

Unified Scalable Equivalent Formulations for Schatten Quasi-Norms

Fanhua Shang, *Member, IEEE*, Yuanyuan Liu, and James Cheng

Abstract

The Schatten quasi-norm ($0 < p < 1$) can be used to bridge the gap between the nuclear norm and rank function. However, most existing algorithms are too slow or even impractical for large-scale problems, due to the singular value decomposition (SVD) or eigenvalue decomposition (EVD) of the whole matrix in each iteration. In this paper, we rigorously prove that for any $p, p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$, the Schatten- p quasi-norm of any matrix is equivalent to the minimization of the product of the Schatten- p_1 norm (or quasi-norm) and Schatten- p_2 norm (or quasi-norm) of its two factor matrices. Then we present and prove the equivalence relationship between the product formula of the Schatten quasi-norm and its sum formula for the two cases of p_1 and p_2 : $p_1 = p_2$ and $p_1 \neq p_2$. Especially, when $p > 1/2$, there is an equivalence between the Schatten- p quasi-norm (or norm) of any matrix and the Schatten- $2p$ norms of its two factor matrices, where the widely used equivalent formulation of the nuclear norm, i.e., $\|X\|_* = \min_{X=UV^T} (\|U\|_F^2 + \|V\|_F^2)/2$, can be viewed as a special case. That is, various Schatten- p quasi-norm minimization problems with $p > 1/2$ can be transformed into the one only involving the smooth, convex norms of two factor matrices, which can lead to simpler and more efficient algorithms than conventional methods.

We further extend the theoretical results of two factor matrices to the cases of three and more factor matrices, from which we can see that for any $0 < p < 1$, the Schatten- p quasi-norm of any matrix is the minimization of the mean to the power of $(\lfloor 1/p \rfloor + 1)$ of the Schatten- $(\lfloor 1/p \rfloor + 1)p$ norms of all factor matrices, where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$. In other words, for any $0 < p < 1$, the Schatten- p quasi-norm minimization can be transformed into an optimization problem only involving the smooth, convex norms of multiple factor matrices. In addition, we also present some representative examples for two and three factor matrices. Naturally, the bi-nuclear and Frobenius/nuclear quasi-norms defined in our previous paper [1] and the tri-nuclear quasi-norm defined in our previous paper [2] are three important special cases.

Index Terms

Schatten quasi-norm, nuclear norm, rank function, factor matrix, equivalent formulations.

F. Shang, Y. Liu and J. Cheng are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: {fhshang, yliu, jcheng}@cse.cuhk.edu.hk.

CUHK Technical Report CSE-ShangLC20160307, March 7, 2016.

I. 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 low-rank matrix completion (LRMC) [3], robust principal component analysis (RPCA) [4], low-rank representation [5], multivariate regression [6], multi-task learning [7] and system identification [8]. To efficiently solve such problems, we mainly relax the rank function to its tractable convex envelope, i.e., the nuclear norm (sum of the singular values, also known as the trace norm or Schatten-1 norm), which leads to a convex optimization problem [3, 9, 10, 11]. In fact, the nuclear norm of one matrix is the ℓ_1 -norm of the vector of its singular values, and thus it can motivate a low-rank solution. However, it has been shown in [12, 13] that the ℓ_1 -norm over-penalizes large entries of vectors, and therefore results in a solution from a possibly biased solution space. Recall from the relationship between the ℓ_1 -norm and nuclear norm, the nuclear norm penalty shrinks all singular values equally, which also leads to over-penalize large singular values. That is, the nuclear norm may make the solution deviate from the original solution as the ℓ_1 -norm does. Compared with the nuclear norm, the Schatten- p quasi-norm with $0 < p < 1$ is non-convex, but it can give a closer approximation to the rank function. Thus, the Schatten- p quasi-norm minimization has received a significant amount of attention from researchers in various communities, such as images recovery [14, 15], collaborative filtering [16, 17] and MRI analysis [18].

Recently, two classes of iterative reweighted least squares (IRLS) algorithms in [19] and [20] were proposed to approximate associated Schatten- p quasi-norm minimization problems, respectively. In addition, Lu et al. [14] proposed a family of iteratively reweighted nuclear norm (IRNN) algorithms to solve various non-convex surrogate, including the Schatten quasi-norm, minimization problems. In [14, 15, 16, 21, 22], the Schatten- p quasi-norm has been shown to be empirically superior to the nuclear norm for many different problems. Moreover, [23] theoretically proved that Schatten- p quasi-norm minimization requires significantly fewer measurements than conventional nuclear norm minimization. However, existing algorithms mentioned above have to be solved iteratively and involve singular value decomposition (SVD) or eigenvalue decomposition (EVD) in each iteration. Thus they suffer from high computational cost and are even not applicable for large-scale problems [1, 2].

On the contrary, the nuclear norm has a scalable equivalent formulation, also known as the bilinear spectral penalty [11, 24, 25], which has been successfully applied in many large-scale applications, such

as collaborative filtering [17, 26, 27]. In addition, Zuo et al. [28] proposed a generalized shrinkage-thresholding operator to iteratively solve ℓ_p quasi-norm minimization with arbitrary p values, i.e., $0 \leq p < 1$. Since the Schatten- p quasi-norm of one matrix is equivalent to the ℓ_p quasi-norm on its singular values, we may naturally ask the following question: can we design a unified scalable equivalent formulation to the Schatten- p quasi-norm with arbitrary p values, i.e., $0 < p < 1$.

In this paper, we first present and prove the equivalence relationship between the Schatten- p quasi-norm of any matrix and the minimization of the product of the Schatten- p_1 norm (or quasi-norm) and Schatten- p_2 norm (or quasi-norm) of its two factor matrices, for any $p, p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$. In addition, we also prove the equivalence relationship between the product formula of the Schatten quasi-norm and its sum formula for the two cases of p_1 and p_2 : $p_1 = p_2$ and $p_1 \neq p_2$. When $p > 1/2$ and by setting the same value for p_1 and p_2 , there is an equivalence between the Schatten- p quasi-norm (or norm) of any matrix and the Schatten- $2p$ norms of its two factor matrices, where a representative example is the widely used equivalent formulation of the nuclear norm, i.e., $\|X\|_* = \min_{X=UV^T} (\|U\|_F^2 + \|V\|_F^2)/2$. In other words, various Schatten- p quasi-norm minimization problems with $p > 1/2$ can be transformed into the one only involving the smooth convex norms of two factor matrices, which can lead to simpler and more efficient algorithms than conventional methods [14, 15, 16, 19, 20, 21, 22].

We further extend the theoretical results of two factor matrices to the cases of three and more factor matrices, from which we can know that for any $0 < p < 1$, the Schatten- p quasi-norm of any matrix is equivalent to the minimization of the mean to the power of $(\lfloor 1/p \rfloor + 1)$ of the Schatten- $(\lfloor 1/p \rfloor + 1)p$ norms of all factor matrices, where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$. Note that the norms of all factor matrices are convex and smooth. Besides the theoretical results, we also present several representative examples for two and three factor matrices. Naturally, the bi-nuclear and Frobenius/nuclear quasi-norms defined in our previous paper [1] and the tri-nuclear quasi-norm defined in our previous paper [2] are three important special cases.

II. NOTATIONS AND 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*

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

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

When $p \geq 1$, Definition 1 defines a natural norm, for instance, the Schatten-1 norm is the so-called nuclear norm, $\|X\|_*$, and the Schatten-2 norm is the well-known Frobenius norm, whereas it defines a quasi-norm for $0 < p < 1$. As the non-convex surrogate for the rank function, the Schatten- p quasi-norm is the better approximation than the nuclear norm [23], analogous to the superiority of the ℓ_p quasi-norm to the ℓ_1 -norm [20, 29].

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

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_{S_p}^p, \quad \text{subject to } \mathcal{A}(X) = b \quad (2)$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^l$ is a general linear operator. Alternatively, the Lagrangian version of (2) is

$$\min_{X \in \mathbb{R}^{m \times n}} \lambda \|X\|_{S_p}^p + f(\mathcal{A}(X) - b) \quad (3)$$

where $\lambda > 0$ is a regularization parameter, and the loss function $f(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}$ generally denotes certain measurement for characterizing the loss $\mathcal{A}(X) - b$. For instance, \mathcal{A} is the linear projection operator \mathcal{P}_Ω , and $f(\cdot) = \|\cdot\|_2^2$ in LRMC problems [14, 19, 22, 30], where \mathcal{P}_Ω is the orthogonal projection onto the linear subspace of matrices supported on $\Omega := \{(i, j) | D_{ij} \text{ is observed}\}$: $\mathcal{P}_\Omega(D)_{ij} = D_{ij}$ if $(i, j) \in \Omega$ and $\mathcal{P}_\Omega(D)_{ij} = 0$ otherwise. In addition, for RPCA problems [4, 31, 32, 33], \mathcal{A} is the identity operator and $f(\cdot) = \|\cdot\|_1$. In the problem of multivariate regression [34], $\mathcal{A}(X) = AX$ with A being a given matrix, and $f(\cdot) = \|\cdot\|_F^2$. $f(\cdot)$ may be chosen as the Hinge loss in [24] or the ℓ_p quasi-norm in [16].

Generally, the Schatten- p quasi-norm minimization problem, such as (2) and (3), is non-convex, non-smooth and even non-Lipschitz [35]. So far, only few algorithms, such as IRLS [19, 20] and IRNN [14], have been developed to solve such challenging problems. However, since most existing Schatten- p quasi-norm minimization algorithms involve SVD or EVD of the whole matrix in each iteration, they suffer from a high computational cost of $O(n^2m)$, which severely limits their applicability to large-scale problems [1, 2]. While there have been many efforts towards fast SVD or EVD computation such as partial SVD [36], the performance of those methods is still unsatisfactory for many real applications [37]. Therefore, it is a very important problem that how to transform the challenging problems such as (2) and (3) into a more tractable one, which can be solved by simpler and more efficient algorithms.

III. MAIN RESULTS

In this section, we first present and prove the equivalence relationship between the Schatten- p quasi-norm of any matrix and the Schatten- p_1 and Schatten- p_2 quasi-norms (or norms) of its two factor matrices,

where $1/p = 1/p_1 + 1/p_2$ with any $p_1 > 0$ and $p_2 > 0$. Moreover, we prove the equivalence relationship between the product formula of the Schatten quasi-norm and its sum formula for the two cases of p_1 and p_2 : $p_1 = p_2$ and $p_1 \neq p_2$. For any $1/2 < p \leq 1$, the Schatten- p quasi-norm (or norm) of any matrix is equivalent to the minimization of the squared mean of the Schatten- $2p$ norms of both factor matrices, for instance $\|X\|_* = \min_{X=UV^T} (\|U\|_F^2 + \|V\|_F^2)/2$, which can lead to simpler and more efficient algorithms than conventional methods. Finally, we extend the theoretical results of two factor matrices to the cases of three and more factor matrices. We can see that for any $0 < p < 1$, the Schatten- p quasi-norm of any matrix is the minimization of the mean to the power of $(\lfloor 1/p \rfloor + 1)$ of the Schatten- $(\lfloor 1/p \rfloor + 1)p$ norms of all factor matrices, where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$.

A. Unified Schatten Quasi-Norm Formulations of Two Factor Matrices

Theorem 1. *For any matrix $X \in \mathbb{R}^{m \times n}$ with $\text{rank}(X) = r \leq d$, 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 $0 < p \leq 1$, $p_1 > 0$ and $p_2 > 0$ satisfying $1/p_1 + 1/p_2 = 1/p$, then*

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}}. \quad (4)$$

The detailed proof of Theorem 1 is provided in Section IV-A. From Theorem 1, it is very clear that for any $0 < p \leq 1$ and $p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$, then the Schatten- p quasi-norm (or norm) of any matrix X is equivalent to the minimization of the product of the Schatten- p_1 norm (or quasi-norm) and Schatten- p_2 norm (or quasi-norm) of its two factor matrices.

Naturally, we can see that p_1 and p_2 may have the same value, i.e., $p_1 = p_2 = 2p$, or different values, i.e., $p_1 \neq p_2$. Next, we discuss these two cases for p_1 and p_2 , i.e., $p_1 = p_2$ and $p_1 \neq p_2$.

1) *Case of $p_1 = p_2$:* First, we discuss the case when $p_1 = p_2$. In fact, for any given $0 < p \leq 1$, there exist infinitely many pairs of positive numbers p_1 and p_2 satisfying $1/p_1 + 1/p_2 = 1/p$, such that the equality (4) holds. By setting the same value for p_1 and p_2 , i.e., $p_1 = p_2 = 2p$, we give a unified scalable equivalent formulation for the Schatten- p quasi-norm (or norm) as follows.

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

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_{S_{2p}} + \|V\|_{S_{2p}}}{2} \right)^2. \end{aligned} \quad (5)$$

Remark 1. The detailed proof of Corollary 1 is provided in Section IV-B. From the second equality in (5), we know that, for any $0 < p \leq 1$, the Schatten- p quasi-norm (or norm) minimization problems in many low-rank matrix completion and recovery applications can be transformed into the one of minimizing the mean of the Schatten- $2p$ norms (or quasi-norms) of both much smaller factor matrices. We note that when $1/2 < p \leq 1$, the norms of both much smaller factor matrices are convex and smooth due to $2p > 1$, which can lead to simpler and more efficient algorithms than conventional methods [14, 15, 16, 19, 20, 21, 22].

When $p=1$ and $p_1=p_2=2$, the equalities in Corollary 1 become the following forms.

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

$$\begin{aligned} \|X\|_* &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}, X=UV^T} \|U\|_F \|V\|_F \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}, X=UV^T} \left(\frac{\|U\|_F + \|V\|_F}{2} \right)^2 \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}, X=UV^T} \frac{\|U\|_F^2 + \|V\|_F^2}{2}. \end{aligned} \quad (6)$$

The bilinear spectral penalty in the third equality of (6) has been widely used in many low-rank matrix completion and recovery problems, such as collaborative filtering [11, 24], RPCA [38], online RPCA [?], and image recovery [39]. Note that the well-known equivalent formulations of the nuclear norm in Corollary 2 are just a special case of Corollary 1, i.e., $p=1$ and $p_1=p_2=2$. In the following, we give two more representative examples for the case of $p_1=p_2$.

Example 1: When $p=1/2$, and by setting $p_1=p_2=1$ and using Theorem 1, we have

$$\|X\|_{S_{1/2}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}, X=UV^T} \|U\|_* \|V\|_*.$$

Due to the basic inequality $xy \leq [(x+y)/2]^2$ for any real numbers x and y , we obtain

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

Let $U_* = L_X \Sigma_X^{1/2}$ and $V_* = R_X \Sigma_X^{1/2}$ as in [1, 2], then we have $X = U_* V_*^T$ and

$$\|X\|_{S_{1/2}} = (\text{Tr}^{1/2}(\Sigma_X))^2 = \|U_*\|_* \|V_*\|_* = \left(\frac{\|U_*\|_* + \|V_*\|_*}{2} \right)^2$$

where $\text{Tr}^{1/2}(\Sigma_X) = \sum_i (\Sigma_X)_{i,i}^{1/2}$. Therefore, under the constraint $X = UV^T$, we have the following property [1, 2].

Property 1.

$$\begin{aligned}\|X\|_{S_{1/2}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_* \|V\|_* \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_* + \|V\|_*}{2} \right)^2.\end{aligned}\quad (7)$$

In our previous papers [1, 2], the scalable formulations in the above equalities are known as the bi-nuclear quasi-norm. In other words, the bi-nuclear quasi-norm is also a special case of Corollary 1, i.e., $p=1/2$ and $p_1=p_2=1$.

Example 2: When $p=2/3$, and by setting $p_1=p_2=4/3$ and using Theorem 1, we have

$$\|X\|_{S_{2/3}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{4/3}} \|V\|_{S_{4/3}}.$$

Due to the basic inequality $xy \leq (\frac{x+y}{2})^2$ for any real numbers x and y , then

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

Let $U_\star = L_X \Sigma_X^{1/2}$ and $V_\star = R_X \Sigma_X^{1/2}$, then we have $X = U_\star V_\star^T$ and

$$\|X\|_{S_{2/3}} = (\text{Tr}^{2/3}(\Sigma_X))^{3/2} = \|U_\star\|_{S_{4/3}} \|V_\star\|_{S_{4/3}} = \left(\frac{\|U_\star\|_{S_{4/3}} + \|V_\star\|_{S_{4/3}}}{2} \right)^2.$$

Together with the constraint $X=UV^T$, thus we have the following property.

Property 2.

$$\begin{aligned}\|X\|_{S_{2/3}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{4/3}} \|V\|_{S_{4/3}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_{S_{4/3}} + \|V\|_{S_{4/3}}}{2} \right)^2.\end{aligned}\quad (8)$$

2) *Case of $p_1 \neq p_2$:* In this part, we discuss the case of $p_1 \neq p_2$. Deferent from the case of $p_1 = p_2$, we may set infinitely many different values for p_1 and p_2 . For any given $0 < p \leq 1$, there must exist $p_1, p_2 > 0$, at least one of which is no less than 1 (which means that the norm of at least one factor matrix is convex), such that $1/p = 1/p_1 + 1/p_2$. Indeed, for any $0 < p \leq 1$, the values of p_1 and p_2 may be different, e.g., $p_1 = 1$ and $p_2 = 2$, thus we give the following unified scalable equivalent formulations for the Schatten- p quasi-norm (or norm).

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

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1} + p_1 \|V\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{1/p}. \end{aligned} \quad (9)$$

Remark 2. The detailed proof of Corollary 3 is given in Section IV-C. From Corollary 3, we know that Corollaries 1 and 2 can be viewed as two special cases of Corollary 3, i.e., $p_1 = p_2 = 2p$ and $p_1 = p_2 = 2$, respectively. That is, Corollary 3 is the more general form of Corollaries 1 and 2. From the second equality in (9), we can see that, for any $0 < p \leq 1$, the Schatten- p quasi-norm (or norm) minimization problem can be transformed into the one of minimizing the weighted sum of the Schatten- p_1 norm (or quasi-norm) and Schatten- p_2 norm (or quasi-norm) of two much smaller factor matrices, where the weights of the two terms in the second equality of (9) are $p_2/(p_1+p_2)$ and $p_1/(p_1+p_2)$, respectively.

In the following, we give two representative examples for the case of $p_1 \neq p_2$.

Example 3: When $p = 2/3$, and by setting $p_1 = 1$ and $p_2 = 2$, and using Theorem 1, then

$$\|X\|_{S_{2/3}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_* \|V\|_F.$$

In addition, we have

$$\begin{aligned} \|U\|_* \|V\|_F &= \sqrt{\|U\|_*} \sqrt{\|U\|_* \|V\|_F} \\ &\stackrel{a \leq}{\leq} \left(\frac{\sqrt{\|U\|_*} + \sqrt{\|U\|_*} + \|V\|_F}{3} \right)^3 \\ &= \left(\frac{2\sqrt{\|U\|_*} + \sqrt{\|V\|_F^2}}{3} \right)^3 \\ &\stackrel{b \leq}{\leq} \left(\frac{2\|U\|_* + \|V\|_F^2}{3} \right)^{3/2} \end{aligned}$$

where the inequality $a \leq$ holds due to the fact that $x_1 x_2 x_3 \leq [(x_1 + x_2 + x_3)/3]^3$ for any real numbers x_1, x_2 and x_3 , and the inequality $b \leq$ follows from the Jensen's inequality for the concave function $g(x) = x^{1/2}$.

Let $U_* = L_X \Sigma_X^{2/3}$ and $V_* = R_X \Sigma_X^{1/3}$ as in [1], then we have $X = U_* V_*^T$ and

$$\|X\|_{S_{2/3}} = (\text{Tr}^{2/3}(\Sigma_X))^{3/2} = \|U_*\|_* \|V_*\|_F = \left(\frac{2\|U_*\|_* + \|V_*\|_F^2}{3} \right)^{3/2}.$$

Therefore, together with the constraint $X = UV^T$, we have the following property [1].

Property 3.

$$\begin{aligned}
\|X\|_{S_{2/3}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_* \|V\|_F \\
&= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{2\|U\|_* + \|V\|_F^2}{3} \right)^{3/2}.
\end{aligned} \tag{10}$$

In our previous paper [1], the scalable formulations in the above equalities are known as the Frobenius/nuclear hybrid quasi-norm. It is clear that the Frobenius/nuclear hybrid quasi-norm is also a special case of Corollary 3, i.e., $p=2/3$, $p_1=1$ and $p_2=2$. As shown in the above representative examples and our previous papers [1, 2], we can design more efficient algorithms to solve the Schatten- p quasi-norm with $1/2 \leq p < 1$ than conventional methods [14, 15, 16, 19, 20, 21, 22].

Example 4: When $p=2/5$, and by setting $p_1=1/2$ and $p_2=2$, and using Theorem 1, we have

$$\|X\|_{S_{2/5}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{1/2}} \|V\|_F.$$

Moreover,

$$\begin{aligned}
\|U\|_{S_{1/2}} \|V\|_F &= \left(\|U\|_{S_{1/2}}^{1/4} \right)^4 \|V\|_F \\
&\stackrel{a \leq}{\leq} \left(\frac{4\sqrt{\|U\|_{S_{1/2}}^{1/2} + \sqrt{\|V\|_F^2}}}{5} \right)^5 \\
&\stackrel{b \leq}{\leq} \left(\frac{4\|U\|_{S_{1/2}}^{1/2} + \|V\|_F^2}{5} \right)^{5/2}
\end{aligned}$$

where the inequality $a \leq$ holds due to the familiar inequality of arithmetic and geometric means, and the inequality $b \leq$ follows from the Jensen's inequality for the concave function $g(x)=x^{1/2}$.

Let $U_* = L_X \Sigma_X^{4/5}$ and $V_* = R_X \Sigma_X^{1/5}$, then we have $X = U_* V_*^T$ and

$$\|X\|_{S_{2/5}} = (\text{Tr}^{2/5}(\Sigma_X))^{5/2} = \|U_*\|_{S_{1/2}} \|V_*\|_F = \left(\frac{4\|U_*\|_{S_{1/2}}^{1/2} + \|V_*\|_F^2}{5} \right)^{5/2}.$$

With the constraint $X=UV^T$, thus we have the following property.

Property 4.

$$\begin{aligned}
\|X\|_{S_{2/5}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{1/2}} \|V\|_F \\
&= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{4\|U\|_{S_{1/2}}^{1/2} + \|V\|_F^2}{5} \right)^{5/2}.
\end{aligned} \tag{11}$$

B. Extensions to Multiple Factor Matrices

Theorem 2. For any matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$, 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 $W \in \mathbb{R}^{n \times d}$, i.e., $X = UVW^T$. For any $0 < p \leq 1$ and $p_i > 0$ for all $i = 1, 2, 3$, satisfying $1/p_1 + 1/p_2 + 1/p_3 = 1/p$, then

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}}. \quad (12)$$

The detailed proof of Theorem 2 is provided in Section IV-D. From Theorem 2, we can see that for any $0 < p \leq 1$ and $p_1, p_2, p_3 > 0$ satisfying $1/p_1 + 1/p_2 + 1/p_3 = 1/p$, the Schatten- p quasi-norm (or norm) of any matrix is equivalent to the minimization of the product of the Schatten- p_1 norm (or quasi-norm), Schatten- p_2 norm (or quasi-norm) and Schatten- p_3 norm (or quasi-norm) of these three much smaller factor matrices. Similarly, we extend Theorem 2 to the case of more factor matrices as follows.

Theorem 3. For any matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$, it can be decomposed into the product of multiple much smaller matrices U_i , $i = 1, 2, \dots, M$, i.e., $X = \prod_{i=1}^M U_i$. For any $0 < p \leq 1$ and $p_i > 0$ for all $i = 1, 2, \dots, M$, satisfying $\sum_{i=1}^M 1/p_i = 1/p$, then

$$\|X\|_{S_p} = \min_{X = \prod_{i=1}^M U_i} \prod_{i=1}^M \|U_i\|_{S_{p_i}}. \quad (13)$$

The proof of Theorem 3 is very similar to that of Theorem 2 and is thus omitted. Similar to the case of two factor matrices, for any given $0 < p \leq 1$, there exist infinitely many positive numbers p_1 , p_2 and p_3 such that $1/p_1 + 1/p_2 + 1/p_3 = 1/p$, and the equality (12) holds. 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 unified scalable equivalent formulations for the Schatten- p quasi-norm (or norm).

Corollary 4. Given any matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$, then the following equalities hold:

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \left(\frac{\|U\|_{S_{3p}} + \|V\|_{S_{3p}} + \|W\|_{S_{3p}}}{3} \right)^3. \end{aligned} \quad (14)$$

Remark 3. The detailed proof of Corollary 4 is provided in Section IV-E. From the second equality in (14), we know that, for any $0 < p < 1$, various Schatten- p quasi-norm minimization problems in many low-rank matrix completion and recovery applications can be transformed into the problem of minimizing the mean of the Schatten- $3p$ norms (or quasi-norms) of three much smaller factor matrices. In addition, we note that when $1/3 < p \leq 1$, the norms of the three factor matrices are convex and smooth due to $3p > 1$, which can also lead to some simpler and more efficient algorithms than conventional methods.

Example 5: In the following, we give a representative example. When $p=1/3$ and $p_1=p_2=p_3=3$, the equalities in Corollary 4 become the following forms [2].

Property 5. For any matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$, then the following equalities hold:

$$\begin{aligned} \|X\|_{S_{1/3}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X=UVW^T} \|U\|_* \|V\|_* \|W\|_* \\ &= \min_{U, V, W: X=UVW^T} \left(\frac{\|U\|_* + \|V\|_* + \|W\|_*}{3} \right)^3. \end{aligned} \quad (15)$$

From Property 5, we can see that the tri-nuclear quasi-norm defined in our previous paper [2] is also a special case of Corollary 4.

Remark 4. From Corollary 2, we know that for any $1/2 < p \leq 1$, the Schatten- p quasi-norm (or norm) of any matrix is equivalent to the minimization of the squared mean of the Schatten- $2p$ norms of both factor matrices, as well as Corollary 4 for any $1/3 < p \leq 1$. In other words, if $1/2 < p \leq 1$ or $1/3 < p \leq 1$, the original Schatten- p quasi-norm (or norm) minimization problem can be transformed into a simpler one only involving the convex and smooth norms of two or three factor matrices. In addition, we extend the results of Corollary 2 and Corollary 4 to the case of more factor matrices, as shown in Corollary 5 below. The proof of Corollary 5 is very similar to that of Corollary 4 and is thus omitted. In other words, for any $0 < p < 1$, the Schatten- p quasi-norm of any matrix can theoretically be equivalent to the minimization of the mean to the power of M of the Schatten- (Mp) norms of all M factor matrices, where $M = (\lfloor 1/p \rfloor + 1)$ and $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$.

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

$$\begin{aligned} \|X\|_{S_p} &= \min_{X=\prod_{i=1}^M U_i} \prod_{i=1}^M \|U_i\|_{S_{Mp}} \\ &= \min_{X=\prod_{i=1}^M U_i} \left(\frac{\sum_{i=1}^M \|U_i\|_{S_{Mp}}}{M} \right)^M. \end{aligned} \quad (16)$$

IV. PROOFS

In this section, we give the detailed proofs for some important theorems and corollaries. We first introduce several important inequalities, such as the Jensen's inequality, Hölder's inequality and Young's inequality, that we use throughout our proofs.

Lemma 1 (Jensen's inequality). *Assume that the function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ 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 $i = 1, \dots, n$, then*

$$g\left(\sum_{i=1}^n t_i x_i\right) \geq \sum_{i=1}^n t_i g(x_i). \quad (17)$$

Lemma 2 (Hölder's inequality). *For any $p, q > 1$ satisfying $1/p + 1/q = 1$, then for any x_i and y_i , $i = 1, \dots, n$,*

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q} \quad (18)$$

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

Lemma 3 (Young's inequality). *Let $a, b \geq 0$ and $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. Then*

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab \quad (19)$$

with equality iff $a^p = b^q$.

A. Proof of Theorem 1

Before giving a complete proof for Theorem 1, we first present and prove the following lemma.

Lemma 4. *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 $Z = L_Z \Sigma_Z R_Z^T$, where $L_Z \in \mathbb{R}^{m \times r}$, $R_Z \in \mathbb{R}^{n \times r}$ and $\Sigma_Z \in \mathbb{R}^{r \times r}$. 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 $(A \Sigma_Z A^T)_{k,k} \geq 0$ for all $k = 1, \dots, r$, and*

$$\text{Tr}^p(A \Sigma_Z A^T) \geq \text{Tr}^p(\Sigma_Z) = \|Z\|_{S_p}^p,$$

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

Proof: For any $k \in \{1, \dots, r\}$, we have $(A \Sigma_Z A^T)_{k,k} = \sum_i a_{ki}^2 \sigma_i \geq 0$, where $\sigma_i \geq 0$ is the i -th singular value of Z . Then

$$\text{Tr}^p(A \Sigma_Z A^T) = \sum_k \left(\sum_i a_{ki}^2 \sigma_i\right)^p. \quad (20)$$

Recall that $g(x) = x^p$ with $0 < p < 1$ is a concave function on \mathbb{R}^+ . By using the Jensen's inequality [40], as stated in Lemma 1, and $\sum_i a_{ki}^2 = 1$ for any $k \in \{1, \dots, r\}$, we have

$$\left(\sum_i a_{ki}^2 \sigma_i\right)^p \geq \sum_i a_{ki}^2 \sigma_i^p.$$

Using the above inequality and $\sum_k a_{ki}^2 = 1$ for any $i \in \{1, \dots, r\}$, (20) can be rewritten as

$$\begin{aligned}
\text{Tr}^p(A\Sigma_Z A^T) &= \sum_k \left(\sum_i a_{ki}^2 \sigma_i \right)^p \\
&\geq \sum_k \sum_i a_{ki}^2 \sigma_i^p \\
&= \sum_i \sigma_i^p \\
&= \text{Tr}^p(\Sigma_Z) = \|Z\|_{S_p}^p.
\end{aligned} \tag{21}$$

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

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

which means that the inequality (21) is still satisfied. This completes the proof. \blacksquare

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 $\Sigma_X = \text{diag}([\sigma_1(X), \dots, \sigma_r(X), 0, \dots, 0]) \in \mathbb{R}^{d \times d}$.

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 $\widehat{O}_1 O_1 = O_1^T O_1 = I_d$. 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 $\forall i, j \in \{1, 2, \dots, d\}$, 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 $\sum_i (O_4)_{ij}^2 \leq 1$ and $\sum_j (O_4)_{ij}^2 \leq 1$ for $\forall i, j \in \{1, 2, \dots, d\}$.

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_j denote the i -th and the j -th diagonal elements of Σ_U and Σ_V , respectively. In the following, 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 $1/2 \leq p \leq 1$ and $p_1, p_2 > 0$ satisfying $1/p_1 + 1/p_2 = 1/p$, at least one of p_1 and p_2 must be no less than 1. On the other hand, only if $0 < p < 1/2$, there exist $0 < p_1 < 1$ and $0 < p_2 < 1$ such that $1/p_1 + 1/p_2 = 1/p$, i.e., both of them are smaller than 1.

Case 1. For any $0 < p \leq 1$, there exist $p_1 > 0$ and $p_2 > 0$, at least one of which is no less than 1, such that $1/p_1 + 1/p_2 = 1/p$. Without loss of generality, we assume that $p_2 \geq 1$. Here, we set $k_1 = p_1/p$ and $k_2 = p_2/p$. Clearly, we can know that $k_1, k_2 > 1$ and $1/k_1 + 1/k_2 = 1$. From Lemma 4, we have

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

where the inequality $a \leq$ holds due to the Hölder's inequality [40], as stated in Lemma 2. In addition, the inequality $b \leq$ follows from the basic inequality $xy \leq \frac{x^2+y^2}{2}$ for any real numbers x and y , and the inequality $c \leq$ relies on the facts that $\sum_i (O_3)_{ij}^2 = 1$ and $\sum_i (O_4)_{ji}^2 \leq 1$, and we apply the Jensen's inequality (see Lemma 1) 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

$$\|X\|_{S_p} \leq \|U\|_{S_{p_1}} \|V\|_{S_{p_2}}. \quad (22)$$

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

$$X = U_\star V_\star^T, \quad \|U_\star\|_{S_{p_1}}^{p_1} = \|V_\star\|_{S_{p_2}}^{p_2} = \|X\|_{S_p}^p \quad \text{with} \quad 1/p = 1/p_1 + 1/p_2,$$

and

$$\|X\|_{S_p} = (\text{Tr}^p(\Sigma_X))^{1/p} = \|U_\star\|_{S_{p_1}} \|V_\star\|_{S_{p_2}}.$$

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

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

Case 2. For any $0 < p < 1/2$, there exist $0 < \widehat{p}_1 < 1$ and $0 < \widehat{p}_2 < 1$ such that $1/\widehat{p}_1 + 1/\widehat{p}_2 = 1/p$. Next we prove that the result of Theorem 1 also holds. Naturally, for any given p , there must exist $p_1 > 0$ and $p_2 \geq 1$ such that $1/p_1 + 1/p_2 = 1/p$ and $1/p_1 = 1/\widehat{p}_1 + 1/q$ with $q \geq 1$. Clearly, we can know that $1/\widehat{p}_1 < 1/p_1$. Let $U^* = L_X \Sigma_X^{p/p_1}$, $V^* = R_X \Sigma_X^{p/p_2}$, $U_1^* = L_X \Sigma_X^{p/\widehat{p}_1}$ and $V_1^* = R_X \Sigma_X^{p/\widehat{p}_2}$, then we have

$$X = U^*(V^*)^T = U_1^*(V_1^*)^T$$

from which it follows that

$$\|X\|_{S_p} = \|U^*\|_{S_{p_1}} \|V^*\|_{S_{p_2}} = \|U_1^*\|_{S_{\widehat{p}_1}} \|V_1^*\|_{S_{\widehat{p}_2}}. \quad (23)$$

Since $1/p = 1/p_1 + 1/p_2 = 1/\widehat{p}_1 + 1/\widehat{p}_2$ and $1/p_1 = 1/\widehat{p}_1 + 1/q$, then $1/\widehat{p}_2 = 1/q + 1/p_2$. 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 $V_1 = L_V \Sigma_V^{\widehat{p}_2/p_2}$, where $U_2^T = R_V \Sigma_V^{\widehat{p}_2/q}$, then it is not difficult to verify that

$$\begin{aligned} V &= V_1 U_2, \quad X = U_1 V_1^T, \\ \|V\|_{S_{\widehat{p}_2}} &= \|U_2\|_{S_q} \|V_1\|_{S_{p_2}}, \\ \|U_1\|_{S_{p_1}} &\leq \|U\|_{S_{\widehat{p}_1}} \|U_2\|_{S_q} \end{aligned} \quad (24)$$

where the above inequality follows from (22) with $q \geq 1$. Combining (23) and (24), for any U and V satisfying $X = UV^T$, we have

$$\begin{aligned} \|X\|_{S_p} &= \|U^*\|_{S_{p_1}} \|V^*\|_{S_{p_2}} \\ &\leq \|U_1\|_{S_{p_1}} \|V_1\|_{S_{p_2}} \\ &\leq \|U\|_{S_{\widehat{p}_1}} \|U_2\|_{S_q} \|V_1\|_{S_{p_2}} \\ &= \|U\|_{S_{\widehat{p}_1}} \|V\|_{S_{\widehat{p}_2}} \end{aligned} \quad (25)$$

where the first inequality follows from (22). Recall that

$$\|X\|_{S_p} = \|U_1^*\|_{S_{\widehat{p}_1}} \|V_1^*\|_{S_{\widehat{p}_2}}. \quad (26)$$

Therefore, for any $0 < \widehat{p}_1 < 1$ and $0 < \widehat{p}_2 < 1$ satisfying $1/p = 1/\widehat{p}_1 + 1/\widehat{p}_2$, and by (25) and (26), we also have

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{\widehat{p}_1}} \|V\|_{S_{\widehat{p}_2}}.$$

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

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

This completes the proof. ■

B. Proof of Corollary 1

Proof: Because $p_1 = p_2 = 2p > 0$, and using Theorem 1, we obtain

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

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

$$\begin{aligned} & \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} \\ &\leq \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_{S_{2p}} + \|V\|_{S_{2p}}}{2} \right)^2. \end{aligned}$$

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

$$X = U_\star V_\star^T, \quad \|U_\star\|_{S_{2p}}^{2p} = \|V_\star\|_{S_{2p}}^{2p} = \|X\|_{S_p}^p,$$

which implies that

$$\|U_\star\|_{S_{2p}} \|V_\star\|_{S_{2p}} = \left(\frac{\|U_\star\|_{S_{2p}} + \|V_\star\|_{S_{2p}}}{2} \right)^2.$$

The theorem now follows because

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

This completes the proof. ■

C. Proof of Corollary 3

Proof: For any $0 < p \leq 1$, $p_1 > 0$ and $p_2 > 0$ satisfying $1/p_1 + 1/p_2 = 1/p$, and using Theorem 1, we have

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

Let $k_1 = (p_1 + p_2)/p_2$ and $k_2 = (p_1 + p_2)/p_1$, we can know that $1/k_1 + 1/k_2 = 1$. Then

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

where the above inequality follows from the well-known Young's inequality, as stated in Lemma 3, and the monotone increasing property of the function $g(x) = x^{1/p}$.

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

$$\|U_\star\|_{S_{p_1}}^{k_1} = \|V_\star\|_{S_{p_2}}^{k_2} \quad \text{and} \quad \|X\|_{S_p} = \|U_\star\|_{S_{p_1}} \|V_\star\|_{S_{p_2}}.$$

Then

$$\begin{aligned} \|X\|_{S_p} &= \|U_\star\|_{S_{p_1}} \|V_\star\|_{S_{p_2}} \\ &= \left(\frac{p_2 \|U_\star\|_{S_{p_1}}^{p_1} + p_1 \|V_\star\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{1/p} \end{aligned}$$

which implies that

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1} + p_1 \|V\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{1/p}. \end{aligned} \quad (27)$$

This completes the proof. ■

D. Proof of Theorem 2

Proof: Let $P \in \mathbb{R}^{m \times d}$ and $Q \in \mathbb{R}^{n \times d}$ be any factor matrices such that $X = PQ^T$, and $\hat{p}_1 = p_1 > 0$ and $\hat{p}_2 = p_2 p_3 / (p_2 + p_3) > 0$, which means that $1/\hat{p}_1 + 1/\hat{p}_2 = 1/p$. According to Theorem 1, we obtain

$$\|X\|_{S_p} = \min_{P \in \mathbb{R}^{m \times d}, Q \in \mathbb{R}^{n \times d}: X=PQ^T} \|P\|_{S_{\hat{p}_1}} \|Q\|_{S_{\hat{p}_2}}. \quad (28)$$

Let $Q_1 \in \mathbb{R}^{d \times d}$ and $Q_2 \in \mathbb{R}^{n \times d}$ be factor matrices of Q , i.e., $Q_1 Q_2^T = Q^T$. Since $\hat{p}_2 = p_2 p_3 / (p_2 + p_3)$, then $1/\hat{p}_2 = 1/p_2 + 1/p_3$. By Theorem 1, we also have

$$\|Q\|_{S_{\hat{p}_2}} = \min_{Q_1 \in \mathbb{R}^{d \times d}, Q_2 \in \mathbb{R}^{n \times d}: Q=(Q_1 Q_2^T)^T} \|Q_1\|_{S_{p_2}} \|Q_2\|_{S_{p_3}}. \quad (29)$$

Combining (28) and (29), we obtain

$$\|X\|_{S_p} = \min_{P \in \mathbb{R}^{m \times d}, Q_1 \in \mathbb{R}^{d \times d}, Q_2 \in \mathbb{R}^{n \times d}: X=PQ_1 Q_2^T} \|P\|_{S_{p_1}} \|Q_1\|_{S_{p_2}} \|Q_2\|_{S_{p_3}}.$$

This completes the proof. ■

E. Proof of Corollary 4

Proof: Since $p_1 = p_2 = p_3 = 3p > 0$, and using Theorem 2, we have

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

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

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

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

$$X = U_\star V_\star W_\star^T, \quad \|U_\star\|_{S_{3p}}^{3p} = \|V_\star\|_{S_{3p}}^{3p} = \|W_\star\|_{S_{3p}}^{3p} = \|X\|_{S_p}^p,$$

which implies that

$$\|U_\star\|_{S_{3p}} \|V_\star\|_{S_{3p}} \|W_\star\|_{S_{3p}} = \left(\frac{\|U_\star\|_{S_{3p}} + \|V_\star\|_{S_{3p}} + \|W_\star\|_{S_{3p}}}{3} \right)^3.$$

The theorem now follows because

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

This completes the proof. ■

V. CONCLUSIONS

In general, the Schatten- p quasi-norm minimization is non-convex, non-smooth and even non-Lipschitz. In addition, most existing algorithms are too slow or even impractical for large-scale problems, due to the SVD or EVD of the whole matrix in each iteration. Therefore, it is very important that how to transform such challenging problems into a simpler one, such as a smooth optimization problem. In this paper, we first presented and rigorously proved that for any $p, p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$, the Schatten- p quasi-norm of any matrix is equivalent to the minimization of the product of the Schatten- p_1 norm (or quasi-norm) and Schatten- p_2 norm (or quasi-norm) of two factor matrices. Especially, when $p > 1/2$, there is an equivalence between the Schatten- p quasi-norm (or norm) of any matrix and the Schatten- $2p$ norms of its two factor matrices, e.g., $\|X\|_* = \min_{X=UV^T} (\|U\|_F^2 + \|V\|_F^2)/2$. That is, various Schatten- p quasi-norm minimization problems with $p > 1/2$ can be transformed into the simpler one only involving the

smooth norms of two factor matrices, which can naturally lead to simpler and more efficient algorithms than conventional methods.

We further extended the equivalence relationship of two factor matrices to the cases of three and more factor matrices, from which we can see that for any $0 < p < 1$, the Schatten- p quasi-norm of any matrix is the minimization of the mean to the power of $(\lfloor 1/p \rfloor + 1)$ of the Schatten- $(\lfloor 1/p \rfloor + 1)p$ norms of all factor matrices. In other words, for any $0 < p < 1$, the Schatten- p quasi-norm minimization can be transformed into an optimization problem only involving the smooth norms of multiple factor matrices. Finally, we provided some representative examples for two and three factor matrices. It is clear that the bi-nuclear and Frobenius/nuclear quasi-norms defined in our previous paper [1] and the tri-nuclear quasi-norm defined in our previous paper [2] are three important special cases.

REFERENCES

- [1] F. Shang, Y. Liu, and J. Cheng, “Scalable algorithms for tractable Schatten quasi-norm minimization,” in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 2016–2022.
- [2] ———, “Tractable and scalable Schatten quasi-norm approximations for rank minimization,” in *Proc. 19th Int. Conf. Artif. Intell. Statist.*, 2016, pp. 620–629.
- [3] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [4] E. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [5] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
- [6] M. Yuan, A. Ekici, Z. Lu, and R. Monteiro, “Dimension reduction and coefficient estimation in multivariate linear regression,” *J. R. Statist. Soc. B*, vol. 69, pp. 329–346, 2007.
- [7] A. Argyriou, C. A. Micchelli, M. Pontil, and Y. Ying, “A spectral regularization framework for multi-task structure learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 25–32.
- [8] Z. Liu and L. Vandenberghe, “Interior-point method for nuclear norm approximation with application to system identification,” *SIAM J. Matrix ANAL. Appl.*, vol. 31, no. 3, pp. 1235–1256, 2009.
- [9] M. Fazel, H. Hindi, and S. P. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proc. American Control Conf.*, 2001, pp. 4734–4739.

- [10] E. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.
- [11] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, pp. 471–501, 2010.
- [12] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its Oracle properties,” *J. Am. Statist. Assoc.*, vol. 96, pp. 1348–1361, 2001.
- [13] C. H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
- [14] C. Lu, J. Tang, S. Yan, and Z. Lin, “Generalized nonconvex nonsmooth low-rank minimization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4130–4137.
- [15] Z. Lu and Y. Zhang, “Schatten- p quasi-norm regularized matrix optimization via iterative reweighted singular value minimization,” *arXiv:1401.0869v2*, 2015.
- [16] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, “Robust matrix completion via joint Schatten p -norm and L_p -norm minimization,” in *Proc. IEEE Int. Conf. Data Mining*, 2012, pp. 566–574.
- [17] A. Aravkin, R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, “Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation,” *SIAM J. Sci. Comput.*, vol. 36, no. 5, pp. S237–S266, 2014.
- [18] A. Majumdar and R. K. Ward, “An algorithm for sparse MRI reconstruction by Schatten p -norm minimization,” *Magn. Reson. Imaging*, vol. 29, pp. 408–417, 2011.
- [19] K. Mohan and M. Fazel, “Iterative reweighted algorithms for matrix rank minimization,” *J. Mach. Learn. Res.*, vol. 13, pp. 3441–3473, 2012.
- [20] M. Lai, Y. Xu, and W. Yin, “Improved iteratively reweighted least squares for unconstrained smoothed ℓ_p minimization,” *SIAM J. Numer. Anal.*, vol. 51, no. 2, pp. 927–957, 2013.
- [21] F. Nie, H. Huang, and C. Ding, “Low-rank matrix recovery via efficient Schatten p -norm minimization,” in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 655–661.
- [22] G. Marjanovic and V. Solo, “On ℓ_p optimization and matrix completion,” *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714–5724, 2012.
- [23] M. Zhang, Z. Huang, and Y. Zhang, “Restricted p -isometry properties of nonconvex matrix recovery,” *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4316–4323, 2013.
- [24] N. Srebro, J. Rennie, and T. Jaakkola, “Maximum-margin matrix factorization,” in *Proc. Adv. Neural*

Inf. Process. Syst., 2004, pp. 1329–1336.

- [25] R. Mazumder, T. Hastie, and R. Tibshirani, “Spectral regularization algorithms for learning large incomplete matrices,” *J. Mach. Learn. Res.*, vol. 11, pp. 2287–2322, 2010.
- [26] K. Mitra, S. Sheorey, and R. Chellappa, “Large-scale matrix factorization with missing data under additional constraints,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1642–1650.
- [27] B. Recht and C. Ré, “Parallel stochastic gradient algorithms for large-scale matrix completion,” *Math. Prog. Comp.*, vol. 5, no. 2, pp. 201–226, 2013.
- [28] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, “A generalized iterated shrinkage algorithm for non-convex sparse coding,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 217–224.
- [29] S. Foucart and M. Lai, “Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$,” *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 397–407, 2009.
- [30] Y. Liu, F. Shang, H. Cheng, and J. Cheng, “A Grassmannian manifold algorithm for nuclear norm regularized least squares problems,” in *Proc. 30th Conf. Uncert. in Art. Intel.*, 2014, pp. 515–524.
- [31] F. Shang, Y. Liu, J. Cheng, and H. Cheng, “Robust principal component analysis with missing data,” in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manag.*, 2014, pp. 1149–1158.
- [32] H. Xu, C. Caramanis, and S. Sanghavi, “Robust PCA via outlier pursuit,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 2496–2504.
- [33] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, “Low-rank matrix recovery from errors and erasures,” *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4324–4337, 2013.
- [34] C. Hsieh and P. A. Olsen, “Nuclear norm minimization via active subspace selection,” in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 575–583.
- [35] W. Bian, X. Chen, and Y. Ye, “Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization,” *Math. Program.*, vol. 149, pp. 301–327, 2015.
- [36] R. Larsen, “PROPACK-software for large and sparse SVD calculations,” Available from <http://sun.stanford.edu/srmunk/PROPACK/>, 2005.
- [37] J.-F. Cai and S. Osher, “Fast singular value thresholding without singular value decomposition,” *Methods Anal. appl.*, vol. 20, no. 4, pp. 335–352, 2013.
- [38] R. Cabral, F. Torre, J. Costeira, and A. Bernardino, “Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2488–2495.

- [39] K. H. Jin and J. C. Ye, “Annihilating filter-based low-rank Hankel matrix approach for image inpainting,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3498–3511, Nov. 2015.
- [40] D. S. Mitrinović, *Analytic Inequalities*. Heidelberg: Springer-Verlag, 1970.