Article

A Unified Scalable Equivalent Formulation for Schatten Quasi-Norms

Fanhua Shang^{1,2,*} , Yuanyuan Liu¹, Fanjie Shang¹, Hongying Liu¹ , Lin Kong¹ and Licheng Jiao¹

¹ Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China; yyliu@xidian.edu.cn (Y.L.); fjshang@stu.xidian.edu.cn (F.S.); hylu@xidian.edu.cn (H.L.); lkong@stu.xidian.edu.cn (L.K.); lchjiao@mail.xidian.edu.cn (L.J.)

² Peng Cheng Laboratory, Shenzhen 518066, China

* Correspondence: fhshang@xidian.edu.cn

Received: 7 July 2020; Accepted: 5 August 2020; Published: 10 August 2020

Abstract: The Schatten quasi-norm is an approximation of the rank, which is tighter than the nuclear norm. However, most Schatten quasi-norm minimization (SQNM) algorithms suffer from high computational cost to compute the singular value decomposition (SVD) of large matrices at each iteration. In this paper, we 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 minimizing the product of the Schatten p_1 -(quasi-)norm and Schatten p_2 -(quasi-)norm of its two much smaller factor matrices. Then, we present and prove the equivalence between the product and its weighted sum formulations for two cases: $p_1 = p_2$ and $p_1 \neq p_2$. In particular, when $p > 1/2$, there is an equivalence between the Schatten p -quasi-norm of any matrix and the Schatten $2p$ -norms of its two factor matrices. 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 of the Schatten $(\lfloor 1/p \rfloor + 1)p$ -norms of $\lfloor 1/p \rfloor + 1$ factor matrices, where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$.

Keywords: Schatten quasi-norm; nuclear norm; rank function; factor matrix; equivalent formulations

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 matrix completion [1], robust principal component analysis (RPCA) [2], low-rank representation [3], multivariate regression [4], multi-task learning [5], and system identification [6]. To efficiently solve such problems, we mainly relax the rank function to its tractable convex envelope, that is, the nuclear norm $\|\cdot\|_*$ (sum of the singular values, also known as the trace norm or Schatten 1-norm), which leads to a convex optimization problem [1,7–9]. In fact, the nuclear norm of a 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 [10,11] 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 that the nuclear norm penalty shrinks all singular values equally, which also leads to the over-penalization of large singular values [12]. That is, the nuclear norm may make the solution deviate from the original solution as the ℓ_1 -norm does. Unlike the nuclear norm, the Schatten p -quasi-norm with $0 < p < 1$ is non-convex, and it can give a closer approximation to the rank function. Thus, the Schatten quasi-norm minimization (SQNM) has received

a significant amount of attention from researchers in various communities, such as image recovery [13], collaborative filtering [14,15], and MRI analysis [16].

Recently, several iteratively re-weighted least squares (IRLS) algorithms as in [17,18] have been proposed to approximately solve associated Schatten quasi-norm minimization problems. Lu et al. [13] proposed a family of iteratively re-weighted nuclear norm (IRNN) algorithms to solve various non-convex surrogate (including the Schatten quasi-norm) minimization problems. In [13,14,19,20], the Schatten quasi-norm has been shown to be empirically superior to the nuclear norm for many real-world problems. Moreover, [21] theoretically proved that the SQNM requires significantly fewer measurements than nuclear norm minimization as in [1,9]. However, all the algorithms have to be solved iteratively and involve singular value decomposition (SVD) in each iteration. Thus, they suffer from a high computational complexity of $O(\min(m, n)mn)$ and are not applicable for large-scale matrix recovery problems [22].

It is well known that the nuclear norm has a scalable equivalent formulation (also called the bilinear spectral penalty [9,23,24]), which has been successfully applied in many applications, such as collaborative filtering [15,25,26]. Zuo et al. [27] proposed a generalized shrinkage-thresholding operator to iteratively solve ℓ_p quasi-norm minimization with $0 \leq p < 1$. 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 (note that the first version [28] of this paper appeared on ArXiv, and in the current version we have included experimental results and fixed a few typos.), we first present and prove the equivalence relation on the Schatten p -(quasi-)norm of any matrix and minimization of the product of the Schatten p_1 -(quasi-)norm and Schatten p_2 -(quasi-)norm of its two smaller factor matrices, for any $p, p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$. Moreover, we prove the equivalence between the product formulation of the Schatten (quasi-)norm and its weighted sum formulation 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 of any matrix and the Schatten $2p$ -norms of its two factor matrices, where a representative example is the equivalent formulation of the nuclear norm, i.e.,

$$\|X\|_* = \min_{X=UV^T} \frac{\|U\|_F^2 + \|V\|_F^2}{2}.$$

That is, the SQNM problems with $p > 1/2$ can be transformed into problems only involving the smooth convex norms of two factor matrices, which can lead to simpler and more efficient algorithms (e.g., [12,22,29,30]). Clearly, the resulting algorithms can reduce the per-iteration cost from $O(\min(m, n)mn)$ to $O(mnd)$, where $d \ll m, n$ in general.

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 of the Schatten $(\lfloor 1/p \rfloor + 1)p$ -norms of $\lfloor 1/p \rfloor + 1$ factor matrices, where $\lfloor 1/p \rfloor$ denotes the largest integer not exceeding $1/p$. Note that the norms of all the smaller factor matrices are convex and smooth. Besides the theoretical results, we also present several representative examples for two and three factor matrices. In fact, the bi-nuclear and Frobenius/nuclear quasi-norms defined in our previous work [22] and the tri-nuclear quasi-norm defined in our previous work [29] are three important special cases of our unified scalable formulations for Schatten quasi-norms.

2. Notations and Background

In this section, we present the definition of the Schatten p -norm and discuss some related work about Schatten p -norm minimization. We first briefly recall the definitions of both the norm and the quasi-norm.

Definition 1. A quasi-norm $\|\cdot\|$ on a vector space X is a real-valued map $X \rightarrow [0, \infty)$ with the following conditions:

- $\|x\| \geq 0$ for any vector $x \in X$, and $\|x\| = 0$ if and only if $x = 0$.
- $\|\alpha x\| = |\alpha| \|x\|$ if $\alpha \in \mathbb{R}$ or $\mathbb{C}$, and $x \in X$.
- There is a constant $c \geq 1$ so that if $x_1, x_2 \in X$ we have $\|x_1 + x_2\| \leq c(\|x_1\| + \|x_2\|)$ (triangle inequality).

If the triangle inequality is replaced by $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$, then it becomes a norm.

Definition 2. The Schatten p -norm ($0 < p < \infty$) of a matrix $X \in \mathbb{R}^{m \times n}$ (without loss of generality, we 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 2 defines a natural norm, whereas it defines a quasi-norm for $0 < p < 1$. For instance, the Schatten 1-norm is the so-called nuclear norm, $\|X\|_*$, and the Schatten 2-norm is the well-known Frobenius norm, $\|X\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$. As the non-convex surrogate for the rank function, the Schatten p -quasi-norm is the better approximation than the nuclear norm [21], analogous to the superiority of the ℓ_p quasi-norm to the ℓ_1 -norm [18,31].

To recover a low-rank matrix from a small set of linear observations, $b \in \mathbb{R}^l$, the general SQNM 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

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

where $\lambda > 0$ is a regularization parameter, and $f(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}$ denotes a certain measurement for characterizing the loss $(\mathcal{A}(X) - b)$. For instance, $\mathcal{A}$ is the projection operator $\mathcal{P}_\Omega$, and $f(\cdot) = \|\cdot\|_2^2$ in matrix completion [13,17,20,32], where $\mathcal{P}_\Omega$ is the orthogonal projection onto the linear subspace of matrices supported on $\Omega := \{(i, j) | D_{ij} \text{ is observed}\}$, that is, $\mathcal{P}_\Omega(D)_{ij} = D_{ij}$ if $(i, j) \in \Omega$, and $\mathcal{P}_\Omega(D)_{ij} = 0$ otherwise. For RPCA problems [2,33–36], $\mathcal{A}$ is an identity operator and $f(\cdot) = \|\cdot\|_1$. In the multivariate regression problem in [37], $\mathcal{A}(X) = AX$ with A being a given matrix, and $f(\cdot) = \|\cdot\|_F^2$. In addition, $f(\cdot)$ may be chosen as the Hinge loss in [23] or the ℓ_p quasi-norm in [14].

Generally, SQNM problems (e.g., (2) and (3)) are non-convex, non-smooth, and even non-Lipschitz [38]. So far, only a few algorithms (e.g., IRLS [17,18] and IRNN [13]) have been developed to solve such challenging problems. However, since the SQNM algorithms involve SVD of large matrices in each iteration, they suffer from a high computational cost of $O(mn^2)$, which severely limits their applicability to large-scale problems. While there have been many efforts towards fast SVD computation such as partial SVD [39], the performances of these methods are still unsatisfactory for many real applications [40]. As an alternative to reduce the computational complexity of SVD on a large matrix, one can factorize X into two smaller factor matrices (i.e., $X = UV^T$). According to the unitary invariant property of norms, (2) and (3) can be reformulated into two much smaller matrix optimization problems as in [23,41–44], which are still non-convex and non-smooth. Therefore, it is very important to determine how to transform challenging problems such as (2) and (3) into more tractable ones, which can be solved by simpler and more efficient algorithms as in [30,43].

3. A Unified Formulation for Schatten Quasi-Norms

In this section, we first present and prove an equivalence relation on the Schatten p -(quasi)-norm of any matrix and the Schatten p_1 -(quasi)-norm and Schatten p_2 -(quasi)-norm 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 between the product formulation of the Schatten (quasi)-norm and its weighted sum formulation for the two cases: $p_1 = p_2$ and $p_1 \neq p_2$, respectively. For any $1/2 < p \leq 1$, the Schatten p -(quasi)-norm of any matrix is equivalent to minimizing the mean of the Schatten $2p$ -norms of both factor matrices, which can lead to simpler and more efficient algorithms. Finally, we extend the theoretical results of two factor matrices to the cases of three and more factor matrices. One can see that for any $0 < p < 1$, the Schatten p -(quasi)-norm of any matrix is the minimization of the mean of the Schatten $(\lfloor 1/p \rfloor + 1)p$ -norms of $\lfloor 1/p \rfloor + 1$ factor matrices.

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

In this subsection, we present some unified Schatten quasi-norm scalable formulations for the case of two factor matrices.

Theorem 1. Each matrix $X \in \mathbb{R}^{m \times n}$ with $\text{rank}(X) = r \leq d$ 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 $1/p_1 + 1/p_2 = 1/p$, we have

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

The proof of Theorem 1 is provided in Section 5.1. From Theorem 1, it is clear that for any $0 < p \leq 1$ and $p_1, p_2 > 0$ satisfying $1/p = 1/p_1 + 1/p_2$, the Schatten p -(quasi)-norm of any matrix X is equivalent to minimizing the product of the Schatten p_1 -(quasi)-norm and Schatten p_2 -(quasi)-norm of its two factor matrices.

In fact, we can find 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 the two cases for p_1 and p_2 , that is, $p_1 = p_2$ and $p_1 \neq p_2$.

3.1.1. The Case of $p_1 = p_2$

We first 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 as follows.

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

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_{S_{2p}}^{2p} + \|V\|_{S_{2p}}^{2p}}{2} \right)^{1/p}. \quad (5)$$

Remark 1. The proof of Theorem 2 is provided in Section 5.2. From the equality in (5), we know that for any $0 < p \leq 1$, the Schatten p -(quasi)-norm minimization problems in many low-rank matrix completion and recovery applications can be transformed into the problems of minimizing the mean of the Schatten $2p$ -(quasi)-norms of two smaller factor matrices. It should be noted that when $1/2 < p \leq 1$, the norms of two smaller factor matrices are convex and smooth (see Example 2 below) due to $2p > 1$, which can lead to simpler and more efficient algorithms as in [43].

When $p = 1$ and $p_1 = p_2 = 2$, the equalities in Theorem 2 become the following form [23].

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

$$\|X\|_* = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_F \|V\|_F = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \frac{1}{2} (\|U\|_F^2 + \|V\|_F^2). \quad (6)$$

The bilinear spectral penalty in (6) has been widely used in many low-rank matrix completion and recovery problems, such as collaborative filtering [9,23], RPCA [45], online RPCA [46], and image recovery [47]. Note that the well-known equivalent formulation of the nuclear norm in Corollary 1 is just a special case of Theorem 2 when $p = 1$ and $p_1 = p_2 = 2$. Moreover, 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 the following property, as in our previous work [22,29].

Property 1. For any matrix $X \in \mathbb{R}^{m \times n}$ with $\text{rank}(X) = r \leq d$, the following equalities hold:

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

In [22,29], 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 Theorem 2 when $p = 1/2$ 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 the following property.

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

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

3.1.2. The Case of $p_1 \neq p_2$

In this part, we discuss the case of $p_1 \neq p_2$. Unlike 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, 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 $p_2 = 2$ for $p = 2/3$), and thus we give the following unified scalable equivalent formulations for the Schatten p -(quasi-)norm. Note that after our work [28] appeared on ArXiv, [12] also presented the same result independently as in Theorem 3.

Theorem 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:

$$\begin{aligned} \|X\|_{s_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{s_{p_1}} \|V\|_{s_{p_2}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{p_2 \|U\|_{s_{p_1}}^{p_1} + p_1 \|V\|_{s_{p_2}}^{p_2}}{p_1 + p_2} \right)^{1/p} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{\|U\|_{s_{p_1}}^{p_1}/p_1 + \|V\|_{s_{p_2}}^{p_2}/p_2}{1/p} \right)^{1/p}. \end{aligned} \quad (8)$$

Remark 2. The proof of Theorem 3 is given in Section 5.3. From Theorem 3, we know that Theorem 2 and Corollary 1 can be viewed as two special cases of Theorem 3 when $p_1 = p_2 = 2p$ and $p_1 = p_2 = 2$, respectively. That is, Theorem 3 is the more general form of Theorem 2 and Corollary 1. From the equality in (8), we can see that for any $0 < p \leq 1$, the Schatten p -(quasi)-norm minimization problem can be transformed into the one of minimizing the weighted sum of the Schatten p_1 -(quasi)-norm and Schatten p_2 -(quasi)-norm of two smaller factor matrices (see Examples 3 and 4 below), where the weights of the two terms are $p_2/(p_1 + p_2)$ and $p_1/(p_1 + p_2)$, respectively.

Below 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 $p_2 = 2$, we have the following property, as in our previous work [22].

Property 3. Given any matrix $X \in \mathbb{R}^{m \times n}$ with $\text{rank}(X) = r \leq d$, the following equalities hold:

$$\|X\|_{S_{2/3}} = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_* \|V\|_F = \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \left(\frac{2\|U\|_* + \|V\|_F^2}{3} \right)^{3/2}. \quad (9)$$

The scalable formulation in (9) is known as the Frobenius/nuclear hybrid quasi-norm defined in our previous work [22]. It is clear that the Frobenius/nuclear hybrid quasi-norm is also a special case of Theorem 3 when $p = 2/3$, $p_1 = 1$, and $p_2 = 2$. As shown in [22,29], we can design more efficient algorithms to solve the Schatten p -quasi-norm minimization with $1/2 \leq p < 1$.

Example 4. When $p = 2/5$, and by setting $p_1 = 1/2$ and $p_2 = 2$, we have the following property.

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

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

3.2. Extensions to Multiple Factor Matrices

In this subsection, we extend the unified Schatten quasi-norm formulations of two factor matrices to the case of multiple factor matrices.

Theorem 4. Each matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$ 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 $W \in \mathbb{R}^{n \times d}$, that is, $X = UVW^T$. 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

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X=UVW^T} \left(\frac{\|U\|_{S_{p_1}}^{p_1/p_1} / p_1 + \|V\|_{S_{p_2}}^{p_2/p_2} / p_2 + \|W\|_{S_{p_3}}^{p_3/p_3} / p_3}{1/p} \right)^{1/p}. \end{aligned} \quad (10)$$

The proof of Theorem 4 is provided in Section 5.4. From Theorem 4, one 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 of any matrix is equivalent to minimizing the weighted sum of the Schatten p_1 -(quasi)-norm, Schatten p_2 -(quasi)-norm, and Schatten p_3 -(quasi)-norm of the three smaller factor matrices, where the weights of the three terms are p/p_1 , p/p_2 , and p/p_3 , respectively. Similarly, we extend the results in Theorem 4 to the case of more factor matrices as follows.

Theorem 5. Each matrix $X \in \mathbb{R}^{m \times n}$ of $\text{rank}(X) = r \leq d$ 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

$$\|X\|_{S_p} = \min_{U_i: X = \prod_{i=1}^M U_i} \prod_{i=1}^M \|U_i\|_{S_{p_i}} = \min_{U_i: X = \prod_{i=1}^M U_i} \left(\frac{\sum_{i=1}^M \|U_i\|_{S_{p_i}}^{p_i} / p_i}{1/p} \right)^{1/p}. \quad (11)$$

The proof of Theorem 5 is very similar to that of Theorem 4, and is thus omitted. From Theorem 5, we can know that for any $0 < p \leq 1$ and $p_i > 0$ for all $i = 1, 2, \dots, M$ satisfying $\sum_{i=1}^M 1/p_i = 1/p$, the Schatten p -(quasi-)norm of any matrix is equivalent to minimizing the weighted sum of the Schatten p_i -(quasi-)norms of the smaller factor matrices, where the weights for these terms are p/p_i for all $i = 1, 2, \dots, M$.

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 (10) 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.

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

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p} + \|V\|_{S_{3p}}^{3p} + \|W\|_{S_{3p}}^{3p}}{3} \right)^{1/p}. \end{aligned} \quad (12)$$

Remark 3. The proof of Corollary 2 is provided in Section 5.5. From the equality in (12), we know that for any $0 < p < 1$, Schatten p -quasi-norm minimization can be transformed into the problem of minimizing the mean of the Schatten $3p$ -norms (or quasi-norms) of three much smaller factor matrices. We also note that when $1/3 < p \leq 1$, the norms of the three factor matrices are convex and smooth because $3p > 1$, which can also lead to some simpler and more efficient algorithms.

Example 5. Below we give a representative example. When $p = 1/3$ and $p_1 = p_2 = p_3 = 1$, the equalities in Corollary 2 become the following forms, as in our previous work [29].

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

$$\begin{aligned} \|X\|_{S_{1/3}} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \|U\|_* \|V\|_* \|W\|_* \\ &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: X = UVW^T} \left(\frac{\|U\|_* + \|V\|_* + \|W\|_*}{3} \right)^3. \end{aligned} \quad (13)$$

From Property 5, we can see that the tri-nuclear quasi-norm defined in our previous work [29] is also a special case of Corollary 2.

Theorem 2 shows that for any p satisfying $1/2 < p \leq 1$, the Schatten p -(quasi-)norm of any matrix is equivalent to minimizing the mean of the Schatten $2p$ -norms of both much smaller factor matrices, as well as Corollary 2 for any p satisfying $1/3 < p \leq 1$. In other words, if $1/2 < p \leq 1$ or $1/3 < p \leq 1$, Schatten p -(quasi-)norm minimization can be transformed into a simpler problem only involving the convex and smooth norms of two or three smaller factor matrices. Moreover, we extend the results of Theorem 2 and Corollary 2 to the case of more factor matrices, as shown in Corollary 3 below. The proof of Corollary 3 is very similar to that of Corollary 2, and is thus omitted. That is, for any $0 < p < 1$, the Schatten p -quasi-norm of any matrix can theoretically be equivalent to the minimization

of the mean of the Schatten (Mp)-norms of all M factor matrices, where $M = (\lfloor 1/p \rfloor + 1)$. It needs to be strongly emphasized that the norms of all factor matrices are convex and smooth due to $Mp > 1$, which can help us to design simpler and more efficient algorithms.

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

$$\|X\|_{S_p} = \min_{U_i: X = \prod_{i=1}^M U_i} \prod_{i=1}^M \|U_i\|_{S_{Mp}} = \min_{U_i: X = \prod_{i=1}^M U_i} \left(\frac{\sum_{i=1}^M \|U_i\|_{S_{Mp}}^{Mp}}{M} \right)^{1/p}. \quad (14)$$

4. Numerical Experiments

We conducted numerical experiments to evaluate the performance of our scalable Schatten quasi-norm formulations. For a simple reason, we only briefly consider the following bi-nuclear (BN) quasi-norm and Frobenius/nuclear (FN) quasi-norm regularized least-squares problems:

$$\begin{aligned} & \min_{U,V} \left\{ \frac{\lambda}{2} (\|U\|_* + \|V\|_*) + \frac{1}{2} \|\mathcal{P}_\Omega(UV^T - D)\|_F^2 \right\}, \\ & \min_{U,V} \left\{ \frac{\lambda}{3} (2\|U\|_* + \|V\|_F^2) + \frac{1}{2} \|\mathcal{P}_\Omega(UV^T - D)\|_F^2 \right\}. \end{aligned}$$

Here we introduce the alternating direction method of multipliers (ADMM) for solving the above two scalable Schatten quasi-norm minimization models as in [43]. We compared our BN and FN methods with existing state-of-the-art methods such as weighted nuclear norm minimization (WNNM) [48] (<http://www4.comp.polyu.edu.hk/~cslzhang/>) and iteratively re-weighted nuclear norm (IRNN) [49], which usually have better performance than traditional nuclear norm solvers such as [50]. As suggested in [13], we chose the ℓ_q -norm penalty, where q was chosen from the range of $\{0.1, 0.2, \dots, 1\}$. That is, IRNN is one Schatten q -quasi-norm minimization method. The source code of IRNN can be downloaded at the link: <https://sites.google.com/site/canyilu/>. We selected the same values for the parameters d , ρ , and μ_0 for both our methods (e.g., $d = 9$ as in [51]), and all the experiments were performed on a PC with an Intel i7-7700 CPU and 32 GB RAM.

We report the inpainting results of all the methods on the lenna and pepper images with 50% random missing pixels, as shown in Figures 1 and 2. The results show that both our methods produced significantly better PSNR results than the other methods. As analyzed in [22,29], our quasi-norm formulations required significantly fewer observations than traditional nuclear norm solvers (e.g., [50]), and also led to scalable optimization problems. In particular, the resulting Schatten quasi-norm formulations may be Lipschitz, while the original Schatten quasi-norm minimization is non-Lipschitz and generally NP-hard [38]. Table 1 shows the running time of all the methods. It is clear that both our methods were more than 50 times faster than the well-known methods, especially IRNN, which suffers from high computational costs due to performing SVD in each iteration.

Table 1. Comparison of running time of the methods (in seconds).

	IRNN [49]	WNNM [48]	BN	FN
lenna	2381.04	389.53	7.06	5.27
pepper	2257.39	382.78	7.35	5.31

Figure 1. Image inpainting results of IRNN [49], WNNM [48], and our two methods on the lenna image with 50% random missing pixels (best viewed in color).

Figure 2. Image inpainting results of IRNN [49], WNNM [48], and our two methods on the pepper image with 50% random missing pixels (best viewed in color).

5. Proofs of Main Results

In this section, we give the proofs for some theorems and corollaries. We first introduce several key lemmas, such as Jensen's inequality, Hölder's inequality, and Young's inequality, which are used throughout our proofs.

Lemma 1 (Jensen's inequality). Assume that the function $g(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ 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

$$g\left(\sum_{i=1}^n t_i x_i\right) \geq \sum_{i=1}^n t_i g(x_i). \quad (15)$$

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

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q} \quad (16)$$

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

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab \quad (17)$$

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

5.1. 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 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

$$\text{Tr}^p(A \Sigma_Z A^T) = \sum_k \left(\sum_i a_{ki}^2 \sigma_i \right)^p. \quad (18)$$

Recall that $g(x) = x^p$ with $0 < p < 1$ is a concave function on $\mathbb{R}^+$. By Jensen's inequality [52], 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 $i \in \{1, \dots, r\}$, (18) can be rewritten as

$$\begin{aligned} \text{Tr}^p(A\Sigma_Z A^T) &= \sum_k \left(\sum_i a_{ki}^2 \sigma_i \right)^p \geq \sum_k \sum_i a_{ki}^2 \sigma_i^p \\ &= \sum_i \sigma_i^p = \text{Tr}^p(\Sigma_Z) = \|Z\|_{S_p}^p. \end{aligned} \quad (19)$$

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 (19) is still satisfied. $\square$

Proof of Theorem 1. 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 $\exists O_1, \hat{O}_1 \in \mathbb{R}^{d \times d}$ satisfy $L_X = L_U O_1$ and $L_U = L_X \hat{O}_1$, which implies that $O_1 = L_U^T L_X$ and $\hat{O}_1 = L_X^T L_U$. Thus, $O_1 = \hat{O}_1^T$. Since $L_X = L_U O_1 = L_X \hat{O}_1 O_1$, we have $\hat{O}_1 O_1 = O_1^T O_1 = I_d$. Similarly, we have $O_1 \hat{O}_1 = O_1 O_1^T = I_d$. 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 $O_3 O_3^T = O_3^T O_3 = I_d$, that is, $\sum_i (O_3)_{ij}^2 = \sum_j (O_3)_{ij}^2 = 1$ for $\forall i, j \in \{1, 2, \dots, d\}$, where a_{ij} denotes the element in the i -th row and the j -th column of the matrix A . 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, 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 q_i and τ_j denote the i -th and 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 , that is, 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 $1/p_1 + 1/p_2 = 1/p$, that is, 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

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

where the inequality $a \leq$ holds due to the Hölder's inequality [52], as stated in Lemma 2. 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 Jensen's inequality (see Lemma 1) for the convex function $h(x) = x^{p_2}$ with $p_2 \geq 1$.

For any matrices $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ satisfying $X = UV^T$, we have

$$\|X\|_{s_p} \leq \|U\|_{s_{p_1}} \|V\|_{s_{p_2}}. \quad (20)$$

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

$$X = U_* V_*^T, \|U_*\|_{s_{p_1}}^{p_1} = \|V_*\|_{s_{p_2}}^{p_2} = \|X\|_{s_p}^p \text{ with } 1/p = 1/p_1 + 1/p_2,$$

and

$$\|X\|_{s_p} = (\text{Tr}^p(\Sigma_X))^{1/p} = \|U_*\|_{s_{p_1}} \|V_*\|_{s_{p_2}}.$$

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 $0 < \hat{p}_1 < 1$ and $0 < \hat{p}_2 < 1$ such that $1/\hat{p}_1 + 1/\hat{p}_2 = 1/p$. Next, we prove that the result of Theorem 1 also holds. In fact, 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 $1/p_1 = 1/\hat{p}_1 + 1/q$ with $q \geq 1$. Clearly, we can know that $1/\hat{p}_1 < 1/p_1$. Let $U^* = L_X \Sigma_X^{p/p_1}$, $V^* = R_X \Sigma_X^{p/p_2}$, $U_1^* = L_X \Sigma_X^{p/\hat{p}_1}$ and $V_1^* = R_X \Sigma_X^{p/\hat{p}_2}$, then we have $X = U^* (V^*)^T = U_1^* (V_1^*)^T$, which implies that

$$\|X\|_{s_p} = \|U^*\|_{s_{p_1}} \|V^*\|_{s_{p_2}} = \|U_1^*\|_{s_{\hat{p}_1}} \|V_1^*\|_{s_{\hat{p}_2}}. \quad (21)$$

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

$$V = V_1 U_2, X = U_1 V_1^T, \|V\|_{s_{\hat{p}_2}} = \|U_2\|_{s_q} \|V_1\|_{s_{p_2}}, \|U_1\|_{s_{p_1}} \leq \|U\|_{s_{\hat{p}_1}} \|U_2\|_{s_q}, \quad (22)$$

where the above inequality follows from (20) with $q \geq 1$. Combining (21) and (22) and for any U and V satisfying $X = UV^T$, we have

$$\begin{aligned} \|X\|_{s_p} &= \|U^*\|_{s_{p_1}} \|V^*\|_{s_{p_2}} \leq \|U_1\|_{s_{p_1}} \|V_1\|_{s_{p_2}} \\ &\leq \|U\|_{s_{\hat{p}_1}} \|U_2\|_{s_q} \|V_1\|_{s_{p_2}} = \|U\|_{s_{\hat{p}_1}} \|V\|_{s_{\hat{p}_2}}, \end{aligned} \quad (23)$$

where the first inequality follows from (20). Recall that

$$\|X\|_{s_p} = \|U_1^*\|_{s_{\hat{p}_1}} \|V_1^*\|_{s_{\hat{p}_2}}. \quad (24)$$

Therefore, for any $0 < \hat{p}_1 < 1$ and $0 < \hat{p}_2 < 1$ satisfying $1/p = 1/\hat{p}_1 + 1/\hat{p}_2$, and by (23) and (24), we also have

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

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. $\square$

5.2. Proof of Theorem 2

Proof. Since $p_1 = p_2 = 2p > 0$, $1/p_1 + 1/p_2 = 1/p$, and using Theorem 1,

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

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

$$\begin{aligned} \|X\|_{S_p} &= \min_{U,V:X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} = \min_{U,V:X=UV^T} \left(\|U\|_{S_{2p}}^p \|V\|_{S_{2p}}^p \right)^{1/p} \\ &\leq \min_{U,V:X=UV^T} \left(\frac{\|U\|_{S_{2p}}^{2p} + \|V\|_{S_{2p}}^{2p}}{2} \right)^{1/p}. \end{aligned}$$

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

$$\|X\|_{S_p} = \|U_\star\|_{S_{2p}} \|V_\star\|_{S_{2p}} = \left(\frac{\|U_\star\|_{S_{2p}}^{2p} + \|V_\star\|_{S_{2p}}^{2p}}{2} \right)^{1/p}.$$

The theorem now follows because

$$\min_{U,V:X=UV^T} \|U\|_{S_{2p}} \|V\|_{S_{2p}} = \min_{U,V:X=UV^T} \left(\frac{\|U\|_{S_{2p}}^{2p} + \|V\|_{S_{2p}}^{2p}}{2} \right)^{1/p}.$$

This completes the proof. $\square$

5.3. Proof of Theorem 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 $k_2 = (p_1 + p_2)/p_1$, and it is easy to know that $1/k_1 + 1/k_2 = 1$. Then

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

where the above inequality follows from the well-known Young inequality and the monotone increasing property of the function $g(x) = x^{1/p}$.

Let $U_{\star} = L_X \Sigma_X^{p/p_1}$ and $V_{\star} = R_X \Sigma_X^{p/p_2}$, then $X = U_{\star} V_{\star}^T$ and

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

which implies that

$$\begin{aligned} \|X\|_{S_p} &= \min_{U \in \mathbb{R}^{m \times d}, V \in \mathbb{R}^{n \times d}: X=UV^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \\ &= \min_{U, V: X=UV^T} \left(\frac{p_2 \|U\|_{S_{p_1}}^{p_1} + p_1 \|V\|_{S_{p_2}}^{p_2}}{p_1 + p_2} \right)^{1/p} \\ &= \min_{U, V: X=UV^T} \left(\frac{\|U\|_{S_{p_1}}^{p_1}/p_1 + \|V\|_{S_{p_2}}^{p_2}/p_2}{1/p} \right)^{1/p}. \end{aligned} \quad (25)$$

This completes the proof. $\square$

5.4. Proof of Theorem 4

Proof. Let $U \in \mathbb{R}^{m \times d}$ and $\hat{V} \in \mathbb{R}^{n \times d}$ be any factor matrices such that $X = U \hat{V}^T$, 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

$$\|X\|_{S_p} = \min_{U \in \mathbb{R}^{m \times d}, \hat{V} \in \mathbb{R}^{n \times d}: X=U \hat{V}^T} \|U\|_{S_{\hat{p}_1}} \|\hat{V}\|_{S_{\hat{p}_2}}. \quad (26)$$

Let $V \in \mathbb{R}^{d \times d}$ and $W \in \mathbb{R}^{n \times d}$ be factor matrices of $\hat{V}$ (i.e., $VW^T = \hat{V}^T$). Since $\hat{p}_2 = p_2 p_3 / (p_2 + p_3)$, then $1/\hat{p}_2 = 1/p_2 + 1/p_3$ and

$$\|\hat{V}\|_{S_{\hat{p}_2}} = \min_{V \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{n \times d}: \hat{V}=(VW^T)^T} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}}. \quad (27)$$

Combining (26) and (27), we obtain

$$\|X\|_{S_p} = \min_{U, V, W: X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}}.$$

Using the above result, we have

$$\begin{aligned} \|X\|_{S_p} &= \min_{U, V, W: X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\ &= \min_{U, V, W: X=UVW^T} \left(\|U\|_{S_{p_1}}^p \|V\|_{S_{p_2}}^p \|W\|_{S_{p_3}}^p \right)^{1/p} \\ &\leq \min_{U, V, W: X=UVW^T} \left(\frac{p_2 p_3 \|U\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{1/p}, \end{aligned}$$

where the above inequality follows from the well-known Young inequality.

Let $U_\star = L_X \Sigma_X^{p/p_1}$, $V_\star = \Sigma_X^{p/p_2}$, and $W_\star = R_X \Sigma_X^{p/p_3}$. It is easy to verify that $X = U_\star V_\star W_\star^T$. Then, we have

$$\begin{aligned}\|X\|_{S_p} &= \min_{U,V,W:X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} = \|U_\star\|_{S_{p_1}} \|V_\star\|_{S_{p_2}} \|W_\star\|_{S_{p_3}} \\ &= \left(\frac{p_2 p_3 \|U_\star\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V_\star\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W_\star\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{1/p}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\|X\|_{S_p} &= \min_{U,V,W:X=UVW^T} \|U\|_{S_{p_1}} \|V\|_{S_{p_2}} \|W\|_{S_{p_3}} \\ &= \min_{U,V,W:X=UVW^T} \left(\frac{p_2 p_3 \|U\|_{S_{p_1}}^{p_1} + p_1 p_3 \|V\|_{S_{p_2}}^{p_2} + p_1 p_2 \|W\|_{S_{p_3}}^{p_3}}{p_2 p_3 + p_1 p_3 + p_1 p_2} \right)^{1/p} \\ &= \min_{U,V,W:X=UVW^T} \left(\frac{\|U\|_{S_{p_1}}^{p_1}/p_1 + \|V\|_{S_{p_2}}^{p_2}/p_2 + \|W\|_{S_{p_3}}^{p_3}/p_3}{1/p} \right)^{1/p}.\end{aligned}$$

This completes the proof. $\square$

5.5. Proof of Corollary 2

Proof. Since $p_1 = p_2 = p_3 = 3p > 0$ and $1/p_1 + 1/p_2 + 1/p_3 = 1/p$, and using Theorem 4, we have

$$\|X\|_{S_p} = \min_{U,V,W:X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}}.$$

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

$$\begin{aligned}\|X\|_{S_p} &= \min_{U,V,W:X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U,V,W:X=UVW^T} \left(\|U\|_{S_{3p}}^p \|V\|_{S_{3p}}^p \|W\|_{S_{3p}}^p \right)^{1/p} \\ &\leq \min_{U,V,W:X=UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p} + \|V\|_{S_{3p}}^{3p} + \|W\|_{S_{3p}}^{3p}}{3} \right)^{1/p}.\end{aligned}$$

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

$$\begin{aligned}\|X\|_{S_p} &= \|U_\star\|_{S_{3p}} \|V_\star\|_{S_{3p}} \|W_\star\|_{S_{3p}} \\ &= \left(\frac{\|U_\star\|_{S_{3p}}^{3p} + \|V_\star\|_{S_{3p}}^{3p} + \|W_\star\|_{S_{3p}}^{3p}}{3} \right)^{1/p}.\end{aligned}$$

The theorem now follows because

$$\begin{aligned} & \min_{U,V,W:X=UVW^T} \|U\|_{S_{3p}} \|V\|_{S_{3p}} \|W\|_{S_{3p}} \\ &= \min_{U,V,W:X=UVW^T} \left(\frac{\|U\|_{S_{3p}}^{3p} + \|V\|_{S_{3p}}^{3p} + \|W\|_{S_{3p}}^{3p}}{3} \right)^{1/p}. \end{aligned}$$

This completes the proof. $\square$

6. Conclusions

Generally, the Schatten quasi-norm minimization problem is non-convex, non-smooth, and even non-Lipschitz, and thus most existing algorithms are too slow or even impractical for large-scale matrix recovery and completion problems. Therefore, it is very important to know how to transform such challenging problems into simpler ones. 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 (or weighted sum) of the Schatten p_1 -(quasi)-norm and Schatten p_2 -(quasi)-norm of two much smaller factor matrices. In particular, when $p > 1/2$, there is an equivalence between the Schatten p -(quasi)-norm of any matrix and the Schatten $2p$ -norms of its two factor matrices (i.e., Property 1). That is, the Schatten quasi-norm minimization problem with $p > 1/2$ can be transformed into a simpler one only involving the smooth norms of smaller factor matrices, which can naturally lead to simpler and more efficient algorithms (e.g., [12,22,29,30]).

We further extended the equivalence formulation 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 of the Schatten $(\lfloor 1/p \rfloor + 1)p$ -norms of $\lfloor 1/p \rfloor + 1$ factor matrices. In other words, for any $0 < p < 1$, the Schatten quasi-norm minimization problem 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 in our previous work [22] and the tri-nuclear quasi-norm in our previous work [29] are three important special cases. Our future work is the theoretical analysis of the properties of the proposed unified Schatten p -norm formulations compared to the nuclear norm and the Schatten quasi-norm as in [53,54] (e.g., how many observations are sufficient for our models to reliably recover low-rank matrices).

Author Contributions: Methodology and Formal analysis, F.S. (Fanhua Shang); Formal analysis, Y.L.; Investigation, L.K.; Software, F.S. (Fanjie Shang); Visualization, H.L.; Review & editing, L.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 61876220, 61876221, 61976164, 61836009 and U1701267, and 61871310), the Project supported the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61621005), the Program for Cheung Kong Scholars and Innovative Research Team in University (No. IRT_15R53), the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) (No. B07048), the Science Foundation of Xidian University (Nos. 10251180018 and 10251180019), the National Science Basic Research Plan in Shaanxi Province of China (Nos. 2019JQ-657 and 2020JM-194), and the Key Special Project of China High Resolution Earth Observation System-Young Scholar Innovation Fund.

Acknowledgments: We thank all the reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Candès, E.; Recht, B. Exact Matrix Completion via Convex Optimization. *Found. Comput. Math.* **2009**, *9*, 717–772. [\[CrossRef\]](#)
2. Candès, E.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 1–37. [\[CrossRef\]](#)

3. Liu, G.; Lin, Z.; Yu, Y. Robust Subspace Segmentation by Low-Rank Representation. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 663–670.
4. Yuan, M.; Ekici, A.; Lu, Z.; Monteiro, R. Dimension reduction and coefficient estimation in multivariate linear regression. *J. R. Stat. Soc. B* **2007**, *69*, 329–346. [[CrossRef](#)]
5. Argyriou, A.; Micchelli, C.A.; Pontil, M.; Ying, Y. A Spectral Regularization Framework for Multi-Task Structure Learning. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 25–32.
6. Liu, Z.; Vandenberghe, L. Interior-Point Method for Nuclear Norm Approximation with Application to System Identification. *SIAM J. Matrix Anal. Appl.* **2009**, *31*, 1235–1256. [[CrossRef](#)]
7. Fazel, M.; Hindi, H.; Boyd, S.P. A Rank Minimization Heuristic with Application to Minimum Order System Approximation. In Proceedings of the 2001 American Control Conference, Arlington, VA, USA, 25–27 June 2001; pp. 4734–4739.
8. Candès, E.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **2010**, *56*, 2053–2080. [[CrossRef](#)]
9. Recht, B.; Fazel, M.; Parrilo, P.A. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Rev.* **2010**, *52*, 471–501. [[CrossRef](#)]
10. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its Oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1361. [[CrossRef](#)]
11. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [[CrossRef](#)]
12. Xu, C.; Lin, Z.; Zha, H. A Unified Convex Surrogate for the Schatten- p Norm. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 926–932.
13. Lu, C.; Tang, J.; Yan, S.; Lin, Z. Generalized Nonconvex Nonsmooth Low-Rank Minimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 4130–4137.
14. Nie, F.; Wang, H.; Cai, X.; Huang, H.; Ding, C. Robust Matrix Completion via Joint Schatten p -Norm and L_p -Norm Minimization. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10 December 2012; pp. 566–574.
15. Aravkin, A.; Kumar, R.; Mansour, H.; Recht, B.; Herrmann, F.J. Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation. *SIAM J. Sci. Comput.* **2014**, *36*, S237–S266. [[CrossRef](#)]
16. Majumdar, A.; Ward, R.K. An algorithm for sparse MRI reconstruction by Schatten p -norm minimization. *Magn. Reson. Imaging* **2011**, *29*, 408–417. [[CrossRef](#)]
17. Mohan, K.; Fazel, M. Iterative Reweighted Algorithms for Matrix Rank Minimization. *J. Mach. Learn. Res.* **2012**, *13*, 3441–3473.
18. Lai, M.; Xu, Y.; Yin, W. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_p minimization. *SIAM J. Numer. Anal.* **2013**, *51*, 927–957. [[CrossRef](#)]
19. Nie, F.; Huang, H.; Ding, C. Low-Rank Matrix Recovery via Efficient Schatten p -Norm Minimization. In Proceedings of the Twenty-sixth AAAI conference on artificial intelligence, Toronto, AB, Canada, 22–26 July 2012; pp. 655–661.
20. Marjanovic, G.; Solo, V. On ℓ_p Optimization and Matrix Completion. *IEEE Trans. Signal Process.* **2012**, *60*, 5714–5724. [[CrossRef](#)]
21. Zhang, M.; Huang, Z.; Zhang, Y. Restricted p -Isometry Properties of Nonconvex Matrix Recovery. *IEEE Trans. Inform. Theory* **2013**, *59*, 4316–4323. [[CrossRef](#)]
22. Shang, F.; Liu, Y.; Cheng, J. Scalable Algorithms for Tractable Schatten Quasi-Norm Minimization. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 2016–2022.
23. Srebro, N.; Rennie, J.; Jaakkola, T. Maximum-Margin Matrix Factorization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 13–18 December 2004; pp. 1329–1336.
24. Mazumder, R.; Hastie, T.; Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **2010**, *11*, 2287–2322. [[PubMed](#)]

25. Mitra, K.; Sheorey, S.; Chellappa, R. Large-Scale Matrix Factorization with Missing Data under Additional Constraints. In Proceedings of the 24th Annual Conference on Neural Information Processing Systems 2010, Vancouver, BC, Canada, 6–9 December 2010; pp. 1651–1659.
26. Recht, B.; Ré, C. Parallel stochastic gradient algorithms for large-scale matrix completion. *Math. Prog. Comp.* **2013**, *5*, 201–226. [\[CrossRef\]](#)
27. Zuo, W.; Meng, D.; Zhang, L.; Feng, X.; Zhang, D. A Generalized Iterated Shrinkage Algorithm for Non-convex Sparse Coding. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 217–224.
28. Shang, F.; Liu, Y.; Cheng, J. Unified Scalable Equivalent Formulations for Schatten Quasi-Norms. *arXiv* **2016**, arXiv:1606.00668.
29. Shang, F.; Liu, Y.; Cheng, J. Tractable and Scalable Schatten Quasi-Norm Approximations for Rank Minimization. In Proceedings of the Artificial Intelligence and Statistics, Cadiz, Spain, 9–11 May 2016; pp. 620–629.
30. Yang, L.; Pong, T.K.; Chen, X. A Nonmonotone Alternating Updating Method for a Class of Matrix Factorization Problems. *SIAM J. Optim.* **2018**, *28*, 3402–3430. [\[CrossRef\]](#)
31. Foucart, S.; Lai, M. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 397–407.
32. Liu, Y.; Shang, F.; Cheng, H.; Cheng, J. A Grassmannian Manifold Algorithm for Nuclear Norm Regularized Least Squares Problems. In Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, Quebec City, QC, Canada, 23–27 July 2014; pp. 515–524.
33. Shang, F.; Liu, Y.; Cheng, J.; Cheng, H. Robust Principal Component Analysis with Missing Data. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 3–7 November 2014; pp. 1149–1158.
34. Xu, H.; Caramanis, C.; Sanghavi, S. Robust PCA via outlier pursuit. In Proceedings of the 24th Annual Conference on Neural Information Processing Systems 2010, Vancouver, BC, Canada, 6–9 December 2010; pp. 2496–2504.
35. Chen, Y.; Jalali, A.; Sanghavi, S.; Caramanis, C. Low-rank matrix recovery from errors and erasures. *IEEE Trans. Inform. Theory* **2013**, *59*, 4324–4337. [\[CrossRef\]](#)
36. Shang, F.; Liu, Y.; Tong, H.; Cheng, J.; Cheng, H. Robust bilinear factorization with missing and grossly corrupted observations. *Inform. Sci.* **2015**, *307*, 53–72. [\[CrossRef\]](#)
37. Hsieh, C.; Olsen, P.A. Nuclear Norm Minimization via Active Subspace Selection. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 575–583.
38. Bian, W.; Chen, X.; Ye, Y. Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. *Math. Program.* **2015**, *149*, 301–327. [\[CrossRef\]](#)
39. Larsen, R. PROPACK-Software for Large and Sparse SVD Calculations. 2015. Available online: <http://sun.stanford.edu/srmunk/PROPACK/> (accessed on 28 March 2020)
40. Cai, J.F.; Osher, S. Fast Singular Value Thresholding without Singular Value Decomposition. *Methods Anal. Appl.* **2013**, *20*, 335–352. [\[CrossRef\]](#)
41. Liu, G.; Yan, S. Active subspace: Toward scalable low-rank learning. *Neural Comp.* **2012**, *24*, 3371–3394. [\[CrossRef\]](#)
42. Shang, F.; Liu, Y.; Cheng, J. Generalized Higher-Order Tensor Decomposition via Parallel ADMM. In Proceedings of the 28th AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014; pp. 1279–1285.
43. Shang, F.; Cheng, J.; Liu, Y.; Luo, Z.Q.; Lin, Z. Bilinear Factor Matrix Norm Minimization for Robust PCA: Algorithms and Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2066–2080. [\[CrossRef\]](#)
44. Krechetov, M.; Marecek, J.; Maximov, Y.; Takac, M. Entropy-Penalized Semidefinite Programming. *arXiv* **2019**, arXiv:1802.04332.
45. Cabral, R.; Torre, F.; Costeira, J.; Bernardino, A. Unifying Nuclear Norm and Bilinear Factorization Approaches for Low-rank Matrix Decomposition. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2488–2495.
46. Feng, J.; Xu, H.; Yan, S. Online robust PCA via stochastic optimization. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 404–412.

47. Jin, K.H.; Ye, J.C. Annihilating Filter-Based Low-Rank Hankel Matrix Approach for Image Inpainting. *IEEE Trans. Image Process.* **2015**, *24*, 3498–3511.
48. Gu, S.; Xie, Q.; Meng, D.; Zuo, W.; Feng, X.; Zhang, L. Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision. *Int. J. Comput. Vis.* **2017**, *121*, 183–208. [[CrossRef](#)]
49. Lu, C.; Tang, J.; Yan, S.; Lin, Z. Nonconvex Nonsmooth Low Rank Minimization via Iteratively Reweighted Nuclear Norm. *IEEE Trans. Image Process.* **2016**, *25*, 829–839. [[CrossRef](#)]
50. Toh, K.C.; Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.* **2010**, *6*, 615–640.
51. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2117–2130. [[CrossRef](#)] [[PubMed](#)]
52. Mitrinović, D.S. *Analytic Inequalities*; Springer: Berlin/Heidelberg, Germany, 1970.
53. Xu, Z. The minimal measurement number for low-rank matrix recovery. *Appl. Comput. Harmon. Anal.* **2018**, *44*, 497–508. [[CrossRef](#)]
54. Zhang, R.; Li, S. Optimal RIP bounds for sparse signals recovery via ℓ_p minimization. *Appl. Comput. Harmon. Anal.* **2019**, *47*, 566–584. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).