Singular Value Decomposition Applications

Gabriel Hugh Elkaim

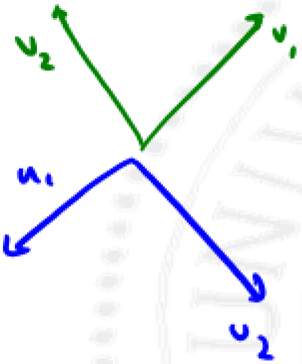
$$A \in \mathbb{R}^{4\times 4} \longrightarrow \text{SVD} \longrightarrow \Sigma = \mathrm{diag}(\underbrace{10, 7}_{\text{effectively } \mathbb{R}^{2\times 2}}, 0.1, 0.05)$$

$$10 \geq \frac{\|Ax\|}{\|x\|} \geq 0.05$$

$$y = Ax$$

$$\mathrm{Rank}(A) = 4$$

$$\kappa(A) = \frac{10}{0.05} = 200.$$



$$\dot{x} = Ax$$

$$y = Cx$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \longrightarrow \boxed{} \longrightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

$\text{rank}(\mathcal{O})$ is full

$$\kappa(\mathcal{O}) \triangleq \frac{\sigma_{max}}{\sigma_{min}} \quad \text{CONDITION \#.} \quad [1 \rightarrow \infty]$$

$$\text{RCOND} \triangleq \frac{\sigma_{min}}{\sigma_{max}} \quad [1 \rightarrow 0]$$

Singular Value Decomposition Applications

- Low Rank approximation via the SVD
- General pseudo-inverse
- Full SVD
- Image of Unit Ball under linear transformation
- SVD in estimation/inversion
- Sensitivity of linear equations to data error

General Pseudo-Inverse (1.3)

eigenvalues : $A = T \Lambda T^{-1}$ $\quad A \in \mathbb{R}^{n \times n}$

symmetric : $A = Q \Lambda Q^T$ $\quad A = A^T \in \mathbb{R}^{n \times n}$

svd : $A = U \Sigma V^T$ $\quad A \in \mathbb{R}^{n \times m}$, $\operatorname{rank}(A) = r$

$U^T U = I$ $\quad (\diagdown)$ $\quad V^T V = I$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$

A^+ defined for skinny, fat, full ranked matrices

General Pseudo-Inverse (2.3)

if A has an SVD : $A = U\Sigma V^T$

$\leftarrow [\diagdown]$ $r \times r$, positive

$A^+ = V\Sigma^{-1}U^T$ $\leftarrow$ pinv

skinny full rank matrix $A \in \mathbb{R}^{n\times m}$

$$\begin{array}{c} \xleftrightarrow{m} \\ n \updownarrow \begin{bmatrix} & \\ & \end{bmatrix} \end{array} \quad n > m$$

$\begin{bmatrix} 1/\sigma_1 & \cdots & 1/\sigma_m \end{bmatrix}$

$$A^+ = (A^TA)^{-1}A^T = \left[(U\Sigma V^T)^T(U\Sigma V^T)\right]^{-1}(U\Sigma V^T)^T$$

$$= \left[V\Sigma U^T U\Sigma V^T\right]^{-1} V\Sigma U^T =$$

$$= \left[V\Sigma^2 V^T\right]^{-1} V\Sigma U^T = V\Sigma^{-2}\boxed{V^TV}\Sigma U^T = \boxed{V\Sigma^{-1}U^T}$$

General Pseudo-Inverse (3.3)

A is fat & full rank $\rightarrow A^+ = A^T(AA^T)^{-1}$ $\qquad (ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

$$A^+ = (U\Sigma V^T)^T\left[(U\Sigma V^T)(U\Sigma V^T)^T\right]^{-1}$$

$$= V\Sigma U^T\left[U\Sigma V^T V\Sigma U^T\right]^{-1} = V\Sigma U^T\left[U\Sigma^2 U^T\right]^{-1}$$

$$= V\Sigma U^T\left[U\Sigma^{-2}U^T\right] = V\Sigma U^T U\Sigma^{-2}U^T = \boxed{V\Sigma^{-1}U^T}$$

Skinny full rank: $\begin{bmatrix} A \end{bmatrix} \rightarrow x_{ls} = A^{+} y = V \Sigma^{-1} U^T y$

Fat full rank: $\begin{bmatrix} A & & \end{bmatrix} \rightarrow x_{ln} = A^{+} y = V \Sigma^{-1} U^T y$

In general

$$X_{ls} = \{ z \mid \|Az - y\| = \min_w \|Aw - y\| \}$$

$x_{pinv} = A^{+} y \in X_{ls}$ has the min. norm.

Pseudo-Inverse via Regularization (1.2)

$\mu > 0$: let x_μ be a unique minimizer of

$\|Ax_\mu - y\|^2 + \mu\|x\|^2 \longrightarrow x_\mu = (A^TA + \mu I)^{-1} A^T y$.

$(A^TA + \mu I) > 0$.

$x_\mu = (V\Sigma^2V^T + \mu VV^T)^{-1} V\Sigma U^T$

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

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

$$V\begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\mu} & & \\ & \ddots & \\ & & \frac{\sigma_r}{\sigma_r^2+\mu} \end{bmatrix} U^T$$

Pseudo-Inverse via Regularization (2.2)

$$\lim_{\mu \to 0} V \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\mu} & & \\ & \ddots & \\ & & \frac{\sigma_r}{\sigma_r^2+\mu} \end{bmatrix} U^T \longrightarrow \underline{A^{\dagger} = V \Sigma^{-1} U^T}$$

Full SVD (1.3)

$A \in \mathbb{R}^{m \times n}$ rank$(A) = r$ $m > n$

$$A = U_1 \Sigma_1 V_1^T = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

$$\Sigma = \left[\begin{array}{c|c} \underbrace{\Sigma_1}_{r \times r} & \underbrace{0}_{r \times (n-r)} \\ \hline \underbrace{0}_{(m-r) \times r} & \underbrace{0}_{(m-r) \times (m-n)} \end{array}\right]$$

$U = [U_1 \vdots U_2] \in \mathbb{R}^{m \times m}$ $R(U) = \mathbb{R}^m$

$V = [V_1 \vdots V_2] \in \mathbb{R}^{n \times n}$ $R(V_1) = N(A)^\perp$

Full SVD (2.3)

$$A = [U_1 | U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ \hline V_2^T \end{bmatrix} = U \Sigma V^T$$

Σ^{-1} no longer allowed

σ_1 unambiguous

σ_{min} ? σ_r or ϕ.

$svd(A) = U \Sigma V^T$

$svd(A, '...') = [U_1, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$

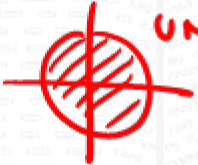
Full SVD (3.3)

$y = Ax$

$$x \xrightarrow[\ ]{n} \boxed{V^T} \xrightarrow[r]{V^T x} \boxed{\Sigma} \xrightarrow{r} \boxed{U} \xrightarrow{m} y = Ax$$

$A = U \Sigma V^T$

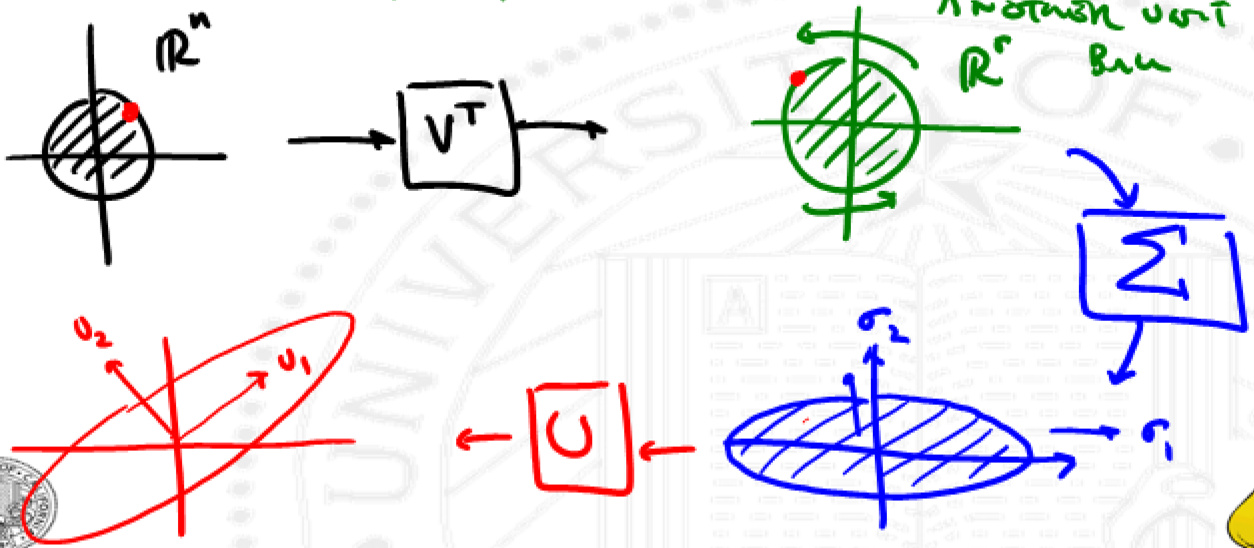
$\{x \mid \|x\| \leq 1\} \longrightarrow$



Unit Ball under Linear Transformation (1.3)

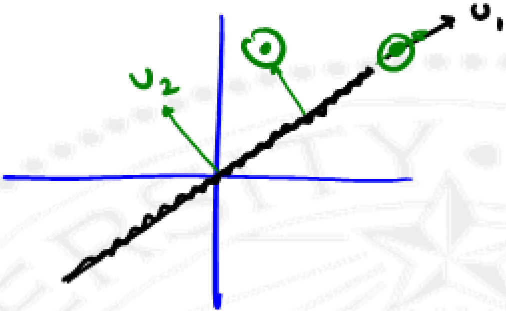
$\{x \mid \|x\| \leq 1\}$

$\|V^T x\| = \|x\|$



Unit Ball under Linear Transformation (2.3)

$$\Sigma = \begin{bmatrix} 2 & \\ & 1/100 \end{bmatrix}$$



$$\Sigma = \begin{bmatrix} 10 & & \\ & 9 & \\ & & 0.01 \end{bmatrix}$$

Unit Ball under Linear Transformation (3.3)

$A = U \Sigma V^T$

$A^+ = V \Sigma^{-1} U^T$

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{bmatrix}$$

SVD in estimation/inversion (1.3)

$$y = Ax + v$$

$y \in \mathbb{R}^m$ measurement

$x \in \mathbb{R}^n$ vector to be estimated

$v \in \mathbb{R}^m$ is an noise $\|v\| < \alpha$.

$\hat{x} = A^{+}y$

$\hat{x} = By$ $BA = I$ unbiased estimator

$\hat{x} = x$ if $v = 0$.

$$\tilde{x} \hat{=} \hat{x} - x = B(Ax + v) - x = \overbrace{BA}^{I}x + Bv - x = Bv$$

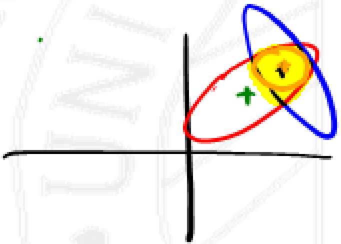
$$\tilde{x} \in \Sigma_{unB} = \{ Bv \mid \|v\| \le \alpha \}$$

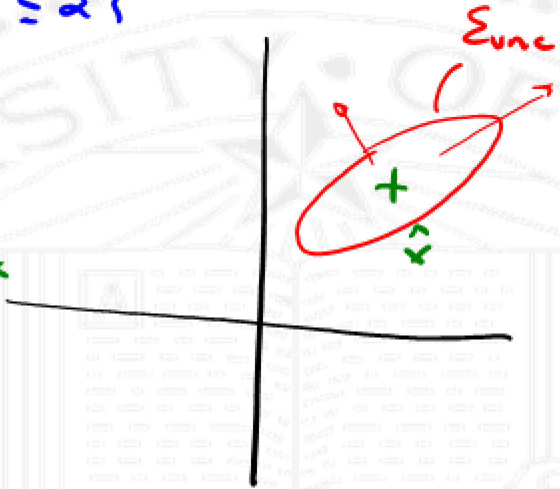
SVD in estimation/inversion (2.3)

$\tilde{x} \in \mathcal{E}_{unc} : \{ Bv \mid \|v\| \le \alpha \}$

$x = \hat{x} - \tilde{x} = \hat{x} + \underline{\mathcal{E}_{unc}}$

True x lies within the $\mathcal{E}_{unc}$ centered @ $\hat{x}$.





iRain

SVD in estimation/inversion (3.3)

Semi-major axes of Σ_{unc} $\quad \propto \sigma_i v_i$ singular value and right singular vector of B

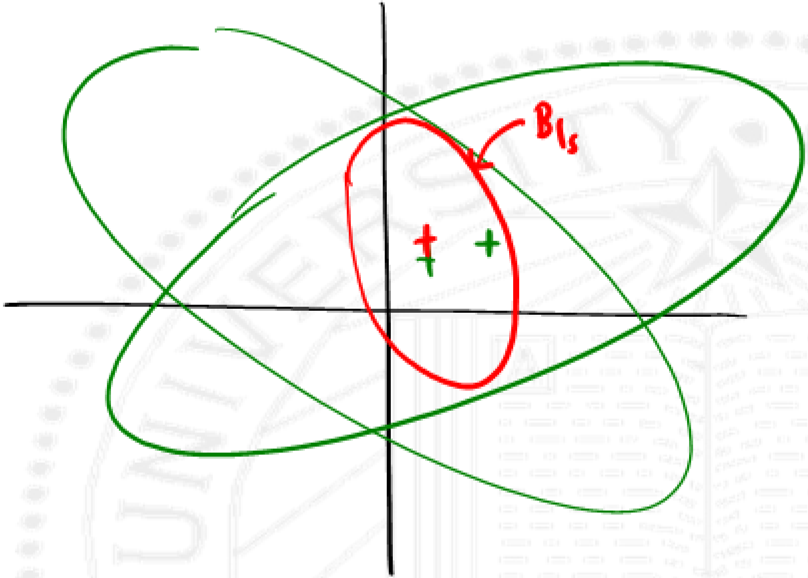
$\|\hat{x} - x\| \leq \alpha \|B\|$ $\qquad BA = I$

$B_{ls} \doteq A^{\dagger}$ $\qquad B_{ls} B_{ls}^T \leq BB^T$

$\sigma_i(B_{ls}) \leq \sigma_i(B) \qquad i = 1 \ldots n$

$\|B_{ls}\| \leq \|B\| \longrightarrow \Sigma_{ls} \subseteq \Sigma_{unc_B}$

Example (1.3)

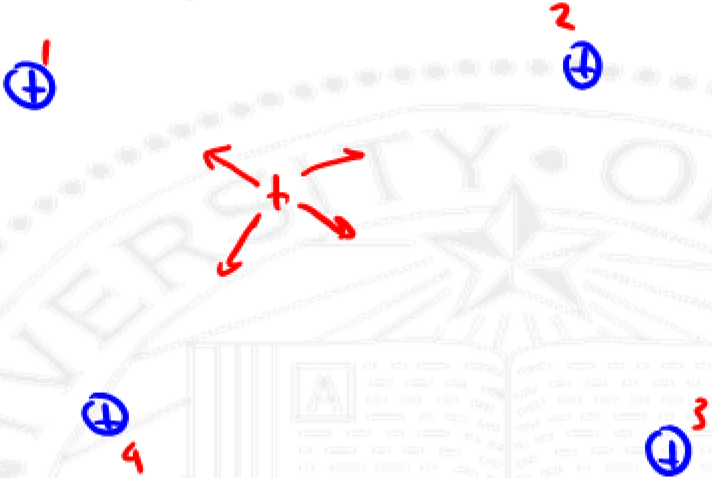


Example (2.3)

$y_i = -k_i^T x + v_i$

$A_1 = \begin{bmatrix} -k_1^T \\ -k_2^T \end{bmatrix}$

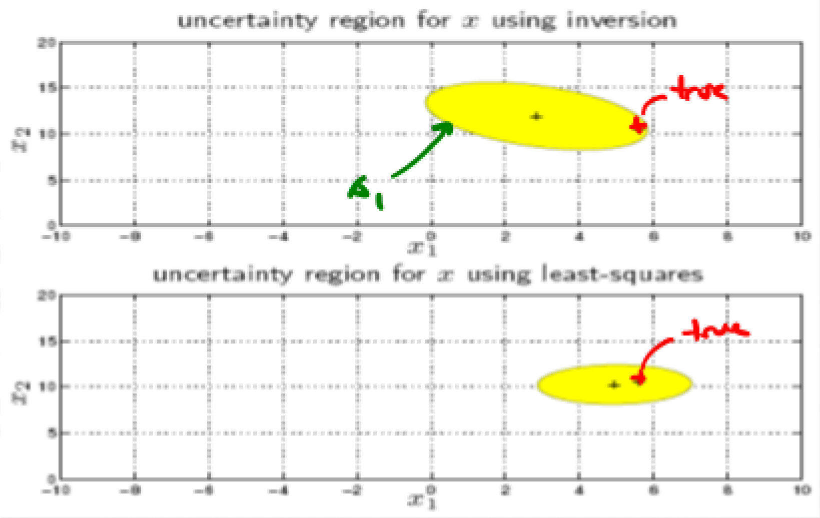
$A_2 = -\begin{bmatrix} k_1^T \\ \vdots \\ k_4^T \end{bmatrix}$



$\alpha = 1$

Example (3.3)

only the first two beacons



LS solution

Proof of Optimality Property (1.3)

$A \in \mathbb{R}^{m \times n}$ $\quad m > n \quad$ full rank

$$A = U \underset{n \times n}{\Sigma} V^T$$

$$B_{ls} = A^{\dagger} = V \Sigma^{-1} U^T$$

$$B_{ls} \cdot A = V \Sigma^{-1} U^T U \Sigma V^T = I \; ✓$$

$$Z \triangleq B - B_{ls} \quad \therefore \quad B = B_{ls} + Z \qquad BA = \overbrace{B_{ls} A}^{} + \overbrace{ZA}^{} = I$$

$$ZA = ZU\Sigma V^T = \{0\} \qquad [ZU\Sigma V^T] V \Sigma^{-1} = 0 \quad \therefore \quad ZU = \{0\}$$

$$Z \in N(U) \longrightarrow Z = \{0\}.$$

smallest [illegible] Gain [illegible]

Proof of Optimality Property (2.3)

$BB^T = (B_k + Z)(B_k + Z)^T$

$= B_k B_k^T + B_k Z^T + Z B_k^T + ZZ^T$

$= B_k B_k^T + ZZ^T$

$Z B_k^T = (ZU) \Sigma' V^T = 0$

$BB^T = B_k B_k^T + ZZ^T \rightarrow B_k B_k^T \leq B_k B_k^T + ZZ^T$

Proof of Optimality Property (3.3)

Sensitivity of Linear Equations to Data Error (1.3)

$y = Ax \qquad A \in \mathbb{R}^{n \times n} \qquad A^{-1}$ exists

$x = A^{-1} y \longrightarrow (x + \delta x) = A^{-1}(y - \delta y)$

$A = U \Sigma V^T$

$A^{-1} = V \Sigma^{-1} U^T = \underbrace{VR}_{\tilde{V}} \tilde{\Sigma} \underbrace{RU^T}_{\tilde{U}^T}$

$R = \begin{bmatrix} & & 1 \\ & ⋰ & \\ 1 & & \end{bmatrix}$

$\tilde{\Sigma} = \begin{bmatrix} 1/\sigma_n & & \\ & \ddots & \\ & & 1/\sigma_1 \end{bmatrix}$

Sensitivity of Linear Equations to Data Error (2.3)

$\|\delta x\| = \|A^{-1}\delta y\| \le \|A^{-1}\|\|\delta y\|$

if $\|A^{-1}\|$ is large:

- small errors in y can lead to big errors in x
- can't solve for x given y with small errors
- A may be full rank, but in PRACTICE is singular.

Sensitivity of Linear Equations to Data Error (3.3)

$y = Ax$ $\quad\|y\| \leq \|A\|\|x\|$

$$\frac{\|\delta x\|}{\|x\|} \leq \underbrace{\|A\|\|A^{-1}\|}\frac{\|\delta y\|}{\|y\|}$$

perturbation in x as %

perturbation in y as a %

$$\kappa(A) = \|A\|\|A^{-1}\| = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \geq 1 \quad \text{"cond"}$$

$$\text{rcond} \triangleq \frac{1}{\kappa(A)}$$

Relative error in solution $\leq K \cdot$ relative error in data.

Low rank approximations (1.3)

bits in solution ≃ # bits in data $- \log_2 \kappa$

κ small $\rightarrow 1$ A "well conditioned"

κ large $\rightarrow \infty$ A "ill conditioned"

$$\kappa(A) \triangleq \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

Low rank approximations (2.3)

$A \in \mathbb{R}^{m \times n}$ $m > n$ $\text{rank}(A) = r$

$$A = U \Sigma V^T = \sum_{i=1}^{r} \underbrace{\sigma_i u_i v_i^T}_{\text{rank 1 diadic expansion}}$$

$\hat{A} \rightarrow \text{rank}(\hat{A}) \leq p < r$ such that $\hat{A} \approx A$

$\min_{\hat{A}} \| A - \hat{A} \|$ $\| A \|_{FR} = \| A(:) \|$

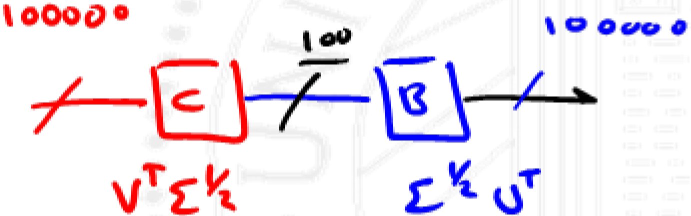
$$\hat{A} = \sum_{i=1}^{p} \sigma_i u_i v_i^T$$

Low rank approximations (3.3)

$$\|A-\hat{A}\| = \left\|\sum_{i=p+1}^{r} \sigma_i u_i v_i^T\right\| = \sigma_{p+1}$$

$u_i v_i^T$ — in order of importance

$$A = BC$$



Distance to Singularity (1.3)

σ_i :

$$\sigma_i = \min\{\|A - B\| \mid \mathrm{rank}(B) \le (i-1)\}$$

$\{15, 10, 3, 2, 1\times10^{-5}\}$ $y = Ax + v$ $\|v\| \le 0.01$

small σ_{min} means that A is close to singular

Distance to Singularity (2.3)

$y = \underbrace{Ax}_{\text{signal}} + \overbrace{v}^{\text{noise}}$

$A \in \mathbb{R}^{100 \times 30}$

$\Sigma = \{10, 7, 2, \tfrac{1}{2}, \tfrac{1}{100}, \tfrac{1}{1000} \ldots\}$

$\|x\|$ is on the order of 1

$\|v\|$ is on the order of 0.1

$\sigma_i u_i v_i^T = 0$ for $i = 5 \ldots 30$.

$$\boxed{y = \sum_{i=1}^{4} \sigma_i u_i v_i^T x + v}$$

Better model

Distance to Singularity (3.3)

$\{a_1 \ldots a_{100}\} \in \mathbb{R}^{10}$

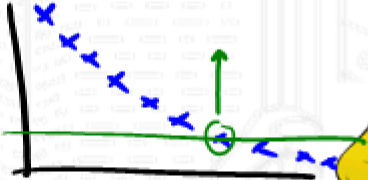
stock price data for last 100 days

$[a_1 \cdots a_{100}] = A \in \mathbb{R}^{10 \times 100}$ $QR \to 10$.

svd (A) → 3 significant singular values

$\log(\sigma_i)$

academia —



Questions?