Projection theorem
Normal Equations

ROB 501

Necmiye Ozay

- Last time: Pre-projection theorem & Gram Schmidt (GS) process
  - Will start with some useful results related to GS and orthogonal subspaces
- Projection theorem
- Normal equations

Orthogonal complement

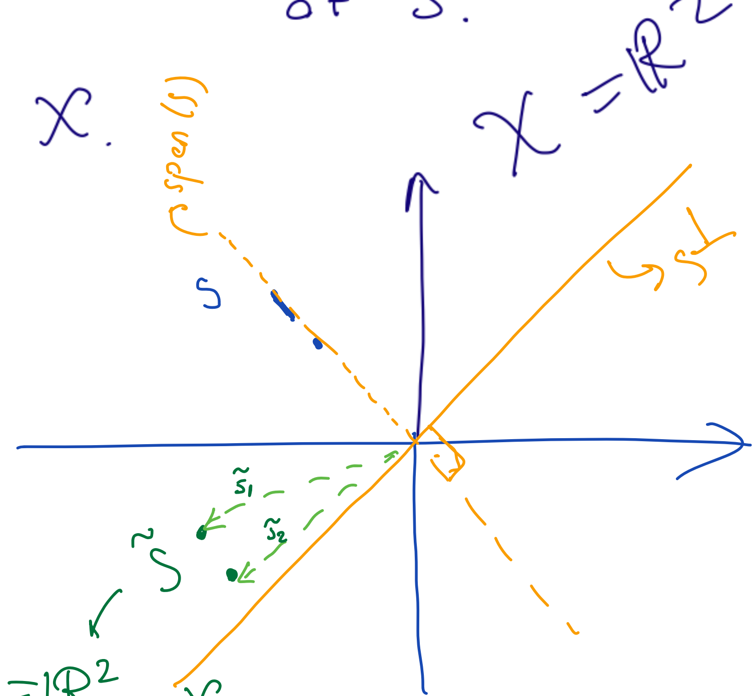
Consider an inner product space $(\mathcal{X}, \mathcal{F}, \langle \cdot, \cdot \rangle)$, where $\mathcal{F} = \mathbb{R}$ or $\mathbb{C}$

__Def:__ Suppose $S \subset \mathcal{X}$ is a subset. Then,

$$S^{\perp} := \{ x \in \mathcal{X} \mid x \perp S \}$$ is the orthagonal complement of S.

__Exercise:__ (1) $S^{\perp}$ is a subspace of $\mathcal{X}$.

(2) $S^{\perp} = (\operatorname{span}(S))^{\perp}$



$\tilde{S}^{\perp} = \{0\}$

$\operatorname{span}(\tilde{S}) = \mathbb{R}^2 = \mathcal{X}$

Proposition: Let $(\mathcal{X}, \mathcal{F} = \mathbb{R} \text{ or } \mathbb{C}, \langle \cdot, \cdot \rangle)$ be an inner product space and finite-dimensional. Let $M \subset \mathcal{X}$ be a subspace. Then,

$$\mathcal{X} = M \oplus M^\perp \qquad \text{"direct sum"}$$

$\forall x \in \mathcal{X}, \exists m_1 \in M, \exists m_2 \in M^\perp$ s.t. $x = m_1 + m_2$

Note: $M \cap M^\perp = \{0\}$

$x \in M$ and $x \in M^\perp \implies \langle x, x \rangle = 0 \iff x = 0$

Proof: Let $k = \dim M$ and $n = \dim(\mathcal{X})$.

- If $k = n$, then $M^\perp = \{0\}$ (since subspace $M = \mathcal{X}$ if $\dim M = \dim \mathcal{X}$), and we are done.

- Assume $1 \leq k < n$ and let $\{y^1, \dots, y^k\}$ be a basis for M.

Complete this basis to be basis $\{y^1, \dots, y^k, y^{k+1}, \dots, y^n\}$ for $\mathcal{X}$. Apply G.S. to produce orthogonal vectors $\{v^1, \dots, v^n\}$ s.t.

$$M = \text{span}\{y^1, \dots, y^k\} = \text{span}\{v^1, \dots, v^k\}$$

and $\{v^1, \dots, v^k\} \perp \{v^{k+1}, \dots, v^n\}$.

$$\mathcal{X} = \underbrace{\text{span}\{v^1, \dots, v^k\}}_{M} \oplus \underbrace{\text{span}\{v^{k+1}, \dots, v^n\}}_{M^\perp} = M \oplus M^\perp$$

i.e. $\forall\, x \in \mathcal{X}$, we have $x = \underbrace{\alpha_1 v^1 + \alpha_2 v^2 + \dots + \alpha_k v^k}_{m_1 \in M} + \underbrace{\alpha_{k+1} v^{k+1} + \dots \alpha_n v^n}_{m_2 \in M^\perp}$

$\square$

Projection theorem

Projection Theorem: Let $(\mathcal{X}, \mathbb{R}, \langle\cdot,\cdot\rangle)$ be a finite-dim inner product space, $M \subset \mathcal{X}$ a subspace and $x \in \mathcal{X}$. Then, $\exists$ a unique $\hat{x} \in M$ s.t.

$$\hat{x} = \underset{m \in M}{\operatorname{argmin}} \|x-m\| \quad \text{and} \quad x - \hat{x} \perp M.$$

(Recall pre-projection theorem: ① p: "$\exists \hat{x} \in M$ s.t. $\hat{x} = \underset{m \in M}{\operatorname{argmin}} \|x-m\|$"

$\Rightarrow \hat{x}$ is unique.

② $p \Leftrightarrow \exists \hat{x} \in M$ s.t. $x - \hat{x} \perp M$.)

Proof: Only thing left to show is existence of $\hat{x}$ s.t. $x - \hat{x} \perp M$ (the rest is shown as part of preprojection theorem)

Since $\mathcal{X}$ is fin. dimensional, this implies
$\mathcal{X} = M \oplus M^\perp$.

$\therefore \exists\, m \in M$ and $\tilde{m} \in M^\perp$ s.t. $x = m + \tilde{m}$

Then $x - m = \tilde{m} \in M^\perp$, hence $\hat{x} = m$.

<u>Normal Equations</u>:

$(\mathcal{X}, \mathbb{R}, \langle \cdot, \cdot \rangle)$, finite dim., $M = \text{span}\{y^1, \ldots, y^k\}$ (subspace)

where $\{y^1, \ldots, y^k\}$ are lin. indep..

Given $x \in \mathcal{X}$, seek an explicit formula for $\hat{x}$.

$$\hat{x} = \underset{m \in M}{\text{argmin}} \|x - m\|$$

We know
$$x - \hat{x} \perp M = \text{span}\{y^1, \ldots, y^k\}$$
$$\Updownarrow$$
$$x - \hat{x} \perp y^i, \quad 1 \leq i \leq k$$
$$\Updownarrow$$
$$\langle x - \hat{x}, y^i \rangle = 0, \quad 1 \leq i \leq k$$
$$\Updownarrow$$
$$\langle x, y^i \rangle - \langle \hat{x}, y^i \rangle = 0, \quad 1 \leq i \leq k$$
$$\Updownarrow$$
$$\langle x, y^i \rangle = \langle \hat{x}, y^i \rangle, \quad 1 \leq i \leq k$$

Recall $\hat{x} \in M \iff \hat{x} = \alpha_1 y^1 + \ldots + \alpha_k y^k$

$$\rightarrow \langle \alpha_1 y^1 + \ldots + \alpha_k y^k, y^i \rangle = \langle x, y^i \rangle \quad 1 \leq i \leq k$$

$$\alpha_1 \langle y^1, y^i \rangle + \dots + \alpha_k \langle y^k, y^i \rangle = \langle x, y^i \rangle \quad 1 \leq i \leq k$$

k equations, one per i. We can write in matrix form:

- $\mathcal{F} = \mathbb{R} \implies \langle y^i, y^j \rangle = \langle y^j, y^i \rangle$

Define

$$\underbrace{\begin{bmatrix} \langle y^1, y^1 \rangle, & \langle y^2, y^1 \rangle, & \dots, & \langle y^k, y^1 \rangle \\ \vdots & & & \\ \langle y^1, y^k \rangle, & \langle y^2, y^k \rangle, & \dots & \langle y^k, y^k \rangle \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}}_{\alpha} = \underbrace{\begin{bmatrix} \langle x, y^1 \rangle \\ \vdots \\ \langle x, y^k \rangle \end{bmatrix}}_{\beta}$$

Note that $[G(y^1, \dots, y^k)]_{ij} := \langle y^i, y^j \rangle$

where $G^T = G$ $\quad \mathcal{F} = \mathbb{R}$

(because $\langle y^i, y^j \rangle = \langle y^j, y^i \rangle$)

$$\boxed{G^T \alpha = \beta}$$ "Normal equations"

↳ solve for α, then get $\hat{x} = \alpha_1 y^1 + \dots + \alpha_k y^k$

Def: The <u>Gram Matrix</u> G and the <u>Gram determinant</u> $g(y^1, \dots, y^k) := \det G$

Proposition: $\{y^1, \ldots, y^k\}$ lin. indep.

$$\iff g(y^1, \ldots, y^k) \neq 0$$

(Gram matrix is invertible)

(in which case, we can uniquely solve for $\alpha = G^{-T}\beta$.)

Proof: From the projection theorem and uniqueness (details will be posted in an handout at the end of this lecture)

Recall Regression: $A\alpha = b$, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, we were solving for

$$\hat{\alpha} := \underset{\alpha \in \mathbb{R}^m}{\operatorname{argmin}} \|A\alpha - b\|_2^2$$

We claimed $\hat{\alpha} = (A^TA)^{-1}A^Tb$. How is this related to "normal equations"

if $\operatorname{rank}(A) = m < n$ (overdetermined)

$$A = [A_1 | A_2 | \cdots | A_m]$$

$$A\alpha = \alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_m A_m$$

Let's formulate using Normal Equations:

$\mathcal{X} = \mathbb{R}^n$, $\mathcal{F} = \mathbb{R}$, $\langle x, y \rangle = x^T y$

$M = \text{span}\{A_1, \dots, A_m\}$

seek $\hat{x} \in M$ s.t. $\|\hat{x} - b\|_2$ is minimized

$$\hat{x} = \underset{m \in M}{\text{argmin}} \|b - m\|_2$$

My y_i's are A_i's.
My x is b.

By normal equations:

Since $\hat{x} \in M$, $\hat{x} = \alpha_1 A_1 + \alpha_2 A_2 + \dots + \alpha_m A_m$

$$G^T \cdot \alpha = \beta$$

$$G_{ij} = G^T_{ij} = \langle A_i, A_j \rangle, \quad \beta_i = \langle b, A_i \rangle$$

$$[A^T A]_{ij} = \left[\begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \end{bmatrix} [A_1 | \dots | A_m] \right]_{ij} = A_i^T A_j = \langle A_i, A_j \rangle$$

$$G = G^T = A^T A$$

$$\beta_i = \langle b, A_i \rangle = b^T A_i \overset{\text{scaler}}{=} (b^T A_i)^T = A_i^T b$$

$$\begin{bmatrix} A_1^T \\ \vdots \\ A_m^T \end{bmatrix} b$$

$$\beta = \begin{bmatrix} A_1^T b \\ A_2^T b \\ \vdots \\ A_m^T b \end{bmatrix} = A^T b$$

$$G^T \alpha = \beta \iff A^T A \alpha = A^T b$$

$$\implies \alpha = (A^T A)^{-1} A^T b$$

$A^T A$ is invertible because $\text{rank}(A) = m$

$$A^T A \alpha = A^T b$$

$$A \alpha = b$$

$$\alpha = A^{-1} b$$

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

$\text{range}(A) \implies$ m dim. subspace of an n-dim space

$\text{range}(A^T) \implies \mathbb{R}^m$

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

$$\mathrm{null}(A) \subset \mathbb{R}^m$$

$$\beta_i = b^T A_i$$

$$\beta = \begin{bmatrix} b^T A_1 \\ \vdots \\ \vdots \\ b^T A_m \end{bmatrix} = \begin{bmatrix} A_1^T b \\ \vdots \\ \vdots \\ A_m^T b \end{bmatrix} = A^T b$$

Prop. $g(y^1, y^2, \cdots, y^k) \neq 0 \Leftrightarrow \{y^1, \cdots, y^k\}$ is linearly independent.

Proof: $g(y^1, y^2, \cdots, y^k) = 0 \leftrightarrow \exists \alpha \neq 0$ such that $G^\top \alpha = 0$.

From our construction of the normal equations, $G^\top \alpha = 0$ if, and only if

$$\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = 0 \quad i = 1, 2, \cdots, k.$$

This is equivalent to

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \perp y^i = 0 \quad i = 1, 2, \cdots, k$$

which is equivalent to

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \perp \text{span}\{y^1, \cdots, y^k\} =: M$$

and thus

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \in M^\perp.$$

Because $\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k \in M$, we have that

$$\left(\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k\right) \in M \cap M^\perp$$

and therefore

$$\alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k = 0.$$

By the linear independence of $\{y^1, \cdots, y^k\}$, we deduce that

$$\alpha_1 = \alpha_2 = \cdots \alpha_k = 0. \ \square$$