# Matrix Reconstruction & Approximation
## Part II: Nuclear Norm Relaxation

Calin Cruceru
ccruceru@inf.ethz.ch

March 20, 2020

ETH Zurich – cil.inf.ethz.ch

# Recall: Problem Statement

- **Given** a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with observed entries $\mathcal{I} \subseteq [m] \times [n]$
- **Goal**: fill in the unobserved entries
- **Assumption** ("learnability"): $\mathbf{A}$ can be explained by fewer than $m \times n$ parameters
  - $\exists \, (\mathbf{U} \in \mathbb{R}^{m \times k}, \mathbf{V} \in \mathbb{R}^{n \times k})$ such that $\mathbf{A} \approx \mathbf{U}\mathbf{V}^\top$

# Recall: Formalizations

- Low-rank matrix recovery

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}}^2 \qquad \textbf{(P1)}$$
$$\text{subject to } \operatorname{rank}(\mathbf{X}) \leqslant k$$

- Exact matrix recovery

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \operatorname{rank}(\mathbf{X}) \qquad \textbf{(P2)}$$
$$\text{subject to } \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}} = 0$$

- Both are NP-hard $\rightarrow$ require approximations!
  - Last time: re-parametrized **(P1)** as

$$\{\mathbf{X}\in\mathbb{R}^{m\times n} : \operatorname{rank}(\mathbf{X}) \leq k\} = \{\mathbf{U}\mathbf{V}^\top : \mathbf{U}\in\mathbb{R}^{m\times k}, \mathbf{V}\in\mathbb{R}^{n\times k}\}$$

and used Alternating Least Squares.

# Convex Relaxation of $\operatorname{rank}(\mathbf{X})$

- Tightest lower-bound

$$\operatorname{rank}(\mathbf{X}) \geqslant \|\mathbf{X}\|_* \quad \text{for } \|\mathbf{X}\|_2 \leqslant 1.$$

- Intuition: $L_1$ norm as the relaxation of the $L_0$ "norm"

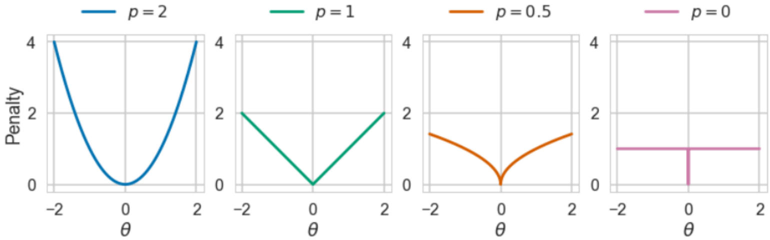


Figure: source: https://arxiv.org/abs/1712.01312

# Inequality Proof (Problem 1.1)

First, recall that multiplying a matrix by an invertible matrix maintains its rank,

$$\begin{aligned} \operatorname{rank}(\mathbf{XY}) = \operatorname{rank}(\mathbf{ZX}) &= \operatorname{rank}(\mathbf{X}) \\ \forall\ \mathbf{Y} \in \mathbb{R}^{n\times n},\ \ \operatorname{rank}(\mathbf{Y}) &= n \\ \forall\ \mathbf{Z} \in \mathbb{R}^{m\times m},\ \operatorname{rank}(\mathbf{Z}) &= m \end{aligned}$$

Then, with $\mathbf{A} = \mathbf{UDV}^\top$,

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{UDV}^\top) = \operatorname{rank}(\mathbf{D}) = \#\{\sigma_i > 0\}.$$

We also have that $\|\mathbf{A}\|_2 = \sigma_{\max}(\leqslant 1)$, so $\sigma_i \leqslant 1$ for all $i$. Thus,

$$\operatorname{rank}(\mathbf{A}) = \#\{\sigma_i > 0\} = \sum_{i:\sigma_i>0} 1 \geq \sum_{i:\sigma_i>0} \sigma_i = \sum_i \sigma_i = \|\mathbf{A}\|_*.$$

# Convexity Proof (Problem 1.2)

The **lazy** (and more insightful) way:

1. Show that $\|\mathbf{A}\|_*$ is a norm
   - (i) $\|a\mathbf{A}\|_* = |a|\|\mathbf{A}\|_*$ (absolute homogeneous) and (ii) $\|\mathbf{A}\|_* = 0 \iff \mathbf{A} = \mathbf{0}$ (positive definite) are trivial
   - Show triangle inequality! (iii) $\|\mathbf{A} + \mathbf{B}\|_* \leqslant \|\mathbf{A}\|_* + \|\mathbf{B}\|_*$
2. Show that all norms are convex,

$$\|\lambda x + (1-\lambda)y\| \leqslant \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\|,$$

by triangle inequality and absolute homogeneity.

The **definition**-based approach: show that

$$\|\lambda\mathbf{A} + (1-\lambda)\mathbf{B}\|_* \leqslant \lambda\|\mathbf{A}\|_* + (1-\lambda)\|\mathbf{B}\|_*$$

for all $\lambda \in [0, 1]$ and $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$.

# Convexity Proof (Problem 1.2)

**Goal**: Show that triangle inequality holds!

First, show that the nuclear and spectral norms are *dual*,[1]

$$\|\mathbf{A}\|_* = \sup_{\mathbf{X}:\|\mathbf{X}\|_2 \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle = \sup_{\sigma_1(\mathbf{X}) \leqslant 1} \operatorname{Tr} \mathbf{X}^\top \mathbf{A}.$$

Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$. We will show that

$$\sum_i \sigma_i \geqslant \sup_{\sigma_1(\mathbf{X}) \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle \quad \text{and} \quad \sum_i \sigma_i \leqslant \sup_{\sigma_1(\mathbf{X}) \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle.$$

We show the easier "$\geqslant$" inequality first:

$$\sup_{\sigma_1(\mathbf{X}) \leqslant 1} \operatorname{Tr} \mathbf{X}^\top \mathbf{A} \overset{\mathbf{X} \overset{!}{=} \mathbf{U}\mathbf{V}^\top}{\geqslant} \operatorname{Tr} \mathbf{V}\mathbf{U}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top = \operatorname{Tr} \mathbf{D} = \sum_i \sigma_i.$$

---

[1] https://en.wikipedia.org/wiki/Dual_norm

# Convexity Proof (Problem 1.2) - *cont.(1)*

We now prove the "$\leqslant$" inequality:

$$
\begin{aligned}
\sup_{\sigma_1(\mathbf{X})\leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle &= \sup_{\sigma_1(\mathbf{X})\leqslant 1} \operatorname{Tr} \mathbf{X}^\top \mathbf{U}\mathbf{D}\mathbf{V}^\top && \text{(matrix inner product)} \\
&= \sup_{\sigma_1(\mathbf{X})\leqslant 1} \operatorname{Tr} \mathbf{D}\mathbf{V}^\top \mathbf{X}^\top \mathbf{U} && \text{(trace cyclicity)} \\
&= \sup_{\sigma_1(\mathbf{X})\leqslant 1} \sum_i \sigma_i \mathbf{u}_i^\top \mathbf{X} \mathbf{v}_i && \text{(trace def., diag. } \mathbf{D}\text{)} \\
&\leqslant \sup_{\sigma_1(\mathbf{X})\leqslant 1} \max_{\|\mathbf{u}\|=\|\mathbf{v}\|=1} \sum_i \sigma_i \mathbf{u}^\top \mathbf{X} \mathbf{v} && \text{(upper bound)} \\
&= \sup_{\sigma_1(\mathbf{X})\leqslant 1} \sum_i \sigma_i \sigma_1(\mathbf{X}) && \text{(largest sing. val.)} \\
&= \left( \sum_i \sigma_i \right) \sup_{\sigma_1(\mathbf{X})\leqslant 1} \sigma_1(\mathbf{X}) && \text{(factor out const.)} \\
&= \sum_i \sigma_i .
\end{aligned}
$$

# Convexity Proof (Problem 1.2) - *cont.(2)*

Hence, we have

$$\|\mathbf{X}\|_* = \sup_{\mathbf{X}:\|\mathbf{X}\|_2 \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle = \sum_i \sigma_i(\mathbf{A}).$$

Then,

$$\begin{aligned}
\|\mathbf{A}+\mathbf{B}\|_* &= \sup_{\sigma_1(\mathbf{X} \leqslant 1} \langle \mathbf{X}, \mathbf{A}+\mathbf{B} \rangle \\
&= \sup_{\sigma_1(\mathbf{X} \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle + \langle \mathbf{X}, \mathbf{B} \rangle \\
&\leqslant \sup_{\sigma_1(\mathbf{X} \leqslant 1} \langle \mathbf{X}, \mathbf{A} \rangle + \sup_{\sigma_1(\mathbf{X} \leqslant 1} \langle \mathbf{X}, \mathbf{B} \rangle \\
&= \|\mathbf{A}\|_* + \|\mathbf{B}\|_*.
\end{aligned}$$

Therefore, $\|\mathbf{A}\|_*$ respects triangle inequality, so it is a norm. Hence it is convex.

# Relaxed Optimization Problems

- We have the following convex relaxation of **(P2)**

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \qquad \textbf{(P2r)}$$
$$\text{subject to } \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}} = 0$$

- ... and of **(P1)**

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}}^2 \quad \textbf{(P1r)}$$
$$\text{subject to } \|\mathbf{X}\|_* \leqslant k$$

- Note the following:
  - the set $\{\mathbf{X}\in\mathbb{R}^{m\times n} : \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}} = 0\}$ is a convex set (check it!)
  - the function $g(\mathbf{X}) = \|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}}^2$ is convex
    - hint: show that $\|\mathbf{X}\|_{\mathcal{I}}$ is a seminorm ($\|\mathbf{X}\|_{\mathcal{I}} = 0$ does not imply $\mathbf{X} = \mathbf{0}$; it does not influence the convexity proof on slide 6)

# Two Further Questions

1. **(Q1)** How well do the relaxed versions approximate the solutions to the original problems?
2. **(Q2)** Can we speed up the minimization of the relaxed formulations?

# Theoretical Guarantees for Nuclear Norm Minimization (Q1)

- For *most* matrices **A** of rank $k$, a minimum $\mathbf{X}^*$ of **(P2r)** perfectly recovers **A** provided that the number of observed entries satifies

$$|\mathcal{I}| \geqslant Ckn^{6/5} \log n.$$

  - Note: $\mathbf{X}^*$ is not necessarily low-rank (empirically, it is).
  - *"Exact Matrix Completion via Convex Optimization"*, Candes & Recht, https://arxiv.org/abs/0805.4471

- Later improved to

$$|\mathcal{I}| \geqslant C\mu^4 k^2 n \log^2 n,$$

  with a new assumption on the incoherence parameter $\mu$.
  - *"The Power of Convex Relaxation: Near-Optimal Matrix Completion"*, Candes & Tao, https://arxiv.org/abs/0903.1476

# Efficient Implementation (Q2)

- The shrinkage operator

$$\mathcal{D}_\tau(\mathbf{Y}) = \arg\min_{\mathbf{X}} \left\{ \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \tau\|\mathbf{X}\|_* \right\}$$

  has the closed-form solution

$$\mathcal{D}_\tau(\mathbf{Y}) = \mathbf{U}\mathcal{D}_\tau(\mathbf{D})\mathbf{V}^\top, \quad \text{with } \mathcal{D}_\tau(\mathbf{D}) = \mathrm{diag}\{(\sigma_i - \tau)_+\}.$$

- The *SVD Thresholding* algorithm:
  1. Input: $\mathbf{A}$ with observed entries in $\Omega$
  2. Initialize $\mathbf{Y}_0 = \mathbf{0}$
  3. For $k = 1, 2, \ldots, K$ do:
     3.1 $\mathbf{X}^{(k)} = \mathcal{D}_\tau(\mathbf{Y}^{(k-1)})$
     3.2 $\mathbf{Y}^{(k)} = \mathbf{Y}^{(k-1)} + \delta_k \mathcal{P}_\Omega(\mathbf{A} - \mathbf{X}^{(k)})$
  4. Output: $\mathbf{X}^{(K)}$

# Efficient Implementation (Q2) - *cont.*

- Can show that the sequence $\{\mathbf{X}^{(k)}\}$ converges to the unique solution of

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2$$
$$\text{subject to } \mathcal{P}_\Omega(\mathbf{A}-\mathbf{X}) = \mathbf{0}$$

  - The constraint is the same as $\|\mathbf{A}-\mathbf{X}\|_{\mathcal{I}} = 0$ (prev. notation)
  - Notice the similarity to **(P2r)**
    - Not exactly the same, but more computationally efficient due to the sparsity of $\mathbf{Y}^{(k)}$ and the (empirically observed) low rank of $\mathbf{X}^{(k)}$.
- *"A Singular Value Thresholding Algorithm for Matrix Completion"*, Candes et. al., https://arxiv.org/abs/0810.3286