Supplementary Material to "Adaptive Rank Estimate in Robust Principal Component Analysis"

In this document, we give the proofs of one lemma and two theorems used in the main paper.

1. Proof of Lemma 1

Lemma 1 If $\boldsymbol{C}, \boldsymbol{D} \in R^{m\times n}$ satisfy $\boldsymbol{C}^T\boldsymbol{D} = 0$, we have

$$\|\boldsymbol{C}+\boldsymbol{D}\|_{\boldsymbol{W}} \geq \|\boldsymbol{C}\|_{\boldsymbol{W}}$$
$$\|\boldsymbol{C}+\boldsymbol{D}\|_F \geq \|\boldsymbol{C}\|_F$$

Proof.

Let $\lambda_k(\boldsymbol{X})$ and $\sigma_k(\boldsymbol{X})$ denote he kth eigenvalue and singular value of matrix $\boldsymbol{X} \in R^{m\times n}$, respectively, and $\mathbb{S}$ is the subspace of R^n. Based on the Courant-Fischer MaxMin Theorem, we can obtain that

$$\begin{aligned}\lambda_k(\boldsymbol{C}^T\boldsymbol{C}+\boldsymbol{D}^T\boldsymbol{D}) &= \max_{dim(\mathbb{S})=k}\min_{\substack{\boldsymbol{Y}\neq 0\\ \boldsymbol{Y}\in\mathbb{S}}}\frac{(\boldsymbol{Y}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{Y})}{\boldsymbol{Y}^T\boldsymbol{Y}}\\ &= \max_{dim(\mathbb{S})=k}\min_{\substack{\boldsymbol{Y}\neq 0\\ \boldsymbol{Y}\in\mathbb{S}}}\frac{\boldsymbol{Y}^T(\boldsymbol{C}^T\boldsymbol{C}+\boldsymbol{D}^T\boldsymbol{D}+\boldsymbol{D}^T\boldsymbol{C}+\boldsymbol{C}^T\boldsymbol{D})\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}}\\ &= \max_{dim(\mathbb{S})=k}\min_{\substack{\boldsymbol{Y}\neq 0\\ \boldsymbol{Y}\in\mathbb{S}}}\frac{\boldsymbol{Y}^T(\boldsymbol{C}+\boldsymbol{D})^T(\boldsymbol{C}+\boldsymbol{D})\boldsymbol{Y}}{\boldsymbol{Y}^T\boldsymbol{Y}}\\ &= {\sigma_k}^2(\boldsymbol{C}+\boldsymbol{D}) \end{aligned} \tag{a.1}$$

$$\begin{aligned}\lambda_k(\boldsymbol{C}^T\boldsymbol{C}) &= \max_{dim(\mathbb{S})=k}\min_{\substack{\boldsymbol{Y}\neq 0\\ \boldsymbol{Y}\in\mathbb{S}}}\frac{(\boldsymbol{Y}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{Y})}{\boldsymbol{Y}^T\boldsymbol{Y}}\\ &= {\sigma_k}^2(\boldsymbol{C})\end{aligned} \tag{a.2}$$

Since $\boldsymbol{Y}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{Y} \geq 0$, we can obtain that

$$\frac{(\boldsymbol{Y}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{Y}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{Y})}{\boldsymbol{Y}^T\boldsymbol{Y}} \geq \frac{(\boldsymbol{Y}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{Y})}{\boldsymbol{Y}^T\boldsymbol{Y}} \tag{a.3}$$

$${\sigma_k}^2(\boldsymbol{C}+\boldsymbol{D}) \geq {\sigma_k}^2(\boldsymbol{C}) \tag{a.4}$$

$$|\sigma_k(\boldsymbol{C}+\boldsymbol{D})| \geq |\sigma_k(\boldsymbol{C})| \tag{a.5}$$

Based on the definition of $\|\cdot\|_F$ and $\|\cdot\|_{\boldsymbol{W}}$, we can obtain that

$$\|\boldsymbol{C}+\boldsymbol{D}\|_{\boldsymbol{W}} = \sum_i |\omega_i\sigma_i(\boldsymbol{C}+\boldsymbol{D})| \geq \sum_i |\omega_i\sigma_i(\boldsymbol{C})| = \|\boldsymbol{C}\|_{\boldsymbol{W}} \tag{a.6}$$

$$\|\boldsymbol{C}+\boldsymbol{D}\|_F^2 = \sum_i {\sigma_k}^2(\boldsymbol{C}+\boldsymbol{D}) \geq \sum_i {\sigma_k}^2(\boldsymbol{C}) = \|\boldsymbol{C}\|_F^2 \tag{a.7}$$

2. Proof of Theorem 1

Theorem 1 Given $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ where $\boldsymbol{Q} = \boldsymbol{U_Q}\boldsymbol{D_Q}\boldsymbol{V_Q}^T$, the minimization problem

$$\arg\min_{\boldsymbol{P}} \frac{1}{2}\|\boldsymbol{P}-\boldsymbol{Q}\|_F^2 + \tau\|\boldsymbol{P}\|_{\boldsymbol{W}} \tag{a.8}$$

has solution $\widehat{\boldsymbol{P}} = \boldsymbol{U_Q}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}\boldsymbol{V_Q}^T$, where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}$ is the solution of the following optimization problem:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}} = \arg\min_{\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}} \frac{1}{2}\left\|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}-\boldsymbol{D_Q}\right\|_F^2 + \tau\left\|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}\right\|_{\boldsymbol{W}} \tag{a.9}$$

Proof.

Define that $\boldsymbol{U}_{\boldsymbol{Q}\perp}$ is the set of orthogonal based of the complementary space of $\boldsymbol{U}_{\boldsymbol{Q}}$, the matrix $\boldsymbol{P}$ can be expressed as $\boldsymbol{P} = \boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{B}_1 + \boldsymbol{U}_{\boldsymbol{Q}\perp}\boldsymbol{B}_2$, where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are the components of $\boldsymbol{P}$ in subspaces $\boldsymbol{U}_{\boldsymbol{Q}\perp}$ and $\boldsymbol{U}_{\boldsymbol{Q}}$, respectively. Then, we can get

$$f(\boldsymbol{P}) = \frac{1}{2}\|\boldsymbol{P}-\boldsymbol{Q}\|_F^2 + \tau\|\boldsymbol{P}\|_{\boldsymbol{W}}$$

$$= \frac{1}{2}\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{B}_1 + \boldsymbol{U}_{\boldsymbol{Q}\perp}\boldsymbol{B}_2 - \boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{Q}}\boldsymbol{V}_{\boldsymbol{Q}}^T\right\|_F^2 + \tau\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{B}_1 + \boldsymbol{U}_{\boldsymbol{Q}\perp}\boldsymbol{B}_2\right\|_{\boldsymbol{W}}$$

Based on Lemma 1, the function can be expressed as

$$f(\boldsymbol{P}) \geq \frac{1}{2}\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{B}_1 - \boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{Q}}\boldsymbol{V}_{\boldsymbol{Q}}^T\right\|_F^2 + \tau\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{B}_1\right\|_{\boldsymbol{W}} \quad (\mathrm{a.10})$$

Similarly, for $\boldsymbol{V}_{\boldsymbol{Q}}$ we can get

$$f(\boldsymbol{P}) \geq \frac{1}{2}\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{\Sigma}_{\boldsymbol{P}}\boldsymbol{V}_{\boldsymbol{Q}}^T - \boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{Q}}\boldsymbol{V}_{\boldsymbol{Q}}^T\right\|_F^2 + \tau\left\|\boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{\Sigma}_{\boldsymbol{P}}\boldsymbol{V}_{\boldsymbol{Q}}^T\right\|_{\boldsymbol{W}} \quad (\mathrm{a.11})$$

Since $\boldsymbol{U}_{\boldsymbol{Q}}$ and $\boldsymbol{V}_{\boldsymbol{Q}}$ will be not change $\|\cdot\|_F$ and $\|\cdot\|_{\mathbf{w},*}$, we can get

$$f(\boldsymbol{P}) \geq \frac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{P}} - \boldsymbol{D}_{\boldsymbol{Q}}\right\|_F^2 + \tau\|\boldsymbol{\Sigma}_{\boldsymbol{P}}\|_{\boldsymbol{W}}$$

Therefore, according to the solution of Eq. (a.11), the solution of Eq. (a.10) can be obtained by $\widehat{\boldsymbol{P}} = \boldsymbol{U}_{\boldsymbol{Q}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}\boldsymbol{V}_{\boldsymbol{Q}}^T$

3. Proof of Theorem 2

Theorem 2 Given $\tau > 0$, $\boldsymbol{P}, \boldsymbol{Q} \in \mathbb{R}^{m\times n}$ where $\boldsymbol{Q} = \boldsymbol{U}_{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{Q}}\boldsymbol{V}_{\boldsymbol{Q}}^T$, $\boldsymbol{D}_{\boldsymbol{Q}} = diag(\delta_{\boldsymbol{Q}1}, \cdots, \delta_{\boldsymbol{Q}r}, \delta_{\boldsymbol{Q}(r+1)}, \cdots, \delta_{\boldsymbol{Q}\ell})$ and $\ell = min(m,n)$. We can define $\boldsymbol{Q} = \boldsymbol{Q}_1 + \boldsymbol{Q}_2$, $\boldsymbol{Q}_1 = \boldsymbol{U}_{\boldsymbol{Q}1}\boldsymbol{D}_{\boldsymbol{Q}1}\boldsymbol{V}_{\boldsymbol{Q}1}^T$ and $\boldsymbol{Q}_2 = \boldsymbol{U}_{\boldsymbol{Q}2}\boldsymbol{D}_{\boldsymbol{Q}2}\boldsymbol{V}_{\boldsymbol{Q}2}^T$, where $\boldsymbol{D}_{\boldsymbol{Q}1} = diag(\delta_{\boldsymbol{Q}1}, \cdots, \delta_{\boldsymbol{Q}r}, 0, \cdots, 0)$, $\boldsymbol{U}_{\boldsymbol{Q}1}$ and $\boldsymbol{V}_{\boldsymbol{Q}1}$ are the singular vector matrices corresponding to the rth largest singular values, $\boldsymbol{D}_{\boldsymbol{Q}2} = diag(0, \cdots, 0, \delta_{\boldsymbol{Q}(r+1)}, \cdots, \delta_{\boldsymbol{Q}\ell})$, $\boldsymbol{U}_{\boldsymbol{Q}2}$ and $\boldsymbol{V}_{\boldsymbol{Q}2}$ corresponding to the singular values from $(r+1)$th to the last. $\|\cdot\|_{\boldsymbol{W}}$ is defined as Eq. (30) and Eq. (31). The optimal solution of the minimization problem

$$\arg\min_{\boldsymbol{P}} \tfrac{1}{2}\|\boldsymbol{P}-\boldsymbol{Q}\|_F^2 + \tau\|\boldsymbol{P}\|_{\boldsymbol{W}}$$

can be expressed as

$$\boldsymbol{P}^* = \mathcal{D}_{\tau,\boldsymbol{W}}[\boldsymbol{Q}] = \boldsymbol{U}_{\boldsymbol{Q}}(\boldsymbol{D}_{\boldsymbol{Q}1} + S_\tau[\boldsymbol{D}_{\boldsymbol{Q}2}])\boldsymbol{V}_{\boldsymbol{Q}}^T = \boldsymbol{Q}_1 + \boldsymbol{U}_{\boldsymbol{Q}2}S_\tau[\boldsymbol{D}_{\boldsymbol{Q}2}]\boldsymbol{V}_{\boldsymbol{Q}2}^T$$

Proof.

Based on the Theorem 1, we can get the equivalent solution of the problem (40), namely $\widehat{\boldsymbol{P}} = \boldsymbol{U}_{\boldsymbol{Q}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}\boldsymbol{V}_{\boldsymbol{Q}}^T$ where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}} = \arg\min_{\boldsymbol{\Sigma}_{\boldsymbol{P}}} \frac{1}{2}\left\|\boldsymbol{\Sigma}_{\boldsymbol{P}} - \boldsymbol{D}_{\boldsymbol{Q}}\right\|_F^2 + \tau\|\boldsymbol{\Sigma}_{\boldsymbol{P}}\|_{\mathbf{W}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{P}} = diag\{\sigma_1(\boldsymbol{P}), \cdots, \sigma_\ell(\boldsymbol{P})\}$. The minimize function of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}}$ can be expressed as

$$\mathcal{T}(\boldsymbol{\Sigma}_{\boldsymbol{P}}) = \frac{1}{2}\sum_{i=1}^{\ell}\left(\sigma_i^2(\boldsymbol{Q}) - 2\sigma_i(\boldsymbol{P})\sigma_i(\boldsymbol{Q}) + \sigma_i^2(\boldsymbol{P})\right) + \tau\sum_{i=r+1}^{\ell}\sigma_i(\boldsymbol{P}) \quad (\mathrm{a.12})$$

$\sigma_i^2(\boldsymbol{Q})$ can be regarded as a constant and ignored. Therefore, the function can be expressed as

$$\mathcal{T}(\boldsymbol{\Sigma}_{\boldsymbol{P}}) = \frac{1}{2}\sum_{i=1}^{\ell}\left(-2\sigma_i(\boldsymbol{P})\sigma_i(\boldsymbol{Q}) + \sigma_i^2(\boldsymbol{P})\right) + \tau\sum_{i=r+1}^{\ell}\sigma_i(\boldsymbol{P})$$

$$= \frac{1}{2}\left(\sum_{i=1}^{r}\left(-2\sigma_i(\boldsymbol{P})\sigma_i(\boldsymbol{Q}) + \sigma_i^2(\boldsymbol{P})\right) + \sum_{i=r+1}^{\ell}\left(-2\sigma_i(\boldsymbol{P})\sigma_i(\boldsymbol{Q}) + \sigma_i^2(\boldsymbol{P}) + 2\tau\sigma_i(\boldsymbol{P})\right)\right) \quad (\mathrm{a.11})$$

Since each $\sigma_i(\boldsymbol{P})$ is independent in Eq. (a.11) and it is easy to see that the minimum of $\mathcal{T}(\boldsymbol{\Sigma}_{\boldsymbol{P}})$ is obtained at $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}} = diag(\hat{\sigma}_1(\boldsymbol{P}), \cdots, \hat{\sigma}_\ell(\boldsymbol{P}))$, where $\hat{\sigma}_i(\boldsymbol{P})$ is defined as

$$\hat{\sigma}_i(\boldsymbol{P}) = \begin{cases} \sigma_i(\boldsymbol{Q}) & if\ i < r+1 \\ max(\sigma_i(\boldsymbol{Q}) - \tau, 0) & otherwise \end{cases}. \tag{a.12}$$

Hence, the solution of Eq. (42) is $\widehat{\boldsymbol{P}} = \boldsymbol{U}_{\boldsymbol{Q}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{P}} \boldsymbol{V}_{\boldsymbol{Q}}^T = \mathcal{D}_{\tau,\boldsymbol{W}}[\boldsymbol{Q}] = \boldsymbol{Q}_1 + \boldsymbol{U}_{\boldsymbol{Q}2} S_\tau[\boldsymbol{D}_{\boldsymbol{Q}2}] \boldsymbol{V}_{\boldsymbol{Q}2}^T$.