The perturbation analysis of nonconvex low-rank matrix robust recovery

Jianwen Huang[1], Wendong Wang[1], Jianjun Wang[2]*, Feng Zhang[1]

[1] *School of Mathematic & Statistics, Tianshui University, Tianshui, Gansu, 741000, China*

[2] *School of Artificial Intelligence, Southwest University, Chongqing, 400715, China*

Abstract. In this paper, we bring forward a completely perturbed nonconvex Schatten p-minimization to address a model of completely perturbed low-rank matrix recovery. The paper that based on the restricted isometry property generalizes the investigation to a complete perturbation model thinking over not only noise but also perturbation, gives the restricted isometry property condition that guarantees the recovery of low-rank matrix and the corresponding reconstruction error bound. 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 $\delta_{2r} < 1$ [21]. The numerical experiments are conducted to show better performance, and provides outperformance of the nonconvex Schatten p-minimization method comparing with the convex nuclear norm minimization approach in the completely perturbed scenario.

Key words. Low rank matrix recovery; Perturbation of linear transformation; Nonconvex Schatten p-minimization

1 Introduction

Low-rank matrix recovery (LMR) 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:

$$y = \mathcal{A}(X) \tag{1.1}$$

where $\mathcal{A}: \mathbb{R}^{m\times n} \to \mathbb{R}^M$ is a known linear transformation (we suppose that $m \leq n$), $y \in \mathbb{R}^M$ is a given observation vector, and $X \in \mathbb{R}^{m\times n}$ is the matrix to be recovered. The objective of LMR is to find the lowest rank matrix based on $(y,\ \mathcal{A})$. If the observation y is corrupted by noise z, model (1.1) is changed into the following form

$$\hat{y} = \mathcal{A}(X) + z \tag{1.2}$$

*Corresponding author, E-mail: wjjmath@gmail.com, wjj@swu.edu.cn(J.J. Wang)

where $\hat{y}$ is the noisy measurement, and z is the additive noise independent of the matrix X. However, more LMR models can be encountered where not only the linear measurement y is contaminated by the noise vector z, but also the linear transformation $\mathcal{A}$ is perturbed by $\mathcal{E}$ for completely perturbed setting, namely, substitute the linear transformation $\mathcal{A}$ with $\hat{\mathcal{A}} = \mathcal{A} + \mathcal{E}$. The completely perturbed appearance arises in remote sensing[1], radar[2], source separation[3], etc. When $m = n$ and the matrix $X = \mathrm{diag}(x)$ ($x \in \mathbb{R}^m$) is diagonal, models (1.1) and (1.2) degenerates to the compressed sensing models

$$y = Ax, \tag{1.3}$$

$$\hat{y} = Ax + z \tag{1.4}$$

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 the problem (1.3) as the sparse signal recovery. For the completely perturbed model, the convex nuclear norm minimization is frequently considered [4] as follows:

$$\min_{\tilde{Z}\in\mathbb{R}^{m\times n}} \|\tilde{Z}\|_* \text{ s.t. } \|\hat{\mathcal{A}}(\tilde{Z}) - \hat{y}\|_2 \le \epsilon'_{\mathcal{A},r,y}, \tag{1.5}$$

where $\|\tilde{Z}\|_*$ is the nuclear norm of the matrix $\tilde{Z}$, that is, the sum of its singular values, and $\epsilon'_{\mathcal{A},r,y}$ is the total noise level. Problem (1.5) can be reduced to the l_1-minimization [5]

$$\min_{\tilde{z}\in\mathbb{R}^{n_1}} \|\tilde{z}\|_1 \text{ s.t. } \|\hat{A}\tilde{z} - \hat{y}\|_2 \le \epsilon'_{A,r,y}, \tag{1.6}$$

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 [7] showed that fewer measurements are required for exact reconstruction if l_1-norm is substituted with l_p-norm. There exist many work regarding reconstructing x via the l_p-minimization [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. In [7], numerical simulations demonstrated that fewer measurements are needed for exact reconstruction than when $p = 1$.

In this paper, we are interested in the completely perturbed model for the nonconvex Schatten p-minimization ($0 < p < 1$)

$$\min_{\tilde{Z}\in\mathbb{R}^{m\times n}} \|\tilde{Z}\|_p^p \text{ s.t. } \|\hat{\mathcal{A}}(\tilde{Z}) - \hat{y}\|_2 \le \epsilon'_{\mathcal{A},r,y}, \tag{1.7}$$

where $\|\tilde{Z}\|_p^p$ is the Schatten p quasi-norm of the matrix $\tilde{Z}$, that is, $\|\tilde{Z}\|_p^p = (\sum_i \sigma_i^p(\tilde{Z}))^{1/p}$ with $\sigma_i(\tilde{Z})$ being ith singular value of $\tilde{Z}$. Problem (1.7) can be returned to the l_p-minimization [6]

$$\min_{\tilde{z}\in\mathbb{R}^{M\times m}} \|\tilde{z}\|_p^p \text{ s.t. } \|\hat{A}\tilde{z} - \hat{y}\|_2 \le \epsilon'_{A,r,y}, \tag{1.8}$$

where $\|\tilde{z}\|_p^p = (\sum_i \tilde{z}_i^p)^{1/p}$ is the l_p-quasi-norm of the vector $\tilde{z}$. To the best of our knowledge, recently researches are considered only in unperturbed situation ($\mathcal{E} = 0$), that is, the linear transformation $\mathcal{A}$ is not perturbed by $\mathcal{E}$ (for related work, see [22], [23], [24], [25], [26], [27], [28], [29]). From the perspective of application, it is more practical to investigate the recovery of low-rank matrices in the scenario of complete perturbation.

In this paper, based on restricted isometry property (RIP), the performance of low-rank matrices reconstruction is showed by the nonconvex Schatten p-minimization in completely perturbed setting. The main

contributions of this paper are as follows. First, we present a sufficient condition for reconstruction of low-rank matrices via the nonconvex Schatten p-minimization. Second, the estimation accurateness between the optimal solution and the original matrix is described by a total noise and a best r-rank approximation error. The result reveals that stable and robust performance concerning reconstruction of low-rank matrices in existence of total noise. Third, numerical experiments are conducted to sustain the gained results, and demonstrate that the performance of nonconvex Schatten p-minimization can be better than that of convex nuclear norm minimization in completely perturbed model.

The rest of this paper is constructed as follows.

2 Notation and main results

Before presenting the main results, we first introduce the notion of RIC of a linear transformation $\mathcal{A}$, which is as follows.

Definition 2.1. *The restricted isometry constant (RIC) δ_r of a linear transformation $\mathcal{A}$ is the smallest constant such that*

$$(1-\delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1+\delta)\|X\|_F^2 \tag{2.9}$$

holds for all r-rank $X \in \mathbb{R}^{m\times n}$ (i.e., $rank(X) \leq r$), where $\|X\|_F := \sqrt{\langle X, X\rangle} = \sqrt{trace(X^\top X)}$ is the Frobenius norm of the matrix X.

Then we provide some notations similar to [4], which quantifying the perturbations $\mathcal{E}$ and z with the bounds:

$$\frac{\|\mathcal{E}\|_{op}}{\|\mathcal{A}\|_{op}} \leq \epsilon_{\mathcal{A}},\ \frac{\|\mathcal{E}\|_{op}^{(r)}}{\|\mathcal{A}\|_{op}^{(r)}} \leq \epsilon_{\mathcal{A}}^{(r)},\ \frac{\|z\|_2}{\|y\|_2} \leq \epsilon_y, \tag{2.10}$$

where $\|\mathcal{A}\|_{op} = \sup\{\|\mathcal{A}(X)\|_2/\|X\|_F :\ X \in \mathbb{R}^{m\times n} \setminus \{0\}\}$ is the operator norm of linear transformation $\mathcal{A}$, and $\|\mathcal{A}\|_{op}^{(r)} = \sup\{\|\mathcal{A}(X)\|_2/\|X\|_F :\ X \in \mathbb{R}^{m\times n} \setminus \{0\}$ and $\text{rank}(X) \leq r\}$, and representing

$$t_r = \frac{\|X_{[r]^c}\|_F}{\|X_{[r]}\|_F},\ s_r = \frac{\|X_{[r]^c}\|_*}{\sqrt{r}\|X_{[r]}\|_F},\ \kappa_{\mathcal{A}}^{(r)} = \frac{\sqrt{1+\delta_r}}{\sqrt{1-\delta_r}},\ \alpha_{\mathcal{A}} = \frac{\|\mathcal{A}\|_{op}}{\sqrt{1-\delta_r}}. \tag{2.11}$$

Here $X_{[r]}$ is the best r-rank approximation of the matrix X, its singular values are composed of r-largest singular values of the matrix X, and $X_{[r]^c} = X - X_{[r]}$. With notations and symbols above, we present our results for reconstruction of low-rank matrices via the completely perturbed nonconvex Schatten p-minimization.

Theorem 2.2. *For given relative perturbations $\epsilon_{\mathcal{A}}$, $\epsilon_{\mathcal{A}}^{(r)}$, $\epsilon_{\mathcal{A}}^{(2r)}$, and ϵ_y in (2.10), suppose the RIC for the linear transformation $\mathcal{A}$ fulfills*

$$\delta_{2ar} < \frac{2+\sqrt{2}a^{1/2-1/p}}{(1+\sqrt{2}a^{1/2-1/p})(1+\epsilon_{\mathcal{A}}^{(2ar)})^2} - 1 \tag{2.12}$$

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

$$t_r + s_r < \frac{1}{\kappa_{\mathcal{A}}^{(r)}}. \tag{2.13}$$

Then a minimizer X^* of problem (1.7) approximates the true matrix X with errors

$$\|X - X^*\|_F^p \leq C_1(\epsilon'_{\mathcal{A},r,y})^p + C_2\frac{\|X_{[r]^c}\|_p^p}{r^{1-p/2}}, \tag{2.14}$$

$$\|X - X^*\|_p^p \leq C'_1 r^{1-p/2}(\epsilon'_{\mathcal{A},r,y})^p + C'_2\|X_{[r]^c}\|_p^p, \tag{2.15}$$

where the total noise is

$$\epsilon'_{\mathcal{A},r,y} = \left[\frac{\epsilon_{\mathcal{A}}^{(r)}\kappa_{\mathcal{A}}^{(r)} + \epsilon_{\mathcal{A}}\alpha_{\mathcal{A}}t_r}{1 - \kappa_{\mathcal{A}}^{(r)}(t_r + s_r)} + \epsilon_y\right]\|y\|_2, \tag{2.16}$$

and

$$C_1 = \frac{2^p(1 + a^{p/2-1})(1 + \hat{\delta}_{(a+1)r})^{p/2}}{(1 - \hat{\delta}_{(a+1)r})^p - a^{p/2-1}(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}, \tag{2.17}$$

$$C_2 = 2a^{p/2-1}[1 + \frac{(1 + a^{p/2-1})(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}{(1 - \hat{\delta}_{(a+1)r})^p - a^{p/2-1}(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}], \tag{2.18}$$

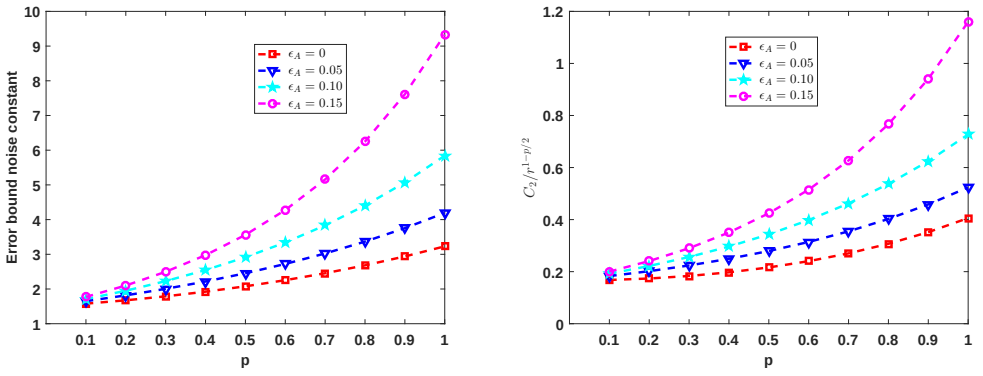
$$C'_1 = \frac{2^{p+1}(1 + a)^{1-p/2}(1 + \hat{\delta}_{(a+1)r})^{p/2}}{(1 - \hat{\delta}_{(a+1)r})^p - a^{p/2-1}(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}, \tag{2.19}$$

$$C'_2 = 2 + \frac{4(1 + a)^{1-p/2}a^{p/2-1}(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}{(1 - \hat{\delta}_{(a+1)r})^p - a^{p/2-1}(\hat{\delta}^2_{(a+1)r} + \hat{\delta}^2_{2ar})^{p/2}}, \tag{2.20}$$

where $\hat{\delta}_{(a+1)r} = (1 + \delta_{(a+1)r})(1 + \epsilon_A^{((a+1)r)})^2 - 1, \quad \hat{\delta}_{2ar} = (1 + \delta_{2ar})(1 + \epsilon_A^{(2ar)})^2 - 1.$

Remark 2.3. *Theorem 2.2 gives a sufficient conditions for the reconstruction of low-rank matrices via nonconvex Schatten p-minimization in completely perturbed scenario. Condition (2.12) of the Theorem extends the assumption of l_p situation in [6] to the nonconvex Schatten p-minimization. Observe that as the value of p becomes large, the bound of RIC δ_{2ar} reduces, which reveals that smaller value of p can induce weaker reconstruction guarantee. Particularly, when $p \to 0$ $(a > 1)$ ((1.7) degenerates to the rank minimization: $\min_{\tilde{Z}\in\mathbb{R}^{m\times n}} rank(\tilde{Z})$ s.t. $\|\hat{\mathcal{A}}(\tilde{Z}) - \hat{y}\|_2 \leq \epsilon'_{\mathcal{A},r,y}$), it leads to the RIP condition $\delta_{2r} < 2/(1 + \epsilon_{\mathcal{A}}^{(2ar)})^2 - 1$ for reconstruction of low-rank matrices via the rank minimization, to the best of our knowledge, the current optimal recovery condition about RIP is $\delta_{2r} < 1$ to ensure exact reconstruction for r-rank matrices via rank minimization [21], therefore the Theorem extends that condition to the scenario of presence of noise and r-rank matrices. Furthermore, when $m = n$ and the matrix $X = diag(x)$ $(x \in \mathbb{R}^m)$ is diagonal, the Theorem reduces to the case of compressed sensing given by [6].*

Remark 2.4. *Under the requirement (2.12), one can easily check that the condition (2.13) is satisfied. Besides, when $rank(X) \leq r$, the condition (2.13) holds. Additionally, the inequalities (2.14) and (2.15) in Theorem 2.2 which exploit two kinds of metrics provide upper bound estimations on the reconstruction 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 aren't noise (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 (2.14), 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. 2.1.*



(a) (b)

Fig. 2.1: The error bound noise constant C_1, and the error bound compressibility constant $C_2/r^{1-p/2}$ versus p for $a = 5, \delta_{(a+1)r} = \delta_{2ar} = 0.05$ in (a) and (b), respectively.

Remark 2.5. *When the matrix X is a strictly r-rank matrix (i.e., $X = X_{[r]}$), a minimizer X^* of problem (1.7) approximates the true matrix X with errors*

$$\|X - X^*\|_F \leq C_1^{1/p} \epsilon'_{\mathcal{A},r,y},$$
$$\|X - X^*\|_p \leq C_1'^{1/p} r^{1/p-1/2} \epsilon'_{\mathcal{A},r,y},$$

where

$$\epsilon'_{\mathcal{A},r,y} = [\epsilon_{\mathcal{A}}^{(r)} \kappa_{\mathcal{A}}^{(r)} + \epsilon_y] \|y\|_2.$$

In the case of $\mathcal{E} = 0$, that is, there doesn't exist perturbation in the linear transformation $\mathcal{A}$, then $\hat{\delta}_{(a+1)r} = \delta_{(a+1)r}$, $\hat{\delta}_{2ar} = \delta_{2ar}$. In the case that $m = n$, the matrix $X = diag(x)$ $(x \in \mathbb{R}^m)$ is diagonal (i.e., the results of Theorem reduce to the case of compressed sensing), $p = 1$ and $a = 1$, our result contains that of Theorem 2 in [5].

3 Proofs of the main results

In this part, we will provide the proofs of main results. In order to prove our main results, we need the following auxiliary lemmas. Firstly, we give Lemma 3.1 which incorporates an important inequality associating with δ_r and $\hat{\delta}_r$.

Lemma 3.1. *(RIP for $\hat{\mathcal{A}}$ [4]) Given the RIC δ_r related with linear transformation $\mathcal{A}$ and the relative perturbation $\epsilon_{\mathcal{A}}^{(r)}$ corresponded with linear transformation $\mathcal{E}$, fix the constant $\hat{\delta}_{r,\max} = (1 + \delta_r)(1 + \epsilon_{\mathcal{A}}^{(r)})^2 - 1$. Then the RIC $\hat{\delta}_r \leq \hat{\delta}_{r,\max}$ for $\hat{\mathcal{A}} = \mathcal{A} + \mathcal{E}$ is the smallest nonnegative constant such that*

$$(1 - \hat{\delta}_r)\|X\|_F^2 \leq \|\hat{\mathcal{A}}(X)\|_2^2 \leq (1 + \hat{\delta}_r)\|X\|_F^2 \tag{3.21}$$

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

We will employ the fact that $\hat{\mathcal{A}}$ maps low-rank orthogonal matrices to nearly sparse orthogonal vectors, which is given by [20].

Lemma 3.2. *([20]) For all $X,\ Y$ satisfying $\langle X, Y\rangle = 0$, and rank$(X) \le r_1$, rank$(Y) \le r_2$,*

$$\left|\left\langle \hat{\mathcal{A}}(X), \hat{\mathcal{A}}(Y)\right\rangle\right| \le \hat{\delta}_{r_1+r_2}\|X\|_F\|Y\|_F. \tag{3.22}$$

Moreover, the following lemma will be utilized in the proof of main result, which combines with Lemma 2.3 [21] and Lemma 2.2 [25].

Lemma 3.3. *Assume that $X,\ Y \in \mathbb{R}^{m\times n}$ obey $X^\top Y = 0$ and $XY^\top = 0$. Let $0 < p \le 1$. Then*

$$\|X+Y\|_p^p = \|X\|_p^p + \|Y\|_p^p,\ \ \|X+Y\|_p \ge \|X\|_p + \|Y\|_p, \tag{3.23}$$

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

For any matrix $X \in \mathbb{R}^{m\times n}$, we represent the singular values decomposition (SVD) of X as

$$X = U\mathrm{diag}(\sigma(X))V^\top,$$

where $\sigma(X) := (\sigma_1(X), \cdots, \sigma_m(X))$ is the vector of the singular values of X, U and V are respectively the left and right singular value matrices of X.

Proof of the theorem 2.2. Let X denote the original matrix to be recovered and X^* denote the optimal solution of (1.7). Let $Z = X - X^*$, and based on the SVD of X, its SVD is given by

$$U^\top ZV = U_1\mathrm{diag}(\sigma(U^\top ZV))V_1^\top,$$

where $U_1,\ V_1 \in \mathbb{R}^{m\times m}$ are orthogonal matrices, and $\sigma(U^\top ZV)$ stands for the vector comprised of the singular values of $U^\top ZV$. Let T_0 is the set composed of the locations of the r largest magnitudes of elements of $\sigma(X)$. We adopt technology similar to the reference [6] to partition $\sigma(U^\top ZV)$ into a sum of vectors $\sigma_{T_i}(U^\top ZV)$ $(i = 0, 1, \cdots, J)$, where T_1 is the set composed of the locations of the ar largest magnitudes of entries of $\sigma_{T_0^c}(U^\top ZV)$, T_2 is the set composed of the locations of the second ar largest magnitudes of entries of $\sigma_{T_0^c}(U^\top ZV)$, and so forth (except possibly T_J). Then $Z = \sum_{i=0}^{J} Z_{T_i}$ where $Z_{T_i} = UU_1\mathrm{diag}(\sigma_{T_i}(U^\top ZV))(VV_1)^\top$, $i = 0, 1, \cdots, J$. One can easily verify that $Z_{T_i}^\top Z_{T_j} = 0$ and $Z_{T_i}Z_{T_j}^\top = 0$ for all $i \ne j$, and $\mathrm{rank}(Z_{T_0}) \le r$, $\mathrm{rank}(Z_{T_j}) \le ar$, $i = 0, 1, \cdots, J$. For simplicity, denote $T_{01} = T_0 \bigcup T_1$. Then, we have (see (22) in [24], Lemma 2.6 [27])

$$\|Z_{T_0^c}\|_p^p \le \|Z_{T_0}\|_p^p + 2\|X_{[r]^c}\|_p^p. \tag{3.24}$$

By the decomposition of Z, for each $l \in T_i,\ k \in T_{i-1},\ i \ge 2$, $\sigma_{T_i}(U^\top ZV)[l] \le \sigma_{T_{i-1}}(U^\top ZV)[k]$, it implies that

$$(\sigma_{T_i}(U^\top ZV)[l])^p \le \frac{\sum_{k=1}^{ar}(\sigma_{T_{i-1}}(U^\top ZV)[l])^p}{ar} = \frac{\|\sigma_{T_{i-1}}(U^\top ZV)\|_p^p}{ar} = \frac{\|Z_{T_{i-1}}\|_p^p}{ar}, \tag{3.25}$$

which deduces

$$\|Z_{T_i}\|_F^2 \le (ar)^{1-\frac{2}{p}}\|Z_{T_{i-1}}\|_p^2. \tag{3.26}$$

Thereby,

$$\|Z_{T_i}\|_F^p \leq (ar)^{\frac{p}{2}-1}\|Z_{T_{i-1}}\|_p^p. \tag{3.27}$$

Notice that $Z_{T_i}^\top Z_{T_j} = 0$ and $Z_{T_i} Z_{T_j}^\top = 0$ for all $i \neq j$, due to Lemma 3.3 and (3.27), then we can get

$$\sum_{i\geq 2}\|Z_{T_i}\|_F^p \leq (ar)^{\frac{p}{2}-1}\sum_{i\geq 2}\|Z_{T_{i-1}}\|_p^p = (ar)^{\frac{p}{2}-1}\|Z_{T_0^c}\|_p^p. \tag{3.28}$$

By the inequality $\|Z_{T_0}\|_F^p \leq \|Z_{T_{01}}\|_F^p$ and Hölder's inequality, we get

$$\|Z_{T_0}\|_p^p \leq r^{1-\frac{p}{2}}\|Z_{T_{01}}\|_F^p. \tag{3.29}$$

From (3.24), (3.28), (3.29) and the inequality that for every fixed $n \in \mathbb{N}$, and any $0 < \alpha \leq 1$, $(\sum_{i=1}^n x)^\alpha \leq \sum_{i=1}^n x^\alpha$ for every $x_i \geq 0,\ i = 1, \cdots, n$, it follows

$$\|Z_{T_{01}^c}\|_F^p = (\sum_{i\geq 2}\|Z_{T_i}\|_F^2)^{\frac{p}{2}} \leq \sum_{i\geq 2}\|Z_{T_i}\|_F^p \leq (ar)^{\frac{p}{2}-1}(r^{1-\frac{p}{2}}\|Z_{T_{01}}\|_F^p + 2\|X_{[r]^c}\|_p^p). \tag{3.30}$$

Since

$$\begin{aligned}\|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^2 =& < \hat{\mathcal{A}}(Z_{T_{01}}), \hat{\mathcal{A}}(Z_{T_{01}}) > \\ =& < \hat{\mathcal{A}}(Z_{T_{01}}), \hat{\mathcal{A}}(Z) > - < \hat{\mathcal{A}}(Z_{T_{01}}), \sum_{i\geq 2}\hat{\mathcal{A}}(Z_{T_i}) > \\ \leq& \|\hat{\mathcal{A}}(Z_{T_{01}})\|_2\|\hat{\mathcal{A}}(Z)\|_2 + \sum_{i\geq 2}| < \hat{\mathcal{A}}(Z_{T_{01}}), \hat{\mathcal{A}}(Z_{T_i}) > |,\end{aligned} \tag{3.31}$$

we get

$$\|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^{2p} \overset{\text{(a)}}{\leq} \|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^p\|\hat{\mathcal{A}}(Z)\|_2^p + \sum_{i\geq 2}| < \hat{\mathcal{A}}(Z_{T_{01}}), \hat{\mathcal{A}}(Z_{T_i}) > |^p, \tag{3.32}$$

where (a) follows from the fact that $(a+b)^p \leq a^p + b^p$ for nonnegative a and b.

Additionally, by the minimality of X^*, we get

$$\|\hat{\mathcal{A}}(Z)\|_2^2 \leq \|\hat{y} - \hat{\mathcal{A}}(X)\|_2^2 + \|\hat{y} - \hat{\mathcal{A}}(X^*)\|_2^2 \leq 2\epsilon'_{\mathcal{A},r,y}. \tag{3.33}$$

Since $Z_{T_{01}}$ is $(a+1)r$-rank and Z_{T_i} is ar-rank, $i \geq 2$, by applying the RIP of $\hat{\mathcal{A}}$ and combination with (3.32) and (3.33), we get

$$\|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^{2p} \leq (2\epsilon'_{\mathcal{A},r,y})^p(1+\hat{\delta}_{(a+1)r})^{\frac{p}{2}}\|Z_{T_{01}}\|_F^p + \sum_{i\geq 2}| < \hat{\mathcal{A}}(Z_{T_{01}}), \hat{\mathcal{A}}(Z_{T_i}) > |^p. \tag{3.34}$$

Because $< Z_{T_i}, Z_{T_j} >$ for all $i \neq j$, and Z_{T_0} is r-rank, by Lemma 3.2 and (3.30), we get

$$\begin{aligned}\|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^{2p} \leq& (2\epsilon'_{\mathcal{A},r,y})^p(1+\hat{\delta}_{(a+1)r})^{\frac{p}{2}}\|Z_{T_{01}}\|_F^p + (\hat{\delta}_{(a+1)r}\|Z_{T_0}\|_F + \hat{\delta}_{2ar}\|Z_{T_1}\|_F)^p \sum_{i\geq 2}\|Z_{T_i}\|_F^p \\ \leq& (2\epsilon'_{\mathcal{A},r,y})^p(1+\hat{\delta}_{(a+1)r})^{\frac{p}{2}}\|Z_{T_{01}}\|_F^p \\ &+ (\hat{\delta}_{(a+1)r}\|Z_{T_0}\|_F + \hat{\delta}_{2ar}\|Z_{T_1}\|_F)^p(ar)^{\frac{p}{2}-1}(r^{1-\frac{p}{2}}\|Z_{T_{01}}\|_F^p + 2\|X_{[r]^c}\|_p^p)\end{aligned} \tag{3.35}$$

From (2.12), one can easily check that

$$a^{\frac{p}{2}-1}(\hat{\delta}^2_{(a+1)r}+\hat{\delta}^2_{2ar})^{\frac{p}{2}} < (1-\hat{\delta}_{(a+1)r})^{\frac{p}{2}}. \tag{3.36}$$

By (3.35), (3.36) and the inequality $\|\hat{\mathcal{A}}(Z_{T_{01}})\|_2^p \geq (1-\hat{\delta}_{(a+1)r})^{\frac{p}{2}}\|Z_{T_{01}}\|_F^p$, one can get

$$\begin{aligned}\|Z_{T_{01}}\|_F^p \leq & \frac{2^p(1+a^{\frac{p}{2}-1})(1+\hat{\delta}_{(a+1)r})^{\frac{p}{2}}}{(1-\hat{\delta}_{(a+1)r})^{\frac{p}{2}}-a^{\frac{p}{2}-1}(\hat{\delta}^2_{(a+1)r}+\hat{\delta}^2_{2ar})^{\frac{p}{2}}}(\epsilon'_{\mathcal{A},r,y})^p \\ & + \frac{2a^{\frac{p}{2}-1}(\hat{\delta}^2_{(a+1)k}+\hat{\delta}^2_{2ar})^{\frac{p}{2}}}{(1-\hat{\delta}_{(a+1)r})^{\frac{p}{2}}-a^{\frac{p}{2}-1}(\hat{\delta}^2_{(a+1)r}+\hat{\delta}^2_{2ar})^{\frac{p}{2}}}\frac{\|X_{[r]^c}\|_p^p}{r^{1-\frac{p}{2}}} \\ =: & \beta(\epsilon'_{\mathcal{A},r,y})^p+\gamma\frac{\|X_{[r]^c}\|_p^p}{r^{1-\frac{p}{2}}},\end{aligned} \tag{3.37}$$

consequently,

$$\begin{aligned}\|Z_{T_0}\|_p^p &\leq r^{1-\frac{p}{2}}\|Z_{T_0}\|_F^p \\ &\leq \beta r^{1-\frac{p}{2}}(\epsilon'_{\mathcal{A},r,y})^p+\gamma\|X_{[r]^c}\|_p^p.\end{aligned} \tag{3.38}$$

Thus, from (3.30) and (3.37), we get

$$\begin{aligned}\|Z\|_F^p &\leq \|Z_{T_{01}}\|_F^p+\|Z_{T_{01}^c}\|_F^p \\ &\leq C_1(\epsilon'_{\mathcal{A},r,y})^p+C_2\frac{\|X_{[r]^c}\|_p^p}{r^{1-\frac{p}{2}}};\end{aligned} \tag{3.39}$$

in addition, a combination of (3.24) and (3.38), one can get

$$\begin{aligned}\|Z\|_p^p &\leq \|Z_{T_0}\|_p^p+\|Z_{T_0^c}\|_p^p \\ &\leq C_1' r^{1-\frac{p}{2}}(\epsilon'_{\mathcal{A},r,y})^p+C_2'\|X_{[r]^c}\|_p^p,\end{aligned} \tag{3.40}$$

where the constants C_1, C_2, C_1' and C_2' are defined in Theorem 2.2. The proof is complete. □

4 Numerical experiments

In this section, we carry out some numerical experiments to sustain verification of our theoretical results, we implement all experiments in MATLAB 2016a running on a PC with an Inter core i7 processor (3.6 GHz) with 8 GB RAM. In order to address the completely perturbed nonconvex Schatten p-minimization model, we employ the alternating direction method of multipliers (ADMM) method, which is often applied in compressed sensing and sparse approximation [30], [31], [32], [33]. The constrained optimization problem (1.7) can be transformed into an equivalent unconstrained form

$$\min_{\tilde{Z}\in\mathbb{R}^{m\times n}} \lambda\|\tilde{Z}\|_p^p+\frac{1}{2}\|\hat{A}\mathrm{vec}(\tilde{Z})-\hat{y}\|_2^2, \tag{4.41}$$

where $\hat{A}\in\mathbb{R}^{M\times mn}$, $\mathrm{vec}(\tilde{Z})$ represents the vectorization of $\tilde{Z}$. Hence, $\hat{A}\mathrm{vec}(\tilde{Z})$ presents the linear map $\hat{\mathcal{A}}(\tilde{Z})$. Then, introducing an auxiliary variable $W\in\mathbb{R}^{m\times n}$, the problem (4.41) can be equivalently turned into

$$\min_{W,\ \tilde{Z}\in\mathbb{R}^{m\times n}} \lambda\|W\|_p^p + \frac{1}{2}\|\hat{A}\text{vec}(\tilde{Z}) - \hat{y}\|_2^2 \text{ s.t. } \tilde{Z} = W. \tag{4.42}$$

The augmented Lagrangian function is provided by

$$L_\rho(\tilde{Z}, W, Y) = \lambda\|W\|_p^p + \frac{1}{2}\|\hat{A}\text{vec}(\tilde{Z}) - \hat{y}\|_2^2 + < Y, \tilde{Z} - W > + \frac{\rho}{2}\|\tilde{Z} - W\|_F^2, \tag{4.43}$$

where $Y \in \mathbb{R}^{m\times n}$ is dual variable, and $\rho > 0$ is a penalty parameter. Then, ADMM used to (4.43) comprises of the iterations as follows

$$\tilde{Z}^{k+1} = \arg\min_{\tilde{Z}} \frac{1}{2}\|\hat{A}\text{vec}(\tilde{Z}) - \hat{y}\|_2^2 + \frac{\rho}{2}\|\tilde{Z} - (W^k - \frac{Y^k}{\rho})\|_F^2, \tag{4.44}$$

$$W^{k+1} = \arg\min_{W} \lambda\|W\|_p^p + \frac{\rho}{2}\|\tilde{Z}^{k+1} - (W - \frac{Y^k}{\rho})\|_F^2, \tag{4.45}$$

$$Y^{k+1} = Y^k + \rho(\tilde{Z}^{k+1} - W^{k+1}). \tag{4.46}$$

All solving processes are concluded in Algorithm 4.1.

Algorithm 4.1 : Solve problem (1.7) 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 $\|E\| = \epsilon_A\|A\|$, $p \in (0, 1]$.

2: Initialize $\hat{\mathcal{Z}}^0 = W^0 = Y^0$, $\gamma = 1.1$, $\lambda_0 = 10^{-6}$, $\lambda_{\max} = 10^{10}$, $\rho = 10^{-6}$, $\varepsilon = 10^{-8}$, $k = 0$.

3: **while** not converged **do**

4: Updated $\tilde{Z}^{k+1}$ by

$$\tilde{z} = (\hat{A}^\top\hat{A} + \rho I)^{-1}\left(\hat{A}^\top\hat{y} + \rho\text{vec}(W^k) - \text{vec}(Y^k)\right);$$

$\tilde{Z}^{k+1} \leftarrow \tilde{z}$: reshape $\tilde{z}$ to the matrix $\tilde{Z}^{k+1}$ of size $m \times n$.

5: Update W^{k+1} by

$$\arg\min_{W} \lambda\|W\|_p^p + \frac{\rho}{2}\|\tilde{Z}^{k+1} - (W - \frac{Y^k}{\rho})\|_F^2;$$

6: Update Y^{k+1} by

$$Y^{k+1} = Y^k + \rho(\tilde{Z}^{k+1} - W^{k+1});$$

7: Update λ_{j+1} by $\lambda_{j+1} = \min(\gamma\lambda_j, \lambda_{\max})$;

8: Check the convergence conditions

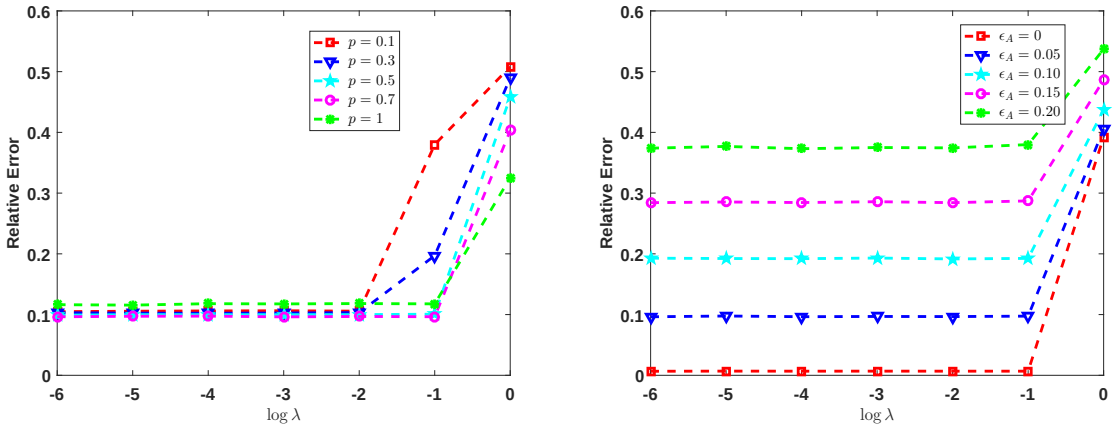
$$\|\tilde{Z}^{k+1} - \tilde{Z}^k\|_\infty \le \varepsilon,\ \|W^{k+1} - W^k\|_\infty \le \varepsilon,$$

$$\|\hat{A}\text{vec}(\tilde{Z}^{k+1}) - \hat{y}\|_\infty \le \varepsilon,\ \|\tilde{Z}^k - W^{j+1}\|_\infty \le \varepsilon.$$

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. We generate $X \in \mathbb{R}^{m\times n}$ 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. We select $M = 660$, $m = n = 30$ and $r = 0.2m$. With X and A, the measurements y are produced by $y = A\text{vec}(X) + z$, where z is the Gaussian

noise. The perturbation matrix E is with its entries following Gaussian distribution, which fulfills $\|E\| = \epsilon_A \|A\|$, where ϵ_A denotes the perturbation level of A and its value is not fixed. The perturbed matrix $\hat{A}$, $\hat{A} = A + E$, is used in (4.44). To avoid the randomness, we perform 100 times independent trails as well as the average result in all test.

To look for a proper parameter λ that derives the better recovery effect, we carry out two sets of trails. Fig. 4.2 (a) and (b) respectively plot the parameter λ and relative error (RelError, $\|X - X^*\|_F / \|X\|_F$) results in different p values and perturbation level ϵ_A. λ ranges from 10^{-6} to 1. Fig. 4.2 (a) and (b) show that $\lambda \in [10^{-6}, 10^{-2}]$ is relatively suitable.



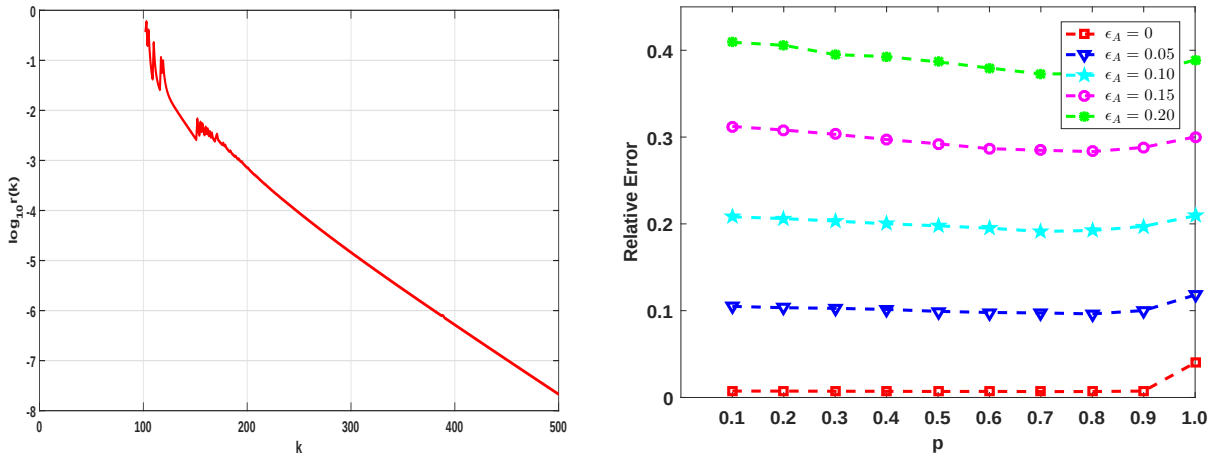
(a) (b)

Fig. 4.2: Parameter selection λ for (a) $\epsilon_A = 0.05$, (b) $p = 0.7$

By giving $\lambda = 10^{-6}$, we consider the convergence of Algorithm 4.1. Fig. 4.3(a) presents the relationship between relative neighboring iteration error (RNIE, $r(k) = \|X^{k+1} - X^k\|_F / \|X^k\|_F$) and number of iterations k. One can easily see that with the increasing of iterations, RNIE decreases quickly, and when $k \geq 250$, $r(k) < 10^{-4}$. The results that relative error versus the values of p in different ϵ_A are showed in Fig. 4.3(b). Fig. 4.3(b) indicates that the proper choice of the size of p will be helpful to facilitate the performance of nonconvex Schatten p-minimization.

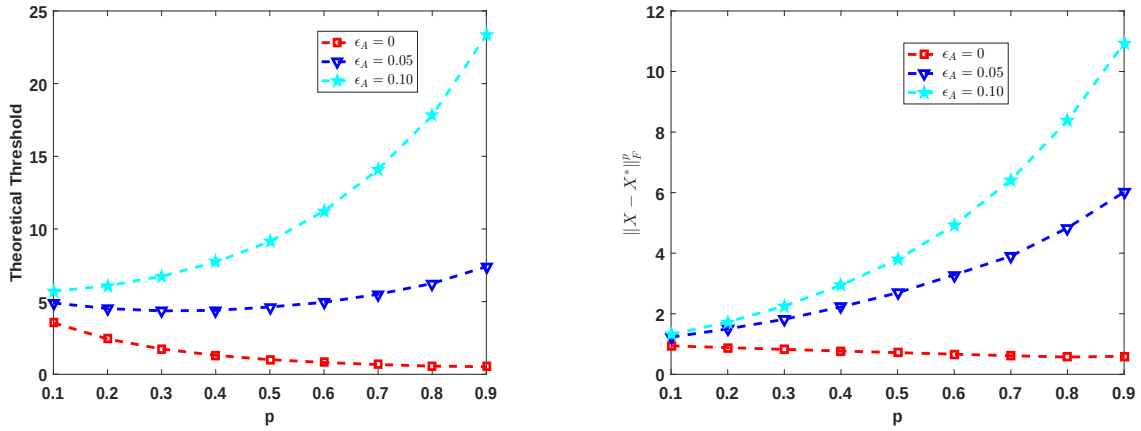
The theoretical error bound and $\|X - X^*\|_F^p$ versus the values of p with $a = 2$, $\delta_{2ar} = \delta_{(a+1)r} = 0.1$ and $r = 6$ in different perturbation level ϵ_A, the results are provided in Figs. 4.4 (a) and (b). The values of p vary from 0.1 to 0.9. From the observation of Fig. 4.4, $\|X - X^*\|_F^p$ is smaller than the theoretical error bound.

In Fig. 4.5, 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. 4.5, with the increase of number of measurements or the decrease of perturbation level, the recovery performance of nonconvex Schatten p-minimization gradually improves. Moreover, Fig. 4.5(b) reveals that the performance of nonconvex Schatten p-minimization is better than that of convex nuclear norm minimization. In Fig. 4.6, we plot the the relative error versus the rank r of the matrix X for different ϵ_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



(a) (b)

Fig. 4.3: Convergence analysis for Algorithm 4.1 and relative error versus p. (a) Convergence analysis, $\epsilon_A = 0.05$, (b)Reconstruction performance of completely perturbed nonconvex Schatten p-minimization, varying p



(a) (b)

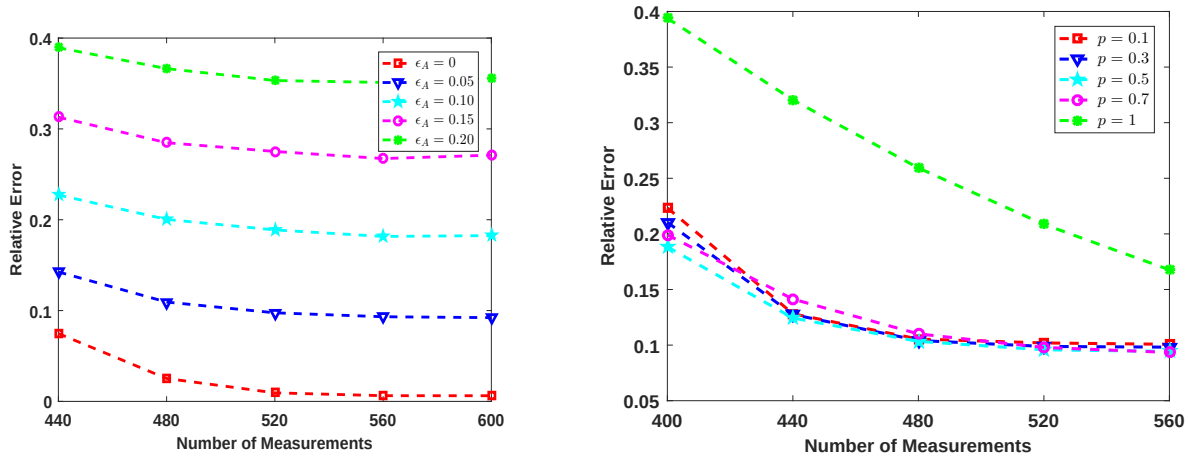
Fig. 4.4: Theoretical error bound and $\|X - X^*\|_F^p$ versus p for (a) $a = 2$, $\delta_{2ar} = \delta_{(a+1)r} = 0.1$, (b) $r = 6$.

the reconstruction effect of nonconvex Schatten p-minimization.

Furthermore, Fig. 4.7 offers the results concerning the recovery performance of the nonconvex method and the convex method for the $\epsilon_A = 0.05$. 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. 4.7 displays that the performance of nonconvex method is superior to that of the convex method.

5 Conclusion

In this paper, we investigate the completely perturbed problem employing the nonconvex Schatten p-minimization for reconstructing low-rank matrices. We derive a sufficient condition and the corresponding upper bounds of error estimation. The gained results reveal the nonconvex Schatten p-minimization has the



(a) (b)

Fig. 4.5: Reconstruction performance of completely perturbed nonconvex Schatten p-minimization versus number of measurements M. (a) $p = 0.7$, (b)$\epsilon_A = 0.05$

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 superior enhance the reconstruction performance, but also can also present a theoretical sustaining to approximation accurateness. Moreover, 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.

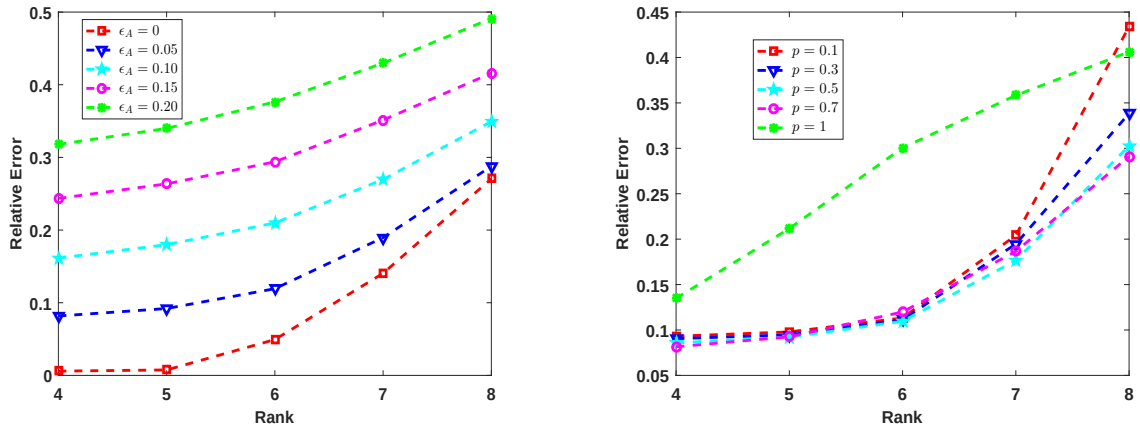
6 Acknowledgements

This work was supported by Natural Science Foundation of China (Grant Nos. 61673015, 61273020), Fundamental Research Funds for the Central Universities (Grant Nos. XDJK2015A007, XDJK2018C076, SWU1809002), Graduate Student Scientific Research Innovation Projects in Chongqing (Grant No. CYB19083), Youth Science and technology talent development project (Grant No. Qian jiao he KY zi [2018]313), Science and technology Foundation of Guizhou province (Grant No. Qian ke he Ji Chu [2016]1161), Guizhou province natural science foundation in China (Grant No. Qian Jiao He KY [2016]255).

References

[1] Fannjiang A C, Strohmer T, Yan P. Compressed remote sensing of sparse objects. SIAM J Imaging Sci 2010; 3(3):595-618.

[2] Herman M A, Strohmer T. High-resolution radar via compressed sensing. IEEE Trans Signal Process 2009; 57(6):2275-84.



(a) (b)

Fig. 4.6: Reconstruction performance of completely perturbed nonconvex Schatten p-minimization varying rank r. (a) $p = 0.7$, (b) $\epsilon_A = 0.05$

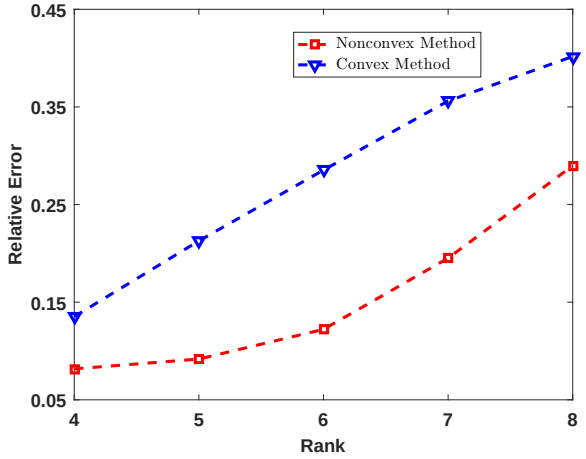


Fig. 4.7: Reconstruction performance of nonconvex Schatten p-minimization and convex nuclear norm minimization, varying rank r for $\epsilon_A = 0.05$

[3] Blumensath T, Davies M. Compressed sensing and source separation. Proceedings of the 7th international conference on Independent component analysis and signal separation, Springer-Verlag Berlin, Heidelberg; 2007. p. 341-8.

[4] Jianwen Huang, Jianjun Wang, Feng Zhang, et al. The perturbation analysis of low-rank matrix stable recovery. Submitted for publication, 2019

[5] Herman M A, Strohmer T. General Deviants: An Analysis of Perturbation in Compressed Sensing. Signal Process 2010; 4(2):342-9.

[6] Ince T, Nacaroglu A. On the perturbation of measurement matrix in non-convex compressed sensing[J]. Signal Processing, 2014, 98: 143-149.

[7] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization[J]. IEEE Signal Processing Letters, 2007, 14(10): 707-710.

[8] Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing[J]. Inverse Problems, 2008, 24(3): 035020.

[9] Foucart S, Lai M J. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q < 1$[J]. Applied and Computational Harmonic Analysis, 2009, 26(3): 395-407.

[10] M Lai, L Liu, A new estimate of restricted isometry constants for sparse solutions (2011). http://www.math.uga.edu/ mjlai/papers/LaiLiu11.pdf

[11] Lai M J, Xu Y, Yin W. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization[J]. SIAM Journal on Numerical Analysis, 2013, 51(2): 927-957.

[12] Wang Y, Wang J, Xu Z. Restricted p-isometry properties of nonconvex block-sparse compressed sensing[J]. Signal Processing, 2014, 104: 188-196.

[13] Song C B, Xia S T. Sparse Signal Recovery by ℓ_q Minimization Under Restricted Isometry Property[J]. IEEE Signal Processing Letters, 2014, 21(9): 1154-1158.

[14] Wang J, Zhang J, Wang W, et al. A perturbation analysis of nonconvex block-sparse compressed sensing[J]. Communications in Nonlinear Science and Numerical Simulation, 2015, 29(1-3): 416-426.

[15] Wen J, Li D, Zhu F. Stable recovery of sparse signals via lp-minimization[J]. Applied and Computational Harmonic Analysis, 2015, 38(1): 161-176.

[16] Wen F, Pei L, Yang Y, et al. Efficient and robust recovery of sparse signal and image using generalized nonconvex regularization[J]. IEEE Transactions on Computational Imaging, 2017, 3(4): 566-579.

[17] Yi Gao, Jigen Peng, Shigang Yue. Stability and robustness of the l_2/l_q-minimization for block sparse recovery. Signal Processing, 2017, 137: 287-297.

[18] Zhang R, Li S. Optimal RIP bounds for sparse signals recovery via ℓ_p minimization[J]. Applied and Computational Harmonic Analysis, 2017.

[19] Wen F, Chu L, Liu P, et al. A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning[J]. IEEE Access, 2018, 6: 69883-69906.

[20] Candes E J, Plan Y. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements[J]. IEEE Transactions on Information Theory, 2011, 57(4): 2342-2359.

[21] Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM review, 2010, 52(3): 471-501.

[22] Mohan K, Fazel M. New restricted isometry results for noisy low-rank recovery[C]. 2010 IEEE International Symposium on Information Theory. IEEE, 2010: 1573-1577.

[23] Dvijotham K, Fazel M. A nullspace analysis of the nuclear norm heuristic for rank minimization[C]. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010: 3586-3589.

[24] M. Zhang, Z. Huang, and Y. Zhang, Restricted p-isometry properties of nonconvex matrix recovery, IEEE Trans. Inf. Theory, vol. 59, no. 7, pp. 4316-4323, Jul. 2013.

[25] Kong L, Xiu N. Exact low-rank matrix recovery via nonconvex Schatten p-minimization[J]. Asia-Pacific Journal of Operational Research, 2013, 30(03): 1340010.

[26] Wang H M, Li S. The bounds of restricted isometry constants for low rank matrices recovery[J]. Science China Mathematics, 2013, 56(6): 1117-1127.

[27] Chen W G, Li Y L. Stable recovery of low-rank matrix via nonconvex Schatten p-minimization[J]. Science China Mathematics, 2015, 58(12): 2643-2654.

[28] Gao Y, Han X, Ma M. Recovery of low-rank matrices based on the rank null space properties[J]. International Journal of Wavelets, Multiresolution and Information Processing, 2017, 15(04): 1750032.

[29] Wang W, Zhang F, Wang J. Low-rank matrix recovery via regularized nuclear norm minimization[J]. arXiv preprint arXiv:1903.01053, 2019.

[30] Lu C, Feng J, Lin Z, et al. Exact low tubal rank tensor recovery from Gaussian measurements[C]. Proceedings of the 27th International Joint Conference on Artificial Intelligence. AAAI Press, 2018: 2504-2510.

[31] Wen F, Liu P, Liu Y, et al. Robust Sparse Recovery in Impulsive Noise via ℓ_p-ℓ_1 Optimization[J]. IEEE Transactions on Signal Processing, 2017, 65(1): 105-118.

[32] Wang W, Wang J, Zhang Z. Block-sparse signal recovery via ℓ_2/ℓ_{1-2} minimisation method[J]. IET Signal Processing, 2017, 12(4): 422-430.

[33] Wang W, Wang J. Enhancing matrix completion using a modified second-order total variation[J]. Discrete Dynamics in Nature and Society, 2018, 2018.