lecture 1

> This notes are based on my personal notes. For an official one, please refer to Piazza.

Question: What is a set?

It is very interesting that rigorously defining the notion of set is a very challenging thing. There is a Wikipedia article for set. In particular, it is very interesting to read about *Cantor's definition* and *Russel's paradox*.

Question: What is a function between two sets?

A function from a set S to a set T, is a subset F of the set $S \times T$ with the following property: whenever $(s_1, t_1), (s_2, t_2) \in F$ with $s_1 = s_2$, then it is also true that $t_1 = t_2$. The function is represented by the notation $f : S \to T$, where the symbol f is defined as $f(s) := t \Leftrightarrow (s, t) \in F$.

Interestingly, this definition requires only the notion of a set. Contrast this to less rigorous definitions, which define a function to be a mapping or a rule from S to T. But then these definitions do not define what is a mapping or a rule.

Definition: A group $(G, *)$ is defined by a set G together with a binary operation $*$, such that the following four properties are satisfied: (Wiki)

- **Closure:** For all a and b in G, $a * b$ is also in G.
- **Associativity:** For all a, b and c in G, we have $(a * b) * c = a * (b * c)$.
- **Identity:** There is some element e from G, called the identity element, such that $e * a = a$ and $a * e = a$ for all a in G.
- **Inverses:** For each a in G, there must exist a unique element b in G, called the inverse of a, that satisfies $a * b = b * a = e$, where e is the identity element.

An abelian group has one more property: (Wiki)

- **Commutativity:** For all a and b in G, $a * b = b * a$.

Example: $(\mathbb{R}^{m\times n}, +)$ is a group, because it satisfy the following properties:

- **Closure:** For all $A, B \in \mathbb{R}^{m\times n}$, $A + B$ is also in $\mathbb{R}^{m\times n}$.
- **Associativity:** For all $a, b, c \in \mathbb{R}^{m\times n}$, we have $(A + B) + C = A + (B + C)$.
- **Identity:** There is a element $0_{m\times n} \in \mathbb{R}^{m\times n}$, such that $0_{m\times n} + A = A$ and $A + 0_{m\times n} = A$ for all A in $\mathbb{R}^{m\times n}$.
- **Inverses:** For each A in $\mathbb{R}^{m\times n}$, there exists a unique element $-A$ in $\mathbb{R}^{m\times n}$ that satisfies $A + (-A) = (-A) + A = 0_{m\times n}$, where $0_{m\times n}$ is the identity element.

If $m = n$, then $\mathbb{R}^{n\times n}$ is already a group under addition but is not a group under multiplication.

We define matrix multiplication as $(A \cdot B)_{ij} = \sum_{k=1}^{n} A_{ik} \cdot B_{kj}$ for integer $i, j \in [1, n]$. The identity element is $e = I_{n\times n}$ since for all $A \in \mathbb{R}^{n\times n}$, $A \cdot e = e \cdot A = A$ holds. However, for $A = 0_{n\times n}$, there is no $B \in \mathbb{R}^{n\times n}$ such that $A \cdot B = e$.

Proposition: If $(G, *)$ is a group, then the identity element e is unique.

Proof: We prove this by contradiction. If there are two different identity elements e and e', then according to identity property, $e = e * e' = e'$. This contradicts with the fact that e and e' are different.

Proposition: If $(G, *)$ is a group, then for all $g \in G$, g^{-1} is unique.

Proof: We prove this by contradiction. If there are two different inverses f and h for g, then according to inverses property, $g * f = f * g = e$ and $g * h = h * g = e$. Then $g * f = g * h$ holds. We multiply h to both sides.

$$\begin{aligned} h * g * f &= h * g * h \\ (h * g) * f &= (h * g) * h \\ e * f &= e * h \\ f &= h \end{aligned}$$

This contradicts with the fact that f and h are different.

Proposition: If $(G, *)$ is a group, then for all g and h in G, $(g * h)^{-1} = h^{-1} * g^{-1}$.

Proof: The inverse of $(g * h)$ is $(g * h)^{-1}$. We need to show $h^{-1} * g^{-1}$ is an inverse of $(g * h)$.

$$\begin{aligned} (g * h) * (h^{-1} * g^{-1}) &= g * (h * (h^{-1} * g^{-1})) \\ &= g * ((h * h^{-1}) * g^{-1}) \\ &= e \end{aligned}$$

By applying the uniqueness of inverse, we get $(g * h)^{-1} = h^{-1} * g^{-1}$.

Homework: Let G be a group with group operation $*$ and identity element e. Suppose that $\forall g \in G$ we have that $g * g = e$. Show that G is an abelian group.

Definition: Given two groups, $(G, *)$ and $(H, \cdot)$, a group homomorphism from $(G, *)$ to $(H, \cdot)$ is a function $f : G \to H$ such that for all u and v in G it holds that $f(u * v) = f(u) \cdot f(v)$ where the group operation on the left hand side of the equation is that of G and on the right hand side that of H. (Wiki)

$$\begin{tikzcd} (g_1, g_2) \arrow[r] \arrow[d, "{(f,f)}"] & g_1 * g_2 \arrow[d, "f"] \\ (f(g_1), f(g_2)) \arrow[r] & f(g_1 * g_2) \end{tikzcd}$$

Figure 1: Commutative diagram.

Homework: Let G, H be groups with group operations $*_G, *_H$ and identity elements e_G, e_H respectively. Let $\phi : G \to H$ be a group homomorphism. Show that $\phi(e_G) = e_H$ and that $\forall g \in G$ we have that $\phi(g^{-1}) = (\phi(g))^{-1}$.

Cont.

Definition: Two groups G and H are called isomorphic if there exist group homomorphisms $\phi : G \to H$ and $\psi : H \to G$, such that applying the two functions one after another in each of the two possible orders gives the identity functions of G and H. That is, $\phi(\psi(h)) = h$ and $\psi(\phi(g)) = g$ for any $g \in G$ and $h \in H$. (Wiki)

Example: Let $(G, +)$ and $(H, +)$ be two groups that $G \in \mathbb{R}^{m\times n}$ and $H \in^{n\times n}$. Let S be a fixed n by m matrix. Let ϕ be a function such that $\phi : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times n}$ and $\phi(A) = S \cdot A$. Then ϕ is a group homomorphism from $(G, +)$ to $(H, +)$.

$$\begin{aligned}\phi(A+B) &= S \cdot (A+B)\\ &= S\cdot A + S\cdot B\\ &= \phi(A) + \phi(B)\end{aligned}$$

If $m = n$ and S is invertible, then ϕ is a group isomorphic.

Homework: Let G, H, K be groups such that G is isomorphic to H, and H is isomorphic to K. Show that G is isomorphic to K.

Definition: A ring is a set R equipped with two binary operations $+$ and $\cdot$ satisfying the following three properties: (Wiki)

- R is an abelian group under $+$, meaning that:
  1. $(a+b)+c = a+(b+c)$ for all $a, b, c \in R$ ($+$ is associative).
  2. $a+b = b+a$ for all $a, b \in R$ ($+$ is commutative).
  3. There is an element 0 in R such that $a + 0 = a$ for all $a \in R$ (0 is the additive identity).
  4. For each $a \in R$ there exists $-a \in R$ such that $a + (-a) = 0$ ($-a$ is the additive inverse of a).
- R is a monoid under $\cdot$, meaning that:
  1. $(a\cdot b)\cdot c = a\cdot(b\cdot c)$ for all $a, b, c \in R$ ($\cdot$ is associative).
  2. There is an element 1 in R such that $a \cdot 1 = a$ and $1 \cdot a = a$ for all $a \in R$ (1 is the multiplicative identity).
- Multiplication is distributive with respect to addition:
  1. $a\cdot(b+c) = (a\cdot b)+(a\cdot c)$ for all $a, b, c \in R$ (left distributivity).
  2. $(b+c)\cdot a = (b\cdot a)+(c\cdot a)$ for all $a, b, c \in R$ (right distributivity).

Example: $\mathbb{R}^{n\times n}$ is a ring because it follows these properties:

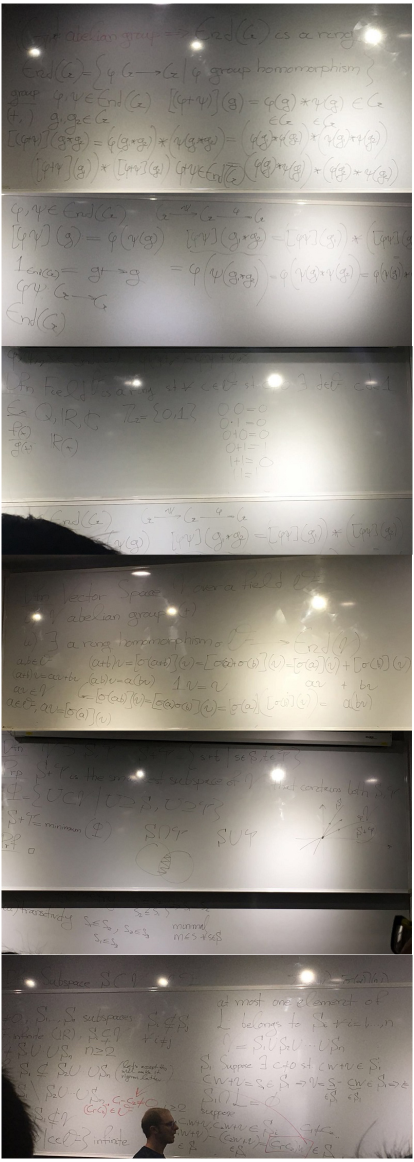
- $\mathbb{R}^{n\times n}$ is an abelian group under $+$.
- R is a monoid under $\cdot$, meaning that:
  1. $(A\cdot B)\cdot C = A\cdot(B\cdot C)$ for all $A, B, C \in \mathbb{R}^{n\times n}$.
  2. There is an identity matrix I in $\mathbb{R}^{n\times n}$ such that $A\cdot I = A$ and $I \cdot A = A$ for all $A \in \mathbb{R}^{n\times n}$.
- Multiplication is distributive with respect to addition:
  1. $A\cdot(B+C) = (A\cdot B)+(A\cdot C)$ for all $A, B, C \in \mathbb{R}^{n\times n}$.

Definition: A ring homomorphism is a function $f : R \to S$ between two rings R and S such that:

- $f(a +_R b) = f(a) +_S f(b)$ for all a and b in R.
- $f(a \cdot_R b) = f(a) \cdot_S f(b)$ for all a and b in R.
- $f(1_R) = 1_S$.

The End.

lecture 2



lecture 3

Dfn $V \supset S_1, S_2, \dots, S_n$ we say V "direct sum" of $S_1, \dots, S_n$

i) $V = S_1 + \dots + S_n$ $\quad V = S_1 \oplus S_2 \oplus \dots \oplus S_n$

ii) $S_i \cap (\sum_{j \neq i} S_j) = 0$ independent subspaces

disjoint subspaces $S_i \cap S_j = 0$

Ex 3 lines in $\mathbb{R}^2$

$l_i \cap l_j = 0$, $l_i + l_j = \mathbb{R}^2$

l_1, l_2, l_3, $\mathbb{R}^2$

$l_k \cap (l_i + l_j) = l_k$

(2)

$V = S_1 + \dots + S_n$ 1) $S_1, \dots, S_n$ independent

$\Leftrightarrow \nexists\, s_1 \in S_1, \dots, s_n \in S_n$ with some $s_i \neq 0$ s.t. $0 = s_1 + s_2 + \dots + s_n \Leftrightarrow \forall v \in V\ \exists!\ v = s_1 + \dots + s_n$ $\quad s_i \in S_i$

Prf 1) $\Rightarrow$ 2) suppose $\exists\, s_i \in S_i$ not all zero s.t. $0 = s_1 + \dots + s_n$ $\exists j$ s.t. $s_j \neq 0$

$s_1 \neq 0$ (without loss of generality) $s_1 = (-s_2) + (-s_3) + \dots + (-s_n) \Rightarrow s_1 = 0 \Rightarrow \Leftarrow$

(3)

2) $\Rightarrow$ 3) $v \in V$ $\quad v = s_1 + \dots + s_n = t_1 + \dots + t_n$, $s_i, t_i \in S_i \rightarrow (s_1 - t_1) + \dots + (s_n - t_n) = 0 \Rightarrow s_i = t_i$

3) $\Rightarrow$ 1) $v \in S_1 \cap (S_2 + \dots + S_n)$ $\quad v = s = s_2 + \dots + s_n$, $s_i \in S_i \rightarrow s_i = 0 \rightarrow v = 0$ $\square$

Dfn $V \supset \mathcal{T}$: subset (could be infinite) $\operatorname{span}(\mathcal{T}) = \{a_1 t_1 + \dots + a_n t_n \mid a_i \in F, t_i \in \mathcal{T}, n \in \mathbb{Z}_+\}$

"subspace spanned by $\mathcal{T}$" $=: \langle \mathcal{T} \rangle$

Dfn $V \supset \mathcal{T}$: subset $\mathcal{T}$ "spanning set" if $\operatorname{span}(\mathcal{T}) = V$

Dfn $\mathcal{T}$: subset V

$\mathcal{T}$ "linearly independent"

$a_1 t_1 + \dots + a_n t_n = 0 \Rightarrow a_i = 0$

$a_i \in F$, $t_i \in \mathcal{T}$

(4)

Thm (1.7/Roman) $\mathcal{T} \subset V$, $\mathcal{T}$: subset 1) $\mathcal{T}$ is l.i. and spans V, $\operatorname{span}(\mathcal{T}) = V$

$\Leftrightarrow$ 2) $\forall v \in V$, $v = a_1 t_1 + \dots + a_n t_n$ and a_i, t_i are unique

$\Leftrightarrow$ 3) $\mathcal{T}$ is minimal spanning set $\Leftrightarrow$ 4) $\mathcal{T}$ is maximal l.i.

Dfn We call such a $\mathcal{T}$ a basis for V

Prf 1) $\Rightarrow$ 2) $v \in V$ $\quad v = a_1 s_1 + \dots + a_n s_n$ $\quad s_i, t_j \in \mathcal{T}$

$= b_1 t_1 + \dots + b_m t_m$ $\quad a_i, b_j \in F$

$= b_1 s_1 + \dots + b_n s_n + b_{n+1} t_{n+1} + \dots + b_m t_m$ $\quad \{s_{n+1}, \dots, s_n\} \cap \{t_{n+1}, \dots, t_m\} = \phi$

$v = a_1 s_1 + \dots + a_n s_n + a_{n+1} s_{n+1} + \dots + a_n s_n$

$v = b_1 s_1 + \dots + b_n s_n + b_{n+1} t_{n+1} + \dots + b_m t_m$

(5)

$(a_1 - b_1) s_1 + \dots + (a_n - b_n) s_n + a_{n+1} s_{n+1} + \dots + a_n s_n - b_{n+1} t_{n+1} - \dots - b_m t_m = 0$ $\quad \mathcal{T}$ is l.i.

$\Rightarrow a_1 = b_1 \dots a_n = b_n$

$0 = a_{n+1} = \dots = a_n = b_{n+1} = \dots = b_m$

2) $\Rightarrow$ 3) suppose not minimal spanning set

$\rightarrow \exists\, t \in \mathcal{T}$ s.t. $\mathcal{T} \setminus \{t\}$ span set $\exists\, a_i \in F, t_i \in \mathcal{T} \setminus \{t\}$ s.t. $t = a_1 t_1 + \dots + a_n t_n \rightarrow \Leftarrow$

3) $\Rightarrow$ 4) suppose not max l.i. $\Rightarrow \exists\, v \in V$ s.t. $\mathcal{T} \cup \{v\}$ is l.i. but $\operatorname{span}(\mathcal{T}) = V \Rightarrow v = a_1 t_1 + \dots + a_n t_n$

$\Rightarrow v - a_1 t_1 - \dots - a_n t_n = 0 \Rightarrow \Leftarrow$ $\quad$ 4) $\Rightarrow$ 1) $\operatorname{span}(\mathcal{T}) \neq V \Rightarrow \exists\, v \notin \operatorname{span}(\mathcal{T})$

$\Rightarrow \mathcal{T} \cup \{v\}$ is l.i. for if not $av + b_1 t_1 + \dots + b_s t_s = 0$ not all $a, b_1, \dots, b_s$ are zero

claim $\square$

if $a = 0 \Rightarrow b_i = 0$

$\Rightarrow a \neq 0 \Rightarrow$

$\Rightarrow \Leftarrow$ $v \in \operatorname{span}(\mathcal{T}) \Leftarrow v = -a^{-1}(b_1 t_1 + \dots + b_s t_s)$

(6)

Set Theory

Zorn's Lemma

$\Leftrightarrow$

Axiom of Choice

$\Leftrightarrow$

Well ordering

Partially Ordered Set S $\leq$

$a \leq a$

$a \leq b$, $b \leq a \Rightarrow b = a$

$a \leq b$, $b \leq c \Rightarrow a \leq c$

* maximal element $m \in S$, if $s \geq m \Rightarrow s = m$

* chain: subset $C \subset S$ s.t. $C, \leq$ totally ordered

* upper bound u of $\mathcal{T} \subset S$

$\forall t \in \mathcal{T}$ $t \leq u$

Thm (0.10/Roman)

If P partially ordered and every chain has upper bound then P has maximal element

Thm (1.9/Roman) Existence of basis, V. $\mathcal{I}$: l.i., $\mathcal{T}$: spanning set

then $\exists$ maximal l.i. B s.t. $\mathcal{I} \subset B \subset \mathcal{T}$ $\quad$ **Prf** $\mathcal{A} = \{E \text{ l.i.} \mid \mathcal{I} \subset E \subset \mathcal{T}\}$

$\mathcal{A} \neq \phi$ because $\mathcal{I} \in \mathcal{A}$, $C = \{\mathcal{I}_k \mid k \in K\}$ $\quad \mathcal{I}_k \in \mathcal{A}$ $\quad$ order: set inclusion $\quad \mathcal{I}_k \supset \mathcal{I}$

index set

$U = \bigcup_{k \in K} \mathcal{I}_k$, claim: U is in $\mathcal{A}$ and U upper bound of C $\quad \mathcal{I}_1 \subset \mathcal{I}_2 \subset \mathcal{I}_3 \dots$

U l.i. $\quad v_1, \dots, v_n \in U$ $\quad v_1 \in \mathcal{I}_1, \dots, v_n \in \mathcal{I}_n \Rightarrow v_1, \dots, v_n \in \mathcal{I}_n$

$\mathcal{I} \subset U \subset \mathcal{T}$

$\mathcal{I} \subset \mathcal{I}_k \subset \bigcup_{k \in K} \mathcal{I}_k = U$

$\Downarrow$

$U \supset \mathcal{I}$

$\mathcal{I}_k \subset \mathcal{T} \Rightarrow \bigcup_{k \in K} \mathcal{I}_k \subset \mathcal{T}$

$U \in \mathcal{A}$ $\forall\, \mathcal{I}_k \in C$

$\mathcal{I}_k \leq U \Rightarrow \mathcal{A}$ has maximal element of $\mathcal{A}$

$\exists$ $\quad B: \mathcal{I} \subset B \subset \mathcal{T}$

B: l.i.

B: max of $\mathcal{A}$

$B \cup \{v\} \in \mathcal{A}$

$v \notin B \Rightarrow \Leftarrow$

$B \cup \{v\} \supset B$

B: ma

suppo

$\exists v \in$

$\boxed{B \cup \{v\}}$

$t_i \in \mathcal{T}$

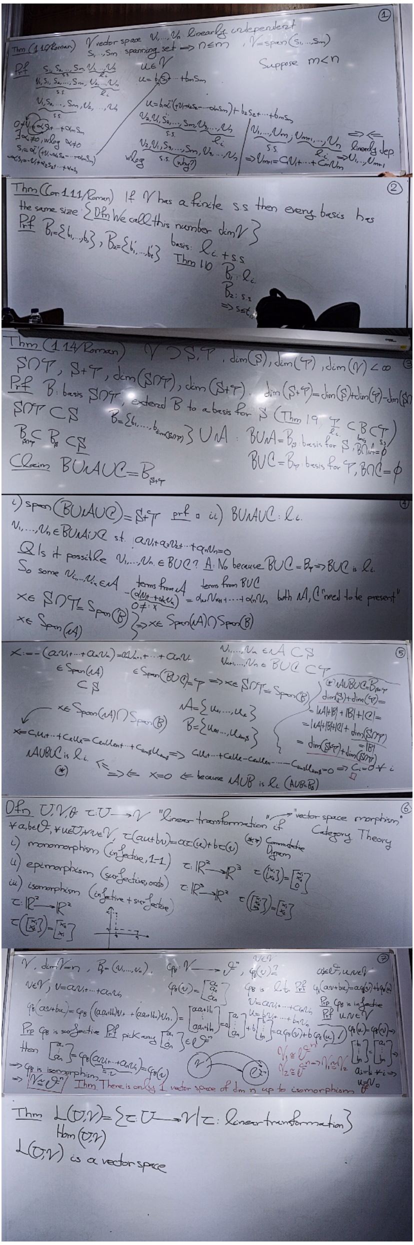
$\cup \langle v \rangle$ is l.i.

4) $\Rightarrow$ 1) $\quad \operatorname{span}(\mathcal{T}) \neq$

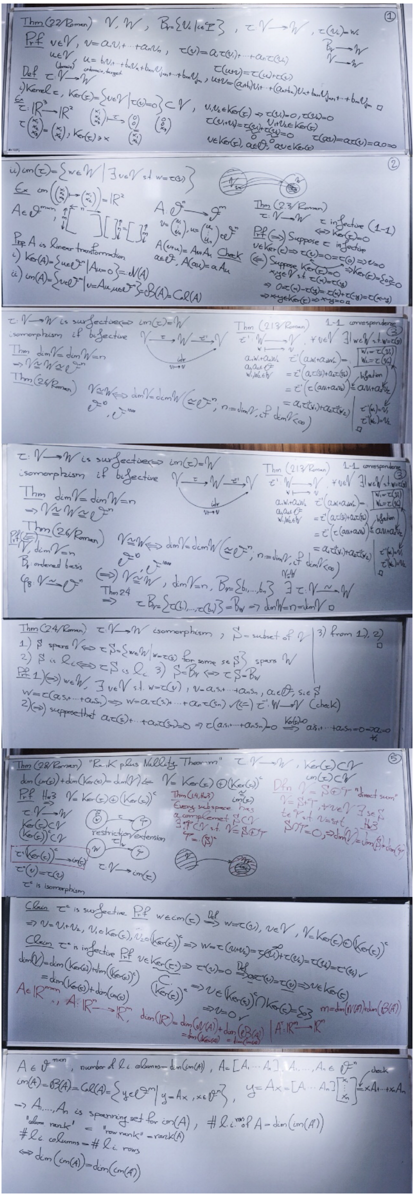
$v + \dots + b_s t_s = 0$ not all

$\Rightarrow \Leftarrow$

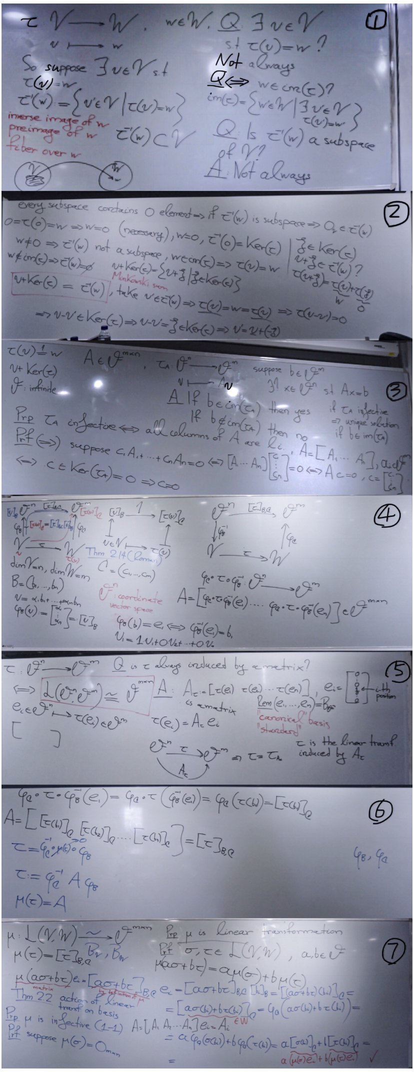
lecture 4



lecture 5



lecture 6



Not always
Q. Is $\tau^{-1}(w)$ a subspace of V?
A. Not always
inverse image of w
preimage of w
fiber over w
Prp τ_A injective $\Leftrightarrow$ all columns of A are l.i.
Q. is τ always induced by a matrix?
canonical basis
"standard"
τ is the linear transf induced by A_τ
$\mu(\tau)=A$
Prp μ is linear transformation
Prp μ is injective (1-1)

lecture 7

①

$B=(b_1,\dots,b_n)$, $C=(c_1,\dots,c_n)$

$$\begin{tikzcd} & \mathcal{F}^n \arrow[d, "\varphi_C\varphi_B^{-1}"] \\ \mathcal{V} \arrow[ur, "\varphi_B"] \arrow[r, "\varphi_C"'] & \mathcal{F}^n \end{tikzcd}$$

$e_i=\begin{bmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{bmatrix}\leftarrow i\text{-th}$

$\tau:\mathcal{V}\longrightarrow\mathcal{W}$ $\quad[\tau]_{B,C}=\left([\tau b_1]_C\ \cdots\ [\tau b_n]_C\right)$ **Thm 2.14**

$B=(b_1,\dots,b_n)$ $\quad[\varphi_B]_B=\left(\underbrace{[\varphi_B(b_1)]_\mathcal{E}}_{=e_1}\ \cdots\ [\varphi_B(b_n)]_\mathcal{E}\right)=I=\begin{pmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{pmatrix}$

$C=(c_1,\dots,c_n)$ $\quad\mathcal{E}=(e_1,e_2,\dots,e_n)$

$M_{B,C}=\left([\varphi_C\varphi_B^{-1}(e_1)]_C\cdots[\varphi_C\varphi_B^{-1}(e_n)]_C\right)=\left([b_1]_C\ [b_2]_C\cdots[b_n]_C\right)$

$[v]_B\longmapsto M_{B,C}[v]_B=[v]_C$ $\quad\varphi_C\varphi_B^{-1}(e_i)=\varphi_C(b_i)=[b_i]_C$

Representation of isomorphism is invertible matrix

$M_{B,C}$ is invertible

$\varphi\circ\psi-id_\mathcal{V}=0$

$\mu(\varphi\circ\psi)=\mu(id_\mathcal{V})$

$\mu(\varphi)\mu(\psi)=I$

$\mu:\mathcal{L}(\mathcal{V},\mathcal{W})\longrightarrow\mathcal{F}^{m\times n}$

$\mathcal{V}=\mathcal{W}$, $\mu:\mathcal{L}(\mathcal{V},\mathcal{V})\longrightarrow\mathcal{F}^{n\times n}$ $\quad\varphi\in\mathcal{L}(\mathcal{V},\mathcal{V})$ $\exists\psi\in\mathcal{L}(\mathcal{V},\mathcal{V})$: $\varphi\circ\psi=\psi\circ\varphi=id_\mathcal{V}$

---

②

Ex 2.4 (Roman) $\quad\mathcal{P}_2=\left\{\alpha_0+\alpha_1x+\alpha_2x^2\ \middle|\ \begin{matrix}\alpha_0,\alpha_1,\alpha_2\in\mathbb{R}\\ x:\text{free over }\mathbb{R}\end{matrix}\right\}\cong\left\{\begin{bmatrix}\alpha_0\\ \alpha_1\\ \alpha_2\end{bmatrix},\alpha_i\in\mathbb{R}\right\}=\mathbb{R}^3$

$D:\mathcal{P}_2\longrightarrow\mathcal{P}_2$, $\quad D(\alpha_0+\alpha_1x+\alpha_2x^2)=\alpha_1+2\alpha_2x$

$B_{\mathcal{P}_2}=\{1,x,x^2\}$, $B_{\mathbb{R}^3}=\{e_1,e_2,e_3\}=\mathcal{E}$

$$\begin{tikzcd} \mathcal{P}_2 \arrow[r, "D"] \arrow[d, "\wr"] & \mathcal{P}_2 \arrow[d, "\wr"] \\ \mathbb{R}^3 \arrow[r, "{[D]_{B_{\mathcal{P}_2},B_{\mathbb{R}^3}}}"] & \mathbb{R}^3 \end{tikzcd}$$

$[D]_{B_{\mathcal{P}_2},B_{\mathbb{R}^3}}=\left([D(1)]_\mathcal{E},[D(x)]_\mathcal{E},[D(x^2)]_\mathcal{E}\right)$

$=\left([0]_\mathcal{E},[1]_\mathcal{E},[2x]_\mathcal{E}\right)$

$=\begin{pmatrix}0&1&0\\0&0&2\\0&0&0\end{pmatrix}$

$p(x)=5+x+2x^2\sim\begin{pmatrix}5\\1\\2\end{pmatrix}$

$\begin{pmatrix}0&1&0\\0&0&2\\0&0&0\end{pmatrix}\begin{pmatrix}5\\1\\2\end{pmatrix}=\begin{pmatrix}1\\4\\0\end{pmatrix}\sim 1+4x$

---

③

Thm 2.16 $\quad\tau:\mathcal{V}\longrightarrow\mathcal{W}$, $\quad B,B'$ for $\mathcal{V}$ $\quad C',[v]_{C'}$

$[\tau]_{B,C}$, $[\tau]_{B',C'}$ $\quad C,C'$ for $\mathcal{W}$

$[\tau]_{B',C'}[v]_{B'}=M_{C,C'}[\tau]_{B,C}M_{B',B}[v]_{B'}$

If $\mathcal{V}=\mathcal{W}\Rightarrow C=B$ $\quad[\tau]_{B',C'}=M_{C,C'}[\tau]_{B,C}M_{B',B}$

$\tau:\mathcal{V}\longrightarrow\mathcal{V}$ $\quad[\tau]_{B'}=M_{B,B'}[\tau]_BM_{B',B}$

Dfn A similar B if $B=SAS^{-1}$

$M_{B,B'}=\left(M_{B',B}\right)^{-1}\Rightarrow[\tau]_{B'}=M_{B,B'}[\tau]_B\left(M_{B,B'}\right)^{-1}=S[\tau]_BS^{-1}$ similarity

---

④

Projections $\quad\mathcal{V}\supset\mathcal{S},\mathcal{T}$ s.t. $\mathcal{V}=\mathcal{S}\oplus\mathcal{T}$: $\mathcal{S}\cap\mathcal{T}=\{0\}$, $\mathcal{V}=\mathcal{S}+\mathcal{T}$

$\rho_{\mathcal{S},\mathcal{T}}:\mathcal{V}\longrightarrow\mathcal{V}$, $\forall v\in\mathcal{V}$ $\exists!\ s\in\mathcal{S},t\in\mathcal{T}$ s.t. $v=s+t$

projection of $\mathcal{V}$ onto $\mathcal{S}$ along $\mathcal{T}$

$v\in\mathcal{V}\longmapsto s=\rho_{\mathcal{S},\mathcal{T}}(v)$

Lem $A\in\mathbb{R}^{m\times n}$, $\underbrace{\mathcal{B}(A^T)}_{\subset\mathbb{R}^n}\cap\underbrace{\operatorname{Ker}(A)}_{\mathbb{R}^n}=0$

Prf

$v\in\mathcal{B}(A^T)\cap\operatorname{Ker}(A)$

$v\in\mathcal{B}(A^T)\Rightarrow\boxed{v=A^Tu}$, for some $u\in\mathbb{R}^m$

$v\in\operatorname{Ker}(A)\Rightarrow\boxed{Av=0}$ $\quad 0=Av=AA^Tu\Rightarrow AA^Tu=0$

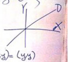
$\mathcal{F}$

Ex 2.5 (Roman) $\mathcal{F}^2=\{(x,y)\mid x,y\in\mathcal{F}\}$

$X=\{(x,0)\mid x\in\mathcal{F}\}$

$Y=\{(0,y)\mid y\in\mathcal{F}\}$

$D=\{(w,w)\mid w\in\mathcal{F}\}$



$\mathcal{F}^2=D\oplus X$ $\quad\rho_{D,X}(x,y)=(y,y)$

$(x,y)=(y,y)+(x-y,0)$, $\rho_{D,Y}$, $\rho_{X,Y}$

Thm 2.22/Roman (algebraic characterization of projections)

$\mathcal{L}(\mathcal{V},\mathcal{V})$ ring $\cong\mathcal{F}^{n\times n}$

$\rho\in\mathcal{L}(\mathcal{V},\mathcal{V})$ is projection $\Leftrightarrow\rho^2=\rho$ (idempotent)

($\Rightarrow$) suppose ρ projection, $\rho=\rho_{\mathcal{S},\mathcal{T}}$ for some $\mathcal{S},\mathcal{T}$

⑤

$\rho^2(v)=\rho(\rho(v))=\rho(\underbrace{\boxed{\rho(\overset{v}{\overline{s+t}})}}_{=s})=\rho(\boxed{s})=s=\rho(\underset{\in\mathcal{S}}{s}+\underset{\in\mathcal{T}}{0})=\rho(v)$ $\quad s=s+0$ $\quad\forall v\Rightarrow\rho^2=\rho$

($\Leftarrow$) Suppose $\rho^2=\rho$ $\quad v=s+t$

we want to find $\mathcal{S},\mathcal{T}$ s.t. $\rho=\rho_{\mathcal{S},\mathcal{T}}$ $\quad$ Idea: $\mathcal{S}=\operatorname{im}(\rho)$, $\mathcal{T}=\operatorname{Ker}(\rho)$

need to show $\mathcal{V}=\operatorname{im}(\rho)\oplus\operatorname{Ker}(\rho)$, check: $v\in\operatorname{im}(\rho)\cap\operatorname{Ker}(\rho)$

$v\in\operatorname{im}(\rho)\Leftrightarrow v=\rho(w)$ for some $w\in\mathcal{V}$

$v\in\operatorname{Ker}(\rho)\Rightarrow 0=\rho(v)=\rho(\rho(w))=\rho^2(w)=\rho(w)=v\Rightarrow v=0$

$\Rightarrow\operatorname{Ker}(\rho)\cap\operatorname{im}(\rho)=0$, $\quad\mathcal{V}\overset{?}{=}\operatorname{im}(\rho)+\operatorname{Ker}(\rho)$

$\rho(v-\rho(v))=\rho(v)-\rho^2(v)$

$=\rho(v)-\rho(v)=0$

so $\mathcal{V}=\operatorname{Ker}(\rho)\oplus\operatorname{im}(\rho)$ $\quad v\in\mathcal{V}$, $v=\underbrace{\rho(v)}_{\in\operatorname{im}(\rho)}+\underbrace{v-\rho(v)}_{\in\operatorname{Ker}(\rho)}$

$v\in\mathcal{V}$, $v=\rho(v)+[v-\rho(v)]$

$\rho(v)=\rho(\rho(v)+[v-\rho(v)])=\rho^2(v)+\rho(\cancelto{0}{v-\rho(v)})=\rho^2(v)=\rho(v)$ ✓

$\dots\ AA^Tu=0$ $\quad\xi=\begin{bmatrix}\xi_1\\ \vdots\\ \xi_n\end{bmatrix}$ $\quad 0=\xi=A^Tu=v$

$[\ u^T\ ]\underbrace{[\ \ ]}_{AA^Tu}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}\Rightarrow u^TAA^Tu=0\Rightarrow\xi^T\xi=0\Leftrightarrow\xi_1^2+\dots+\xi_n^2=0\Rightarrow\xi_i=0\ \forall i\Rightarrow\xi=0$ $\quad\xi_i^2\geq 0$

$\xi:=A^Tu\in\mathbb{R}^n$

---

⑦

$A\in\mathbb{R}^{m\times n}$, column-rank$(A)=\dim(\operatorname{im}(\tau_A))=\dim(\mathcal{B}(A))$

$\mathcal{B}(A)=\{v\in\mathbb{R}^m\mid\exists u\in\mathbb{R}^n\text{ with }v=Au\}=\operatorname{Span}(A_1,\dots,A_n)\subset\mathbb{R}^m$

$A=[A_1\cdots A_n]$

$\mathcal{B}(A^T)$ $\quad$ row-rank$(A)=\dim(\mathcal{B}(A^T))$

$A=[\ \ ]$

$\Rightarrow c_i=0$ because v_i: l.i.

$\Rightarrow Av_1,\dots,Av_s$ are l.i. $\in\mathcal{B}(A)$ $\quad s\leq\dim\mathcal{B}(A)$

Thm row-rank(A) = col-rank(A) =: rank(A) $\quad$ **Prf** (valid only for $\mathbb{R},\mathbb{C}$)

$v_i\in\mathcal{B}(A^T)$

Take a basis $v_1,\dots,v_s$ for $\mathcal{B}(A^T)\Rightarrow v_1,\dots,v_s$ are l.i. $\quad v_i\in\mathbb{R}^n$ $\quad\sum c_iv_i\in\mathcal{B}(A^T)$

$c_1Av_1+c_2Av_2+\dots+c_sAv_s=0$ $\quad\mathbb{R}^m$

$\Rightarrow A(c_1v_1+\dots+c_sv_s)=0\Rightarrow c_1v_1+\dots+c_sv_s\in\operatorname{Ker}(A)\Rightarrow\sum c_iv_i\in\operatorname{Ker}(A)\cap\mathcal{B}(A^T)\overset{\text{Lem}}{=}0$

lecture 8

$A\in\mathbb{R}^{m\times n}$, $\operatorname{rank}(A)=\dim(\operatorname{im}(\tau_A))$, $\tau_A: \mathcal{V}^n \to \mathcal{V}^m$, $\operatorname{rank}(A) = \max \#$ l.i. col (1)
$= \max \#$ l.i. rows

$B\in\mathbb{R}^{n\times p}$, $AB: \mathcal{V}^p \xrightarrow{B} \mathcal{V}^n \xrightarrow{A} \mathcal{V}^m$, $\operatorname{rank}(AB)=$

$\dim(\operatorname{im}(\tau_{AB}))=\dim \mathcal{R}(AB)$, $\operatorname{rank}(AB)\le\operatorname{rank}(A)$

$AB = A[B_{*1}\cdots B_{*p}]$, $AB_{*j} = [A_{*1}\cdots A_{*n}]\begin{bmatrix}B_{1j}\\ \vdots \\ B_{nj}\end{bmatrix} = B_{1j}A_{*1}+\cdots+B_{nj}A_{*n} = (AB)_{*j}$

Prp P, Q: invertible $\Rightarrow \operatorname{rank}(PAQ)=\operatorname{rank}(A)$

Thm (p. 210/Meyer) $\operatorname{rank}(AB)=\operatorname{rank}(B)-\dim\mathcal{N}(A)\cap\mathcal{R}(B)$ (2)

$\mathcal{N}(A)=\{v\in\mathbb{R}^n \mid Av=0\}=\ker(A)$, $\mathcal{R}(B)=\operatorname{Span}(B_{*1},\ldots,B_{*p})$

Prf $B_{\mathcal{N}(A)\cap\mathcal{R}(B)}=\{x_1,\ldots,x_s\}$, $\dim\mathcal{N}(A)\cap\mathcal{R}(B)=s$, $\underbrace{\mathcal{N}(A)\cap\mathcal{R}(B)}_{s}\subset\underbrace{\mathcal{R}(B)}_{s+t}$

$B_{\mathcal{R}(B)}=\{x_1,\ldots,x_s,\overbrace{z_1,\ldots,z_t}^{\{z_1,\ldots,z_t\}}\}$ Goal: $\dim\mathcal{R}(AB)=t=(s+t)-s$

$Az_1,\ldots,Az_t$ is basis for $\mathcal{R}(AB)$

1) S.S. $b\in\mathcal{R}(AB)$, $b=A\underbrace{By}_{\in\mathcal{R}(B)}$, $y\in\mathbb{R}^p$

$By=\xi_1x_1+\cdots+\xi_sx_s+\eta_1z_1+\cdots+\eta_tz_t \overset{b=}{\Rightarrow} ABy=\eta_1Az_1+\cdots+\eta_tAz_t \Rightarrow b\in\operatorname{Span}(Az_1,\ldots,Az_t)$ (3)

2) $\{Az_i\}$: l.i. $\quad 0=\sum\alpha_iAz_i=A(\sum\alpha_iz_i)\Rightarrow\sum\alpha_iz_i\in\mathcal{N}(A)\cap\mathcal{R}(B)$

$\sum\alpha_iz_i=\beta_1x_1+\cdots+\beta_sx_s \Leftrightarrow \beta_1x_1+\cdots+\beta_sx_s+(-\alpha_1)z_1+\cdots+(-\alpha_t)z_t=0$

$\Rightarrow \beta_1=\cdots=\beta_s=\alpha_1=\cdots=\alpha_t=0$

$\mathcal{R}(AB)\subset\operatorname{Span}(Az_1,\ldots,Az_t)\subset\mathcal{R}(AB)$ $\quad\Rightarrow\{Az_i\}$: l.i. ✓

z_i part of basis for $\mathcal{R}(B)\Rightarrow z_i\in\mathcal{R}(B)\overset{\exists w_i}{\Rightarrow} z_i=Bw_i \Rightarrow Az_i=ABw_i\in\mathcal{R}(AB)$

Thm (p. 211/Meyer) $A\in\mathbb{R}^{m\times n}$, $B\in\mathbb{R}^{n\times p}$ i) $\operatorname{rank}(AB)\le\min\{\operatorname{rank}(A),\operatorname{rank}(B)\}$ (4)
ii) $\operatorname{rank}(AB)\ge\operatorname{rank}(A)+\operatorname{rank}(B)-n$

Prf i) $\operatorname{rank}(AB)=\operatorname{rank}(B)-\dim(\mathcal{N}(A)\cap\mathcal{R}(B))\le\operatorname{rank}(B)$, $\quad$ | $\operatorname{rank}(AB)=\operatorname{rank}(B^TA^T)$

$\operatorname{rank}(AB)=\operatorname{rank}((AB)^T)=\operatorname{rank}(B^TA^T)\le\operatorname{rank}(A^T)=\operatorname{rank}(A)$ $\quad$ i) ✓

ii) $\dim\mathcal{N}(A)\cap\mathcal{R}(B)\le\dim\mathcal{N}(A)=\boxed{n-\operatorname{rank}(A)}$ (α)

Rank+Nullity Thm: $\mathcal{V}^n\xrightarrow{A}\mathcal{V}^m$ $\quad \operatorname{rank}(AB)=\operatorname{rank}(B)-\dim\mathcal{N}(A)\cap\mathcal{R}(B)\ge$

$n=\operatorname{rank}(A)+\dim\mathcal{N}(A)$ $\quad\ge\operatorname{rank}(B)+\operatorname{rank}(A)-n$

Thm (2.12/Meyer) $A\in\mathbb{R}^{m\times n}$ i) $\operatorname{rank}(A^TA)=\operatorname{rank}(A)=\operatorname{rank}(AA^T)$
ii) $\mathcal{R}(A^TA)=\mathcal{R}(A^T)$ $\quad \mathcal{V}^n\xrightarrow{A}\mathcal{V}^m$, $\mathcal{V}^m \xrightarrow{A^T} \mathcal{V}^n$
iii) $\mathcal{N}(A^TA)=\mathcal{N}(A)$

Prf {Lem $\mathcal{N}(A)\cap\mathcal{R}(A^T)=0$}

$x\in\mathcal{N}(A)\cap\mathcal{R}(A^T)\Rightarrow x=A^Ty\Rightarrow x=0$

$\Rightarrow 0=Ax=AA^Ty=0$

$y^TAA^Ty=0\Rightarrow\|A^Ty\|_2=0$

$\Rightarrow A^Ty=0$

Prf $\operatorname{rank}(A^TA)=\operatorname{rank}(A)-\underbrace{\dim\mathcal{N}(A^T)\cap\mathcal{R}(A)}_{=0}$ i) ✓

$\mathcal{R}(A^TA)=\mathcal{R}(A^T)$ $\quad$ Lem $\mathcal{V}\subset\mathcal{U}$, $\dim\mathcal{V}=\dim\mathcal{U}$ $\Rightarrow\mathcal{V}=\mathcal{U}$

$x\in\mathcal{R}(A^TA)\overset{\exists y}{\Rightarrow}x=A^T\underbrace{Ay}_{w}=A^Tw\in\mathcal{R}(A^T)$, $w=Ay$

iii) $\mathcal{N}(A^TA)=\mathcal{N}(A)$

$\xi\in\mathcal{N}(A)\Rightarrow A\xi=0\Rightarrow A^TA\xi=0$

Lem $\mathcal{R}(A^TA)\subset\mathcal{R}(A^T)$ Prf take $x\in\mathcal{R}(A^TA)\Rightarrow\exists y\in\mathbb{R}^n$ s.t. $x=A^TAy$

define $w=Ay\in\mathbb{R}^m$ then $x=A^Tw\Rightarrow x\in\mathcal{R}(A^T)$

$\tau:\mathcal{U}\to\mathcal{V}$

$\dim\mathcal{U}=\dim(\operatorname{im}(\tau))+\dim\ker(\tau)$, $\dim\mathcal{V}$?

"R+N", A: $n=\operatorname{rank}(A)+\dim\mathcal{N}(A)$

"R+N", A^TA: $n=\operatorname{rank}(A^TA)+\dim\mathcal{N}(A^TA)$

$Ax=b$ "inconsistent"

$\min\limits_x\|b-Ax\|_2$

$B^TB=I$

[ ]

Normal Equations: $Ax=b$ (*) has a solution $\Leftrightarrow b\in\mathcal{R}(A)$

$\Leftrightarrow b\in\mathcal{R}(A)$ whenever (*) has solution "consistent"

$\underbrace{\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}}_{A}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\end{bmatrix}$

Suppose (*) consistent $\exists x_0\in\mathbb{R}^n$ s.t. $Ax_0=b$

solution set of (*) $=\{x_0+\mathcal{N}(A)\}$

$Ax=b\Rightarrow$

$A^TAx=A^Tb$ (**) $\quad$ Solution (*) $\subset$ Solution (**)

$A^TAx_0=A^Tb$
$\Uparrow$
$Ax_0=b$

Ex A is 1×2
$A=[a_{11}\ a_{12}]$
$[a_{11}\ a_{12}]\begin{bmatrix}x_1\\x_2\end{bmatrix}=b_1$

$\mathbb{R}$

$\mathcal{N}(A)=\left\{\begin{bmatrix}y_1\\y_2\end{bmatrix}:\begin{bmatrix}y_1\\y_2\end{bmatrix}\perp\begin{bmatrix}a_1\\a_2\end{bmatrix}\right\}$

$a_1y_1+a_2y_2=0$

Prp1 (**) always solution Prf need to show that $A^Tb\in\mathcal{R}(A^TA)=\mathcal{R}(A^T)$ ✓

Prp2 If (*) consistent $\Rightarrow$ Sol(*) = Sol(**) Prf Sol(*) $=\{x_0+\mathcal{N}(A)\}=\{x_0+\mathcal{N}(A^TA)\}$, $\mathcal{N}(A)=\mathcal{N}(A^TA)$

$\mathbb{R}^n$, $f:\mathbb{R}^n\to\mathbb{R}_{\ge0}$, $f(x)=\sqrt{x^Tx}=\sqrt{x_1^2+\cdots+x_n^2}$: "Euclidean" norm $f(x)=\|x\|_2$

$=(x_1^2+\cdots+x_n^2)^{1/2}$ "ℓ_2" norm

$g:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$, $g(x,y)=x^Ty$

Vector Norms $\|\cdot\|$
$\|x\|\ge0$, $\|x\|=0\Leftrightarrow x=0$
$\|\alpha x\|=|\alpha|\,\|x\|$ "homogeneity"
$\|x+y\|\le\|x\|+\|y\|$ "triangle inequality"

Cauchy-Schwarz inequality

i) $|\langle x,y\rangle|\le\|x\|\cdot\|y\|$ Prf (ℓ_2-norm, General case p. 287)

ii) $|\langle x,y\rangle|=\|x\|\,\|y\|\Leftrightarrow x,y$ l.d. $\Leftrightarrow y=\alpha x$

Prf $\alpha=\frac{x^Ty}{\|x\|^2}$, $x\ne0$

Observe $\boxed{x^T(\alpha x-y)=0}$ $\quad 0\le\|\alpha x-y\|_2^2=(\alpha x-y)^T(\alpha x-y)=\alpha^2x^Tx-\alpha x^Ty-\alpha y^Tx+y^Ty$

$x^T\left(\left(\frac{x^Ty}{\|x\|^2}\right)x-y\right)=\left(\frac{x^Ty}{\|x\|^2}\right)x^Tx-x^Ty=\left(\frac{x^Ty}{\|x\|^2}\right)\|x\|^2-x^Ty=0$

$=\left(\frac{x^Ty}{\|x\|^2}\right)^2\|x\|^2-2\frac{x^Ty}{\|x\|^2}(x^Ty)+y^Ty$

$=y^Ty-\frac{(x^Ty)^2}{\|x\|^2}=\frac{y^Ty\,\|x\|^2-(x^Ty)^2}{\|x\|^2}=\frac{\|y\|^2\|x\|^2-(x^Ty)^2}{\|x\|^2}\ge0$

$\Rightarrow\|x\|^2\|y\|^2-(x^Ty)^2\ge0$

$\Leftrightarrow(x^Ty)^2\le\|x\|^2\|y\|^2$

$|x^Ty|\le\|x\|\,\|y\|$

lecture 9

①

$\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}$

i) $\|x\| = 0 \iff x = 0$, $\|x\| \geq 0$

ii) $\|\alpha x\| = |\alpha| \, \|x\|$, $\forall \alpha \in \mathbb{R}$

iii) $\|x+y\| \leq \|x\| + \|y\|$

$\|x\|_2 = \sqrt{x^T x} = (x_1^2 + \cdots + x_n^2)^{1/2}$

Prp $\|x+y\|_2 \leq \|x\|_2 + \|y\|_2 \quad \forall x, y \in \mathbb{R}^n$

Prf $\|x+y\|_2^2 = (x+y)^T(x+y) = \|x\|_2^2 + \|y\|_2^2 + 2x^T y$

$\overset{CS}{\leq} \|x\|_2^2 + \|y\|_2^2 + 2\|x\|_2 \|y\|_2 = (\|x\|_2 + \|y\|_2)^2$

$|x^T y| \leq \|x\|_2 \|y\|_2$: CS $\quad \ell_0 : \|x\|_0 \sim \|x\|_1$

Reverse Δ inequality

$\big| \|x\|_2 - \|y\|_2 \big| \leq \|x+y\|_2 \leq \|x\|_2 + \|y\|_2$

Prf $\|x\|_2 = \|x+y-y\|_2 \leq \|x+y\|_2 + \|y\|_2$

$\iff \|x\|_2 - \|y\|_2 \leq \|x+y\|_2$ ✓

Matrix Norms

$\mathbb{R}^{m \times n} \cong \mathbb{R}^{mn}$

$A = [\,\equiv\,] = [\,|\,|\,|\,]$

$A \longmapsto vec(A)$

$Ax = \begin{bmatrix} A_{1*}x \\ \vdots \\ A_{m*}x \end{bmatrix}$ $\quad \begin{bmatrix} A_{1*} \\ A_{2*} \\ A_{m*} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Frobenious

$\|vec(A)\|_2 = [trace(A^T A)]^{1/2}$, $(A^T A)_{ii} = A_{*i}^T A_{*i} = \sum_{j=1}^{n} \alpha_{ji}^2$

$B \in \mathbb{R}^{n \times n}$: $trace(B) = \sum_{i=1}^{n} b_{ii}$

conformable

$\|AB\|_F \leq \|A\|_F \|B\|_F$

Prf

Lem $\|Ax\|_2 \leq \|A\|_F \|x\|_2$

Prf $\|Ax\|_2^2 = \sum (A_{i*} x)^2 \overset{CS}{\leq} \cdots$

CS: $A_{i*}^T x \leq \|A_{i*}\|_2 \|x\|_2$

$\cdots \leq \sum_{i=1}^{m} \|A_{i*}\|_2^2 \|x\|_2^2 = \|x\|_2^2 \left(\sum_{i=1}^{n} \|A_{i*}\|_2^2 \right)$

$= \|x\|_2^2 \, \|A\|_F^2$

②

Prf ($\|AB\|_F \leq \|A\|_F \|B\|_F$)

$AB = \begin{bmatrix} A_{1*} \\ \vdots \\ A_{m*} \end{bmatrix} [B_{*1} \cdots B_{*k}]$

…

Induced matrix norms

vector norms $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$

$\|A\|_{\alpha,\beta} = \max_{\|x\|_\alpha = 1} \|Ax\|_\beta$

spectral norm $\|A\|_{2,2} = \|A\|_2$

Thm $\|A\|_2 = [\lambda_{max}(A^T A)]^{1/2}$

$Ax = \lambda x$

Prf

Prp symmetric matrices are orthogonally diagonalible

$A \in \mathbb{R}^{n \times n}$ $A^T = A$, $A = U \Lambda U^T$, $UU^T = U^T U = I$, Λ: diagonal

$\iff$ there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A

$A U_{*i} = U \Lambda U^T U_{*i} = U \Lambda e_i = U(\lambda_i e_i) = \lambda_i (U e_i) = \lambda_i U_{*i}$

$A^T A = U \Lambda U^T$ take any $x \in \mathbb{R}^n$ $\exists c_i$ s.t. $x = \sum c_i U_{*i}$

$\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2$, $\boxed{\|Ax\|_2^2} = x^T A^T A x = x^T A^T A \left(\sum c_i U_{*i} \right)$

$= x^T \left(\sum c_i A^T A U_{*i} \right) = x^T \left(\sum_{i=1}^{n} c_i \lambda_i U_{*i} \right) = \left(\sum_{j=1}^{n} c_j U_{*j} \right)^T \left(\sum_{i=1}^{n} c_i \lambda_i U_i \right) =$

$= \sum_{i,j} c_i c_j \lambda_i U_{*j}^T U_{*i} = \sum_{i,j} c_i c_j \lambda_i \delta_{ij} = \sum_i c_i^2 \lambda_i$

③

$x^T A^T A x = \sum_{i=1}^{n} c_i^2 \lambda_i$, $x = \sum_i c_i U_{*i} \Rightarrow x^T x = \sum_i c_i^2 = 1$ $\quad \lambda_{max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

$\leq \sum_{i=1}^{n} c_i^2 \lambda_{max} = \lambda_{max} \sum_i c_i^2 = \lambda_{max} \Rightarrow \|A\|_2 \leq \lambda_{max}^{1/2}$, for $x = U_{*1}$, $x^T A^T A x = \lambda_{max}^2$ ✓

Prf(#2) $\mathcal{L}(x, \lambda) = x^T A^T A x + \lambda(-x^T x + 1)$

Lagrangian $\quad$ Lagrange multiplier

Thm optimal points of initial problem are critical points of $\mathcal{L}$

$\nabla_x \mathcal{L}(x, \lambda) = 0$

$\nabla_x \mathcal{L}(x, \lambda) = A^T A x - \lambda x = 0 \iff A^T A x = \lambda x \Rightarrow \lambda$ eigenvalue, x eigenvector of $A^T A$

$A, B \in \mathbb{R}^{m \times n}$ $\quad trace(A^T B) \leq \|A\|_F \|B\|_F$

$vec(A)^T vec(B) \leq \|vec(A)\|_2 \|vec(B)\|_2$

$x, y \in \mathbb{R}^n$ $\quad \cos(\theta) = \frac{x^T y}{\|x\| \, \|y\|} \leq 1$

$\theta = \measuredangle x, y$

④

Gramm-Schmidt $\iff$ QR factorization

orthogonalization

$\alpha_1, \ldots, \alpha_k$ l.i., $\mathbb{R}^n$

$S = span(\alpha_1, \ldots, \alpha_k)$

$\dim(S) = k$

$q_1 = \frac{\alpha_1}{\nu_1}$, $\nu_1 = \|\alpha_1\|_2$

$S_1 := span(q_1)$

$\rho_{S_1} = q_1 q_1^T$

$q_2 = \alpha_2 - \rho_{S_1}(\alpha_2) = \alpha_2 - (q_1 q_1^T)\alpha_2 = \underbrace{\alpha_2 - (q_1^T \alpha_2) q_1}_{\nu_2} \perp q_1$, $\nu_2 := \|\alpha_2 - (q_1^T \alpha_2) q_1\|_2$

suppose $\nu_2 = 0 \Rightarrow \alpha_2 - (q_1^T \alpha_2) q_1 = 0 \iff \alpha_2 - (q_1^T \alpha_2) \frac{\alpha_1}{\|\alpha_1\|_2} = 0 \Rightarrow \Leftarrow$

$q_3 = \underbrace{\alpha_3 - (q_1 q_1^T)\alpha_3 - (q_2 q_2^T)\alpha_3}_{\nu_3}$

$Q = [q_1 \cdots q_k]$, $Q^T Q = I_k$

$V = im(\rho) \oplus Ker(\rho)$

$im(\rho_{S_1}) \perp Ker(\rho_{S_1})$

$(q_1 q_1^T)(q_1 q_1^T) = q_1 \underbrace{(q_1^T q_1)}_{=1} q_1^T = q_1 q_1^T$, $im(\rho_{S_1}) = Span(q_1)$, $Ker(\rho_{S_1}) = (S_1)^\perp$

assumption first r columns of A are l.i.

$A \in \mathbb{R}^{m \times n}$ $\exists Q, R$ s.t. $A = QR$, $Q^T Q = I_{\dim B(\cdot)}$, $R = [\nabla \; \square]$

if A is full-column rank $A = QR$, $Q \in \mathbb{R}^{m \times r}$, $R \in \mathbb{R}^{r \times r}$, R: upper triangular

qr(A) $\quad Ax = b \iff QRx = b \Rightarrow Q^T Q R x = Q^T b \Rightarrow Rx = Q^T b \iff \begin{bmatrix} \ddots & \\ 0 & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} (Q^T b)_1 \\ \vdots \\ (Q^T b)_n \end{bmatrix}$

back-substitution

$q_1 = \frac{\alpha_1}{\nu_1}$

$q_2 = \frac{\alpha_2 - (q_1^T \alpha_2) q_1}{\nu_1}$

$q_3 = \frac{\alpha_3 - (q_1^T \alpha_3) q_1 - (q_2^T \alpha_3) q_2}{\nu_3}$

....

$\iff$

$\alpha_1 = \nu_1 q_1$

$\alpha_2 = (q_1^T \alpha_2) q_1 + \nu_2 q_2$

$\alpha_3 = (q_1^T \alpha_3) q_1 + (q_2^T \alpha_3) q_2 + \nu_3 q_3$

$\iff [\alpha_1 \; \alpha_2 \; \alpha_3] = \underbrace{[q_1 \; q_2 \; q_3]}_{Q} \underbrace{\begin{bmatrix} \nu_1 & q_1^T \alpha_2 & q_1^T \alpha_3 \\ 0 & \nu_2 & q_2^T \alpha_3 \\ 0 & 0 & \nu_3 \end{bmatrix}}_{R}$

$A = [A_{*1} \; A_{*2} \; A_{*3}]$

$vec(A) = \begin{bmatrix} A_{*1} \\ A_{*2} \\ A_{*3} \end{bmatrix}$

lecture 10

$\mathbb{R}^n = \mathcal{S} \oplus \mathcal{T}$, $\mathcal{S} \cap \mathcal{T} = 0$, $\mathbb{R}^n = \mathcal{S} + \mathcal{T}$, $\rho_{\mathcal{S},\mathcal{T}} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $\rho_{\mathcal{S},\mathcal{T}}(v) = s$, $v = s + t$, $s \in \mathcal{S}, t \in \mathcal{T}$ ①

$B_{\mathcal{S}} = [u_1 \cdots u_\ell]$, $B_{\mathcal{T}} = [v_1, \dots, v_m]$, Thm (p386 Meyer) $[\rho_{\mathcal{S},\mathcal{T}}]_{\mathcal{E},\mathcal{E}} = [B_{\mathcal{S}}\ B_{\mathcal{T}}] \begin{bmatrix} I_\ell & 0 \\ 0 & 0 \end{bmatrix} [B_{\mathcal{S}}\ B_{\mathcal{T}}]^{-1}$

$B = [B_{\mathcal{S}}\ B_{\mathcal{T}}]$, $\mathcal{E} = [e_1 \cdots e_n] = I_n$ Prf $[\rho_{\mathcal{S},\mathcal{T}}]_{B,B} = \left([\rho_{\mathcal{S},\mathcal{T}}(u_1)]_B \cdots [\rho_{\mathcal{S},\mathcal{T}}(u_\ell)]_B\ [\rho_{\mathcal{S},\mathcal{T}}(v_1)]_B \cdots [\rho_{\mathcal{S},\mathcal{T}}(v_m)]_B \right)$

$[\rho_{\mathcal{S},\mathcal{T}}]_{\mathcal{E},\mathcal{E}} = M_{B,\mathcal{E}} [\rho_{\mathcal{S},\mathcal{T}}]_{B,B} M_{\mathcal{E},B}$ $\quad = \left([u_1]_B \cdots [u_\ell]_B\ [0]_B \cdots [0]_B \right) = (e_1 \cdots e_\ell\ 0 \cdots 0)$

$M_{B,\mathcal{E}} = \left([u_1]_{\mathcal{E}} \cdots [u_\ell]_{\mathcal{E}} [v_1]_{\mathcal{E}} \cdots [v_m]_{\mathcal{E}} \right) = B = [B_{\mathcal{S}}\ B_{\mathcal{T}}]$ $\quad = \begin{pmatrix} I_\ell & 0 \\ 0 & 0 \end{pmatrix}$

Dfn $\rho_{\mathcal{S},\mathcal{T}}$ is called orthogonal if $\mathcal{S} \perp \mathcal{T}$. $\mathcal{S} = im(\rho_{\mathcal{S},\mathcal{T}})$, $\mathcal{T} = Ker(\rho_{\mathcal{S},\mathcal{T}})$ ②

Dfn $\mathcal{S} \subset \mathbb{R}^n$, $\mathcal{S}^\perp = \{x \in \mathbb{R}^n \mid x \perp \mathcal{S}\} = \{x \in \mathbb{R}^n \mid x \perp y, \forall y \in \mathcal{S}\}$

Prp $\mathbb{R}^n = \mathcal{S} \oplus \mathcal{S}^\perp$ Prf $x \in \mathcal{S} \cap \mathcal{S}^\perp \Rightarrow x \perp x \Leftrightarrow x^T x = 0 \Leftrightarrow \|x\|_2 = 0 \Leftrightarrow x = 0 \Rightarrow \mathcal{S} \cap \mathcal{S}^\perp = 0$

$\mathbb{R}^n = \mathcal{S} + \mathcal{S}^\perp$, $B_{\mathcal{S}} = [v_1 \dots v_\ell]$, $B_{\mathcal{S}}^T B_{\mathcal{S}} = I_\ell$, $x \in \mathbb{R}^n$, $x_1 := \sum_{i=1}^{\ell} (x^T v_i) v_i$, $x_2 := x - x_1$

Cor $\dim(\mathcal{S}^\perp) = n - \dim(\mathcal{S})$ Prp $(\mathcal{S}^\perp)^\perp = \mathcal{S}$, $\mathcal{S} \subset \mathbb{R}^k$

$\mathbb{R}^n = \mathcal{S} \oplus \mathcal{T}$ Prf $x \in (\mathcal{S}^\perp)^\perp$, $x \in \mathbb{R}^k = \mathcal{S} \oplus \mathcal{S}^\perp$

$x_1 \in \mathcal{S}$. to show $x_2 \perp \mathcal{S}$ it is enough to show $x_2 \perp v_k\ \forall k$

$v_k^T x_2 = v_k^T x - v_k^T x_1 = v_k^T x - \sum_{i=1}^{\ell} (x^T v_i)\underbrace{(v_k^T v_i)}_{\delta_{ki}}$

$x = m + n$

$m \in \mathcal{S} \Rightarrow x = m$ $\quad 0 = \langle x, n \rangle = \langle m + n, n \rangle = \langle m, n \rangle + \langle n, n \rangle$ $\quad = v_k^T x - x^T v_k = 0$

$n \in \mathcal{S}^\perp \Rightarrow x \in \mathcal{S}$ $\quad \mathcal{S}^\perp \perp (\mathcal{S}^\perp)^\perp$ $\quad \mathcal{S} \perp \mathcal{S}^\perp$ $\quad = \langle n, n \rangle = \|n\|_2^2 \Rightarrow n = 0$

$\Rightarrow (\mathcal{S}^\perp)^\perp \subset \mathcal{S} \Rightarrow (\mathcal{S}^\perp)^\perp = \mathcal{S}$

Take $\mathcal{T} = \mathcal{S}^\perp$, $\rho_{\mathcal{S},\mathcal{S}^\perp}$, $B_{\mathcal{S}}$, $B_{\mathcal{S}^\perp}$, $B_{\mathcal{S}}^T B_{\mathcal{S}} = I_\ell$, $B_{\mathcal{S}^\perp}^T B_{\mathcal{S}^\perp} = I_{n-\ell}$ ③

$[\rho_{\mathcal{S},\mathcal{S}^\perp}] = [B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}] \begin{bmatrix} I_\ell & 0 \\ 0 & 0 \end{bmatrix} [B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}]^{-1}$, $[B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}]^T \underbrace{[B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}]}_{\in \mathbb{R}^{n \times n}} =$

$= \begin{bmatrix} B_{\mathcal{S}}^T \\ B_{\mathcal{S}^\perp}^T \end{bmatrix} [B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}] = \begin{bmatrix} B_{\mathcal{S}}^T B_{\mathcal{S}} & B_{\mathcal{S}}^T B_{\mathcal{S}^\perp} \\ B_{\mathcal{S}^\perp}^T B_{\mathcal{S}} & B_{\mathcal{S}^\perp}^T B_{\mathcal{S}^\perp} \end{bmatrix} = \begin{bmatrix} I_\ell & 0 \\ 0 & I_{n-\ell} \end{bmatrix}$

$\mathbb{R}^k = \mathcal{S}^\perp \oplus (\mathcal{S}^\perp)^\perp \Rightarrow \dim(\mathcal{S}^\perp)^\perp = k - \dim(\mathcal{S}^\perp) = k - (k - \dim(\mathcal{S})) = \dim(\mathcal{S})$

$[\rho_{\mathcal{S},\mathcal{S}^\perp}]_{\mathcal{E}} = [B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}] \begin{bmatrix} I_\ell & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} B_{\mathcal{S}}^T \\ B_{\mathcal{S}^\perp}^T \end{bmatrix} = [B_{\mathcal{S}}\ B_{\mathcal{S}^\perp}] \begin{bmatrix} B_{\mathcal{S}}^T \\ 0 \end{bmatrix} = B_{\mathcal{S}} B_{\mathcal{S}}^T$

$B_{\mathcal{S}} = [v_1 \dots v_\ell]$ $\quad B_{\mathcal{S}} B_{\mathcal{S}}^T x = [v_1 \dots v_\ell] \begin{bmatrix} v_1^T \\ \vdots \\ v_\ell^T \end{bmatrix} x = [v_1 \dots v_\ell] \begin{bmatrix} v_1^T x \\ \vdots \\ v_\ell^T x \end{bmatrix} = \sum (v_i^T x) v_i$

n, ℓ

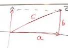
$\rho_{\mathcal{S},\mathcal{S}^\perp} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ $\quad \mathbb{R}^n \xrightarrow[\substack{\text{projection} \\ \text{surjective}}]{B_{\mathcal{S}}^T} \mathbb{R}^\ell \xrightarrow[\substack{\text{embedding} \\ \text{injective}}]{B_{\mathcal{S}}} \mathbb{R}^n$

$[\rho_{\mathcal{S}^\perp,\mathcal{S}}] = P_{\mathcal{S}^\perp} = I - P_{\mathcal{S}}$

$P_{\mathcal{S}} = B_{\mathcal{S}} B_{\mathcal{S}}^T$

$\rho_{\mathcal{S},\mathcal{T}} + \rho_{\mathcal{T},\mathcal{S}} = id$

Pythagorean Theorem ⑤



$c = \sqrt{a^2 + b^2}$

$\|\vec{c}\|_2 = \sqrt{\|\vec{a}\|_2^2 + \|\vec{b}\|_2^2} =$

$= [\|P_x(\vec{c})\|_2^2 + \|P_y(\vec{c})\|_2^2]^{1/2}$

Thm $\mathcal{S} \subset \mathbb{R}^n$, $x \in \mathbb{R}^n$

$\|x\|_2 = [\|P_{\mathcal{S}}(x)\|_2^2 + \|P_{\mathcal{S}^\perp}(x)\|_2^2]^{1/2}$

Prf $x = P_{\mathcal{S}}(x) + P_{\mathcal{S}^\perp}(x)$

$\|x\|_2^2 = x^T x = [(P_{\mathcal{S}}(x))^T + (P_{\mathcal{S}^\perp}(x))^T][P_{\mathcal{S}}(x) + P_{\mathcal{S}^\perp}(x)] =$

$= \|P_{\mathcal{S}}(x)\|_2^2 + \|P_{\mathcal{S}^\perp}(x)\|_2^2 + 2\underbrace{(P_{\mathcal{S}}(x))^T P_{\mathcal{S}^\perp}(x)}_{=0}$

Thm $A \in \mathbb{R}^{m \times n}$, $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$

Prf $x \in \mathcal{R}(A)^\perp \Leftrightarrow x \perp \mathcal{R}(A) \Leftrightarrow x^T A y = 0\ \forall y \in \mathbb{R}^n \Leftrightarrow y^T (A^T x) = 0\ \forall y$

$\Leftrightarrow e_i^T (A^T x) = 0\ \forall i \Leftrightarrow (A^T x)_i = 0\ \forall i \Leftrightarrow A^T x = 0 \Leftrightarrow x \in \mathcal{N}(A^T)$

Cor $\mathbb{R}^m = \mathcal{R}(A) \oplus \mathcal{N}(A^T)$

$A \in \mathbb{R}^{m \times n}$, $rank(A) = n$, $A = [\ ]$, $P_{\mathcal{R}(A)} = A(A^T A)^{-1} A^T = Q Q^T$ ⑥

$A^T A$ is invertible, $A^T A \in \mathbb{R}^{n \times n}$, $rank(A^T A) = rank(A) = n$ $\quad Q$: orthonormal basis for $\mathcal{R}(A)$

$[A(A^T A)^{-1} A^T][A(A^T A)^{-1} A^T] = A \underbrace{(A^T A)^{-1} A^T A}_{I} (A^T A)^{-1} A^T = A (A^T A)^{-1} A^T$ $\quad rank(R) = n \Rightarrow R$: invertible

$A(A^T A)^{-1}(A^T x)$ $\quad \dim \mathcal{R}(A) = n$, $Q \in \mathbb{R}^{m \times n}$, $A = QR$, $Q^T Q = I_n$, $R \in \mathbb{R}^{n \times n}$

$A(A^T A)^{-1} A^T = QR(R^T \underbrace{Q^T Q}_{I} R)^{-1} R^T Q^T = QR(R^T R)^{-1} R^T Q^T = QRR^{-1}(R^T)^{-1} R^T Q^T = QQ^T$

Vandermonde metrices

$Ax = b$, $A \in \mathbb{R}^{m \times n}$, $m > n$, $b \in \mathbb{R}^n$ $\quad \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$ ⑦

$[\ ][\ ] = [\ ]$ suppose inconsistent no solution

tall

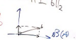
P: orthogonal projection onto $\mathcal{R}(A)$, $P^\perp$: projection onto $\mathcal{R}(A)^\perp$

$J = \|Ax - b\|_2^2 = \|Ax - (Pb + P^\perp b)\|_2^2 = \|P(Ax) - Pb - P^\perp b\|_2^2 =$ $\quad Ax^* = A(A^T A)^{-1} A^T b =$

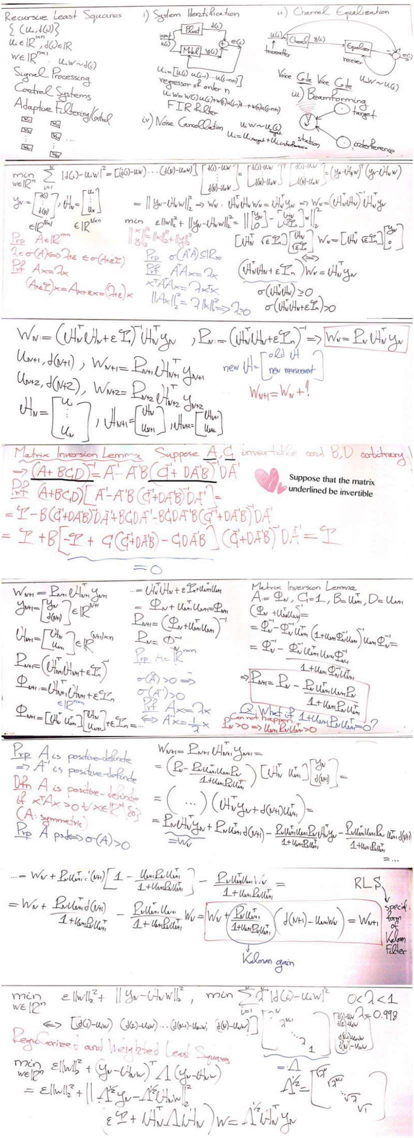
$= \|P(Ax - b) - P^\perp b\|_2^2 = \|P(Ax - b)\|_2^2 + \|P^\perp b\|_2^2$ $\quad = Pb$

$P(Ax - b) = 0 \Leftrightarrow PAx = Pb \Leftrightarrow Ax = Pb \Leftrightarrow Ax = A(A^T A)^{-1} A^T b$ consistent $\quad J_{min} = \|P^\perp b\|_2^2$

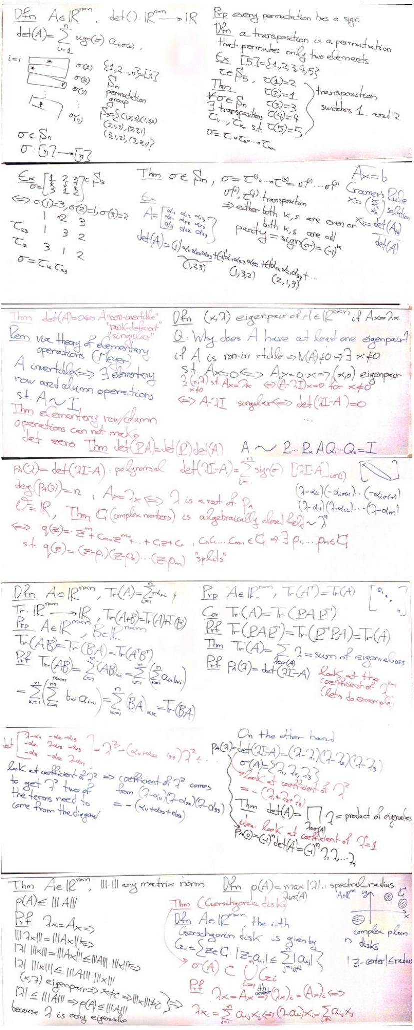
$\Rightarrow A^T A x = A^T b \Rightarrow \boxed{x^* = (A^T A)^{-1} A^T b}$



lecture 11

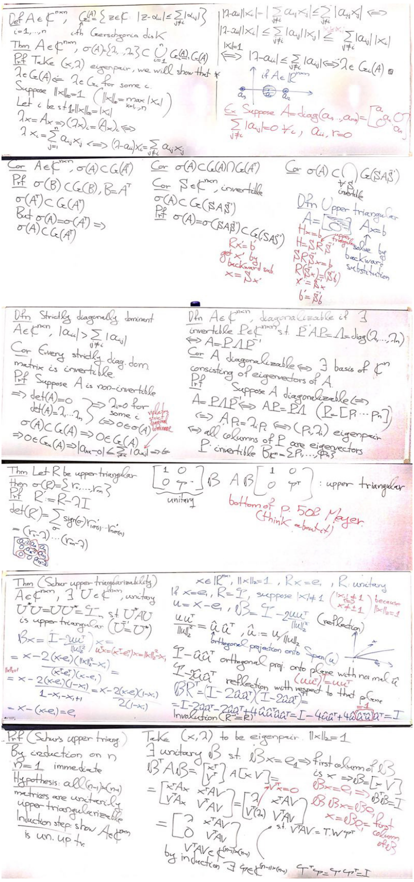

Recursive Least Squares
i) System Identification
ii) Channel Equalization
Signal Processing
Control Systems
Adaptive Filtering/Control
FIR Filter
iii) Beamforming
iv) Noise Cancellation
Matrix Inversion Lemma
Suppose that the matrix underlined be invertible
Kalman gain
RLS
Regularized and Weighted Least Squares

lecture 12

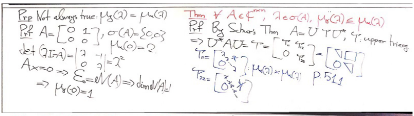


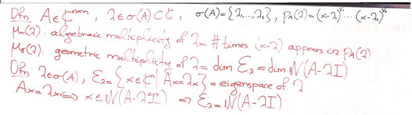
$$\Rightarrow |\lambda - a_{ii}|\,|x_i| \le \sum_{j\neq i} |a_{ij}|\,|x_j|$$

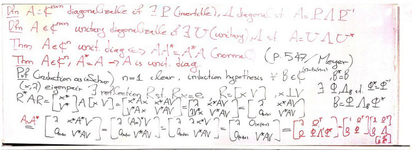
lecture 13

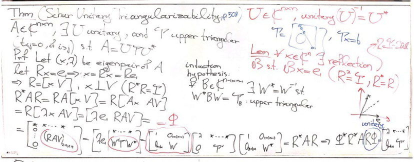

Def A ∈ ℂ^{n×n}; G_i(A) = { z ∈ ℂ : |z − a_ii| ≤ Σ_{j≠i} |a_ij| }
i = 1, …, n   i-th Gerschgorin disk
Thm A ∈ ℂ^{n×n}, σ(A) = {λ_1, …, λ_n} ⊂ ∪_{i=1}^n G_i(A) =: G(A)
Prf Take (x, λ) eigenpair, we will show that *
λ ∈ G(A) ⇐ λ ∈ G_i for some i
Suppose ||x||_∞ = 1 (||x||_∞ = max_{k=1,…,n} |x_k|)
Let i be s.t. 1 = ||x||_∞ = |x_i|
λx = Ax ⇒ (λx)_i = (Ax)_i ⇔
λx_i = Σ_{j=1}^n a_ij x_j ⇔ (λ − a_ii)x_i = Σ_{j≠i} a_ij x_j
|λ − a_ii||x_i| = |Σ_{j≠i} a_ij x_j| ≤ Σ_{j≠i} |a_ij x_j| ⇔
|λ − a_ii||x_i| ≤ Σ_{j≠i} |a_ij||x_j| ≤ Σ_{j≠i} |a_ij||x_i|
|x_i| = 1
⇔ |λ − a_ii| ≤ Σ_{j≠i} |a_ij| ⇔ λ ∈ G_i(A)
if A ∈ ℝ^{n×n}
a_3 a_1 a_2
Ex Suppose A = diag(a_1, …, a_n)
Σ_{j≠i} |a_ij| = 0 ∀i, a_ii, r = 0
Cor A ∈ ℂ^{n×n}, σ(A) ⊂ G(A^T)
Prf σ(B) ⊂ G(B), B = A^T
σ(A^T) ⊂ G(A^T)
But σ(A) = σ(A^T) ⇒
σ(A) ⊂ G(A^T)
Cor σ(A) ⊂ G(A) ∩ G(A^T)
Cor S ∈ ℂ^{n×n}, invertible
σ(A) ⊂ G(SAS^{-1})
Prf σ(A) = σ(SAS^{-1}) ⊂ G(SAS^{-1})
Rx' = b'
get x' by backward sub
x = S^{-1}x'
Cor σ(A) ⊂ ∩_{∀S invertible} G(SAS^{-1})
Dfn Upper triangular
A = [upper triangular], Ax = b
solve by backward substitution
Hx = b
H = SRS^{-1}
upper triangular
SRS^{-1}x = b
R(S^{-1}x) = (S^{-1}b)
x' = S^{-1}x
b' = S^{-1}b
Dfn Strictly diagonally dominant
A ∈ ℂ^{n×n}, |a_ii| > Σ_{j≠i} |a_ij|
Cor Every strictly diag. dom. matrix is invertible.
Prf Suppose A is non-invertible
⇒ det(A) = 0
det(A) = λ_1 ⋯ λ_n
⇒ λ_i = 0 for some i
⇔ 0 ∈ σ(A)
σ(A) ⊂ G(A) ⇒ 0 ∈ G(A)
⇒ 0 ∈ G_k(A) ⇒ |a_kk − 0| ≤ Σ_{j≠k} |a_kj| ⇒ ⇐
violates strict diagonal dominance
Dfn A ∈ ℂ^{n×n}, diagonalizable if ∃ invertible P ∈ ℂ^{n×n} s.t. P^{-1}AP = Λ = diag(λ_1, …, λ_n)
⇔ A = PΛP^{-1}
Cor A diagonalizable ⇔ ∃ basis of ℂ^n consisting of eigenvectors of A.
Prf Suppose A diagonalizable ⇔
A = PΛP^{-1} ⇔ AP = PΛ (P = [p_1 ⋯ p_n])
⇔ Ap_i = λ_i p_i ⇔ (p_i, λ_i) eigenpair
⇔ all columns of P are eigenvectors
P invertible B_{ℂ^n} = {p_1, …, p_n}
Thm Let R be upper triangular then σ(R) = {r_11, …, r_nn}
Prf R' := R − λI
det(R') = Σ_σ sign(σ) r'_{1σ(1)} ⋯ r'_{nσ(n)}
= (r_11 − λ) ⋯ (r_nn − λ)
[1 0; 0 Ψ^T] B^T A B [1 0; 0 Ψ] : upper triangular
unitary
bottom of P. 508 Meyer. (think about it)
Thm (Schur upper-triangularizability)
A ∈ ℂ^{n×n}, ∃ U ∈ ℂ^{n×n}, unitary
U*U = UU* = I_n s.t. U*AU is upper triangular (U^{-1} = U*)
Bx = (I − 2uu^T/||u||_2^2) x =
u^T x = (x^T − e_1^T)x = ||x||_2^2 − x_1
= x − 2(x − e_1)(||x||_2^2 − x_1) / (x^T − e_1^T)(x − e_1) =
||x||_2 = 1
= x − 2(x − e_1)(1 − x_1) / (1 − x_1 − x_1 + 1) = x − 2(x − e_1)(1 − x_1) / 2(1 − x_1)
= x − (x − e_1) = e_1
x ∈ ℝ^{n×1}, ||x||_2 = 1, Rx = e_1, R: unitary
If x = e_1, R = I, suppose |x_1| ≠ 1
(|x_1| ≠ 1 ⇔ x ≠ ±1) because ||x||_2 = 1
u = x − e_1, B = I − 2uu^T/||u||_2^2 (reflection)
uu^T/||u||_2^2 = ûû^T, û := u/||u||_2
orthogonal projection onto Span(u)
I − ûû^T orthogonal proj. onto plane with normal û
I − 2ûû^T reflection with respect to that plane
(uu^T)^T = uu^T
BR^T = (I − 2ûû^T)(I − 2ûû^T)^T =
= I − 2ûû^T − 2ûû^T + 4ûû^Tûû^T = I − 4ûû^T + 4û(û^Tû)û^T = I
Involution (R^{-1} = R)
Prf (Schur's upper triang.)
By induction on n
n = 1 immediate
Hypothesis: all (n−1)×(n−1) matrices are unitarily upper triangularizable
Induction step: show A ∈ ℂ^{n×n} is un. up tr.
Take (x, λ) to be eigenpair. ||x||_2 = 1
∃ unitary B s.t. Bx = e_1 ⇒ first column of B is x ⇒ B = [x V]
B^T A B = [x^T; V^T] A [x V] =
= [x^T Ax x^T AV; V^T Ax V^T AV] = [λ x^T AV; V^T(λx) V^T AV]
V^T x = 0
= [λ x^T AV; 0 V^T AV]
Bx = e_1 ⇒ B^T B = I
B^T Bx = B^T e_1
x = B^T e_1 = first column of B
V^T AV ∈ ℂ^{(n−1)×(n−1)}
by induction ∃ Ψ ∈ ℂ^{(n−1)×(n−1)} Ψ^T Ψ = ΨΨ^T = I
s.t. V^T AV = T.W Ψ^T

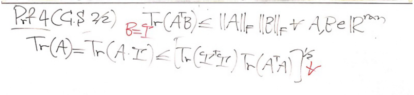
lecture 14

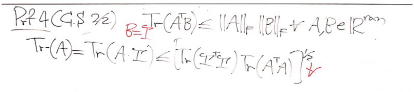

Prp Not always true: μg(λ) = μa(λ)
Thm ∀ A∈ℂ^{n×n}, λ∈σ(A), μg(λ) ≤ μa(λ)
Prf By Schur's Thm
P. 511

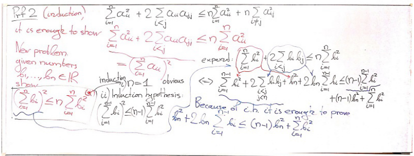

Dfn A∈ℂ^{n×n}, λ∈σ(A)⊂ℂ
μa(λ): algebraic multiplicity of λ = # times (x−λ) appears in pA(λ)
μg(λ): geometric multiplicity of λ = dim Eλ = dim N(A−λI)
Dfn λ∈σ(A), Eλ = {x∈ℂ^n | Ax=λx} = eigenspace of λ

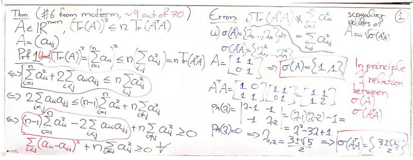

Dfn A∈ℂ^{n×n} diagonalizable if ∃ P (invertible), Λ diagonal s.t. A = PΛP^{-1}
Dfn A∈ℂ^{n×n} unitary diagonalizable if ∃ U (unitary), Λ s.t. A = UΛU*
Thm A∈ℂ^n unit. diag ⟺ AA* = A*A (normal) (P. 547/Meyer)
Thm A∈ℂ^n, A* = A ⟹ A is unit. diag.


Thm (Schur Unitary Triangularizability, p.508)
Lem ∀ x∈ℂ^n ∃ reflection


Prf 4 (C.S 22)
Tr(A^T B) ≤ ||A||_F ||B||_F ∀ A,B ∈ ℝ^{n×n}
Tr(A) = Tr(A·I) ≤ [Tr(I^T I) Tr(A^T A)]^{1/2}




Prf 2 (induction)
it is enough to show
New problem
induction
obvious
expand


Thm (#6 from midterm, ~9 out of 70)
A∈ℝ^{n×n}, (Tr(A))^2 ≤ n Tr(A^T A)
Errors
In principle no relation between σ(A) σ(A^T A)

lecture 15

Dfn $A\in\mathbb{R}^{n\times n}$, $A^T=A$, "positive semi-definite" if $x^TAx\geq 0\ \forall x\in\mathbb{R}^n$, $A\geq 0$

"negative semi-definite" if $x^TAx\leq 0\ \forall x\in\mathbb{R}^n$ $A\leq 0$

"positive-definite", $x^TAx>0\ \forall x\in\mathbb{R}^n\setminus\{0\}$ $A>0$

Dfn A principal submatrix is one obtained by removing columns and rows of the same indices. $\sum_{k=0}^{n-1}\binom{n}{k}=\frac{n!}{k!(n-k)!}$ $A\succ 0$

Thm

1) $A\geq 0 \Leftrightarrow$
2) $\exists B\in\mathbb{R}^{n\times n}$ s.t. $A=B^TB \Leftrightarrow$
3) $\sigma(A)\subset\mathbb{R}_{\geq 0} \Leftrightarrow$
4) every principal minor of is ≥ 0

$1)\Rightarrow 3)\Rightarrow 2)\Rightarrow 1)$

Prf $1)\Rightarrow 3)$ (x,λ), $Ax=\lambda x \Rightarrow 0\leq x^TAx=\lambda x^Tx=\lambda\|x\|_2^2\Rightarrow\lambda\geq 0$

$3)\Rightarrow 2)$ $A^T=A\Rightarrow A=U\Lambda U^T=U\Lambda^{1/2}\Lambda^{1/2}U^T=(U\Lambda^{1/2})(U\Lambda^{1/2})^T=BB^T$

$2)\Rightarrow 1)$ $A=BB^T$, $x^TAx=x^TBB^Tx=(B^Tx)^TB^Tx=\|B^Tx\|_2^2\geq 0$

"G-F" Thm $A^T=A$, $\lambda_1(A)\geq\lambda_2(A)\geq\dots\geq\lambda_n(A)$

$$\boxed{\lambda_k(A)=\max_{\mathcal{V}:\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}} x^TAx}$$

Prf (λ_i,u_i), $u_i^Tu_j=\delta_{ij}$, suppose we have some $\mathcal{V}$ of dimension k

$\mathcal{W}=\underset{k}{\mathcal{V}}\cap\underset{+}{}\operatorname{Span}(u_k,u_{k+1},\dots,u_n)\neq 0$ $\quad n-k+1\ =n+1$

* take $x\in\mathcal{W}$, $\|x\|_2=1$, $x=\sum_{i=k}^n c_iu_i$, $x^TAx=\left(\sum_{j=k}^n c_ju_j^T\right)A\left(\sum_{i=k}^n c_ju_j\right)=\sum_{i=k}^n c_i^2\lambda_i\leq$

$\boxed{\max_{\mathcal{V}}\min_{x} x^TAx\leq\lambda_k(A)}\Leftarrow\lambda_k\sum_{i=k}^n c_i^2=\lambda_k(A)$

* take $\mathcal{V}=\operatorname{Span}(u_1,\dots,u_k)$

$x\in\mathcal{V}$, $\|x\|_2=1$, $x=\sum_{i=1}^k c_iu_i$

$x^TAx=\sum_{i=1}^k\lambda_ic_i^2\overset{\lambda_i\geq\lambda_k}{\geq}\lambda_k\sum_{i=1}^k c_i^2=\lambda_k\Rightarrow\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx\geq\lambda_k\Rightarrow\boxed{\max_{\mathcal{V}}\min_{x\in\mathcal{V}}x^TAx\geq\lambda_k(A)}$

$\forall\mathcal{V}$ s.t. $\dim\mathcal{V}=k$ $\exists\underset{x\neq 0}{x\in}\mathcal{W}:=\mathcal{V}\cap\operatorname{Span}(u_k,\dots,u_n)$

$x=\sum_{i=k}^n c_iu_i\Rightarrow x^TAx=\sum_{i=k}^n c_i^2\lambda_i\leq\lambda_k\Rightarrow\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx\leq\lambda_k$

but this is true $\forall\mathcal{V}$ of dim. k

$\left[\forall\mathcal{V}\text{ s.t. }\dim\mathcal{V}=k\Rightarrow\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx\leq\lambda_k\right]\Rightarrow\max_{\mathcal{V}:\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx\leq\lambda_k$.

Thm (Weyl-I) $A^T=A$, $B^T=B$, $\lambda_1(A)\geq\lambda_2(A)\geq\dots\geq\lambda_n(A)$

$\lambda_1(B)\geq\lambda_2(B)\geq\dots\geq\lambda_n(B)$

$\lambda_k(A)+\lambda_n(B)\leq\lambda_k(A+B)\leq\lambda_k(A)+\lambda_1(B)$

$\lambda_k(A)$
$\lambda_k(B)$
$\lambda_k(A+B)$
$\det(A+B-\lambda I)$
$P_{A+B}(x)$
$P_A(x), P_B(x)$

Prf $\lambda_k(A+B)=\max_{\mathcal{V}:\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^T(A+B)x=\max_{\text{"G-F"}}\min\left(x^TAx+x^TBx\right)$

$\geq\max_{\mathcal{V}:\dim\mathcal{V}=k}\left[\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}(x^TAx)+\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}(x^TBx)\right]\geq\max_{\dim\mathcal{V}=k}\left[\min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}(x^TAx)+\lambda_n(B)\right]$

$\min_x[f(x)+g(x)]\geq\min_x f(x)+\min_x g(x)$

Prf $x_f=\operatorname{argmin}_x f(x)$, $x_g=\operatorname{argmin}_x g(x)$

$x_{fg}=\operatorname{argmin}_x[f(x)+g(x)]$

$\min_x[f(x)+g(x)]=f(x_{fg})+g(x_{fg})\geq f(x_f)+g(x_g)=\min_x f(x)+\min_x g(x)$

$=\max_{\text{"G-F"}}\min(x^TAx)+\lambda_n(B)=\lambda_k(A)+\lambda_n(B)$

this proves $\lambda_k(A)+\lambda_n(B)\leq\lambda_k(A+B)$

for $\lambda_k(A+B)\leq\lambda_k(A)+\lambda_1(B)$ $\quad\lambda_k(A)=\max_{\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx=\min_{\dim\mathcal{V}=n-k+1}\ \max_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^TAx$

$\lambda_k(A+B)=\min_{\text{"G-F"}}\max\left(x^TAx+x^TBx\right)\leq\min_{\text{"G-F"}}\left(\max_{\text{"G-F"}}x^TAx+\max_{\text{"G-F"}}x^TBx\right)\leq\min\left(\max x^TAx+\lambda_1(B)\right)$

$=\min\max(x^TAx)+\lambda_1(B)$

$=\lambda_k(A)+\lambda_1(B)$

Thm ("Interlacing I") $A^T=A$, $\alpha\in\mathbb{R}$, $y\in\mathbb{R}^n$

$\lambda_{k+1}(A)\leq\lambda_k(A+\alpha\underbrace{yy^T}_{\text{rank }1})\leq\lambda_{k-1}(A)$

Prf $\lambda_k(A+\alpha yy^T)=\max_{\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}}x^T(A+\alpha yy^T)x\leq\max_{\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1\\ x\perp y}}x^T(A+\alpha yy^T)x=\max_{\dim\mathcal{V}=k}\ \min_{\substack{x\in\mathcal{V}\\ \|x\|_2=1\\ x\perp y}}x^TAx$

$=\max_{\substack{\mathcal{V}'=\mathcal{V}\cap\operatorname{Span}(y)^\perp\\ \dim\mathcal{V}=k}}\ \min_{\substack{x\in\mathcal{V}'\\ \|x\|_2=1}}x^TAx\leq\max_{\substack{\mathcal{V}'=\mathcal{V}\cap\operatorname{Span}(z)^\perp\\ z\in\mathbb{R}^n,\dim\mathcal{V}=k}}\ \min_{\substack{x\in\mathcal{V}'\\ \|x\|_2=1}}x^TAx$ $\quad\lambda_k(A)\leq\lambda_{k-1}(A)$

$=\max_{\mathcal{V}':\dim\mathcal{V}'=k-1}\ \min_{\substack{x\in\mathcal{V}'\\ \|x\|_2=1}}x^TAx\overset{\text{"G-F"}}{=}\lambda_{k-1}(A)$

Let $(v_1,\dots,v_{n-k})=B_{\mathcal{V}^\perp}$, $x\in\mathcal{V}\Leftrightarrow x\perp v_1,\dots,v_{n-k}$

1) if $z\in\mathcal{V}\Rightarrow\{z,v_1,\dots,v_{n-k}\}$ is l.i. $\Rightarrow\dim\operatorname{Span}(z,v_1,\dots,v_{n-k})^\perp=k-1$

$x\in\mathcal{V}\cap\operatorname{Span}(z)^\perp\Leftrightarrow\begin{matrix}x\perp v_1,\dots,v_{n-k}\\ x\perp z\end{matrix}\Leftrightarrow x\perp z,v_1,\dots,v_{n-k}$

$\Rightarrow x$ is in $(k-1)$-dimensional space

lecture 16

Dfn $A \in \mathbb{C}^{n\times n}$, $\mu_a(\lambda)$: number of times $(x-\lambda)$ appears in $p_A(x) = \det(A - xI)$

$p_A(x) = (x-\lambda)^{\mu_a(\lambda)} g(x)$, $g(\lambda) \neq 0$, $\mu_g(\lambda) = \dim \mathcal{E}_\lambda = \dim N(A - \lambda I)$

Thm $\forall \lambda \in \sigma(A)$, $\mu_g(\lambda) \leq \mu_a(\lambda)$

Proof (Meyer), By Schur's thm $A = UTU^*$, U: unitary, T: upper triangular

$v := \mu_a(\lambda)$, $T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$, $T_{11} = \begin{bmatrix} \lambda & * \\ 0 & \lambda \end{bmatrix} \in \mathbb{C}^{v\times v}$, $T_{22} = \begin{bmatrix} \xi_1 & * \\ 0 & \xi_{n-v} \end{bmatrix}$, $\xi_i \neq \lambda$ $\Rightarrow T_{22} - \lambda I$: invertible

$T = \begin{bmatrix} t_{11} & & * \\ & t_{22} & \\ 0 & & \ddots \\ & & & t_{nn} \end{bmatrix}$

$p_A(x) = p_T(x) = \det(T - xI) = (x - t_{11}) \cdots (x - t_{nn})$

$\operatorname{rank}(A - \lambda I) = \operatorname{rank}(U^*(A - \lambda I)U) = \operatorname{rank}(T - \lambda I) = \operatorname{rank}\begin{pmatrix} T_{11} - \lambda I_v & T_{12} \\ 0 & T_{22} - \lambda I_{n-v} \end{pmatrix}$

$\geq \operatorname{rank}(T_{22} - \lambda I_{n-v}) = n - v$

$\mu_g(\lambda) = \dim N(A - \lambda I) = n - \operatorname{rank}(A - \lambda I) \leq n - (n - v) = v = \mu_a(\lambda)$

Prf (Roman, thm 8.5) $k := \mu_g(\lambda) = \dim \mathcal{E}_\lambda$, let $B_\lambda = (v_1, \ldots, v_k)$ be basis for $\mathcal{E}_\lambda$

extend B_λ to a basis $B = [B_\lambda\ B']$ of $\mathbb{C}^n$, $\tau_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$, $\tau_A(v) = Av$

$A = [\tau_A]_{\mathbb{E}}$, $\mathbb{E} = (e_1, \ldots, e_n)$, $[\tau_A]_B = \begin{pmatrix} \lambda I_k & C \\ 0 & D \end{pmatrix} \Rightarrow p_A(x) = \det\begin{pmatrix} \lambda I_k - xI_k & C \\ 0 & D - xI_{n-k} \end{pmatrix}$

$b_1 \in B_\lambda$, $\tau_A(b_1) = Ab_1 = \lambda b_1$

$B_\lambda = [b_1\ b_2 \cdots b_k]$

$= \det(\lambda I_k - xI_k)\det(D - xI_{n-k})$

$= (\lambda - x)^k \det(D - xI_{n-k}) \Rightarrow \mu_a(\lambda) \geq k := \mu_g(\lambda)$

Thm $A \in \mathbb{C}^{n\times n}$, $(x_1, \lambda_1), \ldots, (x_s, \lambda_s)$: eigenpairs, i.e., $Ax_i = \lambda_i x_i$, $\lambda_i \neq \lambda_j$ for $i \neq j$

$\{x_1, \ldots, x_s\}$ is l.i. (eigenvectors of distinct eigenvalues are l.i.)

Prf Suppose $x_1, \ldots, x_s$ are l.d. and wlog $x_1, \ldots, x_r$ are maximally l.i. $(r < s)$

$x_{r+1} = \sum_{i=1}^{r} \alpha_i x_i \Rightarrow 0 = (A - \lambda_{r+1} I)x_{r+1} = \sum_{i=1}^{r} \alpha_i (A - \lambda_{r+1} I)x_i = \sum_{i=1}^{r} \alpha_i \underbrace{(\lambda_i - \lambda_{r+1})}_{\neq 0} x_i$

$\Rightarrow \alpha_i(\lambda_i - \lambda_{r+1}) = 0 \Rightarrow \alpha_i = 0 \Rightarrow x_{r+1} = 0 \Rightarrow \Leftarrow$ (eigenvector $\neq 0$)

Thm $A \in \mathbb{C}^{n\times n}$, $\sigma(A) = \{\lambda_1, \ldots, \lambda_s\}$, $\lambda_i \neq \lambda_j$, $B_i = (u_1^{(i)}, \ldots, u_{v_i}^{(i)})$: basis for $\mathcal{E}_{\lambda_i}$

$B_1 \cup B_2 \cup \cdots \cup B_s$ is l.i.

Prf Suppose $(\alpha_1^{(1)} u_1^{(1)} + \cdots + \alpha_{v_1}^{(1)} u_{v_1}^{(1)}) + \cdots + (\alpha_1^{(s)} u_1^{(s)} + \cdots + \alpha_{v_s}^{(s)} u_{v_s}^{(s)}) = 0$

$\Rightarrow \underbrace{(\alpha_1^{(1)} u_1^{(1)} + \cdots + \alpha_{v_1}^{(1)} u_{v_1}^{(1)})}_{=0} \ \cdots \ \underbrace{(\alpha_1^{(s)} u_1^{(s)} + \cdots + \alpha_{v_s}^{(s)} u_{v_s}^{(s)})}_{=0}$

$\Rightarrow \alpha_1^{(1)} = \cdots = \alpha_{v_1}^{(1)} = 0$

Dfn $A \in \mathbb{C}^{n\times n}$, diagonalizable if $\exists$ invertible P, diagonal Λ s.t. $A = P\Lambda P^{-1}$

Thm A unitarily diagonalizable $\Leftrightarrow A^T A = AA^T$ (normal)

Thm $A \in \mathbb{C}^{n\times n}$ is diagonalizable $\Leftrightarrow \mu_a(\lambda) = \mu_g(\lambda)\ \forall \lambda \in \sigma(A)$ $\quad \sigma(A) = \{\lambda_1, \ldots, \lambda_s\}$

Prf $(\Rightarrow)$ Suppose $A = P\Lambda P^{-1} \Leftrightarrow AP = P\Lambda \Rightarrow P = [u_1 \cdots u_n] \overset{\text{wlog}}{=} [P_1\ P_2 \cdots P_s]$

u_i: eigenvector $\Leftrightarrow Au_i = \lambda_i u_i$

$P_i \subset \mathcal{E}_{\lambda_i}$, P_i is l.i., $P_i \in \mathbb{C}^{n\times n_i}$ $\Rightarrow \boxed{n_i \leq \mu_g(\lambda_i)}$

$n = \sum_{i=1}^{s} n_i \leq \sum_{i=1}^{s} \mu_g(\lambda_i) \leq \sum_{i=1}^{s} \mu_a(\lambda_i) = n \Rightarrow \sum_{i=1}^{s} \mu_g(\lambda_i) = \sum_{i=1}^{s} \mu_a(\lambda_i)$

$\Rightarrow \mu_g(\lambda_i) = \mu_a(\lambda_i)$

$(\Leftarrow)$ Suppose $\mu_a(\lambda) = \underbrace{\mu_g(\lambda)}_{=: n_i}\ \forall \lambda \in \sigma(A)$

let $P_i \in \mathbb{C}^{n\times n_i}$: basis for $\mathcal{E}_{\lambda_i}$, by previous thm $P = [P_1 \cdots P_s]$ has l.i. columns

but $n = \sum \mu_a(\lambda) = \sum \mu_g(\lambda) = \sum n_i \Rightarrow P \in \mathbb{C}^{n\times n} \Rightarrow P$ invertible

$AP_i = P_i(\lambda_i I_{n_i}) \overset{\forall i}{\Longrightarrow} AP = P\Lambda \Rightarrow A = P\Lambda P^{-1}$

Thm $A = A^T$, $\lambda_i \neq \lambda_j \Rightarrow \mathcal{E}_{\lambda_i} \perp \mathcal{E}_{\lambda_j}$

Prf (1) $A = U\Lambda U^T$, $U = [U_1\ U_2 \cdots U_s]$, U_i: basis for $\mathcal{E}_{\lambda_i}$, $U^T U = I$

$U^T U = \begin{bmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_s^T \end{bmatrix} [U_1 \cdots U_s] = \begin{bmatrix} U_1^T U_1 & U_1^T U_2 & \cdots & U_1^T U_s \\ U_2^T U_1 & U_2^T U_2 & \cdots & U_2^T U_s \\ \vdots \\ U_s^T U_1 & U_s^T U_2 & \cdots & U_s^T U_s \end{bmatrix} = I_n = \begin{bmatrix} I_{n_1} & & \\ & I_{n_2} & \\ & & \ddots & \\ & & & I_{n_s} \end{bmatrix} \Rightarrow U_i^T U_j = 0,\ i \neq j$

(2) (λ_i, x_i), (λ_j, x_j)

$0 = ((A - \lambda_j I)x_j)^T x_i = x_j^T (A^T - \lambda_j I)x_i = x_j^T (A - \lambda_j I)x_i = x_j^T (\lambda_i - \lambda_j)x_i = (\lambda_i - \lambda_j) x_j^T x_i \overset{\lambda_i \neq \lambda_j}{\Longrightarrow} x_i \perp x_j$

Thm $A^T = A \Rightarrow \sigma(A) \subset \mathbb{R}$ **Prf** $Ax = \lambda x \Rightarrow x^* A^* = \lambda^* x^*$

$A = A^*$

$\Rightarrow x^* A = \lambda^* x^* \Rightarrow (x^* A)x = (\lambda^* x^*)x \Rightarrow x^*(\lambda x) = (\lambda^* x^*)x \Rightarrow \lambda(x^* x) = \lambda^*(x^* x)$

$\Rightarrow \lambda \|x\|_2^2 = \lambda^* \|x\|_2^2 \Rightarrow \lambda = \lambda^*$

Thm (Courant-Fischer), $A = A^T$, $p_A(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_n)$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, $\lambda_i = \max_{\dim V = i} \min_{\substack{x \in V \\ \|x\|_2 = 1}} x^T Ax = \min_{\dim V = n-i+1} \max_{\substack{x \in V \\ \|x\|_2 = 1}} x^T Ax$

Cor $i = 1$: $\lambda_1 = \max_{\dim V = 1} \min_{\substack{x \in V \\ \|x\|_2 = 1}} x^T Ax$, if $\dim V = 1 \Rightarrow V = \operatorname{span}(\vec{v})$, $\|v\|_2 = 1$

$= \max_{\|v\|_2 = 1} (v)^T Av = \boxed{\max_{\|x\|_2 = 1} x^T Ax = \lambda_1} = \max_{x \neq 0} \frac{x^T Ax}{x^T x}$: Rayleigh quotient

$\boxed{\lambda_n = \min_{\|x\|_2 = 1} x^T Ax}$

Thm $A^T = A$, $\lambda_{max} = \max_{\|x\|_2 = 1} x^T Ax$, $\lambda_{min} = \min_{\|x\|_2 = 1} x^T Ax$

Prf $A = U\Lambda U^T$, $= \max_{\substack{\|x\|_2 = 1 \\ x = Uy}} x^T Ax = \max_{\|Uy\|_2 = 1} y^T U^T AUy = \max_{\|y\|_2 = 1} y^T U^T U\Lambda U^T Uy$

$\|Uy\|_2^2 = 1 \Leftrightarrow y^T U^T Uy = 1 \Leftrightarrow y^T y = 1$

$= \max_{\|y\|_2 = 1} y^T \Lambda y = \max_{\|y\|_2 = 1} \sum_{i=1}^{n} \lambda_i y_i^2 \leq \lambda_1 \underbrace{\max_{\|y\|_2 = 1} \sum_{i=1}^{n} y_i^2}_{=1} = \lambda_1$

for equality choose $x = v$, where $Av = \lambda_1 v$, $\|v\|_2 = 1$

Thm $A^T = A \Rightarrow A = U\Lambda U^T$ "spectral theorem"

U_i: basis for $\mathcal{E}_{\lambda_i}$, $\sigma(A) = \{\lambda_1, \ldots, \lambda_s\}$, $U = [U_1 \cdots U_s]$ $\quad U_1 U_1^T + \cdots + U_s U_s^T = I$

$A = U\Lambda U^T = [U_1\ U_2 \cdots U_s] \begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \lambda_2 I_{n_2} & \\ & & \ddots & \\ & & & \lambda_s I_{n_s} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_s^T \end{bmatrix} = \lambda_1 U_1 U_1^T + \lambda_2 U_2 U_2^T + \cdots + \lambda_s U_s U_s^T$

orthogonal projection onto $\mathcal{E}_{\lambda_1}$

lecture 17

$\mathcal{V} \subset \mathbb{R}^n$, $y \in \mathbb{R}^n$, $\dim \mathcal{V} \cap \langle y \rangle^\perp \geq \dim \mathcal{V} - 1$ and $\dim \mathcal{V} \cap \langle y \rangle^\perp = \dim \mathcal{V} - 1 \iff \mathcal{V} \not\subset \langle y \rangle^\perp$ ①

Prf If $\langle y \rangle^\perp \supset \mathcal{V} \Rightarrow \mathcal{V} \cap \langle y \rangle^\perp = \mathcal{V}$

If $\mathcal{V} \not\subset \langle y \rangle^\perp$ $\exists x \in \mathcal{V} \setminus \langle y \rangle^\perp \Rightarrow \dim(\langle x \rangle + \langle y \rangle^\perp) = n \Rightarrow \dim(\mathcal{V} + \langle y \rangle^\perp) = n$

$\langle x \rangle + \langle y \rangle^\perp \subset \mathcal{V} + \langle y \rangle^\perp$

$\dim(\mathcal{V} \cap \langle y \rangle^\perp) = \dim \mathcal{V} + \dim \langle y \rangle^\perp - \dim(\mathcal{V} + \langle y \rangle^\perp)$
$= \dim \mathcal{V} + (n-1) - n = \dim \mathcal{V} - 1$

③ Cor (Inter-I) $A^T = A$, $B^T = B$, $\operatorname{rank}(B) = r$
$\lambda_{k+r}(A) \leq \lambda_k(A+B) \leq \lambda_{k-r}(A)$

Prf $B = V \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0 \end{bmatrix} V^T$, $V = [v_1\ v_2 \cdots v_n]$
$B = \lambda_1(B) v_1 v_1^T + \cdots + \lambda_r(B) v_r v_r^T$

$\lambda_k(A+B) = \lambda_k((A + \lambda_1(B) v_1 v_1^T + \cdots + \lambda_{r-1}(B) v_{r-1} v_{r-1}^T) + \underbrace{\lambda_r(B) v_r v_r^T}_{\alpha y y^T}) \overset{\text{Inter-I}}{\leq}$
$\lambda_{k-1}(A + \lambda_1(B) v_1 v_1^T + \cdots + \lambda_{r-1}(B) v_{r-1} v_{r-1}^T) \overset{\text{Inter-I}}{\leq} \lambda_{k-2}(A + \lambda_1(B) v_1 v_1^T + \cdots + \lambda_{r-2}(B) v_{r-2} v_{r-2}^T) \leq \cdots \leq \lambda_{k-r}(A)$

Thm (Interlacing-I) $A^T = A$, $\alpha \in \mathbb{R}$, $y \in \mathbb{R}^n$. $\lambda_{k+1}(A) \leq \lambda_k(A + \alpha y y^T) \leq \lambda_{k-1}(A)$ ②
$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$

Prf $\lambda_k(A + \alpha y y^T) = \max\limits_{\dim \mathcal{V} = k} \min\limits_{\substack{x \in \mathcal{V} \\ \|x\|_2 = 1}} x^T(A + \alpha y y^T)x \leq \max\limits_{\dim \mathcal{V} = k} \min\limits_{\substack{x \in \mathcal{V} \\ \|x\|_2 = 1 \\ x \perp y}} x^T A x = \max\limits_{\substack{\mathcal{V}' = \mathcal{V} \cap \langle y \rangle^\perp \\ \dim \mathcal{V} = k}} \min\limits_{\substack{x \in \mathcal{V}' \\ \|x\|_2 = 1}} x^T A x$
$\leq \max \left\{ \max\limits_{\dim \mathcal{V}' = k} \min\limits_{\substack{x \in \mathcal{V}' \\ \|x\|_2 = 1}} x^T A x,\ \max\limits_{\dim \mathcal{V}' = k-1} \min\limits_{\substack{x \in \mathcal{V}' \\ \|x\|_2 = 1}} x^T A x \right\} = \max\{\lambda_k(A), \lambda_{k-1}(A)\} = \lambda_{k-1}(A)$

$\Rightarrow \lambda_k(A + \alpha y y^T) \leq \lambda_{k-1}(A)$, $A' = A + \alpha y y^T \to \lambda_k(A') \leq \lambda_{k-1}(A' - \alpha y y^T)$
$\forall A, \forall \alpha, \forall y$

$\forall \alpha$, $\alpha' = -\alpha$
$\lambda_k(A') \leq \lambda_{k-1}(A' + \alpha' y y^T)$
$\forall k$, $k' = k+1$

$\lambda_{k+1}(A) \leq \lambda_k(A + \alpha y y^T)$ set $A' = A$, $\alpha' = \alpha$ $\Longleftarrow \lambda_{k+1}(A') \leq \lambda_k(A' + \alpha' y y^T)$

④ Remainder: Thm (Weyl-I) $A^T = A$, $B^T = B$, $\lambda_k(A) + \lambda_n(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_1(B)$
$k=1$: $\lambda_1(A+B) \leq \lambda_1(A) + \lambda_1(B)$

Thm (Weyl-II) $A^T = A$, $B^T = B$, $\lambda_{j+k-1}(A+B) \leq \lambda_j(A) + \lambda_k(B) \leq \lambda_{j+k-n}(A+B)$

Prf $A = U \Lambda U^T = [u_1 \cdots u_n] \begin{bmatrix} \lambda_1(A) & & \\ & \ddots & \\ & & \lambda_n(A) \end{bmatrix} [u_1 \cdots u_n]^T = \sum\limits_{s=1}^{n} \lambda_s(A) u_s u_s^T$

Convention $\lambda_\ell(A) = \begin{cases} -\infty, & \text{if } \ell > n \\ +\infty, & \text{if } \ell < 1 \end{cases}$

$A_{1:j-1} = \sum\limits_{s=1}^{j-1} \lambda_s(A) u_s u_s^T$
$B_{1:k-1} = \sum\limits_{s=1}^{k-1} \lambda_s(B) v_s v_s^T$
$\Rightarrow \lambda_j(A) = \lambda_1(A - A_{1:j-1})$, $\lambda_k(B) = \lambda_1(B - B_{1:k-1})$

Prove it first for $A, B \geq 0$

⑤ … $\lambda_{k+j-1}(A+B) \overset{\text{Inter-I}}{\leq} \lambda_1((A - A_{1:j-1}) + (B - B_{1:k-1})) \overset{\text{Weyl-I}}{\leq} \lambda_1(A - A_{1:j-1}) + \lambda_1(B - B_{1:k-1}) =$
$= \lambda_j(A) + \lambda_k(B)$

$\underbrace{A + B - A_{1:j-1} - B_{1:k-1}}_{\text{rank} \leq j+k-2 := r}$

$k+j-1 \longrightarrow k+j-1-(j+k-2) = 1$

$\lambda_k(H + \delta I) = \lambda_k(H) + \delta$

Suppose A, B general symmetric matrices
$\exists\ \delta \geq \max\{|\lambda_n(A)|, |\lambda_n(B)|\}$
$\Rightarrow \underbrace{A + \delta I}_{\tilde{A}}, \underbrace{B + \delta I}_{\tilde{B}} \geq 0$ (shift) $\Rightarrow$

$\lambda_{k+j-1}(A+B) + 2\delta = \lambda_{k+j-1}(A + B + 2\delta I) = \lambda_{k+j-1}(\tilde{A} + \tilde{B}) \leq \lambda_j(\tilde{A}) + \lambda_k(\tilde{B}) = \lambda_j(A + \delta I) + \lambda_k(B + \delta I)$
$= \lambda_j(A) + \delta + \lambda_k(B) + \delta$

$\lambda_{k+j-1}(A+B) + 2\delta \leq \lambda_j(A) + \lambda_k(B) + 2\delta$

Thm (Interlacing-II) $A^T = A$, B: principal $r \times r$ submatrix obtained by deleting rows columns $i_1, i_2, \ldots, i_{n-r} \Rightarrow \lambda_{n+k-r}(A) \leq \lambda_k(B) \leq \lambda_k(A)$

Prf $\lambda_{n+k-r}(A) = \max\limits_{\dim \mathcal{V} = n+k-r} \min\limits_{\substack{x \in \mathcal{V} \\ \|x\|_2 = 1}} x^T A x \leq \max\limits_{\dim \mathcal{V} = n+k-r} \min\limits_{\substack{x \in \mathcal{V} \\ \|x\|_2 = 1 \\ x \perp e_{i_1}, \ldots, e_{i_{n-r}}}} x^T A x =$
$= \max\limits_{\dim \mathcal{V} = n+k-r} \min\limits_{\substack{x \in \mathcal{V} \\ \|x\|_2 = 1 \\ x \perp e_{i_1}, \ldots, e_{i_{n-r}}}} x^T B' x =$
$= \max\limits_{\substack{\mathcal{V}' = \mathcal{V} \cap \langle e_{i_1}, \ldots, e_{i_{n-r}} \rangle^\perp \\ \dim \mathcal{V} = n+k-r}} \min\limits_{\substack{x \in \mathcal{V}' \\ \|x\|_2 = 1}} x^T B' x$

$\lambda_k(B) = \max \min\limits_{z \in \mathbb{R}^r} z^T B z$

$e_{i_1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i_1$, $x \perp e_{i_1} \Rightarrow x_{i_1} = 0$

$\det(A - \lambda I)$
$\det(B - \lambda I)$

$\mathcal{V} \subset \mathbb{R}^n$
$\mathcal{V} \cap \langle e_{i_1}, \ldots, e_{i_{n-r}} \rangle^\perp$
$\mathbb{R}^r$

without loss of generality $\{i_1, \ldots, i_{n-r}\} = \{r+1, \ldots, n\}$

$B' = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$

$\mathbb{R}^5 \ni [\ ] \to [\ ] \in \mathbb{R}^3$

Thm A subspace $\mathcal{V} \subset \mathbb{R}^n$ of dimension d is the intersection of $n-d$ hyperplanes

dimension

codimension

lecture 18

Thm (Interlacing-II) $A^T=A: n\times n$, $B: r\times r$ obtained by deleting rows/columns $i_1,\dots,i_{n-r}$

$\lambda_{n+k-r}(A)\le\lambda_k(B)\le\lambda_k(A)\quad\forall k=1,\dots,r$

wlog: $\{i_1,\dots,i_{n-r}\}=\{1,\dots,n-r\}$ $\quad A=\begin{bmatrix} * & * \\ * & B\end{bmatrix}$

Prf $\lambda_{n+k-r}(A)=\max\limits_{\substack{\mathcal{V}\subset\mathbb{R}^n\\ \dim\mathcal{V}=n+k-r}}\min\limits_{\substack{x\in\mathcal{V}\\ \|x\|_2=1}} x^TAx\le\max\limits_{\substack{\mathcal{V}\subset\mathbb{R}^n\\ \dim\mathcal{V}=n+k-r}}\min\limits_{\substack{x\in\mathcal{V},\ \|x\|_2=1\\ x\perp e_1,\dots,e_{n-r}}} x^TAx=$

$=\max\limits_{\substack{\mathcal{V}'=\mathcal{V}\cap\langle e_1,\dots,e_{n-r}\rangle^\perp\\ \mathcal{V}\subset\mathbb{R}^n\\ \dim\mathcal{V}=n+k-r}}\min\limits_{\substack{x\in\mathcal{V}'\\ \|x\|_2=1}} x^TAx=\max\limits_{\substack{\mathcal{V}'=\mathcal{V}\cap\langle e_1,\dots,e_{n-r}\rangle^\perp\\ \mathcal{V}\subset\mathbb{R}^n\\ \dim\mathcal{V}=n+k-r}}\min\limits_{\substack{x=\begin{bmatrix}0\\z\end{bmatrix}\in\mathcal{V}'\\ \|x\|_2=\|z\|_2=1\\ z\in\mathbb{R}^r}} z^TBz=$

$x^TAx=\begin{bmatrix}0^T & z^T\end{bmatrix}\begin{bmatrix} * & * \\ * & B\end{bmatrix}\begin{bmatrix}0\\z\end{bmatrix}=z^TBz$

$=\max\limits_{\substack{\mathcal{V}''\subset\mathbb{R}^r\\ \mathcal{V}''=\{z\in\mathbb{R}^r \mid \begin{bmatrix}0\\z\end{bmatrix}\in\mathcal{V}'\}\\ \mathcal{V}'=\mathcal{V}\cap\langle e_1,\dots,e_{n-r}\rangle^\perp\\ \dim\mathcal{V}=n+k-r}}\min\limits_{\substack{z\in\mathcal{V}''\\ \|z\|_2=1}} z^TBz\le\max\limits_{\substack{\mathcal{V}''\subset\mathbb{R}^r\\ \mathcal{V}''=\{z\in\mathbb{R}^r:\begin{bmatrix}0\\z\end{bmatrix}\in\mathcal{V}'\}\\ \mathcal{V}'\subset\mathbb{R}^n\\ \dim\mathcal{V}'\ge k}}\min\limits_{\substack{z\in\mathcal{V}''\\ \|z\|_2=1}} z^TBz=\max\limits_{\substack{\mathcal{V}''\subset\mathbb{R}^r\\ \dim\mathcal{V}''\ge k}}\min\limits_{\substack{z\in\mathcal{V}''\\ \|z\|_2=1}} z^TBz\le$

$\le\lambda_k(B)$

intersection of $n-r$ hyperplanes $[\mathcal{V}\cap\langle e_1\rangle^\perp\cap\langle e_2\rangle^\perp\cap\dots\cap\langle e_{n-r}\rangle^\perp$ $\quad\dim\mathcal{V}=n+k-r$ $\quad\dim\mathcal{V}\cap(\dots)\ge k$

Prp $\langle e_1,\dots,e_{n-r}\rangle^\perp=$

$=\langle e_1\rangle^\perp\cap\dots\cap\langle e_{n-r}\rangle^\perp$

Prf $x\in\langle e_1,\dots,e_{n-r}\rangle^\perp$

$\Leftrightarrow x\perp e_i\ \forall i=1,\dots,n-r$

$\Leftrightarrow x\in\langle e_i\rangle^\perp\ \forall i=1,\dots,n-r$

$\Leftrightarrow x\in\langle e_1\rangle^\perp\cap\dots\cap\langle e_{n-r}\rangle^\perp$

Prp $\mathcal{V}=\langle b_1,\dots,b_s\rangle^\perp$

$\Rightarrow\mathcal{V}=\langle b_1\rangle^\perp\cap\dots\cap\langle b_s\rangle^\perp$

Prf $x\in\mathcal{V}\Leftrightarrow$

$x\perp b_i\ \forall i$

$\Leftrightarrow x\in\langle b_i\rangle^\perp\ \forall i$

$\Leftrightarrow x\in\bigcap\limits_{i=1}\langle b_i\rangle^\perp$

$x\in\mathcal{V}\Leftrightarrow x\perp\mathcal{V}^\perp\Leftrightarrow x\perp b_i\ \forall i\Leftrightarrow x\in\langle b_i\rangle^\perp\ \forall i\Leftrightarrow x\in\bigcap\limits_{i=1}^{n-k}\langle b_i\rangle^\perp$

$\mathcal{V}^\perp=\langle b_1,\dots,b_{n-k}\rangle$

Thm $\mathcal{V}\subset\mathbb{R}^n$, $\dim\mathcal{V}=k$

$\mathcal{V}=$ intersection of $n-k$ hyperplanes

Prf $\mathbb{R}^n=\mathcal{V}\oplus\mathcal{V}^\perp$

$\dim\mathcal{V}^\perp=n-k$

Thm (Variational characterization of sum of eigenvalues)

$A^T=A: n\times n$ $\quad$ i) $\sum\limits_{i=1}^k\lambda_i(A)=\max\limits_{\substack{U: n\times k\\ U^TU=I_k}}\operatorname{tr}(U^TAU)$

ii) $\sum\limits_{i=n-k+1}^n\lambda_i(A)=\min\limits_{\substack{U: n\times k\\ U^TU=I_k}}\operatorname{tr}(U^TAU)$

$U^TAU: k\times k$

$A: n\times n$

Prf given U extend it (via G-S) to $V=[U\ Q]: n\times n$, $V^TV=I_n$

V^TAV is similar to $A\Rightarrow\lambda_i(A)=\lambda_i(V^TAV)$

$V^TAV=\begin{bmatrix}U^T\\Q^T\end{bmatrix}A[U\ Q]=\begin{bmatrix}U^TAU & * \\ * & *\end{bmatrix}$

principal $k\times k$ submatrix of V^TAV

By Interlacing-II $\quad\lambda_i(\underbrace{U^TAU}_{B})\le\lambda_i(\underbrace{V^TAV}_{A})=\lambda_i(A)\quad(*)$

sum $(*)$ for $i=1,\dots,k\Rightarrow\operatorname{tr}(U^TAU)=\sum\limits_{i=1}^k\lambda_i(U^TAU)\le\sum\limits_{i=1}^k\lambda_i(A)\Rightarrow$

$\forall U:\ \operatorname{tr}(U^TAU)\le\sum\limits_{i=1}^k\lambda_i(A)$ $\quad$ Q: How to pick U to achieve equality?

$(u_i,\lambda_i(A))$: eigenpairs $\quad u_i^Tu_j=\delta_{ij}$

$U=[u_1\cdots u_k]$, $\quad U^TAU=\begin{bmatrix}u_1^T\\u_2^T\\ \vdots\\u_k^T\end{bmatrix}A[u_1\cdots u_k]=\begin{bmatrix}u_1^T\\ \vdots\\u_k^T\end{bmatrix}[\lambda_1u_1\cdots\lambda_ku_k]=\begin{bmatrix}\lambda_1u_1^Tu_1 & \lambda_2u_1^Tu_2 & \cdots & \lambda_ku_1^Tu_k\\ \lambda_1u_2^Tu_1 & \lambda_2u_2^Tu_2 & \cdots & \lambda_ku_2^Tu_k\\ \vdots & & & \\ \lambda_1u_k^Tu_1 & \cdots & \cdots & \lambda_ku_k^Tu_k\end{bmatrix}=$

$=\begin{bmatrix}\lambda_1(A) & & \\ & \lambda_2(A) & \\ & & \ddots\ \lambda_k(A)\end{bmatrix}$

Def $x=(x_1,x_2,\dots,x_n)$ $\quad y=(y_1,\dots,y_n)$

$x_1\ge x_2\ge\dots\ge x_n$ $\quad y_1\ge y_2\ge\dots\ge y_n$

"x majorizes y, $x\succ y$" if

$\forall k=1,\dots,n\quad\sum\limits_{i=1}^kx_i\ge\sum\limits_{i=1}^ky_i$

Thm $A^T=A$, $\underline{\lambda}(A)=\begin{bmatrix}\lambda_1(A)\\ \vdots\\ \lambda_n(A)\end{bmatrix}$

$\underline{\alpha}=\begin{bmatrix}\alpha_{i_1i_1}\\ \alpha_{i_2i_2}\\ \vdots\\ \alpha_{i_ni_n}\end{bmatrix}$, $\alpha_{i_1i_1}\ge\alpha_{i_2i_2}\ge\dots\ge\alpha_{i_ni_n}$ Then $\underline{\lambda}(A)\succ\underline{\alpha}$ and we have $\sum\limits_{i=1}^n\lambda_i(A)=\sum\limits_{i=1}^n\alpha_{ii}=\operatorname{tr}(A)$

Prf By induction on n

for $n=1$ $\quad A=\alpha_{11}=\lambda_1(A)$ ✓

suppose the statement is true $\forall$ $(n-1)\times(n-1)$ symmetric matrices

wlog $\alpha_{11}\ge\alpha_{22}\ge\dots\ge\alpha_{nn}$

Let B be the $(n-1)\times(n-1)$ matrix obtained by deleting the n^{th} row/column $\Rightarrow B$: principal

$\lambda_i(A)\overset{\text{Int-II}}{\ge}\lambda_i(B)\Rightarrow$ $\quad A=\begin{bmatrix}B & \vdots\\ \cdots & \end{bmatrix}$

$\sum\limits_{i=1}^k\lambda_i(A)\overset{\text{Int-II}}{\ge}\sum\limits_{i=1}^k\lambda_i(B)\overset{\text{i.h.}}{\ge}\sum\limits_{i=1}^k\alpha_{ii}\quad\forall k=1,\dots,n-1$

for $k=n$ we already know $\sum\limits_{i=1}^n\lambda_i(A)=\operatorname{tr}(A)$ ∎

Thm (Hadamard's Inequality) $A: n\times n$, $A\ge0$, $\Rightarrow\det(A)\le\prod\limits_{i=1}^n\alpha_{ii}$

Prf $A\ge0\Rightarrow\alpha_{ii}\ge0$ (Q: Why? $e_i^TAe_i=\alpha_{ii}\ge0$) $\Rightarrow\prod\limits_{i=1}^n\alpha_{ii}\ge0$ $\quad$ If A singular $\Rightarrow\det(A)=0$ $\Rightarrow$ statement true.

So suppose $A>0\Rightarrow\alpha_{ii}>0$ $\quad D=\begin{bmatrix}1/\sqrt{\alpha_{11}} & & \\ & \ddots & \\ & & 1/\sqrt{\alpha_{nn}}\end{bmatrix}$, $\quad B=DAD\Rightarrow\operatorname{tr}(B)=n$

$\dfrac{\det(A)}{\prod\limits_{i=1}^n\alpha_{ii}}=\det(DAD)=\prod\limits_{i=1}^n\lambda_i(DAD)\le\left(\dfrac1n\sum\limits_{i=1}^n\lambda_i(DAD)\right)^n=\left(\dfrac1n\operatorname{tr}(DAD)\right)^n=\left(\dfrac1n n\right)^n=\underline{1}$ ∎

$\det(B)=\det(DAD)=\det(D)^2\det(A)=\dfrac{\det(A)}{\prod\limits_{i=1}^n\alpha_{ii}}$

Thm (arithmetic/geometric mean inequality)

$(x_1\cdots x_n)^{1/n}\le\dfrac1n(x_1+\dots+x_n)$

Roger Horn "Matrix Analysis"

Hw7 / Prb.1 $\quad\mathcal{U}\subset\mathbb{R}^m$, $\dim\mathcal{U}=d$, $B_{\mathcal{U}}=\{u_1,\dots,u_d\}$ $\quad\mathcal{U}'\subset\mathbb{R}^{m\times n}$, $\mathcal{U}'=\operatorname{Span}\{u_ix^T \mid \forall i=1,\dots,d,\ \forall x\in\mathbb{R}^n\}$

$\mathcal{V}\subset\mathbb{R}^n$, $\dim\mathcal{V}=d$, $B_{\mathcal{V}}=\{v_1,\dots,v_d\}$ $\quad\mathcal{V}'\subset\mathbb{R}^{m\times n}$, $\mathcal{V}'=\operatorname{Span}\{yv_j^T \mid j=1,\dots,d,\ y\in\mathbb{R}^m\}$

$\mathcal{W}=\mathcal{V}'+\mathcal{U}'\subset\mathbb{R}^{m\times n}$, $\dim\mathcal{W}=?$, what is a formula for orthogonal projection onto $\mathcal{W}$, $\mathcal{W}^\perp$ $\quad\dim\mathcal{W}=ms+ns-s^2$

$\pi_{\mathcal{W}}:\mathbb{R}^{m\times n}\longrightarrow\mathbb{R}^{m\times n}$ $\quad P_{\mathcal{U}}: m\times m$ $\quad A\in\mathbb{R}^{m\times n}$ $\quad A\in\mathcal{U}'\mapsto P_{\mathcal{U}}A\in\mathcal{U}'$

$\mathbb{R}^{mn}\longrightarrow\mathbb{R}^{mn}$ $\quad P_{\mathcal{V}}: n\times n$ $\quad A\in\mathcal{V}'\mapsto AP_{\mathcal{V}}$ $\quad\varphi(A)=P_{\mathcal{U}}A+AP_{\mathcal{V}}$ $\quad\varphi^2=\varphi?$

$\dim\mathcal{W}=\dim(\mathcal{V}'+\mathcal{U}')=\dim(\mathcal{V}')+\dim(\mathcal{U}')-\dim(\mathcal{V}'\cap\mathcal{U}')$

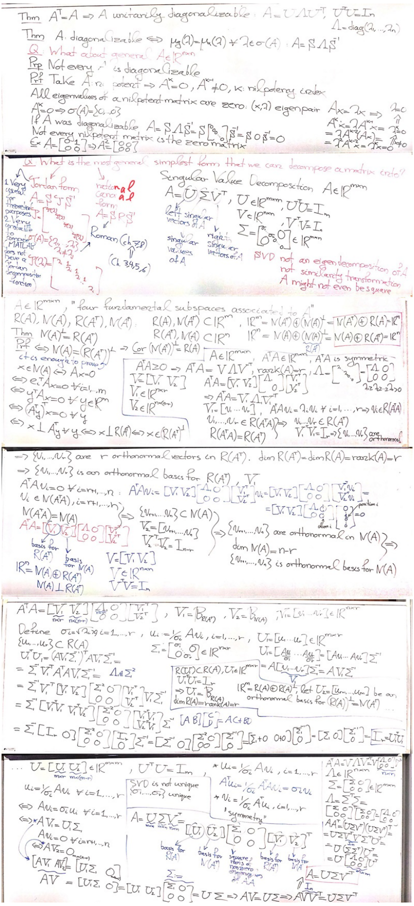
$\varphi(\varphi(A))=\varphi(P_{\mathcal{U}}A+AP_{\mathcal{V}})=\varphi(P_{\mathcal{U}}A)+\varphi(AP_{\mathcal{V}})$

$=P_{\mathcal{U}}A+P_{\mathcal{U}}AP_{\mathcal{V}}+P_{\mathcal{U}}AP_{\mathcal{V}}+AP_{\mathcal{V}}=\varphi(A)+2P_{\mathcal{U}}AP_{\mathcal{V}}$

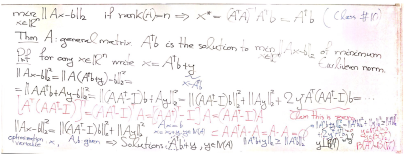
$\pi_{\mathcal{W}}(A)=P_{\mathcal{U}}A+AP_{\mathcal{V}}-P_{\mathcal{U}}AP_{\mathcal{V}}$

$B_{\mathcal{V}'\cap\mathcal{U}'}=\{u_iv_j^T\}\Rightarrow\dim\mathcal{V}'\cap\mathcal{U}'=d^2$

lecture 19


Singular Value Decomposition
SVD not an eigen decomposition of A
not similarity transformation
A might not even be square

lecture 20


Thm A: general matrix. A^+b is the solution to $\min_{x\in\mathbb{R}^n} \|Ax-b\|_2$ of minimum Euclidean norm.

$A \in \mathbb{R}^{n\times n}$ "several ways to measure how close A is to zero"

$0 \in \mathbb{R}^{n\times n}$, $\text{rank}(A) = n-1, n-2, \dots, 1, 0$

"more than one way to measure invertibility of A"

$\text{rank}(A) = n$, A^{-1}; $\text{rank}(A) = n-1$, $Ax = b \Rightarrow x = \underbrace{\text{"}A^{-1}\text{"}}_{\text{some kind of inverse}} b$

$A \in \mathbb{R}^{m\times n}$,

$\tau: V \to W$ τ is invertible $\Leftrightarrow \exists\, \varphi: W \to V$ s.t. $\varphi\circ\tau = 1_V$ left inversion; $\tau\circ\varphi = 1_W$ right inversion

Q: When is a linear transformation left-invertible?

Thm $\tau \in L(V,W)$ left-invertible $\Leftrightarrow \tau$ injective $\Leftrightarrow \ker(\tau) = 0$

Prf

$V \xrightarrow{\tau} W \xrightarrow{\varphi} V$, $\varphi\circ\tau = 1_V$

$V \xrightarrow{\tau} \underbrace{\text{im}(\tau)\oplus\text{im}(\tau)^c}_{W} \xrightarrow{\varphi} V$, $\tau: V \xrightarrow{\sim} \text{im}(\tau)$

Define $\varphi: W \to V$ by $\varphi|_{\text{im}(\tau)} = (\tau)^{-1}$; $\varphi|_{\text{im}(\tau)^c} = 0$

$A \in \mathbb{R}^{m\times n}$, $\tau_A: \mathbb{R}^n \to \mathbb{R}^m$, $\exists\, B: \mathbb{R}^m \to \mathbb{R}^n$ s.t. $BA = I_n \Leftrightarrow \tau_A$ injective

Q: How to test if A is injective? A: $\text{rank}(A) = n \Leftrightarrow$ columns of A linearly independent

$Ax = b$ ↳ suppose it is consistent $b \in \mathcal{R}(A)$

$\underbrace{[\ \ ]}_{A}[x] = \underbrace{[\ ]}_{b}$ if A injective A^TA: invertible because $\text{rank}(A^TA) = \text{rank}(A) = n$

$\Rightarrow Ax = b \Rightarrow A^TAx = A^Tb \Rightarrow \boxed{x = (A^TA)^{-1}A^Tb}$ "inverse" "pseudo-inverse"

Q: When is $\tau \in L(V,W)$ right-invertible?

Thm $\tau \in L(V,W)$ right-invertible $\Leftrightarrow \tau$ surjective

$W \xrightarrow{\varphi} V \xrightarrow{\tau} W$, 1_W

if τ not surjective $\exists\, w \in W$ s.t. $w \notin \text{im}(\tau)$

Thm $A \in \mathbb{R}^{m\times n}$ surjective $\Leftrightarrow \text{rank}(A) = m$

$\text{rank}(A) = m \Rightarrow \underbrace{AA^T}_{m\times m}$ invertible

$Ax = b$

$\boxed{x^* = A^T(AA^T)^{-1}b}$ is a solution "inverse"

$Ax^* = AA^T(AA^T)^{-1}b = b$

$A \in \mathbb{R}^{m\times n}$, $Ax = b$ assume it is consistent

$\text{rank}(A) = n \Leftrightarrow A$ injective: $x = (A^TA)^{-1}A^Tb$: unique solution

$A^\dagger = (A^TA)^{-1}A^T$: $n\times m$

$Ax = b$: overdetermined $[\ \ ][\ ] = [\ ]$ more equations than unknowns

$\text{rank}(A) = m \Leftrightarrow A$ surjective: $x = A^T(AA^T)^{-1}b$: one solution

$A^\dagger = A^T(AA^T)^{-1}$: $n\times m$

$[\ \ ][\ ] = [\ ]$ more unknowns than equations

Properties

injective: 1. $AA^\dagger A = A(A^TA)^{-1}A^TA = A$

2. $A^\dagger AA^\dagger = (A^TA)^{-1}A^TAA^\dagger = A^\dagger$

3. $(AA^\dagger)^T = (A^\dagger)^TA^T = ((A^TA)^{-1}A^T)^TA^T = A(A^TA)^{-1}A^T = AA^\dagger \Rightarrow AA^\dagger$: symmetric

4. $(A^\dagger A)^T = A^\dagger A$

Q: If A injective or surjective $\Rightarrow A$ has full rank because $\text{rank}(A) = \min\{n,m\}$

What if A is not full rank?

$A \in \mathbb{R}^{m\times n}$: any matrix **Dfn** We call $A^\dagger$ the "Moore-Penrose" inverse of A or "pseudo-inverse" or "generalized inverse" if

1. $AA^\dagger A = A$
2. $A^\dagger AA^\dagger = A^\dagger$
3. $(AA^\dagger)^T = AA^\dagger$
4. $(A^\dagger A)^T = A^\dagger A$

Thm $A \in \mathbb{R}^{m\times n}$, $A^\dagger \in \mathbb{R}^{n\times m}$ exists and it is unique.

Prf uniqueness Suppose $B, C \in \mathbb{R}^{n\times m}$ are pseudo-inverses of A, i.e., they satisfy 1–4

Then $AB = AC$ Prf $AB = \underbrace{(ACA)}_{A}B = (AC)(AB) = (AC)^T(AB)^T = (C^TA^T)(B^TA^T) =$ (property 3)

$= C^T(A^TB^TA^T) = C^T\underbrace{(ABA)}_{A}^T = C^TA^T = (AC)^T = AC$ (prp 3)

Also $BA = CA$ Prf $BA = B(ACA) = (BA)(CA) = (BA)^T(CA)^T = (A^TB^T)(A^TC^T) = (ABA)^TC^T = A^TC^T = (CA)^T = CA$

$C = CAC = (CA)C = (BA)C = B(AC) = BAB = B$

Existence: $A = U\Sigma V^T$: SVD of A

"full SVD", $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, $\Sigma = \begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix} \in \mathbb{R}^{m\times n}$

$A = [U_1\ U_2]\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_1^T\\ V_2^T\end{bmatrix} = U_1\Sigma_1V_1^T$: "thin SVD"

U_1: orthonormal basis for $\mathcal{R}(A)$; V_1: o.n. basis for $N(A)$

$U_1 \in \mathbb{R}^{m\times r}$ $U_1^TU_1 = I_r$

$V_1 \in \mathbb{R}^{n\times r}$ $V_1^TV_1 = I_r$

$\Sigma_1 \in \mathbb{R}^{r\times r}$

$r = \text{rank}(A)$

$A = U_1\Sigma_1V_1^T \in \mathbb{R}^{m\times n}$

$A^\dagger = V_1\Sigma_1^{-1}U_1^T \in \mathbb{R}^{n\times m}$

Check: $A(V_1\Sigma_1^{-1}U_1^T)A = \underbrace{(U_1\Sigma_1V_1^T)(V_1\Sigma_1^{-1}U_1^T)}_{U_1\Sigma_1\Sigma_1^{-1}U_1^T = U_1U_1^T}A = \underbrace{U_1U_1^T}_{\text{orthogonal projection onto } \mathcal{R}(A)}A = (U_1\underbrace{U_1^T)(U_1}_{I_r}\Sigma_1V_1^T) = U_1\Sigma_1V_1^T = A$

property 1 ✓

$(A(V_1\Sigma_1^{-1}U_1^T))^T = (U_1\Sigma_1^{-1}V_1^T)A^T =$ property 3 ✓

$= (U_1\Sigma_1^{-1}V_1^T)(V_1\Sigma_1U_1^T) = U_1U_1^T$

$(V_1\Sigma_1^{-1}U_1^T)A = V_1V_1^T$

lecture 21

Another thing: many times data is not numeric $\longrightarrow$ convert to numeric representation

Our setting: dataset = {data points} / similarities between pairs of points $\longleftrightarrow$ graph, weighted, undirected

$\mathcal{X} = \{1, 2, \dots, N\}$, $W \in \mathbb{R}^{N\times N}$

$|w_{ij}|$ large $\Leftrightarrow$ point i is similar to point j

φ: nullspace of "graph Laplacian"

<u>Dfn</u> Graph $G=(V,E,W)$ is a set of "nodes" $V=\{1,\dots,N\}$, a set of edges $E \subset V \times V$ and a set of weights $w_{ij} \in \mathbb{R}$, $w_{ij}=0$ if $(i,j) \notin E$.

<u>Rem</u> A graph is a topological space $\Rightarrow$ it is the union of its "connected components."

<u>Dfn</u> A graph is connected if $\exists$ path from every node to any other node. <u>Ex</u>

<u>Dfn</u> A connected component of G is a connected subgraph G' s.t. no edge exists between G' and $G \setminus G'$

<u>Prb</u> Find the connected components $G = G_1 \cup \dots \cup G_K$

<u>Rem</u> If similarity W is perfect (zero similarity for points in different clusters) then G has at least K components

oversegmentation

Assumption: There is a 1-1 correspondence between data clusters and connected components

$W \in \mathbb{R}^{n \times n}$, $W^T = W$: similarity / affinity matrix

Unormalized Graph $\quad W := \frac{W' + W'^T}{2}$

degree matrix $D = \mathrm{diag}(d_1,\dots,d_N)$, $d_i = \sum_{j=1}^{N} w_{ij}$, Laplacian : $L := D - W$

<u>Prp</u> $(D-W)\mathbb{1}_{N\times 1} = 0_{N\times 1}$, $\mathbb{1}_{N\times 1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

<u>Prf</u> $[(D-W)\mathbb{1}_{N\times 1}]_i = d_i - \sum_{j=1}^{N} w_{ij} = 0$ $\qquad \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in N(L)$, $0 \in \sigma(L)$

<u>Prp</u> $L \geq 0$
$L = D - W$

<u>Prf 1</u> $\left|\sum_{i\neq j} w_{ij} - \lambda\right| \leq \sum_{i \neq j} w_{ij} \Rightarrow \lambda \geq 0$ (Gerschgorin)

<u>Prf 2</u> Int.-II (?!)

<u>Prf 3</u> $\forall x \in \mathbb{R}^N$ $x^T L x = x^T D x - x^T W x = \sum_{i=1}^{N} \underbrace{\left(\sum_{j=1}^{N} w_{ij}\right)}_{d_i} x_i^2 - \sum_i \sum_j w_{ij} x_i x_j$
$= \sum_{i=1}^{N}\left(\sum_{j \neq i} w_{ij}\right) x_i^2 - 2\sum_{i<j} w_{ij} x_i x_j = \sum_{i<j} w_{ij}(x_i - x_j)^2 \geq 0$

$w_{kl} x_k^2 + w_{kl} x_l^2 - 2 w_{kl} x_k x_l = w_{kl}(x_k - x_l)^2$

Thm (Interlacing-II)
$A \in \mathbb{R}^{n\times n}$, $A^T = A$
B: principal $r\times r$ submatrix $\Rightarrow$
$\lambda_k(B) \leq \lambda_k(A)$
$\forall k = 1,\dots,r$

<u>Dfn</u> $G = G_1 \cup G_2 \cup \dots \cup G_K$

$\mathbb{1}_{G_i}(j) = \begin{cases} 0, & \text{if } j \notin G_i \\ 1, & \text{if } j \in G_i \end{cases}$

$\mathbb{1}_{G_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbb{1}_{G_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ indicator vector

G_1, G_2

<u>Rem</u> Each indicator vector gives us one cluster. If we had $\mathbb{1}_{G_1},\dots,\mathbb{1}_{G_n}$ we could get the clusters immediately

<u>Thm</u> $\{\mathbb{1}_{G_1},\dots,\mathbb{1}_{G_K}\}$ is a basis for $N(L)$

$\xi = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_2 \\ \vdots \\ \alpha_K \\ \vdots \end{bmatrix}_N = \alpha_1 \mathbb{1}_{G_1} + \alpha_2 \mathbb{1}_{G_2} + \dots + \alpha_K \mathbb{1}_{G_K}$

$\xi_i = \xi_j = \xi_k = \xi_l$

<u>Prf</u> $\alpha_1 \mathbb{1}_{G_1} + \dots + \alpha_K \mathbb{1}_{G_K} = 0 \Rightarrow \alpha_1 = \dots = \alpha_K = 0$

If $\xi \in \mathbb{R}^N \in N(L)$, $L\xi = 0 \Rightarrow \xi^T L \xi = 0 \Rightarrow \sum_{i\neq j} w_{ij}(\xi_i - \xi_j)^2 = 0 \overset{w_{ij}\geq 0}{\Longrightarrow} \xi_i = \xi_j$ whenever $w_{ij} \neq 0$

$\Rightarrow \xi \in \mathrm{Span}(\mathbb{1}_{G_1},\dots,\mathbb{1}_{G_K})$ <u>Cor</u> $\dim N(L) = k$

$U \in \mathbb{R}^{N\times k}$: orthonormal basis for $N(L)$

$U = [\mathbb{1}_{G_1} \cdots \mathbb{1}_{G_K}]A$ — $k\times k$ coefficient matrix

$U = \begin{bmatrix} U_{(1,:)} \\ U_{(2,:)} \\ \vdots \\ U_{(N,:)} \end{bmatrix} \in \mathbb{R}^{N\times k}$

construct $\varphi: \{1,\dots,N\} \to \mathbb{R}^k$ by $\varphi(j) = U_{(j,:)}$ — run K-means on $\tilde{\mathcal{X}} = \{\varphi(1), \varphi(2), \dots, \varphi(N)\} \subset \mathbb{R}^k$

<u>Prp</u> $\varphi(i) = \varphi(j) \Leftrightarrow i,j$ are in the same connected component

$[\mathbb{1}_{G_1}\, \mathbb{1}_{G_2} \cdots \mathbb{1}_{G_K}] = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} A$

Spectral Clustering = non-linear mapping + K-means

we are given data $\rightarrow$ partition them to K clusters

K-means: Input $\mathcal{X} \in \mathbb{R}^{D\times N}$, $\mathcal{X} = [x_1\, x_2 \cdots x_N]$, $x_j \in \mathbb{R}^D$

$K=3$

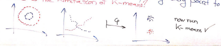
$\hookrightarrow$ Try to cluster $\mathcal{X}$ around K other points $\mu_1,\dots,\mu_K$ Centers

$\min_{\mu_1,\dots,\mu_K \in \mathbb{R}^D} \sum_{j=1}^{N} \min_{i=1,\dots,K} \|x_j - \mu_i\|_2$ $\rightarrow$ solve it with alternating minimization

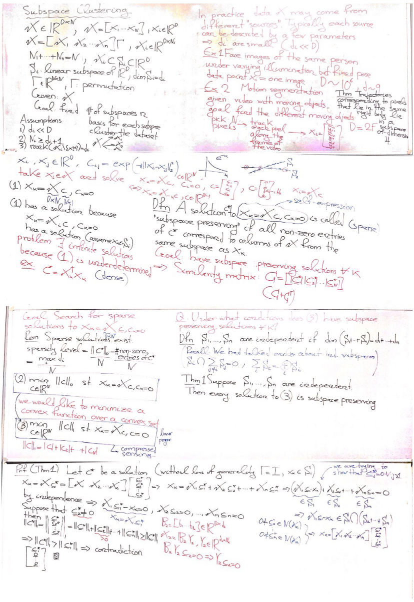
i) pick $\mu_1,\dots,\mu_K$
ii) assign every point to closest center $\Big\}\rightarrow$ converges in a finite number of steps

Q: What is the limitation of K-means?

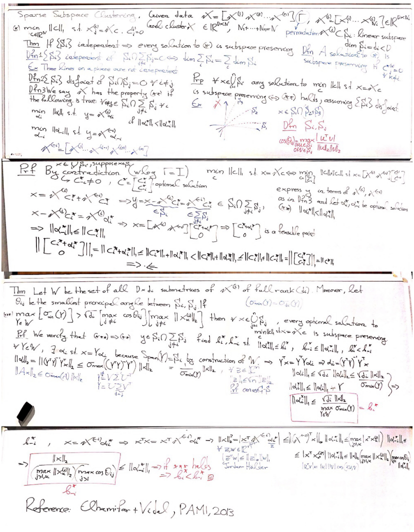
$\mathcal{X}$
$D=2$
$K=2$



lecture 22


Subspace Clustering

lecture 23


Sparse Subspace Clustering
Reference: Elhamifar + Vidal, PAMI, 2013

lecture 24

PCA

2017/12/14 08:25

let $X \in \mathbb{R}^{D\times n}$, usualli D is large. Try to find a subspace S of $\mathbb{R}^D$, s.t. $\dim S = d \ll D$. say, S passes as close as possible from the data in X.
Once S is computed, then find the "low dim repensentation" of X.

Let $U \in \mathbb{R}^{D\times d}$ the orthgonal basis of S, then

$P_s = UU^T \ \mathbb{R}^D \to \mathbb{R}^D$ (Orthogonal Projection: $\mathbb{R}^D \to \mathbb{R}^d \to \mathbb{R}^D$)

$\mathbb{R}^D \xrightarrow{U^T} \mathbb{R}^d \xrightarrow{U} \mathbb{R}^D$

Objective: find the best U, hence S.

$$\min_{\substack{U\in\mathbb{R}^{D\times d}\\ U^TU=I}} \sum_j \| x_j - UU^T x_j \|_2^2$$

Solution (via pure matrix analysis):

$$\| x_j - UU^T x_j \|_2^2 = (x_j - UU^Tx_j)^T (x_j - UU^T x_j)$$
$$= \|x_j\|_2^2 - 2x_j^T UU^T x_j + x_j^T UU^T x_j$$

$\Rightarrow$ (x_j has nothing to do with U)

$$\max_{U^TU=I_d} \sum_j x_j^T UU^T x_j = \max \sum_j tr(x_j^T UU^T x_j)$$
$$= \max \sum_j tr(U^T x_j \cdot x_j^T U)$$
$$= \max tr(\sum_j U^T x_j \cdot x_j^T U)$$
$$= \max tr(U^T R_x U^T) \quad (*)$$

note: $A^T = A$, then $\max_{U^TU=I_d} tr(U^TAU) = \sum_{i=1}^d \lambda_i^{\downarrow}(A)$.

$$\Rightarrow R_x U = U \begin{bmatrix} \lambda_1^{\downarrow}(R_x) & & \\ & \ddots & \\ & & \lambda_d^{\downarrow}(R_x) \end{bmatrix}$$
$$\Leftrightarrow R_x u_i = \lambda_i(R_x) u_i, \quad i = 1 \dots d$$
$$U = [u_1 \dots u_d]$$

remark: If $\operatorname{rank}(X) = d \Leftrightarrow X \subset S$ then $\operatorname{rank} R_x = d$, $\lambda^{\downarrow}_{d+1 \dots D}(R_x) = 0$.

Solution (via optimization):

Thm. all local optimter x^* for Laugranglan must satisfy

$\nabla \mathcal{L}|_{x^*, \lambda^*} = 0$ for some $\lambda^* \in \mathbb{R}^m$ (m constrains.)

$\Rightarrow \nabla f|_{x^*} = -(\lambda^*)^T \nabla g|_{x^*}$, $\nabla g = [\nabla g_1|_{x^*} | \dots | \nabla g_m|_{x^*}]$

the Largrangian of $(*)$:

$$\mathcal{L}(U, \{\lambda_{k,l}\}) = tr(U^T R_x U) + \sum_{k,l} \lambda_{k,l}(u_k^T u_l - \sigma_{k,l})$$

let $\Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1d} \\ \vdots & & \\ \lambda_{1d} & \cdots & \lambda_{dd} \end{bmatrix}$

$$\Rightarrow \mathcal{L}(U, \Lambda) = tr(U^T R_x U) + tr(\Lambda(U^TU - I_d))$$
$$= tr(U^T R_x U) + tr(U\Lambda U^T) - tr(\Lambda), \quad \nabla_U \mathcal{L} = 0$$

Lem. $\nabla_U tr(U^TAU) = \nabla_u \sum_{i=1}^d (U^TAU)_{ii} = \nabla_u \sum_{i=1}^d u_i^T A u_i$ # $u_l = (u_{1l}, u_{2l} \dots u_{dl})$

$$= \frac{\partial}{\partial u_{kl}} (u_l^T A u_l)$$
$$= \left(\frac{\partial u_l}{\partial u_{kl}}\right)^T A u_l + u_l^T A \left(\frac{\partial u_l}{\partial u_{kl}}\right)$$
$$= e_k^T A u_l + u_l^T A e_k$$
$$\overset{A^T=A}{=} 2 \times (k^{th} \text{ entry of } Au_l) \text{ or } 2\times(k,l^{th} \text{ entry of } Au)$$

$\nabla_u \mathcal{L} = R_x U + U\Lambda = 0 \Leftrightarrow R_x U = -U\Lambda$ (need to show that Λ is diagonal)

because $g_{k,l} = g_{l,k}$, we can take Λ symmetric, $\Lambda = V\Lambda' V^T$

$\Rightarrow R_x U = UV\Lambda' V^T \Leftrightarrow R_x(UV) = (UV^T)\Lambda'$

$\Rightarrow UV$ eigenvectors, $\Rightarrow \underset{D\times d}{U} \ \underset{d\times d}{V}$

then, if U is a solution $\Leftrightarrow UV$ is a solution because $UU^T = U\underbrace{VV^T}_{I}U^T$.