The Perturbation Analysis of Nonconvex Low-Rank Matrix Robust Recovery

Jianwen Huang, Feng Zhang, Jianjun Wang, Member, IEEE, Xinling Liu, and Jinping Jia

Abstract—In this article, we bring forward a completely perturbed nonconvex Schatten $p$-minimization to address a model of completely perturbed low-rank matrix recovery (LRMR). This article based on the restricted isometry property (RIP) and the Schatten-$p$ null space property (NSP) generalizes the investigation to a complete perturbation model thinking over not only noise but also perturbation, and it gives the RIP condition and the Schatten-$p$ NSP assumption that guarantee the recovery of low-rank matrix and the corresponding reconstruction error bounds. In particular, the analysis of the result reveals that in the case that $p$ decreases $0$ and $a > 1$ for the complete perturbation and low-rank matrix, the condition is the optimal sufficient condition $\delta_p < 1$ (Recht et al., 2010). In addition, we study the connection between RIP and Schatten-$p$ NSP and discern that Schatten-$p$ NSP can be inferred from the RIP. The numerical experiments are conducted to show better performance and provide outperformance of the nonconvex Schatten $p$-minimization method comparing with the convex nuclear norm minimization approach in the completely perturbed scenario.

Index Terms—Low-rank matrix recovery (LRMR), nonconvex Schatten $p$-minimization, perturbation of linear transformation.

I. INTRODUCTION

LOW-RANK matrix recovery (LRMR) is a rapidly developing topic attracting the interest of numerous researchers in the field of optimization and compressed sensing. Mathematically, we can describe it as follows:

$$y = A(x)$$

where $A: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^m$ is a known linear transformation (we suppose that $m < n$), $y \in \mathbb{R}^m$ is a given observation vector, and $X \in \mathbb{R}^{m \times n}$ is the matrix to be recovered. In practical applications, the matrix $X$ can be the gray-scale image to be recovered or completed, about which see the experimental section of this article, or it can be the scene graph, about which see the literature [1] for a study. The objective of LRMR is to find the low-rank matrix based on $(y, A)$. If the observation $y$ is corrupted by noise $z$, the model (1) is changed into the following form:

$$\hat{y} = A(x) + z$$

where $\hat{y}$ is the noisy measurement, and $z$ is the additive noise independent of the matrix $X$. However, numerous LRMR models can be encountered where not only the linear measurement $y$ is contaminated by the noise vector $z$ but also the linear transformation $A$ is perturbed by $E$ for completely perturbed setting, namely, substitute the linear transformation $A$ with $\hat{A} = A + E$. The completely perturbed appearance arises in remote sensing [2], radar [3], source separation [4], etc. When $m = n$ and the matrix $X = \text{diag}(x)$ ($x \in \mathbb{R}^n$) is diagonal, models (1) and (2) separately degenerate to the compressed sensing models

$$y = Ax$$

$$\hat{y} = Ax + z$$

where $A \in \mathbb{R}^{M \times m}$ is a measurement matrix, and $x \in \mathbb{R}^m$ is an unknown sparse signal. We call problem (3) as the sparse signal recovery. For the completely perturbed model, the convex nuclear norm minimization is frequently considered [5] as follows:

$$\min_{Z \in \mathbb{R}^{m \times n}} \|Z\|_* \text{ s.t. } \|A(\hat{Z}) - \hat{y}\|_2 \leq \epsilon_{A,r,y}$$

where $\|\hat{Z}\|_*$ is the nuclear norm of the matrix $\hat{Z}$, that is, the sum of its singular values, and $\epsilon_{A,r,y}$ is the total noise level. Under the condition of (4), problem (5) can be reduced
to the $l_1$-minimization [6]
\[
\min_{\tilde{z} \in \mathbb{R}^n} \|\tilde{z}\|_1 \quad \text{s.t.} \quad \|\hat{A}\tilde{z} - \hat{y}\|_2 \leq \epsilon_{A,r,y} \tag{6}
\]
where $\|\tilde{z}\|_1$ is the $l_1$-norm of the vector $\tilde{z}$, that is, the sum of absolute value of its coefficients.

Chartrand [8] showed that fewer measurements are required for exact reconstruction if the $l_1$-norm is substituted with the $l_p$-norm ($0 < p \leq 1$). There exist many works regarding reconstructing $x$ via the $l_p$-minimization [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. In [8], numerical simulations demonstrated that fewer measurements are needed for exact reconstruction than when $p = 1$.

In this article, we are interested in the completely perturbed model for the nonconvex Schatten $p$-minimization ($0 < p < 1$)
\[
\min_{\tilde{z} \in \mathbb{R}^n} \|\tilde{z}\|_p^p \quad \text{s.t.} \quad \|\hat{A}\tilde{z} - \hat{y}\|_2 \leq \epsilon_{A,r,y} \tag{7}
\]
where $\|\tilde{z}\|_p$ is the Schatten-$p$ quasi-norm of the matrix $\tilde{z}$, that is, $\|\tilde{z}\|_p = (\sum_i \sigma_i(\tilde{Z}))^{1/p}$ with $\sigma_i(\tilde{Z})$ being the $i$th singular value of $\tilde{Z}$. Problem (7) can be returned to the $l_p$-minimization [7]
\[
\min_{\tilde{z} \in \mathbb{R}^n} \|\tilde{z}\|_p^p \quad \text{s.t.} \quad \|\hat{A}\tilde{z} - \hat{y}\|_2 \leq \epsilon_{A,r,y} \tag{8}
\]
where $\|\tilde{z}\|_p = (\sum_i \tilde{z}_i^p)^{1/p}$ is the $l_p$-quasi-norm of the vector $\tilde{z}$. Aldroubi et al. [21] introduced the $l_p$ null space property (NSP) with respect to an arbitrary basis, which ensures that the solutions to $l_p$ minimization are stable with respect to perturbations of the measurement matrix and the dictionary.

The rest of this article is constructed as follows. In Section II, we give some notations and necessary lemmas. In Sections III and IV, we present the main results and its proofs, respectively. In Section V, we conduct a series of numerical experiments to demonstrate the effectiveness and robustness of the method (7) for low-rank matrix reconstruction, and the verifiability of the theoretical results. In Section VI, we have summarized the whole article.

II. PRELIMINARIES

In this section, we will give some important definitions and lemmas that will be used to prove the main results. We first introduce the notion of restricted isometry property (RIP) of a linear transformation $A$, which is as follows.

Definition 1 [22]: The RIC $\delta_r$ of a linear transformation $A$ is the smallest constant such that
\[
(1 - \delta_r)\|X\|_F^2 \leq \|A(X)\|_2^2 \leq (1 + \delta_r)\|X\|_F^2 \tag{9}
\]
holds for all $X \in \mathbb{R}^{m \times n}$ whose rank is at most $r$ [i.e., rank$(X) \leq r$], where $\|X\|_F := \langle (X^T X) \rangle^{1/2} = \text{trace}(X^T X)^{1/2}$ is the Frobenius norm of the matrix $X$.

Then we provide some notations similar to [5], which quantify the perturbations $\mathcal{E}$ and $\epsilon$ with the bounds
\[
\frac{\|\mathcal{E}\|_{\text{op}}}{\|A\|_{\text{op}}} \leq \epsilon_A, \quad \frac{\|\mathcal{E}\|^{(r)}_{\text{op}}}{\|A\|^{(r)}_{\text{op}}} \leq \epsilon_A^{(r)}, \quad \|\tilde{z}\|_2 \leq \epsilon_y \tag{10}
\]
where $\|A\|_{\text{op}} = \sup \{\|A(X)\|_2 /\|X\|_F : X \in \mathbb{R}^{m \times n} \setminus \{0\}\}$ is the operator norm of linear transformation $A$, and $\|A\|_{r}^{(p)} = \sup \{\|A(X)\|_2 /\|X\|_F : X \in \mathbb{R}^{m \times n} \setminus \{0\} \text{ and } \text{rank}(X) \leq r\}$, and represent

$$
t_r = \frac{\|X_{[r]}\|_F}{\|X\|_F}, \quad s_r = \frac{\|X_{[r]}\|_F}{\sqrt{r} \|X_{[r]}\|_F},
$$

$$
k_{\delta}^{(r)} = \frac{\sqrt{1 + \delta_r}}{\sqrt{1 - \delta_r}}, \quad \alpha_A = \frac{\|A\|_{\text{op}}}{\sqrt{1 - \delta_r}}. \quad (11)
$$

Here, $X_{[r]}$ is the best $r$-rank approximation of the matrix $X$, its singular values are composed of the $r$-largest singular values of the matrix $X$, and $X_{[r]} = X - X_{[r]}$.

To prove our main results, we need the following auxiliary lemmas. First, we give Lemma 1 which incorporates an important inequality associated with $\delta_r$ and $\delta_r$.

\textbf{Lemma 1 (RIP for $A$ [5]):} Given the RIC $\delta_r$, related to linear transformation $A$ and the relative perturbation $\epsilon_{A}^{(r)}$ corresponding to linear perturbation $\delta$, fix the constant $\delta_{r, \max} = (1 + \delta_r)(1 + \epsilon_{A}^{(r)})^2 - 1$. Then the RIC $\delta_r \leq \delta_{r, \max}$ for $\delta = A + \delta E$ is the smallest nonnegative constant such that

$$
(1 - \delta_r)\|X\|_F^2 \leq \|\hat{A}(X)\|_F^2 \leq (1 + \delta_r)\|X\|_F^2 \quad (12)
$$

holds for all the matrices $X \in \mathbb{R}^{m \times n}$ that are $r$-rank.

We will use this fact that $A$ maps low-rank orthogonal matrices to nearly sparse orthogonal vectors, which is given by Candès and Plan [22].

\textbf{Lemma 2 [22]} For all $X, Y$ satisfying $(X, Y) = 0$, and rank($X$) $\leq r_1$, rank($Y$) $\leq r_2$

$$
\left\|\hat{A}(X), \hat{A}(Y)\right\| \leq \delta_{r_1 + r_2} \|X\|_F \|Y\|_F. \quad (13)
$$

Moreover, the following lemma will be used in the proofs of main result, which combines with [23, Lemma 2.3] and [27, Lemma 2.2].

\textbf{Lemma 3:} Assume that $X, Y \in \mathbb{R}^{m \times n}$ obey $XTY = 0$ and $XYT = 0$. Let $0 < p \leq 1$. Then

$$
\|X + Y\|^p_p \leq \|X\|^p_p + \|Y\|^p_p, \quad \|X + Y\|_p \geq \|X\|_p + \|Y\|_p \quad (14)
$$

where $\|X\|^p_p$ and $\|X\|_p$ stand for the nuclear norm of matrix $X$ in the case of $p = 1$.

In addition, another important theoretical tool for analyzing LRMR is the Schatten-$p$ NSP proposed by Gao et al. [30]. This definition is given as follows.

\textbf{Definition 2:} Let $A: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^n$ be a linear measurement operator. Given $r \in \mathbb{Z}^+$, for all $X \in \mathbb{R}^{m \times n}$, there are constants $\rho \in (0, 1)$ and $\tau \in (0, +\infty)$, such that

$$
\|X_{[r]}\|_p \leq \rho \|X_{[r]}\|_p + \|A(X)\|_2 \quad (15)
$$

then $A$ is said to fulfill the Schatten-$p$ NSP of order $r$ with constants $\rho$ and $\tau$.

Besides, they gave a more generalized notion to describe the solution to (7) with $E = 0$.

\textbf{Definition 3:} Let $A: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^n$ be a linear measurement operator and $q \in [1, 2]$. Given $r \in \mathbb{Z}^+$, for all $X \in \mathbb{R}^{m \times n}$, there are constants $\rho \in (0, 1)$ and $\tau \in (0, +\infty)$, such that

$$
\|X_{[r]}\|_q \leq \frac{\rho}{\sqrt{r}^{1/p - 1/q}} \|X_{[r]}\|_p + \tau \|A(X)\|_2 \quad (16)
$$

then $A$ is said to fulfill the $l_q$ Schatten-$p$ NSP (RNSP) of order $r$ with constants $\rho$ and $\tau$.

It is obvious that Definition 3 returns to Definition 2 in the case of $p = q$, and it goes to Definition 3.1 with respect to $l_q$ [59].

The following lemma gives the Schatten-$p$ NSP of the model (7) with noisy measurements.

\textbf{Lemma 4:} Given $\epsilon_A$ defined by (10), the measurement operator $A$ fulfills the Schatten-$p$ NSP with constants $\rho \in (\tau \epsilon_A \|A\|_{\text{op}}, 1)$ and $\tau \in (0, +\infty)$. If for all $X \in \mathbb{R}^{m \times n}$, the singular values of $X$ satisfy (15) and fix constants $\hat{\tau} = \tau/(1 + \tau \epsilon_A \|A\|_{\text{op}})$ and $\hat{\rho} = \rho - (\rho + 1) \tau \epsilon_A \|A\|_{\text{op}}/(1 + \tau \epsilon_A \|A\|_{\text{op}})$, then the measurement operator $A$ satisfies the Schatten-$p$ NSP with constants $\hat{\rho}$ and $\hat{\tau} \in (0, 1)$ and $\hat{\tau} \in (0, +\infty)$.

\textbf{Remark 1:} In the case of $E = 0$, i.e., $\epsilon_A = 0$, which means that the measurement operator $A$ is not perturbed by noise $E$, then $\hat{\tau} = \tau$ and $\hat{\rho} = \rho$. In this situation, Lemma 4 degenerates to [30, Definition 4.3].

\textbf{Proof of the Lemma 4:} To control the upper bound of $\|\hat{A}(X)\|_2$ by an equation containing $\|A(X)\|_2$ and $\epsilon_A$, using (10) and the triangle inequality, we get

$$
\|\hat{A}(X)\|_2 \leq \|A(X)\|_2 + \|E\|_2 \leq \|A(X)\|_2 + \|E\|_{\text{op}} \|X\|_F \leq \|A(X)\|_2 + \epsilon_A \|A\|_{\text{op}} \|X\|_F. \quad (17)
$$

Combining with the Schatten-$p$ NSP of $\hat{A}$, the matrix norm inequality $\|X\|_F \leq \|X\|_p$ for all $X \in \mathbb{R}^{m \times n}$ and (17), we get

$$
\|X_{[r]}\|_p \leq \hat{\rho} \|X_{[r]}\|_p + \hat{\tau} \|\hat{A}(X)\|_2 \leq \hat{\rho} \|X_{[r]}\|_p + \hat{\tau} \|\hat{A}(X)\|_{\text{op}} \|X\|_F \leq (\hat{\rho} + \hat{\tau} \epsilon_A \|A\|_{\text{op}}) \|X_{[r]}\|_p + \hat{\tau} \|A(X)\|_2 + \hat{\tau} \epsilon_A \|A\|_{\text{op}} \|X_{[r]}\|_p.
$$

The above inequality leads to

$$
\|X_{[r]}\|_p \leq \frac{\hat{\rho} + \hat{\tau} \epsilon_A \|A\|_{\text{op}}}{1 - \hat{\tau} \epsilon_A \|A\|_{\text{op}}} \|X_{[r]}\|_p + \frac{\hat{\tau} \epsilon_A \|A\|_{\text{op}}}{1 - \hat{\tau} \epsilon_A \|A\|_{\text{op}}} \|A(X)\|_2.
$$

Therefore, $\hat{\tau} = \tilde{\tau}/(1 + \tau \epsilon_A \|A\|_{\text{op}})$ and $\hat{\rho} = \rho - (\rho + 1) \tau \epsilon_A \|A\|_{\text{op}}/(1 + \tau \epsilon_A \|A\|_{\text{op}})$. For given $\epsilon_A$ determined by (10), $\|A\|_{\text{op}}$ and $\tau > 0$, to make the value of $\hat{\rho}$ fall between 0 and 1, we need to solve the following inequality:

$$
0 < \rho - (\rho + 1) \tau \epsilon_A \|A\|_{\text{op}} < 1
$$

which results in

$$
\tau \epsilon_A \|A\|_{\text{op}} < \rho - 1 + 2 \tau \epsilon_A \|A\|_{\text{op}}.
$$

Combining with the assumption $0 < \rho < 1$, the desired result follows.

III. MAIN RESULTS

With notations and symbols above, we present our results for reconstruction of low-rank matrices via the completely perturbed nonconvex Schatten $p$-minimization.
Theorem 1: For given relative perturbations $e_A, e_A^{(r)}, e_A^{(2r)}$, and $e_y$ in (10), suppose the RIC for the linear transformation $A$ fulfills

$$\delta_{2ar} < \frac{2 + \sqrt{2a^{1/2 - 1/p}}}{(1 + \sqrt{2a^{1/2 - 1/p}}) \left(1 + \epsilon_A^{(2ar)}\right)^2} - 1$$

(18)

for $a > 1$ and that the general matrix $X$ meets

$$t_r + s_r < \frac{1}{\kappa_A}.$$  

(19)

Then a minimizer $X^*$ of problem (7) approximates the true matrix $X$ with errors

$$\|X - X^*\|_F \leq C_1(e_{A,r},y)^p + C_2 \|X^{(r)}\|_p^{1 - p/2}$$

(20)

and

$$\|X - X^*\|_p \leq C_1\epsilon_A^{(r)} + C_2 \|X^{(r)}\|_p^{1 - p/2}$$

(21)

where the total noise is

$$e'_{A,r,y} = \left[e_A^{(r)} + e_{A}a_{A,r}t_r + s_r\right] \|y\|_2$$

(22)

and

$$C_1 = \frac{2^p \left(1 + a^{p/2 - 1}\right) \left(1 + \delta_{(a+1)r}\right)^{p/2}}{(1 - \delta_{(a+1)r})^p - a^{p/2 - 1} \left(\delta_{(a+1)r}^{p/2} + \epsilon_A^{(r)}\right)^{p/2}}$$

(23)

$$C_2 = \frac{2a^{p/2 - 1}}{(1 - \delta_{(a+1)r})^p - a^{p/2 - 1} \left(\delta_{(a+1)r}^{p/2} + \epsilon_A^{(r)}\right)^{p/2}}$$

(24)

where

$$\delta_{(a+1)r} = (1 + \delta_{(a+1)r}) \left(1 + \epsilon_A^{(a+1)r}\right)^2 - 1, \quad \delta_{2ar} = (1 + \delta_{2ar}) \left(1 + \epsilon_A^{(2ar)}\right)^2 - 1.$$  

Remark 2: Theorem 1 gives a sufficient condition for the reconstruction of low-rank matrices via nonconvex Schatten $p$-minimization in completely perturbed scenario. Condition (18) of the theorem extends the assumption of $l_p$ situation in [7] to the nonconvex Schatten $p$-minimization. Observe that as the value of $p$ becomes large, the bound of RIC $\delta_{2ar}$ reduces, which reveals that smaller value of $p$ can induce weaker reconstruction guarantee. Particularly, when $p \to 0$ ($a > 1$) ([7]) degenerates to the rank minimization: $\min_{\tilde{Z} \in \mathbb{R}^{m \times r}} \text{rank}(\tilde{Z}) \text{s.t.} ||A(\tilde{Z}) - Z||_F \leq \epsilon_A$, it leads to the RIP condition $\delta_2 < 2/(1 + \epsilon_A^{(2ar)})^2 - 1$ for reconstruction of low-rank matrices via the rank minimization. To the best of our knowledge, the current optimal recovery condition about RIP is $\delta_2 < 1$ to ensure exact reconstruction for $r$-rank matrices via rank minimization [23]. Therefore, the theorem extends that condition to the scenario of the presence of noise and $r$-rank matrices. Furthermore, when $m = n$ and the matrix $X = \text{diag}(x)$ ($x \in \mathbb{R}^m$) is diagonal, the theorem reduces to the case of compressed sensing given by Ince and Nacaroglu [7].

Remark 3: Under the requirement (18), one can easily check that condition (19) is satisfied. Besides, when rank($X \leq r$), condition (19) holds. In addition, inequalities (20) and (21) in Theorem 1 exploit two kinds of metrics to provide upper bound estimations on the reconstruction error of nonconvex Schatten $p$-minimization. The estimations evidence that reconstruction accurateness can be controlled by the best $r$-rank approximation error and the total noise. In particular, when there are no noises (i.e., $\mathcal{E} = 0$ and $z = 0$), they clear that the $r$-rank matrix can be accurately reconstructed via the nonconvex Schatten $p$-minimization. In (20), both the error-bound noise constant $C_1$ and the error-bound compressibility constant $C_2$ may rely on the value of $p$. Numerical simulations reveal that when we fix the other independent parameters, a smaller value of $p$ will produce a smaller $C_1$ and a smaller $C_2/r^{1-p/2}$. For more details, see Fig. 1.

Remark 4: When the matrix $X$ is a strict rank at most $r$ matrix (i.e., $X = X^{(r)}$), a minimizer $X^*$ of problem (7) approximates the true matrix $X$ with errors

$$\|X - X^*\|_F \leq C_1'\epsilon_A^{r} + C_2'\|X^{(r)}\|_p^{1 - p/2}$$

where

$$\epsilon_A^{r} = \left[e_A^{(r)} + e_y\right] \|y\|_2$$

(27)

In the case of $\mathcal{E} = 0$, that is, there does not exist perturbation in the linear transformation $A$, then $\delta_{(a+1)r} = \delta_{2ar} = \delta_{2ar}$. Below, based on the Schatten-$p$ NSP, we give the reconstruction condition of the model (7) and the corresponding error estimators.

Theorem 2: For known $e_A, e_A^{(r)}, e_A^{(2r)}$, and $e_y$ in (10), assume that the measurement operator $A$ satisfies the Schatten-$p$ NSP of order $r$ with constants $\rho \in (\tau e_A^{(r)} \|A\|_{op}, 1)$ and $\tau \in (0, +\infty)$. Then the solution $X^*$ to the model (7) fulfills

$$\|X - X^*\|_F \leq D_1 \|X^{(r)}\|_p + D_2\epsilon_A^{r} + \epsilon_y$$

(28)

and

$$\|X - X^*\|_p \leq D_1 \|X^{(r)}\|_p + D_2\epsilon_A^{r} + \epsilon_y$$

(29)
where
\[ D_1 = \frac{2^{\frac{r}{2}}(1 + \rho^r)^{\frac{r}{2}}}{(1 - \rho^r)^{\frac{r}{2}}}, \quad D_2 = \frac{2^{\frac{r}{2}}}{(1 - \rho^r)^{\frac{r}{2}}} \]
\[ D'_1 = \frac{2^{\frac{r}{2}}(2\rho^2 + 1)^{\frac{r}{2}}}{(1 - \rho^r)^{\frac{r}{2}}} \]
\[ D'_2 = 2\sqrt{2} \frac{2^{\frac{r}{2}}}{1 + \frac{2^{\frac{r}{2}}(2\rho^2 + 1)}{(1 - \rho^r)^{\frac{r}{2}}}} \]
with \( \rho \) and \( \tau \) defined in Lemma 4.

Remark 5: The theorem shows that the completely perturbed constrained Schatten-\( p \) norm minimization method (7) is stable and robust.

In what follows, the connection between restricted isometric property and the Schatten-\( p \) RNSP will be investigated. We will show that the Schatten-\( p \) RNSP can be inferred from the RIP.

Theorem 3: For any \( t > 1 \), assume that the measurement operator \( \mathcal{A} \) fulfills the restricted isometric property with
\[ \delta_t < \sqrt{\frac{t - 1}{t}}. \]
We have
\[ \|X_r\|_F \leq \frac{\delta_t}{\sqrt{(t - 1)(1 - \delta_t^2)}} \|X_r\|_p \]
\[ + \frac{2}{(1 - \delta_t)\sqrt{1 + b_{tr}}} \|A(X)\|_2 \]
that is, \( \mathcal{A} \) meets the Schatten-\( p \) RNSP with constants \( \rho \) and \( \tau \).

Remark 6: We conjecture that the combination of Theorem 3 and Lemma 1 leads to a new conclusion that guarantees that the low-rank matrix \( X \) in \( \hat{X} = \mathcal{A}(X) + z \) can be robustly reconstructed by the completely perturbed Schatten-\( p \)-miniaturization method (7). Since this problem needs to be explored and studied in depth, we will consider it in the future.

IV. PROOFS OF THE MAIN RESULTS
In this part, we will provide the proofs of main results. For any matrix \( X \in \mathbb{R}^{m \times n} \), we represent the singular values decomposition (SVD) of \( X \) as
\[ X = U\text{diag}(\sigma(X))V^\top \]
where \( \sigma(X) := (\sigma_1(X), \ldots, \sigma_m(X))^\top \) is the vector of the singular values of \( X \), and \( U \) and \( V \) are, respectively, the left and right singular value matrices of \( X \).

Proof of the Theorem 1: Let \( X \) denote the original matrix and \( X^* \) denote the optimal solution of (7). Let \( Z = X - X^* \), and based on the SVD of \( X \), its SVD is given by
\[ U^\top Z V = U_1\text{diag}(\sigma(U^\top Z V))V_1^\top \]
where \( U_1, V_1 \in \mathbb{R}^{m \times m} \) are orthogonal matrices, and \( \sigma(U^\top Z V) \) stands for the vector composed of the singular values of \( U^\top Z V \). Let \( T_i \) be the set composed of the locations of the \( r \) largest magnitudes of elements of \( \sigma(X) \). We adopt the technology similar to [7] to partition \( \sigma(U^\top Z V) \) into a sum of vectors \( \sigma_{T_i}(U^\top Z V) \) \( (i = 0, 1, \ldots, J) \), where \( T_i \) is the set composed of the locations of the \( ar \) largest magnitudes of entries of \( \sigma_{T_i}(U^\top Z V) \), \( T_2 \) is the set composed of the locations of the second \( ar \) largest magnitudes of entries of \( \sigma_{T_i}(U^\top Z V) \), and so forth (except possibly \( T_J \)). Then \( Z = \sum_{i=0}^{J} T_i \), where \( Z_{T_i} = UU_1\text{diag}(\sigma_{T_i}(U^\top Z V))(V V_1)^\top, i = 0, 1, \ldots, J \). One can easily verify that \( Z_{T_i}^\top Z_{T_i} = 0 \) and \( Z_{T_i} Z_{T_i}^\top = 0 \) for all \( i \neq j \). Let \( Z_{T_i}, Z_{T_j}, \|Z_{T_i} \|_{2}, \|Z_{T_j} \|_{2} \) be explored and studied in depth, we will consider it in the future.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
we get
\[ \| \hat{A} (Z_{Ti}) \|_2^p \leq \| \hat{A} (Z_{Ti}) \|_2^2 \| \hat{A} (Z) \|_2^p \]
+ \sum_{i \geq 2} \| (\hat{A} (Z_{Ti}), \hat{A} (Z_T)) \|^p \]
where (a) follows from the fact that \((a + b)^p \leq a^p + b^p\) for nonnegative \(a\) and \(b\).

In addition, by feasibility of \(X^*\), we get
\[ \| \hat{A} (Z) \|_2 \leq \| \hat{y} - \hat{A} (X) \|_2 + \| \hat{y} - \hat{A} (X^*) \|_2 \leq 2 \epsilon_{A,r,y} \]  
(38)

Since \(Z_{Ti}\) is \((a + 1)r\)-rank and \(Z_T\) is \(ar\)-rank, \(i \geq 2\), by applying the RIP of \(\hat{A}\) and in combination with (37) and (38), we get
\[ \| \hat{A} (Z_{Ti}) \|_2^2 \leq (2 \epsilon_{A,r,y})^p (1 + \delta_{(a+1)r}) \| Z_{Ti} \|_F^p \]
+ \sum_{i \geq 2} \| (\hat{A} (Z_{Ti}), \hat{A} (Z_T)) \|^p \]
(39)

Because \((Z_T, Z_T) = 0\) for all \(i \neq j\), and \(Z_{Ti}\) is \(r\)-rank, by Lemma 2 and (35), we get
\[ \| \hat{A} (Z_{Ti}) \|_2^p \leq (2 \epsilon_{A,r,y})^p (1 + \delta_{(a+1)r}) \| Z_{Ti} \|_F^p \]
+ \sum_{i \geq 2} \| (\hat{A} (Z_{Ti}), \hat{A} (Z_T)) \|^p \]
(40)

From (18), one can easily check that
\[ a^\xi - 1 (\delta_{(a+1)r} + \delta_{2ar}) \leq (1 - \delta_{(a+1)r})^\xi. \]
(41)

By (40), (41), and the inequality \(\| \hat{A} (Z_{Ti}) \|_2^p \geq (1 - \delta_{(a+1)r})^{p/2} \| Z_{Ti} \|_F^p\), one can get
\[ \| X_{Ti} \|_F^p \leq \left( \frac{2^p (1 + a^\xi - 1 (1 + \delta_{(a+1)r})^\xi)}{(1 - \delta_{(a+1)r})^\xi - a^\xi - 1 (\delta_{(a+1)r} + \delta_{2ar})^\xi} \right) (\epsilon_{A,r,y})^p \]
+ \frac{2 \xi - 1 (\delta_{(a+1)r} + \delta_{2ar})^\xi}{(1 - \delta_{(a+1)r})^\xi - a^\xi - 1 (\delta_{(a+1)r} + \delta_{2ar})^\xi} \| X_{Ti} \|_F^p \]
(42)

Thus, from (35) and (42), we get
\[ \| X \|_F^p \leq \left( \frac{Z_{Ti}}{F} \right)^p + \left( \frac{Z_T}{F} \right)^p \]
\[ \leq C_1 (\epsilon_{A,r,y})^p + C_2 \frac{\| X_{Ti} \|_F^p}{r^{1 - \xi}}. \]
(43)

In addition, from a combination of (29) and (43), one can get
\[ \| X \|_p^p \leq \left( \frac{Z_{Ti}}{p} \right)^p + \left( \frac{Z_T}{p} \right)^p \]
\[ \leq C_1 (\epsilon_{A,r,y})^p + C_2 \frac{\| X_{Ti} \|_F^p}{r^{1 - \xi}}. \]
(44)

where the constants \(C_1, C_2, C_1', \) and \(C_2'\) are defined in Theorem 1. The proof is complete.

Proof of the Theorem 2: Applying Theorem 1 [60], we get
\[ \| X^* \|_p^p \leq \left( \frac{X - (X - X^*)}{p} \right)^p \]
\[ \geq \sum_{i=1}^m | \sigma_i (X) - \sigma_i (X - X^*) | \]
\[ = \sum_{i=1}^m | \sigma_i (X) - \sigma_i (X - X^*) | + \sum_{i=r+1}^f (| \sigma_i (X) - \sigma_i (X - X^*) | \]
\[ \geq \sum_{i=1}^m | \sigma_i (X) - \sigma_i (X - X^*) | \]
\[ + \sum_{i=r+1}^f (| \sigma_i (X) - \sigma_i (X - X^*) | \]
(45)

Accordingly, by feasibility of \(X^*\), it implies
\[ \| (X - X^*) \|_F^p \]
\[ \leq \| X^* \|_F^p + \| (X - X^*) \|_F^p + \| X_{Ti} \|_F^p \]
\[ \leq \| X^* \|_F^p + \| X_{Ti} \|_F^p + \| X_{Ti} \|_F^p + \| X_{Ti} \|_F^p \]
\[ \leq \left( \frac{(X - X^*)}{F} \right)^p + \| X_{Ti} \|_F^p. \]
(46)

Using the Schatten-\(p\) NSP of \(\hat{A}\) and (38), we get
\[ \| X^* \|_F^p \]
\[ \leq \left( \hat{\rho} \frac{(X - X^*)}{F} + \hat{\epsilon} \cdot (\epsilon_{A,r,y}) \right)^p \]
\[ + \| X_{Ti} \|_F^p. \]
(47)

where (a) follows from the fact that \((a + b)^p < a^p + b^p\) for \(0 < p < 1\) and \(a, b > 0\).

By (46), we get
\[ \| (X - X^*) \|_F^p \leq \frac{1}{1 - \hat{\rho}^p} \left( \hat{\epsilon} \cdot (\epsilon_{A,r,y}) \right)^p + 2 \| X_{Ti} \|_F^p. \]
(48)
From a combination of the Schatten-\(p\) NSP of \(\tilde{A}\) and (47), we get
\[
\|X - X^*\|_p \leq \left( \hat{\rho} \|\tilde{X} - X^*\|_{\rho^\uparrow} + \hat{\tau} \cdot (2 \epsilon'_{A,r,y}) \right)^p + \|\tilde{X} - X^*\|_{\rho^\downarrow}^p \\
\leq (1 + \hat{\rho}^p) \|\tilde{X} - X^*\|_{\rho^\uparrow}^p + \hat{\tau}^p \cdot (2 \epsilon'_{A,r,y})^p \\
\leq 2 \left( 1 + \frac{\hat{\rho}^p + \hat{\tau}^p}{1 - \hat{\rho}^p} \right) \|\tilde{X} - X^*\|_{\rho^\uparrow}^p + \frac{2 \hat{\tau}^p}{1 - \hat{\rho}^p} \epsilon'_{A,r,y}.
\]
which is (27).

In the following, we prove (28). By exploiting the Schatten-
\(p\) NSP of \(\tilde{A}\) and the matrix inequality \(\|X\|_F \leq \|X\|_p\) for all \(X \in \mathbb{R}^{m \times n}\), we get
\[
\|X - X^*\|_p \leq \frac{2^{\frac{p}{p-1}} (1 + \hat{\rho}^p)^\frac{1}{p}}{(1 - \hat{\rho}^p)^\frac{1}{p}} \|\tilde{X} - X^*\|_p + \frac{2 \hat{\tau}^p}{1 - \hat{\rho}^p} \epsilon'_{A,r,y}
\]
which is (27).

Finally, using the vector inequality \(\|x\|_2 \leq \|x\|_1\) for all \(x \in \mathbb{R}^n\), it implies
\[
\|X - X^*\|_F \leq 2 \sqrt{2 \hat{\tau} \epsilon'_{A,r,y}} \left[ 1 + \frac{2 \hat{\tau}^2 (2 \hat{\rho}^2 + 1)}{(1 - \hat{\rho}^p)^\frac{2}{p}} + \frac{2 \hat{\tau}^2 (2 \hat{\rho}^2 + 1)^\frac{1}{2}}{(1 - \hat{\rho}^p)^\frac{1}{p}} \right] \|X\|_{\rho^\uparrow}^p
\]
which is (28). The proof is complete. \(\square\)

**Proof of the Theorem 3:** Unless specifically stated, the notations appearing here have the same meaning as before. Assume \(\text{supp}(\sigma(X_{\rho^\uparrow})) = T_0\). Let
\[
\|X\|_{\rho^\uparrow}^p = \|\sigma r(X)\|_{\rho^\uparrow}^p = r \alpha^p.
\]
The set \(T_0^c\) is divided as
\[
T_0^c = \Lambda_1 \cup \Lambda_2
\]
where
\[
\Lambda_1 = \left\{ i \in T_0^c : |\sigma r(X_{\rho^\uparrow})_i|^p > \frac{\|X\|_{\rho^\uparrow}^p}{(t - 1)r} \right\}
\]
\[
\Lambda_2 = \left\{ i \in T_0^c : |\sigma r(X_{\rho^\uparrow})_i|^p \leq \frac{\|X\|_{\rho^\uparrow}^p}{(t - 1)r} \right\}.
\]
Then
\[
X_{\rho^\uparrow} = X_{\Lambda_1} + X_{\Lambda_2}
\]
where \(X_{\Lambda_i} = \text{U} \text{diag}(\sigma_{\Lambda_i}(X))V^T, i = 1, 2\). Set \(r_1 = \text{rank}(X_{\Lambda_1}) = \|X_{\Lambda_1}\|_{\rho^\uparrow}\).
Since
\[
\|X_{\Lambda_1}\|_{\rho^\uparrow}^p \geq r_1 \frac{\|X\|_{\rho^\uparrow}^p}{(t - 1)r} \geq r_1 \frac{\|X_{\Lambda_1}\|_{\rho^\uparrow}^p}{(t - 1)r}
\]
then we get
\[
r_1 < (t - 1)r.
\]
By the notion of \(\Lambda_i, i = 1, 2\), we can derive that
\[
\|X_{\Lambda_2}\|_{\rho^\uparrow}^p = \|\sigma_{\Lambda_2}\|_{\rho^\uparrow}^p \leq \frac{\|X_{\rho^\uparrow}\|_{\rho^\uparrow}^p}{(t - 1)r} = \left( \frac{\alpha}{(t - 1)^\frac{p}{2}} \right)^p
\]
\[
\|X_{\Lambda_2}\|_{\rho^\uparrow}^p = \|\sigma_{\Lambda_2}\|_{\rho^\uparrow}^p \leq \|X_{\rho^\uparrow}\|_{\rho^\uparrow}^p - \|X_{\Lambda_1}\|_{\rho^\uparrow}^p \\
\leq \|X_{\rho^\uparrow}\|_{\rho^\uparrow}^p - r_1 \frac{\|X_{\rho^\uparrow}\|_{\rho^\uparrow}^p}{(t - 1)r} \\
= [(t - 1)r - r_1] \left( \frac{\alpha}{(t - 1)^\frac{p}{2}} \right)^p
\]
which combined with [19, Lemma 2.2] and \(\sigma_{\Lambda_2}\) can be denoted as the convex hull of \([(t - 1)r - r_1]\)-sparse vectors
\[
\sigma_{\Lambda_2} = \sum_j \mu_j w_j
\]
where \(w_j = [(t - 1)r - r_1]\)-sparse and
\[
\sum_j \mu_j = 1, \quad \mu_j \in [0, 1], \quad j = 1, 2, \ldots, m
\]
with $w_j$ fulfilling
\[
\sum_j \mu_j \|w_j\|_2^2 \leq \frac{\alpha^p}{r-1} \|X_{\Lambda}^\top\|_{2-p}^{2-p} \leq \frac{\alpha^p}{r-1} \|X_{\{r\}^\top}\|_{2-p}^{2-p}
\]
\[
\leq \frac{1}{r-1} \frac{\|X_{\{r\}^\top}\|_p^p}{\|X_{\{r\}^\top}\|_{2-p}^{2-p}} 
\leq \frac{1}{(r-1)r} \|X_{\{r\}^\top}\|_p^2.
\]  
(49)

One can easily check the identity below \cite{31}, \cite{61}
\[
\frac{1}{4\delta_m} \sum_j \mu_j \left( \|A[(1 + \delta_m)X_{T_m,\Lambda}_1 + \delta_n G_j]\|_2^2 
- \|A[(1 - \delta_m)X_{T_m,\Lambda}_1 - \delta_n G_j]\|_2^2 \right)
\leq \frac{\|A(X_{T_m,\Lambda}_1 + \sum_j \mu_j G_j)\|_p}{\|A(X)\|}
= \frac{\|A(X_{T_m,\Lambda}_1), A(X_{T_m,\Lambda}_1 + \sum_j \mu_j G_j)\|}{\|A(X)\|}
\]  
(50)

where $G_j = U diag(w_j) V^\top$ with $G_j$ being $[(r-1)T - r]$. By the ranks of $X_{T_m}$ and $X_{\Lambda}$ being $r$ and $r_1$, separately, we can gain that $(1 + \delta_m)X_{T_m,\Lambda}_1 + \delta_n G_j$ and $(1 - \delta_m)X_{T_m,\Lambda}_1 - \delta_n G_j$ are $p$-rank.

First, we consider the estimation of the upper bound for the right-hand side of (50). By the definition of RIP, we get
\[
\|A(X_{T_m,\Lambda}_1), A(X)\| \leq \|A(X_{T_m,\Lambda}_1)\|_p \|A(X)\|_2 
\leq \sqrt{1 + \delta_m} \|X_{T_m,\Lambda}_1\|_F \|A(X)\|_2.
\]  
(51)

On the other hand, using the $tr$-order RIP, we can control the left-hand side of (50) from underneath by
\[
\frac{1}{4\delta_m} \sum_j \mu_j \left( (1 - \delta_m) \| (1 + \delta_m)X_{T_m,\Lambda}_1 + \delta_n G_j \|_p^2 
- (1 + \delta_m) \| (1 - \delta_m)X_{T_m,\Lambda}_1 - \delta_n G_j \|_p^2 \right)
\geq \frac{1 - \delta_m^2}{2} \|X_{T_m,\Lambda}_1\|_p^2 - \frac{\delta_m^2}{2} \sum_j \mu_j \|w_j\|_2^2
\geq \frac{1 - \delta_m^2}{2} \|X_{T_m,\Lambda}_1\|_p^2 - \frac{\delta_m^2}{2(t-1)r} \|X_{\{r\}^\top}\|_p^2
\]  
(52)

where (a) follows from the fact $(X_{T_m,\Lambda}_1, G_j) = (\sigma_{T_m,\Lambda}_1, w_j) = 0$, and (b) is due to (49).

A combination of (51) and (52) brings about
\[
\frac{1}{2} \|X_{T_m,\Lambda}_1\|_p^2 - \sqrt{1 + \delta_m} \|X_{T_m,\Lambda}_1\|_F \|A(X)\|_2 
- \frac{\delta_m^2}{2(t-1)r} \|X_{\{r\}^\top}\|_p^2 \leq 0.
\]

As a result of $(1 - \delta_m^2)/2 > 0$, we can settle aforementioned second-order inequality and obtain
\[
\|X_{\{r\}^\top}\|_F \leq \|X_{T_m,\Lambda}_1\|_F
\leq \frac{1}{1 - \delta_m} \left( \sqrt{1 + \delta_m} \|X_{T_m,\Lambda}_1\|_F \|A(X)\|_2 
+ \frac{\sqrt{1 + \delta_m} \|X_{T_m,\Lambda}_1\|_F \|A(X)\|_2}{(1 - \delta_m)} \right.
+ \left. \frac{\delta_m^2}{(t-1)r} \|X_{\{r\}^\top}\|_p^2 \right)^2
\]
\[
\leq \frac{2}{(1 - \delta_m) \sqrt{1 + \delta_m}} \|A(X)\|_2 + \frac{\delta_m}{\sqrt{(1 - \delta_m^2) (1 - \delta_m^2)}} \|X_{\{r\}^\top}\|_p^2
\]
\[
=: \tau \|A(X)\|_2 + \rho \|X_{\{r\}^\top}\|_p
\]
where (a) is due to the fact that $(a^2 + b^2)/2 \leq a + b$ for $a, b \in [0, \infty)$. In the case of $\delta_m < ((r-1)/t)^{1/2}$, we gain the wanted inequality with $\rho = \delta_m/((r-1)(1-\delta_m^2))^{1/2} < 1$. \hfill \Box

V. NUMERICAL EXPERIMENTS

In this section, we carry out some numerical experiments to sustain verification of our theoretical results, and we implement all the experiments in MATLAB 2016a running on a PC with an Intel core i7 processor (3.6 GHz) with 8-GB RAM.

A. Solving Problem (7) by ADMM

To address the completely perturbed nonconvex Schatten $p$-minimization model, we use the alternating direction method of multipliers (ADMM), which is often applied in compressed sensing and sparse approximation \cite{62}, \cite{63}, \cite{64}, \cite{65}. The constrained optimization problem (7) can be transformed into an equivalent unconstrained form
\[
\min_{Z \in \mathbb{R}^{m \times n}} \lambda \|Z\|_p^p + \frac{1}{2} \|\hat{A}(Z) - \hat{y}\|_2^2
\]
\[
\text{where } \hat{A} \in \mathbb{R}^{m \times mn} \text{ and vec}(\hat{Z}) \text{ represents the vectorization of } \hat{Z}. \text{ Hence, } \hat{A}(\hat{Z}) \text{ presents the linear measurements } \hat{A}(\hat{Z}). \text{ Then, introducing an auxiliary variable } W \in \mathbb{R}^{m \times n}, \text{ problem (53) can be equivalently turned into}
\min_{W \in \mathbb{R}^{m \times n}} \lambda \|W\|_p^p + \frac{1}{2} \|\hat{A}(\hat{Z}) - \hat{y}\|_2^2
\text{ s.t. } \hat{Z} = W. \quad (54)
\]

The associating augmented Lagrangian function is provided by
\[
L_\rho(\hat{Z}, W, Y) = \lambda \|W\|_p^p + \frac{1}{2} \|\hat{A}(\hat{Z}) - \hat{y}\|_2^2 
+ \langle Y, \hat{Z} - W \rangle + \frac{\rho}{2} \|\hat{Z} - W\|_F^2
\]
\[
\text{where } Y \in \mathbb{R}^{m \times n} \text{ is a dual variable, and } \rho > 0 \text{ is a penalty parameter. Then, ADMM used in (55) comprises the iterations as follows:}
\]
\[
\hat{Z}^{k+1} = \arg \min_{\hat{Z}} \frac{1}{2} \|\hat{A}(\hat{Z}) - \hat{y}\|_2^2 + \frac{\rho}{2} \|\hat{Z} - W^k - \frac{Y^k}{\rho}\|_F^2
\]
\[
W^{k+1} = \arg \min_W \lambda \|W\|_p^p + \frac{\rho}{2} \|\hat{Z}^{k+1} - W - \frac{Y^k}{\rho}\|_F^2
\]
\[
y^{k+1} = y + \rho (\hat{Z}^{k+1} - W^{k+1})
\]
To make our algorithm clear to the reader, we now briefly give the solution of subproblem (57). To facilitate the notation, we reword subproblem (57) as follows:
\[
\min \alpha \|x\|_p^p + \frac{1}{2}\|X - \Phi\|_F^2.
\] (59)

The SVD of X is represented by \(X = U\text{diag}(\sigma(X))V^\top\), where \(\sigma(X)\) has the same meaning as before. Nie et al. [66] have proved that for the optimal solution \(X\) to (59), \(U\) and \(V\) are separately the left and right singular vectors of \(\Phi\), respectively, and the \(i\)th singular value \(\sigma_i\) is the optimal solution of the below problem
\[
\min_{\sigma \geq 0} \alpha \sigma_i^p + \frac{1}{2}(\sigma_i - \theta_i)^2
\] (60)
where \(\theta_i\) is the \(i\)th singular value of \(\Phi\). Represent the goal function in problem (60) by \(f(x)\), i.e.,
\[
f(x) = \alpha x^p + \frac{1}{2}(x - \theta)^2.
\] (61)

One can easily see that the subdifferential of \(f(x)\) is
\[
h(x) = \frac{d}{dx}f(x) = \alpha px^{p-1} + x - \theta.
\] (62)

Set
\[
\mu = \left(\frac{\alpha p(1-p)}{2}\right)^{-\frac{1}{p-1}}.
\] (63)

The optimal solution of (60) can be gained by
\[
\begin{align*}
\hat{h}(\mu) \geq 0, & \quad x^* = 0 \\
\hat{h}(\mu) < 0, & \quad x^* = \arg\min_{x \in [0,x_1]} f(x)
\end{align*}
\] (64)
where \(x_1\) is the root of \(h(x) = 0\), which can be estimated with the Newton approach initiated at \(2\mu\).

All the solving processes are concluded in Algorithm 1.

Algorithm 1 Solve Problem (7) by ADMM

1: Input \(A \in \mathbb{R}^{M \times mn}, \ y \in \mathbb{R}^M\), perturbation \(E \in \mathbb{R}^{M \times mn}\) with \(\|E\| = \epsilon_A\|A\|, \ \rho \in (0,1]\).
2: Initialize \(\hat{Z}^0 = W^0 = Y^0, \ \gamma = 1.1, \ \lambda_0 = 10^{-6}, \ \lambda_{\max} = 10^{10}, \ \rho = 10^{-6}, \ \epsilon = 10^{-8}, \ k = 0\).
3: while not converged do
4: Updated \(\hat{Z}^{k+1}\) by
\[
\hat{Z}^{k+1} \leftarrow \hat{Z}^k + (\hat{A}^\top \hat{A} + \rho I)^{-1}(\hat{A}^\top \hat{y} + \rho \text{vec}(W^k) - \text{vec}(Y^k));
\]
5: Update \(W^{k+1}\) by
\[
\text{arg min}_W \lambda \|W\|_p^p + \frac{\rho}{2}\|\hat{Z}^{k+1} - \left(W - \frac{Y^k}{\rho}\right)\|_F^2;
\]
6: Update \(Y^{k+1}\) by
\[
y^{k+1} = y^k + \rho(\hat{Z}^{k+1} - W^{k+1});
\]
7: Update \(\lambda_{j+1}\) by \(\lambda_{j+1} = \min(\gamma \lambda_j, \lambda_{\max})\).
8: Check the convergence conditions
\[
\|\hat{Z}^{k+1} - \hat{Z}^k\|_\infty \leq \epsilon, \quad \|W^{k+1} - W^k\|_\infty \leq \epsilon, \quad \|\hat{y}^{k+1} - \hat{y}^k\|_\infty \leq \epsilon, \quad \|\hat{Z}^{k+1} - \hat{W}^{k+1}\|_\infty \leq \epsilon.
\]

B. Settings on Synthetic Data Experiments

In our experiments, we generate a measurement matrix \(A \in \mathbb{R}^{M \times mn}\) with i.i.d. Gaussian \(\mathcal{N}(0,1/M)\) elements. And we generate \(X \in \mathbb{R}^{m \times r}\) of rank \(r\) by \(X = PQ\), where \(P \in \mathbb{R}^{m \times r}\) and \(Q \in \mathbb{R}^{r \times n}\) are with its elements being zero-mean, one-variation Gaussian, i.i.d. random variables. As can be seen from Fig. 2, the higher the sampling rate, the better the reconstruction. We fix \(M = 0.4mn\) and select \(m = n = 60\) and \(r = 0.2m\). With fixed \(X\) and \(A\), the measurements \(y\) are produced by \(y = \text{vec}(X) + z\), where \(z\) is the Gaussian noise with its entries obeying \(N(0,0.01)\). The perturbation matrix \(E\) is with its entries following Gaussian distribution, which fulfills \(\|E\| = \epsilon_A\|A\|\), where \(\epsilon_A\) denotes the perturbation level of \(A\) and its value is not fixed. The perturbed matrix \(A, \hat{A} = A + E\), is used in (56). To avoid the randomness, we perform 20 times independent trails and inform the average result in all the tests.

C. Synthetic Experiments

1) Choice of Regularization Parameter \(\lambda\): To look for a proper parameter \(\lambda\) that derives the better recovery effect, we carry out two sets of trails. Fig. 3(a) and (b), respectively, plots the parameter \(\lambda\) and relative error \(\|\|X - \hat{X}\|_F/\|X\|_F\|\) in different \(p\) values and perturbation levels \(\epsilon_A\). The components of \(\lambda\) follow the Gaussian distribution with a mean value of \(0\) and a variance of 0.01. \(\lambda\) ranges from \(10^{-6}\) to \(10^0\). Fig. 3(a) and (b) shows that \(\lambda \in [10^{-6}, 10^{-2}]\) is relatively suitable.

2) Relationship Among Reconstruction Error, Noise Level, and Parameter \(\lambda\): For a given perturbation level \(\epsilon_A = 0.01\), we now explore the relationship among reconstruction error, noise level \(\epsilon\), and regularization parameter \(\lambda\). The variation ranges of the noise \(\sigma\) (that is, the variance of elements of \(z\)) and the regularization parameter \(\lambda\) are \([0.01, 0.03, 0.05, 0.07, 0.10]\) and \([10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}]\), respectively. And
Moreover, for each given noise variance, the regularization parameter \( \lambda \) of results of the relative error for different noise levels and values \( p \) we fix. When the value of \( \lambda \) is fixed, except for \( \lambda = 10^0 \), the relative error gradually decreases as the variance reduces. Moreover, for each given noise variance, the regularization parameter \( \lambda \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\} \) corresponds to a smaller relative error, i.e., it implies a better reconstruction. Consequently, \( \lambda \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\} \) is the better regularization parameter of the model (53). The data in Table I are depicted in Fig. 4, which makes us intuitively see the results of the aforementioned analysis.

3) Convergence of Algorithm: By giving \( \lambda = 10^{-6} \), we consider the convergence of Algorithm 1. Fig. 5(a) and (b) separately presents the relationship between relative neighboring iteration error [RNIE, \( r(k) = \|X^{k+1} - X^k\|_F/\|X^k\|_F \)] and the number of iterations \( k \), as well as the connection between the relative recovery iteration error [RRIE, \( t(k) = \|X^k - X^*\|_F/\|X^*\|_F \)] and the number of iterations \( k \). One can easily see that with the increase in iterations, both RNIE and RRIE decrease quickly. In particular, \( k \geq 200, r(k) < 10^{-4}, \) and \( k \geq 230, t(k) < 10^{-4} \). The results of relative error versus the values of \( p \) in different \( \epsilon_A \) are shown in Fig. 6. Fig. 6 indicates that the proper choice of the size of \( p \) will be helpful to facilitate the performance of nonconvex Schatten \( p \)-minimization.

4) Comparison of Real Error Bound With Theoretical Error Bound: The results of the theoretical error bound (dotted lines) and true error bound \( \|X - X^*\|_F^p \) versus the values of \( p \) with \( a = 2, \delta_{2ar} = \delta_{(a+1)r} = 0.1, \) and \( r = 6 \) in different perturbation levels \( \epsilon_A \) are provided in Fig. 7. The values of \( p \) vary from 0.1 to 0.9. From the observation of Fig. 7, when \( p \) \( \in [0.45, 0.90] \), \( \|X - X^*\|_F^p \) is smaller than the theoretical error bound.

5) Effect of Number of Measurements and Rank of Original Matrix on Reconstruction Efficiency: In Fig. 8, the relative error is plotted versus the number of measurements \( M \) in different \( \epsilon_A = 0, 0.05, 0.10, 0.15, 0.20 \) and \( p = 0.1, 0.3, 0.5, 0.7, 1 \), respectively. From Fig. 8, with the increase in the number of measurements or the decrease in perturbation level, the recovery performance of nonconvex Schatten \( p \)-minimization gradually improves. Moreover, Fig. 8(b) reveals that the performance of nonconvex Schatten \( p \)-minimization is better than that of convex nuclear norm minimization. In Fig. 9, we plot the relative error versus rank \( r \) of the matrix \( X \) for different \( \epsilon_A \) and \( p \), respectively. The results indicate that the smaller the rank of the matrix, the better the recovery performance, and choosing a smaller perturbation level or the values of \( p \) will improve the reconstruction effect of nonconvex Schatten \( p \)-minimization.

6) Comparison of Recovery Performance Between Nonconvex and Convex Methods: Furthermore, Fig. 10 offers the results concerning the recovery performance of the nonconvex method and the convex method for \( \epsilon_A = 0.05 \). The curves of relationship between the relative error and the rank \( r \) are described by nonconvex Schatten \( p \)-minimization and convex nuclear norm minimization, respectively. Fig. 10 displays that the performance of nonconvex method is superior to that of the convex method.
TABLE I

| Relative Error | $\varepsilon^r = 0.380$ | $\varepsilon^r = 1.138$ | $\varepsilon^r = 1.893$ | $\varepsilon^r = 2.548$ | $\varepsilon^r = 3.800$ |
|----------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| $\lambda_1 = 10^0$ | 0.5783 | 0.5708 | 0.5785 | 0.5766 | 0.5724 |
| $\lambda_2 = 10^{-1}$ | 0.2234 | 0.2238 | 0.2238 | 0.2275 | 0.2350 |
| $\lambda_3 = 10^{-2}$ | 0.02415 | 0.02816 | 0.03375 | 0.04101 | 0.05251 |
| $\lambda_4 = 10^{-3}$ | 0.02372 | 0.02766 | 0.03408 | 0.04111 | 0.05235 |
| $\lambda_5 = 10^{-4}$ | 0.02483 | 0.02753 | 0.03344 | 0.04041 | 0.05200 |
| $\lambda_6 = 10^{-5}$ | 0.02501 | 0.02772 | 0.03340 | 0.04104 | 0.05275 |
| $\lambda_7 = 10^{-6}$ | 0.02477 | 0.02885 | 0.03424 | 0.04073 | 0.05184 |

TABLE II

| PSNR|SSIM RESULTS ON LENA BY TWO METHODS |
|----------------|----------------|----------------|----------------|----------------|
| $\varepsilon_A = 0$ | 77.77 | 0.998 | 51.78 | 0.991 | 49.30 | 0.985 | 45.34 | 0.965 |
| $\varepsilon_A = 0.05$ | 56.39 | 0.997 | 51.42 | 0.990 | 47.48 | 0.980 | 44.37 | 0.960 |

Fig. 9. Reconstruction performance of completely perturbed nonconvex Schatten $p$-minimization varying rank $r$. (a) $p = 0.7$. (b) $\varepsilon_A = 0.05$ and $M = 0.5mn$.

Fig. 10. Reconstruction performance of nonconvex Schatten $p$-minimization and convex nuclear norm minimization, varying rank $r$ for $\varepsilon_A = 0.05$.

D. Real Data Experiments

In this section, the nonconvex model and convex model are applied to the reconstruction of real data, i.e., gray images. We use the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) targets to appraise the quality of the reconstructed pictures. For the definition of SSIM target, see [67]. Let $X$ indicate the original image with size $m \times n$ and $X^*$ stand for the estimated image. The definition of PSNR target is

$$\text{PSNR} = 10 \log_{10} \frac{255^2 \times m \times n}{\|X - X^*\|_F^2}.$$ 

In the experiments, the pixel values are normalized to $[0, 255]$. Similar to the case of synthetic data, we first take the classical Lena image as an example to give the curve of the relative error of the reconstruction with respect to the parameter $\lambda$ at different perturbation levels, and the specific results are given in Fig. 11(b). Observing the image shows that $\lambda = 10^{-6}$ is a suitable choice. We now apply two methods to recover Lena image at different perturbation levels, and the detailed results are given in Table II. In addition, two methods are applied to reconstruct the common 20 test images presented in Fig. 12, and the results are provided in Table III. From the above experiments, we conclude that: 1) in the case of partial perturbation (i.e., $\varepsilon = 0$), the nonconvex approach outperforms the convex reconstruction; 2) the recovery performance of both the methods becomes better as the perturbation level gradually decreases; and 3) with the exception of four of these images, overall, the nonconvex approach is more competitive for various perturbation levels in a completely perturbed scenario.

E. Analysis of Time Complexity

In this section, we analyze the time complexity of Algorithm 1. From the basic operations, the time complexity of lines 4 and 5 in the algorithm is $O(m^3n^3)$ and $O(mn^2)$, separately, where the time complexity of $(\tilde{A}^\top \tilde{A} + \rho I)^{-1}$ is $O(mn^2)$. Thereby, the time complexity of the whole algorithm is $O(m^3n^3)$. To reduce the time complexity of the algorithm, we use the matrix inversion lemma for the term $(\tilde{A}^\top \tilde{A} + \rho I)^{-1}$ to have

$$(\tilde{A}^\top \tilde{A} + \rho I)^{-1} = \rho^{-1}I - \rho^{-2}\tilde{A}^\top (\rho^{-1}\tilde{A} \tilde{A}^\top + I)^{-1}\tilde{A}.$$ (65)
VI. CONCLUSION

In this article, we investigate the completely perturbed problem using the nonconvex Schatten $p$-minimization for reconstructing low-rank matrices. We derive two sufficient conditions and the corresponding upper bounds of error estimation. The gained results reveal that the nonconvex Schatten $p$-minimization has the stability and robustness for reconstructing low-rank matrices with the existence of a total noise. The practical meaning of gained results not only can conduct the choice of the linear transformations for reconstructing low-rank matrices, that is, a linear transformation with a smaller RIC instead of a larger one can not only superior enhance the reconstruction performance, but also present a theoretical sustaining to approximation accurateness. Moreover, we consider the relationship between RIP and Schatten-$p$ RNSP, discovering that Schatten-$p$ RNSP can be derived from the condition $\delta_{tr} \leq ((t-1)/t)^{1/2}$ for any given $t > 1$. Finally, the numerical experiments further show the verification of our results, and the performance of nonconvex Schatten $p$-minimization is better than that of convex nuclear norm minimization in the complete perturbation situation. In the future, we will consider how to relax the conditions in this article and reduce the corresponding error estimates to even the tightest. In addition, the reconfigurability theory of the equivalent unconstrained model \((53)\) for model \((7)\), based on the restricted isometric property, is also a topic we will consider in the future.

ACKNOWLEDGMENT

The authors would like to thank Researcher Haozhe Wang, Alibaba, China, for a beneficial discussion of the article. Moreover, they also thank the Editor and anonymous Reviewers for their insightful comments and valuable suggestions, which led to a significant improvement of this article.

REFERENCES

[1] X. Chang, P. Ren, P. Xu, Z. Li, X. Chen, and A. Hauptmann, “A comprehensive survey of scene graphs: Generation and application,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 1–26, Jan. 2023.
[50] S. Liu, Q. Hu, P. Li, J. Zhao, M. Liu, and Z. Zhu, “ Speckle suppression based on weighted nuclear norm minimization and grey theory,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 5, pp. 2700–2708, May 2019.

[51] Y. Yu, Y. Zhang, and S. Yuan, “ Quaternion-based weighted nuclear norm minimization for color image denoising,” Neurocomputing, vol. 332, pp. 283–297, Mar. 2019.

[52] H. Deng, J. Tao, X. Song, and C. Zhang, “ Estimation of the parameters of a weighted nuclear norm model and its application in image denoising,” Inf. Sci., vol. 528, pp. 246–264, Aug. 2020.

[53] L. Liu, W. Huang, and D.-R. Chen, “ Exact minimum rank approximation via schatten $p$-norm minimization,” J. Comput. Appl. Math., vol. 267, pp. 218–227, Sep. 2014.

[54] F. Nie, H. Wang, H. Huang, and C. Ding, “ Joint schatten $p$-norm and $\ell_p$-norm robust matrix completion for missing value recovery,” Knowl. Inf. Syst., vol. 42, no. 3, pp. 525–544, Mar. 2015.

[55] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, “ Hyperspectral image restoration via iteratively regularized weighted schatten $p$-norm minimization,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[56] C. Xu et al., “ A unified convex surrogate for the Schatten-$p$ norm,” in Proc. 31st AAAI Conf. Artif. Intell., Feb. 2017, vol. 31, no. 1, pp. 1–7.

[57] H. Zhang et al., “ Low-rank matrix recovery via modified Schatten-$p$ norm minimization with convergence guarantees,” IEEE Trans. Image Process., vol. 29, pp. 3132–3142, 2019.

[58] H. Ye, H. Li, F. Cao, and L. Zhang, “ A hybrid truncated norm regularization method for matrix completion,” IEEE Trans. Image Process., vol. 28, no. 10, pp. 5171–5186, Oct. 2019.

[59] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege, “ Stable low-rank matrix recovery via null space properties,” Inf. Inference, A J. MA., vol. 5, no. 4, pp. 405–441, Dec. 2016.

[60] M.-C. Yue and A. M.-C. So, “ A perturbation inequality for concave functions of singular values and its applications in low-rank matrix recovery,” Appl. Comput. Harmon. Anal., vol. 40, no. 2, pp. 396–416, Mar. 2016.

[61] H. Ge, W. Chen, D. Li, and F. Wu, “ Performance analysis for unconstrained analysis based approaches,” Inverse Problems, vol. 38, no. 9, Aug. 2022, Art. no. 095004.

[62] C. Lu, J. Feng, Z. Lin, and S. Yan, “ Exact low tubal rank tensor recovery from Gaussian measurements,” in Proc. Int. Joint Conf. Artif. Intell. (IJCAI), Jul. 2018, pp. 2504–2510.

[63] F. Wen, P. Liu, Y. Liu, R. C. Qiu, and W. Yu, “ Robust sparse recovery in impulsive noise via $\ell_1/\ell_1$ optimization,” IEEE Trans. Signal Process., vol. 65, no. 1, pp. 105–118, Jan. 2017.

[64] W. Wang, J. Wang, and Z. Zhang, “ Block-sparse signal recovery via minimisation method,” IET Signal Process., vol. 12, no. 4, pp. 422–430, Jun. 2018.

[65] W. Wang and J. Wang, “ Enhancing matrix completion using a modified second-order total variation,” Discrete Dyn. Nature Soc., vol. 2018, pp. 1–12, Sep. 2018.

[66] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding, “ Robust matrix completion via joint Schatten $p$-norm and $\ell_p$-norm minimization,” in Proc. IEEE 12th Int. Conf. Data Mining, Dec. 2012, pp. 566–574.

[67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “ Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

Feng Zhang received the B.S. degree in information and computation science from the College of Mathematics Science, Inner Mongolia Normal University, Hohhot, China, in 2014, and the M.S. and Ph.D. degrees in statistics from the School of Mathematics and Statistics, Southwest University, Chongqing, China, in 2017 and 2020, respectively.

He is currently an Associate Professor and a Post-Doctoral Researcher with the School of Mathematics and Statistics, Southwest University. His work has been published in the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, Applied and Computational Harmonic Analysis, Inverse Problems, and Signal Processing. His work has been funded by the National Natural Science Foundation of China and the China Postdoctoral Science Foundation. His research interests include compressed sensing and tensor sparsity.

Dr. Zhang served as a Reviewer for Mathematical Reviews, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, Inverse Problems, and IEEE ACCESS.

Jianjun Wang (Member, IEEE) received the B.S. degree in mathematical education and the M.S. degree in fundamental mathematics from Ningxia University, Yinchuan, China, in 2000 and 2003, respectively, and the Ph.D. degree in applied mathematics from the Institute for Information and System Science, Xi’an Jiaotong University, Xi’an, China, in 2006.

He is currently a Professor with the School of Mathematics and Statistics, Southwest University, Chongqing, China. His research interests include machine learning, data mining, neural networks, and sparse learning.

Xinling Liu received the B.S. degree in mathematics and applied mathematics and the M.S. degree in probability theory and mathematical statistics from Southwest University, Chongqing, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics.

His research interests include compressive sensing and machine learning.

Jinpeng Jia received the B.S. degree from the School of Mathematics and Information Science, Northwest Normal University, Lanzhou, China, in 2005, the M.S. degree from the School of Science, Hangzhou Dianzi University, Hangzhou, China, in 2008, and the Ph.D. degree from the School of Mathematics and Statistics, Xidian University, Xi’an, China, in 2019.

He is currently an Associate Professor with the School of Mathematics and Statistics, Tianshui Normal University, Tianshui, China. His current research interests include sampled-data systems, nonlinear systems, stochastic systems, and event-triggered control.