Schur Complement (cont'd)

Previously, if $\mathbf{D}$ and the Schur complement $\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ are nonsingular,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

Now suppose $\mathbf{A}$ and the Schur complement $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are nonsingular. Likewise,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

Compare the above two expressions of $\mathbf{X}^{-1}$. If $\mathbf{A},\mathbf{D}$ and both Schur complements $\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}$, $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are nonsingular, then

$$(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$$

By setting $\mathbf{D}=\mathbf{I}$ and $\mathbf{B}'=-\mathbf{B}$, the above equation leads to the *matrix inversion lemma*

$$(\mathbf{A}+\mathbf{B}'\mathbf{C})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}'(\mathbf{I}+\mathbf{C}\mathbf{A}^{-1}\mathbf{B}')^{-1}\mathbf{C}\mathbf{A}^{-1}$$

Schur Complement of PSD Matrices

Let $\mathbf{X} \in \mathbb{S}^n$ and partition it as

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}$$

where $\mathbf{A} \in \mathbb{S}^n$ and $\mathbf{C} \in \mathbb{S}^{n-m}$

The Schur complements are $\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T$ and $\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$

Properties:

$\succeq \qquad \succeq \Longleftrightarrow \succ \qquad \succeq$

- With nonsingular $\mathbf{C}$, $\mathbf{X} \succ \mathbf{0} \Longleftrightarrow \mathbf{C} \succ \mathbf{0}$ and $\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T \succ \mathbf{0}$
- With $\mathbf{C} \succ \mathbf{0}$, $\mathbf{X} \succeq \mathbf{0} \Longleftrightarrow \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T \succeq \mathbf{0}$
- With nonsingular $\mathbf{A}$, $\mathbf{X} \succ \mathbf{0} \Longleftrightarrow \mathbf{A} \succ \mathbf{0}$ and $\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} \succ \mathbf{0}$
- With $\mathbf{A} \succ \mathbf{0}$, $\mathbf{X} \succeq \mathbf{0} \Longleftrightarrow \mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} \succeq \mathbf{0}$

Example: For any $\mathbf{b} \in \mathbb{R}^n$ and any symmetric and PD $\mathbf{C}$,

$X = \begin{bmatrix} 1 & b^T \\ b & C \end{bmatrix}$

$$1 - \mathbf{b}^T\mathbf{C}^{-1}\mathbf{b} \geq 0 \iff \mathbf{C} - \mathbf{b}\mathbf{b}^T \succeq \mathbf{0}$$

Schur complement of C $\Updownarrow$ Schur complement of 1

$X \succeq 0$

Important Facts for Proving the Properties of Schur Complement

Let $\mathbf{Y} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^{-1}\mathbf{B}^T & \mathbf{I} \end{bmatrix}$, which is nonsingular. Then consider $\mathbf{Y}^T\mathbf{X}\mathbf{Y}$

According to Property 3 on Page 14, Sec 5.1,

$$X \underset{\succeq}{\succ} 0 \iff Y^T X Y \underset{\succeq 0}{\succ 0}$$

$$Y^T X Y = \begin{bmatrix} I & -BC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & 0 \\ -C^{-1}B^T & I \end{bmatrix}$$

$$= \begin{bmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{bmatrix}$$ block diagonal

The eigenvalues of $Y^T X Y$ = the eigenvalues of $A - BC^{-1}B^T$ and the eigenvalues of C

Matrix Computations
Chapter 6: Singular value, Singular Vectors, and Singular Value Decomposition
Section 6.1 Singular Value Decomposition

Jie Lu
ShanghaiTech University

Singular Value Decomposition

Theorem

Given any $\mathbf{A} \in \mathbb{R}^{m \times n}$*, there exists a 3-tuple* $(\mathbf{U}, \Sigma, \mathbf{V}) \in \mathbb{R}^{m \times m} \times \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times n}$ *s.t.*

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$$

where $\mathbf{U}$ *and* $\mathbf{V}$ *are orthogonal matrices and*

$$[\Sigma]_{ij} = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j \end{cases} \quad \textit{with } \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0, \; p = \min\{m, n\}$$

$p = n$

$\sigma_1, \ldots, \sigma_n$ may have zero entry

$m > n$

$\Sigma = \begin{bmatrix} \diagdown \\ 0 \end{bmatrix}$

$m < n$

$\Sigma = \begin{bmatrix} \diagdown & 0 \end{bmatrix}$

$\sigma_1, \ldots, \sigma_m$

$p = m$

- The above decomposition is called the singular value decomposition (SVD)
- σ_i is called the ith singular value
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are called the ith left and right singular vectors, respectively
- Notations denoting singular values of $\mathbf{A}$:

$$\sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \ldots \geq \sigma_p(\mathbf{A}) = \sigma_{\min}(\mathbf{A})$$

Different Forms of SVD

- **Partitioned form**: Let r be the number of nonzero singular values, so that $\sigma_1 \geq \ldots \sigma_r > 0$, $\sigma_{r+1} = \ldots = \sigma_p = 0$. Then,

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \underbrace{\begin{bmatrix} \tilde{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\Sigma} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}$$

  where

  - $\tilde{\Sigma} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r)$
  - $\mathbf{U}_1 = [\ \mathbf{u}_1, \ldots, \mathbf{u}_r\ ] \in \mathbb{R}^{m \times r}$, $\mathbf{U}_2 = [\ \mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\ ] \in \mathbb{R}^{m \times (m-r)}$
  - $\mathbf{V}_1 = [\ \mathbf{v}_1, \ldots, \mathbf{v}_r\ ] \in \mathbb{R}^{n \times r}$, $\mathbf{V}_2 = [\ \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\ ] \in \mathbb{R}^{n \times (n-r)}$

- Thin SVD: $\mathbf{A} = \mathbf{U}_1 \tilde{\Sigma} \mathbf{V}_1^T$ $\quad$ U_1, V_1 semi-orthogonal

- **Outer-product form**: $\mathbf{A} = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

SVD and Eigendecomposition

Note from the SVD $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ that

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}_1\mathbf{U}^T,\ \mathbf{D}_1 = \Sigma\Sigma^T = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{m-p \text{ zeros}}) \qquad (*)$$

$$\mathbf{A}\mathbf{A}^T = U\Sigma \underbrace{V^T \cdot V}\Sigma^T U^T = U\Sigma\Sigma^T U^T$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}_2\mathbf{V}^T,\ \mathbf{D}_2 = \Sigma^T\Sigma = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{n-p \text{ zeros}}) \qquad (**)$$

If $\Sigma = \begin{bmatrix} \boxed{\tilde{D}} & 0 \end{bmatrix}$, $\Sigma^T = \begin{bmatrix} \boxed{\tilde{D}} \\ 0 \end{bmatrix}$. $\Sigma\Sigma^T = \tilde{D}^2 = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{bmatrix}$ $m=p$

$\Sigma^T\Sigma = \begin{bmatrix} \tilde{D}^2 & 0 \\ 0 & 0 \end{bmatrix}$

Observations:

- $(*)$ is an eigendecomposition of $\mathbf{A}\mathbf{A}^T$
- $(**)$ is an eigendecomposition of $\mathbf{A}^T\mathbf{A}$
- The left singular vector matrix $\mathbf{U}$ of $\mathbf{A}$ is the eigenvector matrix of $\mathbf{A}\mathbf{A}^T$
- The right singular vector matrix $\mathbf{V}$ of $\mathbf{A}$ is the eigenvector matrix of $\mathbf{A}^T\mathbf{A}$
- $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$

Insights of the Proof of SVD

- To see the insights of the constructive proof, consider the special case of square nonsingular $\mathbf{A}$

- $\mathbf{A}\mathbf{A}^T$ is PD with eigendecomposition

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\Lambda\mathbf{U}^T, \qquad \lambda_1 \geq \ldots \geq \lambda_n > 0$$

- Let $\Sigma = \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m})$ and $\mathbf{V} = \mathbf{A}^T\mathbf{U}\Sigma^{-1}$

- It can be verified that $\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{A}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

Proof of SVD

Let $A \in \mathbb{R}^{m\times n}$.

Since AA^T is symmetric and PSD, it has eigenvalues

$\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_m$ and eigendecomposition

$$AA^T = U\Lambda U^T = [U_1 \quad U_2]\begin{bmatrix} \tilde{\Lambda} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = U_1\tilde{\Lambda}U_1^T \quad (*)$$

where $\tilde{\Lambda} = \mathrm{Diag}(\lambda_1, \cdots, \lambda_r)$

Note that

$$(U_2^T A)(U_2^T A)^T = U_2^T \underbrace{AA^T}_{(*)} U_2 = U_2^T U_1 \tilde{\Lambda} \underbrace{U_1^T U_2}_{=0} = 0$$

$$\Rightarrow U_2^T A = 0 \quad (**)$$

Proof of SVD (cont'd)

Let $\tilde{\Sigma} = \tilde{\Lambda}^{\frac{1}{2}} = \mathrm{Diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ and $V_1 = A^T U_1 \tilde{\Sigma}^{-1} \in \mathbb{R}^{n\times r}$

(similar to the special case)

$$V_1^T V_1 = \tilde{\Sigma}^{-1} U_1^T \underbrace{A A^T}_{(\star)} U_1 \tilde{\Sigma}^{-1} = \tilde{\Sigma}^{-1} \underbrace{U_1^T U_1} \tilde{\Lambda} \underbrace{U_1^T U_1} \tilde{\Sigma}^{-1}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\lambda_r}} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\lambda_r}} \end{bmatrix} = I$$

Let $V_2 \in \mathbb{R}^{n\times(n-r)}$ be s.t. $V = [V_1 \ \ V_2]$ is orthogonal.

Then, $U_1^T A V_1 = U_1^T \underbrace{A A^T}_{(\star)} U_1 \tilde{\Sigma}^{-1} = \underbrace{U_1^T U_1} \tilde{\Lambda} \underbrace{U_1^T U_1} \tilde{\Sigma}^{-1} = \tilde{\Sigma}$ $(\star)$

$U_1^T A V_2 = \tilde{\Sigma} \underbrace{\tilde{\Sigma}^{-1} U_1^T A}_{V_1^T} V_2 = 0$ $(\star\star)$

Proof of SVD (cont'd)

Finally,

$$U^T A V = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} \underbrace{U_1^T A V_1}_{(*)} & \underbrace{U_1^T A V_2}_{(**)} \\ \underbrace{U_2^T A V_1}_{(**)} & \underbrace{U_2^T A V_2}_{(**)} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma}$$

$$\Longleftrightarrow \quad A = U \Sigma V^T$$

SVD and Subspaces

$A = U\Sigma V^T$

$A^T = V\Sigma^T U^T$

SVD of A^T

Properties:

(a) $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U}_1)$, $\mathcal{R}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{U}_2)$

(b) $\mathcal{R}(\mathbf{A}^T) = \mathcal{R}(\mathbf{V}_1)$, $\mathcal{R}(\mathbf{A}^T)^\perp = \mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{V}_2)$

(c) $\mathrm{rank}(\mathbf{A}) = r$ (the number of nonzero singular values)

the number of nonzero eigenvalues of AA^T and A^TA

$\mathrm{rank}(A) = \dim \mathcal{R}(A) = \dim \mathcal{R}(U_1) = r$

- In practice, SVD can be used a numerical tool for computing bases of $\mathcal{R}(\mathbf{A})$, $\mathcal{R}(\mathbf{A})^\perp$, $\mathcal{R}(\mathbf{A}^T)$, $\mathcal{N}(\mathbf{A})$

- Using SVD, we can easily show the following facts:
  - $\mathrm{rank}(\mathbf{A}^T) = \mathrm{rank}(\mathbf{A})$
  - $\dim \mathcal{N}(\mathbf{A}) = n - \mathrm{rank}(\mathbf{A})$

Matrix Computations
Chapter 6: Singular value, Singular Vectors, and Singular Value Decomposition

Section 6.2 Matrix Norms

Jie Lu
ShanghaiTech University

Matrix Norms

Definition: A function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a matrix norm if (i) $f(\mathbf{A}) \geq 0$ for all $\mathbf{A}$; (ii) $f(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$; (iii) $f(\mathbf{A}+\mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B})$ for any $\mathbf{A}, \mathbf{B}$; (iv) $f(\alpha\mathbf{A}) = |\alpha| f(\mathbf{A})$ for any $\mathbf{A}$ and any scalar α

- For example, the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = [\mathrm{tr}(\mathbf{A}^T\mathbf{A})]^{1/2}$ is a norm
- Induced norm or operator norm: the function

$$f(\mathbf{A}) = \max_{\|\mathbf{x}\|_\beta \leq 1} \|\mathbf{A}\mathbf{x}\|_\alpha$$

  where $\|\cdot\|_\alpha, \|\cdot\|_\beta$ denote any vector norms
- Matrix norms induced by the vector p-norm ($p \geq 1$):

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p \leq 1} \|\mathbf{A}\mathbf{x}\|_p$$

  - $\|\mathbf{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |a_{ij}|$
  - $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |a_{ij}|$
  - $\|\mathbf{A}\|_2 =?$

Matrix 2-Norm

The Matrix 2-norm or spectral norm is given by

$$\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A}) = \sqrt{\lambda_{max}(AA^T)} = \sqrt{\lambda_{max}(A^TA)}$$

Prove this using SVD $A = U\Sigma V^T$

For any $x \in \mathbb{R}^n$ with $\|x\|_2 \le 1$,

$$\|Ax\|_2^2 = \|U\Sigma V^T x\|_2^2 \underset{\substack{\uparrow \\ U \text{ orthogonal}}}{=} \|\Sigma(V^Tx)\|_2^2 \le \sigma_1^2 \|V^Tx\|_2^2$$

$$\underset{\substack{\uparrow \\ V^T \text{ orthogonal}}}{=} \sigma_1^2 \|x\|_2^2 \le \sigma_1^2$$

where the equality holds when $x = v_1$

Implication to linear systems: Let $\mathbf{y} = \mathbf{Ax}$ be a linear system. Under the input energy constraint $\|\mathbf{x}\|_2 \le 1$, the system output energy $\|\mathbf{y}\|_2^2$ is maximized when $\mathbf{x}$ is chosen as the 1st right singular vector

Corollary: $\min_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \sigma_{\min}(\mathbf{A})$ if $m \ge n$

not matrix norm

Properties of Matrix 2-Norm

- $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$
  - In fact, $\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p$ for any $p \geq 1$
- $\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$
  - A special case of the first property
- $\|\mathbf{QAW}\|_2 = \|\mathbf{A}\|_2$ for any orthogonal $\mathbf{Q}, \mathbf{W}$
  - We also have $\|\mathbf{QAW}\|_F = \|\mathbf{A}\|_F$ for any orthogonal $\mathbf{Q}, \mathbf{W}$
- $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{p}\|\mathbf{A}\|_2$ (here $p = \min\{m, n\}$)

$$\|A\|_F^2 = tr(A^TA) \underset{A=U\Sigma V^T}{=} tr(V\Sigma^T\Sigma V^T) = \text{sum of eigenvalues of } \Sigma^T\Sigma$$

eigendecomposition of A^TA

$$= tr(\Sigma^T\Sigma) = \|\Sigma\|_F^2 = \underbrace{\sigma_1^2}_{=\|A\|_2^2} + \cdots + \sigma_p^2 \leq p\sigma_1^2$$

$$\geq \sigma_1^2$$

Schatten p-Norm

- The function

$$f(\mathbf{A}) = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})^p \right)^{1/p}, \qquad p \geq 1,$$

  is a matrix norm called the Schatten p-norm

- Nuclear norm:

$$\|\mathbf{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A}) = \mathrm{tr}(\sqrt{\mathbf{A}^T\mathbf{A}})$$

  - A special case of the Schatten p-norm
  - A way to prove the nuclear norm is a matrix norm:
    - Show that $f(\mathbf{A}) = \max_{\|\mathbf{B}\|_2 \leq 1} \mathrm{tr}(\mathbf{B}^T\mathbf{A})$ is a norm
    - Show that $f(\mathbf{A}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i$
  - Applications in rank approximation, e.g., for compressive sensing and matrix completion[1]

[1] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

Schatten p-Norm

- $\mathrm{rank}(\mathbf{A})$ is *nonconvex* in $\mathbf{A}$ and is arguably hard to do optimization with it
- **Idea**: The rank function can be expressed as

$$\mathrm{rank}(\mathbf{A}) = \sum_{i=1}^{\min\{m,n\}} \mathbb{1}\{\sigma_i(\mathbf{A}) \neq 0\},$$

true 1

false 0

and we may approximate it via

$$f(\mathbf{A}) = \sum_{i=1}^{\min\{m,n\}} \varphi(\sigma_i(\mathbf{A}))$$

for some friendly function φ

- Using $\varphi(z) = z$, $f(\mathbf{A})$ becomes the nuclear norm

$$\|\mathbf{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})$$

which is *convex* in $\mathbf{A}$