Least Squares Solution by QR

Recall 2-norm invariant under orthogonal transformations:

$$\|Qx\|_2^2 = x^T Q^T Q x = x^T x = \|x\|_2^2$$

<u>makes sense</u>: rotating or reflecting a vector doesn't change its length

<u>CASE I</u> : A has independent columns

$$\underset{m\times n}{A} = \underset{m\times m}{Q} \ \underset{m\times n}{R}$$

$$\begin{array}{c} m \end{array} \underbrace{\left(\; A \; \right)}_{n} = \left(\underbrace{\hat{Q}}_{n} \;\middle|\; \underbrace{\tilde{Q}}_{m-n} \right) \begin{array}{c} \left(\begin{array}{c} \hat{R} \\ 0 \end{array} \right) \\ n \end{array} \begin{array}{l} n \\ m-n \end{array}$$

$$A = \hat{Q}\hat{R} + \tilde{Q}\,0$$

$$\|r\|_2^2 = \| b - Ax \|_2^2$$

$$= \| b - \hat{Q}\hat{R}x \|_2^2$$

$$= \| Q^T b - Q^T \hat{Q}\hat{R}x \|_2^2$$

$$= \left\| \begin{pmatrix} \hat{Q}^T b \\ \tilde{Q}^T b \end{pmatrix} - \begin{pmatrix} \hat{R}x \\ 0 \end{pmatrix} \right\|_2^2$$

$$= \| \hat{Q}^T b - \hat{R}x \|_2^2 + \| \tilde{Q}^T b \|_2^2$$

Choose $\hat{R}x = \hat{Q}^T b$

$x = (\hat{R})^{-1}\hat{Q}^T b$ solve by back substit.

$$\| r \|_2^2 = \| \tilde{Q}^T b \|_2^2$$

CASE II.

$$\operatorname{rank}(A) = r$$

$$AP = Q\begin{pmatrix} \hat{R} & S \\ 0 & 0 \end{pmatrix} \qquad \underset{r\times r}{\hat{R}} \text{ nonsingular}$$

$$= (\hat{Q} \,|\, \tilde{Q})\begin{pmatrix} \hat{R} & S \\ 0 & 0 \end{pmatrix}$$

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

$$= \| b - APP^T x \|_2^2$$

$$= \| b - APy \|_2^2$$

$$= \left\| b - Q\begin{pmatrix} \hat{R} & S \\ 0 & 0 \end{pmatrix} y \right\|_2^2 = \left\| Q^T b - \begin{pmatrix} \hat{R} & S \\ 0 & 0 \end{pmatrix} y \right\|_2^2$$

$$= \left\| \begin{pmatrix} \hat{Q}^T b \\ \tilde{Q}^T b \end{pmatrix} - \begin{pmatrix} \hat{R} y_1 + S y_2 \\ 0 \end{pmatrix} \right\|_2^2$$

$$= \| \hat{Q}^T b - (\hat{R} y_1 + S y_2) \|_2^2 + \| \tilde{Q}^T b \|_2^2$$

$$\hat{R} y_1 + S y_2 = \hat{Q}^T b$$

Solution is not unique. One solution:

$$y_1 = \hat{R}^{-1} \hat{Q}^T b \quad \text{(by back subst.)}$$

$$y_2 = 0$$

Least Squares with a Penalty Term

(Strang II.2) <u>regularized</u> <u>least squares</u>

A has dependent columns
($Ax = 0$ has non-trivial solution)

then $A^T A$ is not invertible

Tikhonov regularization (ridge regression)

$$\boxed{\min \| Ax - b \|_2^2 + \delta^2 \| x \|_2^2}$$

Normal equations

$$\begin{pmatrix} Ax \\ \delta x \end{pmatrix} \approx \begin{pmatrix} b \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} A^T & \delta I \end{pmatrix} \begin{pmatrix} A \\ \delta I \end{pmatrix} x = \begin{pmatrix} A^T & \delta I \end{pmatrix} \begin{pmatrix} b \\ 0 \end{pmatrix}$$

$$\Rightarrow \underbrace{(A^T A + \delta^2 I)}_{\text{invertible}} x = A^T b$$

<u>check</u>: $A^T A + \delta^2 I = V \Sigma^T U^T U \Sigma V^T + \delta^2 I$
$= V \Sigma^T \Sigma V^T + \delta^2 V V^T$
$= V (\Sigma^T \Sigma + \delta^2 I) V^T$

Aside:

Section III.4

$$\min \frac{1}{2}\| Ax-b \| + \lambda \| x \|_1$$

instead of $\hat{x}$ with min. norm,

the ℓ^1 norm leads to a <u>sparse</u> $\hat{x}$

First, look at scalar case

① $$(A^T A + \delta^2 I)x = A^T b$$

<u>scalar</u>

$$(\sigma^2 + \delta^2)x = \sigma b$$

$$x = \frac{\sigma}{\sigma^2+\delta^2} b$$

if $\sigma \neq 0$, then $\frac{\sigma}{\sigma^2+\delta^2} \to \frac{1}{\sigma}$ as $\delta \to 0$

if $\sigma = 0$, then $\frac{\sigma}{\sigma^2+\delta^2} = 0$ for any $\delta > 0$

<u>diagonal</u> consider $A = \Sigma$, then

① $\Rightarrow$ $$(\Sigma^T \Sigma + \delta^2 I)^{-1} \Sigma^T$$

is diagonal with entries $(\sigma_i^2 + \delta^2)^{-1}\sigma_i$

same argument as before...

<u>any A</u>

① $$(V\Sigma^T \cancel{U^T} \cancel{U} \Sigma V^T + \delta^2 I)^{-1} V \Sigma^T U^T$$

$$= V(\Sigma^T\Sigma + \delta^2 I)^{-1} \cancel{V^T V} \Sigma^T U^T$$

$$= V(\Sigma^T\Sigma + \delta^2 I)^{-1} \Sigma^T U^T$$

$$\to A^+ \text{ as } \delta \to 0$$

Example

$$\begin{pmatrix} 2 & 0 \\ 0 & 2^{-100} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \begin{matrix} x = \frac{1}{2} \\ y = 2^{100} \end{matrix}$$

regularized

$$\min \left\| \begin{pmatrix} 2 & 0 \\ 0 & 2^{-100} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_2^2 + \delta^2 \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|_2^2$$

$$(A^T A + \delta^2 I) x = A^T b$$

$$\Rightarrow \begin{pmatrix} 2 & 0 \\ 0 & 2^{-100} \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 2^{-100} \end{pmatrix} + \begin{pmatrix} \delta^2 & 0 \\ 0 & \delta^2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2^{-100} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\Rightarrow \left(\begin{pmatrix} 4 & 0 \\ 0 & 2^{-200} \end{pmatrix} + \begin{pmatrix} \delta^2 & 0 \\ 0 & \delta^2 \end{pmatrix} \right) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 2^{-100} \end{pmatrix}$$

$$\begin{pmatrix} 4+\delta^2 & 0 \\ 0 & 2^{-200}+\delta^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 2^{-100} \end{pmatrix}$$

$$\Rightarrow \begin{matrix} (4+\delta^2) x = 2 \\ (2^{-200}+\delta^2) y = 2^{-100} \end{matrix}$$

$$\Rightarrow x = \frac{2}{4+\delta^2}, \quad y = \frac{2^{-100}}{2^{-200}+\delta^2}$$

without regularization y is <u>huge</u>
w/ regularization y is tiny

Tikhonov regularization

+ effective for dealing with null spaces and numerical issues

+ can improve conditioning

- $\hat{x}$ will not satisfy $A\hat{x} = b$

- $(A^T A + \delta^2 I)$ may be poorly conditioned. increasing δ^2 improves conditioning but makes $Ax = b$ less accurate.

Weighted Least Squares

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

means giving equal weight to all the equations

But if we know something more about the data, we can assign greatest weight to most reliable data.

weighted least squares:

$$\min \| W^{\frac{1}{2}} (b - Ax) \|_2^2$$

$$\Rightarrow \quad A^T W A x = A^T W b \qquad \text{normal equations}$$

ex. W is matrix of inv. variances
$\mathrm{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_n^{-2})$

Example Let $x = b_1$ and $x = b_2$ be indep. noisy measurements of x with variances σ_1^2, σ_2^2, respectively. Solve by weighted L.S.

mean $\dfrac{x_1 + x_2 + \cdots + x_n}{n} = \mu$

variance $\sigma^2 = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$

<u>regular L.S.</u>

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} x = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

normal eq: $\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} x = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

$$2x = b_1 + b_2 \Rightarrow \boxed{x = \frac{b_1 + b_2}{2}}$$

<u>weighted L.S.</u>

$$\frac{1}{\sigma_1} x = \frac{b_1}{\sigma_1} \qquad \begin{pmatrix} \frac{1}{\sigma_1} \\ \frac{1}{\sigma_2} \end{pmatrix} x = \begin{pmatrix} \frac{1}{\sigma_1} b_1 \\ \frac{1}{\sigma_1} b_2 \end{pmatrix}$$

$$\frac{1}{\sigma_2} x = \frac{b_2}{\sigma_2}$$

$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) x = \frac{1}{\sigma_1^2} b_1 + \frac{1}{\sigma_2^2} b_2$$

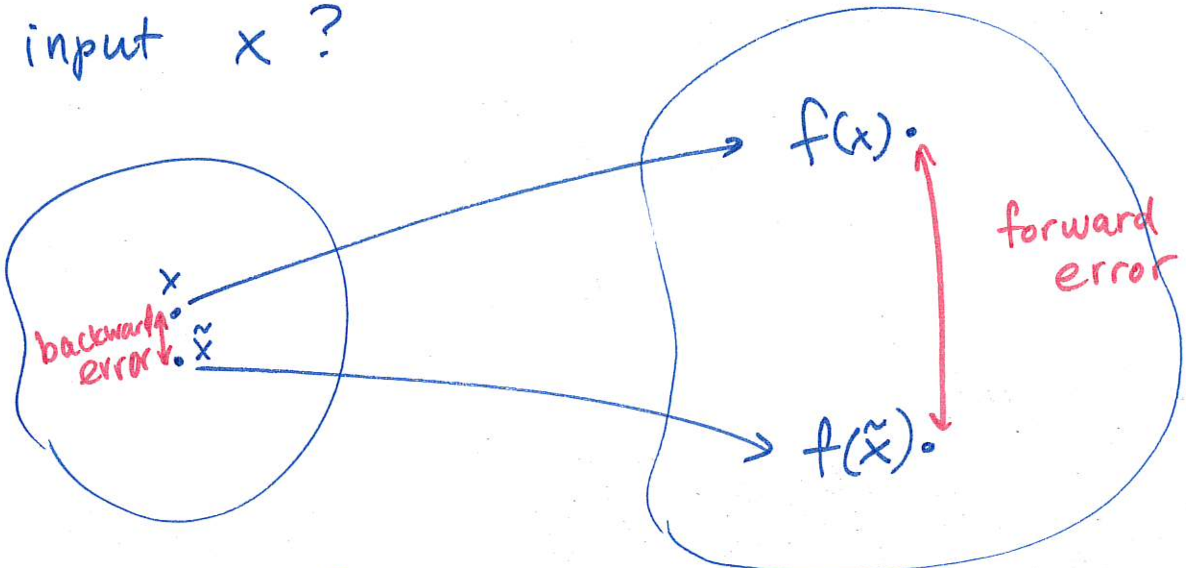
$$\Rightarrow \quad x = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \left(\frac{1}{\sigma_1^2} b_1 + \frac{1}{\sigma_2^2} b_2 \right)$$

$$\boxed{x = \frac{\sigma_2^2 b_1 + \sigma_1^2 b_2}{\sigma_1^2 + \sigma_2^2}}$$

$$\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \frac{\sigma_2^2 + \sigma_1^2}{\sigma_1^2 \sigma_2^2}$$

condition number of function $f(x)$

How sensitive is f to perturbations in its input x?



Does f amplify errors in input?
If it does too much, computing with f is dangerous. Numerical errors (roundoff errors) will be amplified + degrade quality of (accuracy) computed solution.

Quantify how much f amplifies errors in input:

$$\boxed{K_f = \frac{\left[\dfrac{|f(\tilde{x}) - f(x)|}{|f(x)|}\right]}{\left[\dfrac{|\tilde{x} - x|}{|x|}\right]}} \quad \left(\frac{\text{relative error in output, } f.}{\text{relative error in input}}\right)$$

K_f is an <u>amplification factor</u>

Matrix condition number

$$K_A = \frac{\left(\dfrac{\| A\tilde{x} - Ax \|}{\| Ax \|}\right)}{\left(\dfrac{\| \tilde{x} - x \|}{\| x \|}\right)} = \text{cond}(A)$$

rearrange

$$K_A = \frac{\left(\dfrac{\| A(\tilde{x} - x) \|}{\| \tilde{x} - x \|}\right)}{\left(\dfrac{\| Ax \|}{\| x \|}\right)}$$

$$\leq \frac{\| A \|}{\left(\dfrac{\| Ax \|}{\| x \|}\right)}$$

Let $y = Ax \Rightarrow x = A^{-1}y$

$$= \frac{\| A \|}{\left(\dfrac{\| y \|}{\| A^{-1}y \|}\right)} = \| A \| \left(\frac{\| A^{-1}y \|}{\| y \|} \right)$$

$$\leq \| A \| \, \| A^{-1} \|$$

In 2-norm

$$K_A = \| A \|_2 \, \| A^{-1} \|_2$$

$$= \sigma_1 \cdot \frac{1}{\sigma_n}$$

$$\boxed{K_A = \frac{\sigma_1}{\sigma_n}}$$

$$= \text{cond}_2(A)$$

$\text{cond}(A) = 1$ good

$\text{cond}(A) \geq 1$

normal equations square the condition number

$$\operatorname{cond}(A^TA) = \operatorname{cond}(A)^2$$

check

- $A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T$

  $\Rightarrow \operatorname{cond}_2(A^TA) = \dfrac{\sigma_1^2}{\sigma_n^2}$

- $A = U\Sigma V^T$

  $\operatorname{cond}(A) = \dfrac{\sigma_1}{\sigma_n}$

Example

$$A = \begin{pmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} \qquad \varepsilon^2 < \varepsilon_{mach}$$

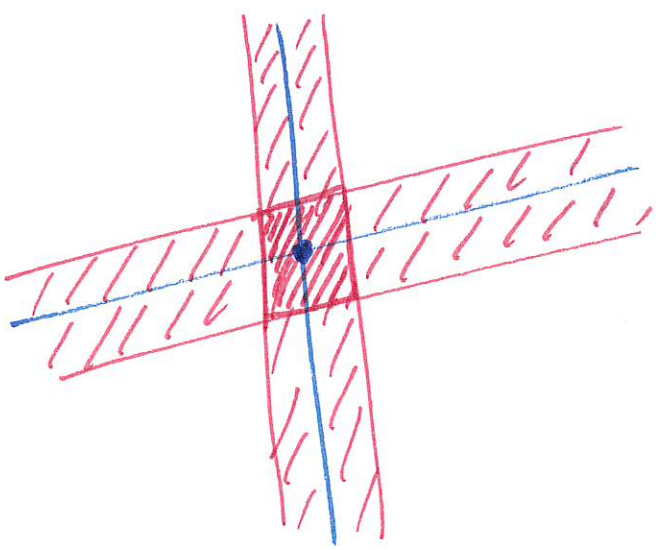
$$A^TA = \begin{pmatrix} 1 & \varepsilon & 0 \\ 1 & 0 & \varepsilon \end{pmatrix}\begin{pmatrix} 1 & 1 \\ \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} = \begin{pmatrix} 1+\varepsilon^2 & 1 \\ 1 & 1+\varepsilon^2 \end{pmatrix}$$

$$\approx \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

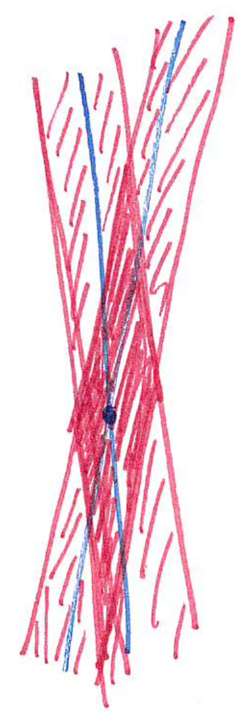
properties

1. $\text{cond}(A) \geq 1$
2. $\text{cond}(I) = 1$
3. $\text{cond}(\alpha A) = \text{cond}(A)$
4. D diagonal, $\text{cond}(D) = \dfrac{\max_i |d_i|}{\min_i |d_i|}$

intuition



well-conditioned



ill-conditioned

$f(x)$ $\qquad$ <u>f differentiable</u>

$$K_f = \frac{|f(\tilde{x}) - f(x)| \,/\, |f(x)|}{|\tilde{x} - x| \,/\, |x|}$$

$$\Delta x = \tilde{x} - x$$

$$K_f = \frac{|f(x+\Delta x) - f(x)| \,/\, |f(x)|}{|\Delta x| \,/\, |x|} \longrightarrow \frac{|f'(x)|\,|x|}{|f(x)|}$$

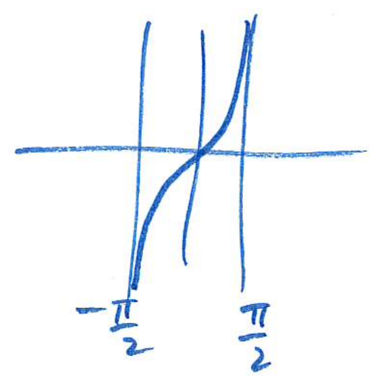
<u>Example</u> $f(x) = \sqrt{x} = x^{\frac{1}{2}}$

$$f'(x) = \frac{1}{2} x^{-\frac{1}{2}}$$

$$K_f = \frac{|\frac{1}{2} x^{-\frac{1}{2}}\, x|}{|x^{\frac{1}{2}}|} = \frac{1}{2}$$

well-conditioned

<u>Example</u> $f(x) = \tan(x) = \frac{\sin(x)}{\cos(x)}$



$$f'(x) = \sec^2(x) = 1 + \tan^2(x)$$

$$K_f = \frac{|x(1+\tan^2(x))|}{|\tan(x)|}$$

very large near $\frac{\pi}{2} = x$