• Least Squares: Theory & Examples.

Announcements:

• Extended Section M/W 10-12

• P/NP / Grade option. Select by 5/6.

Least Squares: Goal is to approximately solve a system of Eqns

$$A\vec{x} \approx \vec{b}$$

known

= Unknown (aka variable).

Least Squares optimization problem:

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2$$

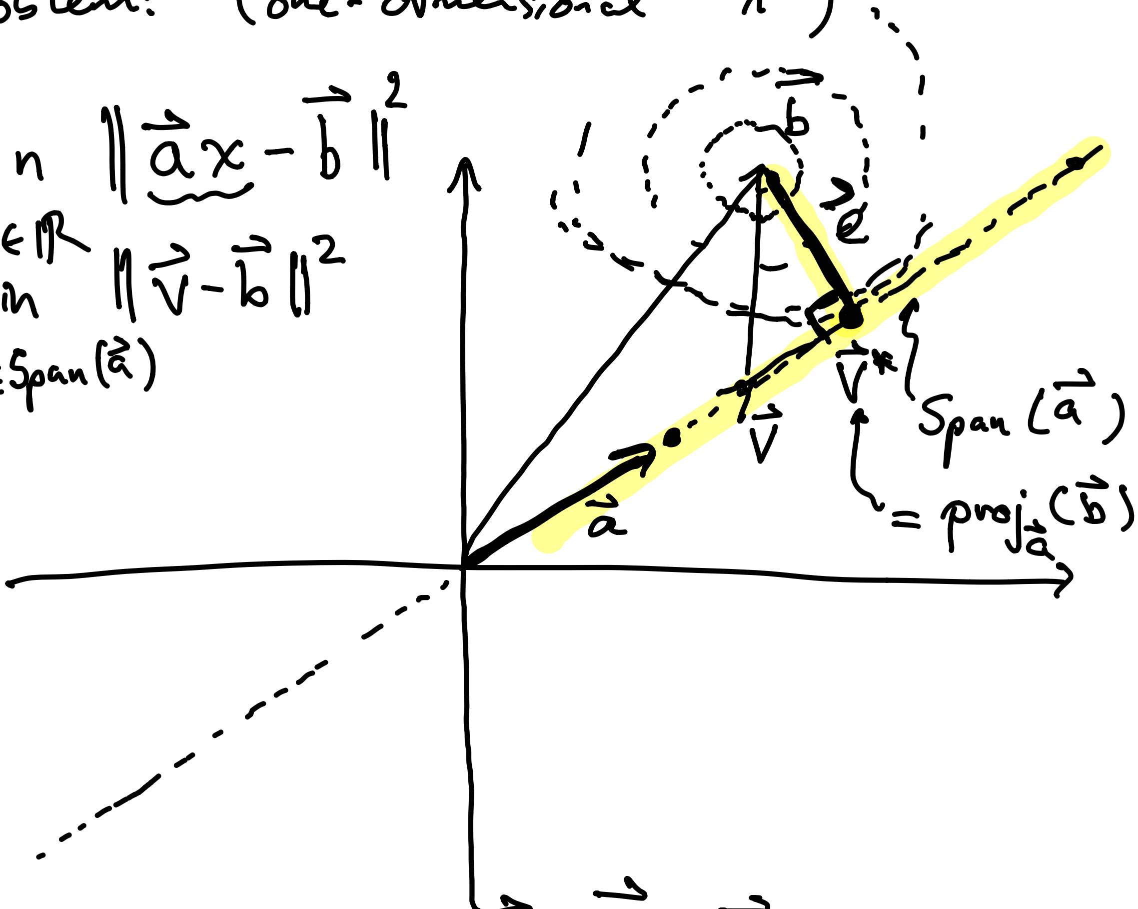
Assumption: We are working w/ Euclidean norm & Inner product.

Lots of Geometric Intuition!

Last time: started with a simpler problem: (one-dimensional $\vec{x}$)

$$\min_{x \in \mathbb{R}} \|\vec{a}x - \vec{b}\|^2$$

$$\equiv \min_{\vec{v} \in \mathrm{Span}(\vec{a})} \|\vec{v} - \vec{b}\|^2$$



Key Idea: error $\vec{e} = \vec{b} - \vec{v}^*$ is orthogonal to Span$(\vec{a})$.

Let's check that if $\vec{e}$ is orth. to Span$(\vec{a})$, then $\vec{v}^*$ must be closest point.

For any other $\vec{v} \in \text{Span}(\vec{a})$ $\vec{v} \in R(A)$

$$\begin{aligned}
\|\vec{v} - \vec{b}\|^2 &= \|(\vec{v} - \vec{v}^*) + (\vec{v}^* - \vec{b})\|^2 \\
&= \|\vec{v} - \vec{v}^*\|^2 + \|\vec{v}^* - \vec{b}\|^2 \\
&\quad + \underbrace{2\langle \underbrace{\vec{v} - \vec{v}^*}_{\in \text{Span}(\vec{a})}, \underbrace{\vec{v}^* - \vec{b}}_{\vec{e}} \rangle}_{=0} \\
&= \|\vec{v} - \vec{v}^*\|^2 + \|\vec{v}^* - \vec{b}\|^2 \\
&\geq \|\vec{v}^* - \vec{b}\|^2 \qquad (*)
\end{aligned}$$

Note: we used:

$$\begin{aligned}
\|\vec{x} + \vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y} \rangle \\
&= \langle \vec{x}, \vec{x} + \vec{y} \rangle + \langle \vec{y}, \vec{x} + \vec{y} \rangle \\
&= \langle \vec{x}, \vec{x} \rangle + \langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{x} \rangle + \langle \vec{y}, \vec{y} \rangle \\
&= \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\langle \vec{x}, \vec{y} \rangle.
\end{aligned}$$

Observe similar to: $(x+y)^2 = x^2 + y^2 + 2xy$.

Conclusion from (*) is that if $\vec{e} = \vec{v}^* - \vec{b}$ is orthogonal to $\text{Span}(\vec{a})$, then $\vec{v}^*$ is solution to

$$\min_{\vec{v} \in \text{Span}(\vec{a})} \|\vec{v} - \vec{b}\|^2.$$

Claim: optimal vector

$$\vec{v}^* = \underbrace{\frac{\langle \vec{b}, \vec{a} \rangle}{\langle \vec{a}, \vec{a} \rangle} \vec{a}}_{\text{proj}_{\vec{a}}(\vec{b})}$$

Proof: Just check for desired orthogonality.

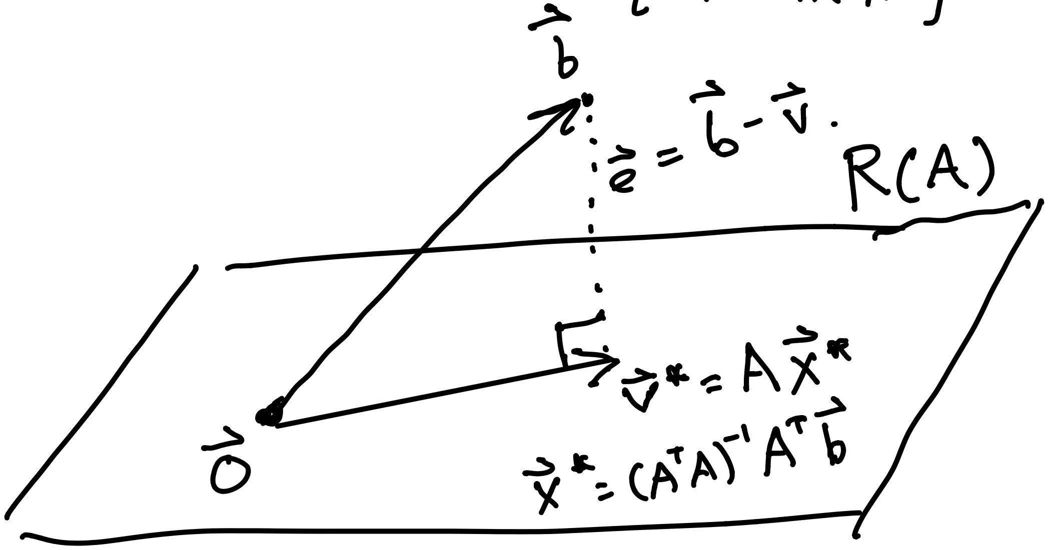
$$\langle \vec{e}, \underbrace{\beta \vec{a}}_{\substack{\text{a vector in} \\ \text{Span}(\vec{a})}} \rangle = \beta \langle \vec{e}, \vec{a} \rangle = \beta \langle \vec{b} - \text{proj}_{\vec{a}}(\vec{b}), \vec{a} \rangle = 0.$$

did last time.

More generally:

$$\min_{\vec{x}} \|A\vec{x}-\vec{b}\|^2 \equiv \min_{\vec{v}\in \underset{\sim}{R}(A)} \|\vec{v}-\vec{b}\|^2$$

$= \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$



Note: If $\vec{e}$ is orthogonal to $R(A)$, then $\vec{v}^*$ is optimal for LS problem

$$\min_{\vec{v}\in \text{Range}(A)} \|\vec{v}-\vec{b}\|^2$$

Logic is same as before: For any vector $\vec{v} \in R(A)$, $\|\vec{v}-\vec{b}\|^2 \geq \|\vec{v}^*-\vec{b}\|^2$.

Thm: If A has zero nullspace, then the LS solution to

$$\min_{\vec{x} \in \mathbb{R}^n} \|\underbrace{A\vec{x}}\, - \vec{b}\|^2$$

is equal to $\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$.

Proof: Just need to check that $\vec{e} = \vec{v}^* - \vec{b}$ is orthogonal to $R(A)$,

where $\vec{v}^* = A\vec{x}^* = A(A^TA)^{-1}A^T\vec{b}$.

First claim that $N(A) = \{\vec{0}\} \Rightarrow (A^TA)$ is invertible. (we'll check this later). $\hookrightarrow$ square matrix.

Fix any $\vec{v} \in R(A)$.

$$\langle \vec{e}, \vec{v} \rangle = \langle \vec{v}^* - \vec{b}, \vec{v} \rangle$$

$$= \langle \underbrace{A(A^TA)^{-1}A^T\vec{b}}_{\vec{v}^*} - \vec{b}, \vec{v} \rangle$$

If $\vec{v} \in R(A) \Rightarrow \vec{v} = A\vec{x}$ for some $\vec{x}$.

$$= \left\langle A(A^TA)^{-1}A^T\vec{b},\ A\vec{x}\right\rangle - \left\langle \vec{b}, A\vec{x}\right\rangle$$

recall
$\langle \vec{a}, \vec{b}\rangle$
$= \vec{a}^T\vec{b}$
$= \vec{b}^T\vec{a}$.

$$= (A\vec{x})^T A(A^TA)^{-1}A^T\vec{b} - (A\vec{x})^T\vec{b}$$

$$= \vec{x}^T \overset{I}{\cancel{(A^TA)(A^TA)^{-1}}} A^T\vec{b} - \vec{x}^T A^T \vec{b}$$

$$= \vec{x}^T A^T \vec{b} - \vec{x}^T A^T\vec{b} = 0 \checkmark$$

Only thing left to do:

Technical Lemma: $\mathcal{N}(A) = \mathcal{N}(A^TA)$, i.e., A^TA is invertible.

Pf: First show $\mathcal{N}(A) \subseteq \mathcal{N}(A^TA)$.

Take $\vec{x} \in \mathcal{N}(A) \implies (A^TA)\vec{x} = A^T(\underbrace{A\vec{x}}_{\vec{0}})$

$$= \vec{0} \implies \vec{x} \in \mathcal{N}(A^TA)$$

Second show: $\mathcal{N}(A^TA) \subseteq \mathcal{N}(A)$

Take $\vec{x} \in \mathcal{N}(A^TA) \Rightarrow A^TA\vec{x} = \vec{0}$

Clever idea:

$$\|A\vec{x}\|^2 = \underbrace{\langle A\vec{x}, A\vec{x}\rangle}_{\vec{x}^T(A^TA)\vec{x}} = \langle \vec{x}, \overbrace{A^TA\vec{x}}^{\vec{0}} \rangle = 0.$$

$$\Rightarrow A\vec{x} = \vec{0} \Rightarrow \vec{x} \in \mathcal{N}(A)$$

$$\begin{aligned}\langle A\vec{x}, A\vec{x}\rangle &= (A\vec{x})^T A\vec{x} \\ &= \vec{x}^T(A^TA)\vec{x} \\ &= \langle \vec{x}, A^TA\vec{x}\rangle.\end{aligned}$$

<u>Remark:</u> If A has non-trivial nullspace, then LS solutions are set of $\vec{x}$ satisfying:

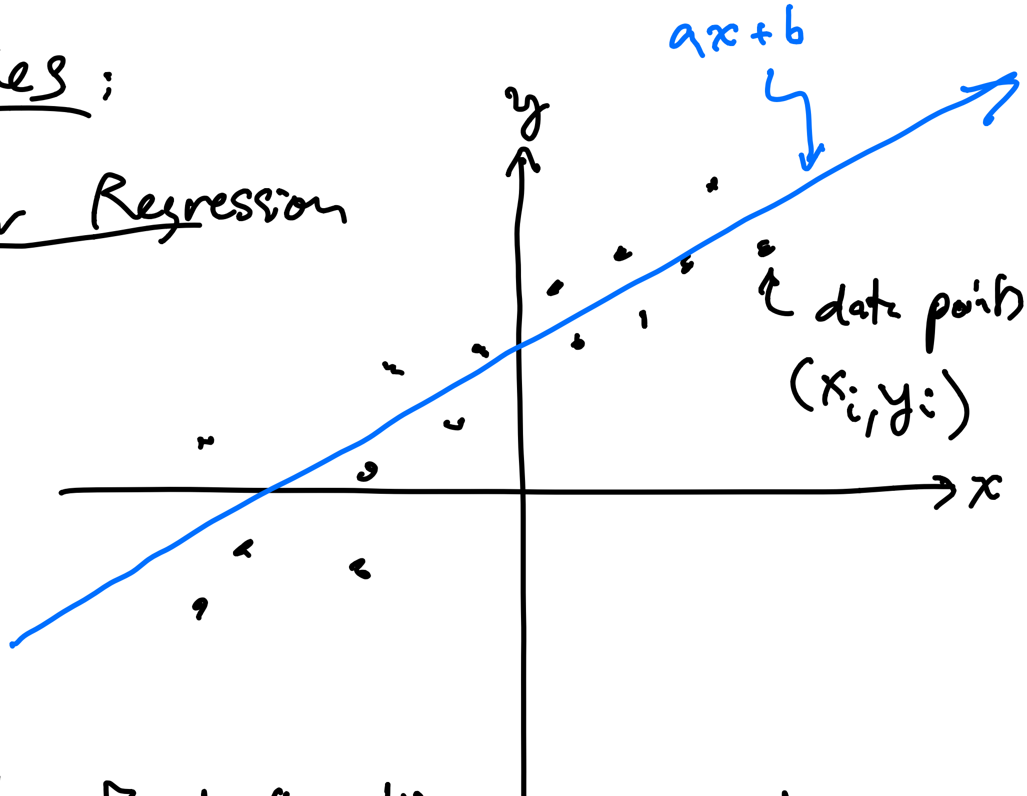
$$(A^TA)\,\vec{x} = A^T\vec{b}. \quad \left(\begin{matrix}\text{normal}\\ \text{equations}\end{matrix}\right)$$

<u>Note</u>: This sysem of Equations is <u>always</u>

consistent (need to show $R(A^T) = R(A^TA)$).

Examples:

Linear Regression



Want: Find "best" a, b such that

$$a x_i + b \approx y_i \quad \text{for } i = 1 \dots m.$$

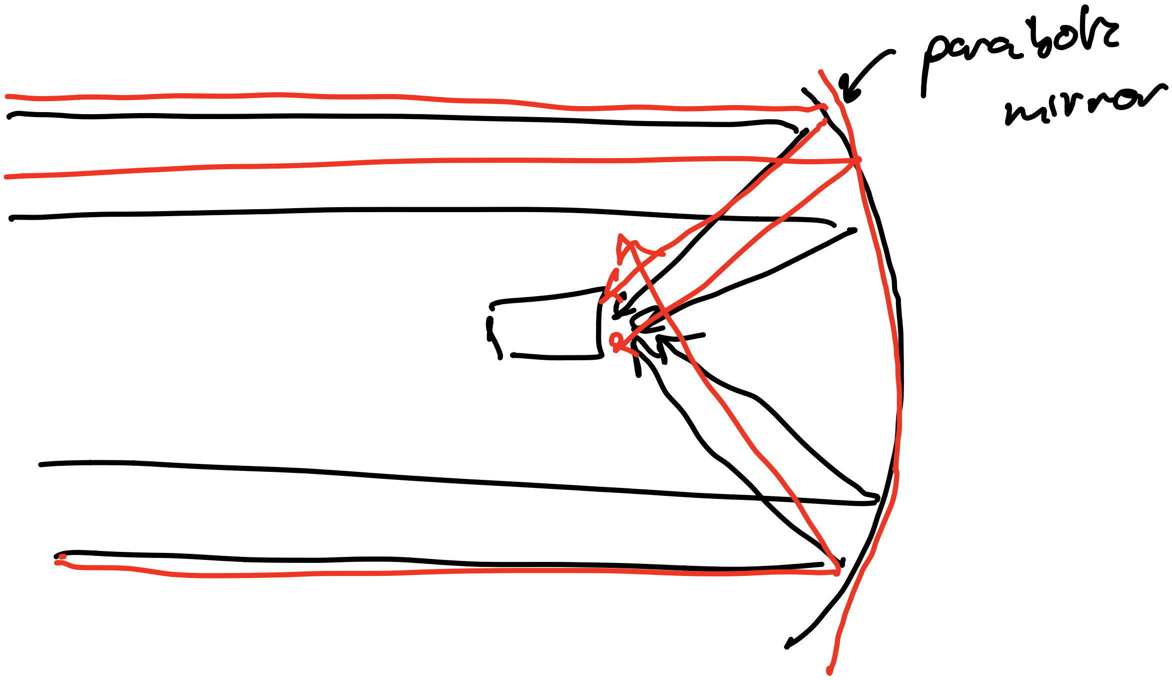
Answer: set up LS problem

$$\min_{a,b} \left\| \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_m & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} - \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{b} \right\|^2 = \min_{a,b} \sum_{i=1}^{m} \| a x_i + b - y_i \|^2$$

known

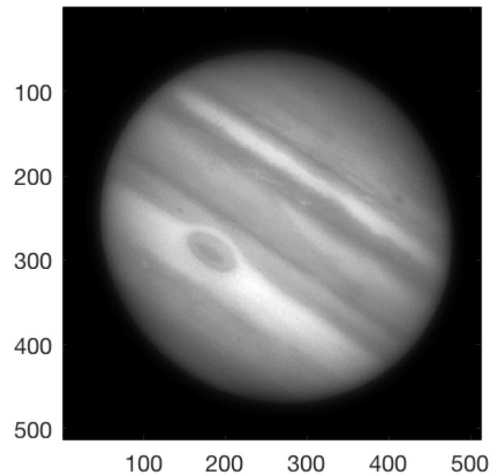
$$\begin{bmatrix} a \\ b \end{bmatrix} = (A^T A)^{-1} A^T \vec{b}.$$

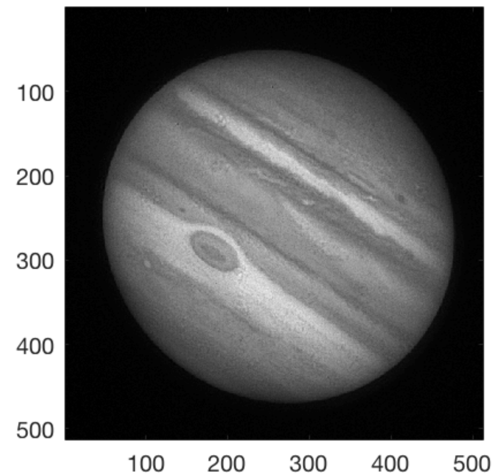
Example: Hubble Space Telescope



imprecise mirror made image blurry.

Image deblurring can be cast as a LS problem.





real image of Jupiter from HST

After designing a deblurring filter using Least Squares!

(See EE120 for details)