

The Method of Least Squares

Setup: you have a matrix equation $Ax=b$ which is (generally) **inconsistent**. What is the **best approximate solution**?

What do we mean by "best approximate solution"?

Def: \hat{x} is a **least squares solution** of $Ax=b$ if $\|b-A\hat{x}\|$ is minimized over all vectors \hat{x} .

This means $A\hat{x}$ is as **close as possible** to b .

NB: $\text{Col}(A) = \{A\hat{x} : \hat{x} \in \mathbb{R}^n\}$, so $A\hat{x}$ is just the **closest vector** to b in $V = \text{Col}(A)$.

This is the **orthogonal projection!**

$$A\hat{x} = b_V \quad \text{for } V = \text{Col}(A)$$

A least-squares solution of $Ax=b$ is just a solution of $A\hat{x} = b_V$ $V = \text{Col}(A)$
(which is consistent)

How do we compute \hat{x} ?

→ Maybe we first compute b_V ?

We do this by solving the normal equation

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

then $A \hat{x} = b_V$ for any solution \hat{x} .

But this just says the solutions of $A \hat{x} = b_V$ are the solutions of $A^T A \hat{x} = A^T b$!

Procedure (Least Squares):

To find the least squares solution(s) of $Ax = b$:

(1) Solve the normal equation $A^T A \hat{x} = A^T b$

Any solution \hat{x} is a least-squares soln,

and $b_V = A \hat{x}$ ($V = \text{Col}(A)$).

Def: The error is the distance from $A \hat{x}$ to b :

$$\text{error} = \|b - A \hat{x}\| = \|b - b_V\| = \|b_{V^\perp}\|$$

minimizing $\|b - A \hat{x}\|$ is the same as
minimizing $\|b - A \hat{x}\|^2$

So we have minimized $\|b - A\hat{x}\|^2 = \|b_{\perp}\|^2$.

The least-squares solution(s) minimize $\|b_{\perp}\|^2$

So if $b_{\perp} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ then we minimized

$$\|b_{\perp}\|^2 = a^2 + b^2 + c^2.$$

This is why it's called a **least squares** solution: we're minimizing the sum of the squares of the entries of $b_{\perp} = b - A\hat{x}$.

Eg: Find the least-squares solution of $Ax = b$

for $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$ $b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$.

$$A^T A = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix} \quad A^T b = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 5 & 3 & 0 \\ 3 & 3 & 6 \end{array} \right) \xrightarrow{\text{ref}} \left(\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 5 \end{array} \right)$$

$$\Rightarrow \hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix} \leftarrow \text{the least-squares soln}$$

$$b_{\perp} = A\hat{x} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

↳ for $V = \text{Col}(A)$

The error is

$$\|b_v\| = \|b - b_v\| = \left\| \begin{pmatrix} 6 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\| \\ = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$

[demo: where is \hat{x} ?]

Eg: Find the least-squares solutions of $Ax=b$ for $A = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

$$A^T A = \begin{pmatrix} 6 & 6 & 6 \\ 6 & 3 & 6 \\ 6 & 6 & 18 \end{pmatrix} \quad A^T b = \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 6 & 6 & 6 & 4 \\ 6 & 3 & 6 & -1 \\ 6 & 6 & 18 & 2 \end{array} \right) \xrightarrow{\text{ref}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 2/3 \\ 0 & 1 & 2 & -1/3 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

$$\xrightarrow{\text{PVF}} \hat{x} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

In this case there are **infinitely many** least-squares solutions!

$$b_v = A\hat{x} \text{ for any } \hat{x} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

$$\text{Take } \hat{x} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix}$$

$$\hookrightarrow b_v = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$b_{v^\perp} = b - b_v = 0$$

So the error is zero — the equation $Ax=b$ was **consistent** after all!

(Compare L11)

Observation 1:

$Ax=b$ has a **unique** least-squares soln

$\iff A$ has **full column rank!**

\rightarrow This is exactly when $A\hat{x}=b_v$ has a unique solution.

Otherwise, there are **infinitely many** least-squares solns. This means $\|b - Ax\|$ is minimized for **any** such \hat{x} :

$$b_v = Ax \text{ for any solution } \hat{x}.$$

(There can't be **zero** least-squares solutions!

$A\hat{x}=b_v$ is **always consistent**.)

Observation 2: If $Ax=b$ is consistent, then
(least squares solutions) = (ordinary solutions).

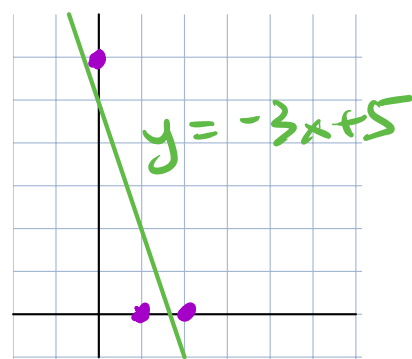
Indeed, a least-squares soln is just a soln of $A\hat{x} = b_V$ ($V = \text{Col}(A)$), and $b = b_V \iff b \in \text{Col}(A) \iff Ax=b$ is consistent.

Least-squares is often useful for fitting data to a model.

Eg (linear regression):

Find the best-fit line $y = Cx + D$ thru the data points $(0,6)$, $(1,0)$, $(2,0)$.

If $(0,6)$ lies on $y = Cx + D$ then substituting $x=0$, $y=6$ would give $6 = C \cdot 0 + D$. So we want to solve:



$$(0,6): 6 = C \cdot 0 + D$$

$$(1,0): 0 = C \cdot 1 + D$$

$$(2,0): 0 = C \cdot 2 + D$$

in the
unknowns $C \Delta D$

ie $Ax = b$ for $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$ $x = \begin{pmatrix} C \\ D \end{pmatrix}$ $b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$

NB: the data points are not collinear \rightarrow
no exact solution! (maybe measurement error).

We found a least-squares solution before:

$$\hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix} \rightarrow \text{best-fit line } y = -3x + 5$$

[demo]

Important Question:

What quantity did we minimize?

We minimized $\|b - A\hat{x}\|^2 = \|b_{\text{r+}}\|^2$

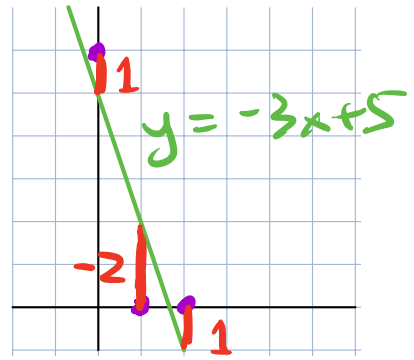
$b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$ = y-values of data pts.
(the y-values we wanted)

$$A\hat{x} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -3 \cdot 0 + 5 \\ -3 \cdot 1 + 5 \\ -3 \cdot 2 + 5 \end{pmatrix} = \begin{matrix} \text{y-values of} \\ y = -3x + 5 \\ \text{at x-values } 0, 1, 2. \end{matrix}$$

(the y-values we got)

$$\text{So } b_{\text{res}} = b - A\hat{x} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

= (vertical distances
from $y = -3x + 5$ to
the data points)



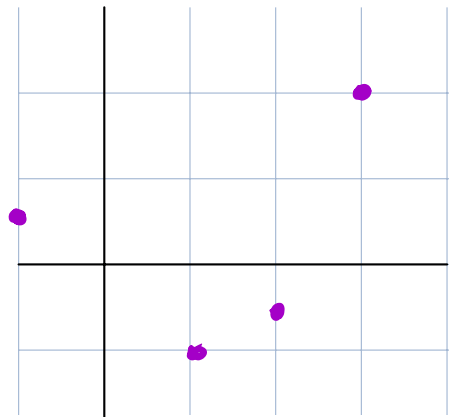
We minimized the sum of the squares of the vertical distances (the error).

Eg (best-fit parabola):

Find the best-fit parabola $y = Bx^2 + Cx + D$

thru the data points $(-1, 1/2), (1, -1), (2, -1/2), (3, 2)$

Substitute the data points for
 x & y you want to solve



$$(-1, 1/2): \quad \frac{1}{2} = B(-1)^2 + C(-1) + D$$

$$(1, -1): \quad -1 = B(1)^2 + C(1) + D$$

$$(2, -1/2): \quad -\frac{1}{2} = B(2)^2 + C(2) + D$$

$$(3, 2): \quad 2 = B(3)^2 + C(3) + D$$

$$\rightarrow Ax = b \quad \text{for } A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \end{pmatrix} \quad b = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix}$$

Let's find the least-squares solution.

$$A^T A = \begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix} \quad A^T b = \begin{pmatrix} 31/2 \\ 7/2 \\ 1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 99 & 35 & 15 & 31/2 \\ 35 & 15 & 5 & 7/2 \\ 15 & 5 & 4 & 1 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 53/88 \\ 0 & 1 & 0 & -379/440 \\ 0 & 0 & 1 & -41/44 \end{array} \right)$$

$$\hat{x} = \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} \rightsquigarrow y = \frac{53}{88}x^2 - \frac{379}{440}x - \frac{41}{44}$$

[demo]

Question: What did we minimize? always $\|b - A\hat{x}\|^2$

$$b = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix} = \text{y-values of data pts.}$$

$$A\hat{x} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} = \begin{pmatrix} 53/88(-1)^2 + \frac{379}{440}(-1) - 41/44 \\ 53/88(1)^2 + \frac{379}{440}(1) - 41/44 \\ 53/88(2)^2 + \frac{379}{440}(2) - 41/44 \\ 53/88(3)^2 + \frac{379}{440}(3) - 41/44 \end{pmatrix}$$

$$= \text{y-values of } y = \frac{53}{88}x^2 - \frac{379}{440}x - \frac{41}{44} \text{ at x-values } -1, 1, 2, 3$$

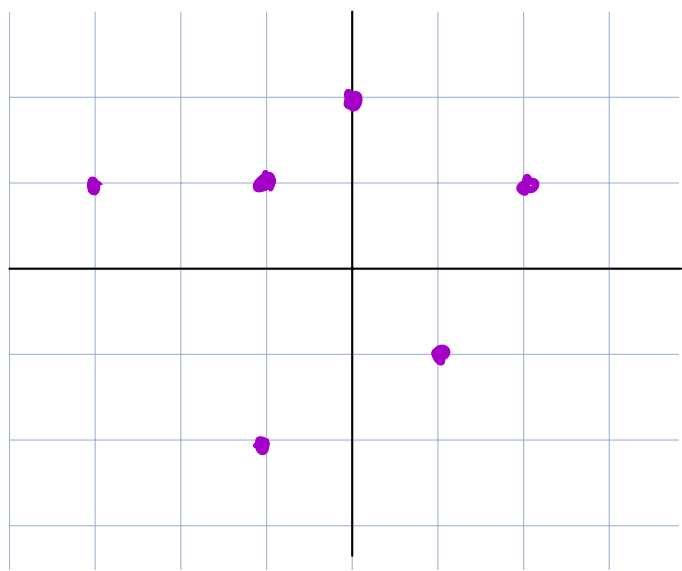
So $b_{\perp} = b - A\hat{x} = \text{vertical distances from the graph to the data points, like before.}$

This same method works to find a best-fit function of the form $y = Af + Bg + Ch + \dots$ where f, g, h, \dots are really any functions! Just plug the x -values of your data points into $f, g, h \dots \rightarrow$ linear equations for A, B, C, \dots

Eg (best-fit trigonometric function):
[demo]

This "real-life" example of Gauss was in the first lecture:

Eg: An asteroid has been observed at coordinates:
 $(0, 2), (2, 1), (1, -1), (-1, -2), (-3, 1), (-1, 1)$



Question: What is the most likely orbit?
Will the asteroid crash into the Earth?

Fact: The orbit is an ellipse.

Equation for an ellipse:

$$x^2 + By^2 + Cxy + Dx + Ey + F = 0$$

For our points to lie on the ellipse, substitute the coordinates into (x,y) \rightsquigarrow these should hold:

$$\begin{array}{l}
 \begin{array}{c} x \\ = \\ (0, 2) \end{array}: \quad 0 + 4B + 0 + 0 + 2E + F = 0 \\
 \begin{array}{c} y \\ = \\ (2, 1) \end{array}: \quad 4 + B + 2C + 2D + E + F = 0 \\
 (1, -1): \quad 1 + B - C + D - E + F = 0 \\
 (-1, -2): \quad 1 + 4B + 2C - D - 2E + F = 0 \\
 (-3, 1): \quad 9 + B - 3C - 3D + E + F = 0 \\
 (-1, 1): \quad 1 + B - C - D + E + F = 0
 \end{array}$$

\curvearrowright constants

Move the constants to the RHS \rightsquigarrow this is $Ax=b$

$$\text{for } A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \\ E \\ F \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \\ -1 \end{pmatrix}$$

Least-squares solution: [demo]

$$\hat{x} = \left(\frac{405}{266}, -\frac{89}{133}, \frac{201}{133}, -\frac{123}{266}, -\frac{687}{133} \right)$$

$$\rightsquigarrow x^2 + \frac{405}{266} y^2 - \frac{89}{133} xy + \frac{201}{133} x - \frac{123}{266} y - \frac{687}{133} = 0$$

What quantity did we minimize? $\|b - A\hat{x}\|^2$ or $\|A\hat{x} - b\|^2$

$$A\hat{x} - b = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & 1 & -3 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{pmatrix} \hat{x} - \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} 0^2 + \frac{405}{266}(2)^2 - \frac{89}{133}(2)(2) + \frac{201}{133}(0) - \frac{123}{266}(2) - \frac{687}{133} \\ 2^2 + \frac{405}{266}(1)^2 - \frac{89}{133}(2)(1) + \frac{201}{133}(2) - \frac{123}{266}(1) - \frac{687}{133} \\ 1^2 + \frac{405}{266}(-1)^2 - \frac{89}{133}(1)(-1) + \frac{201}{133}(1) - \frac{123}{266}(-1) - \frac{687}{133} \\ (-1)^2 + \frac{405}{266}(-2)^2 - \frac{89}{133}(-1)(-2) + \frac{201}{133}(-1) - \frac{123}{266}(-2) - \frac{687}{133} \\ (-3)^2 + \frac{405}{266}(1)^2 - \frac{89}{133}(-3)(1) + \frac{201}{133}(-3) - \frac{123}{266}(1) - \frac{687}{133} \\ (-1)^2 + \frac{405}{266}(1)^2 - \frac{89}{133}(-1)(1) + \frac{201}{133}(-1) - \frac{123}{266}(1) - \frac{687}{133} \end{pmatrix}$$

This was what you get by substituting the x - and y -values of the data points into the LHS of

$$x^2 + \frac{405}{266}y^2 - \frac{89}{133}xy + \frac{201}{133}x - \frac{123}{266}y - \frac{687}{133} = 0$$

It's the distance from zero. [demo]

Upside: You're minimizing $\|b - A\hat{x}\|$; it's up to you to interpret that quantity in your original problem.