Understanding the uniqueness of the solution of the nuclear norm minimization

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: August 24, 2020
Last update : August 26, 2020

Nuclear norm minimization (NNM)

- Given a linear operator $\mathcal{A}$ and a vector $\mathbf{b}$, find a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ by solving the NNM

$$(\mathcal{P}) \ : \ \min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \|\mathbf{X}\|_* \ \text{ s.t. } \ \mathcal{A}(\mathbf{X}) = \mathbf{b}.$$

- This document: understand the uniqueness of the solution of $(\mathcal{P})$.

Prerequisites / things need to know

- Singular value decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{X}$.
- Property of convex function: if f is convex, for any $\mathbf{x}, \mathbf{y} \in \mathrm{dom} f$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle.$$

- The subgradient of nuclear norm.
- Dual characterization of nuclear norm / nuclear norm is dual to the operator norm.

The results generalizes to matrix of complex values, but for simplicity we we stick with matrix with real values.

Nuclear norm minimization for Matrix Completion (MC)

- The constraint $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ represents a general linear constraint on $\mathbf{X}$.
- The problem of Matrix Completion (MC) is a special case of NNM, by specifying the constraint, $(\mathcal{P})$ becomes

$$(\mathcal{P}') \;:\; \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \;\text{ s.t. }\; \mathbf{X}_{ij} = \mathbf{M}_{ij}, \quad (i,j) \in \Omega,$$

  where $\mathbf{M}$ is a partially-observed matrix with indices labeled by the set $\Omega = [1, 2, \ldots, m] \times [1, 2, \ldots, n]$.
- Comparing $(\mathcal{P})$ and $(\mathcal{P}')$, the linear operator $\mathcal{A}$ becomes an identity operator for the elements $(i,j) \in \Omega$, and the vector $\mathbf{b}$ becomes the partially-observed matrix $\mathbf{M}$.
- We will study the uniqueness condition of $(\mathcal{P})$ instead of $(\mathcal{P}')$ since it is more general.

Uniqueness of the NNM

$$(\mathcal{P}) \ : \ \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \ \text{ s.t. } \ \mathcal{A}(\mathbf{X}) = \mathbf{b}.$$

- **Theorem** The matrix $\mathbf{X}_0$ is the unique minimizer of $(\mathcal{P})$ if
  - It is feasible: it satisfies the constraint $\mathcal{A}(\mathbf{X}_0) = \mathbf{b}$.
    - For MC, this translates to $\mathbf{X}_0(i,j) = \mathbf{M}(i,j)$ for all $(i,j) \in \Omega$.
  - The linear operator $\mathcal{A}$ restricted to elements in T is injective, where T is a linear space defined as

$$T = \Big\{\mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top, \mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}\Big\}.$$

    - For MC, this translates to Π_Ω restricted to the elements in T is injective, where Π denotes projection.
  - There exists a dual matrix $\mathbf{P} \in \mathbb{R}^{m\times n}$ such that $\Pi_T\Big(\mathcal{A}^*(\mathbf{P})\Big) = \mathbf{U}\mathbf{V}^\top$ and $\|\Pi_{T^\perp}\Big(\mathcal{A}^*(\mathbf{P})\Big)\|_2 < 1$, where Π denotes the projection operator, $T^\perp$ is the complement of T and $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V} = \mathbf{X}_0$ is the SVD of $\mathbf{X}_0$.

The complications for understanding

- Aft first glance, the following are not directly approachable:
  - What is such $T = \left\{\mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top, \mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}\right\}$?
  - Why the matrix $\mathbf{P}$ is called dual matrix?
  - Why $\Pi_T\Big(\mathcal{A}^*(\mathbf{P})\Big) = \mathbf{U}\mathbf{V}^\top$ and $\|\Pi_{T^\perp}\Big(\mathcal{A}^*(\mathbf{P})\Big)\|_2 < 1$?

  These are related to the "derivative" of $\|\cdot\|_*$.

- Before go to the derivative of $\|\cdot\|_*$, we first understand why we need to know it: this brings us to the proof strategy of the uniqueness of NNM.

Proof of uniqueness by contradiction

- To show $\mathbf{X}_0$ is the unique minimizer to

$$(\mathcal{P}) \ : \ \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \|\mathbf{X}\|_* \ \text{ s.t. } \ \mathcal{A}(\mathbf{X}) = \mathbf{b},$$

  we first assume there exists another another solution $\mathbf{X}_1 \neq \mathbf{X}_0$ to the problem.

- Then we show that

$$\|\mathbf{X}_0\|_* \neq \|\mathbf{X}_1\|_*.$$

  This inequality contradicts to the assumption that both $\mathbf{X}_0$ and $\mathbf{X}_1$ are the optimal solutions to $(\mathcal{P})$, hence the assumption is false: the minimizer has to be unique.

- For the purpose of contradiction, we let $\mathbf{X}_1 = \mathbf{X}_0 + \Delta$, where $\Delta \neq 0$ is a perturbation added to $\mathbf{X}_0$.

- Then now we have to understand Δ.

About the matrix Δ

- As both $\mathbf{X}_0$ and $\mathbf{X}_1$ are feasible points to $(\mathcal{P})$, they satisfy the constraint $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. This gives

$$\mathcal{A}(\mathbf{X}_1) = \mathcal{A}(\mathbf{X}_0 + \Delta) = \mathcal{A}(\mathbf{X}_0) + \mathcal{A}(\Delta) = \mathbf{b} \implies \mathcal{A}(\Delta) = 0.$$

  In other words, Δ has to be inside the null space of $\mathcal{A}$

$$\Delta \in \mathsf{Null}(\mathcal{A}). \tag{1}$$

- For MC with the constraint

$$\mathbf{X}_{ij} \quad = \quad \mathbf{M}_{ij}, \quad (i,j) \in \Omega,$$

  then we have $\Delta_{ij} = 0, \quad (i,j) \in \Omega$, or equivalently $\Pi_\Omega(\Delta) = 0$.

Convexity of $\|\cdot\|_*$

- Recall our goal is to show $\|\mathbf{X}_0\|_* \neq \|\mathbf{X}_0 + \Delta\|_*$. A way to show it is to make use of the facts surrounding the convexity of $\|\cdot\|_*$:
  1. $\|\cdot\|_*$ is a norm, and it is a convex function.
  2. A convex function f satisfies $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$ for any $\mathbf{x}, \mathbf{y} \in$dom$f$. If f is not differentiable, $\nabla f(\mathbf{x})$ is replaced by subgradient.
- We now see a proof strategy: as $\|\cdot\|_*$ is convex, so

$$\|\mathbf{Y}\|_* \geq \|\mathbf{X}\|_* + \langle \nabla \|\mathbf{X}\|_*, \mathbf{Y} - \mathbf{X} \rangle,$$

  where $\nabla\|\mathbf{X}\|_*$ has to be replaced by the subgradient (see next slide). Put $\mathbf{Y} = \mathbf{X}_1$ and $\mathbf{X} = \mathbf{X}_0$ gives

$$\|\mathbf{X}_0 + \Delta\|_* \geq \|\mathbf{X}_0\|_* + \langle \nabla \|\mathbf{X}_0\|_*, \Delta \rangle. \tag{2}$$

- This brings us to the topic of the subgradient of $\|\cdot\|_*$.

Subgradient of $\|\cdot\|_*$

- As $\|\cdot\|_*$ is non-differentiable, we need subgradient instead of gradient.
- The subdifferential (the set of subgradient) of $\|\cdot\|_*$ at a point $\mathbf{M} \in \mathbb{R}^{m\times n}$ is the set

$$\partial\|\mathbf{M}\|_* = \left\{\mathbf{U}\mathbf{V}^\top + \mathbf{W} \;\middle|\; \|\mathbf{W}\|_2 \le 1, \mathbf{U}^\top\mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0\right\}, \quad (3)$$

  where $\mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{M}$ is the SVD of $\mathbf{M}$.

  That is, for any matrix $\mathbf{W}$, as long as it satisfies the conditions $\|\mathbf{W}\|_2 \le 1, \mathbf{U}^\top\mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0$, then the matrix $\mathbf{U}\mathbf{V}^\top + \mathbf{W}$ is a subgradient of $\|\cdot\|_*$ at the point $\mathbf{M}$.

The starting inequality of the proof

- Using the subgradient of $\|\cdot\|_*$, Inequality (2) becomes

$$\|\mathbf{X}_0 + \Delta\|_* \geq \|\mathbf{X}_0\|_* + \langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \Delta \rangle, \tag{4}$$

  where $\mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{X}_0$ and $\mathbf{W}$ is inside the set

$$\Big\{ \mathbf{Z} \mid \|\mathbf{Z}\|_2 \leq 1, \mathbf{U}^\top \mathbf{Z} = 0, \mathbf{Z}\mathbf{V} = 0 \Big\}.$$

- As our goal is to show $\|\mathbf{X}_0 + \Delta\|_* \neq \|\mathbf{X}_0\|_*$, so we want to show

$$\langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \Delta \rangle \neq 0.$$

  To proceed, we need some tools to deal with the inner product $\langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \Delta \rangle$. This term relates to the subgradient of $\|\cdot\|_*$, so this leads us go back to the subgradient of $\|\cdot\|_*$.

- For the matrix $\mathbf{U}\mathbf{V}^\top + \mathbf{W}$, the matrix $\mathbf{U}\mathbf{V}^\top$ is fix and the matrix $\mathbf{W}$ is a free variable, so we need to look at the subgradient from the perspective of this free variable.

Equivalent expressions of the subgradient of $\|\cdot\|_*$

- Let $\mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{M}$, recall the subdifferential of $\|\cdot\|_*$ at a point $\mathbf{M}$ is

$$\partial\|\mathbf{M}\|_* = \left\{\mathbf{U}\mathbf{V}^\top + \mathbf{W} \;\middle|\; \|\mathbf{W}\|_2 \leq 1, \mathbf{U}^\top\mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0\right\}.$$

- The above set can be equivalently expressed as

$$\partial\|\mathbf{M}\|_* = \left\{\mathbf{Z} \;\middle|\; \Pi_T(\mathbf{Z}) = \mathbf{U}\mathbf{V}^\top, \|\Pi_{T^\perp}(\mathbf{Z})\|_2 \leq 1\right\},$$

  with T is the linear space defined as

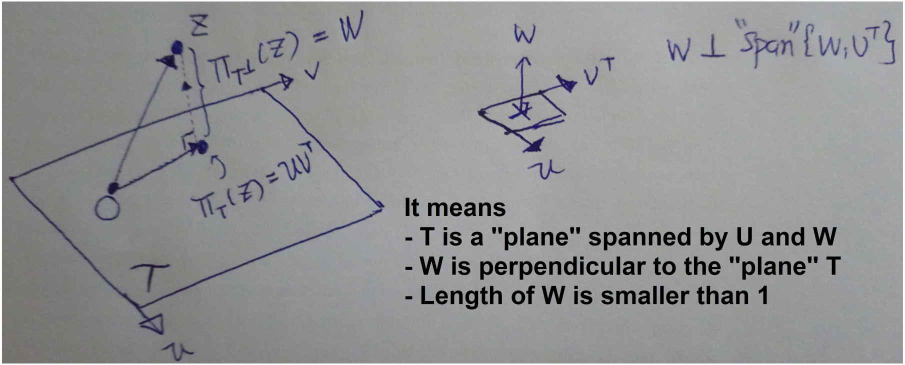
$$T = \left\{\mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top, \mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}\right\}.$$

  That is, for any matrix $\mathbf{Z}$, if it satisfies $\Pi_T(\mathbf{Z}) = \mathbf{U}\mathbf{V}^\top, \|\Pi_{T^\perp}(\mathbf{Z})\|_2 \leq 1$, then it is a subgradient of $\|\cdot\|_*$ at the point $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$.

- We show the equivalence using the dual characterization of $\|\cdot\|_*$.

The geometry of the subgradient of $\|\cdot\|_*$

$$
\begin{aligned}
\partial\|\mathbf{M}\|_* &= \left\{\mathbf{U}\mathbf{V}^\top + \mathbf{W} \;\middle|\; \|\mathbf{W}\|_2 \le 1, \mathbf{U}^\top\mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0\right\} \\
&= \left\{\mathbf{Z} \;\middle|\; \Pi_T(\mathbf{Z}) = \mathbf{U}\mathbf{V}^\top, \|\Pi_{T^\perp}(\mathbf{Z})\|_2 \le 1\right\}, \\
T &= \left\{\mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top, \mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}\right\}.
\end{aligned}
$$



Dual of $\| \cdot \|_*$ ... (1/3)

- The dual of nuclear norm is the operator norm/ 2-norm

$$\|\mathbf{X}\|_* = \sup_{\|\mathbf{Y}\|_2 \leq 1} \langle \mathbf{Y}, \mathbf{X} \rangle$$

- To find $\mathbf{Y}$ that maximizes the inner product $\langle \mathbf{Y}, \mathbf{X} \rangle$ such that $\|\mathbf{Y}\|_2 \leq 1$, we perform orthogonal decomposition of $\mathbf{Y}$ as

$$\mathbf{Y} = \Pi_S(\mathbf{Y}) + \Pi_{S^\perp}(\mathbf{Y}).$$

  with S is some space.

- Based on the geometric understanding of the subgradient of nuclear norm, the space S has to be the linear space T. And then the inner product becomes

$$\begin{aligned} \langle \mathbf{Y}, \mathbf{X} \rangle &= \langle \Pi_T(\mathbf{Y}) + \Pi_{T^\perp}(\mathbf{Y}), \mathbf{X} \rangle \\ &= \langle \Pi_T(\mathbf{Y}), \mathbf{X} \rangle + \langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{X} \rangle \\ &= \langle \Pi_T(\mathbf{Y}), \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \rangle + \langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \rangle \end{aligned}$$

- Now we show that $\langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \rangle = 0$.

Dual of $\|\cdot\|_*$ ... (2/3)

- For $\mathbf{X} = \mathbf{U\Sigma V}^\top$, the projection operators are
  - $\Pi_{\mathbf{U}}$: the orthogonal projection to the column space of $\mathbf{X}_0$.
  - $\Pi_{\mathbf{V}}$: the orthogonal projection to the row space of $\mathbf{X}_0$.
  - $\Pi_T(\mathbf{Y}) = \Pi_{\mathbf{U}}\mathbf{Y} + \mathbf{Y}\Pi_{\mathbf{V}} - \Pi_{\mathbf{U}}\mathbf{Y}\Pi_{\mathbf{V}}$
  - $\Pi_{T^\perp}(\mathbf{Y}) = (\mathbf{I} - \Pi_{\mathbf{U}})\mathbf{Y}(\mathbf{I} - \Pi_{\mathbf{V}})$
- So $\left\langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{U}\Sigma\mathbf{V}^\top \right\rangle = \left\langle (\mathbf{I} - \Pi_{\mathbf{U}})\mathbf{Y}(\mathbf{I} - \Pi_{\mathbf{V}}), \mathbf{U}\Sigma\mathbf{V}^\top \right\rangle$.
- As $\left\langle \mathbf{A}, \mathbf{UB} \right\rangle = \left\langle \mathbf{U}^\top\mathbf{A}, \mathbf{B} \right\rangle$, so
  $\left\langle (\mathbf{I} - \Pi_{\mathbf{U}})\mathbf{Y}(\mathbf{I} - \Pi_{\mathbf{V}}), \mathbf{U}\Sigma\mathbf{V}^\top \right\rangle = \left\langle \mathbf{U}^\top(\mathbf{I} - \Pi_{\mathbf{U}})\mathbf{Y}(\mathbf{I} - \Pi_{\mathbf{V}}), \Sigma\mathbf{V}^\top \right\rangle$
- As $\Pi_{\mathbf{U}} = \mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top$, so

$$\begin{aligned}
\mathbf{U}^\top(\mathbf{I} - \Pi_{\mathbf{U}}) &= \mathbf{U}^\top - \mathbf{U}^\top\Pi_{\mathbf{U}} \\
&= \mathbf{U}^\top - \underbrace{\mathbf{U}^\top\mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}}_{=\mathbf{I}}\mathbf{U}^\top = \mathbf{U}^\top - \mathbf{U}^\top = 0
\end{aligned}$$

- So $\langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{U}\Sigma\mathbf{V}^\top \rangle = 0$.

Dual of $\| \cdot \|_*$ ... (3/3)

- With $\langle \Pi_{T^\perp}(\mathbf{Y}), \mathbf{U}\Sigma\mathbf{V}^\top \rangle = 0$, we have $\langle \mathbf{Y}, \mathbf{X} \rangle = \langle \Pi_T(\mathbf{Y}), \mathbf{U}\Sigma\mathbf{V}^\top \rangle$.
- The inner product is maximized by setting $\Pi_T(\mathbf{Y}) = \mathbf{U}\mathbf{V}^\top$, as this gives $\langle \Pi_T(\mathbf{Y}), \mathbf{U}\Sigma\mathbf{V}^\top \rangle = \langle \mathbf{U}\mathbf{V}^\top, \mathbf{U}\Sigma\mathbf{V}^\top \rangle = \langle \mathbf{I}, \Sigma \rangle = \sum_i \sigma_i = \|\mathbf{X}\|_*$.
- Furthermore, any $\mathbf{W} \in T^\perp$ such that $\|\mathbf{W}\|_2 \leq 1$ can be added inside $\mathbf{Y}$, and the maximum does not change. So now we have $\|\Pi_{T^\perp}(\mathbf{Y})\|_2 = \|\mathbf{W}\|_2 \leq 1$.
- The discussion above showed that the matrix $\mathbf{Y}$ such that $\Pi_T(\mathbf{Y}) = \mathbf{U}\mathbf{V}^\top$ and $\|\Pi_{T^\perp}(\mathbf{Y})\|_2 \leq 1$ is a subgradient of the nuclear norm at $\mathbf{X}$. i.e., we showed the subdifferential of nuclear norm can be expressed as

$$\partial\|\mathbf{M}\|_* = \left\{ \mathbf{Z} \;\middle|\; \Pi_T(\mathbf{Z}) = \mathbf{U}\mathbf{V}^\top, \|\Pi_{T^\perp}(\mathbf{Z})\|_2 \leq 1 \right\},$$

with $T = \left\{ \mathbf{U}\mathbf{X}^\top + \mathbf{Y}\mathbf{V}^\top, \mathbf{X} \in \mathbb{R}^{n\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r} \right\}$.

The proof of uniqueness of the solution NNM

- We are now ready to prove the uniqueness of the solution NNM.
- Let $\mathbf{X}_0$ be an optimal solution to Problem $(\mathcal{P})$. For the purpose of contradiction, let $\mathbf{X}_1 = \mathbf{X}_0 + \Delta$ be another optimal solution to $(\mathcal{P})$. Further assume there exists $\mathbf{P} \in \mathsf{Im}(\mathcal{A}^\top)$ such that

$$\Pi_T(\mathbf{P}) = \mathbf{U}\mathbf{V}^\top, \quad \|\Pi_{T^\perp}(\mathbf{P})\|_2 \leq 1.$$

  And also assume that $\mathcal{A}$ restricted to the element in T is injective.
- First, based on the fact that nuclear norm is convex, then

$$\|\mathbf{X}_0 + \Delta\|_* \geq \|\mathbf{X}_0\|_* + \langle \mathbf{U}\mathbf{V}^\top + \mathbf{W}, \Delta \rangle, \tag{5}$$

  where $\mathbf{W}$ satisfies the conditions $\|\mathbf{W}\|_2 \leq 1, \mathbf{U}^\top \mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0$.
- Next, using $\mathbf{P} = \Pi_T(\mathbf{P}) + \Pi_{T^\perp}(\mathbf{P})$ and the assumption, we have $\mathbf{U}\mathbf{V}^\top = \mathbf{P} - \Pi_{T^\perp}(\mathbf{Y})$, put this into (5)

$$\begin{aligned} \|\mathbf{X}_0 + \Delta\|_* &\geq \|\mathbf{X}_0\|_* + \langle \mathbf{P} - \Pi_{T^\perp}(\mathbf{P}) + \mathbf{W}, \Delta \rangle \\ &= \|\mathbf{X}_0\|_* + \langle \mathbf{P}, \Delta \rangle + \langle \mathbf{W} - \Pi_{T^\perp}(\mathbf{P}), \Delta \rangle. \end{aligned}$$

The proof ... 2/4

- By assumption, $\mathbf{P} \in \mathrm{Im}(\mathcal{A}^\top)$. By the fact that $\mathbf{X}_0$ and $\mathbf{X}_1$ are feasible solutions, $\Delta \in \mathrm{Null}(\mathbf{A})$. Using the linear algebra fact that null space of $\mathcal{A}$ is orthogonal to the range of $\mathbf{A}^\top$, the term $\langle \mathbf{P}, \Delta \rangle$ is zero, and hence

$$\|\mathbf{X}_0 + \Delta\|_* \quad \geq \quad \|\mathbf{X}_0\|_* + \langle \mathbf{W} - \Pi_{T^\perp}(\mathbf{P}), \Delta \rangle$$

- As $\mathbf{W} \in T^\perp$ and hence $\mathbf{W} = \Pi_{T^\perp}(\mathbf{W})$, therefore

$$\begin{aligned}
\|\mathbf{X}_0 + \Delta\|_* \quad &\geq \quad \|\mathbf{X}_0\|_* + \langle \Pi_{T^\perp}(\mathbf{W}) - \Pi_{T^\perp}(\mathbf{P}), \Delta \rangle \\
&= \quad \|\mathbf{X}_0\|_* + \langle \Pi_{T^\perp}(\mathbf{W} - \mathbf{P}), \Delta \rangle \\
&\overset{(*)}{=} \quad \|\mathbf{X}_0\|_* + \langle \mathbf{W} - \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle \\
&= \quad \|\mathbf{X}_0\|_* + \langle \mathbf{W}, \Pi_{T^\perp}(\Delta) \rangle - \langle \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle
\end{aligned}$$

- So

$$\|\mathbf{X}_0 + \Delta\|_* - \|\mathbf{X}\|_* \geq \langle \mathbf{W}, \Pi_{T^\perp}(\Delta) \rangle - \langle \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle.$$

- (*) In next slide we show $\langle \Pi_{T^\perp}(\mathbf{A}), \mathbf{B} \rangle = \langle \mathbf{A}, \Pi_{T^\perp}(\mathbf{B}) \rangle$

Showing $\langle \Pi_{T^\perp}(\mathbf{A}), \Delta \rangle = \langle \mathbf{A}, \Pi_{T^\perp}(\mathbf{B}) \rangle$

Direct proof.

$$\begin{aligned}
\langle \Pi_{T^\perp}(\mathbf{A}), \mathbf{B} \rangle &= \Big\langle (\mathbf{I} - \Pi_{\mathbf{U}})\mathbf{A}(\mathbf{I} - \Pi_{\mathbf{V}}), \mathbf{B} \Big\rangle \\
&= \Big\langle \mathbf{A}, (\mathbf{I} - \Pi_{\mathbf{U}})^\top \mathbf{B} (\mathbf{I} - \Pi_{\mathbf{V}})^\top \Big\rangle \\
&= \Big\langle \mathbf{A}, (\mathbf{I}^\top - \Pi_{\mathbf{U}}^\top) \mathbf{B} (\mathbf{I}^\top - \Pi_{\mathbf{V}}^\top) \Big\rangle \\
&= \Big\langle \mathbf{A}, (\mathbf{I} - \Pi_{\mathbf{U}}) \mathbf{B} (\mathbf{I} - \Pi_{\mathbf{V}}) \Big\rangle \\
&= \langle \mathbf{A}, \Pi_{T^\perp}(\mathbf{B}) \rangle
\end{aligned}$$

where

$$\Pi_{\mathbf{U}} = \mathbf{U}(\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \implies \Pi_{\mathbf{U}}^\top = \Pi_{\mathbf{U}}.$$

The proof ... 3/4

- As $\mathbf{W}$ is a free variable (as long as it satisfies the conditions of subgradient of nuclear norm at $\mathbf{X}_0$), we can set it using $\mathbf{P}$ as

$$\mathbf{W} = \Pi_{T^\perp}(\mathbf{P}), \tag{6}$$

where $\Pi_T(\mathbf{P}) = \mathbf{U}\mathbf{V}^\top$ and $\|\Pi_{T^\perp}(\mathbf{P})\|_2 \leq 1$. Such a dual matrix $\mathbf{P}$ always exists because of the dual characterization

$$\|\mathbf{X}\|_* = \sup_{\|\mathbf{P}\|_2 \leq 1} \langle \mathbf{P}, \mathbf{X} \rangle. \tag{7}$$

- Using (7),

$$\begin{aligned} \|\mathbf{X}_0 + \Delta\|_* - \|\mathbf{X}\|_* &\geq \langle \mathbf{W}, \Pi_{T^\perp}(\Delta) \rangle - \langle \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle. \\ &\overset{(6)}{=} \langle \Pi_{T^\perp}(\mathbf{P}), \Pi_{T^\perp}(\Delta) \rangle - \langle \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle \\ &\overset{(7)}{=} \|\Pi_{T^\perp}(\Delta)\|_* - \langle \mathbf{P}, \Pi_{T^\perp}(\Delta) \rangle \\ &= \|\Pi_{T^\perp}(\Delta)\|_* - \langle \Pi_T(\mathbf{P}) + \Pi_{T^\perp}(\mathbf{P}), \Pi_{T^\perp}(\Delta) \rangle \end{aligned}$$

- Using the same logic as in slide 15, $\langle \Pi_T(\mathbf{P}), \Pi_{T^\perp}(\Delta) \rangle = 0$, hence

$$\|\mathbf{X}_0 + \Delta\|_* - \|\mathbf{X}\|_* \geq \|\Pi_{T^\perp}(\Delta)\|_* - \langle \Pi_{T^\perp}(\mathbf{P}), \Pi_{T^\perp}(\Delta) \rangle.$$

The proof ... 4/4

- Using inequality for dual norms

$$\langle \Pi_{T^\perp}(\mathbf{P}), \Pi_{T^\perp}(\Delta) \rangle \leq \|\Pi_{T^\perp}(\mathbf{P})\|_2 \cdot \|\Pi_{T^\perp}(\Delta)\|_*$$

  Hence

$$\begin{aligned} \|\mathbf{X}_0 + \Delta\|_* - \|\mathbf{X}\|_* &\geq \|\Pi_{T^\perp}(\Delta)\|_* - \|\Pi_{T^\perp}(\mathbf{P})\|_2 \cdot \|\Pi_{T^\perp}(\Delta)\|_* \\ &= \Big(1 - \|\Pi_{T^\perp}(\mathbf{P})\|_2\Big) \|\Pi_{T^\perp}(\Delta)\|_*. \end{aligned}$$

- By assumption $\|\Pi_{T^\perp}(\mathbf{P})\|_2 < 1$ so $\|\mathbf{X}_0 + \Delta\|_* > \|\mathbf{X}\|_*$ unless $\|\Pi_{T^\perp}(\Delta)\|_* = 0$.
- If $\|\Pi_{T^\perp}(\Delta)\|_* = 0$, then $\Delta \in T$. Then $\mathcal{A}(\Delta) \stackrel{(1)}{=} 0$ implies $\Delta = 0$ due to injectivity assumption on $\mathcal{A}$.
- Therefore, $\|\mathbf{X}_0 + \Delta\|_* > \|\mathbf{X}\|_*$ unless $\Delta = 0$. The proof of uniqueness is completed.

End of document.