Low-rank matrix recovery via regularized nuclear norm minimization

Wendong Wang[a], Feng Zhang[a], Jianjun Wang[a,*]

[a]*School of Mathematics and Statistics, Southwest University, Chongqing, 400715, China*

Abstract

In this paper, we theoretically investigate the low-rank matrix recovery problem in the context of the unconstrained regularized nuclear norm minimization (RNNM) framework. Our theoretical findings show that, one can robustly recover any matrix X from its few noisy measurements $\boldsymbol{b} = \mathcal{A}(X) + \boldsymbol{n}$ with a bounded constraint $\|\boldsymbol{n}\|_2 \leq \epsilon$ via the RNNM, if the linear map $\mathcal{A}$ satisfies restricted isometry property (RIP) with

$$\delta_{tk} < \sqrt{\frac{t-1}{t}}$$

for certain fixed $t > 1$. Recently, this condition with $t \geq 4/3$ has been proved by Cai and Zhang (2014) to be sharp for exactly recovering any rank-k matrices via the constrained nuclear norm minimization (NNM). To the best of our knowledge, our work first extends nontrivially this recovery condition for the constrained NNM to that for its unconstrained counterpart. Furthermore, it will be shown that similar recovery condition also holds for regularized ℓ_1-norm minimization, which sometimes is also called Basis Pursuit DeNoising (BPDN).

Keywords: Low-rank matrix recovery, regularized nuclear norm minimization, restricted isometry property, basis pursuit denoising

---

1. Introduction

Over the past decade, low-rank matrix recovery (LRMR) problem has attracted considerable interest of researchers in many fields, including computer vision [1], recommender systems [2], and machine learning [3] to name a few. Mathematically, this problem aims to recover an unknown low-rank matrix $X \in \mathbb{R}^{n_1 \times n_2}$ from

$$\boldsymbol{b} = \mathcal{A}(X) + \boldsymbol{n},$$

*Corresponding author

Email addresses: d.sylan@foxmail.com (Wendong Wang), zhangf@email.swu.edu.cn (Feng Zhang), wjj@swu.edu.cn (Jianjun Wang)

where $\boldsymbol{b} \in \mathbb{R}^m (m \ll n_1 n_2)$ is an observed vector, $\boldsymbol{n} \in \mathbb{R}^m$ is the unknown noise, and $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a known linear map defined as

$$\mathcal{A}(X) = [\mathrm{tr}(X^T A^{(1)}), \mathrm{tr}(X^T A^{(2)}), \cdots, \mathrm{tr}(X^T A^{(m)})]^T. \tag{1}$$

Here, $\mathrm{tr}(\cdot)$ is the trace function and $A^{(i)} \in \mathbb{R}^{n_1 \times n_2}$ is the ith measurement matrix.

A popular approach for the LRMR problem is to solve the nuclear norm minimization (NNM)

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \|X\|_*, \quad s.t. \quad \|\boldsymbol{b} - \mathcal{A}(X)\|_2 \le \epsilon, \tag{2}$$

So far, much work has been done to find the explicit conditions under which the exact/robust recovery of any low-rank matrices can be guaranteed [4, 5, 6, 7]. As one of the most powerful and widely used theoretical tools, restricted isometry property (RIP) captures particular attention.

Definition 1 ([5])**.** *A linear map $\mathcal{A}$ defined by* (1) *is said to satisfy the RIP with restricted isometry constant (RIC) of order k, denoted by δ_k*[1]*, if δ_k is the smallest value $\delta \in (0, 1)$ such*

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

for every rank-k matrix $X \in \mathbb{R}^{n_1 \times n_2}$, i.e., the signal whose rank is at most k.

There exist many RIP-based sufficient conditions for the exact recovery (i.e., the case when $\boldsymbol{n} = 0$ and $\epsilon = 0$) of any rank-k matrices through (2). These include $\delta_{4k} < \sqrt{2} - 1$ [5], $\delta_{4k} < 0.558$, and $\delta_{3k} < 0.4721$ [8], $\delta_{2k} < 0.4931$ [9], $\delta_{2k} < 1/2$ and $\delta_k < 1/3$ [10]. In particular, the sharpest conditions with the form of $\delta_{tk} < \delta^*$ for $t > 0$ have been completely given by Cai and Zhang [11] and Zhang and Li [12], where $\delta^* = \sqrt{(t-1)/t}$ for $t \ge 4/3$ and $\delta^* = t/(4-t)$ otherwise, and they have also proved that under these conditions, one can still robustly reconstruct any (low-rank) matrices.

An alternative approach to the constrained NNM (2) is to solve its unconstrained counterpart, i.e., the following Regularized NNM (RNNM):

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \|X\|_* + \frac{1}{2\lambda}\|\boldsymbol{b} - \mathcal{A}(X)\|_2^2. \tag{3}$$

Compared to the constrained problem (2), this unconstrained problem is much more suitable for noisy measurements and approximately low-rank matrix recovery [13]. Currently, almost all the researches are focus on the algorithms induced by (3), see, e.g., [13, 14, 15]. To the best of our knowledge, Candès and Plan [5] provided the first RIP-based performance guarantee for (3), and their results show that, when the noise $\boldsymbol{n}$ obeys $\|\mathcal{A}^*(\boldsymbol{n})\| \triangleq \|\sum_{i=1}^m \boldsymbol{n}_i \cdot A^{(i)}\| \le \lambda/2$,

[1] When k is not an integer, we define δ_k as $\delta_{\lceil k \rceil}$.

and the map $\mathcal{A}$ satisfies $\delta_{4k} < (3\sqrt{2}-1)/17$, the robust recovery of any rank-k matrices can be guaranteed through (3). However, after their initial work, the theoretical investigation of (3) is rarely reported. Note that their noise setting is based on the Dantzig selector rather than the often used ℓ_2-norm setting (i.e., $\|\boldsymbol{n}\|_2 \le \epsilon$), and the obtained sufficient condition still has room to improve.

In this paper, we theoretically investigate the RIP-based performance guarantee of the constrained problem (3) when the noise $\boldsymbol{n}$ obeys $\|\boldsymbol{n}\|_2 \le \epsilon$. We show that if $\mathcal{A}$ satisfies $\delta_{tk} < \sqrt{(t-1)/t}$ for certain $t > 1$, one can robustly recover any (low-rank) matrices from (3). The obtained results first extend the recovery condition recently obtained by Cai and Zhang [11] for the constrained problem (2) to that for its unconstrained counterpart. It should be also noted that similar condition also holds for the well-known Basis Pursuit DeNoising (BPDN) [16] to guarantee the robust recovery of any (sparse) signals.

The remainder of the paper is organized as follows. Section II introduces some notations and useful lemmas. Section III presents the main results. Section IV gives the related proofs. Finally, conclusion and future works are given in Section V.

2. Notations and Preliminaries

2.1. Notations

We assume w.l.o.g. that $n_1 \le n_2$ and the SVD of $X \in \mathbb{R}^{n_1 \times n_2}$ is $X = \sum_{i=1}^{n_1} \sigma_i(X) \cdot \boldsymbol{u}_X^{(i)} \cdot (\boldsymbol{v}_X^{(i)})^T$, where $\boldsymbol{u}_X^{(i)}$ and $\boldsymbol{v}_X^{(i)}$ are the left and right singular value vectors of X, respectively, and $\sigma_i(X)$ is the ith largest singular value of X. For any positive integer s, we denote $[s] = \{1, 2, \cdots, s\}$, and $E^c = [n_1] \backslash E$ for any $E \subset [n_1]$. We also denote $\sigma_E(X)$ as a vector whose element $(\sigma_E(X))_i = \sigma_i(X)$ for $i \in E$ and $(\sigma_E(X))_i = 0$ otherwise, and $X_E = \sum_{i \in E} \sigma_i(X) \cdot \boldsymbol{u}_X^{(i)} \cdot (\boldsymbol{v}_X^{(i)})^T$ and $X_{[s]} = \sum_{i=1}^{s} \sigma_i(X) \boldsymbol{u}_X^{(i)} (\boldsymbol{v}_X^{(i)})^T$. Besides, we denote $\|\cdot\|_\alpha^\beta = (\|\cdot\|_\alpha)^\beta$ where $\|\cdot\|_\alpha$ is certain (quasi-)norm. Then clearly $\|\sigma_E(X)\|_1 = \|X_E\|_*$. In the end, $\|\boldsymbol{x}\|_0$ is defined to be the number of the nonzero elements in $\boldsymbol{x}$.

2.2. Three key lemmas

Before presenting our main results, we need some lemmas.

Lemma 1 ([11])**.** *For a positive number α and a positive integer k, define the polytope $T(\alpha, k) \subset \mathbb{R}^n$ by $T(\alpha, k) = \{\boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{v}\|_\infty \le \alpha, \|\boldsymbol{v}\|_1 \le k\alpha\}$. For any $\boldsymbol{v} \in \mathbb{R}^n$, define the set $U(\alpha, k, \boldsymbol{v}) \subset \mathbb{R}^n$ by $U(\alpha, k, \boldsymbol{v}) = \{\boldsymbol{u} \in \mathbb{R}^n : supp(\boldsymbol{u}) \subseteq supp(\boldsymbol{v}), \|\boldsymbol{u}\|_0 \le k, \|\boldsymbol{u}\|_1 = \|\boldsymbol{v}\|_1, \|\boldsymbol{u}\|_\infty \le \alpha\}$. Then $\boldsymbol{v} \in T(\alpha, k)$ iff $\boldsymbol{v}$ is in the convex hull of $U(\alpha, k, \boldsymbol{v})$. In particular, any $\boldsymbol{v} \in T(\alpha, k)$ can be expressed as*

$$\boldsymbol{v} = \sum_{i=1}^{c} \gamma_i \boldsymbol{u}_i$$

where $\boldsymbol{u}_i \in U(\alpha, k, \boldsymbol{v})$ and $0 \le \gamma_i \le 1$, $\sum_{i=1}^{c} \gamma_i = 1$.

Lemma 2. *If the map $\mathcal{A}$ obeys the RIP of order $tk(t>1)$ with RIC $\delta_{tk}\in(0,1)$, then for any matrix $H\in\mathbb{R}^{n_1\times n_2}$ and $E\subset[n_1]$ with $|E|=k$, it holds that*

$$\|H_E\|_F \le \beta_1\|\mathcal{A}(H)\|_2+\beta_2\frac{\|H_{E^c}\|_*}{\sqrt{k}}, \tag{4}$$

where

$$\beta_1 \triangleq \frac{2}{(1-\delta_{tk})\sqrt{1+\delta_{tk}}},\ \text{ and } \beta_2 \triangleq \frac{\delta_{tk}}{\sqrt{(1-(\delta_{tk})^2)(t-1)}}.$$

Lemma 3. *Assume that $X^\sharp$ is the solution of* (3) *and $H=X^\sharp-X$. If the noisy measurements $\boldsymbol{b}=\mathcal{A}(X)+\boldsymbol{n}$ are observed with the noise level $\|\boldsymbol{n}\|_2\le\epsilon$, then for any subset $E\subset[n_1]$ with $|E|=k$, we have*

$$\|\mathcal{A}(H)\|_2^2-2\epsilon\|\mathcal{A}(H)\|_2 \le 2\lambda(\|H_E\|_*-\|H_{E^c}\|_*+2\|X_{E^c}\|_*) \tag{5}$$

and

$$\|H_{E^c}\|_* \le \|H_E\|_*+2\|X_{E^c}\|_*+\frac{\epsilon}{\lambda}\|\mathcal{A}(H)\|_2. \tag{6}$$

3. Main results

With previous preparations in mind, we now present our main results.

Theorem 4. *For any observed vector $\boldsymbol{b}=\mathcal{A}(X)+\boldsymbol{n}$ with a bounded constraint $\|\boldsymbol{n}\|_2\le\lambda/2$, if the map $\mathcal{A}$ satisfies RIP with*

$$\delta_{tk}<\sqrt{\frac{t-1}{t}} \tag{7}$$

for certain fixed $t>1$, then we have

$$\|\mathcal{A}(X^\sharp-X)\|_2 \le C_1\|X-X_{[k]}\|_*+C_2, \tag{8}$$

$$\|X^\sharp-X\|_F \le C_3\|X-X_{[k]}\|_*+C_4, \tag{9}$$

where $X^\sharp$ is the optimal solution of (3)*, and*

$$\begin{aligned}
C_1 &= \frac{2\lambda}{\sqrt{k}\beta_1\lambda+\epsilon},\quad C_2=2\sqrt{k}\beta_1\lambda+2\epsilon,\\
C_3 &= \frac{2\sqrt{k}\beta_1(2\sqrt{k}+1+\beta_2)\lambda+2(\sqrt{k}\beta_2+2\beta_2+\sqrt{k})\epsilon}{k\beta_1(1-\beta_2)\lambda},\\
C_4 &= \frac{2(k+\sqrt{k})\beta_1\lambda+(\beta_2+2\sqrt{k}-\sqrt{k}\beta_2)\epsilon}{\sqrt{k}(1-\beta_2)\lambda(\sqrt{k}\beta_1\lambda+\epsilon)^{-1}}.
\end{aligned}$$

Remark 1. The condition (7) has been obtained recently by Cai and Zhang in [11] for exact/robust signal recovery from (2), and it was proved to be sharp for the exact rank-k matrix recovery when $t > 4/3$. To the best of our knowledge, we first extend nontrivially this condition from the constrained problem (2) to its unconstrained counterpart. When compared to some existing results, e.g, [17], our upper bound estimate for $\|X^\sharp - X\|_F$ seems relatively loose. However it can be further improved by using the skills in [17].

Remark 2. BPDN is closely related to (3), and there are some recovery conditions for this BPDN, see, e.g., [17, 18, 19]. However, most of these conditions are unsatisfactory. In fact, by combing Lemma 2 (with setting D be an identity matrix) in [20] and also using the techniques in proof of our Theorem 4, one will obtain a new and much weaker recovery condition for the BPDN. Besides, our theoretical results can still be extended to deal with the noise under Dantzig Selector settings for both sparse signal and low-rank matrix recovery.

Remark 3. There are some special cases of Theorem 4 which can be used to cope with several different LRMR tasks. For examples, one can set $\boldsymbol{n} = 0$ and $\epsilon = 0$ for the noiseless recovery. In this case, the error will almost disappear if one chooses the parameter λ as small as possible, and this result is also coincident with the results obtained in [17, 20]; one can consider the rank-k matrix recovery in presence of noise; similar with [5, 17, 20], one can also associate ϵ with λ, and set $\epsilon = \lambda/2$.

4. Proofs

4.1. Proof of Lemma 2

Proof. The proof mainly follows from [20]. When tk is not an integer, let $t' = \lceil tk \rceil / k$, then $t' > t$ and $t'k$ is an integer. In view of this, we here only need to prove Lemma 2 when tk is a positive integer for a given $t > 1$. To do so, we first denote the SVD of H as

$$H = \sum_{i=1}^{n_1} \sigma_i(H) \cdot \boldsymbol{u}_H^{(i)} \cdot (\boldsymbol{v}_H^{(i)})^T.$$

We also denote $\alpha = \|H_{E^c}\|_* / ((t-1)k)$, and

$$E_1 = \{i \in E^c : \sigma_i(H) > \alpha\}, \quad E_2 = \{i \in E^c : \sigma_i(H) \le \alpha\}.$$

Then clearly $E_1 \cup E_2 = E^c$ and $E_1 \cap E_2 = \emptyset$. We will begin with proving

$$\|H_{E \cup E_1}\|_F \le \beta_1 \|\mathcal{A}(H)\|_2 + \frac{\beta_2}{\sqrt{k}} \|H_{E^c}\|_* \tag{10}$$

Before this, we will show that $s \triangleq |E_1| < (t-1)k$. In fact it holds naturally for $E_1 = \emptyset$. When $E_1 \neq \emptyset$, we know that

$$\|\sigma_{E_1}(H)\|_1 = \|H_{E_1}\|_* > s\alpha = s\frac{\|H_{E^c}\|_*}{(t-1)k} \ge \frac{s}{(t-1)k}\|H_{E_1}\|_* = \frac{s}{(t-1)k}\|\sigma_{E_1}(H)\|_1.$$

Thus a quick simplification of the above inequality yields the desired result.

On the other hand, in terms of $\sigma_{E_2}(H)$, we have

$$\|\sigma_{E_2}(H)\|_1 = \|H_{E_2}\|_* = \|H_{E^c}\|_* - \|H_{E_1}\|_* \leq (t-1)k\alpha - s\alpha = ((t-1)k - s)\alpha,$$

and $\|\sigma_{E_2}(H)\|_\infty = \max_{i \in E_2} \sigma_i(H) \leq \alpha$. Then using Lemma 1, we have

$$\sigma_{E_2}(H) = \sum_{i=1}^{l} \gamma_i \boldsymbol{z}^{(i)},$$

where l is a certain positive integer, $\boldsymbol{z}^{(i)} \in U(\alpha, (t-1)k - s, \sigma_{E_2}(H))$ and $0 \leq \gamma_i \leq 1$, $\sum_{i=1}^{l} \gamma_i = 1$. By further defining

$$\begin{aligned}
&\boldsymbol{b}^{(i)} = (1+\delta_{tk})\sigma_{E\cup E_1}(H) + \delta_{tk}\boldsymbol{z}^{(i)}, \quad \boldsymbol{d}^{(i)} = (1-\delta_{tk})\sigma_{E\cup E_1}(H) - \delta_{tk}\boldsymbol{z}^{(i)}, \\
&Z^{(i)} = \sum_{j=1}^{n_1} (\boldsymbol{z}^{(i)})_j \cdot \boldsymbol{u}_H^{(j)} \cdot (\boldsymbol{v}_H^{(j)})^T, \quad B^{(i)} = \sum_{j=1}^{n_1} (\boldsymbol{b}^{(i)})_j \cdot \boldsymbol{u}_H^{(j)} \cdot (\boldsymbol{v}_H^{(j)})^T, \\
&D^{(i)} = \sum_{j=1}^{n_1} (\boldsymbol{d}^{(i)})_j \cdot \boldsymbol{u}_H^{(j)} \cdot (\boldsymbol{v}_H^{(j)})^T,
\end{aligned}$$

we can easily induce that both $\boldsymbol{b}^{(i)}$ and $\boldsymbol{d}^{(i)}$ are all tk-sparse, and

$$H_{E_2} = \sum_{i=1}^{l} \gamma_i Z^{(i)}, \quad B^{(i)} = (1+\delta_{tk})H_{E\cup E_1} + \delta_{tk} Z^{(i)}, \quad D^{(i)} = (1-\delta_{tk})H_{E\cup E_1} - \delta_{tk} Z^{(i)}.$$

Now applying Definition 1, we will estimate the upper and lower bounds of

$$\rho \triangleq \sum_{i=1}^{l} \gamma_i \Big(\|\mathcal{A}(B^{(i)})\|_2^2 - \|\mathcal{A}(D^{(i)})\|_2^2 \Big).$$

As to the upper bound of ρ, we have

$$\begin{aligned}
\rho &= 4\delta_{tk} \langle \mathcal{A}(H_{E\cup E_1}), \mathcal{A}(H_{E\cup E_1} + \sum_{i=1}^{l} \gamma_i Z^{(i)}) \rangle \\
&= 4\delta_{tk} \langle \mathcal{A}(H_{E\cup E_1}), \mathcal{A}(H) \rangle \leq 4\delta_{tk} \|\mathcal{A}(H_{E\cup E_1})\|_2 \|\mathcal{A}(H)\|_2 \\
&\leq 4\delta_{tk} \sqrt{1+\delta_{tk}} \|H_{E\cup E_1}\|_F \|\mathcal{A}(H)\|_2. \qquad (11)
\end{aligned}$$

As to the lower bound of ρ, we have

$$\begin{aligned}
\rho &\geq \sum_{i=1}^{l} \gamma_i \left((1-\delta_{tk}) \|\boldsymbol{b}^{(i)}\|_2^2 - (1+\delta_{tk}) \|\boldsymbol{d}^{(i)}\|_2^2 \right) \\
&= 2\delta_{tk}(1-(\delta_{tk})^2) \|\sigma_{E\cup E_1}(H)\|_2^2 - 2(\delta_{tk})^3 \sum_{i=1}^{l} \gamma_i \|\boldsymbol{z}_i\|_2^2 \\
&\geq 2\delta_{tk}(1-(\delta_{tk})^2) \|H_{E\cup E_1}\|_F^2 - \frac{2(\delta_{tk})^3}{(t-1)k} \|H_{E^c}\|_*^2, \qquad (12)
\end{aligned}$$

where we used $\langle \sigma_{E\cup E_1}(H), \boldsymbol{z}^{(i)}\rangle = 0$ for the equation,and

$$\|\boldsymbol{z}^{(i)}\|_2^2 \le \|\boldsymbol{z}^{(i)}\|_0 \|\boldsymbol{z}^{(i)}\|_\infty^2 \le ((t-1)k - s)\alpha^2 = \frac{\|H_{E^c}\|_*^2}{(t-1)k}$$

for the last inequality. Combing (11) and (12) yields

$$(1-(\delta_{tk})^2)\|H_{E\cup E_1}\|_F^2 - 2\sqrt{1+\delta_{tk}}\|\mathcal{A}(H)\|_2\|H_{E\cup E_1}\|_F - \frac{(\delta_{tk})^2}{(t-1)k}\|H_{E^c}\|_*^2 \le 0.$$

Therefore,

$$\begin{aligned}\|H_{E\cup E_1}\|_F \le & \frac{2\sqrt{1+\delta_{tk}}\|\mathcal{A}(H)\|_2}{2(1-(\delta_{tk})^2)} + \frac{\sqrt{(2\sqrt{1+\delta_{tk}}\|\mathcal{A}(H)\|_2)^2 + 4(1-(\delta_{tk})^2)\frac{(\delta_{tk})^2}{(t-1)k}\|H_{E^c}\|_*^2}}{2(1-(\delta_{tk})^2)} \\ \le & \frac{2(1-\delta_{tk})^{-1}}{\sqrt{1+\delta_{tk}}}\|\mathcal{A}(H)\|_2 + \frac{\delta_{tk}}{\sqrt{(1-(\delta_{tk})^2)(t-1)}}\frac{\|H_{E^c}\|_*}{\sqrt{k}},\end{aligned}$$

where we used $\sqrt{x^2+y^2} \le |x| + |y|$ for the last inequality. Then combing (10) and $\|H_E\|_F \le \|H_{E\cup E_1}\|_F$ directly leads to (4), which completes the proof.

4.2. Proof of Lemma 3

Proof. Since $X^\sharp$ is the optimal solution of (3), we have

$$\|X^\sharp\|_* + \frac{1}{2\lambda}\|\boldsymbol{b} - \mathcal{A}(X^\sharp)\|_2^2 \le \|X\|_* + \frac{1}{2\lambda}\|\boldsymbol{b} - \mathcal{A}(X)\|_2^2,$$

which is equivalent to

$$\|\mathcal{A}(H)\|_2^2 - 2\langle \boldsymbol{n}, \mathcal{A}(H)\rangle \le 2\lambda(\|X\|_* - \|X^\sharp\|_*). \tag{13}$$

As to the left-hand side of (13), we have

$$\|\mathcal{A}(H)\|_2^2 - 2\langle \boldsymbol{n}, \mathcal{A}(H)\rangle \ge \|\mathcal{A}(H)\|_2^2 - 2\epsilon\|\mathcal{A}(H)\|_2. \tag{14}$$

As to the right-hand side of (13), we know

$$\begin{aligned}\|X^\sharp\|_* - \|X\|_* =& \sum_{i=1}^{n_1}\sigma_i(X+H) - (\|X_E\|_* + \|X_{E^c}\|_*) \\ \ge& \sum_{i=1}^{n_1}|\sigma_i(X) - \sigma_i(-H)| - (\|X_E\|_* + \|X_{E^c}\|_*) \\ \ge& \sum_{i\in E}(\sigma_i(X) - \sigma_i(H)) + \sum_{i\in E^c}(\sigma_i(H) - \sigma_i(X)) - (\|X_E\|_* + \|X_{E^c}\|_*) \\ =& -\|H_E\|_* + \|H_{E^c}\|_* - 2\|X_{E^c}\|_*, \end{aligned} \tag{15}$$

where we used Theorem 1 in [21] for the first inequality. Then combing (13), (14), and (15) leads to the desired result (5), and (6) follows trivially from (5).

85 *4.3. Proof of Theorem 4*

Proof. We start with Denoting $E=[k]$ and $H=X^\sharp-X$. Then by Lemma 2 and Lemma 3, we have

$$\begin{aligned}\|\mathcal{A}(H)\|_2^2-2\epsilon\|\mathcal{A}(H)\|_2 \leq& 2\lambda(\sqrt{k}\|H_E\|_F-\|H_{E^c}\|_*+2\|X_{E^c}\|_*)\\ \leq& 2\sqrt{k}\lambda(\beta_1\|\mathcal{A}(H)\|_2+\frac{\beta_2}{\sqrt{k}}\|H_{E^c}\|_*)-2\lambda\|H_{E^c}\|_*+4\lambda\|X_{E^c}\|_*\\ =& 2\sqrt{k}\beta_1\lambda\|\mathcal{A}(H)\|_2-2(1-\beta_2)\lambda\|H_{E^c}\|_*+4\lambda\|X_{E^c}\|_*\end{aligned}\tag{16}$$

According to the condition (7), we know

$$1-\beta_2=1-\frac{\delta_{tk}}{\sqrt{(1-(\delta_{tk})^2)(t-1)}}>1-\frac{\sqrt{(t-1)/t}}{\sqrt{(1-(t-1)/t)(t-1)}}=0.$$

Therefore we can further know from (16) that

$$\|\mathcal{A}(H)\|_2^2-2(\sqrt{k}\beta_1\lambda+\epsilon)\|\mathcal{A}(H)\|_2-4\lambda\|X_{E^c}\|_*\leq 0,$$

which implies that

$$\begin{aligned}\|\mathcal{A}(H)\|_2 \leq& (\sqrt{k}\beta_1\lambda+\epsilon)+\sqrt{(\sqrt{k}\beta_1\lambda+\epsilon)^2+4\lambda\|X_{E^c}\|_*}\\ \leq& (\sqrt{k}\beta_1\lambda+\epsilon)+(\sqrt{k}\beta_1\lambda+\epsilon)+\frac{2\lambda\|X_{E^c}\|_*}{(\sqrt{k}\beta_1\lambda+\epsilon)}\\ \leq& \frac{2\lambda}{\sqrt{k}\beta_1\lambda+\epsilon}\|X_{E^c}\|_*+2\sqrt{k}\beta_1\lambda+2\epsilon.\end{aligned}$$

This completes (8). Based on (6) and (8), we now give a new upper bound estimate for $\|H_{E^c}\|_*$, i.e.,

$$\|H_{E^c}\|_*\leq\sqrt{k}\|H_E\|_F+\frac{2(\sqrt{k}\beta_1\lambda+2\epsilon)}{\sqrt{k}\beta_1\lambda+\epsilon}\|X_{E^c}\|_*+\frac{2\epsilon}{\lambda}(\sqrt{k}\beta_1\lambda+\epsilon),\tag{17}$$

86 where we used $\|H_E\|_*\leq\sqrt{k}\|H_E\|_F$.

On the other hand, using (4), (8), and (17), we can also give a new upper bound estimate for $\|H_E\|_F$, i.e.,

$$\|H_E\|_F\leq\beta_2\|H_E\|_F+\frac{2\sqrt{k}\beta_1(1+\beta_2)\lambda+4\beta_2\epsilon}{k\beta_1\lambda+\sqrt{k}\epsilon}\|X_{E^c}\|_*+2(\beta_1+\frac{\beta_2\epsilon}{\sqrt{k}\lambda})(\sqrt{k}\beta_1\lambda+\epsilon),$$

which is equivalent to

$$\|H_E\|_F\leq\frac{2\sqrt{k}\beta_1(1+\beta_2)\lambda+4\beta_2\epsilon}{(1-\beta_2)(k\beta_1\lambda+\sqrt{k}\epsilon)}\|X_{E^c}\|_*+\frac{2(\sqrt{k}\beta_1\lambda+\beta_2\epsilon)(\sqrt{k}\beta_1\lambda+\epsilon)}{\sqrt{k}(1-\beta_2)\lambda}.\tag{18}$$

Combining (17), (18), and $\|H_{E^c}\|_F \leq \|H_{E^c}\|_*$, we have

$$\begin{aligned}\|H\|_F \leq & \|H_E\|_F + \|H_{E^c}\|_F \\ \leq & (\sqrt{k}+1)\|H_E\|_F + \frac{2(\sqrt{k}\beta_1\lambda + 2\epsilon)}{\sqrt{k}\beta_1\lambda + \epsilon}\|X_{E^c}\|_* + \frac{2\epsilon}{\lambda}(\sqrt{k}\beta_1\lambda + \epsilon) \\ \leq & C_3\|X_{E^c}\|_* + C_4,\end{aligned}$$

where C_3 and C_4 are defined in Theorem 4. This completes the proof.

5. Conclusion and future works

The goal of this work was to provide a theoretical investigation for the LRMR problem in the context of the unconstrained RNNM framework. In particular, using the powerful RIP tool, we have established a series of sufficient conditions (related to the δ_{tk}) of this RNNM model for recovery of any (low-rank) matrices with the ℓ_2-norm bounded noise. One of our future works will focus on deriving the new recovery conditions on the δ_{tk} for $0 < t \leq 1$. Besides, extending the current theoretical results to more unconstrained convex/nonconvex models for vector/matrix/tensor recovery will be another future work.

References

[1] E. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J. ACM 58 (3) (2011) 1–37.

[2] R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices, J. Mach. Learn. Res. 11 (2010) 2287–2322.

[3] A. Argyriou, T. Evgeniou, M. Pontil, Convex multitask feature learning, Mach. Learn. 73 (3) (2008) 243–272.

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

[5] E. Candès, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, IEEE Trans. Inf. Theory 57 (4) (2011) 2342–2359.

[6] M.-J. Lai, W. Yin, Augmented ℓ_1 and nuclear-norm models with a globally linearly convergent algorithm, SIAM J. Imaging Sci. 6 (2) (2013) 1059–1091.

[7] R. Zhang, S. Li, Optimal rip bounds for sparse signals recovery via ℓ_p minimization, Appl. Comput. Harmon. Anal.doi:10.1016/j.acha.2017.10.004.

[8] K. Mohan, M. Fazel, New restricted isometry results for noisy low-rank recovery, in: Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, IEEE, 2010, pp. 1573–1577.

[9] H. Wang, S. Li, Augmented ℓ_1 and nuclear-norm models with a globally linearly convergent algorithm, Sci. China, Math. 56 (6) (2012) 1117–1127.

[10] T. Cai, A. Zhang, Sharp rip bound for sparse signal and low-rank matrix recovery, Appl. Comput. Harmon. Anal. 35 (2013) 74–93.

[11] T. Cai, A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, IEEE Trans. Inf. Theory 60 (1) (2014) 122–132.

[12] R. Zhang, S. Li, A proof of conjecture on restricted isometry property constants $\delta_{tk}(0 < t < \frac{4}{3})$, IEEE Trans. Inf. Theory 64 (3) (2018) 1699–1705.

[13] M.-J. Lai, Y. Y. Xu, W. T. Yin, Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, SIAM J. Numer. Anal. 51 (2013) 927–957.

[14] K. Toh, S. Yun, An accelerated proximal gradient algorithms for nuclear norm regularized least squares problems, Pacific J. Optimization 6 (2010) 615–640.

[15] D. Goldfarb, S. Ma, Convergence of fixed-point continuation algorithms for matrix rank minimization, Found. Comput. Math. 11 (2) (2011) 183–210.

[16] S. Chen, D. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 43 (1) (2001) 129–159.

[17] Y. Shen, B. Han, E. Braverman, Stable recovery of analysis based approaches, Appl. Comput. Harmon. Anal. 39 (1) (2015) 161–172.

[18] C. Zhu, Stable recovery of sparse signals via regularized minimization, IEEE Trans. Inf. Theory 54 (7) (2008) 3364–3367.

[19] Z. Ben-Haim, Y. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise, IEEE Trans. Signal Process. 58 (10) (2010) 5030–5043.

[20] H. Ge, J. Wen, W. Chen, J. Weng, M.-J. Lai, Stable sparse recovery with three unconstrained analysis based approaches, http://alpha.math.uga.edu/ mjlai/papers/20180126.pdf (2018).

[21] M. Yue, A. So, A perturbation inequality for concave functions of singular values and its applications in low-rank matrix recovery, Appl. Comput. Harmon. Anal. 40 (2) (2016) 396–416.