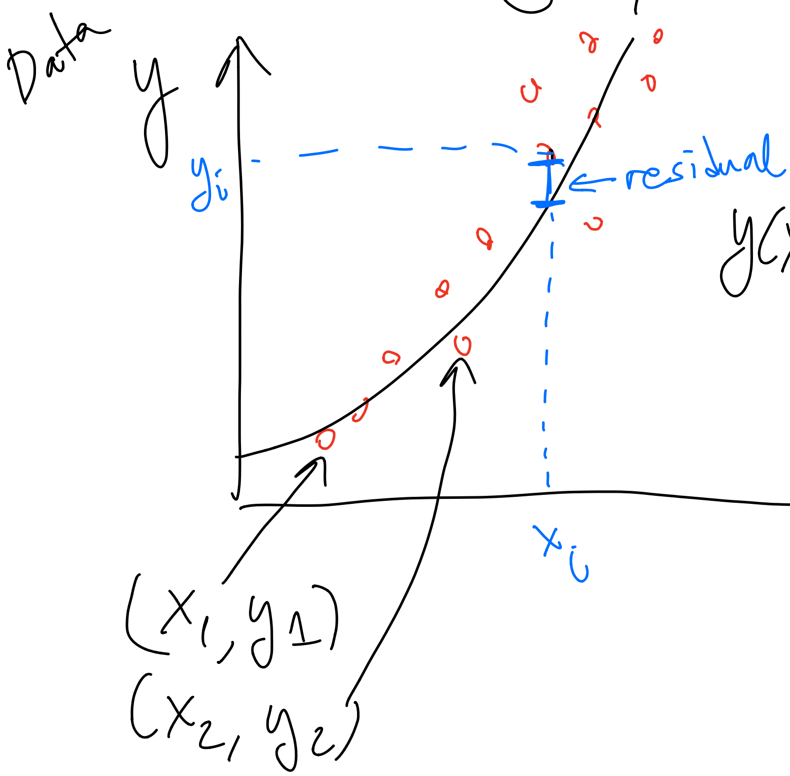
Linear Least Squares

A. Donev, Spring 2021

Fitting / regression



Model:

$$y(x) = a + bx + cx^2$$

Locations (Parameters)

$$y_i = a + b x_i + c x_i^2 + \varepsilon_i$$

$$|\varepsilon_i| << |y_i|$$

"noise" or "modeling error"

"Best fit"

$$r_i = y_i - (a + b x_i + c x_i^2) = \varepsilon_i$$

$\uparrow$ residual

$$(a,b,c)_{best} = \underset{a,b,c}{\arg\min} \|\vec{r}\|_2^2$$

$$\|\vec{r}\|_2^2 = \sum_{i=1}^{n} |r_i|^2$$

Least squares fit

$$\overleftrightarrow{X} = \begin{array}{c} 1 \rightarrow \\ \\ \\ n \rightarrow \end{array} \begin{array}{c} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \vdots & \\ 1 & x_n & x_n^2 \end{bmatrix} \\ \begin{array}{ccc} a & b & c \end{array} \end{array} \qquad \vec{p} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\vec{r} = \overleftrightarrow{X}\vec{p} - \vec{y}$$

$$r_1 = (a + b x_1 + c x_1^2) - y_1$$

$$\vec{y}_{model}(\vec{p}) = \underbrace{\overleftrightarrow{X}\,\vec{p}}_{\text{linear mapping}}$$

linear least squares

Instead: $\cancel{y} = e^{-ax} \cdot \cos(bx)$

non linear least squares

$$y = a\cos(x) + b e^{x} \quad \checkmark$$

$$\boxed{\vec{p} = \arg\min_{\vec{p}} \left\| \overleftrightarrow{X}\,\vec{p} - \vec{y} \right\|_2^2}$$

How to find p?

$y = X p$ (notation)

Overdetermined

"p solves $y = Xp$ in the least squares sense"

$p = X \backslash y$ in Matlab works

$$Ax = b$$

Formula for the fit and x data

$$A = [m \times n]$$

$m \gg n$

$x = [n \times 1]$ unknown parameters

$b = [m \times 1]$ y data

Linear Space of all solutions is $\mathrm{im}(A)$

If $b \in \mathrm{im}(A)$ there $\exists x$

$$\begin{cases} A x_1 = b \\ A x_2 = b \end{cases}$$

Imagine two distinct solutions

$$A x_1 - A x_2 = 0$$

$$A(x_1 - x_2) = 0$$

If A is full-rank $x_1 = x_2$

$$Ax = 0$$

$$\begin{cases} - \text{ infinitely many} \\ - \text{ no nonzero solutions} \end{cases}$$

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$A = [m \times n] \quad m > n$$

If cols are linearly independent, then $\text{rank}(A) = n$ (full-rank matrix)

$$\Rightarrow x = 0$$

$\Rightarrow$ If A is full-rank,
and $b \in \mathrm{im}(A)$,
then x is <u>unique</u>

$$y = a\,x + bx^2 + cx$$
$$= (a+c)x + bx^2$$

$\Rightarrow$ a/c not unique

$$A = \begin{bmatrix} x_1 & x_1^2 & x_1 \\ x_2 & x_2^2 & x_2 \\ - & - & - \\ x_n & x_n^2 & x_n \end{bmatrix}$$

$$y = x_1 f(x) + x_2 g(x) + x_3 h(x) \ldots$$

$f(x)$, $g(x)$ and $h(x)$ are linearly independent

$$f(x) = \cos(x)$$
$$g(x) = \sin(x)$$
$$h(x) = \cos(x + \pi/4)$$
$$= .. \cos(x) + .. \sin(x)$$

$Ax = b$, $A = [m \times n]$
$m > n$
full-rank

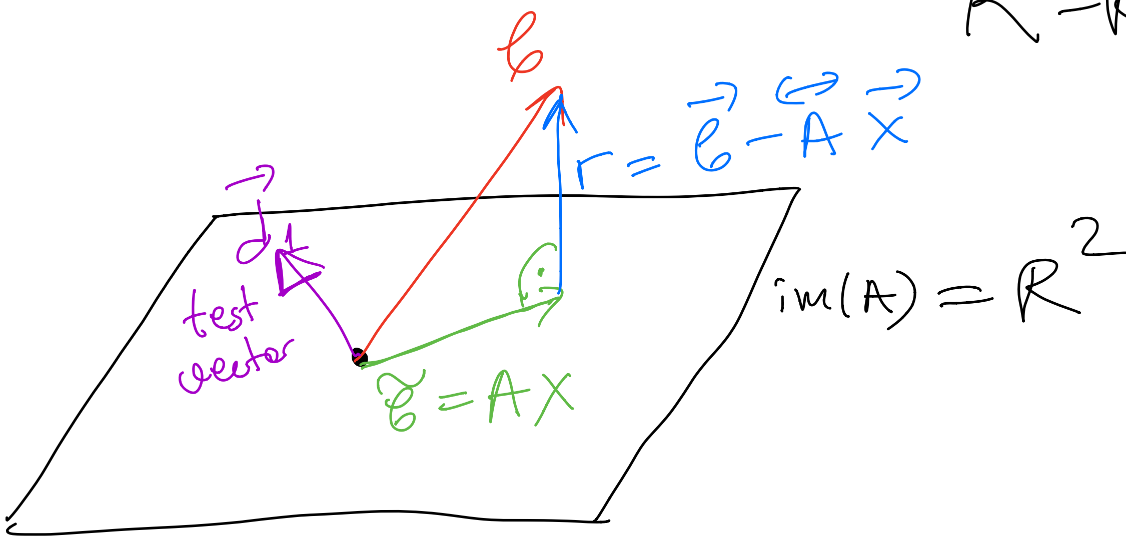
If $b \in \mathrm{Im}(A) \Rightarrow x$ is unique

$$Ax = \tilde{b} = \mathrm{Proj}_{\mathrm{Im}(A)}\, b$$

||| equivalent

least squares problem

$R^m = R^3$



$$r \perp im(A)$$

$$(b - Ax) \perp im(A)$$

$$r \cdot im(A) = 0$$

$$\begin{cases} r \cdot d_1 = 0 \\ r \cdot d_2 = 0 \end{cases}$$

d_1 & d_2 lin. ind.

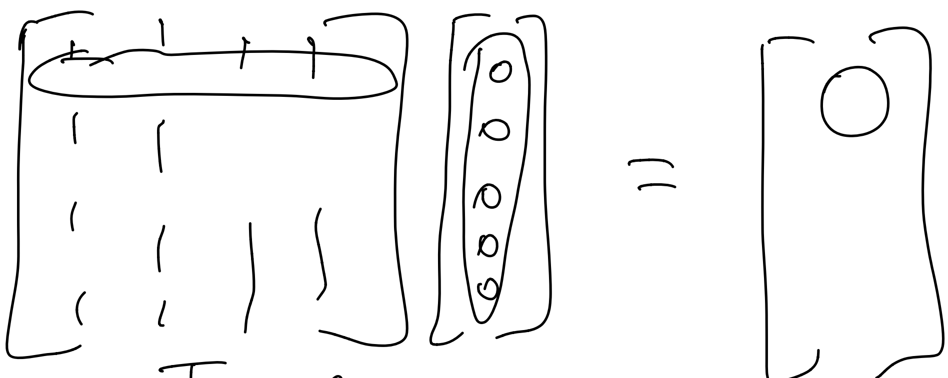
$\{d_1, d_2\}$ is a basis im(A)

$\{a_1, a_2, \ldots, a_n\}$ are the basis of $\mathrm{Im}(A)$

$$a_i \cdot (b - Ax) = 0 \quad \forall i$$

$i = 1, \ldots, n$

$\uparrow$ columns of A $=$ rows of A^T



$$A^T(b - Ax) = 0$$

$\underbrace{[n \times m]}[m \times 1] = [n \times 1]$

$$(A^T A)\, x = A^T b$$

normal equations

$A^T A = [n \times m][m \times n] = [n \times n]$

Square linear system of size n

If A is full-rank $\Rightarrow$
$A^T A$ is non-singular

$$B = A^T A$$

$$B^T = (A^T)(A^T)^T = A^T A = B$$

$B^T = B$, B is symmetric,
positive definite
(all eigs are real & positive)

$$B = L L^T$$ Cholesky factorization

diagonal entries of L are positive

In Matlab

$$(A^T A)\, x = A^T b$$

$$X = (A' * A) \backslash (A' * b)$$

will use Cholesky

computes the same answer as

$$X = A \backslash b$$

<u>Cost</u> of normal equations method

$$B = A^T A$$

$$A = [m \times n]$$

$$A^T = [n \times m]$$

$$\begin{array}{cc} & 4 \\ 3 & \begin{bmatrix} & & & \\ & & & \circ \\ & & & \end{bmatrix} \end{array}$$

n^2 answers

$$\text{FLOPS} = O(m n^2)$$

$$\tilde{b} = A^T b = O(mn) \text{ FLOPS}$$

$$[n \times m] \times [m \times 1]$$

$$\underline{mn^2} > mn$$

$$B x = \tilde{b} \xrightarrow{\text{solve}} \boxed{O(n^3)}$$

$$[n \times n] \times [n \times 1]$$

$$\begin{cases} mn^2 \gg n^3 \\ m \gg n \end{cases}$$

Total cost $\simeq O(mn^2)$

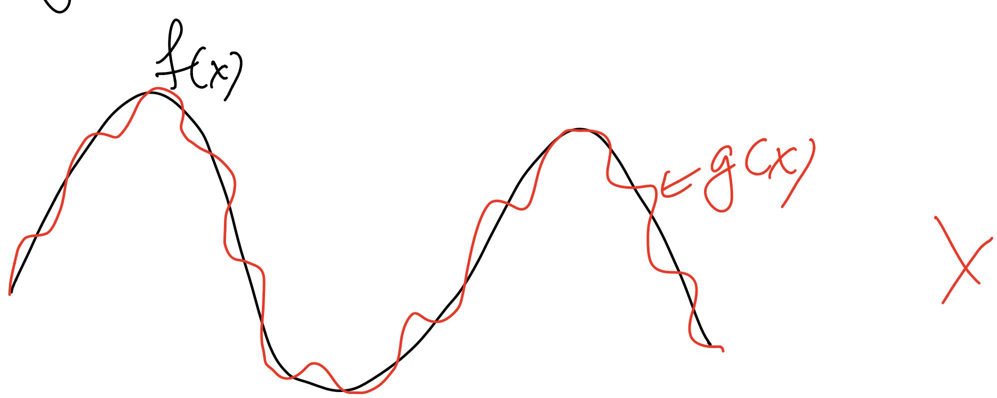
QR Factorization

$$\kappa_2(A^T A) \underline{\underline{\approx}} \left(\kappa_2(A)\right)^2$$

$$\kappa(A^T A) = 10^6 \gg \kappa(A) = 10^3$$

Ill-conditioned even if $Ax = b$ is well conditioned

$$y(x) = \underline{\underline{a}}\, f(x) + \underline{\underline{b}}\, g(x)$$



General idea: Use orthogonal matrices

If $Q^T Q = I$, $K_2(Q) = 1$

Q is perfectly well-conditioned

$K_2 \geq 1$

Find orthonormal basis for im(A)

$\{q_1, \ldots, q_n\}$

$$q_i \cdot q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$Q = [q_1 | \ldots | q_n]$$

$$Q^T Q = I = Q Q^T$$

$$Q^{-1} = Q^T$$

Understand:

→ $Ax = b$, A is square & invertible

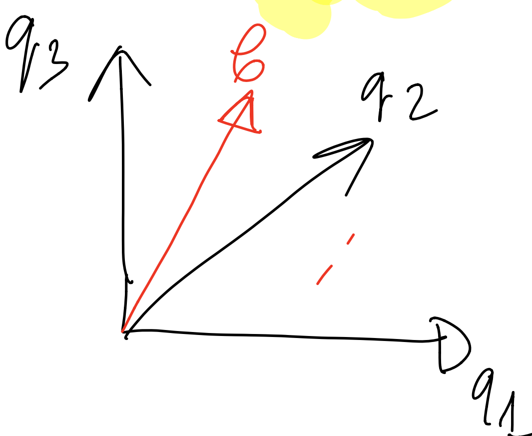
Finding x means:

Express b in the basis formed by columns of A

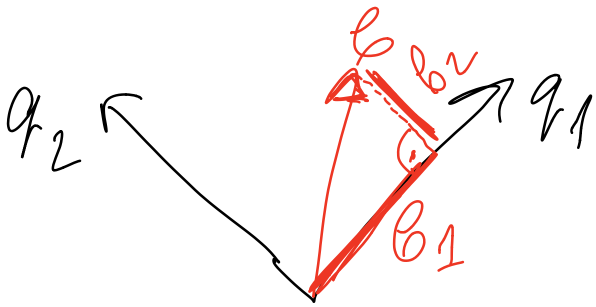
$$x_i = \vec{b} \cdot q_i$$

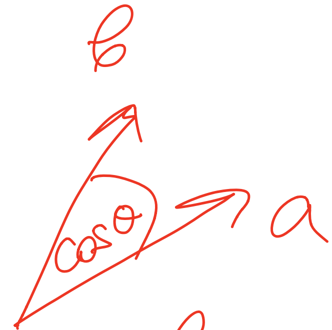
$$x = Q^T b$$

$$Qx = b$$



$$b = \underbrace{(\vec{b} \cdot \vec{q_1})}_{x_1} q_1 + \underbrace{(\vec{b} \cdot \vec{q_2})}_{x_2} q_2 + \underbrace{(\vec{b} \cdot \vec{q_3})}_{x_3} q_3$$

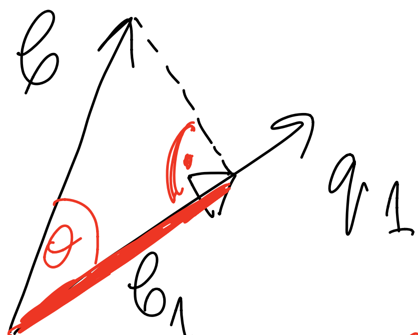




$$\cos\theta = \frac{a \cdot b}{ab}$$

$$\underline{a \cdot b} = ab\cos\theta$$

$$b_1 = b \cdot q_1 \qquad b_2 = b \cdot q_2$$

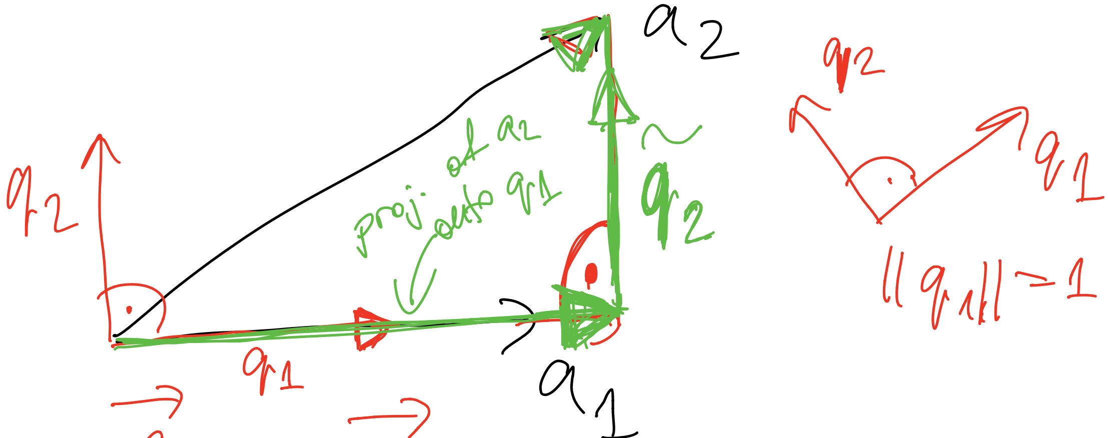


$$\|b_1\| = \|b\| \cos\theta$$

$$= \cancel{\|b\|} \, \frac{b \cdot q_1}{\cancel{\|b\|} \, \cancel{\|q_1\|}_{1}}$$

$$\|b_1\| = b \cdot q_1$$

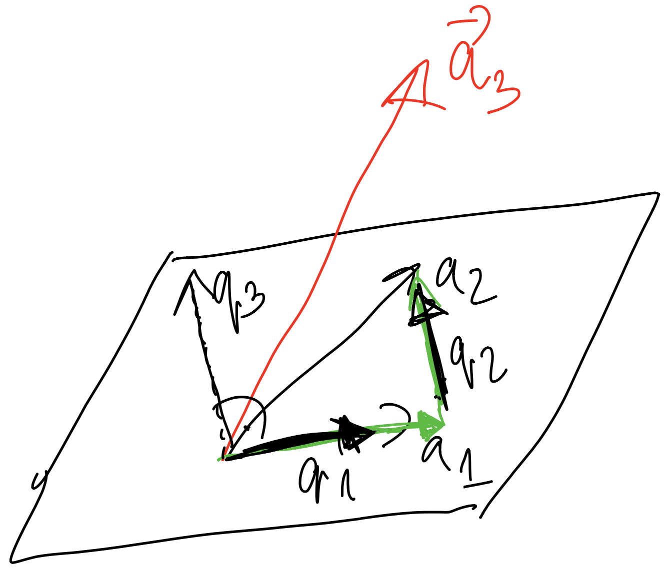
$$\begin{cases} \text{Im}(A) = \text{Im}(Q) \\ Q^{-1} = Q^T \end{cases}$$



$$\vec{q}_1 = \frac{\vec{a}_1}{\|a_1\|}$$

$$\tilde{q}_2 = a_2 - (a_2 \cdot q_1)\vec{q}_1$$

$$q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$$



$$\vec{\tilde{q}}_3 = \vec{a}_3 - (\vec{a}_3 \cdot \vec{q}_1)\vec{q}_1$$
$$- (a_3 \cdot \vec{q}_2)\,\vec{q}_2$$

$$\vec{q}_3 = \frac{\tilde{q}_3}{\| q_3 \|}$$

Gram-Schmidt process

$$\begin{cases} \vec{q}_1 = \dfrac{\vec{a}_1}{\|\vec{a}_1\|} \\ \tilde{q}_{k+1} = \vec{a}_{k+1} - \displaystyle\sum_{j=1}^{k} (\vec{a}_{k+1} \cdot \vec{q}_j)\vec{q}_j \\ q_{k+1} = \dfrac{\tilde{q}_{k+1}}{\|q_{k+1}\|} \end{cases}$$

Standard GS

Unstable to roundoff error

$$\vec{q}_{k+1} \cdot \vec{q}_2 \neq 0$$

Lookup in Wiki: Modified GS

$$r_{11} = \| a_1 \|_2$$

$$r_{12} = a_2 \cdot q_1$$

$$r_{22} = \| a_2 - r_{12} q_1 \|$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

Upper triangular

$$A = QR$$

QR factorization of matrix A

$$[m \times n] = \underbrace{[m \times n]}_{\text{ortogonal columns}} \cdot \underbrace{[n \times n]}_{\text{upper triangular}}$$

$$A x = b$$

take A square & invertible

$$A = QR$$

$$A^{-1} = R^{-1} Q^{-1} = \underbrace{R^{-1} Q^{T}}$$

$$x = A^{-1} b = R^{-1} Q^{T} b$$

$$x = R^{-1} Q^{T} b$$

MATLAB
$$\left\{ \begin{array}{l} [Q, R] = qr(A) \quad O(mn^2) \\ x = R \backslash (Q' * b) \end{array} \right.$$

$x = A \backslash b$ forward substitution

$O(n^2)$ cheap

Claim: Same code works for linear least squares

$$\begin{cases} y = Q^T b & [n \times m][m \times 1] = [n \times 1] \\ \text{solve } \boxed{Rx = y} \end{cases}$$

$$[n \times n]\,[n \times 1] = [n \times 1]$$

$$(A^T A)x = A^T b \quad \text{(normal eqs.)}$$

$A = QR$

$$(R^T \underbrace{Q^T Q}_{\text{Identity}} R)x = R^T Q^T b$$

$$R^{-T} \mid R^T R x = R^T Q^T b$$

$$Rx = Q^T b \quad \checkmark$$

Theorem: If A is full rank then R is invertible

$$A^T A x = A^T b$$

$$\min \| Ax - b \|_2^2$$

$$\min_x (Ax - b)^T W (Ax - b) =$$

$$= (x^T A^T - b^T)(Ax - b) =$$

$$= x^T A^T A x - \underline{x^T A^T b} - \underbrace{b^T A x}_{(Ax,\, b)} + b^T b$$

$$\min_{\vec{x}} \; \vec{x}^T A^T A \vec{x} - 2 x^T A^T b + b^T b$$

$$f(x,y) \quad \begin{cases} \partial f / \partial x = 0 \\ \partial f / \partial y = 0 \end{cases}$$

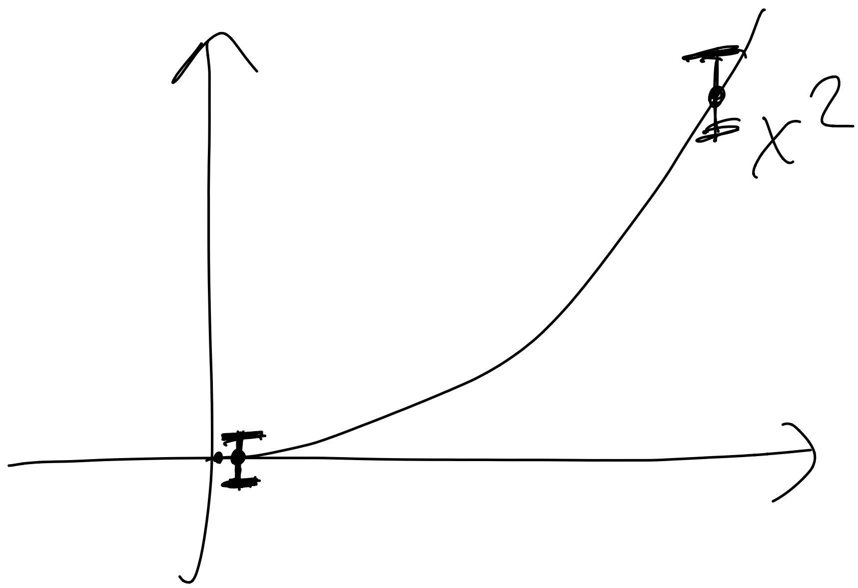
\nabla $\rightarrow$ $$\nabla_x \left(x^T A^T A x - 2x^T A^T b + \underline{\underline{b^T b}} \right)$$

$$\cancel{2} A^T A x - \cancel{2} A^T b = 0$$

$$A^T A x = A^T b$$

$$\nabla_x (x \cdot b) = b$$

$$\nabla_x (x^T A x) = A x$$



x^2