Supplementary Material to "Weighted Nuclear Norm Minimization with Application to Image Denoising"

Shuhang Gu[1], Lei Zhang[1], Wangmeng Zuo[2], Xiangchu Feng[3]
[1]Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[3]Dept. of Applied Mathematics, Xidian University, Xi'an, China

Here we give the proofs of the 3 lemmas used in the main paper. Since the proof of Lemma 2 will use the conclusion of Lemma 3, we prove Lemma 3 before proving Lemma 2 in the following proofs.

Lemma 1. *$\forall \boldsymbol{A}, \boldsymbol{B} \in \Re^{m\times n}$ that satisfy $\boldsymbol{A}^T\boldsymbol{B} = 0$, we have*

$$
\begin{aligned}
&(1)\|\boldsymbol{A}+\boldsymbol{B}\|_{\boldsymbol{w},*} \geq \|\boldsymbol{A}\|_{\boldsymbol{w},*};\\
&(2)\|\boldsymbol{A}+\boldsymbol{B}\|_F \geq \|\boldsymbol{A}\|_F.
\end{aligned}
$$

Proof. Denote by $\lambda_k(\boldsymbol{X})$ and $\sigma_k(\boldsymbol{X})$ the k-th eigenvalue and singular value of matrix $\boldsymbol{X}$, respectively, and denote by $\mathbb{S}$ the subspace of $\Re^n$. Based on the *Courant-Fisher MaxMin Theorem* [1], we have

$$
\begin{aligned}
\lambda_k(\boldsymbol{A}^T\boldsymbol{A}+\boldsymbol{B}^T\boldsymbol{B}) &= \max_{dim(\mathbb{S})=k,}\ \min_{\boldsymbol{Y}\neq 0, \boldsymbol{Y}\in\mathbb{S}} \frac{\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}};\\
&= \max_{dim(\mathbb{S})=k,}\ \min_{\boldsymbol{Y}\neq 0, \boldsymbol{Y}\in\mathbb{S}} \frac{\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{B}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{A}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}}\\
&= \sigma_k^2(\boldsymbol{A}+\boldsymbol{B})\\
\lambda_k(\boldsymbol{A}^T\boldsymbol{A}) &= \max_{dim(\mathbb{S})=k,}\ \min_{\boldsymbol{Y}\neq 0, \boldsymbol{Y}\in\mathbb{S}} \frac{\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}}\\
&= \sigma_k^2(\boldsymbol{A}).
\end{aligned}
$$

Since $\boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Y} \geq 0$, we have

$$
\begin{gathered}
\frac{\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}} \geq \frac{\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}},\\
\sigma_k^2(\boldsymbol{A}+\boldsymbol{B}) \geq \sigma_k^2(\boldsymbol{A}),\\
|\sigma_k(\boldsymbol{A}+\boldsymbol{B})| \geq |\sigma_k(\boldsymbol{A})|.
\end{gathered}
$$

Based on the definition of F-norm and weighted nuclear norm, we have

$$
\begin{gathered}
\|\boldsymbol{A}+\boldsymbol{B}\|_{w,*} = \sum_i |w_i\sigma_i(\boldsymbol{A}+\boldsymbol{B})| \geq \sum_i |w_i\sigma_i(\boldsymbol{A})| = \|\boldsymbol{A}\|_{w,*};\\
\|\boldsymbol{A}+\boldsymbol{B}\|_F^2 = \sum_i \sigma_i^2(\boldsymbol{A}+\boldsymbol{B}) \geq \sum_i \sigma_i^2(\boldsymbol{A}) = \|\boldsymbol{A}\|_F^2.
\end{gathered}
$$

□

Lemma 3. *$\forall \boldsymbol{A} \in \Re^{n\times n}$ and a diagonal non-negative matrix $\boldsymbol{W} \in \Re^{n\times n}$ with non-ascending ordered diagonal elements, let $\boldsymbol{A} = \boldsymbol{X\Phi Y}^T$ be the SVD of $\boldsymbol{A}$, we have*

$$
\sum_i \sigma_i(\boldsymbol{A})\sigma_i(\boldsymbol{W}) = \max_{\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}, \boldsymbol{V}^T\boldsymbol{V}=\boldsymbol{I}} tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}],
$$

where $\boldsymbol{I}$ *is the identity matrix,* $\sigma_i(\boldsymbol{A})$ *and* $\sigma_i(\boldsymbol{W})$ *are the i-th singular values of matrices* $\boldsymbol{A}$ *and* $\boldsymbol{W}$*, respectively. When* $\boldsymbol{U}=\boldsymbol{X}$ *and* $\boldsymbol{V}=\boldsymbol{Y}$*,* $tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}]$ *reaches its maximum value.*

Proof. Based on the definition of eigenvalue and the triangle inequality, we have

$$\begin{aligned} tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}] &= |\sum_i \lambda_i(\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V})| \\ &\le \sum_i |\lambda_i(\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V})|. \end{aligned}$$

Based on the Theorem 3.3.13 and Theorem 3.3.14 in [2], we have

$$\begin{aligned} \sum_i |\lambda_i(\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V})| &\le \sum_i \sigma_i(\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}) \\ &\le \sum_i \sigma_i(\boldsymbol{W})\sigma_i(\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}) \\ &\le \sum_i \sigma_i(\boldsymbol{W})\sigma_i(\boldsymbol{A}) \end{aligned}$$

From the above derivation, we have $\sum_i \sigma_i(\boldsymbol{W})\sigma_i(\boldsymbol{A}) \ge tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}]$. For a special case, let $\boldsymbol{U}=\boldsymbol{X}$ and $\boldsymbol{V}=\boldsymbol{Y}$, $tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}]$ reaches its maximum:

$$\sum_i \sigma_i(\boldsymbol{A})\sigma_i(\boldsymbol{W}) = \max_{\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I},\boldsymbol{V}^T\boldsymbol{V}=\boldsymbol{I}} tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{V}].$$

$\square$

Lemma 2. $\forall \boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$ *with* $\boldsymbol{A} \in \Re^{m\times m}$ *and* $\boldsymbol{D} \in \Re^{n\times n}$*, if the weights satisfy* $w_1 \ge \cdots \ge w_{m+n} \ge 0$*, we have*

$$\|\boldsymbol{M}\|_{\boldsymbol{w},*} \ge \|\boldsymbol{A}\|_{\boldsymbol{w}_1,*} + \|\boldsymbol{D}\|_{\boldsymbol{w}_2,*},$$

where $\boldsymbol{w} = [w_1,\ldots,w_{m+n}]$*,* $\boldsymbol{w}_1 = [w_1,\ldots,w_m]$ *and* $\boldsymbol{w}_2 = [w_{m+1},\ldots,w_{m+n}]$*.*

Proof. Based on the conclusion of Lemma 3, assume that $\boldsymbol{W}$ is a diagonal matrix and its diagonal elements are the corresponding elements in the weight vector $\boldsymbol{w}$. If the weights satisfy $w_1 \ge \ldots \ge w_n \ge 0$, $\forall \boldsymbol{X} \in \Re^{n\times n}$ we have

$$\|\boldsymbol{X}\|_{w,*} = \sum_i \sigma_i(\boldsymbol{X})\sigma_i(\boldsymbol{W}) = \max_{\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I},\boldsymbol{V}^T\boldsymbol{V}=\boldsymbol{I}} tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}].$$

For matrix $\boldsymbol{M}$, with the above equality we have

$$\begin{aligned} \|\boldsymbol{M}\|_{w,*} = \left\| \begin{matrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{matrix} \right\|_{w,*} &= \max_{\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I},\boldsymbol{V}^T\boldsymbol{V}=\boldsymbol{I}} tr[\boldsymbol{W}\boldsymbol{U}^T\boldsymbol{M}\boldsymbol{V}] \\ &\ge \max_{\boldsymbol{U}_1^T\boldsymbol{U}_1=\boldsymbol{I},\boldsymbol{V}_1^T\boldsymbol{V}_1=\boldsymbol{I},\boldsymbol{U}_2^T\boldsymbol{U}_2=\boldsymbol{I},\boldsymbol{V}_2^T\boldsymbol{V}_2=\boldsymbol{I}} tr[\begin{pmatrix} \boldsymbol{w}_1 & 0 \\ 0 & \boldsymbol{w}_2 \end{pmatrix}\begin{pmatrix} \boldsymbol{U}_1^T & 0 \\ 0 & \boldsymbol{U}_2^T \end{pmatrix}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}\begin{pmatrix} \boldsymbol{V}_1 & 0 \\ 0 & \boldsymbol{V}_2 \end{pmatrix}] \\ &= \max_{\boldsymbol{U}_1^T\boldsymbol{U}_1=\boldsymbol{I},\boldsymbol{V}_1^T\boldsymbol{V}_1=\boldsymbol{I}} tr[\boldsymbol{W}_1\boldsymbol{U}_1^T\boldsymbol{A}\boldsymbol{V}_1] + \max_{\boldsymbol{U}_1^T\boldsymbol{U}_1=\boldsymbol{I},\boldsymbol{V}_2^T\boldsymbol{V}_2=\boldsymbol{I}} tr[\boldsymbol{W}_2\boldsymbol{U}_2^T\boldsymbol{D}\boldsymbol{V}_2] \\ &= \|\boldsymbol{A}\|_{\boldsymbol{w}_1,*} + \|\boldsymbol{D}\|_{\boldsymbol{w}_2,*}. \end{aligned}$$

$\square$

References

[1] G. H. Golub and C. F. Van Loan. Matrix computations (johns hopkins studies in mathematical sciences). 1996.

[2] R. A. Horn. *Topics in Matrix Analysis*. Cambridge university press, 1994.