Proving Ky Fan norm (nuclear norm) $\|\mathbf{X}\|_*$ is a norm

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft : December 28, 2018
Last update : October 16, 2019

Quick recall of Singular Value Decomposition

The SVD of a given matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a factorization in the form

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top,$$

- $\mathbf{U} \in \mathbb{R}^{m \times m}$
  - the columns of $\mathbf{U}$ are the left-singular vectors of $\mathbf{A}$
  - these vectors are a set of orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^\top$
- $\Sigma \in \mathbb{R}^{m \times n}$
  - Σ is diagonal
  - the diagonal elements of Σ, denoted as $\sigma_i, i \in [1, 2, \ldots, \min\{m, n\}]$, are the singular values of $\mathbf{A}$
  - σ_i are all non-negative
  - Non-zero σ_i are the the square roots of the non-zero eigenvalues of both $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$
  - convention : σ_i are sorted as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$
- $\mathbf{V} \in \mathbb{R}^{n \times n}$
  - the columns of $\mathbf{V}$ are the right-singular vectors of $\mathbf{A}$
  - these vectors are a set of orthonormal eigenvectors of $\mathbf{A}^\top\mathbf{A}$

Ky Fan norm (a.k.a. the Nuclear norm)

Ky Fan norm[1] of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as

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

Note. As $\sigma_i \geq 0$, we can drop the absolute value sign.
Ky Fan norm of a matrix is the sum of the singular values of that matrix.

The Ky Fan k norm is defined as the sum of the k largest singular values.

Short hand notation : let $\sigma_{\mathbf{A}}$ be the vector holding all the singular values of $\mathbf{A}$, we can express Ky Fan norm as the l_1 norm of $\sigma_{\mathbf{A}}$

$$\|\mathbf{A}\|_* = \|\sigma_{\mathbf{A}}\|_1.$$

[1] Ky Fan. "Maximum properties and inequalities for the eigenvalues of completely continuous operators". Proceedings of the National Academy of Sciences of the United States of America. 37 (11): 760 - 766. 1951

Properties of Ky Fan norm

- It is a norm, and therefore :
  - $\|\cdot\|_*$ is a convex function on the set of $m \times n$ matrices
  - $\|\cdot\|_*$ satisfies the triangle inequality
- It is not differentiable
- The dual norm of $\|\cdot\|_*$ is the spectral norm $\|\cdot\|_2$
- $\langle \mathbf{X}, \mathbf{Y} \rangle \leq \|\mathbf{X}\|_* \|\mathbf{Y}\|_2$
- It is the special case of Schatten p-norm where $p = 1$

This document : show the proof that Ky Fan norm is a norm.

Proving Ky Fan norm is a norm

Let the space $\mathbb{R}^{m\times n}$ be V. Let $f(\mathbf{X}) = \|\mathbf{X}\|_*$ be a function on V.
To show $\|\mathbf{X}\|_*$ is a norm, we need to show

1. f is a non-negative real-value function defined on V
2. $f(\mathbf{X}) = 0$ if and only if $\mathbf{X} = \mathbf{0}$
3. $f(\alpha\mathbf{X}) = |\alpha| f(\mathbf{X})$ for all $\mathbf{X} \in V$ and scalar $\alpha \in \mathbb{R}$
4. $f(\mathbf{X}+\mathbf{Y}) \leq f(\mathbf{X}) + f(\mathbf{Y})$ for all $\mathbf{X}, \mathbf{Y} \in V$

Items 1-3 are easy to show :

- On 1 : by definition of Ky Fan norm as a sum of non-negative singular values
- On 2 : the singular values of zero matrix $\mathbf{0}$ are all zero, so $f(\mathbf{0}) = 0$. Furthermore, as singular values are always non-negative, there does not exist a matrix $\mathbf{A}$ with negative singular values, so for $f(\mathbf{X}) = 0$, $\mathbf{X}$ can only be $\mathbf{0}$.
- On 3 : by the fact that $-\mathbf{X}$, $\mathbf{X}$ and $k\mathbf{X}$ have the same set of singular values

Proving Ky Fan norm is a norm

To show $\|\mathbf{X}\|_*$ is a norm, the hard part is to show the function $f(\mathbf{X}) = \|\mathbf{X}\|_*$ satisfies the triangle inequality on V :

$$f(\mathbf{X}+\mathbf{Y}) \leq f(\mathbf{X}) + f(\mathbf{Y}), \ \forall\, \mathbf{X}, \mathbf{Y} \in V.$$

To prove this we need an equality between Ky Fan norm and a function :

$$\sup_{\sigma_1(\mathbf{Q}) \leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle = \|\mathbf{A}\|_*.$$

The next 3 slides will prove this by

- Showing $\displaystyle\sup_{\sigma_1(\mathbf{Q}) \leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle \geq \|\mathbf{A}\|_*$
- Showing $\displaystyle\sup_{\sigma_1(\mathbf{Q}) \leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle \leq \|\mathbf{A}\|_*$
- Conditions $\geq$ and $\leq$ means $=$

Note : this is exactly this reply made by David Speyer's on a question in stackexchange.. For the reference of the whole proof process for the next 4 slides, see Michael Grant's reply on this stackexchange thread.

Showing $\sup_{\sigma_1(\mathbf{Q})\le 1} \langle \mathbf{Q}, \mathbf{A}\rangle \ge \|\mathbf{A}\|_*$

Let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top$, now consider a matrix $\mathbf{Q}$ constructed as

$$\mathbf{Q} = \mathbf{U}\Sigma'\mathbf{V}^\top = \mathbf{U}[\mathbf{I}\ \mathbf{0}]\mathbf{V}^\top = \mathbf{U}\mathbf{V}^\top,$$

where $\Sigma' = [\mathbf{I}\ \mathbf{0}]$ is a matrix in $\mathbb{R}^{m\times n}$ with all diagonal elements equal to 1. Note that the largest singular value of $\mathbf{Q}$ is 1. Now the inner product between $\mathbf{Q}$ and $\mathbf{A}$ is

$$\langle \mathbf{Q}, \mathbf{A}\rangle\Big|_{\mathbf{Q}=\mathbf{U}\mathbf{V}^\top} = \mathrm{Tr}(\mathbf{Q}^\top\mathbf{A}) \overset{(1)}{=} \mathrm{Tr}(\mathbf{V}\mathbf{U}^\top\mathbf{U}\Sigma\mathbf{V}^\top) \overset{(2)}{=} \mathrm{Tr}\Sigma = \|\mathbf{A}\|_*,$$

where (1) is due to $\mathbf{U}$ is unitary $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_m$ and (2) is due to the property $\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB})$ and $\mathbf{V}$ is also unitary.

Showing $\sup_{\sigma_1(\mathbf{Q}) \leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle \geq \|\mathbf{A}\|_*$

Note that it is universal true that, for all function $f(x)$ and a set C, we always have the inequality :

$$\sup_{x \in C} f(x) \geq f(x_0), \ \forall x_0 \in C.$$

Therefore, the expression $\langle \mathbf{Q}, \mathbf{A} \rangle \Big|_{\mathbf{Q}=\mathbf{U}\mathbf{V}^\top} = \|\mathbf{A}\|_*$ can be treated as a function f on $\mathbf{Q}$ evaluated at the specific $\mathbf{Q}_0 = \mathbf{U}\mathbf{V}^\top$. Hence, we have

$$\sup_{\sigma_1(\mathbf{Q}) \leq 1} f(\mathbf{Q}) \geq f(\mathbf{Q}_0) = f(\mathbf{U}\mathbf{V}^\top) = \|\mathbf{A}\|_*$$

The inequality above means, for all possible $\mathbf{Q}$ such that $\sigma_1(\mathbf{Q}) \leq 1$ (spectral norm of $\mathbf{Q}$ is at most 1), the function f at a point $\mathbf{Q}_0 = \mathbf{U}\mathbf{V}^\top$ (which fulfil $\sigma_1(\mathbf{Q}_0) \leq 1$), is lower bounded by the supremum on f over all possible $\mathbf{Q}$ such that $\sigma_1(\mathbf{Q}) \leq 1$.
That is, we now have

$$\sup_{\sigma_1(\mathbf{Q}) \leq 1} f(\mathbf{Q}) \geq \|\mathbf{A}\|_* \tag{1}$$

Showing $\sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle \leq \|\mathbf{A}\|_*$

$$
\begin{aligned}
\sup_{\sigma_1(\mathbf{Q})\leq 1} f(\mathbf{Q}) &= \sup_{\sigma_1(\mathbf{Q})\leq 1} \mathrm{Tr}\big(\mathbf{Q}^\top \mathbf{A}\big) && \text{Definition of inner product} \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \mathrm{Tr}\big(\mathbf{Q}^\top \mathbf{U}\Sigma\mathbf{V}^\top\big) && \text{SVD of } \mathbf{A} \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \mathrm{Tr}\big(\mathbf{V}^\top\mathbf{Q}^\top \mathbf{U}\Sigma\big) && \text{Property of trace of product} \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \mathrm{Tr}\big((\mathbf{U}\mathbf{Q}\mathbf{V})^\top \Sigma\big) \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{U}\mathbf{Q}\mathbf{V}, \Sigma \rangle && \text{Definition of inner product} \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \sum_i (\mathbf{U}\mathbf{Q}\mathbf{V})_{ii}\sigma_i && \Sigma \text{ is diagonal} \\
&= \sup_{\sigma_1(\mathbf{Q})\leq 1} \sum_i \sigma_i \mathbf{u}_i^\top \mathbf{Q}\mathbf{v}_i \\
&\leq \sup_{\sigma_1(\mathbf{Q})\leq 1} \sum_i \sigma_i \sigma_1(\mathbf{Q}) \\
&= \textstyle\sum_i \sigma_i = \|\mathbf{A}\|_*
\end{aligned}
$$

So we now have $\sup_{\sigma_1(\mathbf{Q})\leq 1} f(\mathbf{Q}) \leq \|\mathbf{A}\|_*$, together with (1) we showed

$$\sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle = \|\mathbf{A}\|_*.$$

Ky Fan norm satisfies the triangle inequality

Now we have

$$\sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle = \|\mathbf{A}\|_*,$$

we can now prove Ky Fan norm satisfies the triangle inequality.
Now consider two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, apply the equality we just proved : replace $\mathbf{A}$ by $\mathbf{A}+\mathbf{B}$, we get

$$\|\mathbf{A}+\mathbf{B}\|_* \quad = \quad \sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A}+\mathbf{B} \rangle$$

The supremum of inner product itself obeys the triangle inequality, thus

$$\sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A}+\mathbf{B} \rangle \quad \leq \quad \sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle + \sup_{\sigma_1(\mathbf{Q})\leq 1} \langle \mathbf{Q}, \mathbf{B} \rangle$$

Therefore,

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

That is, Ky Fan norm satisfies the triangle inequality.

Last page - summary

- The Ky Fan norm of matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$
$$\|\mathbf{A}\|_* = \sum_{i}^{\min\{m,n\}} \sigma_i .$$
- Proving $\|\cdot\|_*$ is a norm :
  - It satisfies $\|\mathbf{A}\|_* = 0$ only if $\mathbf{A} = \mathbf{0}$ and $\|t\mathbf{A}\|_* = |t|\|\mathbf{A}\|_*$
  - It satisfies $\|\mathbf{A} + \mathbf{B}\|_* \leq \|\mathbf{A}\|_* + \|\mathbf{B}\|_*$.
    The proof based on the equality $\sup\limits_{\sigma_1(\mathbf{Q}) \leq 1} \langle \mathbf{Q}, \mathbf{A} \rangle = \|\mathbf{A}\|_*$
- Hence $\|\cdot\|_*$ is a convex function on matrices

What's next : showing the sub-differential of the Ky Fan norm is the set

$$\partial\|\mathbf{X}\|_* = \left\{ \mathbf{U}\mathbf{V}^\top + \mathbf{W} \,\middle|\, \mathbf{W} \in \mathbb{R}^{m \times n}, \mathbf{U}^\top\mathbf{W} = 0, \mathbf{W}\mathbf{V} = 0, \|\mathbf{W}\|_2 \leq 1 \right\}$$

End of document