B24 July 16 Lec 2 Notes

Proof: (of Gram-schmidt orthogonalization algorithm)

Induct on n (length of $x_1, \dots, x_n$)

Base case: $(n=1)$ This is trivial

I.H: Assume algorithm works for lists up to length n.

I.S: Consider $x_1, \dots, x_n$ L.I.

$$\langle v_{n+1}, v_j \rangle = \left\langle x_{n+1} - \sum_{k=1}^{n} \frac{\langle x_{n+1}, v_k \rangle}{\| v_k \|^2} v_k \, , \, v_j \right\rangle$$

$$= \langle x_{n+1}, v_j \rangle - \sum_{k=1}^{n} \frac{\langle x_{n+1}, v_k \rangle}{\| v_k \|^2} \underbrace{\langle v_k \, , \, v_j \rangle}_{\substack{\text{By I.H.,} \\ \text{this is } = 0 \\ \text{unless } k=j}}$$

$$= \langle x_{n+1}, v_j \rangle - \frac{\langle x_{n+1}, v_j \rangle}{\| v_j \|^2} \langle v_j, v_j \rangle$$

$$= 0$$

Remains to show that:

$$\operatorname{span}\{x_1, \dots, x_{n+1}\} = \operatorname{span}\{v_1, \dots, v_{n+1}\}$$

We know from I.H.:

$$\operatorname{span}\{x_1, \dots, x_n\} = \operatorname{span}\{v_1, \dots, v_n\}$$

Proof: $\operatorname{span}\{x_1, \dots, x_{n+1}\} \subseteq \operatorname{span}\{v_1, \dots, v_{n+1}\}$

$$x_{n+1} = \underbrace{x_{n+1} - \sum_{k=1}^{n} \frac{\langle x_{n+1}, v_k \rangle}{\| v_k \|^2} v_k}_{= v_{n+1}} + \underbrace{\sum_{k=1}^{n} \frac{\langle x_{n+1}, v_k \rangle}{\| v_k \|^2} v_k}_{\in \operatorname{span}(v_1, \dots, v_n)}$$

$$\in \operatorname{span}(v_1, \dots, v_{n+1}) \quad \boxtimes$$

Proof: $\operatorname{span}\{x_1, \dots, x_{n+1}\} \supseteq \operatorname{span}\{v_1, \dots, v_{n+1}\}$

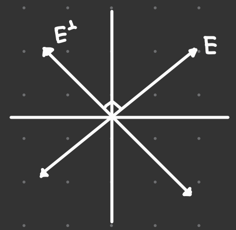
$$v_{n+1} = \underbrace{x_{n+1} - \sum_{k=1}^{n} \frac{\langle x_{n+1}, v_k \rangle}{\| v_k \|^2} v_k}_{\in \operatorname{span}(v_1, \dots, v_n) = \operatorname{span}(x_1, \dots, x_n)}$$

$\boxtimes$

Definition:

If $E \subset V$ is a subspace (V is IPS), we define the orthogonal complement of E by

$$E^{\perp} := \{ x \in V : x \perp E \}$$



Theorem:

If $E \subset V$ is a subspace (V is IPS) , $v \in V$, there exists unique $v_1 \in E$, $v_2 \in E^{\perp}$ s.t.

$$v = v_1 + v_2$$

Proof:

Existence: Let $v_1 = P_E v$, and
$v_2 = v - P_E v$

$v_2 \perp E$ and $v_1 \in E$ so that existence is proven , since $v = v_1 + v_2$

Uniqueness:

If $v_1 \in E$, $v_2 \in E^{\perp}$ are s.t. $v = v_1 + v_2$, then $v_1 = v - v_2 \in E$ and $v - v_1 = v_2 \perp E$

$\Rightarrow v_1 = P_E v$
$\Rightarrow v_2 = v - P_E v$

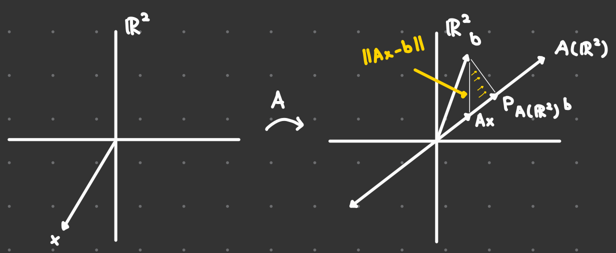
$\boxtimes$

Question:

Given $A : V \to W$ and $b \in W$, how do we minimize $\overbrace{\|Ax - b\|}^{\text{Distance between } Ax \text{ and } b}$ $(x \in V)$?

(for instance if A is a "cost" function and $b = 0$, this is minimizing cost.)

If $Ax = b$ has a solution x_0 , then $\|Ax - b\|$ is minimized at $\|Ax_0 - b\| = 0$.
Otherwise ...



We suspect $Ax = P_{A(\mathbb{R}^2)} b$ will minimize $\|Ax - b\|$, and indeed.

Proposition:

Let V, W be IPS, $b \in W$ and $A: V \to W$ a L.T. Then

$$\inf_{x \in V} \|Ax - b\| = \|P_{range(A)} b - b\|$$

Proof:

Minimizing $\|Ax-b\|$ is equivalent to minimizing $\|Ax-b\|^2$, and

$$\|Ax-b\|^2 = \|\underbrace{Ax - P_{range(A)} b}_{\substack{Ax \in \text{range } A \\ P_{\text{range}A} b \in \text{range } A}} + \underbrace{P_{range(A)} b - b}_{\in (\text{range } A)^\perp}\|^2$$

$$= \|Ax - P_{range(A)} b\|^2 + \|P_{range(A)} b - b\|^2 \quad \text{Pythagorean thm.}$$

$$\geq \|P_{range(A)} b - b\|^2$$

Remark:

Because minimizing $\|Ax-b\|$ is equivalent to minimizing $\|Ax-b\|^2$, and.

$Ax = P_{\text{range}A} b$ is called the "least squares solution".

Question:

How do we find $P_{\text{range}A}\, b$?

Method 1:

Find an orthogonal basis for range(A) (by GS-alg.) and use a formula for $P_{\text{range}A}\, b$.

Definition:

Let A be an $m \times n$ matrix:

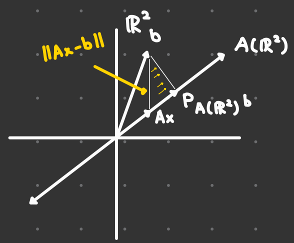
$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

Then the Hermitan adjoint or adjoint of A is defined by the $n \times m$ matrix:

$$A^* := \begin{bmatrix} \overline{a_{11}} & \cdots & \overline{a_{m1}} \\ \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{mn}} \end{bmatrix}$$

Method 2: (A is an m×n matrix)

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$



Since $Ax \in \text{range } A \;\; \forall x \in \mathbb{F}^n$, we have:

$$Ax = P_{\text{range} A}\, b \quad \text{iff} \quad b - Ax \perp \text{range } A$$

Recall the columns of A span range A, so $b - Ax \perp \text{range } A$ iff

$$\left\langle b-Ax, \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \right\rangle = \cdots = \left\langle b-Ax, \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \right\rangle = 0$$

and

$$\left\langle b-Ax, \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix} \right\rangle = \underbrace{\begin{bmatrix} \overline{a_{1k}} & \cdots & \overline{a_{mk}} \end{bmatrix}}_{1 \times m} \cdot \underbrace{(b-Ax)}_{\in \mathbb{F}^m}$$

So

$$\left\langle b-Ax, \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \right\rangle = \cdots = \left\langle b-Ax, \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \right\rangle = 0$$

iff

$$\begin{bmatrix} \overline{a_{1k}} & \cdots & \overline{a_{mk}} \end{bmatrix} \cdot (b-Ax) = 0 \quad \text{for } 1 \le k \le n$$

iff

$$A^*(b-Ax) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

iff

$$\underline{A^*Ax = A^*b}$$

This is called the normal equation, and its solution x satisfies $Ax = P_{\text{range} A}\, b$ i.e. gives the least squares solution

If A^*A is invertible, then

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

So $Ax = A(A^*A)^{-1}A^*b = P_{\text{range} A}\, b$

Note: A is not square

Since this holds for arbitrary $b \in \mathbb{F}^m$, $P_{\text{range}(A)} = A(A^*A)^{-1}A^*$ (if A^*A is invertible)

Theorem:

For an $m \times n$ matrix A,

$$\operatorname{Ker}(A) = \operatorname{Ker}(A^*A)$$

Proof:

If $x \in \operatorname{Ker}(A)$, then $A^*Ax = A^*0 = 0$
i.e. $x \in \operatorname{Ker}(A^*A)$

If $x \in \operatorname{Ker}(A^*A)$, consider the identity

$$\begin{aligned} \|Ax\|^2 &= \langle Ax, Ax \rangle \\ &= \langle A^*Ax, x \rangle, \quad \forall x \in \mathbb{F}^n \end{aligned}$$

So $A^*Ax = 0 \Rightarrow \langle A^*Ax, x \rangle = 0$
$\Rightarrow Ax = 0$
$\Rightarrow x \in \operatorname{Ker}(A)$

$\boxtimes$

Corollary:

For an $m \times n$ matrix A, A^*A is invertible iff $\operatorname{rank}(A) = n$

Proof:

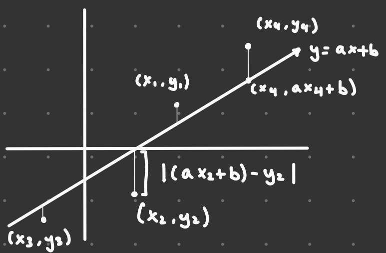
A^*A is invertible iff $\operatorname{Ker}(A^*A) = \{0\}$

iff $\operatorname{Ker}(A) = 0$ By prev. thm.

iff $\operatorname{rank}(A) = n$

$\boxtimes$

Suppose we conjecture that two quantities y, x are related linearly i.e. $y = ax+b$ for some $a,b \in \mathbb{R}$, and we want to find a, b from some data points $(x_1, y_1), \ldots, (x_n, y_n)$



In general, there won't be any $a,b \in \mathbb{R}$ s.t. $y_i = ax_i + b$ for $1 \leq i \leq n$ (for instance due to error in our experiment which produces the data points (x_i, y_i)), but to find the "best" line approximation, we can aim to minimize:

$$\sum_{k=1}^{n} \left| (ax_k + b) - y_k \right|^2 = \left\| \begin{bmatrix} ax_1 + b - y_1 \\ \vdots \\ ax_n + b - y_n \end{bmatrix} \right\|^2$$

$$= \Bigg\| \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{x} - \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{b} \Bigg\|^2$$

Ex 1:

If our data points are $(1,1), (2,3), (3,2)$

We want to minimize

$$= \left\| \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\|^2$$

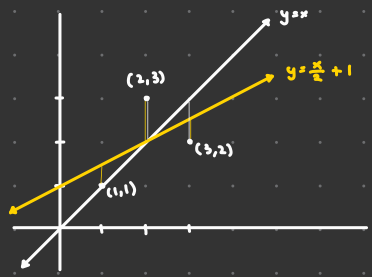
Method 2 tells us to solve $A^*Ax = A^*b$ or

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 14a + 6b \\ 6a + 3b \end{bmatrix} = \begin{bmatrix} 13 \\ 6 \end{bmatrix}$$

$$\Rightarrow a = 1/2, \quad b = 1$$

Ex 1 continued... :



$y=\frac{x}{2}+1$ has error of $(\frac{1}{2})^2 + 1^2 + (\frac{1}{2})^2 = \frac{3}{2}$

$y=x$ has error of $0^2 + 1^2 + 1^2 = 2$