Least squares with the SVD

Note Title 2024-11-13

$$\min_{x \in \mathbb{R}^n} \|Ax - y\|_2 \qquad A \in \mathbb{R}^{m \times n} = USV^T$$

$$x = v_1 \frac{1}{\sigma_1} u_1^T y + v_2 \frac{1}{\sigma_2} u_2^T y + \dots + v_n \frac{1}{\sigma_n} u_n^T y$$

Small σ_i are more sensitive to noise, computational errors

$\leftrightarrow$ least important features

$$x_{reg} = v_1 \frac{1}{\sigma_1} u_1^T y + \dots + v_k \frac{1}{\sigma_k} u_k^T y$$

we stop at k, if σ_{k+1} is too small!

"truncated SVD solution".

Variant: ridge regression, Tikhonov regularization.

$$x_{tik} = \min_{x \in \mathbb{R}^n} \|Ax - y\|^2 + \alpha^2 \|x\|^2$$

$$= \min_{x \in \mathbb{R}^n} \left\| \begin{bmatrix} A \\ \alpha I_n \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2 \qquad \left\| \begin{bmatrix} Ax - y \\ \alpha x \end{bmatrix} \right\|^2$$

$$\underbrace{\begin{bmatrix} A \\ \boxtimes \end{bmatrix}}_{\hat{A}} x - \underbrace{\begin{bmatrix} y \\ 0 \end{bmatrix}}_{\hat{y}}$$

$$x_{tik} = (\hat{A}^T \hat{A})^{-1} \hat{A}^T y = \left(\begin{bmatrix} A^T & \alpha I \end{bmatrix} \begin{bmatrix} A \\ \alpha I \end{bmatrix} \right)^{-1} \begin{bmatrix} A^T & \alpha I \end{bmatrix} \begin{bmatrix} y \\ 0 \end{bmatrix}$$

$$= \underline{(A^T A + \alpha^2 I)^{-1} A^T y}$$

Similar to the formula for A^+ with normal equations, but with an added $\alpha^2 I$

$$A^TA + \alpha^2 I = V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^T + \alpha^2 \underbrace{I}_{VV^T} = V \begin{bmatrix} \sigma_1^2+\alpha^2 & & & \\ & \sigma_2^2+\alpha^2 & & \\ & & \ddots & \\ & & & \sigma_n^2+\alpha^2 \end{bmatrix} V^T$$

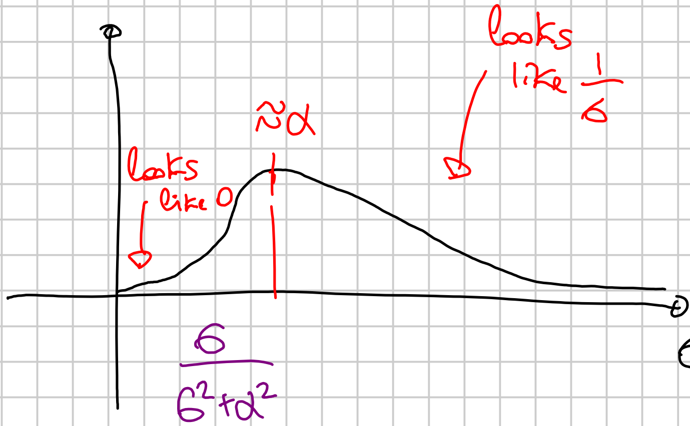
eigenvalues σ_i^2 can be very small in A^TA

eigenvalues of $A^TA+\alpha^2 I$ are at least α^2
$\Rightarrow$ can't be too close to 0

$$\hat{A} = \begin{bmatrix} A \\ \alpha I \end{bmatrix}$$ always has full column rank!

$$x_{tik} = v_1 \boxed{\frac{\sigma_1}{\sigma_1^2+\alpha^2}} u_1^T y + \dots + v_n \boxed{\frac{\sigma_n}{\sigma_n^2+\alpha^2}} u_n^T y$$

The inverses $\frac{1}{\sigma_i}$ are replaced by $\frac{\sigma_i}{\sigma_i^2+\alpha^2}$



When $\sigma \gg \alpha$,

$$\frac{\sigma}{\sigma^2+\alpha^2} \approx \frac{\sigma}{\sigma^2} = \frac{1}{\sigma}$$

When $\sigma \ll \alpha$

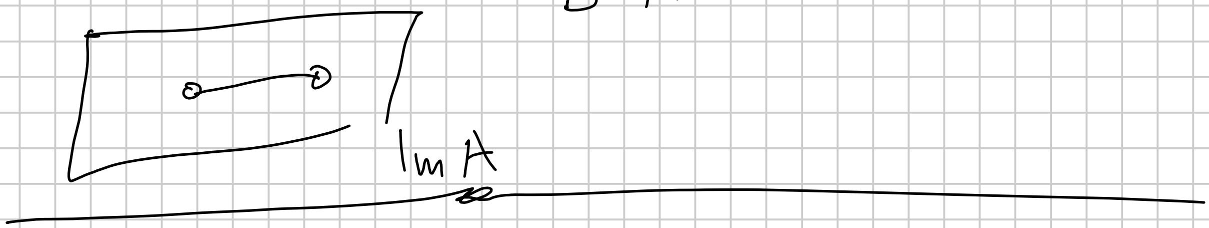
$$\frac{\sigma}{\sigma^2+\alpha^2} \approx \frac{\sigma}{\alpha^2} \approx 0$$

Similar effect to that of truncated SVD:

$$\frac{\sigma}{\sigma^2+\alpha^2} \approx \begin{cases} \frac{1}{\sigma} & \text{if } \sigma \text{ large} \\ 0 & \text{if } \sigma \text{ small} \end{cases}$$

$$[0, \underbrace{\|A\|}_{\sigma_1}]$$

$b = Ax$



Conditioning, sensitivity, accuracy of algorithms

Condition number of a problem:

$y = f(x)$: computing y from x

If I change x to a nearby $\tilde{x}$, how do $\tilde{y} = f(\tilde{x})$ and y relate to each other?

ex: $y = x^2$ $\qquad \tilde{x} = x + \delta \qquad \tilde{y} = (x+\delta)^2 = x^2 + 2x\delta + \delta^2$

$$\tilde{y} - y \approx 2x\delta$$

Error δ is amplified to $\boxed{2x\delta}$

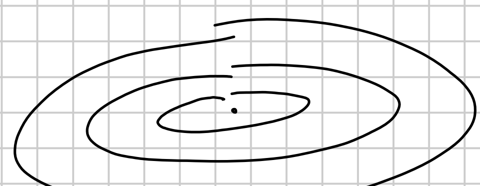
derivative $f'(x)$

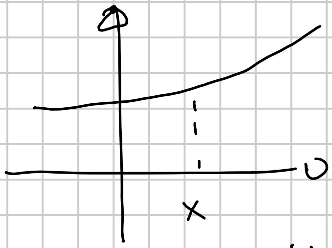
Absolute condition number: it's the number K such that

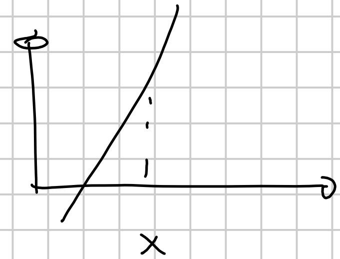
$$\overset{2x\delta}{|\tilde{y} - y|} = \overset{2x}{K} \cdot \overset{\delta}{|\tilde{x} - x|} + O(\delta^2)$$

$$K = |f'(x)|$$

Let us now assume $F : \mathbb{R}^m \to \mathbb{R}^n$. We can formulate similar computations, but now the result depends on direction:







depending on the tomography / direction chosen, the derivative changes.

Def: K (absolute cond. number) is the smallest number such that

$$\|\tilde{y} - y\| \leq K \|\tilde{x} - x\| + O(\delta^2) \qquad \delta = \|\tilde{x} - x\|$$

$$y = F(x)$$
$$\tilde{y} = F(\tilde{x})$$

along all possible directions of approach.

$$K_{abs}(F, x) = \lim_{\delta \to 0} \sup_{\substack{\tilde{x} \text{ s.t.} \\ \|\tilde{x} - x\| \leq \delta}} \frac{\|F(\tilde{x}) - F(x)\|}{\|\tilde{x} - x\|}.$$

For regular functions, $K_{abs}(F, x) = \| J_{F,x} \|$, norm of the Jacobian (Jacobian matrix, i.e. matrix of partial derivatives).

Absolute vs. relative errors:

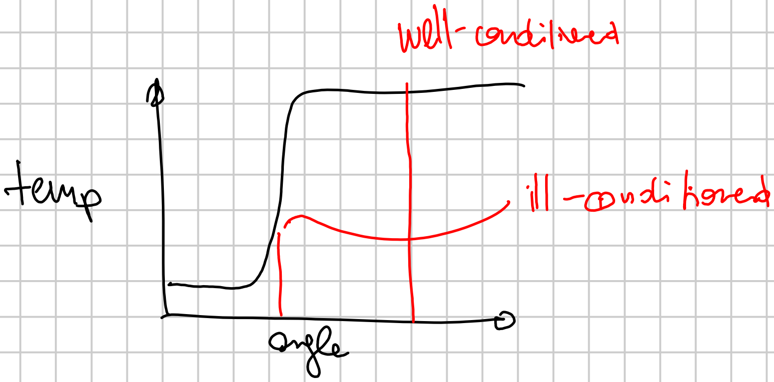
$\dfrac{\|\tilde{x} - x\|}{\|x\|}$ better than $\|\tilde{x} - x\|$, as it gives order-of-magnitude information

$$\frac{\|\tilde{x} - x\|}{\|x\|} \approx 10^{-5} \Leftrightarrow 5 \text{ exact digits (approximately)}$$

If $\|\tilde{y} - y\| \le \kappa_{abs}(f,x)\, \|\tilde{x} - x\|$

$$\frac{\|\tilde{y} - y\|}{\|y\|} \le \boxed{\frac{\kappa_{abs}(f,x)}{\|y\|} \cdot \|x\|} \cdot \frac{\|\tilde{x} - x\|}{\|x\|}$$

$$\kappa_{rel}(f,x) = \frac{\|J_{f,x}\| \cdot \|x\|}{\|y\|}. \qquad \left[\text{cfr. } \frac{f'(x)\,x}{f(x)}\right]$$



Condition numbers are properties of problems, not algorithms

Condition number of solving square linear systems:

$A \in \mathbb{R}^{n \times n} \qquad Ax = y \qquad A\tilde{x} = \tilde{y}$

condition number of the solution x w.r.t. changes in the right-hand side y.

$$\|\tilde{x} - x\| = \|A^{-1}\tilde{y} - A^{-1}y\| = \|A^{-1}(\tilde{y} - y)\| \boxed{\le} \|A^{-1}\| \cdot \|\tilde{y} - y\|$$

$$\|y\| = \|Ax\| \le \|A\| \cdot \|x\| \iff \|x\| \boxed{\ge} \frac{\|y\|}{\|A\|}$$

$$\leadsto \frac{\|\tilde{x} - x\|}{\|x\|} \le \underbrace{\|A\| \cdot \|A^{-1}\|}_{\kappa(A)} \cdot \frac{\|\tilde{y} - y\|}{\|y\|}$$

"condition number of A"

$\Leftrightarrow$ condition number of solving lin. systems w.r.t changes in y.

Condition number w.r.t. changes in A:

$$\tilde{A} = A + E \qquad Ax = y \qquad (A+E)\tilde{x} = y \qquad \tilde{x} = x + \delta$$

$$\cancel{y} = (A+E)(x+\delta) = \cancel{Ax} + Ex + A\delta + \boxed{E\delta}$$

product of two errors, much smaller than the others if $\|E\|, \|\delta\|$ are $<<1$

$$\delta = -A^{-1}Ex$$

$$\frac{\|\tilde{x}-x\|}{\|x\|} = \frac{\|\delta\|}{\|x\|} \leq \frac{\|A^{-1}\| \cdot \|E\| \cdot \cancel{\|x\|}}{\cancel{\|x\|}} = \underbrace{\|A^{-1}\| \cdot \|A\|}_{\kappa(A)} \cdot \frac{\|\tilde{A}-A\|}{\|A\|}$$

$\Rightarrow$ the cond. number w.r.t. changes in A, is once again $\kappa(A)$.

Condition number and singular values

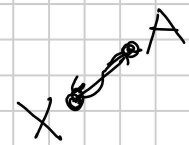
$$\|A\| = \sigma_1 \qquad A = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} V^T$$

$$A^{-1} = \underbrace{V}_{\text{orth.}} \underbrace{\begin{bmatrix} 1/\sigma_1 & & & 0 \\ & 1/\sigma_2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_n \end{bmatrix}}_{\text{diag}} \underbrace{U^T}_{\text{orth}} = \begin{bmatrix} v_n & \cdots & v_1 \end{bmatrix} \cdot \begin{bmatrix} \boxed{1/\sigma_n} & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_1 \end{bmatrix} \cdot \begin{bmatrix} u_n & \cdots & u_1 \end{bmatrix}^T$$

$$\|A^{-1}\| = \text{largest s.v. of } A^{-1} = \frac{1}{\sigma_n}$$

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = \frac{\sigma_1}{\sigma_n}.$$

Property: $\dfrac{1}{\kappa(A)} = \min\limits_{X \text{ singular}} \dfrac{\|X-A\|}{\|A\|}$

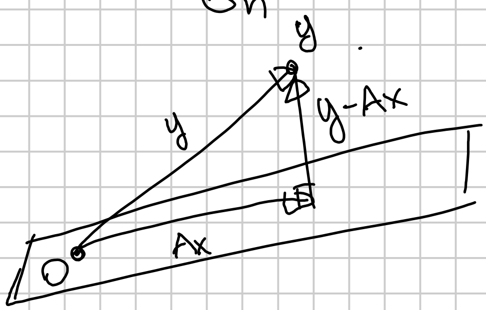


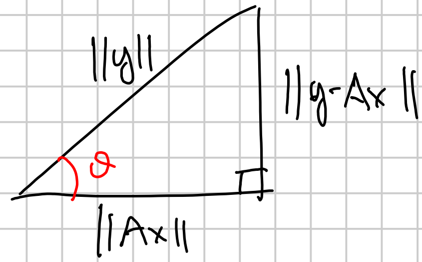
Condition number of least squares problems

$\min\limits_{x} \|Ax-y\|$ $\quad A \in \mathbb{R}^{m\times n}, \quad m>n$

1) $\kappa(A) = \dfrac{\sigma_1}{\sigma_n}$ (definition)

2)



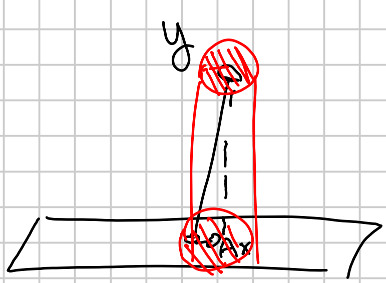


$$\vartheta = \arcsin \frac{\|y-Ax\|}{\|y\|}$$

Theorem:

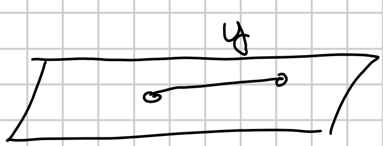
cond. number w.r.t. changes in y $\leq \dfrac{\kappa(A)}{\cos\vartheta}$

cond. number w.r.t. changes in A $\leq \kappa(A) + (\kappa(A))^2 \cdot \tan\vartheta$



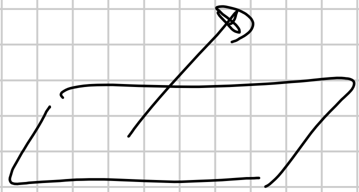
if $\vartheta \approx 90°$, Ax is much smaller than y

$\Rightarrow$ small changes in y become large in x.



If $\vartheta \approx 0°$,

$\kappa(A) + \kappa(A)^2 \tan\vartheta \approx \kappa(A)$



If θ is neither $\approx 0°$ nor $\approx 90°$

$$K(A) + K(A)^2 \tan\theta \approx K(A)^2$$