Characterizations of nuclear norm

Using operator norm and Frobenius norm

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft : Janurary 9, 2020
Last update : August 28, 2020

Characterizations of nuclear norm

- Given a matrix $\mathbf{X}$, the nuclear norm is

$$\|\mathbf{X}\|_* := \sum_i \sigma_i(\mathbf{X}),$$

  where σ_i is the i-th singular value of $\mathbf{X}$.
- A (dual) characterization of nuclear norm

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

  where $\|\mathbf{B}\|_2$ is the spectral norm/operator norm/2-norm of $\mathbf{B}$: the largest singular value of $\mathbf{B}$, i.e., $\|\mathbf{B}\|_2 = \sigma_1(\mathbf{B})$, and

$$\langle \mathbf{X}, \mathbf{B} \rangle := \sum_{i,j} X_{ij} B_{ij} = \mathrm{Tr}\left(\mathbf{B}^\top \mathbf{X}\right) = \mathrm{Tr}\left(\mathbf{X}^\top \mathbf{B}\right) \tag{1}$$

  is the definition of matrix inner product.
- Nuclear norm can also be characterized as

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{L}\mathbf{R}} \frac{1}{2}\Big(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2\Big).$$

This document : shows the proof of these characterizations.

The proof of $\|\mathbf{X}\|_* = \max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle$ ... (1/3)

Theorem Given a matrix $\mathbf{X}$, we have $\|\mathbf{X}\|_* = \max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle$.

Proof. Let $\mathbf{X} \overset{SVD}{=} \mathbf{U}\Sigma\mathbf{V}^\top$. Let $\mathbf{B} = \mathbf{U}\mathbf{V}^\top$ so $\|\mathbf{B}\|_2 \leq 1$ holds.

$$
\begin{array}{rcll}
\max_{\|\mathbf{B}\| \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle & \geq & \langle \mathbf{X}, \mathbf{B} \rangle & \text{by definition of max} \\
& = & \mathrm{Tr}\left(\mathbf{B}^\top \mathbf{X}\right) & \text{by definition of trace} \\
& = & \mathrm{Tr}\left(\mathbf{V}\mathbf{U}^\top \mathbf{U}\Sigma\mathbf{V}^\top\right) & \mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top, \mathbf{B} = \mathbf{U}\mathbf{V}^\top \\
& = & \mathrm{Tr}\left(\mathbf{V}\Sigma\mathbf{V}^\top\right) & \mathbf{U}^\top\mathbf{U} = \mathbf{I} \\
& = & \mathrm{Tr}\left(\mathbf{V}^\top\mathbf{V}\Sigma\right) & \mathrm{Tr}\left(\mathbf{ABC}\right) = \mathrm{Tr}\left(\mathbf{CAB}\right) \\
& = & \mathrm{Tr}\left(\Sigma\right) & \mathbf{V}^\top\mathbf{V} = \mathbf{I} \\
& = & \sum_i \sigma_i(\mathbf{X}) = \|\mathbf{X}\|_* &
\end{array}
$$

So we have

$$
\max_{\|\mathbf{B}\| \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle \geq \|\mathbf{X}\|_* \qquad (2)
$$

The proof of $\|\mathbf{X}\|_* = \max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle$ ... (2/3)

$$
\begin{aligned}
\langle \mathbf{X}, \mathbf{B} \rangle &= \mathrm{Tr}\,(\mathbf{B}^\top \mathbf{X}) && \text{by definition of trace} \\
&= \mathrm{Tr}\,(\mathbf{B}^\top \mathbf{U}\Sigma\mathbf{V}^\top) && \mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top \\
&= \mathrm{Tr}\,(\mathbf{V}^\top \mathbf{B}^\top \mathbf{U}\Sigma) && \mathrm{Tr}\,(\mathbf{ABC}) = \mathrm{Tr}\,(\mathbf{CAB}) \\
&= \mathrm{Tr}\,\big((\mathbf{U}^\top \mathbf{B}\mathbf{V})^\top \Sigma\big) && \\
&= \langle \mathbf{U}^\top \mathbf{B}\mathbf{V}, \Sigma \rangle && \text{by definition of inner product} \\
&= \textstyle\sum_{i,j} [\mathbf{U}^\top \mathbf{B}\mathbf{V}]_{ij} \Sigma_{ij} && \text{by definition of inner product} \\
&= \textstyle\sum_{i} [\mathbf{U}^\top \mathbf{B}\mathbf{V}]_{ii} \sigma_i && \Sigma \text{ is diagonal} \\
&= \textstyle\sum_{i} \mathbf{u}_i^\top \mathbf{B}\mathbf{v}_i \sigma_i &&
\end{aligned}
$$

To continue we need

$$
\mathbf{u}_i^\top \mathbf{B}\mathbf{v}_i \leq \max_{\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1} \mathbf{u}^\top \mathbf{B}\mathbf{v} = \sigma_1(\mathbf{B}).
$$

Explanation. Here $\mathbf{u}$ and $\mathbf{v}$ are *any vector* with unit norm: including the singular vector of $\mathbf{B}$ which gives $\sigma_1(\mathbf{B})$.

The proof of $\|\mathbf{X}\|_* = \max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle$ ... (3/3)

Continue from last equality

$$\langle \mathbf{X}, \mathbf{B} \rangle = \sum_i \mathbf{u}_i^\top \mathbf{B} \mathbf{v}_i \sigma_i.$$

Take max, and using the singular vector of $\mathbf{B}$ as $\mathbf{u}$ and $\mathbf{v}$, we have

$$\begin{aligned}
\max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle &= \max_{\|\mathbf{B}\|_2 \leq 1} \textstyle\sum_i \mathbf{u}_i^\top \mathbf{B} \mathbf{v}_i \sigma_i \\
&\leq \max_{\|\mathbf{B}\|_2 \leq 1} \textstyle\sum_i \sigma_1(\mathbf{B}) \cdot \sigma_i \\
&\leq \textstyle\sum_i \sigma_i \qquad \text{by condition } \|\mathbf{B}\|_2 = \sigma_1(\mathbf{B}) \leq 1 \\
&= \|\mathbf{X}\|_*
\end{aligned}$$

So we have $\max_{\|\mathbf{B}\|_2 \leq 1} \langle \mathbf{X}, \mathbf{B} \rangle \leq \|\mathbf{X}\|_*$, together with (2), implies

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

The proof of $\min_{\mathbf{X}} \|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{LR}} \frac{1}{2}\Big(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2\Big)$

Theorem Given a matrix $\mathbf{X}$, we have

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* = \min_{\mathbf{X}=\mathbf{LR}} \frac{1}{2}\Big(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2\Big).$$

Proof. Let $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top$ and let $\mathbf{L} = \mathbf{U}\Sigma^{\frac{1}{2}}$ and $\mathbf{R} = \Sigma^{\frac{1}{2}}\mathbf{V}^\top$.
Then $\|\mathbf{L}\|_F^2 = \mathrm{Tr}\left(\mathbf{L}^\top\mathbf{L}\right) = \mathrm{Tr}\left((\mathbf{U}\Sigma^{\frac{1}{2}})^\top\mathbf{U}\Sigma^{\frac{1}{2}}\right) = \mathrm{Tr}\left(\Sigma\right) = \sum_i \sigma_i^2 = \|\mathbf{X}\|_*$.
Similarly, $\|\mathbf{R}\|_F^2 = \|\mathbf{X}\|_*$. Hence $\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2 = 2\|\mathbf{X}\|_*$, and the $\frac{1}{2}$ makes them equal.

End of document