Sharp Restricted Isometry Property Bounds for Low-rank Matrix Recovery Problems with Corrupted Measurements

1 Ziye Ma, 2 Yingjie Bi*, 2 Javad Lavaei, 1,3 Somayeh Sojoudi
1 Department of Electrical Engineering and Computer Science
2 Department of Industrial Engineering and Operations Research
3 Department of Mechanical Engineering
University of California, Berkeley
ziyema, yingjiebi, lavaei, sojoudi}@berkeley.edu

Abstract

In this paper, we study a general low-rank matrix recovery problem with linear measurements corrupted by some noise. The objective is to understand under what conditions on the restricted isometry property (RIP) of the problem local search methods can find the ground truth with a small error. By analyzing the landscape of the non-convex problem, we first propose a global guarantee on the maximum distance between an arbitrary local minimizer and the ground truth under the assumption that the RIP constant is smaller than 1/2. We show that this distance shrinks to zero as the intensity of the noise reduces. Our new guarantee is sharp in terms of the RIP constant and is much stronger than the existing results. We then present a local guarantee for problems with an arbitrary RIP constant, which states that any local minimizer is either considerably close to the ground truth or far away from it. The developed results demonstrate how the noise intensity and the RIP constant of the problem affect the locations of the local minima relative to the true solution.

1 Introduction

Low-rank matrix recovery problems arise in various applications, such as matrix completion [1][2], phase synchronization/retrieval [3][5], robust PCA [6], and several others [7][8]. In this paper, we study a class of low-rank matrix recovery problems, where the goal is to recover a symmetric and positive semidefinite ground truth matrix \( \mathbf{M}^* \) with \( \text{rank}(\mathbf{M}^*) = r \) from certain linear measurements corrupted by noise. This problem can be formulated as the following optimization problem:

\[
\begin{align*}
\min_{\mathbf{M} \in \mathbb{R}^{n \times n}} & \quad \frac{1}{2} \| \mathcal{A}(\mathbf{M}) - b + w \|^2 \\
\text{s.t.} & \quad \text{rank}(\mathbf{M}) \leq r, \quad \mathbf{M} \succeq 0.
\end{align*}
\]

Here, \( \mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m \) is a linear operator whose action on a matrix \( \mathbf{M} \) is given by

\( \mathcal{A}(\mathbf{M}) = [(\mathbf{A}_1, \mathbf{M}), \ldots, (\mathbf{A}_m, \mathbf{M})]^T, \)

where \( \mathbf{A}_1, \ldots, \mathbf{A}_m \in \mathbb{R}^{n \times n} \) are called sensing matrices. In addition, \( b = \mathcal{A}(\mathbf{M}^*) \) represents the perfect measurement on the ground truth \( \mathbf{M}^* \) and \( w \) comes from an arbitrary probability distribution.

Although it is possible to solve the problem [1] based on convex relaxations [1][2][9], the computational complexity associated with solving a semidefinite program presents a major challenge for large-scale...
As compared above, there is an evident gap between the state-of-the-art results for the noiseless and
(Theorem 3.1) Suppose that
\[ M \]

Theorem 1.

A

The Kronecker product between

Theorem 31 in [6] further improves the above result by replacing the

The special noiseless case of the problem (2) can be obtained by setting

The unconstrained problem (2) is often solved by local search methods such as gradient descent.

\[ X \]

\[ M \]

\[ A \]

\[ \sigma \]

In this paper,

1.2 Notations

\[ \delta \]

\[ \lambda \]

\[ I \]

1.1 Related works

The general noisy problem, the relation

For the general noisy problem, the relation

Definition 1. The linear operator

\[ A(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^m \] is said to satisfy the \( \delta \)-RIP\(_{2r} \) property for some constant \( \delta \in [0, 1) \) if the inequality

\[
(1 - \delta)\| M \|_F^2 \leq \| A(M) \|_2^2 \leq (1 + \delta)\| M \|_F^2,
\]

holds for all \( M \in \mathbb{R}^{n \times n} \) with \( \text{rank}(M) \leq 2r \).

In the recent paper [20], the author developed a sharp bound on the absence of spurious local minima for the noiseless case of problem (2), which says that the problem has no spurious local minima if the measurement operator \( A \) satisfies the \( \delta \)-RIP\(_{2r} \) property with \( \delta < 1/2 \). This result is tight since there is a known counterexample [23] having spurious local minima under \( \delta = 1/2 \).

For the general noisy problem, the relation \( X^* X^{*T} = M^* \) is unlikely to be satisfied, where \( X^* \)

denotes a global minimizer of problem (2). However, in this situation, \( X^* X^{*T} \) should be close to the ground truth \( M^* \) if the noise \( w \) is small. As a generalization of the above-mentioned results for the noiseless problem, it is natural to study whether all local minimizers, including the global minimizers, are close to the ground truth \( M^* \) under the RIP assumption. One such result is presented in [11] and given below.

Theorem 1. ([17], Theorem 3.1) Suppose that \( w \sim \mathcal{N}(0, \sigma^2_w I_n) \) and \( A(\cdot) \) has the \( \delta \)-RIP\(_{4r} \) property with \( \delta < 1/10 \). Then, with probability at least \( 1 - 10/\sqrt{n^2} \), any local minimizer \( \hat{X} \) of problem (2) satisfies the inequality

\[
\| \hat{X} \hat{X}^T - M^* \|_F \leq 20 \sqrt{\frac{\log(n)}{m}} \sigma_w.
\]

Theorem 31 in [6] further improves the above result by replacing the \( \delta \)-RIP\(_{4r} \) property with the \( \delta \)-RIP\(_{2r} \) property. [24] studies a similar noisy low-rank matrix recovery problem with \( l_1 \) norm.

As compared above, there is an evident gap between the state-of-the-art results for the noiseless and noisy problems. The result for the noiseless problem only requires the RIP constant \( \delta < 1/2 \), but the one for the noisy problem requires \( \delta < 1/10 \) no matter how small the noise is. This gap will be addressed in this paper by showing that a major generalization of Theorem 1 holds for the noisy problem under the same RIP assumption as in the sharp bound for the noiseless problem.

1.2 Notations

In this paper, \( I_n \) refers to the identity matrix of size \( n \times n \). The notation \( M \succeq 0 \) means that \( M \) is a symmetric and positive semidefinite matrix. \( \sigma_i(M) \) denotes the \( i \)-th largest singular value of a matrix \( M \), and \( \lambda_i(M) \) denotes the \( i \)-th largest eigenvalue of \( M \). \( \| v \| \) denotes the Euclidean norm of a vector \( v \), while \( \| M \|_F \) and \( \| M \|_2 \) denote the Frobenius norm and induced \( l_2 \) norm of a matrix \( M \), respectively. \( \langle A, B \rangle \) is defined to be \( \text{tr}(A^T B) \) for two matrices \( A \) and \( B \) of the same size. The Kronecker product between \( A \) and \( B \) is denoted as \( A \otimes B \). For a matrix \( M \), \( \text{vec}(M) \) is the
We first present the global guarantee on the local minimizers of the problem (2). To simplify the explanation, consider arbitrary constants \( \epsilon > 0 \) and assume that the linear operator \( A \) satisfies the \( \delta \)-RIP property. Theorem 2 states that for every arbitrary local minimizer \( \hat{X} \) of problem (2), there exist upper bounds on the distance to the ground truth \( X^* \) as shown in the previous literature. Moreover, when the RIP constant \( \delta \) further decreases from \( 1/2 \), the upper bound on \( \| \hat{X} X^T - M^* \|_F \) will also decrease, which means that a local minimizer found by local search methods will be closer to the ground truth \( M^* \). This suggests that the RIP condition is able to not only guarantee the absence of spurious local minima as shown in the previous literature but also mitigate the influence of the noise in the measurements.

In the special case of rank \( r = 1 \), the conditions (3a) and (3b) in Theorem 2 can be substituted with a simpler condition as shown below.

**Theorem 3.** Consider the case \( r = 1 \) and assume that the linear operator \( A \) satisfies the \( \delta \)-RIP property with \( \delta < 1/2 \). For every \( \epsilon > 0 \), with probability at least \( \mathbb{P}(\| A^T w \| \leq \epsilon) \), every arbitrary local minimizer \( \hat{X} \) satisfies

\[
\| \hat{X} X^T - M^* \|_F \leq \frac{3(1 + \sqrt{2})\epsilon(1 + \delta)}{1 - 2\delta}.
\]

In the case when the RIP constant \( \delta \) is not less than \( 1/2 \), it is not possible to achieve a global guarantee similar to Theorem 2 or Theorem 3 since it is known that the problem may have a spurious solution even in the noiseless case. Instead, we turn to local guarantees by showing that every arbitrary local minimizer \( \hat{X} \) of problem (2) is either close to the ground truth \( M^* \) or far away from it in terms of the distance \( \| \hat{X} X^T - M^* \|_F \).

**Theorem 4.** Assume that the linear operator \( A \) satisfies the \( \delta \)-RIP property for some \( \delta \in [0, 1) \). Consider arbitrary constants \( \epsilon > 0 \) and \( \tau \in (0, 2(\sqrt{2} - 1)) \) such that

\[
\delta < \sqrt{1 - \frac{3 + 2\sqrt{2}}{4}} \tau^2.
\]

Every arbitrary local minimizer \( \hat{X} \) of problem (2) satisfying

\[
\| \hat{X} X^T - M^* \|_F \leq \tau \lambda_r(M^*)
\]
will also satisfy
\[
\| \hat{X} \hat{X}^T - M^* \|_F \leq \epsilon (1 + \delta) C(\tau, M^*) \left( \sqrt{1 - \frac{3 + 2\sqrt{2}}{4} \tau^2} - \delta \right)^{-1}
\] (6)
with probability at least \(\mathbb{P}(\|A^T w\| \leq \epsilon)\), where
\[
C(\tau, M^*) = \sqrt{\frac{2(\lambda_1(M^*) + \tau \lambda_r(M^*))}{(1 - \tau) \lambda_r(M^*)}}.
\]
The upper bounds in (5) and (6) define an outer ball and an inner ball centered at the ground truth \(M^*\).
Theorem 4 states that there is no local minimizer in the ring between the two balls. As \(\epsilon\) approaches zero, the inner ball shrinks to the ground truth. This theorem shows that bad local minimizers are located outside the outer ball. Note that the problem could be highly non-convex when \(\delta\) is close to 1, while this theorem shows a benign landscape in a local neighborhood of the solution. Furthermore, all the theorems in this section are applicable to arbitrary noise models since they make no explicit use of the probability distribution of the noise. The only required information is the probability \(\mathbb{P}(\|A^T w\| \leq \epsilon)\), which can be computed or bounded when the probability distribution of the noise is given as illustrated in Section 4.

3 Proofs of main results

Before presenting the proofs, we first compute the gradient and the Hessian of the objective function \(f(\hat{X})\) of the problem (2):
\[
\nabla f(\hat{X}) = \hat{X}^T A^T (Ae + w), \\
\nabla^2 f(\hat{X}) = 2I_r \otimes \text{mat}_{S}(A^T (Ae + w)) + \hat{X}^T A^T A \hat{X},
\]
where
\[
e = \text{vec}(\hat{X} \hat{X}^T - M^*),
\]
and \(\hat{X} \in \mathbb{R}^{n \times nr}\) is the matrix satisfying
\[
\hat{X} \text{vec}(U) = \text{vec}(\hat{X} U^T + U \hat{X}^T), \quad \forall U \in \mathbb{R}^{n \times r}.
\]
The first step in the proofs is to derive necessary conditions for a matrix \(\hat{X} \in \mathbb{R}^{n \times nr}\) to be a local minimizer, which depend on the linear operator \(A\), the noise \(w \in \mathbb{R}^n\) and the solution \(\hat{X}\).

Lemma 1. Assume that \(\hat{X} \in \mathbb{R}^{n \times nr}\) is a local minimizer of the problem (2). Then, it must satisfy the following inequalities:
\[
\| \hat{X}^T He\| \leq 2\|\hat{X}\|_2 A^T w, \quad (7a) \\
2I_r \otimes \text{mat}_{S}(He) + \hat{X}^T H \hat{X} \succeq -2\|A^T w\|_F I_{nr}, \quad (7b)
\]
where \(H = A^T A\).

Proof. To obtain condition (7a), notice that \(\nabla f(\hat{X}) = 0\) implies that
\[
\| \hat{X}^T He\| = \| \hat{X}^T A^T w\| \leq \|\hat{X}\|_2 A^T w \leq 2\|\hat{X}\|_2 A^T w,
\]
in which the last inequality is due to
\[
\| \hat{X} \text{vec}(U) \| = \| \hat{X} U^T + U \hat{X}^T \|_F \leq 2\|\hat{X}\|_2 |U|_F, \quad \forall U \in \mathbb{R}^{n \times r}.
\]
Similarly, \(\nabla^2 f(\hat{X}) \geq 0\) implies that
\[
2I_r \otimes \text{mat}_{S}(He) + \hat{X}^T H \hat{X} \succeq -2I_r \otimes \text{mat}_{S}(A^T w).
\]
On the other hand, the eigenvalues of \(I_r \otimes \text{mat}_{S}(A^T w)\) are the same those of \(\text{mat}_{S}(A^T w)\), and each eigenvalue of \(\text{mat}_{S}(A^T w)\) further satisfies
\[
|\lambda_i(\text{mat}_{S}(A^T w))| \leq \|\text{mat}_{S}(A^T w)\|_F \leq \frac{1}{2} \|\text{mat}(A^T w)\|_F + \frac{1}{2} \|\text{mat}(A^T w)^T\|_F = \|A^T w\|_F,
\]
which proves condition (7b). \(\square\)
The proof of Theorem 2 consists of studying two cases. The following lemma deals with the first case in which \( \hat{X} \) is a local minimizer with \( \sigma_r(\hat{X}) \) being close to zero.

**Lemma 2.** Assume that \( \hat{X} \in \mathbb{R}^{n \times r} \) is a local minimizer of the problem (2). Given an arbitrary constant \( \epsilon > 0 \), the inequalities

\[
\sigma_r(\hat{X}) \leq \sqrt{\frac{\epsilon}{1 + \delta}}
\]

and \( \|A^T w\| \leq \epsilon \) will together imply the inequality (3a).

**Proof.** Let \( G = \text{mat}_S(He) \) and \( u \in \mathbb{R}^n \) be a unit eigenvector of \( G \) corresponding to its minimum eigenvalue, i.e.,

\[
\|u\| = 1, \quad Gu = \lambda_{\min}(G)u.
\]

In addition, let \( v \in \mathbb{R}^r \) be a singular vector of \( \hat{X} \) such that

\[
\|v\| = 1, \quad \|Xv\| = \sigma_r(\hat{X}).
\]

Let \( U = \text{vec}(uv^T) \). Then, \( \|U\| \leq 1 \). Since \( \hat{X} \) is a local minimizer, (7b) implies that

\[
-2\epsilon \leq 2U^T(I_r \otimes \text{mat}_S(He))U + U^T \hat{X}^THXU
\]

\[
\leq 2 \text{tr}(uv^T Gu^T) + (1 + \delta)\|\hat{X}vu^T + uv^T \hat{X}^T\|_F^2
\]

\[
\leq 2\lambda_{\min}(G) + 4(1 + \delta)\sigma_r(\hat{X})^2
\]

\[
\leq 2\lambda_{\min}(G) + 2\epsilon.
\]

(8)

On the other hand,

\[
(1 - \delta)\|\hat{X}^T - M^*\|^2_\mathcal{F} \leq \epsilon^THe = \text{vec}(\hat{X}^T)\text{He} - \text{vec}(M^*)^THe
\]

\[
= \frac{1}{2} \text{vec}(\hat{X})^T \text{He} - \langle M^*, \text{mat}_S(He) \rangle
\]

\[
\leq \frac{1}{2} \|\hat{X}\|_2^2 \|\hat{X}^T\| + 2\epsilon \text{tr}(M^*)
\]

\[
\leq \epsilon\|\hat{X}\|_2^2 + 2\epsilon \text{tr}(M^*),
\]

in which the second last inequality is due to (8) and the last inequality is due to (7b). Furthermore, the right-hand side of the above inequality can be relaxed as

\[
\epsilon\|\hat{X}\|_2^2 + 2\epsilon \text{tr}(M^*) \leq \epsilon\|\hat{X}^T\|_F^2 + 2\epsilon\sqrt{\mathcal{F}}\|M^*\|_\mathcal{F} \leq \epsilon\|\hat{X}^T - M^*\|_F + (1 + 2\sqrt{\mathcal{F}})\epsilon\|M^*\|_F,
\]

which leads to the inequality (3a). \( \square \)

The remaining case with

\[
\sigma_r(\hat{X}) > \sqrt{\frac{\epsilon}{1 + \delta}}
\]

will be handled in the following lemma using a different method.

**Lemma 3.** Assume that the linear operator \( \mathcal{A} \) satisfies the \( \delta \)-RIP property with \( \delta < 1/2 \) and \( \hat{X} \in \mathbb{R}^{n \times r} \) is a local minimizer of the problem (2). Given an arbitrary constant \( \epsilon > 0 \), the inequalities

\[
\sigma_r(\hat{X}) > \sqrt{\frac{\epsilon}{1 + \delta}}
\]

and \( \|A^T w\| \leq \epsilon \) will together imply the inequality (3b).

The proofs of both Lemma 3 and the local guarantee in Theorem 4 generalize the proof of the absence of spurious local minima for the noiseless problem in [16, 20]. For a fixed solution \( \hat{X} \) and noise \( w \), one can find an operator \( \mathcal{A} \) satisfying the \( \delta \)-RIP property with the smallest possible \( \delta \) such that \( \hat{X} \) and \( \mathcal{A} \) satisfy the necessary conditions stated in Lemma 1. Let \( \delta^*(\hat{X}) \) be the RIP constant of the found measurement operator \( \mathcal{A} \) in the worst-case scenario. Then, if \( \hat{X} \) is a local minimizer of the current problem with the linear operator \( \mathcal{A} \) satisfying the \( \delta \)-RIP property, it holds that \( \delta \geq \delta^*(\hat{X}) \), which can further lead to an upper bound on the distance \( \|\hat{X}^T - M^*\|_F \).
To compute $\delta^*(\hat{X})$ defined above, we define $q = A^Tw$ and solve the following optimization problem whose optimal value is $\delta^*(\hat{X})$:

$$
\min_{\delta, H} \delta \\
\text{s.t.} \quad \|\hat{X}^T H^e\| \leq 2\|\hat{X}\|_2\|q\|, \\
               2I_r \odot \text{mat}_S(H^e) + \hat{X}^T \hat{H} \hat{X} \succeq -2\|q\|I_{nr}, \\
H \text{ is symmetric and satisfies the } \delta\text{-RIP}_{2r} \text{ property.}
$$  \hfill (9)

Note that a matrix $\hat{H} \in \mathbb{R}^{n^2 \times n^2}$ is said to satisfy the $\delta$-RIP$_{2r}$ property if

$$(1 - \delta)\|U\|^2 \leq U^T \hat{H} U \leq (1 + \delta)\|U\|^2$$

holds for every matrix $U \in \mathbb{R}^{n \times n}$ with rank($U$) $\leq 2r$ and $U = \text{vec}(U)$. Obviously, for a linear operator $A$, $\hat{H} = \hat{A}^T \hat{A}$ satisfies the $\delta$-RIP$_{2r}$ property if and only if $\hat{A}$ satisfies the $\delta$-RIP$_{2r}$ property.

However, since problem (9) is non-convex due to the RIP constraint, we instead solve the following convex reformulation:

$$
\min_{\delta, H} \delta \\
\text{s.t.} \quad \|\hat{X}^T H^e\| \leq 2\|\hat{X}\|_2\|q\|, \\
               2I_r \odot \text{mat}_S(H^e) + \hat{X}^T \hat{H} \hat{X} \succeq -2\|q\|I_{nr}, \\
               (1 - \delta)I_{nr} \preceq \hat{H} \preceq (1 + \delta)I_{nr}.
$$  \hfill (10)

Lemma 14 in [12] proves that problem (9) and problem (10) have the same optimal value. The remaining step in the proof of Lemma 3 is to solve the optimization problem (10) for given $\hat{X}$ and $q$. The complete proof of Lemma 3 is lengthy and deferred to Appendix A. Finally, Theorem 2 is a direct consequence of Lemma 2 and Lemma 3. The proof of Theorem 3 is very similar to that of Lemma 3 and is also given in Appendix A.

Now, we turn to the proof of the local guarantee in Theorem 4.

**Proof of Theorem 4.** First, we relax the optimization problem (10) by dropping the constraint related to the second-order necessary optimality condition. This gives rise to the optimization problem

$$
\min_{\delta, H} \delta \\
\text{s.t.} \quad \|\hat{X}^T H^e\| \leq 2\|\hat{X}\|_2\|q\|, \\
               (1 - \delta)I_{nr} \preceq \hat{H} \preceq (1 + \delta)I_{nr}.
$$  \hfill (11)

To further simplify the problem (11), one can replace its decision variable $\delta$ with $\eta$ and introduce the following optimization problem:

$$
\max_{\eta, H} \eta \\
\text{s.t.} \quad \|\hat{X}^T H^e\| \leq 2\|\hat{X}\|_2\|q\|, \\
               \eta I_{nr} \preceq \hat{H} \preceq I_{nr}.
$$  \hfill (12)

Given any feasible solution $(\delta, \hat{H})$ to (11), the tuple

$$
\left( \frac{1 - \delta}{1 + \delta}, \frac{1}{1 + \delta} \hat{H} \right)
$$

is a feasible solution to problem (12). Therefore, if the optimal value of (11) is denoted as $\delta^*_f(\hat{X})$ and the optimal value of (12) is denoted as $\eta^*_f(\hat{X})$, then it holds that

$$
\eta^*_f(\hat{X}) \geq \frac{1 - \delta^*_f(\hat{X})}{1 + \delta^*_f(\hat{X})} \geq \frac{1 - \delta^*(\hat{X})}{1 + \delta^*(\hat{X})} \geq \frac{1 - \delta}{1 + \delta},
$$  \hfill (13)
in which the last inequality is implied by $\delta \geq \delta^*(\hat{X})$ as shown above. To prove the inequality (6), we need to bound $\eta^*_f(\hat{X})$ from above, which can be achieved by finding a feasible solution to the dual problem of (12) given below:

\[
\min_{U_1, U_2, G, \lambda, y} \quad \text{tr}(U_2) + 4\|\hat{X}\|_2^2\|q\|^2\lambda + \text{tr}(G)
\]
\[\text{s.t.} \quad \text{tr}(U_1) = 1, \quad (\hat{X}y)e^T + e(\hat{X}y)^T = U_1 - U_2, \quad (14)\]

\[
\begin{bmatrix}
G & -y \\
-y^T & \lambda
\end{bmatrix} \succeq 0, \quad U_1 \succeq 0, \quad U_2 \succeq 0.
\]

For any matrix $\hat{X} \in \mathbb{R}^{n \times r}$ satisfying $\|\hat{X}\hat{X}^T - M^*\|_F \leq \tau\lambda_r(M^*)$, we have $\hat{X} \neq 0$, and it has been shown in the proof of Lemma 19 in [17] that there exists $y \neq 0$ satisfying the inequalities

\[
\|\hat{X}y\|^2 \geq 2\lambda_r(\hat{X}\hat{X}^T)\|y\|^2, \quad (15a)
\]
\[
\cos \theta \geq \sqrt{1 - \frac{3 + 2\sqrt{2}}{4} r^2}, \quad (15b)
\]

where $\theta$ is the angle between $\hat{X}y$ and $e$. Define

\[
M = (\hat{X}y)e^T + e(\hat{X}y)^T,
\]

and decompose $M$ as $M = [M]_+ - [M]_-$ with $[M]_+ \succeq 0$ and $[M]_- \succeq 0$. Then, it is easy to verify that $(U_1^*, U_2^*, G^*, \lambda^*, y^*)$ defined as

\[
y^* = \frac{y}{\text{tr}([M]_+)}, \quad U_1^* = \frac{[M]_+}{\text{tr}([M]_+)}, \quad U_2^* = \frac{[M]_-}{\text{tr}([M]_+)}, \quad G^* = \frac{y^*(y^*)^T}{\lambda^*}, \quad \lambda^* = \frac{\|y^*\|_2^2}{2\|\hat{X}\|_2\|q\|},
\]

forms a feasible solution to the dual problem (14) with the objective value

\[
\frac{\text{tr}([M]_-) + 4\|\hat{X}\|_2\|q\|^2\|y\|}{\text{tr}([M]_+)}. \quad (16)
\]

Furthermore, rank$(M^*) = r$ implies that $\lambda_r(M^*) > 0$. By the Wielandt–Hoffman theorem,

\[
|\lambda_r(\hat{X}\hat{X}^T) - \lambda_r(M^*)| \leq \|\hat{X}\hat{X}^T - M^*\|_F \leq \tau\lambda_r(M^*),
\]
\[
|\lambda_1(\hat{X}\hat{X}^T) - \lambda_1(M^*)| = \|\hat{X}\|_2^2 - \lambda_1(M^*) \leq \|\hat{X}\hat{X}^T - M^*\|_F \leq \tau\lambda_r(M^*).
\]

Thus, using the above two inequalities and inequality (15a), we have

\[
\frac{2\|\hat{X}\|_2\|y\|}{\|\hat{X}y\|} \leq \frac{2\|\hat{X}\|_2}{\sqrt{2\lambda_r(\hat{X}\hat{X}^T)}} \leq \sqrt{\frac{2(\lambda_1(M^*) + \tau\lambda_r(M^*))}{(1 - \tau)\lambda_r(M^*)}} = C(\tau, M^*). \quad (17)
\]

Next, according to Lemma 14 of [15], one can write

\[
\text{tr}([M]_+) = \|\hat{X}y\|\|e\|(1 + \cos \theta), \quad \text{tr}([M]_-) = \|\hat{X}y\|\|e\|(1 - \cos \theta).
\]

Substituting the above two equations and (17) into the dual objective value (16), one can obtain

\[
\eta^*_f(\hat{X}) \leq \frac{1 - \cos \theta + 2C(\tau, M^*)\|q\|/\|e\|}{1 + \cos \theta},
\]

which together with (13) implies that

\[
\|e\| \leq (1 + \delta)C(\tau, M^*)\|q\|(\cos \theta - \delta)^{-1}.
\]

The inequality (6) can then be proved by combining the above inequality and (15b) under the probabilistic event that $\|q\| \leq \epsilon$. \qed
The upper bound derived from inequality (3a).

The upper bound derived from inequality (3b).

Figure 1: Comparison of the upper bounds given by Theorem 2 for the distance $\|\hat{X}\hat{X}^T - M^*\|_F$ with $\hat{X}$ being an arbitrary local minimizer.

4 Numerical illustration

In this section, we will empirically study the developed probabilistic guarantees and demonstrate the distance $\|\hat{X}\hat{X}^T - M^*\|_F$ between any local minimizer $\hat{X}$ and the ground truth $M^*$ as well as the value of the RIP constant $\delta$ required to be satisfied by the linear operator $A$.

Before delving into the numerical illustration, note that the probability $P(\|w\| \leq w_0)$ used in both Theorem 2 and Theorem 4 can be lower bounded by the probability $P(\|w\| \leq w_0)$ with $w_0 = \epsilon/\|A\|_2$. The latter probability can be easily estimated when the probability distribution of the noise $w$ is given. As an example, in the simplest case when $w$ is sampled from an isotropic Gaussian distribution, i.e., $w \sim N(0, \sigma^2 I_m)$, the random variable $\|w/\sigma\|_2$ follows the chi-square distribution and one can apply the Chernoff bound to obtain

$$P(\|w\| \leq w_0) = 1 - P\left(\|w/\sigma\|^2 \geq \frac{w_0^2}{\sigma^2}\right) \geq 1 - \inf_{0 \leq t < 1/2} (1 - 2t)^{-m/2}e^{-tw_0^2/\sigma^2}.$$

After solving the minimization problem in the above equation, we obtain

$$1 - \left(\frac{2m\sigma^2}{w_0^2}\right)^{-m/2}e^{m\frac{w_0^2}{16\sigma^2}} \leq P(\|w\| \leq w_0).$$

More generally, if $w$ is $(\sigma/\sqrt{m\delta})$-sub-Gaussian vector, then applying Lemma 1 in [25] leads to

$$1 - 2e^{-\frac{w_0^2}{16\sigma^2}} \leq P(\|w\| \leq w_0).$$

For numerical illustration, assume that $n = 50$, $m = 10$ and $\|A\|_2 \leq 2$, while the noise $w$ is a $(0.05/\sqrt{m})$-sub-Gaussian vector. We also assume that the ground truth $M^*$ is of rank 5 with the largest eigenvalue being 1.5 and the smallest eigenvalue being 1.

First, we explore the two inequalities (3a) and (3b) presented in Theorem 2 to obtain two upper bounds on $\|\hat{X}\hat{X}^T - M^*\|_F$, where $\hat{X}$ denotes any arbitrary (worst) local minimizer. Figure 1 gives the contour plots of the two upper bounds on $\|\hat{X}\hat{X}^T - M^*\|_F$, which hold with the given probability on the $y$-axis and the given RIP constant $\delta$ from 0 to $1/2$ on the $x$-axis. While final bound on $\|\hat{X}\hat{X}^T - M^*\|_F$ is often determined by the inequality (3b), the inequality (3a) is needed theoretically to deal with the case when $\hat{X}$ has a singular value close to 0.

Next, we illustrate the bounds given by Theorem 2 and Theorem 4. Figure 2 shows the contour plots of the maximum RIP constant $\delta$ that is necessary to guarantee that each local minimizer $\hat{X}$ (satisfying the inequality (5) when Theorem 4 is applied) lies within a certain neighborhood of the ground truth (measured via the distance $\|\hat{X}\hat{X}^T - M^*\|_F$ on the $x$-axis) with a given probability on the $y$-axis, as implied by the respective global and local guarantees. Figure 2 clearly shows how a smaller RIP
constant $\delta$ leads to a tighter bound on the distance $\|\hat{X}^T - M^*\|_F$ with a higher probability. In addition, the local guarantee generally requires a looser RIP assumption as it still holds even when $\delta > 1/2$. However, as the parameter $\tau$ in Theorem 4 increases, the local bound also degrades quickly, sometimes becoming worse than the global bound as illustrated in Figure 2(d).

5 Conclusion

In this paper, we develop global and local analyses for the locations of the local minima of the low-rank matrix recovery problem with noisy linear measurements. Unlike the existing results, the probability distribution of the noise is arbitrary and the RIP constant of the problem is free to take arbitrary values. The developed results encompass the state-of-the-art results on the non-existence of spurious solutions in the noiseless case. Our analyses show how the value of the RIP constant and the intensity of noise affect the landscape of the non-convex learning problem and the locations of the local minima relative to the ground truth.

References


A Proofs of Lemma 3 and Theorem 3

Proof of Lemma 3. Let \( Z \in \mathbb{R}^{n \times r} \) be a matrix satisfying \( ZZ^T = M^* \). Similar to the proof of Theorem 4, we introduce an optimization problem as follows:

\[
\max_{\eta, \hat{H}} \eta \\
\text{s.t.} \quad \|\hat{X}^T \hat{H} e\| \leq 2\|\hat{X}\|_2 \epsilon, \quad 2I_r \otimes \text{mat}_S(\hat{H} e) + \hat{X}^T \hat{X} \succeq -2\epsilon I_{nr}, \quad \eta I_{n^2} \preceq \hat{H} \preceq I_{n^2}, \tag{18}
\]

where its optimal value \( \eta^*(\hat{X}) \) satisfies the inequality

\[
\eta^*(\hat{X}) \geq \frac{1 - \delta^*(\hat{X})}{1 + \delta^*(\hat{X})} \geq 1 - \delta \tag{19}
\]

In the remaining part, we will prove the following upper bound on \( \eta^*(\hat{X}) \):

\[
\eta^*(\hat{X}) \leq \frac{1}{3} + \frac{2\epsilon \sqrt{\tau} + 2\sqrt{2\epsilon (1 + \delta)} \|\hat{X}\|_2}{\|e\|}. \tag{20}
\]

The inequality (3b) is a consequence of (19), (20) and the inequality

\[
\|\hat{X}\|_2 \leq \|\hat{X} \hat{X}^T\|_{F}^{1/2} \leq \|\hat{X} \hat{X}^T - M^*\|_{F}^{1/2} + \|M^*\|_{F}^{1/2}.
\]

The proof of the upper bound (20) can be completed by finding a feasible solution to the dual problem of (18):

\[
\min_{U_1, U_2, W, G, \lambda, y} \text{tr}(U_2) + \langle \hat{X}^T \hat{X}, W \rangle + 2\epsilon \text{tr}(W) + 4\|\hat{X}\|_2^2 \lambda + \text{tr}(G) \\
\text{s.t.} \quad \text{tr}(U_1) = 1, \\
(\hat{X} y - w)e^T + e(\hat{X} y - w)^T = U_1 - U_2, \\
\begin{bmatrix}
G & -y^T \\
y^T & \lambda
\end{bmatrix} \succeq 0,
\]

\[
U_1 \succeq 0, \quad U_2 \succeq 0, \quad W = \begin{bmatrix} W_{1,1} & \cdots & W_{r,1} \\
\vdots & \ddots & \vdots \\
W_{r,1} & \cdots & W_{r,r} \end{bmatrix} \succeq 0,
\]

\[
w = \sum_{i=1}^r \text{vec}(W_{i,i}). \tag{21}
\]

Before describing the choice of the dual feasible solution, we need to represent the error vector \( e \) in a different form. Let \( P \in \mathbb{R}^{n \times n} \) be the orthogonal projection matrix onto the range of \( \hat{X} \), and \( P_\perp \in \mathbb{R}^{n \times n} \) be the orthogonal projection matrix onto the orthogonal complement of the range of \( \hat{X} \). Then, \( Z \) can be decomposed as \( Z = PZ + P_\perp Z \), and there exists a matrix \( R \in \mathbb{R}^{r \times r} \) such that \( PZ = \hat{X} R \). Note that

\[
ZZ^T = PZZ^T P + PZZ^T P_\perp + P_\perp ZZ^T P + P_\perp ZZ^T P_\perp.
\]

Thus, if we choose

\[
\hat{Y} = \frac{1}{2} \hat{X} - \frac{1}{2} \hat{X} RR^T - P_\perpZR^T, \quad \hat{y} = \text{vec}(\hat{Y}), \tag{22}
\]

then it can be verified that

\[
\hat{X} \hat{Y}^T + \hat{Y} \hat{X}^T - P_\perp ZZ^T P_\perp = \hat{X} \hat{X}^T - ZZ^T, \\
\langle \hat{X} \hat{Y}^T + \hat{Y} \hat{X}^T, P_\perp ZZ^T P_\perp \rangle = 0.
\]

\]

Moreover, we have
\[
\|\hat{X} \hat{Y}^T + \hat{Y} \hat{X}^T\|^2_F = 2 \text{tr}(\hat{X}^T \hat{Y} \hat{X}^T \hat{Y}) + \text{tr}(\hat{X}^T \hat{Y} \hat{X}^T \hat{Y}) + \text{tr}(\hat{Y} \hat{X}^T \hat{X}) \\
\geq 2 \text{tr}(\hat{X}^T \hat{Y} \hat{X}^T \hat{Y}) \geq 2\sigma_2^2(\hat{X})\|\hat{Y}\|_F^2, \tag{23}
\]
in which the first inequality is due to
\[
\text{tr}(\hat{X}^T \hat{Y} \hat{X}^T \hat{Y}) = \frac{1}{4} \text{tr}((\hat{X}^T \hat{X}(I_r - RR^T))^2) = \frac{1}{4} \text{tr}((\hat{X}(I_r - RR^T)\hat{X})^2) \geq 0.
\]
Assume first that \(Z = Z_\perp\), the other case will be handled at the end of this proof. In the case when \(Z \neq 0\), we also have \(\hat{X} \hat{Y}^T + \hat{Y} \hat{X}^T \neq 0\). Otherwise, the inequality (23) and the assumption \(\sigma_2(\hat{X}) > 0\) imply that \(\hat{Y} = 0\). The orthogonality and the definition of \(\hat{Y}\) in (22) then give rise to
\[
\hat{X} - \hat{X} RR^T = 0, \quad \mathcal{P}_\perp Z R = 0.
\]
The first equation above implies that \(R\) is invertible since \(\hat{X}\) has full column rank, which contradicts \(Z \neq 0\). Now, define the unit vectors
\[
\hat{u}_1 = \frac{\hat{X} \hat{y}}{\|\hat{X} \hat{y}\|}, \quad \hat{u}_2 = \frac{\text{vec}(Z_\perp^{-1} Z_\perp^T)}{\|Z_\perp^{-1} Z_\perp^T\|_F}.
\]
Then, \(\hat{u}_1 \perp \hat{u}_2\) and
\[
eq \|e\||(1 - \alpha^2 \hat{u}_1 - \alpha \hat{u}_2)
\]
with
\[
\alpha = \frac{\|Z_\perp^{-1} Z_\perp^T\|_F}{\|\hat{X} \hat{X}^T - ZZ^T\|_F} \tag{25}
\]
We first describe our choices of the dual variables \(W\) and \(y\) (which will be scaled later). Let
\[
\hat{X}^T \hat{X} = QSQ^T, \quad Z_\perp Z_\perp^T = PGP^T,
\]
with orthogonal matrices \(Q, P\) and diagonal matrices \(S, G\) such that \(S_{11} = \sigma_2^2(\hat{X})\). Fix a constant \(\gamma \in [0, 1]\) that is to be determined and define
\[
V_i = k^{1/2}G_{i_i}^{1/2}PE_{i_i}Q^T, \quad \forall i = 1, \ldots, r,
\]
\[
W = \sum_{i=1}^r \text{vec}(V_i) \text{vec}(V_i)^T, \quad y = \mathbf{l} \hat{y},
\]
with \(\hat{y}\) defined in (22) and
\[
k = \frac{\gamma}{\|e\|\|Z_\perp^{-1} Z_\perp^T\|_F}, \quad l = \sqrt{1 - \gamma^2} \|\hat{X}\|_{F},
\]
Here, \(E_{ij}\) is the elementary matrix of size \(n \times r\) with the \((i, j)\)-entry being 1. By our construction, \(\hat{X}^T V_i = 0\), which implies that
\[
\langle \hat{X}^T \hat{X}, W \rangle = \sum_{i=1}^r \|\hat{X} V_i^T + V_i \hat{X}^T\|^2_F = 2 \sum_{i=1}^r \text{tr}(\hat{X}^T \hat{X} V_i^T V_i) = 2k\sigma_2^2(\hat{X}) \sum_{i=1}^r G_{i_i} = 2 \beta \gamma, \tag{26}
\]
with
\[
\beta = \frac{\sigma_2^2(\hat{X}) \text{tr}(Z_\perp^{-1} Z_\perp^T)}{\|\hat{X} \hat{X}^T - ZZ^T\|_F \|Z_\perp^{-1} Z_\perp^T\|_F} \tag{27}
\]
In addition,
\[
\text{tr}(W) = \sum_{i=1}^r \|V_i\|^2_F = k \sum_{i=1}^r G_{i_i} = k \text{tr}(Z_\perp^{-1} Z_\perp^T) \leq \sqrt{r} \|e\|, \tag{28}
\]
and
\[
w = \sum_{i=1}^r \text{vec}(W_{i, i}) = \sum_{i=1}^r V_i V_i^T = k Z_\perp Z_\perp^T.
\]
Therefore,
\[ \hat{X}y - w = \frac{1}{\|e\|}(\sqrt{1 - \gamma^2} \hat{u}_1 - \gamma \hat{u}_2), \]
which together with (24) implies that
\[ \|e\|\|\hat{X}y - w\| = 1, \quad \langle e, \hat{X}y - w \rangle = \gamma \alpha + \sqrt{1 - \gamma^2} \sqrt{1 - \alpha^2} = \psi(\gamma). \tag{29} \]
Next, the inequality (23) and the assumption on \( \sigma_r(\hat{X}) \) imply that
\[ \|y\| \leq \frac{\sqrt{1 - \gamma^2}}{\sqrt{2\sigma_r(\hat{X})\|e\|}} \leq \frac{\sqrt{1 + \delta}}{\sqrt{2\|e\|}}. \tag{30} \]
Define
\[ M = (\hat{X}y - w)e^T + e(\hat{X}y - w)^T, \]
and decompose \( M \) as \( M = [M]_+ - [M]_- \) in which both \([M]_+ \geq 0\) and \([M]_- \geq 0\). Let \( \theta \) be the angle between \( e \) and \( \hat{X}y - w \). By Lemma 14 in [15], we have
\[ \text{tr}([M]_+) = \|e\|\|\hat{X}y - w\|(1 + \cos \theta), \quad \text{tr}([M]_-) = \|e\|\|\hat{X}y - w\|(1 - \cos \theta). \]
Now, one can verify that \( (U_1^*, U_2^*, W^*, G^*, \lambda^*, y^*) \) defined as
\[
\begin{align*}
U_1^* &= \frac{[M]_+}{\text{tr}([M]_+)}; \quad U_2^* = \frac{[M]_-}{\text{tr}([M]_+)}, \quad y^* = \frac{y}{\text{tr}([M]_+)}; \\
W^* &= \frac{\lambda^*}{2\|\hat{X}\|_2 \|\|y^*\|\}}; \quad \lambda^* = \frac{\|y^*\|_2 \|\|e\|\}}{\beta \gamma}; \quad G^* = \frac{1}{\lambda^*} y^* y^{*T}
\end{align*}
\]
forms a feasible solution to the dual problem (21) whose objective value is equal to
\[ \frac{\text{tr}([M]_-) + \langle \hat{X}^T \hat{X}, W \rangle + 2\epsilon \text{tr}(W) + 4\|\hat{X}\|_2 \|y\|}{\text{tr}([M]_+)} \]
Substituting (26), (28), (29) and (30) into the above equation, we obtain
\[
\eta^*(\hat{X}) \leq \frac{2\beta \gamma + 1 - \psi(\gamma) + (2\epsilon \sqrt{\gamma} + 2\sqrt{2\epsilon (1 + \delta)}\|\hat{X}\|_2)/\|e\|}{1 + \psi(\gamma)} \\
\leq \frac{2\beta \gamma + 1 - \psi(\gamma)}{1 + \psi(\gamma)} + \frac{2\epsilon \sqrt{\gamma} + 2\sqrt{2\epsilon (1 + \delta)}\|\hat{X}\|_2}{\|e\|}.
\]
Choosing the best value of the parameter \( \gamma \in [0, 1] \) to minimize the far right-side of the above inequality leads to
\[ \frac{2\beta \gamma + 1 - \psi(\gamma)}{1 + \psi(\gamma)} \leq \eta_0(\hat{X}), \]
with
\[ \eta_0(\hat{X}) := \begin{cases} 
1 - \sqrt{1 - \alpha^2}, & \text{if } \beta \geq \frac{\alpha}{1 + \sqrt{1 - \alpha^2}}, \\
\frac{\beta \alpha}{1 - \beta \alpha}, & \text{if } \beta \leq \frac{\alpha}{1 + \sqrt{1 - \alpha^2}}.
\end{cases} \]
Here, \( \alpha \) and \( \beta \) are defined in (25) and (27), respectively. In the proof of Theorem 1.2 in [20], it is shown that \( \eta_0(\hat{X}) \leq 1/3 \) for every \( \hat{X} \) with \( \hat{X} \hat{X}^T \neq ZZ^T \), which gives the upper bound (20).

Finally, we still need to deal with the case when \( P_{\perp} Z = 0 \). In this case, we know that \( \hat{X}y = e \) with \( \hat{y} \) defined in (22). Then, it is easy to check that \( (U_1^*, U_2^*, W^*, G^*, \lambda^*, y^*) \) defined as
\[
\begin{align*}
U_1^* &= \frac{e e^T}{\|e\|^2}, \quad U_2^* = 0, \quad y^* = \frac{\hat{y}}{2\|e\|^2}; \\
W^* &= 0, \quad \lambda^* = \frac{\|y^*\|_2}{2\|\hat{X}\|_2 \epsilon}; \quad G^* = \frac{1}{\lambda^*} y^* y^{*T}
\end{align*}
\]
forms a feasible solution to the dual problem (21) whose objective value is \(4\|\hat{X}\|_2\|y^*\|\). By the inequality (23), we have

\[
\eta^*(\hat{X}) \leq 4\|\hat{X}\|_2\|y^*\| \leq \frac{\sqrt{2}\epsilon\|\hat{X}\|_2}{\sigma_r(\hat{X})\|e\|} \leq \frac{\sqrt{2}(1 + \delta)\|\hat{X}\|_2}{\|e\|}.
\]

Hence, the upper bound (20) still holds in this case. \(\square\)

**Proof of Theorem 3.** The proof of Theorem 3 is similar to the above proof of Lemma 3, and we will only emphasize the difference here. In the case when \(\hat{X} \neq 0\), after constructing the feasible solution to the dual problem (21), we have

\[
\frac{1 - \delta}{1 + \delta} \leq \eta^*(\hat{X}) \leq \frac{\text{tr}([M_-] + \langle \hat{X}^T \hat{X}, W \rangle) + 2\epsilon \text{tr}(W) + 4\|\hat{X}\|_2\epsilon\|y\|}{\text{tr}([M_+] + 1)}.
\]  

(31)

Note that in the rank-1 case, one can write \(\sigma_r(\hat{X}) = \|\hat{X}\|_2\) and

\[
\|y\| \leq \frac{\|\hat{g}\|}{\|e\|\|\hat{X}\|_2} \leq \frac{1}{\sqrt{2}\|\hat{X}\|_2\|e\|},
\]

in which the last inequality is due to (23). Substituting (26), (28), (29) and the above inequality into (31) and choosing an appropriate \(\gamma\) as shown in the proof of Lemma 3 we obtain

\[
\frac{1 - \delta}{1 + \delta} \leq \eta^*(\hat{X}) \leq \frac{2\beta\gamma + 1 - \psi(\gamma) + (2\epsilon + 2\sqrt{2}\epsilon)/\|e\|}{1 + \psi(\gamma)}
\]

\[
\leq \frac{1}{3} + \frac{2\epsilon + 2\sqrt{2}\epsilon}{\|e\|},
\]

which implies inequality (4) under the probabilistic event that \(\|q\| \leq \epsilon\).

In the case when \(\hat{X} = 0\), \((U_1^*, U_2^*, W^*, \lambda^*, y^*)\) defined as

\[
U_1^* = \frac{ee^T}{\|e\|^2}, \quad U_2^* = 0, \quad y^* = 0,
\]

\[
W^* = \frac{ZZ^T}{\|e\|^2}, \quad \lambda^* = 0, \quad G^* = 0
\]

forms a feasible solution to the dual problem (21), which shows that

\[
\frac{1 - \delta}{1 + \delta} \leq \eta^*(\hat{X}) \leq \frac{\epsilon}{\|e\|}.
\]

The above inequality also implies inequality (4) under the probabilistic event that \(\|q\| \leq \epsilon\). \(\square\)