \tilde{d}x - \lambda_1 x^2}{\tilde{e}}$$

is a parabola.

Case 2: determinant is non-zero

$$\begin{vmatrix} a & b \\ b & c \end{vmatrix} \neq 0$$

$\rightsquigarrow$ *equation by translation.*

$$ax^2 + 2bxy + cy^2 = g$$

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = g$$

$\rightsquigarrow$ *Q orthogonal such that*

$$Q^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} Q = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \text{ with } \lambda_1, \lambda_2 \neq 0$$

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = Q^t \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = Q \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}$$

$$\begin{pmatrix} \tilde{x} & \tilde{y} \end{pmatrix} Q^t \begin{pmatrix} a & b \\ b & c \end{pmatrix} Q \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = g$$

$$\lambda_1 \tilde{x}^2 + \lambda_2 \tilde{y}^2 = g$$

Case 2a: $g = 0$

$$\tilde{y}^2 = -\frac{\lambda_1}{\lambda_2} \tilde{x}^2$$

if $\operatorname{sign} \lambda_1 = \operatorname{sign} \lambda_2 \implies \tilde{x} = \tilde{y} = 0$*. If* $\operatorname{sign} \lambda_1 \neq \operatorname{sign} \lambda_2 \rightsquigarrow \tilde{y} = \pm \frac{\lambda_1}{\lambda_2} \tilde{x}$*.*

$$\frac{\lambda_1}{g} \tilde{x}^2 + \frac{\lambda_2}{g} \tilde{y}^2 = 1$$

Classification: $A > 0$ *or* $A < 0$ *gives an ellipsis, point or the empty set. If* $A \geq 0$ *or* $A \leq 0$*, a parabola or line or the empty set is given. If* A *is indefinite, a hyperbola is given.*

Example 12.14. *Analogously: Quadric in* $\mathbb{R}^3$*. A three-dimensional conic section. Compare with Figure 17.*

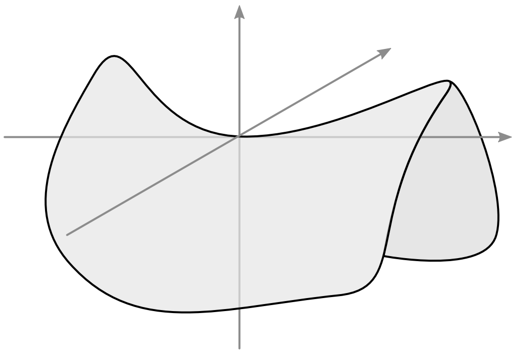


Figure 17: Quadric

Theorem 12.15. *Let $A \geq 0$. Then $\exists! B \geq 0 : B^2 = A$.*

$$B := A^{\frac{1}{2}}$$

Proof. We prove the existence of B.

A is self-adjoint, so there exists a unitary matrix U such that $U^* A U = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ with all $\lambda_i \geq 0$.

$$A = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} U^*$$

$$B = U \begin{bmatrix} +\sqrt{\lambda_1} & & \\ & \ddots & \\ & & +\sqrt{\lambda_n} \end{bmatrix} U^*$$

$M \geq 0 \implies C^* M C' \geq 0$. B satisfies the condition $B^2 = A$ with $B \geq 0$.

Conditionless $B \geq 0$: $\rightsquigarrow 2^{\operatorname{rank}(A)}$ different solutions.

We prove uniqueness of B.

Let $B \geq 0$ and $B^2 = a$. Let $u_1, \dots, u_n$ be orthonormal basis of eigenvectors for eigenvalues $\mu_1, \dots, \mu_n$ of B with $\lambda_i \geq 0$.

$$\implies Bu_i = \mu_i u \qquad \mu_i \geq 0$$

$$Au_i = B^2 u_i = \mu_i^2 u_i \implies \mu_i^2 = \lambda_i \implies \mu_i = +\sqrt{\lambda_i}$$

$$\implies U^* B U = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{bmatrix} \land U^* A U = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

□

Remark (Another solution). *Find B such that $B^* B = A$. There are infinitely many solutions.*

12.3 Cholesky decomposition

Person. *André-Louis Cholesky (1875–1918) descending from the Cholewski (Polish) family*

Theorem 12.16 (Cholesky decomposition)**.** *Let $A > 0 \iff \exists C \in \mathbb{C}^{n\times n}$ (lower triangular matrix) such that $A = C \cdot C^*$ and C has positive diagonal entries. Compare with LU decomposition.*

Remark (Algorithm)**.** *Determine the matrix C or abort if A is not positive definite.*

Remark (Schur complement)**.**

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$\implies M/D := A - BD^{-1}C$$

is the Schur complement of block D in M where M is a $(p+q)\times(p+q)$ matrix and M/D is a $p \times p$ matrix.

Lemma 12.17. *Let $A \in \mathbb{C}^{n\times n}$. $A > 0$, $b \in \mathbb{C}^n$, $\gamma > 0$.*

$$\det \left[\begin{array}{c|c} A & B \\ \hline b^* & \gamma \end{array}\right] = \det A \cdot (\gamma - b^* A^{-1} b)$$

$$\begin{vmatrix} A & b \\ b^* & \gamma \end{vmatrix} = \begin{vmatrix} I & 0 \\ -b^*A^{-1} & 1 \end{vmatrix} \cdot \begin{vmatrix} A & b \\ b^* & \gamma \end{vmatrix} = \begin{vmatrix} A & b \\ -\underbrace{b^*A^{-1}A + b^*}_{=0} & -b^*A^{-1}b + \gamma \end{vmatrix} = \det A \cdot (-b^*A^{-1}b + \gamma)$$

Proof by Cholesky. By complete induction:

Case n = 1:

$$[a_{11}] > 0 \qquad \implies a_{11} > 0$$

$$C = [e^{i\theta}\sqrt{a_{11}}] \text{ is unique} \qquad C^* = [e^{-i\theta}\sqrt{a_n}]$$

Case k → k + 1:

$$A_{k+1} = \begin{bmatrix} A_k & b \\ b^* & \gamma \end{bmatrix} \qquad A_k \in \mathbb{C}^{n\times n}, A_k > 0$$

By induction hypothesis: $\exists C_k : A_k = C_k C_k^*$.

$$\text{Find } C_{k+1} = \begin{pmatrix} & 0 \\ C_k & \vdots \\ & 0 \\ C^* & \alpha \end{pmatrix} \text{ such that } C_{k+1}C_{k+1}^* = A_{k+1}, \alpha > 0$$

$$C_{k+1}C_{k+1}^* = \begin{bmatrix} & 0 \\ C_k & \vdots \\ & 0 \\ C^* & \alpha \end{bmatrix} \begin{bmatrix} & C_k^* & & c \\ & & & \\ 0 & \dots & 0 & \alpha \end{bmatrix} = \left[\begin{array}{c|c} C_kC_k^* & C_k \cdot c \\ C^*C_k^* & C^*C + \alpha^2 \end{array}\right] \overset{!}{=} \begin{bmatrix} A_k & b \\ b^* & \gamma \end{bmatrix}$$

Requirement: $C_k c\dot{c} = b$. $c^*c + \alpha^2 = \gamma$. Choose $c = C_k^{-1}b$.

$$\begin{aligned}\alpha^2 &= \gamma - c^*c = \gamma - b^*C_k^{*-1}C_k{}^{-1}b \\ &= \gamma - b^*A^{-1}b = \frac{\det A_{k+1}}{\det A_k} > 0\end{aligned}$$

$$A_k = C_k C_k^* \qquad A_k^{-1} = C_k^{*-1}C_k^{-1}$$

$$\implies \alpha = \sqrt{\gamma - b^*A^{-1}b} = \sqrt{\frac{\det A_{k+1}}{\det A_k}}$$

□

Remark (Application). *Find C such that $C \cdot C^* = A$. $c_{ij} = 0$ if $j > i$. $c_{ii} > 0$.*

$$a_{ij} = \sum_{k=1}^{n} c_{ik}(c^*)_{kj} = \sum_{k=1}^{n} c_{ik}\overline{c_{jk}} = \sum_{k=1}^{\min(i,j)} c_{ik}\overline{c_{jk}}$$

The algorithm fills up columns from top to bottom and the columns are fill from left to right.

1st column

$$\begin{aligned}a_{11} &= c_{11} \cdot \overline{c_{11}} = c_{11}^2 \implies c_{11} = \sqrt{a_{11}} \\ a_{21} &= c_{21} \cdot \overline{c_{11}} = c_{21} \cdot c_{11} \implies c_{21} = \frac{a_{21}}{c_{11}} \\ a_{31} &= \sum_{k=1}^{\min(3,1)} c_{3k}\overline{c_{jk}} = c_{31}c_{11} \implies c_{31} = \frac{a_{31}}{c_4} \\ c_{n1} &= \frac{a_{n1}}{c_{11}} = \frac{a_{n1}}{\sqrt{a_{11}}}\end{aligned}$$

$$c_{11} = \sqrt{a_{11}} \qquad c_{21} = \frac{a_{21}}{\sqrt{a_{11}}} \qquad \dots \qquad c_{n1} = \frac{a_{n1}}{\sqrt{a_{11}}}$$

Induction *Assume columns $1, \dots, j-1$ are determined, i.e. c_{ik} is known for $k = 1, \dots, j-1$ and $i = 1, \dots, n$.*

$$a_{ji} = \sum_{k=1}^{j} c_{jk}\overline{c_{jk}} = \sum_{k=1}^{j-1} \left|c_{jk}\right|^2 + c_{jj}^2$$

$$\implies c_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} \left|c_{jk}\right|^2}$$

for $i > j$:

$$a_{ij} = \sum_{k=1}^{j} c_{ik}\overline{c_{jk}} = \sum_{k=1}^{j-1} c_{ik}\overline{c_{jk}} + c_{ij}c_{jj}$$

$$c_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} c_{ik}\overline{c_{jk}}}{c_{jj}}$$

where c_{jj} was already determined and the values of the enumerator are already known.

Remark. $A > 0 \iff$ *congruent to unit matrix* $\iff$ *A has full rank. See Corollary 8.30.*

↓ *This lecture took place on 2018/06/18.*

Cholesky decomposition:

1. Given $A > 0$
2. Find C such that $A = CC^*$. Recursively:

$$c_{ij} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} |c_{ik}|^2}$$

$$c_{ij} = \frac{1}{c_{jj}} \left(a_{ij - \sum_{k=1}^{j-1} c_{ik}\overline{c_{jk}}} \right)$$

Example 12.18.

$$A = \begin{bmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} c_{11} & & \\ c_{21} & c_{22} & \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \cdot \begin{bmatrix} c_{11} & \overline{c_{21}} & \overline{c_{31}} \\ 0 & c_{22} & \overline{c_{32}} \\ 0 & 0 & c_{33} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11}^2 & c_{11}\overline{c_{21}} & c_{11}\overline{c_{31}} \\ c_{21}c_{11} & |c_{21}|^2 + c_{22}^2 & \\ c_{31}c_{11} & c_{31}\overline{c_{21}} + c_{32}c_{22} & |c_{31}|^2 + |c_{32}|^2 + c_{33}^2 \end{bmatrix}$$

with

$$\begin{aligned}
c_{11} &= \sqrt{4} = 2 \\
12 &= c_{11} \cdot \overline{c_{21}} \implies c_{21} = \frac{12}{c_{11}} = 6 \\
-16 &= c_{11}\overline{c_{31}} \implies c_{31} = -\frac{16}{c_{11}} = -8 \\
|c_{21}|^2 &= 37 \\
c_{22} &= \sqrt{37 - 6^2} = 1 \\
c_{31}\overline{c_{21}} + c_{32}c_{22} &= -43 \\
c_{32} &= \frac{-43 - c_{31}\overline{c_21}}{c_{22}} = \frac{-43 + 8 \cdot 6}{1} = 5 \\
|c_{31}|^2 + |c_{32}|^2 + c_{33}^2 = 98 \implies c_{33} &= \sqrt{98 - (-8)^2 - 5^2} = 3
\end{aligned}$$

If some root of a negative number needs to be taken, the matrix was not positive definite.

Remark 12.19. *If $A \in \mathbb{R}^{n \times n}$, then*

1. *also $C \in \mathbb{R}^{n \times n}$.*
2. *C is uniquely determined.*
3. *Cholesky decomposition is an LU decomposition*

Corollary 12.20. *If $A > 0$, then $\det A \leq a_{11}a_{22}\dots a_{nn}$.*

Proof. $A = CC^*$ is a Cholesky decomposition.

$$c_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{j-1} |c_{ik}|^2} \leq \sqrt{a_{ii}}$$

We apply the product law.

$$\begin{aligned}
\det(A) &= \det(CC^*) \\
&= \det(C) \cdot \det(C^*) \\
&= \left|\det(C)\right|^2 \\
&= \left|\prod_{i=1}^{n} c_{ii}\right|^2 = \prod c_{ii}^2 \leq \prod a_{ii}
\end{aligned}$$

□

Person. *J. S. Hadamard (1865–1963)*

Corollary 12.21 (Hadamard's inequality)**.** *Let $A \in \mathbb{C}^{n\times n}$ with column vectors $a_1, \dots, a_n$.*

$$|\det A| \leq \prod_{i=1}^{n} ||a_i|| \qquad ||a_i|| := \sqrt{\langle a_i, a_i\rangle} = \sqrt{a_i^* \cdot a_i} = \textit{Euclidean norm}$$

Proof. **Case 1:** A is singular, thus trivial ($\det(A) = 0$).

Case 2: A is invertible $\implies$ $B = A^*A$ is invertible $\implies$ positive definite.

$$\begin{aligned}
\det B \leq \prod_{i=1}^{n} b_{ii} &= \det(A^*A) = |\det A|^2 \\
b_{ii} &= \text{ i-th row of } A^* \times \text{ i-th colum of } A \; = a_i^* \cdot a_i = ||a_i||^2 \\
\implies |\det A|^2 \leq \prod_{i=1}^{n} ||a_i||^2
\end{aligned}$$

□

In this chapter: $AA^* = A^*A$.
There exists an orthonormal basis of eigenvectors:

$$\begin{aligned}
Ax_i &= \lambda_i x_i \\
\left\langle x_i, x_j \right\rangle &= \delta_{ij} \\
x \in \mathbb{C}^n \rightsquigarrow x &= \sum \alpha_i x_i \qquad \alpha_i = \langle x, x_i\rangle \\
Ax &= \sum \alpha_i A x_i = \sum \alpha_i \lambda_i x_i \\
&= \sum \lambda_i \langle x, x_i\rangle x_i \\
\implies A &= \sum_{i=1}^{n} \lambda \langle \cdot, x_i\rangle x_i
\end{aligned}$$

where $\cdot$ is a placeholder. It represents the map $\langle \cdot, x_i \rangle : \mathbb{C}^n \to \mathbb{C}$ with $x \mapsto \langle x, x_i \rangle$.

Remark. *Let $z \in \mathbb{C}$.*

$$z = r \cdot e^{i\theta}$$

$$r = |z| = \sqrt{z\bar{z}}$$
$$e^{i\theta} = \frac{z}{|z|}$$

Theorem 12.22 (Polar decomposition)**.** *Let $A \in \mathbb{C}^{n \times n}$.*

$$|A| := (A^*A)^{\frac{1}{2}} \qquad (\textit{unique, positive semidefinite root})$$

Then $\exists U \in \mathcal{U}(n)$ such that $A = U \cdot |A|$.

Proof. **Case 1: A is invertible** AA^* is positive definite.

$$\rightsquigarrow A^*A = V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} V^* \qquad \lambda_i > 0$$

$$|A| := (A^*A)^{\frac{1}{2}} = V \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} V^*$$

$$U = A \cdot |A|^{-1} \text{ is unitary}$$
$$U = A \cdot (A^*A)^{-\frac{1}{2}}$$
$$U^*U = (A(A^*A)^{-\frac{1}{2}})^*(A(A^*A)^{-\frac{1}{2}}) = ((A^*A)^{-\frac{1}{2}} \cdot A^*)(A(A^*A)^{-\frac{1}{2}}) = (A^*A)^{-\frac{1}{2}}(A^*A)(A^*A)^{-\frac{1}{2}}$$
$$= V \begin{bmatrix} \lambda_1^{-\frac{1}{2}} & & \\ & \ddots & \\ & & \lambda_n^{-\frac{1}{2}} \end{bmatrix} V^*V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} V^*V \begin{bmatrix} \lambda_1^{-\frac{1}{2}} & & \\ & \ddots & \\ & & \lambda_n^{-\frac{1}{2}} \end{bmatrix} V^* = V \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} V^* = I$$
$$U \cdot |A| = A \cdot |A|^{-1} \cdot |A| = A$$

Case 2: A is singular Incomprehensible. Let $A^*A \geq 0$ and some $\lambda_i = 0$. By change of basis, $A = V \operatorname{diag}(\lambda_1, \dots, \lambda_k, 0, \dots, 0)V^*$.

$$|A| = V \operatorname{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_k^{\frac{1}{2}}, 0, \dots, 0)V^*$$

$$|A| = \begin{bmatrix} P & \\ & 0 \end{bmatrix} \qquad U = V \begin{bmatrix} P^{-1} & \\ & 0 \end{bmatrix} \simeq \begin{bmatrix} \tilde{U} & \\ & I \end{bmatrix}$$
$$\implies A = U|A|$$

□

12.4 Singular value decomposition

Remark (Singular value decomposition).

$$A \in \mathbb{C}^{n\times n} \qquad A = U(A^*A)^{\frac{1}{2}} \qquad U \textit{ unitary}$$

$$(A^*A)^{\frac{1}{2}} \geq 0 \textit{ with eigenvalue } s_i \geq 0 \textit{ called singular values of } A$$

$$(A^*A)^{\frac{1}{2}} = \sum s_i \langle \cdot, x_i \rangle \, x_i \textit{ where } x_i \textit{ is an orthonormal basis}$$

$$A = U \cdot (A^*A)^{\frac{1}{2}} = \sum_{i=1}^{n} s_i \langle \cdot, x_i \rangle \overbrace{U_{x_i}}^{=:y_i} = \sum_{i=1}^{n} s_i \langle \cdot, x_i \rangle \, y_i$$

$$y_i = Ux_i \textit{ is also an orthonormal basis}$$

$$A \cdot y_i = s_i \cdot x_i \rightarrow \textit{ numerically stable}$$

$$A^{-1} = \sum s_i^{-1} \langle \cdot, y_i \rangle \, x_i$$

Remark. *Singular value decomposition is numerically stable and therefore very desirable.*

It furthermore has a very important application in medicine: CT scans. Inside the CT tube, X-rays are sent from all directions to all directions. You can only determine how much the

$$\int_\gamma f(x(t), y(t))\, dt = Rf(\psi, \Theta)$$

Is linear: $R(f_1 + f_2) = Rh + Rf_2$. *Radon transformation.* $\lambda_i \to 0$.

You need to invert the integral. The SVD is the only numerically stable method to achieve it. Other methods will trigger numerical errors that will amplify and therefore give wrong images.

13 Eigenvalue estimates

Definition 13.1. *Let* $A \in \mathbb{C}^{n\times n}$.

$$W(A) = \{\langle Ax, x\rangle \mid \|x\| = 1\} \subseteq \mathbb{C} \textit{ is called } \text{numerical range of } A$$

$$w(A) = \sup\{|z| \mid z \in W(A)\} \textit{ is called } \text{numerical radius of } A$$

Lemma 13.2.

$$\operatorname{spec}(A) \subseteq W(A)$$

Proof. $\lambda \in \operatorname{spec}(A)$, eigenvector x such that $Ax = \lambda x$, $\|x\| = 1$.

$$\implies \langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda \in W(A)$$

□

Theorem 13.3 (Theorem by Toeplitz-Hausdorff). *$W(A)$ is convex.*

Remark. *The following implications are left as an exercise to the reader: A is normal* $\implies$ $W(A) = \underbrace{\textit{convex}\ \operatorname{spec}(A)}_{\textit{convex hull}} = \left\{ \sum_{i=1}^{n} \alpha_i \underbrace{\lambda_i}_{\textit{eigenvalue}} \;\middle|\; 0 \le \alpha_i \le 1, \sum \alpha_i = 1 \right\}$ = *convex set that contains* $\operatorname{spec}(A)$.

Person. *J. W. Strutt (1842–1919) aka 3rd Lord Rayleigh*

Person. *W. Ritz (1878–1909), discovered the element Argon (→ Nobel prize)*

13.1 Rayleigh–Ritz Theorem

Theorem 13.4 (Rayleigh–Ritz Theorem)**.** *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Then*

$$\lambda_1 = \min_{x\neq 0} \underbrace{\frac{\langle Ax, x\rangle}{\langle x, x\rangle}}_{\textit{Rayleigh quotient}} = \min\left\{ \langle Ax, x\rangle \mid \|x\| = 1 \right\} = \min W(A)$$

$$\vdots$$

$$\lambda_n = \max_{x\neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \max\left\{ \langle Ax, x\rangle \mid \|x\| = 1 \right\} = \max W(A)$$

where

$$A = A^* \implies \langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$$

Proof. $\lambda_1, \lambda_n \in W(A)$ because $\operatorname{spec}(A) \subseteq W(A)$.

$$\implies \min W(A) \le \lambda_1 \qquad \max W(A) \ge \lambda_n$$

Show that $\forall x : \lambda_1 \le \frac{\langle Ax,x\rangle}{\langle x,x\rangle} \le \lambda_n$.
Let $u_1, \dots, u_n$ be an orthonormal basis of eigenvectors. Let $x \in \mathbb{C}^n$, $\|x\| = 1$. $x = \sum_{i=1}^{n} \alpha_i x_i \implies \|x\|^2 = \sum_{i=1}^{n} |\alpha_i|^2 = 1$.

$$\begin{aligned}
\langle Ax, x\rangle &= \left\langle A \sum \alpha_i x_i, \sum \alpha_j x_j \right\rangle \\
&= \left\langle \sum \alpha_i \lambda_i x_i, \sum \alpha_j x_j \right\rangle \\
&= \sum_i \sum_j \lambda_i \alpha_i \overline{\alpha}_j \left\langle x_i, x_j \right\rangle \text{ with } \left\langle x_i, x_j \right\rangle = \delta_{ij} \\
&= \sum_i \lambda_i |\alpha_i|^2 && \le \max(\lambda_i) \sum |\alpha_i|^2 = \max \lambda_i \\
& && \ge \min(\lambda_i) \sum |\alpha_i|^2 = \min \lambda_i
\end{aligned}$$

□

Remark 13.5. *How can we obtain the other eigenvalues?*

For $x \in \mathcal{L}(u_2, \dots, u_n)$, $\langle Ax, x\rangle \ge \lambda_2 \|x\|^2$.
For $x \in \mathcal{L}(u_1, \dots, u_{n-1}) = \{ x \mid \langle x, u_n\rangle = 0 \} = \{u_n\}^\perp$, $\langle Ax, x\rangle \le \lambda_{n-1}$.

Person. *Richard Courant (1888–1972)*
E. Fischer (1875–1954)
H. Weyl (1885–1955)

All of them worked in Göttingen, Germany.

13.2 Courant–Fischer–Weyl min–max principle

Theorem 13.6 (Courant–Fischer–Weyl min–max principle). *Let $A \in \mathbb{C}^{n \times n}$ be self-adjoint.*

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

Then it holds that

1. $\lambda_k = \max_{\substack{W \subseteq V \\ \dim W = k-1}} \min_{x \in W^\perp \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$.
   Special case $k = 1$*:* $W = \{0\} \to \lambda_i = \min_x \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$.
2. $\lambda_{n+1-k} = \min_{\substack{W \subseteq V \\ \dim W = k-1}} \max_{x \in W^\perp \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$.
   Special case $k = 1$*:* $\lambda_n = \max_x \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$.

This theorem is more generic than the Rayleigh–Ritz Theorem.

Proof. For $W \subseteq V$. Let $m_A(W) = \min\left\{ \frac{\langle Ax, x \rangle}{\langle x, x \rangle} \,\middle|\, x \in W^\perp \setminus \{0\} \right\}$. For vectors $m_A(w_1, \dots, w_k) = m_A(\mathcal{L}(w_1, \dots, w_k))$. For some orthonormal basis $u_1, \dots, u_n$ of eigenvectors,

$$\lambda_k = m_A(u_1, \dots, u_{k-1}) \overset{\substack{\text{see proof of} \\ \text{Theorem 13.4}}}{=} \min\left\{ \frac{\langle Ax, x \rangle}{\langle x, x \rangle} \,\middle|\, x \in \mathcal{L}(u_k, \dots, u_n) \right\}$$

with $\mathcal{L}(u_k, \dots, u_n) = \{u_1, \dots, u_{k-1}\}^\perp$.

$$\implies \lambda_k \leq \max_{\substack{W \subseteq V \\ \dim W = k-1}} m_A(W)$$

Show that: $\forall W \subseteq V$ with $\dim(W) = k - 1$, $m_A(w) \leq \lambda_k$.

$$\dim W^\perp = n - k + 1 \implies W^\perp \cap \mathcal{L}(u_1, \dots, u_k) \neq \{0\}$$

$$v = \sum_{i=1}^{k} \alpha_i u_i \in W^\perp \cap \mathcal{L}(u_1, \dots, u_k) \text{ with } \|v\| = 1$$

$$\langle Av, v \rangle = \sum_{i=1}^{k} |\alpha_i|^2 \cdot \lambda_i \leq \lambda_k \underbrace{\sum |\alpha_i|^2}_{=1} = \lambda_k$$

$$m_A(w) = \min_{x \in W^\perp} \langle Ax, x \rangle \leq \langle Av, v \rangle = \lambda_k$$

□

↓ *This lecture took place on 2018/06/20.*

The Rayleigh-Ritz coefficient gives us $\lambda_1 = \min_{\langle x,x\rangle=1} \langle Ax, x\rangle = \min W(A)$ and $\lambda_n = \max_{\langle x,x\rangle=1} \langle Ax, x\rangle$ with $A = A^*$ and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. $x = \sum \alpha_i u_i$ with $\sum |\alpha_i|^2 = 1$ is an orthonormal basis $u_1, \dots, u_n$ of eigenvectors. $\langle Ax, x\rangle = \sum_1^n \lambda_i |\alpha_i|^2$. For $x \in \mathcal{L}(u_2, \dots, u_n) \to \geq \lambda_2$.

This is used by the Courant–Fischer–Weyl min–max principle.

1. First statement:
$$\lambda_k = \max_{\substack{W \leq V \\ \dim W = k-1}} \min_{\substack{x \in W^\perp \\ \langle x,x\rangle=1}} \langle Ax, x\rangle$$
2. Second statement:
$$\lambda_{n+1-k} = \min_{\substack{W \leq V \\ \dim W = k-1}} \max_{\substack{x \in W^\perp \\ \langle x,x\rangle=1}} \langle Ax, x\rangle$$

Proof of Theorem 13.6 continued. The second statement follows from the first: $-A$ has eigenvalues: $-\lambda_n \leq -\lambda_{n-1} \leq \dots \leq -\lambda_2 \leq -\lambda_1$.

We apply the first statement on $-A$.

$$\begin{aligned}
\underbrace{\lambda_k(-A)}_{-\lambda_{n+1-k}} &= \max_{\substack{W \subseteq V \\ \dim W = k-1}} \min_{\substack{x \in W^\perp \\ \langle x,x\rangle=1}} \langle -Ax, x\rangle \\
&= \max_{\substack{W \subseteq V \\ \dim W = k-1}} \left(- \max_{\substack{x \in W^\perp \\ \langle x,x\rangle=1}} \langle Ax, x\rangle \right) \\
&= - \min_{\substack{W \subseteq V \\ \dim W = k-1}} \max_{\substack{x \in W^\perp \\ \langle x,x\rangle=1}} \langle Ax, x\rangle
\end{aligned}$$

□

13.3 Cauchy interlacing theorem

Corollary 13.7 (Cauchy interlacing theorem (dt. Schachtelungssatz von Cauchy))**.** *Let $A \in \mathbb{C}^{n\times n}$, $A = A^*$. $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Let $B = [a_{ij}]_{i,j=1,\dots,n-1}$. Thus, the last row and column was removed. The dimension is reduced by 1. $B = B^*$. Eigenvalues: $\mu_1 \leq \mu_2 \leq \dots \leq \mu_{n-1}$. Then it holds that $\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \lambda_3 \leq \dots \leq \lambda_{n-1} \leq \mu_{n-1} \leq \lambda_n$. In general: If P is an orthogonal projection on a subspace of dimension $n-1$.*

For example,

$$P = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{bmatrix}$$

then the eigenvalues (except for eigenvalue 0) of PAP are nested like above.

Proof.

$$A = \begin{bmatrix} [B] & b \\ b^* & \gamma \end{bmatrix}$$

Let $w_1, \dots, w_{n-1} \in \mathbb{C}^{n-1}$ be an orthonormal basis of eigenvectors of B.

$$Bw_i = \mu_i w_i \qquad u_i = \begin{pmatrix} w_i \\ 0 \end{pmatrix} \in \mathbb{C}^n \qquad u_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

$\rightsquigarrow u_1, \dots, u_{n-1}, u_n$ is an orthonormal basis of $\mathbb{C}^n$. Attention! There is no eigenvector of A.

$$W_k = \mathcal{L}(u_1, \dots, u_{k-1}, u_n)$$

$$\text{By Theorem 13.6} \implies \lambda_{k+1} = \max_{\substack{W \subseteq \mathbb{C}^n \\ \dim W = k}} \min_{x \in W^\perp \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} \geq \min_{x \in W_k^\perp \setminus \{0\}} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

$$x \in W_k^\perp \iff \langle x, u_i \rangle = 0 \qquad i = 1, \dots, k-1 \land \langle x, u_n \rangle = x_n = 0$$

$$\iff x = \begin{bmatrix} y \\ 0 \end{bmatrix}, y \in \{w_1, \dots, w_{k-1}\}^\perp \underbrace{=}_{w_1, \dots, w_{n-1} \text{ is ONB of } \mathbb{C}^{n-1}} \mathcal{L}(w_k, \dots, w_{n-1}) \subseteq \mathbb{C}^{n-1}$$

$$y = \sum_{i=k}^{n-1} \alpha_i w_i$$

$$\frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \frac{\left\langle \begin{bmatrix} B & b \\ b^* & \gamma \end{bmatrix} \begin{bmatrix} y \\ 0 \end{bmatrix}, \begin{bmatrix} y \\ 0 \end{bmatrix} \right\rangle}{\left\langle \begin{bmatrix} y \\ 0 \end{bmatrix}, \begin{bmatrix} y \\ 0 \end{bmatrix} \right\rangle} = \frac{\left\langle \begin{bmatrix} B & y \\ b^* & -y \end{bmatrix}, \begin{bmatrix} y \\ 0 \end{bmatrix} \right\rangle}{\langle y, y \rangle} = \frac{\langle By, y \rangle}{\langle y, y \rangle}$$

$$= \frac{\left\langle B \sum_{i=k}^{n-1} \alpha_i w_i, \sum_{j=k}^{n-1} \alpha_j w_j \right\rangle}{\sum_{i=1}^{n-1} |\alpha_i|^2} = \frac{\sum_{i=k}^{n-1} \mu_i |\alpha_i|^2}{\sum_{i=1}^{n-1} |\alpha_i|^2} \geq \mu_k$$

Inversion: $-A$ has eigenvalues $-\lambda_n \leq -\lambda_{n-1} \leq \dots \leq -\lambda_2 \leq -\lambda_1$. $-B$ has eigenvalues $-\mu_{n-1} \leq -\mu_{n-2} \leq \dots \leq -\mu_2 \leq -\mu_1$.

We apply step 1 on $-A$ and $-B$:

$$\lambda_{k+1}(-A) \geq \lambda_k(-B)$$

$$-\lambda_{n-k} \geq -\mu_{n-k}$$

$$\implies \lambda_{n-k} \leq \mu_{n-k} \quad \forall k = 1, \dots, n-1$$

$$\implies \lambda_k \leq \mu_k \quad \forall k \forall k = 1, \dots, n-1$$

□

Corollary 13.8. *A, B are self-adjoint $\in \mathbb{C}^{n \times n}$.*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B)$$

Proof.

$$\lambda_1(B) \leq \frac{\langle Bx, x\rangle}{\langle x, x\rangle} \leq \lambda_n(B)$$

because of Theorem 13.4.

$$\lambda_k(A+B) = \max_{\dim W = k-1} \min_{x \in W^\perp \setminus \{0\}} \frac{\langle (A+B)x, x\rangle}{\langle x, x\rangle}$$

$$\frac{\langle Ax, x\rangle}{\langle x, x\rangle} + \lambda_1(B) \leq \frac{\langle (A+B)x, x\rangle}{\langle x, x\rangle} \leq \frac{\langle Ax, x\rangle}{\langle x, x\rangle} + \lambda_n(B)$$

This goes on and on $\rightsquigarrow \lambda_k(A) + \lambda_l(B)$.

$$\leq \lambda_l(A) + \lambda_n(B)$$

□

Corollary 13.9. *If $B \geq 0$, $A = A^*$, then $\lambda_k(A) \leq \lambda_k(A+B) \forall k$.*

Person. *Semën Aranovič Geršgorin (1901–1933)*

13.4 Geršgorin theorem

Theorem 13.10 (Geršgorin Theorem)**.** *Let $A \in \mathbb{C}^{n \times n}$.*

$$\begin{bmatrix} |a_{11}| & |a_{12}| & \dots & |a_{1n}| \\ |a_{21}| & |a_{22}| & \dots & |a_{2n}| \\ \vdots & & \ddots & \vdots \\ |a_{n1}| & |a_{n2}| & \dots & |a_{nn}| \end{bmatrix} \quad \to \quad r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \left|a_{ij}\right| \quad \to \quad \begin{bmatrix} \sum_{j \neq 1} \left|a_{1j}\right| \\ \sum_{j \neq 2} \left|a_{2j}\right| \\ \vdots \\ \sum_{j \neq n} \left|a_{nj}\right| \end{bmatrix}$$

So, we remove the diagonal elements and consider the row sum norm. This yields the so-called Geršgorin discs. *The theorem claims:*

$$\operatorname{spec}(A) \subseteq \bigcup_{k=1}^{n} \{z \in \mathbb{C} \mid |z - a_{ii}| \leq r_i\}$$

Proof. Show that $\forall \lambda \in \operatorname{spec}(A) \exists i : |\lambda - a_{ii}| \leq r_i$. Let x be an eigenvector: $Ax = \lambda x$. Without loss of generality: $\max_i |x_i| = 1$. Hence $\forall j : \left|x_j\right| \leq 1$ and $\exists i : |x_i| = 1$.

$$\underbrace{(Ax)_i}_{\sum_j a_{ij} x_j} = \lambda \cdot x_i \text{ because it is an eigenvector}$$

$$\implies \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j = \lambda x_i - a_{ii} x_i \qquad \implies \underbrace{|\lambda x_i - a_{ii} x_i|}_{=|\lambda - a_{ii}| \cdot |x_i|} = \left|\sum_{j \neq i} a_{ij} x_j\right|$$

$$|\lambda - a_{ii}| \cdot \underbrace{|x_i|}_{=1} \leq \sum_{j \neq i} \left|a_{ij}\right| \underbrace{\left|x_j\right|}_{\leq 1}$$

$$|\lambda - a_{ii}| \leq \sum_{j \neq i} \left|a_{ij}\right| = r_i$$

□

Remark 13.11. *For dimension greater* 5*, it becomes infeasible to determine the characteristic polynomial to retrieve the eigenvalues.*

In theory:

1. $\chi_A(x) = \det(x \cdot I - A)$ *(difficult if precision is necessary)*
2. *find roots (cumbersome or infeasible)*
3. *find eigenvectors (numerically unstable)*

In practice, we apply an iterative approach:

Given A

Find *x such that* $Ax = \lambda x$, $A^2x = \lambda Ax$, $A^3x = \lambda A^2 x \rightsquigarrow A^\infty x = \lambda A^\infty x$ *then* $A^\infty x$ *is eigenvector. As mathematicians we need to ask ourselves, whether* $A^\infty x$ *converges? The answer is no, not always. We need to fix x to converge to infinity, not* 0*.*

Choose initial vector x_0 *with* $\|x_0\| = 1$*.*

$$w_{k+1} = Ax_k \qquad x_{k+1} = \frac{w_{k+1}}{\left\|w_{k+1}\right\|} \qquad \implies \forall n : \|x_n\| = 1$$

Thus we achieved that all the points lie inside the unit sphere. By the Heine-Borel Theorem, a sequence in a bounded and closed (thus compact) space has a convergent subsequence.

Claim. x_n *converges assuming* $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$ *and A diagonalizable.*

Let $v_1, \ldots, v_n$ *be a basis of eigenvectors.*

$$x_0 = \alpha_0 v_1 + \cdots + \alpha_n v_n$$

$$x_k = \frac{A^k x_0}{\left\|A^k x_0\right\|}$$

converges towards v_1*.*

$$A^k v_i = \lambda_i^k v_i$$

$$\begin{aligned}
A^k x_0 &= \alpha_1 A^k v_1 + \cdots + A^k v_n \\
&= \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \cdots + \alpha_n \lambda_n^k v_n \\
x_k = \frac{A^k x_0}{\left\|A^k x_0\right\|} &= \frac{\alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \cdots + \alpha_n \lambda_n^k v_k}{\left\|\alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \cdots + \alpha_n \lambda_n^k v_n\right\|} \\
&= \frac{\lambda_1 \cdot \lambda_1^k}{\left|\alpha_1 \cdot \lambda_1^k\right|} \cdot \frac{v_1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n}{\left\|v_1 + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_n}\right)^k v_n\right\|} \\
&\overset{k\to\infty}{\approx} \frac{\alpha_1 \lambda_1^k}{\left|\lambda_1^k\right|} \cdot v_1 \text{ because } |\lambda_i| > |\lambda_1| \, \forall i \geq 2 \implies \frac{|\lambda_i|}{|\lambda_1|} < 1
\end{aligned}$$

The smaller $\frac{|\lambda_2|}{|\lambda_1|}$ *is, the faster it converges.*

$$\left(\frac{|\lambda_1|}{|\lambda_i|}\right)^k \xrightarrow{k\to\infty} 0$$

Spectral gap *is the difference between the moduli of the two largest eigenvalues of a matrix.*

14 Matrix norms

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

Norm on a vector space is a map $||\cdot|| : V \to [0, \infty[$:

1. $||v|| = 0 \iff v = 0$
2. $||\lambda v|| = |\lambda| \cdot ||v||$
3. $||v + w|| \leq ||v|| + ||w||$

Definition 14.1. *Let V, W be vector spaces with norms $||\cdot||_V$ and $||\cdot||_W$. A norm on* $\mathrm{Hom}(V, W)$ *is* compatible *with $||\cdot||_V$ and $||\cdot||_W$ if*

$$\forall v \in V \forall f \in \mathrm{Hom}(V, W) : \left\|f(v)\right\|_W \leq \left\|f\right\| \cdot ||v||_V$$

$$v = x - y \implies \left\|f(x) - f(y)\right\|_W = \left\|f(x - y)\right\|_W \leq \left\|f\right\| \cdot \left\|x - y\right\|_V$$

Hence f is Lipschitz continuous with constant $\leq \left\|f\right\|$. Specifically, we define the Lipschitz constant uniquely with

$$\inf\left\{C \mid \left\|f(v)\right\|_W \leq C \cdot ||v||_W \ \forall v \in V\right\}$$

Example 14.2. *In $\mathbb{C}^{n\times n}$, one scalar product is defined by $\langle A, B\rangle := \mathrm{Tr}(B^*A)$. The Frobenius norm (or Schur norm) (or Hilbert-Schmidt norm) is defined as*

$$||A||_F := \sqrt{\mathrm{Tr}(A^*A)} = \left(\sum_{ij} \left|a_{ij}\right|^2\right)^{\frac{1}{2}}$$

It is compatible with the Euclidean norm on $\mathbb{C}^n$.

Let $x \in \mathbb{C}^n$.

$$||Ax||_2^2 = \sum_{i=1}^{n} \left|(Ax)_i\right|^2 = \sum_{i=1}^{n} \left|\sum_j a_{ij}x_j\right|^2$$

$$\text{(CBS inequality, Theorem 8.20)} \leq \sum_{i=1}^{n} \sum_j \left|a_{ij}\right|^2 \cdot \sum_k |x_k|^2 = ||A||_F^2 : ||x||_2^2$$

Lemma 14.3. *Let V, W be normed vectorspaces. $f \in \mathrm{Hom}(V, W)$. Then $\left\|f\right\|_{V,W}$ defines a compatible norm and is called* induced norm.

$$\left\|f\right\|_{V,W} := \inf\left\{C \mid \left\|f(v)\right\|_W \leq C \cdot ||v||_V \ \forall v \in V\right\}$$

And specifically,

$$\left\|f\right\|_{V,W} = \sup\left\{ \frac{\left\|f(v)\right\|_W}{\|v\|} \,\middle|\, v \in V \setminus \{0\} \right\}$$
$$= \sup\left\{ \left\|f(v)\right\|_W \,\middle|\, v \in V, \|v\| = 1 \right\}$$

Proof. 1.

$$\sup_{v\neq 0} \frac{\left\|f(v)\right\|_W}{\|v\|_V} = \sup_{\|v\|_V=1} \left\|f(v)\right\|_W \text{ is immediate}$$

2. Let $M_f = \inf\left\{ C \mid \left\|f(v)\right\|_W \leq \|v\|_V \,\forall v \in V \right\}$. Show that $M_f = \sup_{v\neq 0} \frac{\|f(v)\|_W}{\|v\|_V}$.

Let $C \geq M_f$. Then $\left\|f(v)\right\|_W \leq C \cdot \|v\|_V \quad \forall v \in V$.

$$\implies \frac{\left\|f(v)\right\|_W}{\|v\|_V} \leq C \quad \forall v \neq 0 \qquad \implies \sup_{v\neq 0} \frac{\left\|f(v)\right\|}{\|v\|_V} \leq C \qquad \implies \sup \frac{\left\|f(v)\right\|}{\|v\|} w \leq M_f$$

□

↓ *This lecture took place on 2018/06/25.*

A norm on $\mathbb{K}^{m\times n}$ (with $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$) is *compatible* with norms on $\mathbb{K}^n$ and $\mathbb{K}^m$ if

$$\underbrace{\|Ax\|_m}_{\substack{\text{vector norm}\\ \text{on } \mathbb{K}^m}} \leq \underbrace{\|A\|_{m\times n}}_{\text{matrix norm}} \cdot \underbrace{\|x\|_n}_{\substack{\text{vector space}\\ \text{on } \mathbb{K}^n}}$$

$$\|A\|_F = \operatorname{Tr}(A^* A)^{\frac{1}{2}} = \left(\sum \left|a_{ij}\right|^2\right)^{\frac{1}{2}}$$

is compatible with the Euclidean norm.

Optimal norm on $\mathbb{K}^{m\times n}$.

$$\|A\| = \inf\{C > 0 \mid \forall x \in \mathbb{K}^n : \|A \cdot x\|_m \leq C \cdot \|x\|_n\} := \sup_{\substack{x\\ \|x\|_n \leq 1}} \|Ax\|_m$$

$$f : V \to W \qquad \left\|f\right\| = \sup_{\substack{x\in V\\ \|x\|_V \leq 1}} \left\|f(x)\right\|_W$$

Exercise for the practicals:

$$\underbrace{V}_{\|\cdot\|_V} \xrightarrow{f} \underbrace{W}_{\|\cdot\|_W} \xrightarrow{g} \underbrace{Z}_{\|\cdot\|}$$

Example 14.4. 1. *The Frobeniusnorm $\|A\|_F = \mathrm{Tr}(A^*A)^{\frac{1}{2}}$ is not optimal.*

$$\mathrm{id} : \mathbb{C}^n \to \mathbb{C}^n \qquad \|\mathrm{id}\|_{2\to 2} = \sup_{\|x\|_2 \leq 1} \|x\|_2 = 1 \qquad \mathrm{id}_F = \|I\|_F = \mathrm{Tr}(I^2)^{\frac{1}{2}} = \sqrt{n}$$

2. *The norm induced by the Euclidean norm*

$$\begin{aligned}
\|A\|_{2\to 2} &= \sup \{ \|Ax\|_2 \mid \|x\|_2 \leq 1 \} \\
&= \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{x\neq 0} \frac{\langle Ax, Ax\rangle^{\frac{1}{2}}}{\langle x, x\rangle^{\frac{1}{2}}} \\
&= \sup_{x\neq 0} \frac{\langle A^*Ax, x\rangle^{\frac{1}{2}}}{\langle x, x\rangle^{\frac{1}{2}}} = \sqrt{\sup_{x\neq 0} \frac{\langle A^*Ax, x\rangle}{\langle x, x\rangle}} \\
&= \sqrt{\textit{largest eigenvalue of } A^*A} = \sqrt{\textit{largest singular value of } A}
\end{aligned}$$

3. $\|A\|_{\infty\to\infty}$ *on* $\mathbb{K}^n : \|x\|_\infty = \max |x_i|$ *and* $\|x\|_1 = \sum |x_i|$.

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \xrightarrow{p\to\infty} \max |x_i| \qquad (1 \leq p \leq \infty)$$

This convergence is left as an exercise to the reader.

$$\|A\|_{\infty\to\infty} = \sup \{ \|Ax\|_\infty \mid \|x\|_\infty \leq 1 \}$$

$$\begin{aligned}
\|Ax\|_\infty &= \max_i \left|(Ax)i\right| = \max_i \left| \sum_{j=1}^n a_{ij} x_j \right| \\
&\leq \max_i \sum_{j=1}^n \left|a_{ij}\right| \underbrace{\left|x_j\right|}_{\leq \max_j |x_j|} \leq \underbrace{\max_j \left|x_j\right|}_{=\|x\|_\infty} \max_i \sum_{j=1}^n \left|a_{ij}\right| \\
&\implies \forall x \in \mathbb{K}^n : \|Ax\|_\infty \leq \max_i \sum_j \left|a_{ij}\right| \cdot \|x\|_\infty \\
&\implies \|A\|_{\infty\to\infty} \leq \max_i \sum_j \left|a_{ij}\right|
\end{aligned}$$

Claim. $\|A\|_{\infty\to\infty} = \max_i \sum_j \left|a_{ij}\right|$

Proof. Find vector $\tilde{x}$ such that $\|A\tilde{x}\|_\infty = \max_i \sum_j \left|a_{ij}\right| \cdot \|\tilde{x}\|_\infty$. Choose i_0 such that $\sum_i \left|a_{ij}\right| = \max!$.

$$\tilde{x}_j = \begin{cases} \frac{|a_{i_0 j}|}{a_{i_0 j}} & a_{i_0 j} \neq 0 \\ 0 & \text{else} \end{cases}$$

$\tilde{x}_j$ are not all zero, $\left|\tilde{x}_j\right| \in \{0,1\}\,\forall j$.

$$\begin{aligned}(A\cdot\tilde{x})_{i_0} &= \sum_j a_{i_0 j}\tilde{x}_j = \sum_j a_{i_0 j} a_{i_0 j}\frac{\left|a_{i_0 j}\right|}{a_{i_0 j}} = \sum_j \left|a_{i_0 j}\right| = \max_i \sum \left|a_{ij}\right| \\ &\implies \|A\tilde{x}\|_\infty \geq \left|(A\tilde{x})_{i_0}\right| = \max_i \sum \left|a_{ij}\right| \cdot \underbrace{\|\tilde{x}\|_\infty}_{=1} \\ &\implies \|A\|_{\infty\to\infty} \geq \max_i \sum_j \left|a_{ij}\right| \\ &\implies \|A\|_{\infty\to\infty} = \max_i \sum_j \left|a_{ij}\right| = \max\{\|z_i\|_1 \mid z_i \text{ row of } A\}\end{aligned}$$

□

4. $\|A\|_{1\to 1} = \max_j \sum_i \left|a_{ij}\right|$. *The proof is left as an exercise to the reader.*

$$\begin{aligned}Ax = \lambda x &\implies \|Ax\| = |\lambda|\cdot\|x\| \\ &\implies \sup_{x\neq 0}\frac{\|Ax\|}{\|x\|} \geq |\lambda|\end{aligned}$$

Definition 14.5.

$$\rho(A) := \max\{|\lambda| \mid \lambda \in \operatorname{spec}(A)\} \quad \textit{is called } \text{spectral radius } \textit{of } A.$$

Remark. *Why is it called radius? Consider the complex plane. This value is the radius of the smallest circle with center* 0 *that contains all eigenvalues.*

Lemma 14.6.

1. *For every compatible matrix norm:* $\|A\| \geq \rho(A)$
2. $\rho(A)$ *is not a matrix norm (for some nilpotent matrix,* $\rho(A) = 0$ *but* $A \neq 0$*, hence it cannot be a norm)*
3. $\forall A \in \mathbb{C}^{n\times n}\,\forall \varepsilon > 0\,\exists$ *norm on* $\mathbb{C}^n$ *: the induced matrix norm satisfies* $\|A\| \leq \rho(A) + \varepsilon$

Proof of point 3.

$$\exists \text{ invertible } T \in \mathbb{C}^{n\times n} : T^{-1}AT = J = \begin{bmatrix}\lambda_1 & \eta_k & & & 0 \\ & \ddots & \eta_{23} & & \\ & & \ddots & \ddots & \\ & & & \ddots & \eta_{n-1,n} \\ 0 & & & & \lambda_n\end{bmatrix} \qquad \eta_{ij} \in \{0,1\}$$

$$D_\varepsilon = \operatorname{diag}(1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1})$$

$$D_\varepsilon^{-1} J D_\varepsilon = \operatorname{diag}(1, \frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}, \ldots, \frac{1}{\varepsilon^{n-1}})\begin{bmatrix}\lambda_1 & \eta_{12} & & & \\ & \lambda_2 & \eta_{23} & & \\ & & \ddots & & \\ & & \lambda_{n-1} & \eta_{n-1,n} & \\ & & & \eta_n & \end{bmatrix}\operatorname{diag}(1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1})$$

$$
= \begin{bmatrix}
\lambda_1 & \eta_{12}\varepsilon & & & & \\
& \frac{\lambda_2}{\varepsilon}\varepsilon & \frac{\eta_{23}}{\varepsilon}\varepsilon^2 & & & \\
& & \frac{\varepsilon_3}{\varepsilon^2}\varepsilon^2 & \frac{\eta_{34}}{\varepsilon^2}\varepsilon^3 & & \\
& & & \ddots & & \\
& & & & \ddots & \eta_{n-1}\varepsilon \\
& & & & & \lambda_n
\end{bmatrix}
$$

Define a norm on $\mathbb{C}^n$.

$$\|x\|_\varepsilon = \left\|D_\varepsilon^{-1}T^{-1}x\right\|_\infty$$

1. $\|x\|_\varepsilon \geq 0$ is immediate.
   $\|x\|_\varepsilon = 0 \implies D_\varepsilon^{-1}T^{-1}x = 0 \implies x = 0$
2. $\|\lambda x\|_\varepsilon = \left\|D_\varepsilon^{-1}T^{-1}\lambda x\right\| = |\lambda| \cdot \|x\|_\varepsilon$
3. $\left\|x+y\right\|_\varepsilon = \left\|D_\varepsilon^{-1}T^{-1}(x+y)\right\|_\varepsilon \leq \left\|D_\varepsilon^{-1}T^{-1}x\right\|_\infty + \left\|D_\varepsilon^{-1}T^{-1}y\right\|_\infty = \|x\|_\varepsilon + \left\|y\right\|_\varepsilon$

$$\|A\|_{\varepsilon\to\varepsilon} = \sup_{x\neq 0} \frac{\|Ax\|_\varepsilon}{\|x\|_\varepsilon}$$

$$x = \underbrace{TD_\varepsilon}_{\text{invertible}} y = \sup_{y\neq 0} \frac{\left\|ATD_\varepsilon y\right\|_\varepsilon}{\left\|TD_\varepsilon y\right\|_\varepsilon} = \sup_{y\neq 0} \frac{\left\|D_\varepsilon^{-1}T^{-1}\left(ATD_\varepsilon y\right)\right\|_\infty}{\left\|D_\varepsilon^{-1}T^{-1}\left(TD_\varepsilon y\right)\right\|_\infty} = \sup_{y\neq 0} \frac{\left\|J_\varepsilon \cdot y\right\|_\infty}{=} \sup_{y\neq 0} \frac{\left\|J_\varepsilon y\right\|_\infty}{\left\|y\right\|_\infty}$$

$$\leq \|J_\varepsilon\|_{\infty\to\infty} = \max_i \sum_j \left|(J_\varepsilon)_{ij}\right| \leq \underbrace{\max|\lambda_i| + \varepsilon}_{=\rho(A)+\varepsilon}$$

□

Person. *Israel Gelfand (1913–2009)*

Remark 14.7.

$$\rho(A) \leq \|A\|$$

$$\rho(A^2) = \rho(A)^2 \qquad \rho(A^3) = \rho(A)^3$$

$$\operatorname{spec}(A^2) = \left\{\lambda^2 \,\middle|\, \lambda \in \operatorname{spec}(A)\right\}$$

$$\max(|\lambda_i|^2) = (\max|\lambda_i|)^2$$

$$\rho(A^2) \leq \|A\|^2 \quad (\leq \|A\|)^2$$

$$\rho(A^3) \leq \|A\|^3 \qquad \underbrace{\rho(A^k)}_{\rho(A)^k} \leq \left\|A^k\right\|$$

$$\implies \rho(A) \leq \left\|A^k\right\|^{\frac{1}{k}} \; \forall k$$

Geldfand showed:

$$\lim_{k\to\infty} \left\|A^k\right\|^{\frac{1}{k}} = \rho(A)$$

Lemma 14.8. 1. $||A||_{2\to 2} = \rho(A^*A)^{\frac{1}{2}}$ *is the* spectral norm

2. *If A is normal, then $||A||_{2\to 2} = \rho(A)$*

Proof. 1. Already shown

2. $\operatorname{spec}(A^*A) = \left\{|\lambda|^2 \,\middle|\, \lambda \in \operatorname{spec}(A)\right\}$ is left to be proved by the reader. Then $= \left\{\overline{\lambda} \cdot \lambda \,\middle|\, \lambda \in \operatorname{spec}(A)\right\} \implies \rho(A^*A) = \rho(A)^2$

□

Remark 14.9. *Let V be a vector space with dimension ∞ with a norm. For example, $l^2 = \left\{(x_n) \in \mathbb{R}^\infty \,\middle|\, \sum_1^\infty |x_n|^2 < \infty\right\}$ with norm $\left\|(x_n)\right\| = \left(\sum_1^\infty |x_n|^2\right)^{\frac{1}{2}}$ is a* Hilbert space.

$$l^\infty = \left\{(x_n) \,\middle|\, \sup |x_n| < \infty\right\} \qquad \left\|(x_n)\right\|_\infty = \sup |x_n|$$

Example.

$$V = C^\infty[0, 1]$$

$$\frac{d}{dx} : f \mapsto f' \text{ is linear}$$

$\left\|\frac{d}{dx}\right\| = \infty$ *it does not matter which norm on $C^\infty[0, 1]$ is defined.*

$$\rho\left(\frac{d}{dx}\right) = \infty \qquad f' = \lambda f \qquad f(x) = e^{\lambda x} \in C^\infty \implies \operatorname{spec}\left(\frac{d}{dx}\right) = \mathbb{C}$$

Person. *Carl Neumann (1832–1925)*

Corollary 14.10. *For compatible norms, if $||A|| < 1$ then $I - A$ is invertible and $(I - A)^{-1} = \frac{1}{I-A} = \sum_{n=0}^\infty A^n$ and $\left\|(I - A)^{-1}\right\| \leq \frac{1}{1-||A||}$*

1. $\rho(A) \leq ||A|| < 1 \implies 1 \notin \operatorname{spec}(A) \implies I - A$ *invertible*
2. *Neumann series:* $\frac{1}{1-x} = \sum_{n=0}^\infty x^n$ *with* $|x| < 1$

Remark.

$$y' = \lambda y \qquad \int y' = \lambda \cdot \int y \implies \int y = \frac{1}{\lambda} \cdot y$$

Every differentiation can be converted into integration. An integration is bounded (λ becomes $\frac{1}{\lambda}$).

Proof. $\rho(A) < 1 \implies I - A$ invertible.

Claim.

$$\sum_{n=0}^{\infty} A^n \text{ converges}$$

$$\sum \sum_{n=1}^{N+m} A^n - \sum_{n=1}^{N} A^n = \left\| \sum_{n=N+1}^{N+m} A^n \right\| \leq \sum_{n=N+1}^{N+m} \|A^n\| \leq \sum_{n=N+1}^{N+m} \|A\|^n \leq \sum_{n=N+1}^{\infty} \|A\|^n$$
$$\leq \frac{\|A\|^{N+1}}{1-\|A\|} \xrightarrow{N\to\infty} 0$$

Thus, the sequence of partial sums is Cauchy and therefore convergent.

$$(I-A)\sum_{n=0}^{\infty} A^n = \sum_{n=0}^{\infty} A^n - \underbrace{\sum_{n=0}^{\infty} A^{n+1}}_{=\sum_{n=1}^{\infty} A^n} = A^0 = I$$

□

$$Ax = \tilde{b} = b + \text{ error}$$
$$\tilde{x} = A^{-1}\tilde{b} \qquad x = A^{-1}b \qquad \tilde{x} - x = A^{-1}(\tilde{b} - b) \implies \|\tilde{x} - x\| \leq \left\|A^{-1}\right\| \cdot \left\|\tilde{b} - b\right\|$$
$$\|\tilde{x} - x\| \overset{?}{\leq} C \cdot \left\|\tilde{b} - b\right\|$$

Corollary 14.11. *Let A be invertible and B be arbitrary with $\|B\| < \left\|A^{-1}\right\|^{-1}$. $\implies$ $A+B$ is invertible and*

$$\left\|(A+B)^{-1}\right\| \leq \frac{\left\|A^{-1}\right\|}{1 - \|A^{-1} \cdot \|B\|\|}$$

$$\|A\| \sim \max |\lambda_i|$$
$$\left\|A^{-1}\right\| \sim \max \frac{1}{|\lambda_i|} = \frac{1}{\min |\lambda_i|}$$
$$\left\|A^{-1}\right\|^{-1} = \min |\lambda_i|$$

Compare with $A = I$ as special case (this was covered in Corollary 14.10).

$$\|B\| < 1 \qquad \left\|(I+B)^{-1}\right\| \leq \frac{1}{1-\|B\|}$$

$$A + B = A \cdot (I + A^{-1}B)$$
$$(A+B)^{-1} = (I + A^{-1}B)^{-1} \cdot A^{-1}$$

$I + A^{-1}B$ is invertible because $\left\|A^{-1}B\right\| \leq \left\|A^{-1}\right\| \cdot \|B\| < 1$.

$$\implies \left\|(A+B)^{-1}\right\| \leq \left\|(I+A^{-1}B)^{-1}\right\| \cdot \left\|A^{-1}\right\| \leq \frac{1}{1-\|A^{-1}B\|}\left\|A^{-1}\right\| \leq \frac{1}{1-\|A^{-1}\|\,\|B\|} \cdot \left\|A^{-1}\right\|$$

$$x < y \qquad 1 - x > 1 - y$$
$$\frac{1}{1-x} < \frac{1}{1-y}$$

↓ *This lecture took place on 2018/06/30.*

$$\|A\| < 1 \implies I - A \text{ invertible}$$

$$(I - A)^{-1} = \sum_{n=0}^{\infty} A^n$$

$$\limsup \|A^n\|^{\frac{1}{n}} = g(A) < 1$$

$\sum a_n$ converges absolutely if $\limsup |a_n|^{\frac{1}{n}} < 1$. So it holds.

A_0 is invertible, $\|A\| < \left\|A_0^{-1}\right\|^{-1}$.

$$A_0 + A \text{ invertible}$$

$$\left\|(A_0 + A)^{-1}\right\| \leq \frac{\left\|A_0^{-1}\right\|}{1 - \left\|A_0^{-1}\right\| \cdot \|A\|} < 1$$

Remark 14.12 (Sensitivity of linear equation systems). *Instead of $Ax = b$, we consider $\tilde{A}\tilde{x} = \tilde{b}$ with $\left\|A - \tilde{A}\right\|$ and $\left\|b - \tilde{b}\right\|$ as "small" values. Does this imply $\|x - \tilde{x}\|$ is small?*

Let A be a invertible matrix and $\|\cdot\|$ is a compatible matrix norm. Let x be the unique solution of $Ax = b$. Let $\tilde{x}$ be the unique solution of $A\tilde{x} = \tilde{b}$.

$$\implies \tilde{b} = b + \underbrace{\triangle b}_{\text{error}}$$

$$\triangle x = \tilde{x} - b = \text{error in the solution}$$

The relative error is interesting. The error relative to the solution is given by:

$$\frac{\|\triangle x\|}{x} \leq \underbrace{\|\triangle\|}_{\substack{\text{largest} \\ \text{singular value}}} \cdot \underbrace{\left\|A^{-1}\right\|}_{\substack{\text{reciprocal of} \\ \text{smallest} \\ \text{singular value}}} \cdot \frac{\|\triangle b\|}{\|b\|}$$

Definition 14.13.

$$\kappa(A) := \|A\| \cdot \left\|A^{-1}\right\| \geq 1$$

is called condition number of A. *If $\kappa(A)$ is large, then the problem is ill-conditioned.*

Example 14.14. *For diagonal matrices:*

$$\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

$$\|\Lambda\|_{2\to 2} = \max |\lambda_i|$$

$$\left\|\Lambda^{-1}\right\|_{2\to 2} = \frac{1}{\min |\lambda_i|}$$

$$\implies \kappa(A) = \frac{\max |\lambda_i|}{\min |\lambda_i|}$$

Proof of Remark 14.12.

$$Ax = b \qquad A\tilde{x} = \tilde{b}$$

$$\begin{aligned} A(x + \triangle b) &= b + \triangle b \\ \iff Ax + A\triangle x &= b + \triangle b \\ \iff A\triangle x &= \triangle b \\ \implies \triangle x &= A^{-1}\triangle b \\ \implies \|\triangle x\| &\leq \left\|A^{-1}\right\| \cdot \|\triangle b\| \end{aligned}$$

$$\begin{aligned} \|b\| = \|Ax\| &\leq \|A\| \cdot \|x\| \\ \implies \frac{1}{\|x\|} &\leq \frac{\|A\|}{\|b\|} \\ \implies \frac{\|\triangle x\|}{\|x\|} &= \left\|A^{-1}\right\| \cdot \|\triangle b\| \cdot \frac{\|A\|}{\|b\|} \end{aligned}$$

□

Remark 14.15 (General case)**.**

$$\tilde{A} = A + \triangle A \qquad Ax = b$$

$$\tilde{b} = b + \triangle b \qquad \tilde{A}\tilde{x} = \tilde{b}$$

$$\triangle\tilde{x} = \tilde{x} + x$$

Requirement:

$$\begin{aligned} \|\triangle A\| &< \left\|A^{-1}\right\|^{-1} \text{ such that } \tilde{A} \text{ invertible} \\ \implies \frac{\|\triangle A\|}{\|A\|} &\leq \frac{1}{\|A\| \cdot \|A^{-1}\|} = \frac{1}{\kappa(A)} \end{aligned}$$

Then this modified system becomes solvable.
All these times, we use the inequality:

$$\|A \cdot B\| \leq \|A\| \cdot \|B\|$$

$$\|A \cdot x\| \leq \|A\| \cdot \|x\|$$

$$(A + \triangle A)(x + \triangle x) = b + \triangle b$$

$$\begin{aligned} \iff (I + A^{-1}\triangle A)(x + \triangle x) &= x + A^{-1}\triangle b \qquad (4) \\ \iff x + \triangle x + A^{-1}\triangle Ax + A^{-1}\triangle A\triangle x &= x + A^{-1}\triangle b \\ \iff \triangle x &= A^{-1}\triangle b - A^{-1}\triangle A(x + \triangle x) \end{aligned}$$

$$\|\triangle x\| \leq \underbrace{\left\|A^{-1}\right\| \cdot \|\triangle b\|}_{Eq. (5)} + \underbrace{\left\|A^{-1}\right\| \cdot \|\triangle A\| \cdot \|x + \triangle x\|}_{Eq. (6)}$$

$$
\begin{aligned}
\left\|A^{-1}\right\| \cdot \|\triangle b\| &= \frac{\left\|A^{-1}\right\| \cdot \|b\| \cdot \|\triangle b\|}{\|b\|} \\
&\leq \frac{\left\|A^{-1}\right\| \cdot \|Ax\| \cdot \|\triangle b\|}{\|b\|} \\
&\leq \left\|A^{-1}\right\| \cdot \|A\| \cdot \|x\| \cdot \frac{\|\triangle b\|}{\|b\|} \\
&\leq \kappa(A) \cdot \|x\| \cdot \frac{\|\triangle b\|}{\|b\|}
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
\|x + \triangle x\| &\overset{(4)}{=} \left\|(1 + A^{-1}\triangle A)^{-1} \cdot (x + A^{-1}\triangle b)\right\| \\
&\leq \left\|(1 + A^{-1}\triangle A)^{-1}\right\| \cdot \left(\|x\| + \left\|A^{-1}\right\| \cdot \|\triangle b\|\right) \\
&\leq \frac{1}{1 - \|A^{-1}\triangle A\|} \cdot \|x\| \cdot \left(1 + \|x\| \cdot \left\|A^{-1}\right\| \cdot \|b\| \cdot \frac{\|\triangle b\|}{\|b\|}\right) \\
&\leq \frac{1}{1 - \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\triangle A\|}{\|A\|}} \cdot \|x\| \cdot \left(1 + \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|}\right) \\
&\leq \frac{1}{1 - \kappa(A) \cdot \frac{\|\triangle A\|}{\|A\|}} \cdot \|x\| \cdot \left(1 + \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|}\right) \\
&\leq \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|} \cdot \|x\| + \frac{1}{1 - \kappa(A) \cdot \frac{\|\triangle A\|}{\|A\|}} \left(1 + \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|}\right) \cdot \|x\| \qquad (7)
\end{aligned} \tag{6}
$$

$$
\iff \frac{\|\triangle x\|}{\|x\|} \leq \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|} + \frac{1 + \kappa(A) \cdot \frac{\|\triangle b\|}{\|b\|}}{1 - \kappa(A) \cdot \frac{\|\triangle A\|}{\|A\|}}
$$

Remark 14.16 (Sensitivity of eigenvalues)**.** *Let $A \in \mathbb{C}^{n \times n}$, diagonalizable.*

$$\to B^{-1}AB = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

$$\tilde{A} = A + \triangle A$$

$$\lambda \in \operatorname{spec}(\tilde{A}) \implies \exists j : \left|\lambda - \lambda_j\right| \leq \|B\| \cdot \left\|B^{-1}\right\| \cdot \|\triangle A\|$$

($\|\cdot\| = \|\cdot\|_{2\to 2}$ or $\|\cdot\|_{\infty\to\infty}$)

Lemma 14.17 (Special case)**.**

$$A \text{ normal} \implies B \text{ unitary} \implies \text{isometric} \qquad \|u_x\| = \|x\|\, \forall x$$

$$\implies \|B\| = 1, \left\|B^{-1}\right\| = 1$$

Proof. 1. If $\lambda \in \operatorname{spec}(A)$, then nothing to show.

2. If $\lambda \notin \operatorname{spec}(A)$, then $A - \lambda I$ is invertible, $\tilde{A} - \lambda I$ is non-invertible.

$$
\begin{aligned}
\underbrace{\tilde{A} - \lambda I}_{\text{not invertible}} &= (A - \triangle A) - \lambda I \\
&= (A - \lambda I)(A - \lambda I)^{-1} \cdot (A - \lambda I + \triangle A) \\
&= (A - \lambda I)(I + (A - \lambda I)^{-1} \triangle A) \\
&\implies I + (A - \lambda I)^{-1} \cdot \triangle A \text{ is } \textit{not} \text{ invertible} \\
&\xrightarrow{\text{Neumann negated}} \left\| (A - \lambda I)^{-1} \cdot \triangle A \right\| \geq 1
\end{aligned}
$$

$$
\begin{aligned}
1 &\leq \left\| (A - \lambda I)^{-1} \triangle A \right\| \\
&= \left\| (B \Lambda B^{-1} \cdot \lambda B B^{-1})^{-1} \cdot \triangle A \right\| \\
&= \left\| \left(B \left(\Lambda - \lambda I \right) B^{-1} \right)^{-1} \cdot \triangle A \right\| \\
&= \left\| (B(\Lambda - \lambda I^{-1}) B^{-1}) \triangle A \right\| \\
&\leq \|B\| \cdot \left\| (\Lambda - \lambda I)^{-1} \right\| \cdot \left\| B^{-1} \right\| \cdot \|\triangle A\| \\
&= \|B\| \cdot \left\| B^{-1} \right\| \cdot \frac{1}{\min |\lambda_i - \lambda|} \cdot \|\triangle A\| \\
&\implies \min |\lambda_i - \lambda| \leq \kappa(B) \cdot \|\triangle A\|
\end{aligned}
$$

Recall that,

$$
\begin{aligned}
(\Lambda - \lambda I)^{-1} &= \begin{bmatrix} \lambda_1 - \lambda & & \\ & \ddots & \\ & & \lambda_n - \lambda \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \frac{1}{\lambda_1 - \lambda} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n - \lambda} \end{bmatrix} \\
\left\| (\Lambda - \lambda I)^{-1} \right\| &= \max \left| \frac{1}{\lambda_i - \lambda} \right| = \frac{1}{\min |\lambda_i - \lambda|}
\end{aligned}
$$

□

15 Non-negative matrices

Definition 15.1. *$A \in \mathbb{K}^{n \times n}$ is called* non-negative *if $a_{ij} \geq 0 \forall i, j$. We denote $A \geq 0$. Do not mix this up with positive definiteness!*

Example (Markov chains)**.**

$$a_{ij} = W_s k$$

Manhattan: $a_{ij} = W_s k$ that you can reach node j from node i.

$$a_{ij} \geq 0 \forall i, j$$

For fixed i: $\sum_j a_{ij} = 1$.

Matrix A is called row-stochastic.

A has eigenvector:

$$A \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\implies 1 \textit{ is eigenvalue, } \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} =: \Gamma \textit{ is eigenvalue of } 1$$

A directed graph is called strongly connected *if you can reach every node from each other.*

Theorem 15.2 (Perron–Frobenius theorem)**.** *If the graph is strongly connected, then eigenvalue* 1 *has geometric and algebraic multiplicity* 1*, all the other eigenvalues satisfy* $|\lambda| < 1$.

$$A_v^n \rightarrow_{n\rightarrow\infty} \mathbf{1} \forall v$$

Index