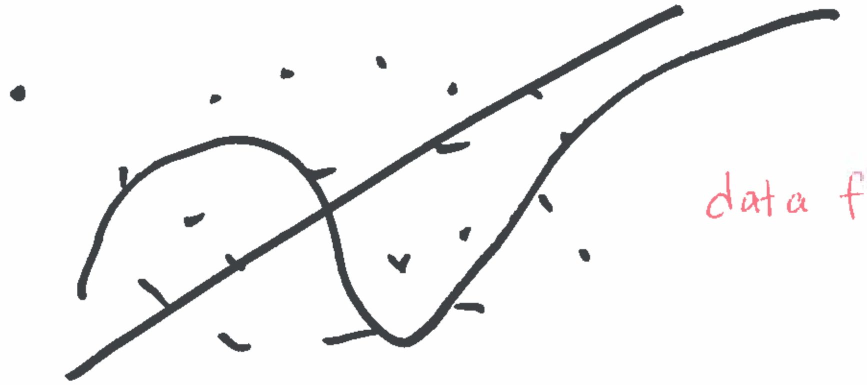
lecture 16 3.18.2019

§3-3 Least squares solution and data fitting

Background:

- $A \in \mathbb{R}^{m\times n}$ ($m > n$)

to solve $A\underline{x} = \underline{b} \in \mathbb{R}^m$ for $\underline{x} \in \mathbb{R}^n$



Review:

matrix-vector multiplication.

$$A = (a_{ij}) = [\underline{a}_1, \underline{a}_2, \cdots \underline{a}_n] \in \mathbb{R}^{m\times n}$$

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

⊛ $A\underline{x} = x_1\underline{a}_1 + x_2\underline{a}_2 \cdots + x_n\underline{a}_n \in \mathbb{R}^m$

$$\text{range}(A) = \{ x_1 \underline{a_1} + x_2 \underline{a_2} + \cdots + x_n \underline{a_n} : \forall\, x_1, x_2 \cdots x_n \in \mathbb{R}^n \}$$

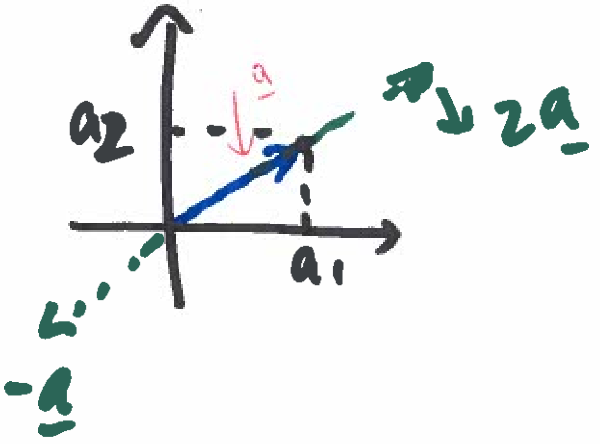
↳ a linear space

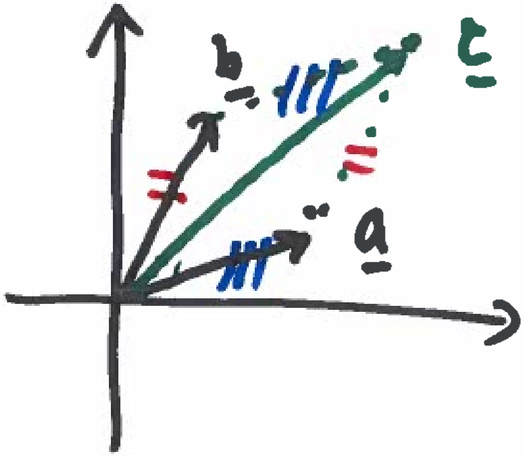
$\underline{a_1}, \underline{a_2} \cdots \underline{a_n}$: linearly independent

Geometric interpretation

Example: $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$

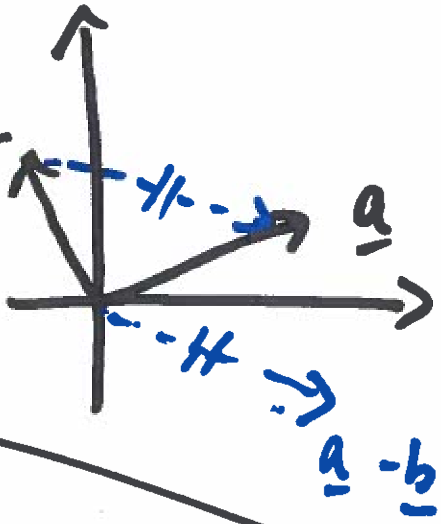
linearly independent





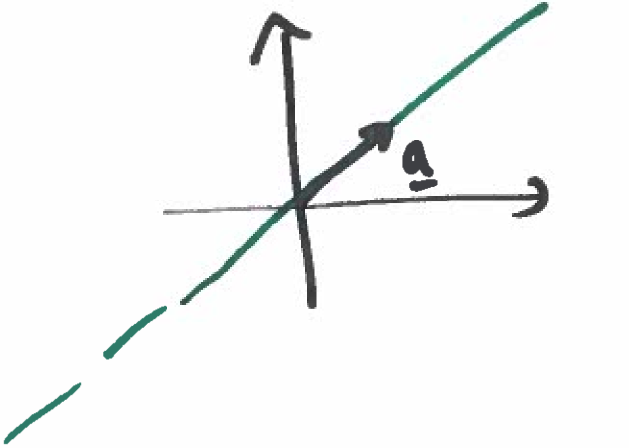
$\underline{a} + \underline{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} = \underline{c}$

$\underline{a} - \underline{b} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \end{pmatrix}$



$A = [\underline{a}] \in \mathbb{R}^{2 \times 1}$

$\text{Range}(A) = \{ x\underline{a} : \forall\, x \in \mathbb{R} \}$

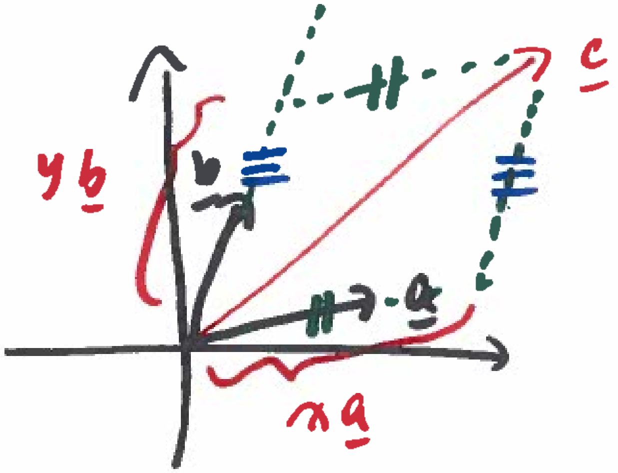


= straight line generated by the vector $\underline{a}$

$A = [\underline{a}, \underline{b}] \in \mathbb{R}^{2\times 2}$

$\text{range}(A) = \{ x\underline{a} + y\underline{b} : \forall x, y \in \mathbb{R} \}$

$= \mathbb{R}^2$ (entire plane)



Given any $\underline{c} \in \mathbb{R}^2$, one can find x, y s.t.
$\underline{c} = x\underline{a} + y\underline{b}$

Example: $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3$

linearly independant

$A = [\underline{a}]$

range(A) : straight line generated by $\underline{a}$.

$A = [\underline{a}, \underline{b}]$

range(A) = $\{ x\underline{a} + y\underline{b} : \forall x, y \in \mathbb{R} \}$

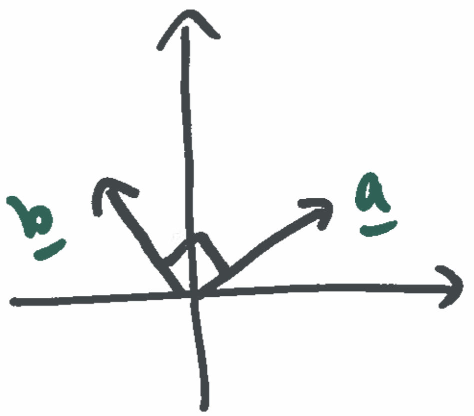
= a plane spaned by $\underline{a}$ and $\underline{b}$.

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

range(A) : a linear space in $\mathbb{R}^m$, generated by column vectors of A

Dot product (inner product),
being orthogonal

In $\mathbb{R}^2$ $\quad \underline{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$



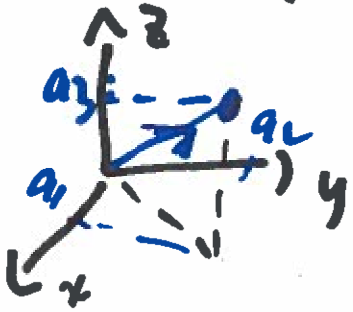
the angle between
$\underline{a}$ and $\underline{b}$ is
right angle, that is

$\underline{a}$ and $\underline{b}$ are orthogonal

$\Leftrightarrow$ $a_1b_1 + a_2b_2 = 0$

we also write $\underline{a} \perp \underline{b}$

In $\mathbb{R}^3$ $\quad \underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$.



$$\underline{a} \perp \underline{b} \Leftrightarrow \underbrace{a_1 b_1 + a_2 b_2 + a_3 b_3}_{\underline{a} \cdot \underline{b}} = 0$$

In general in $\mathbb{R}^n$:

$$\underline{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \underline{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

define dot product (inner product) of $\underline{a}$ and $\underline{b}$

$$\underline{a} \cdot \underline{b} := \underline{a}^T \underline{b} = \sum_{j=1}^{n} a_j b_j$$

$\underline{a}$ is said to be orthogonal to $\underline{b}$, denoted as $\underline{a} \perp \underline{b}$, if $\underline{a} \cdot \underline{b} = \underline{a}^T \underline{b} = 0$

2- norm of $\underline{a} \in \mathbb{R}^n$

$$\|\underline{a}\|_2 = \left(\sum_{j=1}^{n} |a_j|^2 \right)^{\frac{1}{2}} = \sqrt{\underline{a} \cdot \underline{a}}$$

Example: Consider

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 - x_2 = 1 \\ x_1 + x_2 = 3 \end{cases}$$

Solve for x_1, x_2

matrix-vector form

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \underline{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

we want to find $\underline{x} \in \mathbb{R}^2$. s.t.

$$A\underline{x} = \underline{b}.$$

One can see $\underline{x}$ does not exist.

Another way to look at the problem.

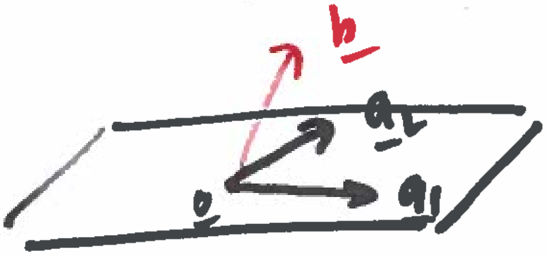
$$A = [\underline{a_1}, a_2]$$

Find x_1, x_2

$$A\underline{x} = \underline{b} \iff x_1 \underline{a_1} + x_2 \underline{a_2} = \underline{b}$$

Here $\underline{x}$ is solvable

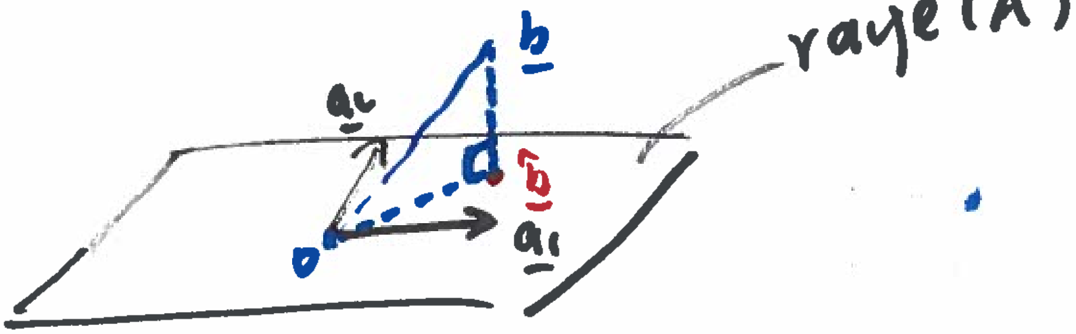
$$\iff \underline{b} \in \text{range}(A).$$



For the specific example

$\underline{b} \notin \text{range}(A)$

$\notin$



- $\hat{\underline{b}}$: projection of $\underline{b}$ into range(A)

  $\hat{\underline{b}} \in$ range(A)

- Find $\hat{\underline{x}}$ s.t. $A\hat{\underline{x}} = \hat{\underline{b}}$.

  $\hat{\underline{x}}$: will be the least squares Solution of $A\underline{x} = \underline{b}$

- $\hat{\underline{b}} \in$ range(A)
- $\underbrace{\underline{b} - \hat{\underline{b}}}$ $\perp Ax \quad \forall x \in \mathbb{R}^2$

($\Leftrightarrow \underline{b} - \hat{\underline{b}} \perp$ range(A))

$\Leftrightarrow \| \underline{b} - \hat{\underline{b}} \|_2 = \min \| \underline{b} - \underline{y} \|_2 \quad \forall \underline{y} \in$ range(A).

↑
not obvious,
can be proved.

Here $\underline{S} \subset \mathbb{R}^n$, $\underline{e} \in \mathbb{R}^n$

$$\underline{e} \perp S \iff \underline{e} \perp \underline{x} \quad \forall \underline{x} \in S$$

---

more general setting

Consider $A \in \mathbb{R}^{m \times n}$ $(m > n)$

$\underline{b} \in \mathbb{R}^m$.

$A\underline{x} = \underline{b}$ may or may not be solvable

Least squares solution: $\hat{\underline{x}} \in \mathbb{R}^n$. satisfying

$$A\hat{\underline{x}} = \hat{\underline{b}}, \quad \text{here}$$

- $\hat{b} \in \text{range}(A)$
- $\underline{b} - \hat{\underline{b}} \perp \text{range}(A)$

$$\iff \|\underline{b} - \hat{\underline{b}}\|_2 = \min_{\underline{y} \in \text{range}(A)} \|\underline{b} - \underline{y}\|_2 .$$

If $\hat{\underline{x}}$ exits, how to find it ?

$$\underline{b} - \hat{\underline{b}} = \underline{b} - A\hat{\underline{x}} \perp \text{range}(A)$$

$$\Leftrightarrow (\underline{b} - A\hat{\underline{x}}) \perp A\underline{x}, \ \forall \underline{x} \in \mathbb{R}^n$$

$$\Leftrightarrow (A\underline{x})^T(\underline{b} - A\hat{\underline{x}}) = 0 \quad \forall \underline{x} \in \mathbb{R}^n$$

$$\Leftrightarrow \underline{x}^T \underbrace{A^T(\underline{b} - A\hat{\underline{x}})}_{\text{called } \underline{w} \in \mathbb{R}^n} = 0 \quad \forall \underline{x} \in \mathbb{R}^n$$

$$\Leftrightarrow \underline{x}^T \underline{w} = 0 \quad \forall \underline{x} \in \mathbb{R}^n$$

take $\underline{x} = \underline{w}$, then

$$\underline{w}^T \underline{w} = \|\underline{w}\|_2^2 = 0$$

$$\Rightarrow \underline{w} \equiv \underline{0}$$

$$\Rightarrow A^T\underline{b} = A^T A \hat{\underline{x}}.$$

or

$$\boxed{\underbrace{A^T A}_{\in \mathbb{R}^{n \times n}.} \hat{\underline{x}} = A^T \underline{b}.}$$

"normal equation"

(square system)

existence and uniqueness of $\hat{x}$

$\Leftrightarrow$ $A^T A$ is invertible

$\overset{\text{lemma}}{\Longleftrightarrow}$ columns of A are linearly independent.

proof:

$\Rightarrow$ Suppose $A^T A$ is invertible.

let $\underline{x} \in \mathbb{R}^n$, satisfying

$$A\underline{x} = \underline{0} \Leftrightarrow (x_1 \underline{a_1} + \cdots + x_n \underline{a_n} = \underline{0})$$

$$\left.\begin{aligned} &\Rightarrow \quad A^T A \underline{x} = A^T \underline{0} = \underline{0}. \\ &\quad\ A^T A \text{ being invertible} \end{aligned}\right\}$$

$\Rightarrow$ $\underline{x} = \underline{0}$.

$\Rightarrow$ columns of A are linearly independent.

$\Leftarrow$ Suppose columns of A are linearly independent, we want to show $A^T A$ is invertible.

By contradiction, Otherwise $\exists\ \underline{y} \in \mathbb{R}^n$, $\underline{y} \neq \underline{0}$

$$A^T A \underline{y} = \underline{0}$$

$$\Rightarrow \utilde{\underline{y}^T} A^T A \underline{y} = 0$$

$$\Leftrightarrow (A\underline{y})^T A\underline{y} = 0$$

$$\Leftrightarrow \| A\underline{y} \|_2^2 = 0$$

$$\Leftrightarrow A\underline{y} \equiv \underline{0}\ .$$

columns of A are linearly independent

$\Rightarrow \underline{y} \equiv \underline{0}$ contradiction

$\Rightarrow A^T A$ is invertible.

Back to normal equation

$$\boxed{A^T A \hat{\underline{x}} = A^T \underline{b}}$$

it is uniquely solvable if columns of A are linearly independent.

$\hat{\underline{x}}$: least square solution of $A\underline{x} = \underline{b}$.

How to understand the "next best", in what sense?

Lemma: $\hat{\underline{b}} \in \text{range}(A)$, satisfying

$\underline{b} - \hat{\underline{b}} \perp \text{range}(A)$

$$\Leftrightarrow \quad \|\underline{b} - \underset{\underset{A\hat{\underline{x}}}{\shortparallel}}{\hat{\underline{b}}}\|_2 = \min_{\underline{y} \in \text{range}(A)} \|\underline{b} - \underline{y}\|_2 \quad ★$$

$$\| \underline{b} - \underset{\underset{A\hat{\underline{x}}}{\|}}{\hat{\underline{b}}} \|_2 = \min_{\underline{y} \in \text{range}(A)} \| \underline{b} - \overset{\nearrow A\underline{x}}{\underline{y}} \|_2$$

$$\Leftrightarrow \quad \| \underline{b} - A\hat{\underline{x}} \|_2 = \min_{\underline{x} \in \mathbb{R}^n.} \| b - A\underline{x} \|_2$$

residual: $\underline{b} - A\underline{x}$ is minimized with respect to 2-norm. leading to the least squares solution $\hat{\underline{x}}$.

(or : the least square solution $\hat{\underline{x}} \in \mathbb{R}^n$ minimizes the residual $\underline{b} - A\underline{x}$ in 2-norm)

Example: Revisit the starting example

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \underline{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

find least squares (LS) solution of $A\underline{x} = \underline{b}$ for $\hat{\underline{x}} \in \mathbb{R}$

Solution:

LS solution $\Leftrightarrow$ the normal equations

$$A^T A \hat{\underline{x}} = A^T \underline{b}$$

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

$$= \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$ SPD (strictly diagonally dominant) invertible

$$A^T \underline{b} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

Solve $\hat{\underline{x}}$ from $\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$

$$\Rightarrow \hat{\underline{x}} = \begin{bmatrix} 7/4 \\ 3/4 \end{bmatrix}$$

One can check residual

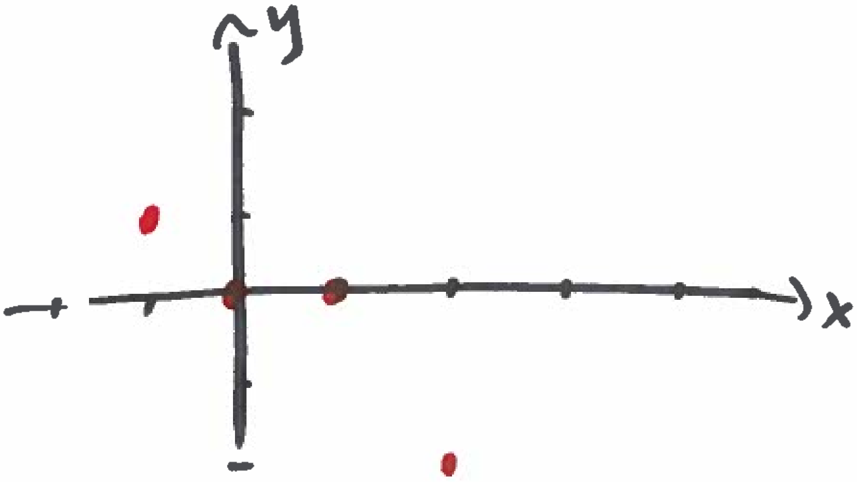
$$\underline{r} = \underline{b} - A\underline{\hat{x}}$$

$$= \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - A\begin{bmatrix} 7/4 \\ 3/4 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \\ 0.5 \end{bmatrix}$$

$$\neq \underline{0}.$$

Application.: data fitting

Example given 4 data points



j	(x_j, y_j)
1	(−1 1)
2	(0 0)
3	(1 0)
4	(2 −2)

- Find the "best" linear polynomial to fit the data.
- Find the best parabola to fit the data.

The best is in the 2-norm hence least squares sense.

Solution:

look for $f(x) = a_1 + a_2 x$

we hope to find a_1, a_2 s.t.

$$f(x_i) = y_i \quad i = 1, 2, 3, 4.$$

over-determined :

matrix-vector form . of $\left\{ \begin{array}{l} f(x_i) = y_i \\ i = 1, 2, 3\ 4 \end{array} \right.$

$$\Rightarrow \quad A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \underline{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -2 \end{pmatrix}$$

$$\underline{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Try to solve $\underline{a}$ from

$$\boxed{A\,\underline{a} = \underline{b}}$$

we look for the LS solution.

normal eqn: $\underbrace{A^T A}\,\underline{\hat{a}} = A^T \underline{b}$

$$\begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix} \underline{\hat{a}} = \begin{pmatrix} -1 \\ -5 \end{pmatrix}$$

$$\underline{\hat{a}} = \begin{bmatrix} 0.2 \\ -0.9 \end{bmatrix} \Rightarrow \boxed{f(x) = 0.2 - 0.9x}$$

Another way to see this solution

$$\|\underline{b} - A\hat{\underline{a}}\|_2 = \min_{\underline{x}\in\mathbb{R}^2} \|\overbrace{\underbrace{\underline{b} - A\underline{x}}_{\underline{r}}}\|_2$$

Or ~~alternatively~~ equivalently

find a_1, a_2 s.t.

$$\underline{r} = \begin{pmatrix} y_1 - f(x_1) \\ y_2 - f(x_2) \\ y_3 - f(x_3) \\ y_4 - f(x_4) \end{pmatrix}$$

is minimized. in 2-norm.

$f(x) = a_1 + a_2 x$