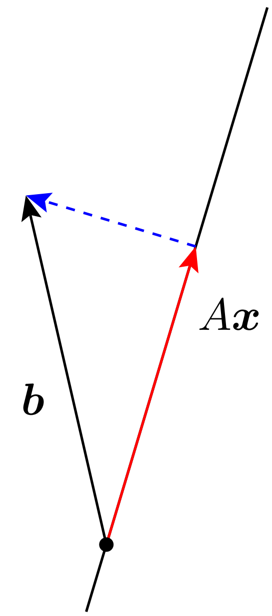
Least squares

Definition 1. $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$ if $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible.

- If $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then

- Interesting case: $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent.

(in other words: the system is overdetermined)



Idea. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent $\Longleftrightarrow$

So, if $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, we

- replace $\boldsymbol{b}$ with
- solve $A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$. (consistent by construction!)

Example 2. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Solution.

Theorem 3. $\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$
$\Longleftrightarrow$ $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Proof.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$

$\Longleftrightarrow$ $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible

$\vdots$

$\vdots$

Armin Straub
astraub@illinois.edu

$\vdots$

$\iff A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ □

Example 4. (again) Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

Solution.

Example 5. Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

What is the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(A)$?

Solution.

Application: least squares lines

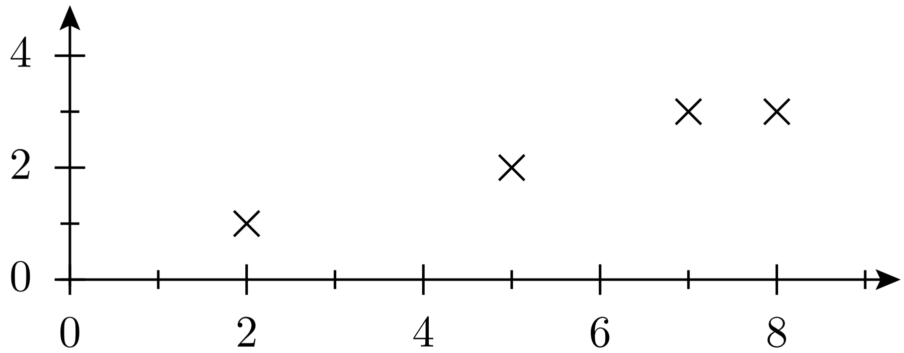
Experimental data: (x_i, y_i)

Wanted: parameters β_1, β_2 such that $y_i \approx \beta_1 + \beta_2 x_i$ for all i

This approximation should be so that

$$\sum_i [y_i - (\beta_1 + \beta_2 x_i)]^2 \quad \text{is}$$

Example 6. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2, 1)$, $(5, 2), (7, 3), (8, 3)$.



Solution. The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Hence, the least squares line is

Armin Straub
astraub@illinois.edu

How well does the line fit the data $(2,1),(5,2),(7,3),(8,3)$?

- The error at a point (x_i, y_i) is $\varepsilon_i = y_i - (\beta_1 + \beta_2 x)$.

  Here:

- The **residual sum of squares** is $\sum \varepsilon_i$.

  Here:

Other curves

We can also fit the experimental data (x_i, y_i) using other curves.

Example 7. $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$.

Multiple regression

The experimental data might be of the form (v_i, w_i, y_i), where now y_i depends on two variables v_i, w_i (instead of just one x_i).

Fitting a linear relationship $y_i \approx \beta_1 + \beta_2 v_i + \beta_3 w_i$, we get:

Armin Straub
astraub@illinois.edu