Alternative proof of the SVT operator theorem

using von Neumann trace inequality

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft: September 21, 2019
Last update : February 5, 2020

SVT theorem

Given a matrix $\mathbf{Y} \in \mathbb{R}^{m\times n}$ and $\lambda > 0$, the solution to the following optimization problem

$$\mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_*$$

is

$$\mathbf{X}^* = \mathsf{SVT}_\lambda(\mathbf{Y}) := \mathbf{U}[\Sigma - \lambda\mathbf{I}]_+\mathbf{V}^\top$$

where $\mathbf{U}\Sigma\mathbf{V}^\top$ =SVD$(\mathbf{Y})$ and $[\,\cdot\,]_+ = \max(\,\cdot\,, 0)$.

The operator $\mathsf{SVT}_\lambda(\cdot)$ is called the Singular Value Thresholding operator (SVT).
Recall the definition of nuclear norm

$$\|\mathbf{X}\|_* = \sum_i \sigma_i \tag{1}$$

What the SVT does : step-by-step

1. Perform SVD on $\mathbf{Y}$ to get $\mathbf{U}\Sigma\mathbf{V}^\top$
2. Subtract all the diagonal value of Σ by λ, denoted as $\Sigma - \lambda\mathbf{I}$
3. Replace negative value in $\Sigma - \lambda\mathbf{I}$ by zero, denoted by $[\Sigma - \lambda\mathbf{I}]_+$
4. Multiply $\mathbf{U}$, $[\Sigma - \lambda\mathbf{I}]_+$ and $\mathbf{V}^\top$

For the proof of the SVT theorem, see here for the proof based on sub-differential of $\|\cdot\|_*$.

The proof by sub-differential may be too complicated, here we show another way to prove the SVT theorem, using von Neumann trace inequality (or Ky Fan inequality)

$$\text{Tr}(\mathbf{X}^\top\mathbf{Y}) \leq \sum_i \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y})$$

Proving the SVT operator via von Neumann inequality

First we have

$$
\begin{aligned}
\frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 &= \frac{1}{2}\|\mathbf{X}\|_F^2 - \mathrm{Tr}\left(\mathbf{X}^\top \mathbf{Y}\right) + \frac{1}{2}\|\mathbf{Y}\|_F^2 \\
&= \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{X})\Big)^2 - \mathsf{Tr}(\mathbf{X}^\top \mathbf{Y}) + \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{Y})\Big)^2,
\end{aligned}
$$

where the last equality is due to the relation of trace and singular values.

Now a tricky step by von Neumann trace inequality,

$$
\begin{aligned}
\frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 &\leq \sum_i \Big(\sigma_i(\mathbf{X})\Big)^2 - \sum_i \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) + \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{Y})\Big)^2 \\
&= \sum_i \left\{\frac{1}{2}\Big(\sigma_i(\mathbf{X})\Big)^2 - \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) + \frac{1}{2}\Big(\sigma_i(\mathbf{Y})\Big)^2\right\} \\
&= \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Y})\Big)^2.
\end{aligned}
$$

Continue from last page

$$\begin{aligned} \mathbf{X}^* &= \underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_* \\ &\leq \underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Y})\Big)^2 + \lambda\|\mathbf{X}\|_* \end{aligned}$$

The equality of von Neumann inequality holds if $\mathbf{X}, \mathbf{Y}$ share the same left and right singular vectors : so $\mathbf{X}^*$ has the form $\mathbf{U}\Sigma_{\mathbf{X}}\mathbf{V}^\top$ where $\mathbf{U}, \mathbf{V}^\top$ come from SVD of $\mathbf{Y}$.

Recall that

$$\mathsf{SVT}_\lambda(\mathbf{Y}) := \mathbf{U}[\Sigma - \lambda\mathbf{I}]_+\mathbf{V}^\top$$

so now we have proved the blue parts, the remaining is to prove the middle part.

By selecting $\mathbf{X}^*$ in the form $\mathbf{U}\Sigma_{\mathbf{X}}\mathbf{V}^\top$ where $\mathbf{U}, \mathbf{V}$ come from SVD$(\mathbf{Y})$, the inequality becomes equality and thus we have

$$
\begin{aligned}
\mathbf{X}^* &= \operatorname*{argmin}_{\mathbf{X}} \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Y})\Big)^2 + \lambda\|\mathbf{X}\|_* \\
&\overset{(1)}{=} \operatorname*{argmin}_{\mathbf{X}} \frac{1}{2}\sum_i \Big(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Y})\Big)^2 + \lambda\sum_i \sigma_i \\
&= \sum_i \left\{ \operatorname*{argmin}_{\sigma_i(\mathbf{X})} \frac{1}{2}(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Y}))^2 + \lambda|\sigma_i(\mathbf{X})| \right\}.
\end{aligned}
$$

For each i, the function is in the form

$$
\min_x \frac{1}{2}(x-y)^2 + \lambda|x|,
$$

which can be solved by soft-thresholding. Thus by applying soft-thresholding on the singular values of $\mathbf{Y}$, the proof is finished. □

End of document