Some Results of Schatten-p Norm

1 Introduction

The affine rank minimization problem arises directly in various areas of science and engineering including statistics, machine learning, information theory, data mining, medical imaging and computer vision. Some representative applications include:

1. low rank matrix completion (LRMC)

$$\min_X rank\,(X) \quad s.t.\ X_{ij} = M_{ij}, (i,j) \in \Omega \tag{1}$$

   or

$$\min_X rank\,(X) \quad s.t.\ \mathcal{P}_\Omega\,(X) = M \tag{2}$$

2. robust principal component analysis (RPCA)

$$\min_E \|E\|_F^2 \quad s.t.\ rank\,(L) \leqslant r, X = L + E \tag{3}$$

   or

$$\min_E \|E\|_1 \quad s.t.\ rank\,(L) \leqslant r, X = L + E \tag{4}$$

3. low rank representation (LRR)

$$\min_Z rank\,(Z) \quad s.t.\ X = XZ + E \tag{5}$$

4. $\cdots$

Question: Why does low rank structure exist in high dimensional data?

2 Background

Definition 1. *The Schatten-p norm* $(0 < p < \infty)$ *of a matrix* $X \in \mathbb{R}^{m\times n}$ *(without loss of generality, we can assume that* $m \geq n$*) is defined as*

$$\|X\|_{S_p} = \left(\sum_{i=1}^{n} \sigma_i^p\,(X)\right)^{\frac{1}{p}} \tag{6}$$

where $\sigma_i(X)$ denotes the i-th singular value of X.

By $H\ddot{o}lder's\ \ and\ \ Minkowski's$ inequality, $p \geq 1, |\cdot|_p$ is a norm, for instance, it is nuclear norm when $p = 1$, it is Frobenius norm when $p = 2$ norm, however it is quasi-norm for $0 < p < 1$, for example:

Example 1.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{2} \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 0 & \dfrac{1}{2} \end{bmatrix} \tag{7}$$

By SVD, we can get the Σs as flowing:

$$\Sigma_A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \Sigma_B = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, (\Sigma_A + \Sigma_B) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

If $p = \dfrac{1}{2}$,

$$\begin{aligned} (\Sigma_A)_{\frac{1}{2}} &= \left(\sigma_1^{\frac{1}{2}} + \sigma_2^{\frac{1}{2}}\right)^2 = \left(1 + \left(\frac{1}{2}\right)^{\frac{1}{2}}\right)^2 = \frac{3}{2} + \sqrt{2} \\ (\Sigma_B)_{\frac{1}{2}} &= \left(\sigma_1^{\frac{1}{2}} + \sigma_2^{\frac{1}{2}}\right)^2 = \left(2^{\frac{1}{2}} + \left(\frac{1}{2}\right)^{\frac{1}{2}}\right)^2 = \frac{9}{2} \\ (\Sigma_{A+B})_{\frac{1}{2}} &= \left(\sigma_1^{\frac{1}{2}} + \sigma_2^{\frac{1}{2}}\right)^2 = \left(3^{\frac{1}{2}} + (1)^{\frac{1}{2}}\right)^2 = 4 + 2\sqrt{3} \end{aligned} \tag{9}$$

but

$$4 + 2\sqrt{3} = (\Sigma_A + \Sigma_B)_{\frac{1}{2}} \geqslant (\Sigma_A)_{\frac{1}{2}} + (\Sigma_B)_{\frac{1}{2}} = 6 + \sqrt{2} \tag{10}$$

Now, recall that $\|\cdot\|_p$ is norm if and only if

1. $\|0\|_p = 0$
2. $\|a\alpha\|_p = a|\alpha|_p \quad \forall a \in R$
3. $\|\alpha + \beta\|_p \leqslant |\alpha|_p + |\alpha|_p$

Due to the example above $\|\cdot\|_p$ where $0 < p < 1$ is not a norm.

But the fowling lemma give us a different view.

Lemma 1. *If $p \in (0, 1)$ and $a, b \geq 0$, then*

$$(a + b)^p \leqslant a^p + b^p \tag{11}$$

with equality if and only if either a or b is 0.

Proof. Define a function $f(t) = (1 + t)^p - 1 - t^p$, and it is easy to show that $f(t) \leqslant 0$, then let $t = \dfrac{a}{b}$. □

To recover a low-rank matrix from a small set of linear observations, $b \in \mathbb{R}^l$, the general $Schatten{-}p$ quasi-norm minimization (SQNM) problem is formulated as follows:

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_{S_p}^p, \quad s.t.\ \mathcal{A}(X) = b \tag{12}$$

where $\mathcal{A} : \mathbb{R}^{m\times n} \to R^l$ is a general linear operator.

Alternatively, the Lagrangian version of $Eq.$12 is

$$\min_{X \in \mathbb{R}^{m \times n}} \lambda \|X\|_{S_p}^p + f\left(\mathcal{A}\left(X\right) - b\right) \tag{13}$$

where $\lambda > 0$ is a regularization parameter, and the loss function $f(\cdot)$: $\mathbb{R}^l \to \mathbb{R}$ generally denote certain measurement for characterizing the loss $\mathcal{A}(X) - b$, Such as $\mathcal{A}$ is the linear projection operator $\mathcal{P}_\Omega$, and $f(\cdot) = \|\cdot\|_2^2$ in low rank matrix complement (LRMC); $\mathcal{A}$ is the identity operator and $f(\cdot) = \|\cdot\|_1$; $\mathcal{A}(x) = AX$ and $f(\cdot) = \|\cdot\|_F^2$ in multivariate regression.

Why the authors present their ideas in there papers?

1. Most exiting Schatten Norm minimization algorithms involve SVD or EVD of the whole matrix in each iteration, they suffer from a high computational cost of $O(n^2 m)$, which severely limits their application to large-scale problems.
2. They present a method which factorize X into two or more smaller factor matrices, i.e. $X = UV$ or $X = \prod_{i=1}^{M} U_i$, get a lower computational cost.

Questions:

1. Is $\|X\|_{S_p}^p$ a continuous function over p?
2. Does $\lim_{p\to 0} \|X\|_{S_p}^p$ exist?
3. If it exists, $\lim_{p\to 0} \|X\|_{S_p}^p \overset{?}{=} \|X\|_{S_0}^0 = rank(X)$

3 Main results

Theorem 1. *For any matrix $X \in \mathbb{R}^{m\times n}$ with the $rank(X) = r \le d$, it can be decomposed into the product of two much smaller matrices $U \in \mathbb{R}^{m\times d}$ and $V \in \mathbb{R}^{n\times d}$, i.e., $X = UV^T$, For any $0 < p \le 1, p_1 > 0$ and $p_2 > 0$ satisfying $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, then*

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m\times d}, V \in \mathbb{R}^{n\times d}, X = UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \tag{14}$$

Remark 1. *From Thm.1, it is very clear that for any $0 < p \leq 1$, and $p_1, p_2 > 0$, satisfying $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$, then the Schatten-p norm of any matrix X is equivalent to minimizing the product of the Schatten-p_1 and Schatten-p_2 norm of its two factor matrix.*

Naturally, $p_1 = p_2$ or $p_1 \neq p_2$, We discuss it respectively. if $p_1 = p_2 = 2p$

Theorem 2. *Given any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \leq d$, then the flowing equalities hold:*

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \left(\frac{\|U\|_{2p}^{2p} + \|V\|_{2p}^{2p}}{2} \right)^{\dfrac{1}{p}} \end{aligned} \tag{15}$$

Remark 2. *From the 2nd equation in Eq.15, for any $0 < p \leq 1$, the SQNM in many LRMC and low rank matrix recovery application can be transformed into the one of the minimizing the mean of the Schatten-2p norms of both much smaller factor matrices. Notice that $1 < 2p \leqslant 2$, i.e. $\dfrac{1}{2} < p \leqslant 1$, Schatten-2p are convex and smooth.*

When $p = 1$, and $p_1 = p_2 = 2$, the $Eq.15$ in $Thm.2$ become the flowing corollary.

Corollary 1. *Given any matrix $X \in \mathbb{R}^{m\times n}$ with the $rank(X) = r \leq d$, the flowing equalities hold:*

$$\begin{aligned} \|X\|_* &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_F \cdot \|V\|_F \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \frac{\|U\|_F^2 + \|V\|_F^2}{2} \end{aligned} \tag{16}$$

Remark 3. *The bilinear spectral penalty in the 2nd equality of Eq.16 has been wildly used in many LRMC and LRM recovery. Recall that the model of regularized singular value decomposition* [3]*:*

$$J_1 = \left\|X - UV^T\right\|_F^2 + \lambda \|U\|_F^2 + \lambda \|V\|_F^2 \tag{17}$$

when $p = \dfrac{1}{2}$, and by setting $p_1 = p_2 = 1$, we can get that

Corollary 2. *Given any matrix $X \in \mathbb{R}^{m\times n}$ with the $rank(X) = r \leq d$, the flowing equalities hold:*

$$\begin{aligned} \|X\|_{S_{\frac{1}{2}}} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_* \|V\|_* \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \left(\frac{\|U\|_* + \|V\|_*}{2} \right)^2 \end{aligned} \tag{18}$$

Remark 4. *The above equalities are known as the bi-nuclear quasi-norm.*

when $p = \dfrac{2}{3}$, and by setting $p_1 = p_2 = \dfrac{4}{3}$, we can get that

Corollary 3. *Given any matrix $X \in \mathbb{R}^{m\times n}$ with the $rank(X) = r \le d$, the flowing equalities hold:*

$$\begin{aligned} \|X\|_{S_{\frac{2}{3}}} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_{S_{\frac{4}{3}}} \|V\|_{S_{\frac{4}{3}}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_{S_{\frac{4}{3}}} \left(\|U\|_{S_{\frac{4}{3}}}^{\frac{4}{3}} + \|V\|_{S_{\frac{4}{3}}}^{\frac{4}{3}} \right)^{\frac{3}{2}} \end{aligned} \tag{19}$$

if $p_1 \neq p_2$

Theorem 3. *Given any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \le d$, and any$0 < p \le 1, p_1 > 0$ and $p_2 > 0$ satisfying $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, then the following equalities hold:*

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1} + p_1 \|V\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{\frac{1}{p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \left(\frac{\frac{\|U\|_{S_{p_1}}^{p_1}}{p_1} + \frac{\|V\|_{S_{p_2}}^{p_2}}{p_2}}{\dfrac{1}{p}} \right)^{\dfrac{1}{p}} \end{aligned} \tag{20}$$

Remark 5. *The SPQM problem can be transformed in to the one of minimizing the weighed sum of the Schatten-p_1 and Schatten-p_2 norm of two much smaller factor matrices, i.e.*

$$\frac{\dfrac{1}{p_1}}{\dfrac{1}{p}} = \frac{1}{p_1} \cdot p = \frac{1}{p_1} \cdot \frac{p_1 \cdot p_2}{p_1 + p_2} = \frac{p_2}{p_1 + p_2} \quad \textit{and} \quad \frac{\dfrac{1}{p_2}}{\dfrac{1}{p}} = \frac{p_1}{p_1 + p_2} \tag{21}$$

when $p = \dfrac{2}{3}$, and by setting $p_1 = 1$ and $p_2 = 2$, we can get that

Corollary 4. *Given any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \le d$, then*

$$\begin{aligned} \|X\|_{S_{\frac{2}{3}}} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \|U\|_* \|V\|_F \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}, X=UV^T} \left(\frac{2\|U\|_* + \|V\|_F^2}{3} \right)^{\frac{3}{2}} \end{aligned} \tag{22}$$

Remark 6. *It is called Frobenius/nuclear hybrid quasi-norm in the 2nd equality of Eq.22.*

when $p=\dfrac{2}{5}$, and by setting $p_1=\dfrac{1}{2} and p_2=2$, we can get that

Corollary 5. *Given any matrix $X\in\mathbb{R}^{m\times n}$ of $rank(X)=r\leq d$, then*

$$\begin{aligned}\|X\|_{S_{\frac{2}{5}}} &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{n\times d},X=UV^T} \|U\|_{S_{\frac{1}{2}}}\|V\|_F \\ &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{n\times d},X=UV^T} \left(\frac{4\|U\|_{S_{\frac{1}{2}}}^{\frac{1}{2}}+\|V\|_F^2}{5}\right)^{\frac{5}{2}}\end{aligned} \tag{23}$$

Generalization:

Theorem 4. *For any matrix $X\in\mathbb{R}^{m\times n}$ of $rank(X)=r\leq d$, it can be decomposed into the product of three much smaller matrices $U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}$ and $W\in\mathbb{R}^{n\times d}, X=UVW^T$. For any $0<p\leq 1$ and $p_i>0$ for all $i=1,2,3$, satisfying $\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}=\frac{1}{p}$, then*

$$\begin{aligned}\|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{d\times d},W\in\mathbb{R}^{n\times d}X=UVW^T} \|U\|_{S_{p_1}}\|V\|_{S_{p_2}}\|W\|_{S_{p_3}} \\ &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{d\times d},W\in\mathbb{R}^{n\times d}X=UVW^T} \left(\frac{p_2p_2\|U\|_{S_{p_1}}^{p_1}+p_1p_3\|V\|_{S_{p_2}}^{p_2}+p_1p_2\|W\|_{S_{p_3}}^{p_3}}{p_2p_3+p_1p_3+p_1p_2}\right)^{\frac{1}{p}} \\ &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{d\times d},W\in\mathbb{R}^{n\times d}X=UVW^T} \left(\frac{\frac{\|U\|_{S_{p_1}}^{p_1}}{p_1}+\frac{\|V\|_{S_{p_2}}^{p_2}}{p_2}+\frac{\|W\|_{S_{p_3}}^{p_3}}{p_3}}{\frac{1}{p}}\right)^{\dfrac{1}{p}}\end{aligned} \tag{24}$$

Remark 7. *$\forall 0<p\leq 1$ and $p_1,p_2,p_3>0$ satisfying $\dfrac{1}{p_1}+\dfrac{1}{p_2}+\dfrac{1}{p_3}=\dfrac{1}{p}$, the Schatten-p quasi-norm of any matrix is equivalent to minimizing the weighted sum of the Schatten-p_1 norm, Schatten-p_2 norm and Schatten-p_3 norm of these three much smaller factor matrices, where the weights of the three term are $\dfrac{p}{p_1}$, $\dfrac{p}{p_2}$ and $\dfrac{p}{p_3}$, respectively.*

By setting the same value for p_1,p_2 and $p_3, i.e.\ p_1=p_2=p_3=3p$, we give the following united scalable equivalent formulations for the Schatten-p quasi-norm.

Corollary 6. *Given any matrix $X\in\mathbb{R}^{m\times n}$ of $rank(X)=r\leq d$, then the flowing equalities hold:*

$$\begin{aligned}\|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{d\times d},W\in\mathbb{R}^{n\times d}X=UVW^T} \|U\|_{S_{3p}}\|V\|_{S_{3p}}\|W\|_{S_{3p}} \\ &= \min_{U\in\mathbb{R}^{m\times d},V\in\mathbb{R}^{d\times d},W\in\mathbb{R}^{n\times d}X=UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p}+\|V\|_{S_{3p}}^{3p}+\|W\|_{S_{3p}}^{3p}}{3}\right)^{\frac{1}{p}}\end{aligned} \tag{25}$$

Remark 8. *From the 2nd of Eq.25, we know that $\forall 0 < p < 1$, various SPQM problems in many LRMC and low rank matrix recovery applications can be transformed into the problems of minimizing the mean of the Schatten-3p norms of three much smaller factor matrices. And we note that when $1 \le 3p \le 3$, i.e. $\frac{1}{3} \le p \le 1$, the norms of the three factor matrices are convex and smooth.*

when $p = \frac{1}{3}$ and $p_1 = p_2 = p_3 = 1$, we can get the flowing corollary,

Corollary 7. *For any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \le d$, then the flowing equalities hold:*

$$
\begin{aligned}
\|X\|_{S_{\frac{1}{3}}} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d} X=UVW^T} \|U\|_*\|V\|_*\|W\|_* \\
&= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d} X=UVW^T} \left(\frac{\|U\|_* + \|V\|_* + \|W\|_*}{3}\right)^3
\end{aligned}
\tag{26}
$$

Similarly, we extend $Thm.4$ to the case of more factor matrices as follows.

Theorem 5. *For any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \le d$, it can be decomposed into the product of multiple much smaller matrices $U_i, i = 1, 2, ..., M$, i.e. $X = \prod_{i=1}^{M} U_i$. For any $0 < p \le 1$ and $p_i > 0$ for all $i = 1, 2, ..., M$. Satisfying $\sum_{i=1}^{M} \frac{1}{p_i} = \frac{1}{p}$, then*

$$
\begin{aligned}
\|X\|_{S_p} &= \min_{U_i: \prod_{i=1}^{M} U_i} \prod_{i=1}^{M} \|U_i\|_{S_{p_i}} \\
&= \|X\|_{S_p} = \min_{U_i: \prod_{i=1}^{M} U_i} \left(\frac{\dfrac{\sum_{i=1}^{M} \|U_i\|_{S_{p_i}}^{p_i}}{p_i}}{\dfrac{1}{p}} \right)^{\dfrac{1}{p}}
\end{aligned}
\tag{27}
$$

Corollary 8. *Given any matrix $X \in \mathbb{R}^{m\times n}$ of $rank(X) = r \le d$, then the flowing equalities hold:*

$$
\begin{aligned}
\|X\|_{S_p} &= \min_{U_i: X=\prod_{i=1}^{M} U_i} \prod_{i=1}^{M} \|U_i\|_{S_{M_p}} \\
&= \min_{U_i: X=\prod_{i=1}^{M} U_i} \left(\frac{\sum_{i=1}^{M} \|U_i\|_{S_{M_p}}^{M_p}}{M} \right)^{\frac{1}{p}}
\end{aligned}
\tag{28}
$$

Remark 9. *From $Thm.2$, we can get that $\forall\ 1 < (\leq)\ 2p$, ($i.e.$ $\frac{1}{2} < (\leqslant) p$), the Schatten-p quasi-norm of any matrix is equivalent to minimizing the mean of Schatten-2p norms of both factor matrix, as well as $Cor.6$, $\forall\ 1 < (\leq)\ 3p$, ($i.e.$ $\frac{1}{3} < (\leqslant) p$), the original SPNM can be transformed into a simpler one only involving the convex and smooth (convex but not smooth)norms of two or three factor matrices. To extend $Thm.2$ and $Corol.6$, then we can get $Cor.8$, $\forall\ 1 < (\leq)\ Mp$, i.e. $M = (\lfloor 1/p \rfloor + 1)$ where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $\frac{1}{p}$, then the original SPNM can be transformed into a simpler one only involving the convex and smooth (convex but not smooth)norms of M factor matrices. And we also set $p \leq 1$.*

4 Proof

We can get a simple general proof in [2], but it is not from special to general case.

Lemma 2 (Jensen's inequality). *Assume that the function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous concave function on $[0, +\infty)$. For all $t_i \geq 0$ satisfying $\sum_i t_i = 1$, and any $x_i \in \mathbb{R}^+$ for $i = 1, ..., n$, then*

$$g\left(\sum_{i=1}^{n} t_i x_i\right) \geqslant \sum_{i=1}^{n} t_i g\left(x_i\right) \tag{29}$$

Lemma 3 ($H\ddot{o}lder's$ inequality). *For any $p, q > 1$ satisfying $\frac{1}{p}+\frac{1}{q} = 1$, for any x_i and $y_i, i = 1, ..., n$*

$$\sum_{i=1}^{n} |x_i y_i| \leqslant \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q} \tag{30}$$

with equality if and only if there is a constant $c \neq 0$ such that each $x_i^p = c y_i^q$.

Lemma 4 (Young's inequality). *Let $a, b \geq 0$ and $1 < p, q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$, Then*

$$\frac{a^p}{p} + \frac{b^q}{q} \geqslant ab \tag{31}$$

with equality if and only if $a^p = b^q$.

Lemma 5. *Suppose that $Z \in \mathbb{R}^{m\times n}$ is a matrix of rank $r \leq min(m, n)$, and we denote its thin SVD by $Z = L_Z \sum_Z R_Z^T$, where $L_Z \in \mathbb{R}^{m\times r}$, $\sum_Z \in \mathbb{R}^{r\times r}$ and $R_Z \in \mathbb{R}^{n\times r}$. For any $A \in \mathbb{R}^{r\times r}$ satisfying $AA^T = A^T A = I_{r\times r}$, and the given $p(0 < p \leq 1)$, then $\left(A \sum_Z A^T\right)_{k,k} \geqslant 0$ for all $k = 1, ..., r$, and*

$$Tr^p\left(A \Sigma_Z A^T\right) \geqslant Tr^p\left(\Sigma_Z\right) = \|Z\|_{S_p}^p \tag{32}$$

where $Tr^p\left(B\right) = \sum_i B_{ii}^p$.

Proof. $\forall k \in \{1, ..., r\}$, we have $\left(A \Sigma_Z A^T\right)_{k,k} = \sum_i a_{k,i}^2 \sigma_i \geqslant 0$, where $\sigma_i \geq 0$ is the i-th sigular value of Z. Then

$$Tr^p\left(A \Sigma_Z A^T\right) = \sum_k \left(\sum_i a_{k,i}^2 \sigma_i\right)^p \tag{33}$$

Recall that $g(x) = x^p$ with $0 < p < 1$ $g'' < 0$ on $\mathbb{R}^+$, so it is a concave function on $\mathbb{R}^+$, by using the Jensen's inequality, as stated in Lemma 2, and $\sum_k a_{ki}^2 = 1$ for any $k \in \{1, 2, ...r\}$, we have that

$$\begin{aligned} Tr^p\left(A\Sigma_Z A^T\right) = \sum_k \left(\sum_i a_{ki}^2 \sigma_i\right)^p &\geqslant \sum_k \sum_i a_{ki}^2 \sigma_i^p \\ &= \sum_i \sum_k a_{ki}^2 \sigma_i^p \\ &= \sum_i \sigma_i^p \\ &= Tr^p\left(\sigma_Z\right) \\ &= \|Z\|_{S_p}^p \end{aligned} \tag{34}$$

In addition, when g(x)=x, i.e. $p = 1$, we obtain that

$$\left(\sum_i a_{ki}^2 \sigma_i\right)^p = \sum_i a_{ki}^2 \sigma_i \tag{35}$$

which means that the inequality (34) is still satisfied. □

Proof of Theorem 1

Proof. Let $U = L_U \Sigma_U R_U^T$ and $V = L_V \Sigma_V R_V^T$ be the thin SVDs of U and V, respectively, where $L_U \in \mathbb{R}^{m\times d}, L_V \in \mathbb{R}^{n\times d}$, and $R_U, \Sigma_U, R_V \Sigma_V \in \mathbb{R}^{d\times d}$. $X = L_x \Sigma_X R_X^T$, where the columns of $L_x \in \mathbb{R}^{m\times d}$ and $R_X \in \mathbb{R}^{n\times d}$ are the left and right singular vectors associated with the top d singular values of X with rank at most $r(r \leq d)$, and $\Sigma_X = diag([\sigma_1(X), \cdots, \sigma_r(X), 0, \cdots, 0]) \in \mathbb{R}^{d\times d}$.

Recall that $X = UV^T$, i.e. $L_X \Sigma_X R_X^T = L_U \Sigma_U R_U^T R_V \Sigma_V L_V^T$, then $\exists O_1 \widehat{O_1} \in \mathbb{R}^{d\times d}$ satisfy $L_X = L_U O_1$ and $L_U = L_X \widehat{O_1}$, which implies that $O_1 = L_U^T L_X$ and $\widehat{O_1} = L_X^T L_U$. Thus, $O_1 = \widehat{O_1}^T$. Since $L_X = L_U O_1 = L_X \widehat{O_1} O_1$, we have $\widehat{O_1} O_1{}^T = O_1{}^T O_1 = I_d$. Similarly, we have $O_1 \widehat{O_1} = O_1 O_1^T = I_d$. In addition,$\exists O_2 \in \mathbb{R}^{d\times d}$ satisfies $R_X = L_V O_2$ with $O_2 O_2^T = O_2^T O_2 = I_d$. Let $O_3 = O_2 O_1^T \in \mathbb{R}^{d\times d}$, then we have $O_3 O_3^T = O_3^T O_3 = I_d$, i.e. $\sum_i (O_3)_{ij}^2 = \sum_j (O_3)_{ij}^2 = 1$ for $i, j \in \{1, 2, ..., d\}$, where $a_{i,j}$ denotes the element of the matrix A in the i-th row and the j-th column. In addition, let $O_4 = R_U^T R_V$, we have $\sum_i (O_4)_{ij}^2 \leqslant 1$ and $\sum_j (O_4)_{ij}^2 \leqslant 1$ for $\forall i, j \in \{1, 2, ..., d\}$.

According to the above analysis, then we have $O_2 \Sigma_X O_2^T = O_2 O_1^T \Sigma_U O_4 \Sigma_V = O_3 \Sigma_U O_4 \Sigma_V$. Let ρ_i and τ_i denote the i-th and j-th diagonal elements of Σ_U and Σ_V, respectively. In flowing, we consider the two cases of p_1 and p_2, i.e. at least one of p_1 and p_2 must be no less than 1, or both of them are smaller than 1. It is clear that for any $\frac{1}{2} \leqslant p \leqslant 1$ and $p_1, p_2 > 0$ satisfying $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, at least one p_1 and p_2 must be no less than 1. On the other hand, only if $0 < p < \dfrac{1}{2}$, there exist $0 < p_1 < 1$ and $0 < p_2 < 1$ such that $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, i.e. both of them are smaller than 1.

1. Case 1: For any $\frac{1}{2} \leqslant p \leqslant 1$ and $p_1, p_2 > 0$ satisfying $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, at least one p_1 and p_2 must be no less than 1.
   Without loss of generality, we assume that that $p_2 \geq 1$. Here, we set $k_1 = \frac{p_1}{p}$ and $k_2 = \frac{p_2}{p}$,. Clearly, we can know that $k_1, k_2 > 1$ and $\frac{1}{k_1} + \frac{1}{k_2} = 1$. From Lemma 5, we have

$$\begin{aligned}
\|X\|_{S_p} &\leqslant \left(Tr^p\left(O_2 \Sigma_X O_2^T\right)\right)^{\frac{1}{p}} = \left(Tr^p\left(O_2 O_1^T \Sigma_U O_4 \sum_V\right)\right)^{\frac{1}{p}} = \left(Tr^p\left(O_3 \Sigma_U O_4 \Sigma_V\right)\right)^{\frac{1}{p}} \\
&= \left(\sum_{i=1}^{d}\left[\sum_{j=1}^{d} \tau_j (O_3)_{ij} (O_4)_{ji\rho_i}\right]^p\right)^{\frac{1}{p}} \\
&\underbrace{\leqslant}_{a} \left(\left[\sum_{i=1}^{d} (\rho_i^p)^{k_1}\right]\left[\sum_{i=1}^{d}\left(\sum_{j=1}^{d} \tau_j (O_3)_{ij} (O_4)_{ji}\right)^{p\times k_2}\right]^{\frac{1}{k_2}}\right)^{\frac{1}{p}} \\
&= \left(\sum_{i=1}^{d} \rho_i^{p_1}\right)^{\frac{1}{p_1}}\left[\sum_{i=1}^{d}\left(\sum_{j=1}^{d} \tau_j (O_3)_{ij} (O_4)_{ji}\right)^{p_2}\right]^{\frac{1}{p_2}} \\
&\underbrace{\leqslant}_{b} \left(\sum_{i=1}^{d} \rho_i^{p_1}\right)^{\frac{1}{p_1}}\left[\sum_{i=1}^{d}\left(\sum_{j=1}^{d} \tau_j \frac{(O_3)_{ij}^2 + (O_4)_{ji}^2}{2}\right)^{p_2}\right]^{\frac{1}{p_2}} \\
&\underbrace{\leqslant}_{c} \left(\sum_{i=1}^{d} \rho_i^{p_1}\right)^{\frac{1}{p_1}}\left[\tau_j^{p_2}\right]^{\frac{1}{p_2}}
\end{aligned} \tag{36}$$

   where the inequality $\underbrace{\leqslant}_{a}$ holds due to $H\ddot{o}lder's$ inequality, as stated in Lemma 3. In addition the inequality $\underbrace{\leqslant}_{b}$ follows from the basic inequality $xy \leqslant \frac{x^2+y^2}{2}$ for any real numbers x and y, and the inequality $\underbrace{\leqslant}_{c}$ relies on the facts that $\sum_i (O_3)_{ij}^2 = 1$ and $\sum_i (O_4)_{ji}^2 \leqslant 1$, and we apply the Jensen's inequality Lemma 2 for the convex function $h(x) = x^{p_2}$ with $p_2 \geq 1$.
   Thus, for any matrices $U \in \mathbb{R}^{m\times d}$ and $V \in \mathbb{R}^{n\times d}$ satisfying $X = UV^T$, we have

$$\|X\|_{S_p} \leqslant \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \tag{37}$$

   On the other hand, let $U_* = L_X \Sigma_X^{\frac{p}{p_1}}$, and $V_* = R_X \Sigma_X^{\frac{p}{p_2}}$, where Σ_X^p is entry-wise power to p, then we obtain

$$X = U_* V_*^T,\ \|U_*\|_{S_{p_1}}^{p_1} = \|V_*\|_{S_{p_2}}^{p_2} = \|X\|_{S_p}^p \quad with \ \frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} \tag{38}$$

   and

$$\|X\|_{S_p} = \left(Tr^p \Sigma_X\right)^{\frac{1}{p}} = \|U_*\|_{S_{p_1}} \|V_*\|_{S_{p_2}} \tag{39}$$

   Therefore, under the constraint $X = UV^T$, we have

$$\|X\|_{S_p} = \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \tag{40}$$

2. Case 2: For any $0 < p < \frac{1}{2}$, there exist $0 < \widehat{p_1} < 1$ and $0 < \widehat{p_2} < 1$ such that $\frac{1}{\widehat{p_1}} + \frac{1}{\widehat{p_1}} = \frac{1}{p}$. Naturally, for any given p, there must exist $p_1 > 0$ and $p_2 > 1$ such that $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$ and $\frac{1}{p_1} = \frac{1}{\widehat{p_1}} + \frac{1}{q}$ with $q \geq 1$. Clearly, we can know that $\frac{1}{\widehat{p_1}} \leqslant \frac{1}{p_1}$. Let $U^* = L_X \sum_X^{\frac{p}{p_1}}, V^* = R_X \sum_X^{\frac{p}{p_2}}, U_1^* = L_X \sum_X^{\frac{p}{\widehat{p_1}}}$ and $V_1^* = R_X \sum_X^{\frac{p}{\widehat{p_2}}}$, then we have

$$X = U^*(V^*)^T = U_1^*(V_1^*)^T \tag{41}$$

from which it follows that

$$\|X\|_{S_p} = \|U^*\|_{S_{p_1}} \|V^*\|_{S_{p_2}} = \|U_1^*\|_{S_{\widehat{p_1}}} \|V_1^*\|_{S_{\widehat{p_2}}} \tag{42}$$

Since $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{\widehat{p_1}} + \frac{1}{\widehat{p_2}}$ and $\frac{1}{p_1} = \frac{1}{\widehat{p_1}} + \frac{1}{q}$, then $\frac{1}{\widehat{p_2}} = \frac{1}{q} + \frac{1}{p_2}$. Consider any factor matrices U and V satisfying $X = UV^T, V = L^V \Sigma_V R_V^T$ is the thin SVD of V. Let $U_1 = UU_2^T$ and $V_1 = L_V \Sigma_V^{\frac{\widehat{p_2}}{p_2}}$, where $U_2^T = R_V \Sigma_V^{\frac{\widehat{p_2}}{q}}$, then it is easy to get that

$$\begin{aligned} & V = V_1 U_2, X = U_1 V_1^T \\ & \|V\|_{S_{\widehat{p_2}}} = \|U_2\|_{S_q} \|V_1\|_{S_{p_2}} \\ & \|U_1\|_{S_{p_1}} \leqslant \|U\|_{S_{\widehat{p_1}}} \|U_2\|_{S_q} \end{aligned} \tag{43}$$

where the above inequality follows from (37) with $q \geq 1$. Combining (42) and (43), for any U and V satisfying $X = UV^T$, we have

$$\begin{aligned} \|X\|_{S_p} &= \|U^*\|_{S_{p_1}} \|V^*\|_{S_{p_2}} \\ &\leq \|U_1\|_{S_{p_1}} \|V_1\|_{S_{p_2}} \\ &\leq \|U\|_{S_{\widehat{p_1}}} \|U_2\|_{S_q} \|V_1\|_{S_{p_2}} \\ &= \|U\|_{S_{\widehat{p_1}}} \|V\|_{S_{\widehat{p_2}}} \end{aligned} \tag{44}$$

where the 1st inequality follows from (37). Recall that

$$\|X\|_{S_p} = \|U_1^*\|_{S_{\widehat{p_1}}} \|V_1^*\|_{S_{\widehat{p_2}}} \tag{45}$$

Therefore, for any $0 < \widehat{p_1} < 1$ and $0 < \widehat{p_2} < 1$ satisfying $\frac{1}{p} = \frac{1}{\widehat{p_1}} + \frac{1}{\widehat{p_2}}$, and by (44) and (45). we also have

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X = UV^T} \|U\|_{S_{\widehat{p_1}}} \|V\|_{S_{\widehat{p_2}}} \tag{46}$$

In summary, for any $0 < p \leq 1, p_1 > 0$ and $p_2 > 0$ satisfying $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$, we have

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X = UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \tag{47}$$

□

Proof of Theorem 2

Proof. $\because p_1 = p_2 = 2p > 0$ and $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, and use Thm.1, we obtain that

$$\|X\|_{S_p} = \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} \tag{48}$$

Due to basic inequality $xy \leqslant \frac{x^2+y^2}{2}$ for any real numbers x and y, we have

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\|U\|_{S_{2p}}^p \|V\|_{S_{2p}}^p\right)^{\frac{1}{p}} \\ &\leq \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{\|U\|_{S_{2p}}^{2p} + \|V\|_{S_{2p}}^{2p}}{2}\right)^{\frac{1}{p}} \end{aligned} \tag{49}$$

Let $U_* = L_X \Sigma_X^{\frac{1}{2}}$ and $V_* = R_X \Sigma_X^{\frac{1}{2}}$, where $\Sigma_X^{\frac{1}{2}}$ is entry-wise power to $\frac{1}{2}$, then we obtain

$$X = U_* V_*^T, \|U_*\|_{S_{2p}}^{2p} = \|V_*\|_{S_{2p}}^{2p} = \|X\|_{S_p}^p \tag{50}$$

which implies that

$$\|X\|_{S_p}^p = \|U_*\|_{S_{2p}} \|V_*\|_{S_{2p}} = \left(\frac{\|U_*\|_{S_{2p}}^{2p} + \|V_*\|_{S_{2p}}^{2p}}{2}\right)^{\frac{1}{p}} \tag{51}$$

The Thm.2 now follows because

$$\min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} = \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{\|U\|_{S_{2p}}^{2p} + \|V\|_{S_{2p}}^{2p}}{2}\right)^{\frac{1}{p}} \tag{52}$$

□

Proof of Theorem 3

Proof. For any $0 < p \leq 1, p_1 > 0$ and $p_2 > 0$ satisfying $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$, and using Thm.1, we have

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\|U\|_{S_{p_1}}^p \|V\|_{S_{p_2}}^p\right)^{\frac{1}{p}} \\ &\leq \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{\|U\|_{S_{p_1}}^{pk_1}}{k_1} + \frac{\|V\|_{S_{p_2}}^{pk_2}}{k_2}\right)^{\frac{1}{p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1}}{p_1} + \frac{p_1 \|V\|_{S_{p_2}}^{p_2}}{p_2}\right)^{\frac{1}{p}} \end{aligned} \tag{53}$$

where the above inequality follows from Young's inequality in Lemma 4, and the monotone increasing property of the function $g(x) = x^{\frac{1}{p}}$.

Let $U_* = L_X \Sigma_X^{\frac{p}{p_1}}$ and $V_* = R_X \Sigma_X^{\frac{p}{p_2}}$, then $X = U_* V_*^T$. Using Thm.1, we have

$$\begin{aligned}
\|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\
&= \|U_*\|_{S_{p_1}} \|V_*\|_{S_{p_2}} \\
&= \left(\frac{p_2 \|U_*\|_{S_{p_1}}^{p_1} + p_1 \|V_*\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{\frac{1}{p}}
\end{aligned} \tag{54}$$

which implies that

$$\begin{aligned}
\|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\
&= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1} + p_1 \|V\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{\frac{1}{p}} \\
&= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{n\times d}: X=UV^T} \left(\frac{\frac{\|U\|_{S_{p_1}}^{p_1}}{p_1} + \frac{\|V\|_{S_{p_2}}^{p_2}}{p_2}}{\frac{1}{p}} \right)^{\frac{1}{p}}
\end{aligned} \tag{55}$$

□

Proof of Theorem 4

Proof. Let $U \in \mathbb{R}^{m\times d}$ and $\widehat{V} \in \mathbb{R}^{n\times d}$ be any factor matrices such that $X = U\widehat{V}^T$, and $\widehat{p_1} = p_1 > 0$ and $\widehat{p_2} = \frac{p_2 p_3}{p_2 + p_3} > 0$, which means that $\frac{1}{\widehat{p_1}} + \frac{1}{\widehat{p_2}} = \frac{1}{p}$. According to Thm.1, we obtain

$$\|X\|_{S_p} = \min_{U\in\mathbb{R}^{m\times d}, \widehat{V}\in\mathbb{R}^{n\times d}: X=U\widehat{V}^T} \|U\|_{S_{\widehat{p_1}}} \left\|\widehat{V}\right\|_{S_{\widehat{p_2}}} \tag{56}$$

Let $V \in \mathbb{R}^{d\times d}$ and $W \in \mathbb{R}^{n\times d}$ be factor matrices of $\widehat{V}$, i.e., $VW^T = \widehat{V}^T$. Since $\widehat{p_2} = \frac{p_2 p_3}{p_2 + p_3}$, then $\frac{1}{\widehat{p_2}} = \frac{1}{p_2} + \frac{1}{p_3}$. Using Thm.1, we also have

$$\|X\|_{S_p} = \min_{V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, \widehat{V}=(VW^T)^T} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \tag{57}$$

Combining (56) and (57), we obtain

$$\|X\|_{S_p} = \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \tag{58}$$

Using the above result, we have

$$\begin{aligned}
\|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\
&= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\|U\|_{S_{p_1}}^{p} \|V\|_{S_{p_2}}^{p} \|W\|_{S_{p_3}}^{p} \right)^{\frac{1}{p}} \\
&\le \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\frac{p_2 p_3 \|U\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{\frac{1}{p}}
\end{aligned} \tag{59}$$

where the above inequality follows from Lemma 4.

Let $U_* = L_X \Sigma_X^{\frac{p}{p_2}}, V = \Sigma_X^{\frac{p}{p_2}}$ and $W_* = R_X \Sigma_X^{\frac{p}{p_3}}$, it is easy to get $X = U_* V_* W_*^T$. Then we have

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\ &= \|U_*\|_{S_{p_1}} \|V_*\|_{S_{p_2}} \|W_*\|_{S_{p_1}} \\ &= \left(\frac{p_2 p_3 \|U_*\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V_*\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W_*\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{\frac{1}{p}} \end{aligned} \tag{60}$$

Therefore, we have

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\frac{p_2 p_3 \|U\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{\frac{1}{p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\frac{\frac{\|U\|_{S_{p_1}}^{p_1}}{p_1} + \frac{\|V\|_{S_{p_2}}^{p_2}}{p_2} + \frac{\|W\|_{S_{p_3}}^{p_3}}{p_3}}{\frac{1}{p}} \right)^{\frac{1}{p}} \end{aligned} \tag{61}$$

□

Proof of Corollary 6

Proof. Since $p_1 = p_2 = p_3 = 3p > 0$ and $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = \frac{1}{p}$, and using Thm.4, we have that

$$\|X\|_{S_p} = \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \tag{62}$$

From the basic inequality $xyz \leqslant \frac{x^3+y^3+z^3}{3}$ for any positive real numbers x,y and z, we obtain

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\|U\|_{S_{3p}}^p \|V\|_{S_{3p}}^p \|W\|_{S_{3p}}^p \right)^{\frac{1}{p}} \\ &\leq \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p} + \|V\|_{S_{3p}}^{3p} + \|W\|_{S_{3p}}^{3p}}{3} \right)^{\frac{1}{p}} \end{aligned} \tag{63}$$

Let $U_* = L_X \Sigma_X^{\frac{1}{3}}, V_* = \Sigma_X^{\frac{1}{3}}$ and $W_* = R_X \Sigma_X^{\frac{1}{3}}$, where $\Sigma_X^{\frac{1}{3}}$ is entry-wise power to $\frac{1}{3}$, then we have

$$X = U_* V_* W^T, \|U_*\|_{S_{3p}}^{3p} = \|V_*\|_{S_{3p}}^{3p} = \|W_*\|_{S_{3p}}^{3p} = \|X\|_{S_p}^p \tag{64}$$

so,

$$\begin{aligned} \|X\|_{S_p} &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U\in\mathbb{R}^{m\times d}, V\in\mathbb{R}^{d\times d}, W\in\mathbb{R}^{n\times d}, X=UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p} + \|V\|_{S_{3p}}^{3p} + \|W\|_{S_{3p}}^{3p}}{3} \right)^{\frac{1}{p}} \end{aligned} \tag{65}$$

□

5 Conclusions

In general, the SQNM is non-convex, non-smooth, and even non-Lipschitz. Can you give us some examples?

For any $0 < p \leq 1$, the SQNM can be transformed into an optimization problem only involving the smooth norms of multiple factor matrices.

6 Reference

[1] F. Shang, Y. Liu, and J. Cheng, "Unified Scalable Equivalent Formulations for Schatten Quasi-Norms," *arXiv:1606.00668* , 2016.

[2] C. Xu, Z. Lin and H. Zha, "A Unified Convex Surrogate for the Schatten-p Norm," *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* , 2017, pp. 926–932.

[3] S. Zheng, C. Ding and F. Nie, "Regularized Singular Value Decomposition and Application to Recommender System," *arXiv:1804.05090v1* , 2018.

[4] F. Shang, Y. Liu, and J. Cheng , " Tractable and Scalable Schatten Quasi-Norm Approximations for Rank Minimization," *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016,pp. 620–629.

[5] F. Shang, Y. Liu, and J. Cheng, "Scalable algorithms for tractable Schatten quasi-norm minimization," *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016, pp. 2016–2022.

[6] E. Elhamifar and R. Vidal, "Sparse Subspace Clustering: Algorithm, Theory, and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, Vol. 35(11), pp. 2765–2781.

[7] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust Recovery of Subspace Structures by Low-Rank Representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013, Vol. 35(1), pp. 171–184.

[8] H.Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "LRR for Subspace Segmentation via Tractable Schatten-p Norm Minimization and Factorization," *IEEE Transactions on Cybernetics*, 2018, pp. 1–13.

[9] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, S. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 2017, Vol. 39, pp. 818–832.

[10] F. Shang, Y. Liu, J. Cheng, Z. Luo, and Z. Lin , "Bilinear Factor Matrix Norm Minimization for Robust PCA: Algorithms and Applications," *IEEE Trans. Pattern Analysis and Machine Intelligence* , 2017.

August 9, 2018