

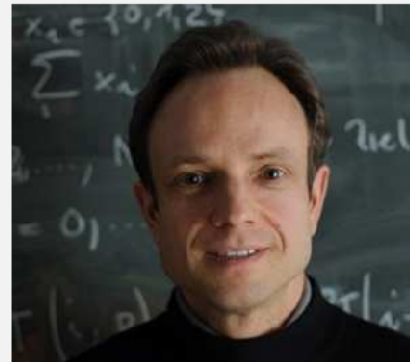
Course Video

Section 3.1.2: Normal Equations

Prof. R. Hiptmair, SAM, ETH Zurich

Date: September 21, 2020

(C) Seminar für Angewandte Mathematik, ETH Zürich



Dependencies. Section 3.1.1

Duration: 30 minutes



Note: Possible minor *mismatch of video and tablet notes!*

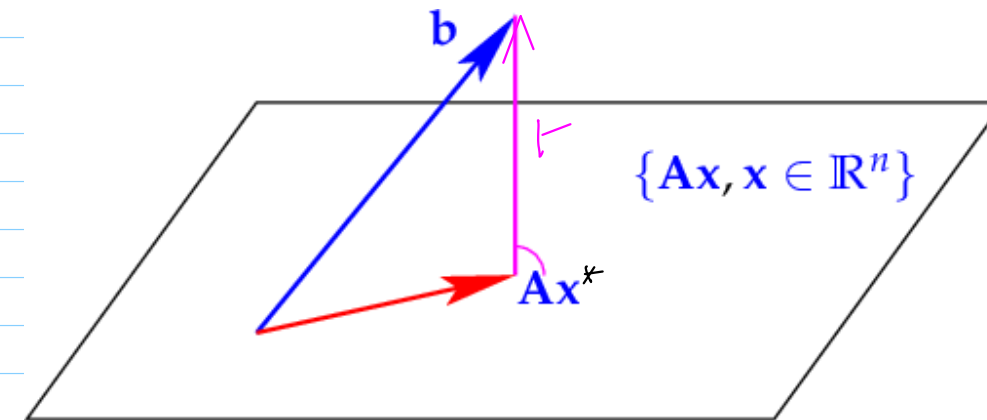
[Corrections and updates can be incorporated into tablet notes only]

III. Direct Methods for Linear Least Squares Problems

3.1. Least Squares Solution Concepts

Setting: $OD\text{-}LSE$ $Ax = b$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, $m \geq n$

3.1.2. Normal Equations



$$x^* \in \text{lsq}(A, b)$$

② residual $r := b - Ax^* \perp \mathcal{R}(A) = \{Ay : y \in \mathbb{R}^n\}$

$$(Ay)^T (b - Ax^*) = 0 \quad \forall y \in \mathbb{R}^n$$

$$\Rightarrow A^T (b - Ax^*) = 0$$

Theorem 3.1.2.1. Obtaining least squares solutions by solving normal equations

The vector $x \in \mathbb{R}^n$ is a least squares solution ($\rightarrow$ Def. 3.1.1.1) of the linear system of equations $Ax = b$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, if and only if it solves the **normal equations (NEQ)**

$$x \in \text{lsq}(A, b) \iff A^T A x = A^T b \quad n \times n \text{ square LSE} \quad (3.1.2.2)$$



NEQ : Size of system matrix

$$\begin{bmatrix} A^T \end{bmatrix} \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} A^T \end{bmatrix} \begin{bmatrix} b \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} A^T A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} A^T \end{bmatrix} \begin{bmatrix} b \end{bmatrix} \in \mathbb{R}^n$$

$\in \mathbb{R}^{n,n}$

Rem 3.1.2.5 : "Analysis-style derivation"

Define : $J : \mathbb{R}^n \rightarrow \mathbb{R}$

$$J(y) := \|b - Ay\|_2^2$$

$$J(y) = y^T A^T A y - 2b^T A y + b^T b$$

$$= \sum_{i=1}^n \sum_{j=1}^n (A^T A)_{ij} y_i y_j - 2 \sum_{i=1}^m \sum_{j=1}^n b_i (A)_{ij} y_j + \sum_{i=1}^m b_i^2, \quad y = [y_1, \dots, y_n]^T \in \mathbb{R}^n.$$

multi-variate polynomial in y : J smooth, C^∞

$$x^* \in \arg\min_{x \in \mathbb{R}^n} J(x) \Rightarrow \underbrace{\text{grad } J(x^*)}_{\mathbb{R}^n \rightarrow \mathbb{R}^n} = 0 \quad \text{(NEQ)}$$

$$\text{grad } J(y) = \left[\frac{\partial J}{\partial y_i}(y) \right]_{i=1}^n = 2A^T A y - 2A^T b$$

Ex 3.1.2.4 : NEQ for linear regression

$$\begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} a \\ \beta \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \Leftrightarrow Ax = b, \quad A \in \mathbb{R}^{m,n+1}, b \in \mathbb{R}^m, x \in \mathbb{R}^{n+1}, \quad (3.0.1.5)$$

Tool: Block matrix multiplication

$$A = [X^T \ 1], \quad X = [x_1, x_2, \dots, x_m] \in \mathbb{R}^{n,n}$$

$$1 = [1, \dots, 1]^T \in \mathbb{R}^n$$

③

NEQ:

$$A^T A = \begin{bmatrix} X X^T & X \cdot \mathbf{1} \\ \mathbf{1}^T X^T & m \end{bmatrix} \begin{bmatrix} \frac{a}{p} \\ \frac{b}{p} \end{bmatrix} = \begin{bmatrix} X Y \\ \mathbf{1}^T Y \end{bmatrix}$$

▷ Note: System matrix of NEQ is symmetric



Explicit NEQ for 1D regression

Issue: Uniqueness of LSQ sols.

Theorem 3.1.2.9. Kernel and range of $A^T A$

For $A \in \mathbb{R}^{m,n}$, $m \geq n$, holds

$$(I) \quad \mathcal{N}(A^T A) = \mathcal{N}(A), \quad (3.1.2.10)$$

$$\mathcal{R}(A^T A) = \mathcal{R}(A^T). \quad (3.1.2.11)$$

Proof of (I): 

$$\begin{aligned} z \in \mathcal{N}(A^T A) &\Rightarrow z^T A^T A z = 0 \Leftrightarrow \|A z\|^2 = 0 \Leftrightarrow A z = 0 \\ &\Leftrightarrow z \in \mathcal{N}(A) \quad \square \end{aligned}$$

Corollary 3.1.2.13. Uniqueness of least squares solutions

If $m \geq n$ and $\mathcal{N}(A) = \{0\}$, then the linear system of equations $Ax = b$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, has a unique least squares solution ($\rightarrow$ 3.1.1.1)

$$x = (A^T A)^{-1} A^T b, \quad (3.1.2.14)$$

that can be obtained by solving the normal equations (3.1.2.2).

$$\mathcal{N}(A^T A) = \{0\} \Rightarrow A^T A \text{ is regular!}$$

$$\text{Note: } \mathcal{N}(A) = \{0\} \Rightarrow A^T A \text{ s.p.d.} \quad \text{II}$$

$$\text{Criterion: } A \in \mathbb{R}^{m,n} : \mathcal{N}(A) = \{0\} \Leftrightarrow \underbrace{\text{rank}(A) = n}_{\text{Full-rank condition (FRC)}}$$

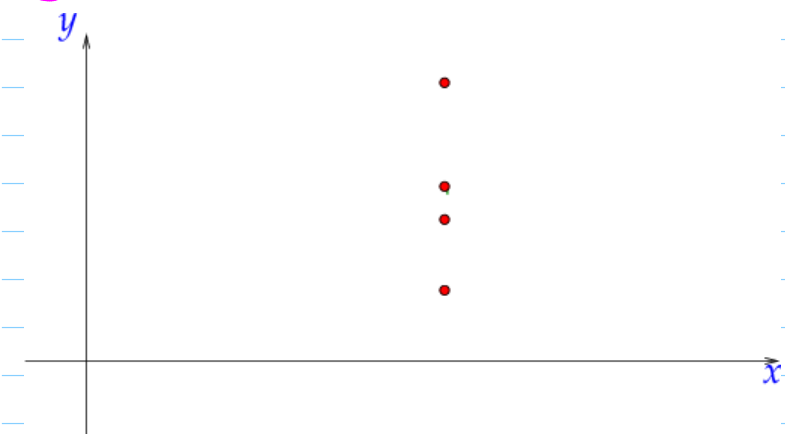
Full-rank condition (FRC)

Ex: 3.1.2.17: FRC for 1D linear regression

$$\text{rank} \begin{pmatrix} \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix} & \begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix} \end{pmatrix} = 2 \Leftrightarrow \exists i, j \in \{1, \dots, m\} : x_i \neq x_j,$$

$$\Leftrightarrow \exists \text{ two linearly indep rows}$$

④



▷ No meaningful regression line

Rem.: Failure of FRC points to

- flawed mathematical model
- useless data

Review questions :

Try to answer Review questions 3.1.2.23 without consulting other parts of the lecture document.

5

7