On the Optimization Landscape of Burer-Monteiro Factorization: When do Global Solutions Correspond to Ground Truth?

Jianhao Ma* and Salar Fattahi+

Department of Industrial and Operations Engineering
University of Michigan
*jianhao@umich.edu, +fattahi@umich.edu

February 21, 2023

Abstract

In low-rank matrix recovery, the goal is to recover a low-rank matrix, given a limited number of linear and possibly noisy measurements. Low-rank matrix recovery is typically solved via a nonconvex method called *Burer-Monteiro factorization* (BM). If the rank of the ground truth is known, BM is free of sub-optimal local solutions, and its true solutions coincide with the global solutions—that is, the true solutions are *identifiable*. When the rank of the ground truth is unknown, it must be over-estimated, giving rise to an over-parameterized BM. In the noiseless regime, it is recently shown that over-estimation of the rank leads to progressively fewer sub-optimal local solutions while preserving the identifiability of the true solutions. In this work, we show that with noisy measurements, the global solutions of the over-parameterized BM no longer correspond to the true solutions, essentially transmuting over-parameterization from blessing to curse. In particular, we study two classes of low-rank matrix recovery, namely *matrix completion* and *matrix sensing*. For matrix completion, we show that even if the rank is only slightly over-estimated and with very mild assumptions on the noise, none of the true solutions are local or global solutions. For matrix sensing, we show that to guarantee the correspondence between global and true solutions, it is necessary and sufficient for the number of samples to scale linearly with the over-estimated rank, which can be drastically larger than its optimal sample complexity that only scales with the true rank.

1 Introduction

We study the optimization landscape of *low-rank matrix recovery*, where the goal is to recover a matrix $X^\star \in \mathbb{R}^{d_1 \times d_2}$ with rank $r \leq \min\{d_1, d_2\}$ from m linear and possibly noisy measurements:

$$\text{find } X^\star \qquad \text{such that} \qquad \text{rank}(X^\star) = r, \quad y = \mathcal{A}(X^\star) + \epsilon.$$

The measurement operator $\mathcal{A} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^m$ is defined as $\mathcal{A}(X) = [\langle A_1, X\rangle \quad \dots \quad \langle A_m, X\rangle]^\top$, and $\epsilon \in \mathbb{R}^m$ is the additive noise vector. Low-rank matrix factorization plays a central role in many modern machine learning problems, including motion detection in video frames [BZ14], face recognition [LFL+14], and collaborative filtering in recommender systems [LZXZ14]. Despite its widespread applications, this problem is inherently nonconvex due to its rank constraint $\text{rank}(X^\star) =$

r, and it is commonly solved via a computationally tractable technique called *Burer-Monteiro factorization* (BM). In BM, the target low-rank matrix is modeled as $X^\star = W_1^\star W_2^\star$, where $W_1^\star \in \mathbb{R}^{d_1 \times k}$ and $W_2^\star \in \mathbb{R}^{k \times d_2}$ are unknown factors each with search rank k. When $k = r$, the rank constraint $\operatorname{rank}(X^*) = r$ is imposed implicitly via the decomposition $W_1 W_2$. Given this factorized model, the problem of recovering the true solution is typically formulated as follows:

$$\min_{\mathbf{W}=(W_1,W_2)} f_{\ell_q}(\mathbf{W}) := \frac{1}{m} \|y - \mathcal{A}(W_1 W_2)\|_{\ell_q}^q . \tag{BM}$$

There are two fundamental challenges when solving BM: first, despite its nonconvexity, the most efficient way to solve it is via local-search algorithms originally designed for convex problems. In the absence of convexity, local-search algorithms may converge to local solutions that are not optimal globally. Second, in the event that these algorithms can recover a globally optimal solution, the recovered solution may correspond to a solution overfitted to noise rather than the ground truth $X^\star$.

Despite the aforementioned challenges, recent studies have postulated that different variants of BM are endowed with two important properties. First, they have *benign landscape*, that is, they are devoid of spurious, sub-optimal local solutions. Second, their true solutions are *identifiable*, i.e., any global solution $\mathbf{W}^\star = (W_1^\star, W_2^\star)$ satisfies $W_1^\star W_2^\star = X^\star$. The success of local-search algorithms can be explained when both conditions are met: the benign landscape ensures that the algorithm eventually recovers a globally optimal solution [LPP+19, JGN+17, FLZ19], while the identifiability ensures that the recovered solution corresponds to the ground truth. The notion of the benign landscape appears to be omnipresent in low-rank matrix recovery (see recent surveys [CLC19, ZQW20]). Alternatively it can also be enforced locally [ZZ20, MBLS22] or through different regularization techniques [ZLTW21, GJZ17]. In contrast, the identifiability of the true solutions has been largely overlooked in the literature. Indeed, a benign landscape for BM without identifiable true solutions would lead to fundamentally flawed and often imperceptible global solutions that are easy to recover, but potentially very far from the ground truth $X^\star$. An important question thus arises:

Under what conditions does BM have identifiable true solutions?

1.1 Burer-Monteiro or Convex Relaxation?

Towards answering the above question, the works [BNS16, GLM16] showed that, under some conditions, two important variants of BM, namely *matrix completion* and *matrix sensing*, have benign landscape and identifiable true solutions. These results led to a flurry of follow-up papers showing that other variants of low-rank matrix recovery also enjoy similar properties [SQW18, SQW16, ZLK+17, FS20, ZLTW18].

Despite the recent advances in characterizing the global geometry of the low-rank matrix recovery, their effectiveness has remained limited due to at least one of the following assumptions. First, it is assumed that the true rank r of $X^\star$ is known *a priori* and coincides with the search rank k of the factorized model $W_1 W_2$. Second, it is assumed that the measurements are either noiseless or corrupted with light-tailed noise (e.g., Gaussian). Neither of these conditions is satisfied in real-world instances; in practice, the rank of $X^\star$ is rarely known and instead over-estimated with $k \gg r$, and the measurements are often corrupted with heavy-tailed noise.

At first glance, one may speculate that these assumptions can be lifted. First, when the true rank is unknown and the measurements are noisy, the so-called *convex relaxation* of the low-rank matrix recovery, which consists in dropping the rank constraint and instead penalizing it

in the objective with a convex proxy, often enjoys a benign landscape (due to its convexity) and identifiability of its true solutions. More precisely, for matrix sensing, the global solutions of the convex relaxation coincide with the ground truth up to a minimax optimal error [CP11] provided that the measurement size scales with the true dimension of $X^\star$, that is, $m \gtrsim \max\{d_1, d_2\}r$ which is information-theoretically optimal. Similarly, convex relaxation techniques can solve another—arguably more popular—variant of low-rank matrix recovery, called matrix completion, with the same near-optimal sample complexity even when the true rank is unknown and a subset of measurements is grossly corrupted with heavy-tailed noise [CLMW11, CSPW11, CCF+19]. Thus, fundamentally, it is possible to guarantee identifiability even if the true rank is unknown and the measurements are corrupted with heavy-tailed noise.[1] Second, for the nonconvex BM with ℓ_1-loss, it has been recently shown that the (sub-)gradient method with small initialization correctly recovers $X^\star$ with the same optimal sample complexity as the convex relaxation techniques, even when the search rank is over-estimated and the measurements are grossly corrupted with noise [MF22]. Collectively, these observations provide strong evidence that the true solutions may be identifiable in the over-parameterized BM, potentially under the same minimal conditions as those required for convex relaxation.

1.2 Summary of Contributions

Contrary to the above observations, we show that the conditions for the identifiability of its true solutions in BM become increasingly restrictive as the search rank exceeds the true rank and the measurements become noisy. In particular, we study the optimization landscape of BM with a general ℓ_q-loss and under a fairly general noise model for two notable classes of low-rank matrix recovery, matrix sensing and matrix completion, and show the following results:

- For the asymmetric matrix completion with search rank $k > r$, none of the true solutions are global minima of BM with ℓ_q-loss for any $q \geq 1$. This result holds even when the full matrix $X^\star$ is observed and only an arbitrarily small fraction of measurements are corrupted with noise. Similarly, for the asymmetric matrix sensing with the measurement size $m \lesssim \max\{d_1, d_2\}k$, none of the true solutions are global minima of BM with ℓ_q-loss for any $q \geq 1$, provided that a subset of the measurements are corrupted with noise. Therefore, to guarantee the identifiability of the true solutions, the number of measurements must scale at least linearly with the search rank k. This can be drastically larger than its information-theoretically optimal sample complexity that only scales with the true rank r. More severely, under the same conditions, there exists at least one true solution that is not even a critical point of BM for asymmetric matrix sensing and completion.

- For the symmetric matrix completion with a positive semidefinite ground truth $X^\star \in \mathbb{R}^{d\times d}$ and $k > r$, none of the true solutions are local or global minima of BM, so long as least one of the observed diagonal entries of $X^\star$ is corrupted with noise. Similarly, for the symmetric matrix sensing with $m \lesssim dk$, none of the true solutions are local or global minima of BM provided that a subset of the measurements are corrupted with noise. Despite these negative results, BM enjoys a flatter landscape around its true solutions in the symmetric setting.

[1] The main advantage of BM over its convex counterpart is a reduction in the variable size; from $(d_1 + d_2)^2$ to $(d_1 + d_2)k$ variables. In scenarios where $k \ll d_1 + d_2$ and $d_1 + d_2$ are in the order of millions, this reduction in the variable size is precisely the reason behind the tractability of the low-rank matrix recovery.

- The provided necessary identifiability conditions are also sufficient for the special case of asymmetric and symmetric matrix sensing with ℓ_1-loss. More precisely, with sample sizes satisfying $m \gtrsim \max\{d_1, d_2\}k$ or $m \gtrsim dk$, the true solutions of asymmetric or symmetric matrix sensing coincide with the global solutions, provided that less than half of the measurements are corrupted with noise.

Our results highlight the fragility of identifiability conditions for BM in the presence of noise and over-parameterization. In the noiseless regime, it is shown that over-parameterization can lead to progressively fewer spurious local solutions, while preserving the identifiability of the true solutions [Zha22]. Our findings is in stark contrast with this result, showing that noisy measurements can drastically change the landscape of BM around its true solutions, essentially transmuting over-parameterization from blessing to curse.

1.3 Related Work

For the noiseless symmetric matrix sensing, BM with ℓ_2-loss and $k = r$ satisfies the strict saddle property, that is, all of its stationary points are either global solutions or strict saddle points that are easily escapable via gradient-based algorithms [BNS16, JGN+17]. This result also holds for other variants of low-rank matrix recovery, such as matrix completion [GLM16], phase retrieval [SQW18], as well as those with more general smooth loss functions [ZLTW18]. In the asymmetric setting, the earlier results show that BM augmented with a balancing regularizer satisfies the same strict saddle property [GJZ17, ZLTW21]. Recently, it has been shown that the balancing regularization is in fact not needed for this property to hold [LLZ+20]. When the strict saddle property does not hold globally, it can be established locally around the true solutions with milder conditions [ZZ20, MBLS22]. Another line of work has studied the landscape of BM with ℓ_1-loss, which is known to be robust against outlier noise. Due to the nonsmoothness of the ℓ_1-loss, the strict saddle property cannot be applied. Nonetheless, it is proven that when $k = r = 1$, most first-order stationary points correspond to the ground truth for matrix completion [FS20, JL22] and matrix sensing [MF21].

Despite their significance, the existing results on the landscape of BM face major breakdowns when the search rank is over-estimated and the measurements are noisy. For instance, undesirable local (or even global) solutions are ubiquitous in the noisy and over-parameterized matrix sensing with ℓ_1-loss [MF22]. Instead, a recent body of work has focused on the *trajectory analysis* of gradient-based methods on BM. It has been recently shown that the trajectories of gradient-based algorithms enjoy implicit regularization [GWB+17, ACHL19, MF19] and incremental learning [MGF22]. For BM with ℓ_2-loss, different variants of gradient descent with small initialization can recover the ground truth, provided that the measurements are noiseless [SS21, XSCM23]. This result has been recently extended to the noisy settings, where it is shown that similar algorithms converge to solutions up to some error [ZKHC21, ZFZ21]. Finally, the recent work [MF22] shows that the sub-gradient method with small initialization applied to BM with ℓ_1-loss converges to the ground truth, even if the does not correspond to the global solution.

In light of the existing literature, our findings call into question the promise of global landscape analysis of BM, especially when spurious local solutions *do exist* or when the true and global solutions *do not coincide*. Instead, we advocate for a finer-grained trajectory analysis of local-search algorithms for BM, since they do not rely on the equivalence between the true and global solutions if initialized properly. Finally, we highlight a recent paper [YMLS22] that compares the performance of BM with the convex relaxation, showing that these techniques are essentially incomparable when applied to

certain classes of matrix completion and matrix sensing.

Notations. For a matrix M, its operator, Frobenius, and element-wise ℓ_q norms are denoted as $\|M\|$, $\|M\|_F$, and $\|M\|_{\ell_q}$, respectively. The symbol $0_{m\times n}$ refers to $m \times n$ zero matrix. Similarly, $I_{m\times m}$ is used to denote $m \times m$ identity matrix. We define $I_{m\times n} = [I_{m\times m}, 0_{m,n-m}]$ if $n > m$, and $I_{m\times n} = [I_{n\times n}, 0_{n,m-n}]^\top$ if $n < m$. For two matrices X and Y of the same size, their inner product is defined as $\langle X, Y\rangle = \mathrm{Tr}(X^\top Y)$, where $\mathrm{Tr}(\cdot)$ is the trace operator. The notation $\mathcal{B}_{\bar{M}}(\gamma)$ is used to denote the unit norm ball with radius γ centered at $\bar{M} \in \mathbb{R}^{m\times n}$, i.e., $\mathcal{B}_{\bar{M}}(\gamma) := \{M \in \mathbb{R}^{m\times n} : \|M - \bar{M}\|_F \le \gamma\}$. When there is no ambiguity, we omit the subscript when the center is assumed to be the origin.

The $\mathrm{Sign}(\cdot)$ function is defined as $\mathrm{Sign}(x) = x/|x|$ if $x \neq 0$, and $\mathrm{Sign}(0) \in [-1, 1]$. Given two sequences $f(n)$ and $g(n)$, the notation $f(n) \lesssim g(n)$ implies that there exists a universal constant C satisfying $f(n) \le Cg(n)$. Moreover, the notation $f(n) \asymp g(n)$ implies that $f(n) \lesssim g(n)$ and $g(n) \lesssim f(n)$. Throughout the paper, the symbols $C, c_1, c_2, \ldots$ refer to universal constants whose precise value may change according to the context.

2 Preliminaries

For a function $f : \mathbb{R}^n \to \mathbb{R}$, a point $\bar{x}$ is called a *global solution* if it corresponds to its global minimizer. Moreover, a point $\bar{x}$ is called a *local solution* if it corresponds to the minimum of $f(x)$ within an open ball centered at $\bar{x}$. The directional derivative of f at point x in the feasible direction d is defined as

$$f'(x, d) = \lim_{t\to 0^+} \frac{f(x + td) - f(x)}{t}, \tag{1}$$

provided that the limit exists. If $f'(x, d) < 0$, then d is called a *descent direction*. For a locally Lipschitz function f, the *Clarke generalized directional derivative* at the point x in the feasible direction d is defined as

$$f^\circ(x, d) := \limsup_{\substack{y\to x \\ t\to 0^+}} \frac{f(y + td) - f(y)}{t} \tag{2}$$

provided that the limit exists. Note the difference between the ordinary directional derivative $f'(x, d)$ and its Clarke generalized counterpart: in the latter, the limit is taken with respect to a *variable* vector y that approaches x, rather than taking the limit exactly at x. It is a well-known fact that $f'(x, d) \le f^\circ(x, d)$ for every direction d, if both $f'(x, d)$ and $f^\circ(x, d)$ exist [Cla75, Proposition 1.4]. A function f is called *subdifferentially regular* if $f'(x, d) = f^\circ(x, d)$ for every direction d [Cla90, Definition 2.3.4]. The *Clarke subdifferential* of f at x is defined as the following set (see [Cla75, Definition 1.1 and Proposition 1.4]):

$$\partial f(x) := \{\psi | f^\circ(x, d) \ge \langle \psi, d\rangle, \forall d \in \mathbb{R}^n\}. \tag{3}$$

A point $\bar{x}$ is called *critical* if $0 \in \partial f(\bar{x})$, or equivalently, $f^\circ(\bar{x}, d) \ge 0$ for every feasible direction d. The following properties of the critical points are adapted from [LSM20] and will be used in our subsequent arguments.

Lemma 1. *For a real-valued locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, we have:*

1. *Every local solution of f is critical.*
2. *Given a point x, suppose that $f'(x,d) \geq 0$ for every feasible $d \in \mathbb{R}^n$. Then, x is a critical point of f.*
3. *Suppose that $f = \max_{1\leq i\leq m} g_i$ for some smooth functions $g_1, \ldots, g_m$. Given some point x, suppose that $f'(x,d) < 0$ for some feasible direction $d \in \mathbb{R}^n$. Then, x is not a critical point of f.*

Proof. The first property is a direct consequence of [Cla90, Proposition 2.3.2]. The second property follows from the basic inequality $f'(x,d) \leq f^\circ(x,d)$ [Cla75, Proposition 1.4] and the definition of the critical points. Finally, due to [RW09, Example 7.28], f is differentiably regular, and hence, $f^\circ(x,d) = f'(x,d)$ for every feasible direction d. This implies that x is not critical due to the existence of a direction d for which $f^\circ(x,d) = f'(x,d) < 0$. □

3 Main Results

In this work, we study the landscape of BM for two classes of low-rank matrix recovery, namely matrix sensing and matrix completion. In matrix sensing, the measurement matrices are designed according to the following model:

Assumption 1 (Matrix Sensing Model)**.** *For every $1 \leq i \leq m$, the entries of the measurement matrix A_i are independently drawn from a standard Gaussian distribution with zero mean and unit variance.* [2]

Recall that in BM, the true solution is modeled as $W_1 W_2$, where $W_1 \in \mathbb{R}^{d_1\times k}$ and $W_1 \in \mathbb{R}^{k\times d_2}$ for some search rank $k \geq r$. BM adapted to matrix sensing is thus formulated as:

$$\min_{W_1\in\mathbb{R}^{d_1\times k}, W_2\in\mathbb{R}^{k\times d_2}} f^{\mathrm{s}}_{\ell_q}(\mathbf{W}) := \frac{1}{m}\sum_{i=1}^{m} |y_i - \langle A_i, W_1W_2\rangle|^q, \text{ where } A_i(k,l) \sim \mathcal{N}(0,1),\ \forall i,k,l. \tag{MS-asym}$$

In *symmetric* matrix sensing, the true solution $X^\star$ is additionally assumed to be positive semidefinite with dimension $d \times d$. In this setting, the true solution can be modeled as $X^\star = W^\star W^{\star\top}$ with $W^\star \in \mathbb{R}^{d\times k}$ and BM can be reformulated as:

$$\min_{W\in\mathbb{R}^{d\times k}} f^{\mathrm{s}}_{\ell_q}(W) := \frac{1}{m}\sum_{i=1}^{m} \left|y_i - \left\langle A_i, WW^\top\right\rangle\right|^q, \text{ where } A_i(k,l) \sim \mathcal{N}(0,1), \forall i,k,l. \tag{MS-sym}$$

Another important subclass of low-rank matrix recovery is matrix completion, where the linear operator $\mathcal{A}(\cdot)$ is assumed to be element-wise projection:

Assumption 2 (Matrix completion model)**.** *Each index pair (k,l) belongs to a measurement set Ψ with a sampling probability $0 \leq s \leq 1$, i.e., $\mathbb{P}((x,y) \in \Psi) = s$ for every $1 \leq x \leq d_1$ and $1 \leq y \leq d_2$. For every $(x',y') \in \Psi$, there exists $1 \leq i \leq |\Psi|$ such that the measurement matrix A_i is defined as $A_i(x,y) = 1$ if $(x,y) = (x',y')$ and $A_i(x,y) = 0$ otherwise.*

[2] In the literature, matrix sensing is typically defined as a class of low-rank matrix recovery that satisfies the so-called *restricted isometry property* (RIP). It is well-known that Gaussian measurements satisfy RIP. In fact, our results can naturally be extended to measurements that satisfy RIP. However, we omit this extension here since it does not have any direct implication on our results.

Based on the above model, we define $Y \in \mathbb{R}^{d_1 \times d_2}$ as $Y_{xy} = y_i$ if $A_i(x,y) = 1$ for some i and (x,y), and $Y_{xy} = 0$ otherwise. Similarly, we define $E \in \mathbb{R}^{d_1 \times d_2}$ as $E_{xy} = \epsilon_i$ if $A_i(x,y) = 1$ for some i and (x,y), and $E_{xy} = 0$ otherwise. If the measurements follow matrix completion model, BM can be written as:

$$\min_{W_1 \in \mathbb{R}^{d_1 \times k}, W_2 \in \mathbb{R}^{k \times d_2}} f_{\ell_q}^{\mathrm{c}}(\mathbf{W}) := \frac{1}{m} \sum_{(x,y) \in \Psi} |Y_{xy} - (W_1 W_2)_{xy}|^q \,, \tag{MC-asym}$$

In symmetric matrix completion, the true solution $X^\star$ is additionally assumed to be positive semidefinte with dimension $d \times d$. Under this assumption, BM can be written as:

$$\min_{W \in \mathbb{R}^{d \times k}} f_{\ell_q}^{\mathrm{c}}(W) := \frac{1}{m} \sum_{(x,y) \in \Psi} \left| Y_{xy} - \left(WW^\top\right)_{xy} \right|^q . \tag{MC-sym}$$

Finally, we present our noise model.

Assumption 3 (Noise Model)**.** *Each measurement is independently corrupted with noise with a corruption probability $0 \le p \le 1$. Let the set of noisy measurements be denoted as $\mathcal{S}$. For each entry $i \in \mathcal{S}$, the value of ϵ_i is drawn from a distribution P_o. Moreover, a random variable ζ under the distribution P_o satisfies $\mathbb{P}(|\zeta| \ge t_0) \ge p_0$ for some constants $0 < t_0, p_0 \le 1$.*

Our considered noise model relies on very minimal assumptions and includes all the "typical" noise models, including Gaussian and outlier noise models. Intuitively, it only requires a nonzero mass at a nonzero value. For example, suppose that $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$. A basic anti-concentration inequality implies that $\mathbb{P}(|\epsilon_i| \ge t_0) = 1 - \mathbb{P}(|X| < t_0) \ge 1 - \sigma t_0$, which satisfies Assumption 3 with $(t_0, p_0) = \left((2\sigma)^{-1}, \sigma/2\right)$. Similarly, any sparse outlier noise is expected to satisfy Assumption 3.

A point $\mathbf{W}^\star = (W_1^\star, W_2^\star)$ is called a *true solution* if it satisfies $W_1^\star W_2^\star = X^\star$. Accordingly, $\mathcal{W}$ collects the set of all true solutions, i.e., $\mathcal{W} = \{(W_1, W_2) : W_1 W_2 = X^\star\}$. The set of true solutions for the symmetric case is defined as $\mathcal{W} = \{W : WW^\top = X^\star\}$.

3.1 Identifiability Conditions for Matrix Sensing

Our next two theorems provide necessary and sufficient conditions for the identifiability of the true solutions for asymmetric matrix sensing.

Theorem 1 (Unidentifiability of true solutions for asymmetric matrix sensing)**.** *Consider MS-asym with measurement matrices satisfying Assumption 1 and the noise satisfying Assumption 3 with corruption probability $0 < p \le 1$ and parameters $0 < t_0, p_0 \le 1$. Suppose that $m \lesssim \max\{d_1, d_2\}k$. With probability at least $1 - \exp(-\Omega(p_0 p m))$, the following statements hold:*

- *None of the true solutions in $\mathcal{W}$ are global solutions of $f_{\ell_q}^{\mathrm{s}}(\mathbf{W})$. More precisely, we have*

$$f_{\ell_q}^{\mathrm{s}}(\mathbf{W}^\star) - \min_{\mathbf{W}} f_{\ell_q}^{\mathrm{s}}(\mathbf{W}) \gtrsim t_0^q \cdot p p_0, \quad \text{for all } \mathbf{W}^\star \in \mathcal{W} \text{ and any } q \ge 1.$$

- *There exists at least one true solution $\mathbf{W}^\star \in \mathcal{W}$ that is not a critical point of $f_{\ell_q}^{\mathrm{s}}(\mathbf{W})$. More precisely, for any $0 < \gamma \lesssim \sqrt{pp_0} t_0$, we have*

$$\min_{\mathbf{W} \in \mathcal{B}_{\mathbf{W}^\star}(\gamma)} \left\{ f_{\ell_q}^{\mathrm{s}}(\mathbf{W}) - f_{\ell_q}^{\mathrm{s}}(\mathbf{W}^\star) \right\} \lesssim -t_0^{q-1} \cdot \sqrt{pp_0} \cdot \gamma.$$

Theorem 1 provides a necessary condition on the identifiability of the true solutions for the asymmetric matrix sensing: to ensure the global optimality of the true solutions in $\mathcal{W}$, the number of measurements must scale with the search rank k. In fact, Theorem 1 provides a stronger result: for any choice of ℓ_q-loss with $q \geq 1$ and arbitrarily small corruption probability $p > 0$, there exists at least one true solution that is not even a critical point of BM. As an immediate implication of this result, BM is unlikely to have identifiable true solutions with the same sample complexity as the convex relaxation. Our next theorem shows that the measurement size $m \gtrsim \max\{d_1, d_2\}k$ is in fact sufficient for the identifiability of the true solutions for BM with ℓ_1-loss (also known as *robust matrix recovery*).

Theorem 2 (Identifiability of the true solutions for asymmetric matrix sensing with ℓ_1-loss)**.** *Consider MS-asym with ℓ_1-loss and measurement matrices satisfying Assumption 1. Suppose that the noise satisfies Assumption 3 with a corruption probability $0 \leq p < 1/2$ and parameters $0 < t_0, p_0 \leq 1$. Moreover, suppose that $m \gtrsim \frac{\max\{d_1,d_2\}k}{(1-2p)^4}$. With probability at least $1 - \exp(-\Omega(\max\{d_1, d_2\}k))$, we have*

$$\mathbf{W}^\star \in \mathcal{W} \qquad \Longleftrightarrow \qquad \mathbf{W}^\star \in \arg\min_{\mathbf{W}} f^{\mathrm{s}}_{\ell_1}(\mathbf{W}).$$

Theorems 1 and 2 show that the landscape of MS-asym with ℓ_1-loss and sparse (but possibly heavy-tailed) noise undergoes a sharp transition as the measurement size exceeds beyond a threshold: if $m \lesssim \max\{d_1, d_2\}k$, none of the true solutions in $\mathcal{W}$ are global, and at least one of them is non-critical. As soon as $m \gtrsim \max\{d_1, d_2\}k$, all the true solutions coincide with the global solutions of MS-asym.

Next, we consider the symmetric matrix sensing, defined as MS-sym, where the true solution $X^\star \in \mathbb{R}^{d\times d}$ is assumed to be positive semidefinite. Our next theorem characterizes the local landscape of MS-sym around the true solutions.

Theorem 3 (Unidentifiability of true solutions for symmetric matrix sensing)**.** *Consider MS-sym with measurement matrices satisfying Assumption 1 and the noise satisfying Assumption 3 with a corruption probability $0 < p \leq 1$ and parameters $0 < t_0, p_0 \leq 1$. Suppose that $m \lesssim \frac{pp_0 d \min\{k-r, d/2\}}{4^q}$. With probability at least $1 - \exp(-\Omega(d(k-r)))$, the following statements hold:*

- *None of the true solutions in $\mathcal{W}$ are global minima of $f^{\mathrm{s}}_{\ell_q}(W)$. More precisely, we have*

$$f^{\mathrm{s}}_{\ell_q}(W^\star) - \min_{W} f^{\mathrm{s}}_{\ell_q}(W) \gtrsim \frac{(t_0/2)^q}{(d(k-r))^{3/2}}, \quad \text{for all } W^\star \in \mathcal{W} \text{ and any } q \geq 1.$$

- *None of the true solutions in $\mathcal{W}$ are local minima of $f^{\mathrm{s}}_{\ell_q}(W)$. More precisely, for any $W^\star \in \mathcal{W}$ and $0 < \gamma^2 \lesssim \frac{t_0}{q2^q(d(k-r))^{3/2}}$, we have*

$$\max_{W^\star \in \mathcal{W}} \min_{W \in \mathcal{B}_{W^\star}(\gamma)} \left\{ f^{\mathrm{s}}_{\ell_q}(W) - f^{\mathrm{s}}_{\ell_q}(W^\star) \right\} \lesssim -q t_0^{q-1} \gamma^2.$$

Suppose that $2r \leq k \leq d/2$ and p, p_0, q are fixed. Then, Theorem 3 shows that none of the true solutions are global or local solutions so long as $m \lesssim dk$. However, unlike the asymmetric case, we do not show the existence of a non-critical true solution. Our next result complements this observation by showing that, despite their sub-optimality, all true solutions emerge as critical points of BM even if m *does not* scale with the search rank k.

Theorem 4 (Identifiability of true solutions for symmetric matrix sensing with ℓ_1-loss)**.** *Consider MS-sym with ℓ_1-loss and measurement matrices satisfying Assumption 1. Suppose that the noise satisfies Assumption 3 with a corruption probablity $0 \le p < 1/2$ and parameters $0 < t_0, p_0 \le 1$. The following statements hold:*

- *Suppose that $m \gtrsim \frac{dr}{(1-2p)^4}$. With probability at least $1 - \exp(\Omega(dr))$, all true solutions in $\mathcal{W}$ are critical. More precisely, we have*

$$\min_{W^\star \in \mathcal{W}} \min_{W \in \mathcal{B}_{W^\star}(\gamma)} \left\{ f_{\ell_1}^{\mathrm{s}}(W) - f_{\ell_1}^{\mathrm{s}}(W^\star) \right\} \gtrsim -\left(\sqrt{\frac{2}{\pi}} + \sqrt{\frac{dk}{m}} \right) \gamma^2, \qquad \text{for any } \gamma \ge 0.$$

- *Suppose that $m \gtrsim \frac{dk}{(1-2p)^4}$. With probability at least $1 - \exp(\Omega(dk))$, we have*

$$W^\star \in \mathcal{W} \qquad \Longleftrightarrow \qquad W^\star \in \arg\min_{W} f_{\ell_1}^{\mathrm{s}}(W).$$

The above theorem shows that the symmetric matrix sensing has a flatter landscape around its true solution compared to the asymmetric case. In particular, when $dr \lesssim m \lesssim dk$, all the true solutions become critical points and the steepest descent direction can reduce the loss by at most $\mathcal{O}(\gamma^2)$ within a γ-neighborhood of any true solution. This is an order of magnitude smaller than the $\mathcal{O}(\gamma)$ reduction in the asymmetric case.

3.2 Identifiability Conditions for Matrix Completion

Next, we study the local landscape of symmetric and asymmetric matrix completion around its true solutions.

Theorem 5 (Unidentifiability of true solutions for asymmetric matrix completion)**.** *Consider MC-asym with measurement matrices satisfying 2 with sampling probability $0 < s \le 1$. Suppose that the noise satisfies Assumption 3 with corruption probability $0 \le p \le 1$ and parameters $0 < t_0, p_0 \le 1$. Moreover, suppose that $k > r$. With probability at least $1 - \exp(-\Omega(spp_0 \max\{d_1, d_2\}(k-r)))$, the following statements hold:*

- *None of the true solutions in $\mathcal{W}$ are global solutions of $f_{\ell_q}^{\mathrm{s}}(\mathbf{W})$. More precisely, we have*

$$f_{\ell_q}^{\mathrm{c}}(\mathbf{W}^\star) - \min_{\mathbf{W}} f_{\ell_q}^{\mathrm{c}}(\mathbf{W}) \gtrsim t_0^q \cdot \sqrt{\frac{k-r}{d_2}} \cdot \sqrt{\frac{pp_0}{sd_1d_2}}, \quad \text{for all } \mathbf{W}^\star \in \mathcal{W} \text{ and any } q \ge 1.$$

- *There exists at least one true solution $\mathbf{W}^\star \in \mathcal{W}$ that is not a critical point of $f_{\ell_q}^{\mathrm{c}}(\mathbf{W})$. More precisely, for any $0 < \gamma \le t_0$, we have*

$$\min_{\mathbf{W} \in \mathcal{B}_{\mathbf{W}^\star}(\gamma)} \left\{ f_{\ell_q}^{\mathrm{s}}(\mathbf{W}) - f_{\ell_q}^{\mathrm{s}}(\mathbf{W}^\star) \right\} \lesssim -t_0^{q-1} \cdot \sqrt{\frac{k-r}{d_2}} \cdot \sqrt{\frac{pp_0}{sd_1d_2}} \cdot \gamma.$$

Theorem 5 shows that the true solutions are unlikely to correspond to the global solutions of MC-asym, so long as the measurements are corrupted with noise and the search rank is strictly greater than the true rank. This result holds even in a near-ideal scenario, where the sampling probability is one (i.e., the entire matrix $X^\star$ can be observed) and an arbitrarily small fraction of the observed entries is corrupted with noise. Finally, we consider the symmetric matrix completion.

Theorem 6 (Unidentifiability of true solutions for symmetric matrix completion)**.** *Consider* MC-sym *with measurement matrices satisfying Assumption 2 with sampling probability* $0 < s \leq 1$*. Suppose that* $k > r$ *and there exists* $(k, k) \in \Psi$ *such that* $E_{kk} \geq t_0$*. Then, the following statements hold:*

- *None of the true solutions in* $\mathcal{W}$ *are global solutions of* $f^{\mathrm{c}}_{\ell_q}(\mathbf{W})$*. More precisely, we have*

$$f^{\mathrm{c}}_{\ell_q}(W^\star) - \min_{W} f^{\mathrm{c}}_{\ell_q}(W) \geq t_0^q, \qquad \text{for all } W^\star \in \mathcal{W} \text{ and any } q \geq 1.$$

- *None of the true solutions in* $\mathcal{W}$ *are local solutions of* $f^{\mathrm{c}}_{\ell_q}(W)$*. More precisely, for any* $W^\star \in \mathcal{W}$ *and* $0 < \gamma \lesssim \sqrt{t_0}$*, we have*

$$\max_{W^\star \in \mathcal{W}} \min_{W \in \mathcal{B}_{W^\star}(\gamma)} \left\{ f^{\mathrm{c}}_{\ell_q}(W) - f^{\mathrm{c}}_{\ell_q}(W^\star) \right\} \leq -t_0^{q-1}\gamma^2.$$

Unlike our previous results, the necessary condition for the identifiability of the true solutions for symmetric matrix completion is purely deterministic: none of the true solutions are local or global so long as $k > r$ and at least one of the diagonal entries of $X^\star$ is observed with a positive noise. The latter condition is indeed very mild and holds with a high probability provided that the noise takes a positive value with a nonzero probability. We also note that, when the true rank is known and equal to one, the true solutions of MC-sym correspond to the global minima even if up to a constant fraction of the measurements are grossly corrupted with noise [FS20, Theorem 6]. Therefore, the over-estimation of the rank is indeed crucial in Theorem 6 and cannot be relaxed.

The rest of the paper is organized as follows: in Section 4, we prove the unidentifiability results for the asymmetric case (Theorems 1 and 5). In Section 5, we extend our analysis to the symmetric settings (Theorems 3 and 6). Finally, in Section 6, we present our proofs for the ℓ_1-loss (Theorems 2 and 4). Throughout our arguments, will make extensive use of basic concentration results on random variables and processes. To streamline the presentation, we defer these results to Appendix A.

4 Asymmetric Case: Non-criticality via Structured Perturbations

At the core of our results lies a class of structured perturbations that can be used to show the existence of a non-critical true solution for both asymmetric matrix sensing and matrix completion. To this goal, we first present the following crucial lemma which provides a sufficient condition for the non-criticality of a point $\bar{\mathbf{W}}$.

Lemma 2. *Given a point* $\bar{\mathbf{W}}$ *and any* $q \geq 1$*, suppose that there exists a direction* $\Delta\mathbf{W}$ *such that* $f'_{\ell_q}(\bar{\mathbf{W}}, \Delta\mathbf{W}) < 0$*. Then,* $\bar{\mathbf{W}}$ *is not a critical point of* $f_{\ell_q}(\mathbf{W})$*.*

Proof. For any $q > 1$, the function $f_{\ell_q}(\mathbf{W})$ is differentiable, and hence, its Clarke subdifferential coincides with its gradient [Cla75, Proposition 1.13]. On the other hand, $f'_{\ell_q}(\bar{\mathbf{W}}, \Delta\mathbf{W}) < 0$ implies that the gradient of $f_{\ell_q}(\mathbf{W})$ at $\bar{\mathbf{W}}$ is negative, which in turn implies that $\bar{\mathbf{W}}$ is non-critical. Now, consider $q = 1$. Define $\mathcal{M}$ as the class of binary functions $\sigma : \{1, \ldots, m\} \to \{-1, +1\}$. It is easy to see that $f_{\ell_q}(\mathbf{W}) = \max_{\sigma \in \mathcal{M}} g_\sigma(\mathbf{W})$, where $g_\sigma(\mathbf{W}) = \sum_{i \in m} \sigma(i) \left(y_i - \langle A_i, W_1 W_2 \rangle\right)$. Note that $g_\sigma(\mathbf{W})$ is smooth for any choice of $\sigma \in \mathcal{M}$. Therefore, the third property of Lemma 1 can be invoked to complete the proof. □

The above lemma implies that, in order to show the non-criticality of a point $\mathbf{W}$, it suffices to obtain a perturbation $\Delta\mathbf{W} = (\Delta W_1, \Delta W_2)$ such that $f'_{\ell_q}(\mathbf{W}, \Delta\mathbf{W}) < 0$. In fact, we will show a stronger result: with high probability and for any $\gamma \le \gamma_0$, there exists $\Delta\mathbf{W} \in \mathcal{B}_F(\gamma)$ such that $f_{\ell_q}(\mathbf{W} + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W}) = -\Omega(\gamma)$. As will be shown later, this result automatically implies the non-criticality of $\mathbf{W}$. Recall that $\mathcal{S}$ is the index set of the noisy measurements, and define $\bar{\mathcal{S}} = [m]\backslash\mathcal{S}$ as the set of *clean* measurements. Moreover, define $\mathcal{S}_t = \{i : |\epsilon_i| \ge t_0\}$. Given any $\mathbf{W}^\star \in \mathcal{W}$, one can write

$$\begin{aligned}
f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W}^\star) =& \frac{1}{m}\sum_{i\in\bar{\mathcal{S}}} |\langle A_i, (W_1^\star + \Delta W_1)(W_2^\star + \Delta W_2) - W_1^\star W_2^\star\rangle|^q \\
&+ \frac{1}{m}\sum_{i\in\mathcal{S}} \left(|\langle A_i, (W_1^\star + \Delta W_1)(W_2^\star + \Delta W_2) - W_1^\star W_2^\star\rangle - \epsilon_i|^q - |\epsilon_i|^q\right) \\
=& \frac{1}{m}\sum_{i\in\bar{\mathcal{S}}} |\langle A_i, \Delta W_1 W_2^\star + W_1^\star \Delta W_2 + \Delta W_1 \Delta W_2\rangle|^q \\
&+ \frac{1}{m}\sum_{i\in\mathcal{S}} \left(|\langle A_i, \Delta W_1 W_2^\star + W_1^\star \Delta W_2 + \Delta W_1 \Delta W_2\rangle - \epsilon_i|^q - |\epsilon_i|^q\right).
\end{aligned}$$

To analyze the infimum of the loss difference over $\Delta\mathbf{W} \in \mathcal{B}_F(\gamma)$, we consider the following set of *structured perturbations* at $\mathbf{W}^\star \in \mathcal{W}$:

$$\begin{aligned}
\mathcal{U}_{\mathbf{W}^\star}(\gamma) := \big\{(U_1, 0_{k\times d_2}) :\ & \|U_1\|_F \le \gamma, \\
& \langle A_i, U_1 W_2^\star\rangle = 0, \qquad \forall i \in \bar{\mathcal{S}}, \\
& \langle A_i, U_1 W_2^\star\rangle \epsilon_i \ge 0, \qquad \forall i \in \mathcal{S}, \\
& |\langle A_i, U_1 W_2^\star\rangle| \le |\epsilon_i|, \quad \forall i \in \mathcal{S}\big\}.
\end{aligned} \tag{4}$$

When there is no ambiguity, we drop the subscript $\mathbf{W}^\star$ from $\mathcal{U}_{\mathbf{W}^\star}(\gamma)$. Evidently, $\mathcal{U}(\gamma)$ is non-empty since $(0_{d_1\times k}, 0_{k\times d_2}) \in \mathcal{U}(\gamma)$, and we have $\mathcal{U}(\gamma) \subseteq \mathcal{B}_F(\gamma)$. The following lemma provides a more tractable upper bound on $\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \big\{f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W})\big\}$ when the search space is restricted to $\mathcal{U}(\gamma)$.

Lemma 3. *We have*

$$\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \big\{f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W})\big\} \le -\frac{q t_0^{q-1}}{q+1} \cdot \max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle|\right\}. \tag{5}$$

Proof. One can write

$$\begin{aligned}
\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \left\{f_{\ell_q}(\mathbf{W}^\star+\Delta\mathbf{W})-f_{\ell_q}(\mathbf{W})\right\} &\le \min_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{f_{\ell_q}(\mathbf{W}^\star+\Delta\mathbf{W})-f_{\ell_q}(\mathbf{W})\right\} \\
&\overset{(a)}{\le} -\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}} |\epsilon_i|^q - |\langle A_i, \Delta W_1 W_2^\star\rangle - \epsilon_i|^q\right\} \\
&\overset{(b)}{\le} -\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\epsilon_i|^q - (|\epsilon_i| - |\langle A_i, \Delta W_1 W_2^\star\rangle|)^q\right\} \\
&\overset{(c)}{\le} -\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\cdot\frac{q}{q+1}\sum_{i\in\mathcal{S}_t} |\epsilon_i|^{q-1}|\langle A_i, \Delta W_1 W_2^\star\rangle|\right\} \\
&\overset{(d)}{\le} -\frac{q t_0^{q-1}}{q+1}\cdot\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle|\right\},
\end{aligned}$$

where (a) and (b) follow from the definition of $\mathcal{U}(\gamma)$, (c) is a direct application of a basic inequality for polynomials (see Lemma 26 in Appendix C), and (d) follows from $|\epsilon_i| \ge t_0$ for every $i \in \mathcal{S}_t$. □

As will be shown next, controlling the right-hand side of (4) is much more tractable. Consider the singular value decomposition of $X^\star = V_1 \Sigma V_2^\top$, where $V_1 \in \mathbb{R}^{d_1\times r}$ and $V_2 \in \mathbb{R}^{d_2\times r}$ are orthonormal matrices and $\Sigma \in \mathbb{R}^{r\times r}$ is a diagonal matrix collecting the nonzero singular values of $X^\star$. Let the solution $\mathbf{W}^\star = (W_1^\star, W_2^\star)$ be defined as:

$$W_1^\star = \begin{bmatrix} V_1\Sigma & 0_{d_1\times(k-r)}\end{bmatrix}, \quad W_2^\star = \begin{bmatrix} V_2 & Z\end{bmatrix}^\top, \quad \text{for some } Z\in\mathbb{R}^{d_2\times(k-r)}. \tag{6}$$

We have $W_1^\star W_2^\star = V_1\Sigma V_2^\top = X^\star$ for any arbitrary $Z \in \mathbb{R}^{(k-r)\times d_2}$, and hence, $\mathbf{W}^\star \in \mathcal{W}$. To streamline the presentation, we assume without loss of generality that $d_1 \ge d_2$.[3] We show that, depending on the measurement matrices, the matrix Z can be designed such that the right-hand side of (5) can be upper bounded by $-\Omega(\gamma)$. Our next lemma achieves this goal for matrix completion.

Lemma 4. *Suppose that the measurement matrices follow the matrix completion model in Assumption 2. Consider the true solution $\mathbf{W}^\star$ defined as (6) with $Z = I_{d_2\times(k-r)}$. For any $\gamma \le t_0$, we have*

$$\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle|\right\} \ge \frac{1}{3}\cdot\sqrt{\frac{k-r}{d_2}}\cdot\sqrt{\frac{pp_0}{sd_1d_2}}\cdot\gamma,$$

with probability at least $1-\exp(-\Omega(spp_0d_1(k-r)))$.

Proof. We prove this lemma by designing an explicit point $\bar{\mathbf{U}} \in \mathcal{U}(\gamma)$ that attains the aforementioned

[3] If $d_1 < d_2$, $\mathbf{W}^\star$ can be defined as follows without affecting our subsequent analysis:

$$W_1^\star = \begin{bmatrix} V_1 & Z\end{bmatrix}, \quad W_2^\star = \begin{bmatrix} V_2\Sigma & 0_{d_2\times(k-r)}\end{bmatrix}^\top, \quad \text{for some } Z\in\mathbb{R}^{d_1\times(k-r)}.$$

lower bound. Let $\bar{\Psi} = \{(x,y) \in \Psi : y \leq k-r, |E_{xy}| \geq t_0\}$ and consider the point

$$\bar{\mathbf{U}} = (\bar{U}_1, 0_{k\times d_2}), \text{where } \bar{U}_1 = \begin{bmatrix} 0_{d_1\times r} & \bar{U}_{12} \end{bmatrix}, \bar{U}_{12} \in \mathbb{R}^{d_1\times(k-r)}, \bar{U}_{12}(x,y) = \begin{cases} \dfrac{\text{Sign}(E_{xy})}{\sqrt{|\bar{\Psi}|}} & \text{if } (x,y) \in \bar{\Psi} \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

With this definition and the choice of Z, we have $\bar{U}_1 W_2^\star = \begin{bmatrix} \bar{U}_{12} & 0_{d_1\times(d_2-k+r)} \end{bmatrix}$. Moreover, simple calculation reveals that $\gamma\bar{\mathbf{U}} \in \mathcal{U}(\gamma)$. Therefore, we have

$$\max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{ \frac{1}{m} \sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle| \right\} \geq \frac{\gamma}{m} \sum_{i\in\mathcal{S}_t} |\langle A_i, \bar{U}_1 W_2^\star\rangle| = \frac{\gamma}{m} \sum_{(x,y)\in\bar{\Psi}} |\bar{U}_{12}(x,y)| = \frac{\sqrt{|\bar{\Psi}|}}{m} \cdot \gamma. \tag{8}$$

To finish the proof, we provide lower and upper bounds for $|\bar{\Psi}|$ and m, respectively. For any $1 \leq x \leq d_1, 1 \leq y \leq d_2$, we have $(x,y) \in \Psi$ with probability s. Therefore, $m = |\Psi|$ has a binomial distribution with parameters $(d_1 d_2, s)$. Therefore, according to Lemma 23 in Appendix A.3, we have $m \leq (3/2)d_1 d_2 s$ with probability at least $1 - \exp(-\Omega(d_1 d_2 s))$. On the other hand, conditioned on $(x,y) \in \Psi$, we have $(x,y) \in \bar{\Psi}$ with probability $pp_0(k-r)/d_2$. Therefore, for any $1 \leq x \leq d_1, 1 \leq y \leq d_2$, we have $(x,y) \in \bar{\Psi}$ with probability $spp_0(k-r)/d_2$, which in turn implies that $|\bar{\Psi}|$ also has a binomial distribution with parameters $(d_1 d_2, spp_0(k-r)/d_2)$. Again, Lemma 23 in Appendix A.3 can be invoked to show that $|\bar{\Psi}| \geq spp_0 d_1(k-r)/4$ with probability at least $1 - \exp(-\Omega(spp_0 d_1(k-r)))$. Therefore, a simple union bound implies that

$$\frac{\sqrt{|\bar{\Phi}|}}{m} \geq \frac{1}{3}\sqrt{\frac{k-r}{d_2}}\sqrt{\frac{pp_0}{sd_1d_2}} \implies \max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{ \frac{1}{m} \sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle| \right\} \geq \frac{1}{3}\sqrt{\frac{k-r}{d_2}}\sqrt{\frac{pp_0}{sd_1d_2}} \cdot \gamma$$

with probability at least $1 - \exp(-\Omega(spp_0 d_1(k-r)))$. This completes the proof. □

Now we are ready to provide the proof of Theorem 5.

Proof of Theorem 5. Consider the true solution $\mathbf{W}^\star$ defined as (6) with $Z = I_{d_2\times(k-r)}$. Combining Lemmas 3 and 4, we have

$$\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \left\{ f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}^\star) - f_{\ell_q}(\mathbf{W}^\star) \right\} \leq -\frac{qt_0^{q-1}}{3(q+1)} \cdot \sqrt{\frac{k-r}{d_2}} \cdot \sqrt{\frac{pp_0}{sd_1d_2}} \cdot \gamma, \tag{9}$$

for any $\gamma \leq t_0$, with probability at least $1 - \exp(\Omega(spp_0 d_1(k-r)))$. In particular, the specific choice of $\bar{\mathbf{U}}$ in (7) is indeed a descent direction, i.e, it satisfies $f'_{\ell_q}(\mathbf{W}^\star, \bar{\mathbf{U}}) < 0$. Therefore, $\mathbf{W}^\star$ is not a critical point in light of Lemma 2. This completes the proof of the second statement. Now, let $\widehat{\mathbf{W}} = \mathbf{W}^\star + \Delta\mathbf{W}^\star$, where $\Delta\mathbf{W}^\star$ is the minimizer of the left-hand side of (9) with $\gamma = t_0$. We have

$$f_{\ell_q}(\mathbf{W}^\star) - \min_{\mathbf{W}}\{f_{\ell_q}(\mathbf{W})\} \geq f_{\ell_q}(\mathbf{W}^\star) - f_{\ell_q}(\widehat{\mathbf{W}}) \geq \frac{qt_0^q}{3(q+1)} \cdot \sqrt{\frac{k-r}{d_2}} \cdot \sqrt{\frac{pp_0}{sd_1d_2}},$$

thereby completing the proof of the first statement. □

Next, we extend our analysis to the asymmetric matrix sensing. Recall that, for asymmetric matrix completion, we provided an upper bound on $\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \left\{ f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W}) \right\}$ by

choosing an explicit point in $\mathcal{U}(\gamma)$ that lead to the desirable bound (Lemma 4). However, when the measurement matrices do not follow the matrix completion model, the explicit solution defined in (7) may no longer belong to $\mathcal{U}(\gamma)$. Under such circumstances, it may be non-trivial to explicitly construct a solution in $\mathcal{U}(\gamma)$. To address this issue, we consider a further refinement of $\mathcal{U}_{\mathbf{W}^\star}(\gamma)$:

$$\begin{aligned}\mathcal{V}_{\mathbf{W}^\star}(\gamma,\zeta) := \big\{(U_1,U_2) : U_2 = 0_{k\times d_2},\ \|U_1\|_F \le \gamma,& \\ \langle A_i, U_1 W_2^\star\rangle = 0, \qquad & \forall i \in \bar{\mathcal{S}} \cup (\mathcal{S}\backslash\mathcal{S}_t), \\ \langle A_i, U_1 W_2^\star\rangle = \zeta \mathrm{Sign}(\epsilon_i), \quad & \forall i \in \mathcal{S}_t\big\}.\end{aligned} \tag{10}$$

When there is no ambiguity, we drop the subscript $\mathbf{W}^\star$ from $\mathcal{V}_{\mathbf{W}^\star}(\gamma,\zeta)$. Evidently, we have $\mathcal{V}(\gamma,\zeta) \subseteq \mathcal{U}(\gamma) \subseteq \mathcal{B}_F(\gamma)$ for every $\zeta \le t_0$. Therefore, assuming that $\mathcal{V}(\gamma,\zeta)$ is non-empty, we have

$$\begin{aligned}\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)} \big\{f_{\ell_q}(\mathbf{W}^\star+\Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W})\big\} &\le -\frac{qt_0^{q-1}}{q+1}\cdot \max_{\Delta\mathbf{W}\in\mathcal{U}(\gamma)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle|\right\} \\ &\le -\frac{qt_0^{q-1}}{q+1}\cdot \max_{\Delta\mathbf{W}\in\mathcal{V}(\gamma,\zeta)} \left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} |\langle A_i, \Delta W_1 W_2^\star\rangle|\right\} \\ &\le -\frac{qt_0^{q-1}}{q+1}\cdot \frac{|\mathcal{S}_t|}{m}\cdot\zeta,\end{aligned} \tag{11}$$

where the last inequality follows from the definition of $\mathcal{V}(\gamma,\zeta)$. Therefore, for an appropriate choice of ζ, the above inequality achieves the desired result. However, unlike $\mathcal{U}(\gamma)$, the set $\zeta(\gamma,\zeta)$ is not guaranteed to be non-empty. To see this, note that the norm of any solution for the system of linear equations in (10) increases with ζ; as a result, the condition $\|U_1\|_F \le \gamma$ would be violated for sufficiently large ζ. The following lemma provides sufficient conditions under which $\mathcal{V}(\gamma,\zeta)$ is non-empty.

Lemma 5. *The set $\mathcal{V}(\gamma,\zeta)$ is non-empty if the following conditions are satisfied:*

(C1) The matrix $\mathbf{A} = \left[\mathrm{vec}\left(A_1 W_2^{\star\top}\right) \quad \dots \quad \mathrm{vec}\left(A_m W_2^{\star\top}\right)\right]^\top$ *is full row-rank.*

(C2) We have $\zeta \le \left(\sqrt{|\mathcal{S}_t|}\left\|\mathbf{A}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\|\right)^{-1}\gamma$.

Proof. For every $1 \le i \le m$, we have $\langle A_i, U_1 W_2^\star\rangle = \langle A_i W_2^{\star\top}, U_1\rangle$. Therefore, to show the non-emptiness of $\mathcal{V}(\gamma,\zeta)$, it suffices to show that the following system of linear equations has a solution U_1 that satisfies $\|U_1\|_F \le \gamma$:

$$\begin{cases}\left\langle A_i W_2^{\star\top}, U_1\right\rangle = 0, & \forall i \in \bar{\mathcal{S}} \cup (\mathcal{S}\backslash\mathcal{S}_t), \\ \left\langle A_i W_2^{\star\top}, U_1\right\rangle = \zeta\mathrm{Sign}(\epsilon_i), & \forall i \in \mathcal{S}_t.\end{cases} \tag{12}$$

The above system of linear equations can be written as

$$\mathbf{A}u = b, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix}\mathrm{vec}\left(A_1 W_2^{\star\top}\right)^\top \\ \vdots \\ \mathrm{vec}\left(A_m W_2^{\star\top}\right)^\top\end{bmatrix},\ u = \mathrm{vec}(U_1),\ b_i = \begin{cases}\zeta\mathrm{Sign}(\epsilon_i) & \text{if } i \in \mathcal{S}_t \\ 0 & \text{if } i \notin \mathcal{S}_t\end{cases}. \tag{13}$$

Since $\mathbf{A}$ is assumed to be full row-rank, the above set of equations is guaranteed to have a solution. To show the existence of a solution U_1 that satisfies $\|U_1\|_F \leq \gamma$, we consider the minimum norm solution of $\mathbf{A}u = b$, which is equal to $u^\star = \mathbf{A}^\dagger b$, where $\mathbf{A}^\dagger$ is the pseudo-inverse of $\mathbf{A}$ defined as $\mathbf{A}^\dagger = \mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1}$. This in turn implies that $\|u^\star\| \leq \left\|\mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\| \|b\| \leq \sqrt{|\mathcal{S}_t|}\left\|\mathbf{A}^\top \left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\| \zeta \leq \gamma$, where the last inequality follows from Condition C2. This completes the proof. □

Next, we apply the above lemma to asymmetric matrix sensing and show that $\mathcal{V}(\gamma, \zeta)$ is non-empty with $\zeta = \mathcal{O}(\gamma)$, provided that the number of measurements is not too large.

Lemma 6. *Suppose that the measurement matrices satisfy Assumption 1. Consider the true solution $\mathbf{W}^\star$ in (6), where $Z = V_2' \in \mathbb{R}^{(k-r)\times d_2}$ is chosen such that $W_2^\star = \begin{bmatrix} V_2 & V_2' \end{bmatrix}^\top \in \mathbb{R}^{k\times d_2}$ has orthonormal rows. Suppose that $m \leq (1/16)d_1 k$ and $\zeta \leq (1/3)\sqrt{1/(pp_0)} \cdot \gamma$. With probability at least $1 - \exp(-\Omega(mpp_0))$, the set $\mathcal{V}(\gamma, \zeta)$ is non-empty.*

Proof. We will show that the conditions of Lemma 5 hold with high probability. Since $W_2^\star$ is orthonormal, the elements of $\left\{A_i W_2^{\star\top}\right\}_{i=1}^m$ are i.i.d. and have standard normal distribution. As a result, $\mathbf{A}$ is full row-rank almost surely and Condition C1 of Lemma 5 is satisfied. On the other hand, due to a standard concentration bound on Gaussian matrices (see Lemma 22 in Appendix A.3), we have $\left\|(d_1 k)^{-1/2}\mathbf{A}\right\| \leq 4/3$ and $\left\|(d_1 k)\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\| \leq 6$ with probability at least $1 - \exp(-\Omega(d_1 k))$ provided that $m \leq d_1 k/16$. Moreover, since $|\mathcal{S}_t|$ has a binomial distribution with parameters (m, pp_0), we have $|\mathcal{S}_t| \leq (3/2)mpp_0$ with probability at least $1 - \exp(-\Omega(mpp_0))$. This implies that

$$\sqrt{|\mathcal{S}_t|}\left\|\mathbf{A}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\| \leq \sqrt{\frac{|\mathcal{S}_t|}{d_1 k}}\left\|(d_1 k)^{-1/2}\mathbf{A}\right\|\left\|(d_1 k)\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\| \leq 8\sqrt{\frac{(3/2)mpp_0}{d_1 k}} \leq 3\sqrt{pp_0},$$

with probability at least $1 - \exp(-\Omega(d_1 k)) - \exp(-\Omega(mpp_0)) = 1 - \exp(-\Omega(mpp_0))$. This yields

$$\left(\sqrt{|\mathcal{S}_t|}\left\|\mathbf{A}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\right\|\right)^{-1}\gamma \geq \frac{1}{3}\sqrt{\frac{1}{pp_0}} \cdot \gamma \geq \zeta,$$

where the second inequality follows from the assumed upper bound on ζ. Therefore, Condition C2 of Lemma 5 is also satisfied, implying that the set $\mathcal{V}(\gamma, \zeta)$ is non-empty. □

Based on the above lemma, we present the proof of Theorem 1.

Proof of Theorem 1. Consider the true solution $\mathbf{W}^\star$ in (6), where $Z = V_2' \in \mathbb{R}^{(k-r)\times d_2}$ is chosen such that $W_2^\star = \begin{bmatrix} V_2 & V_2' \end{bmatrix}^\top \in \mathbb{R}^{k\times d_2}$ has orthonormal rows. For this specific choice of $\mathbf{W}^\star$, Lemma 6 implies that $\mathcal{V}(\gamma, \zeta)$ is non-empty with probability at least $1 - \exp(-\Omega(mpp_0))$, provided that $m \leq (1/16)d_1 k$ and $\zeta = (1/3)\sqrt{1/(pp_0)} \cdot \gamma$. On the other hand, it is assumed that $\gamma \leq 3\sqrt{pp_0}t_0$, which in turn leads to $\zeta \leq t_0$ and $\mathcal{V}(\gamma, \zeta) \subseteq \mathcal{U}(\gamma) \subseteq B_F(\gamma)$. Therefore, (11) can be invoked to show that with probability at least $1 - \exp(-\Omega(mpp_0))$

$$\min_{\Delta\mathbf{W}\in\mathcal{B}_F(\gamma)}\left\{f_{\ell_q}(\mathbf{W}^\star + \Delta\mathbf{W}) - f_{\ell_q}(\mathbf{W}^\star)\right\} \leq -\frac{q t_0^{q-1}}{q+1} \cdot \frac{|\mathcal{S}_t|}{m} \cdot \zeta \leq -\frac{q t_0^{q-1}}{6(q+1)} \cdot \sqrt{pp_0} \cdot \gamma, \tag{14}$$

where in the last inequality we used $\zeta = (1/3)\sqrt{1/(pp_0)} \cdot \gamma$ and the fact that $|\mathcal{S}_t| \geq (1/2)mpp_0$ with probability at least $1 - \exp(-\Omega(mpp_0))$. As an immediate implication, any point $\mathbf{U} \in \mathcal{V}(\gamma, \zeta)$ is a descent direction with $f'_{\ell_q}(\mathbf{W}^\star, \mathbf{U}) < 0$. Therefore, $\mathbf{W}^\star$ is not a critical point in light of Lemma 2. This completes the proof of the second statement. Now, let $\widehat{\mathbf{W}} = \mathbf{W}^\star + \Delta\mathbf{W}^\star$, where $\Delta\mathbf{W}^\star$ is the minimizer of the left-hand side of (14) with $\gamma = 3\sqrt{pp_0}t$. We have

$$f_{\ell_q}(\mathbf{W}^\star) - \min_{\mathbf{W}}\{f_{\ell_q}(\mathbf{W})\} \geq f_{\ell_q}(\mathbf{W}^\star) - f_{\ell_q}(\widehat{\mathbf{W}}) \geq \frac{qt_0^q}{2(q+1)} \cdot pp_0 \cdot \gamma, \tag{15}$$

where in the last inequality, we used (14) with $\gamma = 3\sqrt{pp_0}t$. This completes the proof of the first statement. □

5 Symmetric Case: Sub-optimality via Second-order Perturbations

Our next goal is to characterize the landscape of the symmetric matrix sensing and completion around its true solutions. To this goal, we first note that the set of structured perturbations (4) is no longer feasible for the symmetric setting. To address this issue, we instead rely on a class of *second-order perturbations*. Let $X^\star = V\Sigma V^\top$ be the eigen-decomposition of $X^\star$, where $V \in \mathbb{R}^{d\times r}$ is an orthonormal matrix and $\Sigma \in \mathbb{R}^{r\times r}$ is a diagonal matrix collecting the nonzero eigenvalues of $X^\star$. Moreover, let $\mathcal{O}_{r\times k} := \{R \in \mathbb{R}^{r\times k} : RR^\top = I_r\}$ (note that $\mathcal{O}_{r\times k}$ is non-empty since $k \geq r$). The following lemma provides another characterization of the set $\mathcal{W} = \{W \in \mathbb{R}^{d\times k} : WW^\top = X^\star\}$.

Lemma 7. *We have*

$$W \in \mathcal{W} \qquad \Longleftrightarrow \qquad W = V\Sigma^{1/2}R \text{ for some } R \in \mathcal{O}_{r\times k}.$$

Proof. Indeed, if $W = V\Sigma^{1/2}R$ for some $R \in \mathcal{O}_{r\times k}$, then $WW^\top = X^\star$ and $W \in \mathcal{W}$. Now, suppose that $W \in \mathcal{W}$. We have

$$\begin{aligned} WW^\top = V\Sigma V^\top &\Longrightarrow \Sigma^{-1/2}V^\top WW^\top V\Sigma^{-1/2} = I_r \\ &\Longrightarrow \Sigma^{-1/2}V^\top W = R \text{ for some } R \in \mathcal{O}_{r\times k} \\ &\Longrightarrow VV^\top W = V\Sigma^{1/2}R \text{ for some } R \in \mathcal{O}_{r\times k} \end{aligned}$$

Moreover, let $V^\perp$ be the orthogonal complement of V. We have

$$WW^\top = V\Sigma V^\top \Longrightarrow {V^\perp}^\top WW^\top V^\perp = 0_r \Longrightarrow {V^\perp}^\top W = 0 \Longrightarrow V^\perp {V^\perp}^\top W = 0$$

Combining the above two equalities, we have

$$WW^\top = VV^\top W + V^\perp {V^\perp}^\top W = V\Sigma^{1/2}R \text{ for some } R \in \mathcal{O}_{r\times k}.$$

This completes the proof. □

Based on the above lemma, the set of true solutions can be characterized as $\mathcal{W} = \{V\Sigma^{1/2}R : R \in \mathcal{O}_{r\times k}\}$. This new characterization of $\mathcal{W}$ will be useful in our subsequent analysis. Similar to the asymmetric setting, our goal is to provide an upper bound for $\min_{\Delta W \in \mathcal{B}_F(\gamma)} \left\{f_{\ell_q}(W^\star + \Delta W) - f_{\ell_q}(W^\star)\right\}$. We consider the following set of *second-order perturbations*:

$$\widehat{\mathcal{U}}_{W^\star}(\gamma) := \left\{UR' : U \in \mathbb{R}^{d\times(k-r)}, \|U\|_F \leq \gamma, R' \in \mathcal{O}_{(k-r)\times k}, W^\star {R'}^\top = 0\right\}. \tag{16}$$

Note that any perturbation $\Delta W = UR' \in \widehat{\mathcal{U}}_{W^\star}(\gamma)$ is indeed second-order since:

$$\left\|(W^\star + \Delta W)(W^\star + \Delta W)^\top - W^\star W^{\star\top}\right\|_F = \left\|UU^\top\right\|_F \leq \gamma^2. \tag{17}$$

Evidently, we have $\widehat{\mathcal{U}}_{W^\star}(\gamma) \subseteq \mathcal{B}_F(\gamma)$. Moreover, $\widehat{\mathcal{U}}_{W^\star}(\gamma)$ is non-empty for every $\gamma \geq 0$ since, in light of $\mathcal{W} = \{V\Sigma^{1/2}R : R \in \mathcal{O}_{r\times k}\}$, there always exists $R' \in \mathcal{O}_{(k-r)\times k}$ such that $W^\star R'^\top = 0$. On the other hand, note that

$$\begin{aligned}\max_{W^\star\in\mathcal{W}} \min_{\Delta W\in\mathcal{B}_F(\gamma)} \left\{f_{\ell_q}(W^\star + \Delta W) - f_{\ell_q}(W^\star)\right\} &\leq \max_{W^\star\in\mathcal{W}} \min_{\Delta W\in\widehat{\mathcal{U}}_{W^\star}(\gamma)} \left\{f_{\ell_q}(W^\star + \Delta W) - f_{\ell_q}(W^\star)\right\} \\ &= \min_{U\in\mathcal{B}_F^{k-r}(\gamma)} \left\{\frac{1}{m}\sum_{i=1}^m \left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q\right\},\end{aligned} \tag{18}$$

where $\mathcal{B}_F^{k-r}(\gamma) := \{U : U \in \mathbb{R}^{d\times(k-r)}, \|U\|_F \leq \gamma\}$. Therefore, to prove our main result, it suffices to control the right hand-side of (18). This can be easily done for symmetric matrix completion, as shown below.

Proof of Theorem 6. According to our assumption, there exists $(\bar{x}, \bar{x}) \in \Psi$ such that $E_{\bar{x}\bar{x}} \geq t_0$. Define $\bar{U} \in \mathbb{R}^{d\times(k-r)}$ as

$$\bar{U}_{xy} = \begin{cases}\gamma & \text{if } (x,y) = (\bar{x}, 1) \\ 0 & \text{otherwise}\end{cases}.$$

Indeed, we have $\bar{U} \in \mathcal{B}_F^{k-r}(\gamma)$. Therefore,

$$\begin{aligned}\min_{U\in\mathcal{B}_F^{k-r}(\gamma)} \left\{\frac{1}{m}\sum_{i=1}^m \left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q\right\} &\leq \frac{1}{m}\sum_{i=1}^m \left|\left\langle A_i, \bar{U}\bar{U}^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q \\ &= \left(E_{\bar{x}\bar{x}} - \left(\bar{U}\bar{U}^\top\right)_{\bar{x}\bar{x}}\right)^q - E_{\bar{x}\bar{x}}^q \\ &= -\gamma^2 \sum_{i=0}^{q-1} E_{\bar{x}\bar{x}}^i \left(E_{\bar{x}\bar{x}} - \left(\bar{U}\bar{U}^\top\right)_{\bar{x}\bar{x}}\right)^{q-1-i} \\ &\leq -t_0^{q-1}\gamma^2.\end{aligned}$$

This inequality combined with (18) completes the proof. □

Next, we extend our analysis to the symmetric matrix sensing. When the measurement matrices do not follow the matrix completion model, the explicit perturbation defined in the proof of Theorem 6 may no longer lead to a decrease in the objective value. To address this issue, we provide a more delicate upper bound for (18).

Lemma 8. *Suppose that* $\gamma \leq \sqrt{\frac{t_0}{\max_i\{\|A_i\|_F\}}}$. *We have*

$$\min_{U\in\mathcal{B}_F^{k-r}(\gamma)} \left\{\frac{1}{m}\sum_{i=1}^m \left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q\right\} \leq \min_{U\in\mathcal{B}_F^{k-r}(\gamma)} \left\{A\left(UU^\top\right) + B\left(UU^\top\right) + C\left(UU^\top\right)\right\}, \tag{19}$$

where

$$A(X) = \frac{1}{m}\sum_{i\in\mathcal{S}_t} q\mathrm{Sign}(\epsilon_i)^q \epsilon_i^{q-1} \left\langle A_i, X\right\rangle,$$
$$B(X) = \frac{1}{m}\sum_{i\in\mathcal{S}_t} 2^{q-2}\binom{q}{2}|\epsilon_i|^{q-2}\left\langle A_i, X\right\rangle^2$$
$$C(X) = \frac{1}{m}\sum_{i\notin\mathcal{S}_t} q(2t_0)^{q-1}|\left\langle A_i, X\right\rangle|.$$

Proof. We have

$$\frac{1}{m}\sum_{i=1}^m \left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q = \underbrace{\frac{1}{m}\sum_{i\in\mathcal{S}_t}\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q}_{E_1} + \underbrace{\frac{1}{m}\sum_{i\notin\mathcal{S}_t}\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q}_{E_2}.$$

To bound E_2, we first note that for every $i\notin\mathcal{S}_t$

$$\begin{aligned}\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q &\overset{(a)}{=} \left(\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right| - |\epsilon_i|\right)\sum_{k=0}^{q-1}\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^k |\epsilon_i|^{q-1-k}\\ &\overset{(b)}{\le} \left|\left\langle A_i, UU^\top\right\rangle\right|\sum_{k=0}^{q-1}\left(\left|\left\langle A_i, UU^\top\right\rangle\right| + |\epsilon_i|\right)^k |\epsilon_i|^{q-1-k}\\ &\overset{(v)}{\le} q(2t_0)^{q-1}\left|\left\langle A_i, UU^\top\right\rangle\right|\end{aligned}$$

where (a) follows from the Binomial Theorem, (b) is due to triangular inequality, and (c) follows from the fact that $|\left\langle A_i, UU^\top\right\rangle| \le \|A_i\|_F \|U\|_F^2 \le t_0$ for every $i = 1,\dots,m$. This in turn implies that $E_2 \le C\left(UU^\top\right)$. Next, we provide an upper bound for E_1. To this goal, we write for every $i\in\mathcal{S}_t$

$$\begin{aligned}\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q &\overset{(d)}{=} \mathrm{Sign}(\epsilon_i)^q\left(\left(\epsilon_i - \left\langle A_i, UU^\top\right\rangle\right)^q - \epsilon_i^q\right)\\ &\overset{(e)}{=} \mathrm{Sign}(\epsilon_i)^q\sum_{k=1}^q\binom{q}{k}\epsilon_i^{q-k}\left\langle A_i, UU^\top\right\rangle^k\\ &= q\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}\left\langle A_i, UU^\top\right\rangle + \mathrm{Sign}(\epsilon_i)^q\sum_{k=2}^q\binom{q}{k}\epsilon_i^{q-k}\left\langle A_i, UU^\top\right\rangle^k\\ &\le q\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}\left\langle A_i, UU^\top\right\rangle + \sum_{k=2}^q\binom{q}{k}|\epsilon_i|^{q-k}\left|\left\langle A_i, UU^\top\right\rangle\right|^k\\ &\overset{(f)}{\le} q\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}\left\langle A_i, UU^\top\right\rangle + \binom{q}{2}\left(\left|\left\langle A_i, UU^\top\right\rangle\right| + |\epsilon_i|\right)^{q-2}\left\langle A_i, UU^\top\right\rangle^2\\ &\overset{(g)}{\le} q\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}\left\langle A_i, UU^\top\right\rangle + 2^{q-2}\binom{q}{2}|\epsilon_i|^{q-2}\left\langle A_i, UU^\top\right\rangle^2\end{aligned}$$

where (d) is due to $\left|\left\langle A_i, UU^\top\right\rangle\right| \le t_0$, (e) follows from the Binomial Theorem, (f) is due to a basic inequality derived from binomial expansion (see Lemma 27 in Appendix C), and (g) is again due to $\left|\left\langle A_i, UU^\top\right\rangle\right| \le t_0$. This implies that $E_1 \le A\left(UU^\top\right) + B\left(UU^\top\right)$, thereby completing the proof. $\square$

To use the provided upper bound in Lemma 8, we need to show that $A\left(UU^\top\right) + B\left(UU^\top\right) + C\left(UU^\top\right) \le -\Omega(\gamma^2)$ for some $U \in \mathcal{B}_F^{k-r}(\gamma)$. Note that $B\left(UU^\top\right), C\left(UU^\top\right) \ge 0$ for every $U \in \mathcal{B}_F^{k-r}(\gamma)$, so our hope is to show that $A\left(UU^\top\right)$ can take sufficiently negative value to dominate both $B\left(UU^\top\right)$ and $C\left(UU^\top\right)$. As will be shown in our next lemma, this can be done for the symmetric matrix sensing. In particular, we show that, when the measurement matrices follow the matrix sensing model, there exists $U \in \mathcal{B}_F^{k-r}$ and coefficients $C_a, C_b, C_c \ge 0$ such that

$$A\left(UU^\top\right) \le -C_a\gamma^2, \quad B\left(UU^\top\right) \le C_b\gamma^2, \quad C\left(UU^\top\right) \le C_c\gamma^2, \quad \text{for some } C_a > C_b + C_c. \tag{20}$$

Lemma 9. *Define $r_0 = \min\{k-r, d/2\}$. Suppose that the measurement matrices follow the matrix sensing model in Assumption 1 and $\gamma^2 \lesssim \frac{t_0}{q2^q m\sqrt{dr_0}}$. Conditioned on the noise $\{\epsilon_i\}_{i=1}^m$ and with probability at least $1-\exp(-\Omega(dr_0))$ over the randomness of the measurement matrices, there exists $U \in \mathcal{B}_F^{k-r}$ such that (20) holds with*

$$C_a = cq\sqrt{\frac{dr_0\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}}{m^2}}, \qquad C_b = c(q/2)\sqrt{\frac{dr_0\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}}{m^2}}, \qquad C_c = q(2t_0)^{q-1},$$

for some constant $c > 0$.

The proof of the above lemma is provided in Appendix B. Equipped with this lemma, we are ready to provide the proof of Theorem 3.

Proof of Theorem 3. According to Lemmas 8 and 9, we have with probability at least $1-\exp(-\Omega(dr_0))$

$$\begin{aligned}
\max_{W^\star\in\mathcal{W}}\min_{\Delta W\in\mathcal{B}_F(\gamma)}\left\{f_{\ell_q}(W^\star+\Delta W) - f_{\ell_q}(W^\star)\right\} &\overset{(a)}{=} \min_{U\in\mathcal{B}_F^{k-r}(\gamma)}\left\{\frac{1}{m}\sum_{i=1}^m\left|\left\langle A_i, UU^\top\right\rangle - \epsilon_i\right|^q - |\epsilon_i|^q\right\} \\
&\overset{(b)}{\le} \min_{U\in\mathcal{B}_F^{k-r}(\gamma)}\left\{A\left(UU^\top\right) + B\left(UU^\top\right) + C\left(UU^\top\right)\right\} \\
&\overset{(c)}{\le} -(C_a - C_b - C_c)\gamma^2 \\
&\le \left(-c_1(q/2)t_0^{q-1}\sqrt{\frac{dr_0\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}}{m^2}} + q(2t_0)^{q-1}\right)\gamma^2 \\
&\le -c_1(q/4)t_0^{q-1}\sqrt{\frac{dr_0pp_0}{m}}\cdot\gamma^2 \\
&\le -c_1(q/4)t_0^{q-1}\gamma^2
\end{aligned}$$

provided that

$$\gamma^2 \le c_2\cdot\min\left\{\frac{t_0}{q2^q m\sqrt{dr_0}}, \frac{t_0}{\max_i\{\|A_i\|_F\}}\right\}, \quad \text{and} \quad m \le \frac{c_1^2}{4^{q+1}}\cdot pp_0dr_0.$$

for some constants $0 < c_1, c_2 \leq 1$. In the above inequality, (a) is due to (18), (b) follows from Lemma 8, and (c) follows from Lemma 9. On the other hand, Lemma 21 in Appendix A.3 implies that $\max_i\{\|A_i\|_F\} \leq \sqrt{2}d$ with probability at least $1 - \exp(-\Omega(d^2))$. This implies that

$$c_2 \cdot \min\left\{\frac{t_0}{q2^q m\sqrt{dr_0}}, \frac{t_0}{\max_i\{\|A_i\|_F\}}\right\} \geq \frac{c_2 t_0}{q2^q (d(k-r))^{3/2}} \geq \gamma^2$$

with probability at least $1 - \exp(-\Omega(d^2))$, where the last inequality follows from our assumed upper bound on γ^2. This completes the proof of the second statement. To prove the first statement, define $\widehat{W} = W^\star + \Delta W^\star$, where $W^\star \in \mathcal{W}$ and $\Delta W^\star = \arg\min_{\Delta W \in \mathcal{B}_F(\gamma)} \left\{f_{\ell_q}(W^\star + \Delta W) - f_{\ell_q}(W^\star)\right\}$ with $\gamma^2 = \frac{c' t_0}{q2^q (d(k-r))^{3/2}}$. We have

$$f^{\rm s}_{\ell_q}(W^\star) - \min_W f^{\rm s}_{\ell_q}(W) \geq f^{\rm s}_{\ell_q}(W^\star) - f^{\rm s}_{\ell_q}(\widehat{W}) \geq c_1(q/4)t_0^{q-1} \cdot \frac{c_2 t_0}{q2^q (d(k-r))^{3/2}} = c_3 \cdot \frac{(t_0/2)^q}{(d(k-r))^{3/2}}$$

for $c_3 = c_1 c_2/4$. This completes the proof of the first statement. □

6 Matrix Sensing with ℓ_1-loss: Matching Lower Bounds

In this section, we show that the provided necessary conditions for the identifiability of the true solutions are also sufficient for the symmetric and asymmetric matrix sensing with ℓ_1-loss (Theorems 2 and 4). We first show that if $m \gtrsim \max\{d_1, d_2\}k/(1-2p)^4$, then all the true solutions coincide with the global solutions of `MS-asym`. To this goal, it suffices to show that for any $\mathbf{W} = (W_1, W_2) \notin \mathcal{W}$ and $\mathbf{W}^\star = (W_1^\star, W_2^\star) \in \mathcal{W}$, we have $f_{\ell_1}(\mathbf{W}) - f_{\ell_1}(\mathbf{W}^\star) > 0$. Let $\Delta X = W_1 W_2 - W_1^\star W_2^\star$. One can write

$$\begin{aligned} f_{\ell_1}(\mathbf{W}) - f_{\ell_1}(\mathbf{W}^\star) &\geq \frac{1}{m}\sum_{i\in\bar{\mathcal{S}}} |\langle A_i, \Delta X\rangle| + \frac{1}{m}\sum_{i\in\mathcal{S}} |\langle A_i, \Delta X\rangle - \epsilon_i| - |\epsilon_i| \\ &\geq \underbrace{\frac{1}{m}\sum_{i\in\bar{\mathcal{S}}} |\langle A_i, \Delta X\rangle| - \frac{1}{m}\sum_{i\in\mathcal{S}} |\langle A_i, \Delta X\rangle|}_{F} \end{aligned}$$

where the second inequality follows from the triangle inequality. We will show that if the corruption probability p is strictly less than half, then $F > 0$ with high probability. To this goal, we make use of the well-known *ℓ_1/ℓ_2-restricted isometry property* (ℓ_1/ℓ_2-RIP), which is provided in the following lemma.

Lemma 10 (ℓ_1/ℓ_2-RIP [LZMCSV20, CCD+21])**.** *Suppose that the measurements follow the matrix sensing model in Assumption 1. For all rank-$2r'$ matrices $X \in \mathbb{R}^{d_1 \times d_2}$ and any $\delta > 0$, we have with probability at least $1 - \exp(-\Omega(\delta^2 m))$:*

$$\left(\sqrt{\frac{2}{\pi}} - \delta\right)\|X\|_F \leq \frac{1}{m}\sum_{i=1}^m |\langle A_i, X\rangle| \leq \left(\sqrt{\frac{2}{\pi}} + \delta\right)\|X\|_F,$$

provided that $m \geq c\max\{d_1, d_2\}r'/\delta^2$ for some constant $c > 0$.

Equipped with the above lemma, we are ready to present the proof of Theorem 2

Proof of Theorem 2. Our goal is to show that, if the measurement matrices follow the matrix sensing model, then with an overwhelming probability, we have $F \gtrsim \|\Delta X\|_F$. This in turn implies that if $\Delta X \neq 0$ (or equivalently $\mathbf{W} \notin \mathcal{W}$), then $f_{\ell_1}(\mathbf{W}) - f_{\ell_1}(\mathbf{W}^\star) > 0$. Notice that $\mathrm{rank}(\Delta X) \leq k + r$. Let $\mathcal{S}_{k+r}^{d_1 \times d_2} = \{X : X \in \mathbb{R}^{d_1 \times d_2}, \mathrm{rank}(X) \leq k + r, \|X\|_F = 1\}$. It is easy to see that a sufficient condition for $F \gtrsim \|\Delta X\|_F$ is to have

$$\min_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \bar{\mathcal{S}}} |\langle A_i, \Delta X \rangle| - \frac{1}{m} \sum_{i \in \mathcal{S}} |\langle A_i, \Delta X \rangle| \right\} \gtrsim 1.$$

To prove the above inequality, we write

$$\begin{aligned} &\min_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \bar{\mathcal{S}}} |\langle A_i, \Delta X \rangle| - \frac{1}{m} \sum_{i \in \mathcal{S}} |\langle A_i, \Delta X \rangle| \right\} \\ &\qquad \geq \underbrace{\min_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \bar{\mathcal{S}}} |\langle A_i, \Delta X \rangle| \right\}}_{F_1} - \underbrace{\max_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \mathcal{S}} |\langle A_i, \Delta X \rangle| \right\}}_{F_2}. \end{aligned}$$

We next show that $F_1 - F_2 \geq 0$ with an overwhelming probability. To this goal, we provide separate lower and upper bounds for F_1 and F_2. Conditioned on $|\bar{\mathcal{S}}|$, ℓ_1/ℓ_2-RIP (Lemma 10) with $\delta = \sqrt{\frac{c \max\{d_1, d_2\}(k+r)}{|\bar{\mathcal{S}}|}}$ can be invoked to show that

$$\begin{aligned} \min_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \bar{\mathcal{S}}} |\langle A_i, \Delta X \rangle| \right\} &= \frac{|\bar{\mathcal{S}}|}{m} \min_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{|\bar{\mathcal{S}}|} \sum_{i \in \bar{\mathcal{S}}} |\langle A_i, \Delta X \rangle| \right\} \\ &\geq \frac{|\bar{\mathcal{S}}|}{m} \left(\sqrt{\frac{2}{\pi}} - \sqrt{\frac{c \max\{d_1, d_2\}(k+r)}{|\bar{\mathcal{S}}|}} \right) \end{aligned}$$

with probability at least $1 - \exp(-\Omega(\max\{d_1, d_2\}(k+r)))$. With the same probability, we have

$$\begin{aligned} \max_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{m} \sum_{i \in \mathcal{S}} |\langle A_i, \Delta X \rangle| \right\} &= \frac{|\mathcal{S}|}{m} \max_{\Delta X \in \mathcal{S}_{k+r}^{d_1 \times d_2}} \left\{ \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} |\langle A_i, \Delta X \rangle| \right\} \\ &\leq \frac{|\mathcal{S}|}{m} \left(\sqrt{\frac{2}{\pi}} + \sqrt{\frac{c \max\{d_1, d_2\}(k+r)}{|\mathcal{S}|}} \right). \end{aligned}$$

Combining these two inequalities, we have

$$\begin{aligned} E_2 - E_3 &\geq \sqrt{\frac{2}{\pi}} \left(\frac{|\bar{\mathcal{S}}| - |\mathcal{S}|}{m} \right) - \left(\sqrt{\frac{|\bar{\mathcal{S}}|}{m}} + \sqrt{\frac{|\mathcal{S}|}{m}} \right) \sqrt{\frac{c \max\{d_1, d_2\}(k+r)}{m}} \\ &\geq \sqrt{\frac{2}{\pi}} \left(\frac{m - 2|\mathcal{S}|}{m} \right) - 2\sqrt{\frac{2c \max\{d_1, d_2\} k}{m}} \end{aligned} \tag{21}$$

with probability at least $1-\exp(-\Omega(\max\{d_1,d_2\}(k+r)))$. On the other hand, since $|\mathcal{S}|$ has a binomial distribution with parameters (m,p) and $0\le p<1/2$, Lemma 24 in Appendix A.3 can be invoked to show that $\frac{m-2|\mathcal{S}|}{m}\ge \min\left\{\frac{1}{4},(1-2p)^2\right\}$ with probability at least $1-\exp(-\Omega((1-2p)^2m))$. Combining this inequality with (21), we have

$$F_1-F_2\ge\sqrt{\frac{2}{\pi}}\min\left\{\frac{1}{4},(1-2p)^2\right\}-2\sqrt{\frac{2c\max\{d_1,d_2\}k}{m}}\ge\left(\sqrt{\frac{2}{\pi}}-1\right)\min\left\{\frac{1}{4},(1-2p)^2\right\}>0,$$

with probability at least $1-\exp(-\Omega(\max\{d_1,d_2\}(k+r)))-\exp(-\Omega((1-2p)^2m)=1-\exp(-\Omega(\max\{d_1,d_2\}k))$, where the last inequality holds provided that $m\ge\frac{512c\max\{d_1,d_2\}k}{(1-2p)^4}$. This completes the proof. □

Our next goal is to provide the proof of Theorem 4. We note that the proof of the second statement of Theorem 2 is almost identical to the proof of Theorem 4, and hence, it is omitted for brevity. To prove the second statement, we first note that $\operatorname{rank}(\Delta X)\le k+r$ and $m\gtrsim\frac{dr}{(1-2p)^4}$. Therefore ℓ_1/ℓ_2-RIP cannot be directly used to control the deviation of the loss uniformly over the perturbation set $\mathcal{S}_{k+r}^{d\times d}$. To circumvent this issue, we further decompose ΔX into two parts. Note that $W\in\mathcal{B}_{W^\star}(\gamma)$ if and only if $W=W^\star+\Delta W$ for some $\|\Delta W\|_F\le\gamma$. Therefore, we have

$$\Delta X=WW^\top-W^\star W^{\star\top}=\underbrace{W^\star\Delta W^\top+\Delta WW^{\star\top}}_{\Delta X_1}+\underbrace{\Delta W\Delta W^\top}_{\Delta X_2}.$$

Evidently, we have $\operatorname{rank}(\Delta X_1)\le 2r$. As a result, the effect of ΔX_1 can be controlled via ℓ_1/ℓ_2-RIP with $m\gtrsim\frac{dr}{(1-2p)^4}$. On the other hand, $\operatorname{rank}(\Delta X_2)$ may be as large as k, but its magnitude is much smaller than ΔX_1, since $\|\Delta X_2\|_F\le\|\Delta W\|_F^2\le\gamma^2$. With this insight, we present the proof of the first statement of Theorem 4.

Proof of Theorem 4 (first statement). We have

$$\begin{aligned}
f_{\ell_1}^{\mathrm{s}}(W)-f_{\ell_1}^{\mathrm{s}}(W^\star)&=\frac{1}{m}\sum_{i\in\bar{S}}|\langle A_i,\Delta X_1+\Delta X_2\rangle|+\frac{1}{m}\sum_{i\in S}\left(|\langle A_i,\Delta X_1+\Delta X_2\rangle+\epsilon_i|-|\epsilon_i|\right)\\
&\ge\frac{1}{m}\sum_{i\in\bar{S}}|\langle A_i,\Delta X_1+\Delta X_2\rangle|-\frac{1}{m}\sum_{i\in S}|\langle A_i,\Delta X_1+\Delta X_2\rangle|\\
&\ge\frac{1}{m}\sum_{i\in\bar{S}}|\langle A_i,\Delta X_1\rangle|-\frac{1}{m}\sum_{i\in S}|\langle A_i,\Delta X_1\rangle|-\frac{1}{m}\sum_{i=1}^{m}|\langle A_i,\Delta X_2\rangle|,
\end{aligned}$$

where the first and second inequalities follow from triangle inequality. Note that, for every $W\in\mathcal{B}_{W^\star}(\gamma)$, we have $\operatorname{rank}(\Delta X_1)\le 2r$ and $\|\Delta X_1\|_F\le 2\sqrt{\lambda_1}\gamma$, where λ_1 is the maximum eigenvalue of $X^\star$. Similarly, we have $\operatorname{rank}(\Delta X_2)\le k$ and $\|\Delta X_2\|_F\le\gamma^2$. This implies that

$$\min_{W^\star\in\mathcal{W}}\min_{W\in\mathcal{B}_{W^\star}(\gamma)}\left\{f_{\ell_1}^{\mathrm{s}}(W)-f_{\ell_1}^{\mathrm{s}}(W^\star)\right\}\ge\underbrace{\min_{\|\Delta X_1\|_F\le 2\sqrt{\lambda_1}\gamma}\left\{\frac{1}{m}\sum_{i\in\bar{S}}|\langle A_i,\Delta X_1\rangle|-\frac{1}{m}\sum_{i\in S}|\langle A_i,\Delta X_1\rangle|\right\}}_{G_1}$$
$$-\underbrace{\max_{\|\Delta X_2\|_F\le\gamma^2}\left\{\frac{1}{m}\sum_{i=1}^{m}|\langle A_i,\Delta X_2\rangle|\right\}}_{G_2}.$$

Similar to the proof of Theorem 2, one can show that that $G_1 > 0$ with probability at least $1 - \exp(-\Omega(dr))$ (recall that $\operatorname{rank}(\Delta X_1) \le 2r$). On the other hand, applying ℓ_1/ℓ_2-RIP to F_2 with $\delta = \sqrt{\frac{cdk}{m}}$[4], we have

$$F_2 \lesssim \left(\sqrt{\frac{2}{\pi}} + \sqrt{\frac{dk}{m}}\right)\gamma^2,$$

with probability at least $1 - \exp(-\Omega(dk))$. Combining the above inequalities leads to

$$\min_{W^\star \in \mathcal{W}} \min_{W \in \mathcal{B}_{W^\star}(\gamma)} \left\{ f^{\mathrm{s}}_{\ell_1}(W) - f^{\mathrm{s}}_{\ell_1}(W^\star) \right\} \ge F_1 - F_2 \gtrsim -\left(\sqrt{\frac{2}{\pi}} + \sqrt{\frac{dk}{m}}\right)\gamma^2,$$

with probability at least $1 - \exp(-\Omega(dr)) - \exp(-\Omega(dk)) = 1 - \exp(-\Omega(dr))$. This completes the proof. □

Acknowledgements

We thank Richard Y. Zhang and Tiffany Wu for helpful feedback. This research is supported, in part, by NSF Award DMS-2152776, ONR Award N00014-22-1-2127, and MICDE Catalyst Grant.

References

[ACHL19] Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. Advances in Neural Information Processing Systems, 32, 2019.

[BNS16] Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. arXiv preprint arXiv:1605.07221, 2016.

[BZ14] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. Computer Vision and Image Understanding, 122:22–34, 2014.

[CCD+21] Vasileios Charisopoulos, Yudong Chen, Damek Davis, Mateo Díaz, Lijun Ding, and Dmitriy Drusvyatskiy. Low-rank matrix recovery with composite optimization: good conditioning and rapid convergence. Foundations of Computational Mathematics, 21(6):1505–1593, 2021.

[CCF+19] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. arXiv preprint arXiv:1902.07698, 2019.

[CL06] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. Internet mathematics, 3(1):79–127, 2006.

[4]Note that $\delta > 1$ if $m < cdk$, but ℓ_1/ℓ_2-RIP can be still applied.

[Cla75] Frank H Clarke. Generalized gradients and applications. Transactions of the American Mathematical Society, 205:247–262, 1975.

[Cla90] Frank H Clarke. Optimization and nonsmooth analysis. SIAM, 1990.

[CLC19] Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. IEEE Transactions on Signal Processing, 67(20):5239–5269, 2019.

[CLMW11] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? Journal of the ACM (JACM), 58(3):1–37, 2011.

[CP11] Emmanuel J Candes and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. IEEE Transactions on Information Theory, 57(4):2342–2359, 2011.

[CSPW11] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. SIAM Journal on Optimization, 21(2):572–596, 2011.

[FLZ19] Cong Fang, Zhouchen Lin, and Tong Zhang. Sharp analysis for nonconvex sgd escaping from saddle points. In Conference on Learning Theory, pages 1192–1234. PMLR, 2019.

[FS20] Salar Fattahi and Somayeh Sojoudi. Exact guarantees on the absence of spurious local minima for non-negative rank-1 robust principal component analysis. Journal of machine learning research, 2020.

[GJZ17] Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In International Conference on Machine Learning, pages 1233–1242. PMLR, 2017.

[GLM16] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. arXiv preprint arXiv:1605.07272, 2016.

[GWB+17] Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. Advances in Neural Information Processing Systems, 30, 2017.

[JGN+17] Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In International conference on machine learning, pages 1724–1732. PMLR, 2017.

[JL22] Cédric Josz and Lexiao Lai. Nonsmooth rank-one matrix factorization landscape. Optimization Letters, 16(6):1611–1631, 2022.

[LFL+14] Xiao Luan, Bin Fang, Linghui Liu, Weibin Yang, and Jiye Qian. Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion. Pattern Recognition, 47(2):495–508, 2014.

[LLZ+20] Shuang Li, Qiuwei Li, Zhihui Zhu, Gongguo Tang, and Michael B Wakin. The global geometry of centralized and distributed low-rank matrix recovery without regularization. IEEE Signal Processing Letters, 27:1400–1404, 2020.

[LPP+19] Jason D Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I Jordan, and Benjamin Recht. First-order methods almost always avoid strict saddle points. Mathematical programming, 176:311–337, 2019.

[LSM20] Jiajin Li, Anthony Man Cho So, and Wing Kin Ma. Understanding notions of stationarity in nonsmooth optimization: A guided tour of various constructions of subdifferential for nonsmooth functions. IEEE Signal Processing Magazine, 37(5):18–31, 2020.

[LT91] Michel Ledoux and Michel Talagrand. Probability in Banach Spaces: isoperimetry and processes, volume 23. Springer Science & Business Media, 1991.

[LZMCSV20] Xiao Li, Zhihui Zhu, Anthony Man-Cho So, and Rene Vidal. Nonconvex robust low-rank matrix recovery. SIAM Journal on Optimization, 30(1):660–686, 2020.

[LZXZ14] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. IEEE Transactions on Industrial Informatics, 10(2):1273–1284, 2014.

[MBLS22] Ziye Ma, Yingjie Bi, Javad Lavaei, and Somayeh Sojoudi. Sharp restricted isometry property bounds for low-rank matrix recovery problems with corrupted measurements. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 7672–7681, 2022.

[MF19] Jianhao Ma and Salar Fattahi. Blessing of depth in linear regression: Deeper models have flatter landscape around the true solution. In Advances in Neural Information Processing Systems, 2019.

[MF21] Jianhao Ma and Salar Fattahi. Sign-rip: A robust restricted isometry property for low-rank matrix recovery. arXiv preprint arXiv:2102.02969, 2021.

[MF22] Jianhao Ma and Salar Fattahi. Global convergence of sub-gradient method for robust matrix recovery: Small initialization, noisy measurements, and over-parameterization. arXiv preprint arXiv:2202.08788, 2022.

[MGF22] Jianhao Ma, Lingjun Guo, and Salar Fattahi. Behind the scenes of gradient descent: A trajectory analysis via basis function decomposition. arXiv preprint arXiv:2210.00346, 2022.

[RW09] R Tyrrell Rockafellar and Roger J-B Wets. Variational analysis, volume 317. Springer Science & Business Media, 2009.

[SQW16] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere i: Overview and the geometric picture. IEEE Transactions on Information Theory, 63(2):853–884, 2016.

[SQW18] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. Foundations of Computational Mathematics, 18(5):1131–1198, 2018.

[SS21] Dominik Stöger and Mahdi Soltanolkotabi. Small random initialization is akin to spectral learning: Optimization and generalization guarantees for overparameterized low-rank matrix reconstruction. Advances in Neural Information Processing Systems, 34:23831–23843, 2021.

[Sza82] Stanislaw J Szarek. Nets of grassmann manifold and orthogonal group. In Proceedings of research workshop on Banach space theory (Iowa City, Iowa, 1981), volume 169, page 185. University of Iowa Iowa City, IA, 1982.

[Ver18] Roman Vershynin. High-dimensional probability: An introduction with applications in data science, volume 47. Cambridge university press, 2018.

[VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.

[Wai19] Martin J Wainwright. High-dimensional statistics: A non-asymptotic viewpoint, volume 48. Cambridge University Press, 2019.

[XSCM23] Xingyu Xu, Yandi Shen, Yuejie Chi, and Cong Ma. The power of preconditioning in overparameterized low-rank matrix sensing. arXiv preprint arXiv:2302.01186, 2023.

[YMLS22] Baturalp Yalcin, Ziye Ma, Javad Lavaei, and Somayeh Sojoudi. Semidefinite programming versus burer-monteiro factorization for matrix sensing. arXiv preprint arXiv:2208.07469, 2022.

[ZFZ21] Jialun Zhang, Salar Fattahi, and Richard Y Zhang. Preconditioned gradient descent for over-parameterized nonconvex matrix factorization. Advances in Neural Information Processing Systems, 34:5985–5996, 2021.

[Zha22] Richard Y Zhang. Improved global guarantees for the nonconvex burer–monteiro factorization via rank overparameterization. arXiv preprint arXiv:2207.01789, 2022.

[ZKHC21] Jiacheng Zhuo, Jeongyeol Kwon, Nhat Ho, and Constantine Caramanis. On the computational and statistical complexity of over-parameterized matrix sensing. arXiv preprint arXiv:2102.02756, 2021.

[ZLK+17] Yuqian Zhang, Yenson Lau, Han-wen Kuo, Sky Cheung, Abhay Pasupathy, and John Wright. On the global geometry of sphere-constrained sparse blind deconvolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4894–4902, 2017.

[ZLTW18] Zhihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B Wakin. Global optimality in low-rank matrix optimization. IEEE Transactions on Signal Processing, 66(13):3614–3628, 2018.

[ZLTW21] Zhihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B Wakin. The global optimization geometry of low-rank matrix optimization. IEEE Transactions on Information Theory, 67(2):1308–1331, 2021.

[ZQW20] Yuqian Zhang, Qing Qu, and John Wright. From symmetry to geometry: Tractable nonconvex problems. arXiv preprint arXiv:2007.06753, 2020.

[ZZ20] Jialun Zhang and Richard Zhang. How many samples is a good initial point worth in low-rank matrix recovery? Advances in Neural Information Processing Systems, 33:12583–12592, 2020.

A Useful Concentration Bounds

In this section, we review some basic concentration results for different random variables and random processes that are used throughout our proofs.

A.1 Concentration Bounds for Orlicz Random Variables and Processes

First, we define the notion of Orlicz norm (see [Wai19, Section 5.6] and [LT91, Section 11.1] for more details).

Definition 1 (Orlicz function and Orlicz norm)**.** *A function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is called an Orlicz function if ψ is convex, increasing, and satisfies*

$$\psi(0) = 0, \quad \psi(x) \to \infty \text{ as } x \to \infty. \tag{22}$$

For a given Orlicz function ψ, the Orlicz norm of a random variable X is defined as

$$\|X\|_\psi := \inf\{t > 0 : \mathbb{E}[\psi(|X|/t)] \le 1\}. \tag{23}$$

We will use Orlicz functions and norms to study the concentration of both sub-Gaussian and sub-exponential random variables, which are defined as follows.

Definition 2 (Sub-Gaussian and sub-exponential random variables)**.** *Define the Orlicz function $\psi_p(x) = e^{x^p} - 1$. A random variable X is called sub-Gaussian if $\|X\|_{\psi_2} < \infty$. Moreover, a random variable X is called sub-exponential if $\|X\|_{\psi_1} < \infty$. Accordingly, the quantities $\|X\|_{\psi_2}$ and $\|X\|_{\psi_1}$ refer to the sub-Gaussian norm and sub-exponential norm of X, respectively.*

Lemmas 11-14 establish basic properties of sub-Gaussian and sub-exponential random variables.

Lemma 11 (Tail bounds for sub-Gaussian and sub-exponential random variables, Section 5.6 in [Wai19])**.** *For a random variable X with finite Orlicz norm $\|X\|_{\psi_p} < \infty$ where $\psi_p(x) = e^{x^p} - 1$ and $p \ge 1$, we have*

$$\mathbb{P}\left(|X| \ge \delta\right) \le 2\exp\left(-\frac{\delta^p}{\|X\|_{\psi_p}^p}\right). \tag{24}$$

Lemma 12 (Sum of independent random variables, Theorems 2.6.2 and 2.8.1 in [Ver18])**.** *Let $X_1, \cdots, X_n$ be zero-mean independent random variables.*

- *If each X_i is sub-Gaussian with sub-Gaussian norm $\|X_i\|_{\psi_2}$, we have*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge \delta\right) \le 2\exp\left(-\frac{n^2\delta^2}{\sum_{i=1}^n \|X_i\|_{\psi_2}^2}\right). \tag{25}$$

- *If each X_i is sub-exponential with sub-exponential norm $\|X_i\|_{\psi_1}$, we have*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \delta\right) \leq 2\exp\left(-n\min\left\{\frac{\delta^2}{\max_i \|X_i\|_{\psi_1}^2}, \frac{\delta}{\max_i \|X_i\|_{\psi_1}}\right\}\right). \tag{26}$$

Lemma 13 (Lemma 2.6.8 in [Ver18]). *Suppose that X is sub-Gaussian. We have $\|X - \mathbb{E}[X]\|_{\psi_2} \leq C\|X\|_{\psi_2}$ for some constant $C > 0$.*

Lemma 14 (Lemma 2.7.7 in [Ver18]). *Let X, Y be sub-Gaussian random variables. Then XY is sub-exponential with its sub-exponential norm satisfying $\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}$.*

Next, we introduce the notion of Orlicz process, which will be useful in our subsequent analysis.

Definition 3 (Orlicz process, Definition 5.35 in [Wai19]). *A set of zero-mean random variables $\{X_t, t \in \mathcal{T}\}$ (also known as as stochastic process) is a ψ_p-process with respect to a metric d if*

$$\|X_t - X_{t'}\|_{\psi_p} \leq d(t, t') \quad \text{for all } t, t' \in \mathcal{T}. \tag{27}$$

According to the above definition, the size of the set $\mathcal{T}$ may be infinite. To control the behavior of a ψ_p-process, we first need the notions of ϵ-covering and covering number.

Definition 4 (Covering and covering number). *A set $\mathcal{N}_\epsilon(\mathcal{T})$ is called an ϵ-covering on $(\mathcal{T}, d)$ if for every $t \in \mathcal{T}$, there exists $\pi(t) \in \mathcal{N}_\epsilon(\mathcal{T})$ such that $d(t, \pi(t)) \leq \epsilon$. The covering number $\mathcal{N}(\mathcal{T}, d, \epsilon)$ is defined as the smallest cardinality of an ϵ-covering for $(\mathcal{T}, d)$:*

$$\mathcal{N}(\mathcal{T}, d, \epsilon) := \inf\{|\mathcal{N}_\epsilon(\mathcal{T})| : \mathcal{N}_\epsilon(\mathcal{T}) \text{ is an } \epsilon\text{-covering for } (T, d)\}.$$

Theorem 7 (Concentration of Orlicz process, Theorem 5.36 in [Wai19]). *Let $\{X_t, t \in \mathcal{T}\}$ be a ψ_p-process with respect to the metric d. Then there is a universal constant $C > 0$ such that*

$$\mathbb{P}\left(\sup_{t\in T} X_t \geq C\mathcal{J}_p(0, D) + \delta\right) \leq \exp(-\delta^p/D^p).$$

Here $D = \sup_{t,t'\in T} d(t, t')$ is the diameter of the set $\mathcal{T}$ and $\mathcal{J}_p(\delta, D) = \int_\delta^D \log^{1/p}(1 + \mathcal{N}(\mathcal{T}, d, \epsilon))\, d\epsilon$ is called the generalized Dudley entropy integral.

By providing tractable upper bounds on D and $\mathcal{J}_p(\delta, D)$, one can control the supremum of a ψ_p-process. In our next section, we provide specific sub-classes of ψ_p-process for which these quantities can be controlled efficiently.

A.2 Concentration Bounds for Gaussian Processes over Grassmannian Manifold

In the proof of Lemma 9, we will make extensive use of Gaussian processes, which are special types of Orlicz processes, over Grassmannian manifold. We first start with the definition of a Gaussian process.

Definition 5 (Gaussian process). *A random process $\{X_t\}_{t\in\mathcal{T}}$ is called a Gaussian process if, for any finite subset $\mathcal{T}_0 \subset \mathcal{T}$, the random vector $\{X_t\}_{t\in\mathcal{T}_0}$ has a normal distribution.*

Our next lemma characterizes the expectation of the supremum of a Gaussian process in terms of its covering number.

Theorem 8 (Sudakov's Minoration Inequality, Theorem 5.30 in [Wai19]). *Let $\{X_t\}_{t\in\mathcal{T}}$ be a zero-mean Gaussian process. Then, for any $\epsilon \geq 0$, we have*

$$\mathbb{E}\left[\sup_{t\in\mathcal{T}} X_t\right] \geq \frac{\epsilon}{2}\sqrt{\log \mathcal{N}(\mathcal{T}, d, \epsilon)}. \tag{28}$$

Here, the metric d is defined as $d(t_1, t_2) = \left(\mathbb{E}\left[(X_{t_1} - X_{t_2})^2\right]\right)^{1/2}$ for any $t_1, t_2 \in \mathcal{T}$.

Lemma 15 (Concentration of Gaussian process, Lemma 6.12 in [VH14]). *Let $\{X_t\}_{t\in T}$ be a Gaussian process. Then $\sup_{t\in T} X_t$ is sub-Gaussian with sub-Gaussian norm $\sup_{t\in T} \|X_t\|_{\psi_2}$.*

Consider a Grassmannian defined as

$$\mathcal{G}(n, p) = \{P \in \mathbb{R}^{n\times n} : P^\top = P, P^2 = P, \mathrm{rank}(P) = p\}.$$

Our next lemma characterizes the covering number of the Grassmannian manifold $\mathcal{G}(n, p)$.

Lemma 16 (Covering number for Grassmannian manifold, Proposition 8 in [Sza82]). *For a Grassmannian manifold $\mathcal{G}(n, p)$, define the metric $d(P_1, P_2) = \|P_1 - P_2\|_F$ for any $P_1, P_2 \in \mathcal{G}(n, p)$. Then, for any $0 < \epsilon < \sqrt{2\min\{p, n-p\}}$, we have*

$$\left(\frac{c_1\sqrt{p}}{\epsilon}\right)^{p(n-p)} \leq \mathcal{N}(G(n, p), d, \epsilon) \leq \left(\frac{c_2\sqrt{p}}{\epsilon}\right)^{p(n-p)}, \tag{29}$$

for some constants $c_1, c_2 > 0$.

Lemma 17 (lower bound of Gaussian process over Grassmannian manifold). *Consider the Gaussian process $\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle$, where A is a standard Gaussian matrix and $p \leq n/2$. Then, with probability at least $1 - \exp(-\Omega(np))$, we have*

$$\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle \geq \frac{c_3}{2}\sqrt{np^2}, \tag{30}$$

for $c_3 = \sqrt{\log(c_1)/8}$, where c_1 is the constant appeared in Lemma 16.

Proof. We first use Sudakov's Minoration Inequality (Theorem 8) to characterize the expectation of the supremum:

$$\begin{aligned}
\mathbb{E}\left[\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle\right] &\geq \sup_{0<\epsilon} \frac{\epsilon}{2}\sqrt{\log \mathcal{N}\left(\mathcal{G}(n, p), \|\cdot\|_F, \epsilon\right)} \\
&\overset{(a)}{\geq} \sup_{0<\epsilon\leq\sqrt{2p}} \frac{\epsilon}{2}\sqrt{p(n-p)\log\left(\frac{c_1\sqrt{p}}{\epsilon}\right)} \\
&\overset{\text{set } \epsilon=\sqrt{p}}{\geq} \frac{1}{2}\sqrt{(n-p)p^2\log(c_1)} \\
&\geq c_3\sqrt{np^2}.
\end{aligned} \tag{31}$$

In (a), we invoked the lower bound on the covering number $\mathcal{N}\left(\mathcal{G}(n,p), \|\cdot\|_F, \epsilon\right)$ from Lemma 16. Next, we apply Lemma 15 to control the deviation of the supremum from its expectation. To this goal, we first note that for any $X \in \mathcal{G}(n,p)$, we have $\|\langle A, X\rangle\|_{\psi_2} = \|X\|_F = \sqrt{p}$. Then, according to Lemma 15, $\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle$ is sub-Gaussian with norm $\sqrt{p}$. Applying Lemma 2 yields

$$\begin{aligned}
\mathbb{P}\left(\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle - c_3\sqrt{np^2} \le -\delta\right) &\le \mathbb{P}\left(\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle - \mathbb{E}\left[\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle\right] \le -\delta\right) \\
&\le \exp\left\{-\frac{\delta^2}{\left\|\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle\right\|_{\psi_2^2}}\right\} \\
&\le \exp\left\{-\delta^2/p\right\}.
\end{aligned} \tag{32}$$

Upon setting $\delta = \frac{c_3}{2}\sqrt{np^2}$, we know that with probability at least $1 - \exp(-\Omega(np))$

$$\sup_{X\in\mathcal{G}(n,p)} \langle A, X\rangle \ge \frac{c_3}{2}\sqrt{np^2}, \tag{33}$$

which completes the proof. □

Our next lemma provides another useful bound on the supremum of Gaussian processes over Grassmannian manifold.

Lemma 18. *Suppose that $\{A_i\}_{i=1}^m$ are i.i.d. standard Gaussian matrices and $\{w\}_{i=1}^m$ are non-negative weights. Then, we have with probability at least $1 - \exp(-\Omega(np))$*

$$\sup_{X\in\mathcal{G}(n,p)} \frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2 \le \frac{cnp^2}{m}\sum_{i=1}^m w_i. \tag{34}$$

For some constant $c > 0$.

Proof. We first verify that $\frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2$ is indeed a ψ_1-process. To see this, for arbitrary two matrices $X, X' \in \mathcal{G}(n,p)$, we have

$$\begin{aligned}
\left\|\frac{1}{m}\sum_{i=1}^m w_i\left(\langle A_i, X\rangle^2 - \langle A_i, X'\rangle^2\right)\right\|_{\psi_1} &\le \frac{1}{n}\sum_{i=1}^n w_i \left\|\langle A_i, X\rangle^2 - \langle A_i, X'\rangle^2\right\|_{\psi_1} \\
&\le \frac{1}{m}\sum_{i=1}^m w_i \left\|\langle A_i, X + X'\rangle \langle A_i, X - X'\rangle\right\|_{\psi_1} \\
&\overset{(a)}{\le} \frac{1}{m}\sum_{i=1}^m w_i \left\|\langle A_i, X + X'\rangle\right\|_{\psi_2} \left\|\langle A_i, X - X'\rangle\right\|_{\psi_2} \\
&\le \frac{1}{m}\left(\sum_{i=1}^m w_i\right)\left\|X + X'\right\|_F \left\|X - X'\right\|_F \\
&\le \frac{2\sqrt{p}}{m}\left(\sum_{i=1}^m w_i\right)\left\|X - X'\right\|_F,
\end{aligned} \tag{35}$$

where in (a), we used Lemma 14. Therefore, $\frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2$ is a ψ_1-process. On the other hand, we have $\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2\right] = \frac{1}{m}\sum_{i=1}^m w_i$. Next, we apply Theorem 7 to control the concentration of $\sup_{X\in\mathcal{G}(n,p)} \frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2$. For notational simplicity, we denote $\beta = \frac{2\sqrt{p}}{m}\sum_{i=1}^m w_i$. We have

$$\mathbb{P}\left(\sup_{X\in\mathcal{G}(n,p)} \frac{1}{m}\sum_{i=1}^m \left(w_i \langle A_i, X\rangle^2 - w_i\right) \geq C\beta\mathcal{J}_1(0,D) + \beta\delta\right) \leq \exp(-\delta/D), \tag{36}$$

where the diameter satisfies $D = \sup_{X,X'\in\mathcal{G}(n,p)} \|X - X'\|_F \leq 2\sqrt{p}$ and the generalized Dudley's integral can be bounded as

$$\begin{aligned}
\mathcal{J}_1(0,D) &= \int_0^D \log\left(1 + \mathcal{N}(\mathcal{G}(n,p), \|\cdot\|_F, \epsilon)\right) d\epsilon \\
&\overset{(a)}{\leq} \int_0^{2\sqrt{p}} 2(n-p)p \log\left(\frac{c_1\sqrt{p}}{\epsilon}\right) d\epsilon \\
&\leq \int_0^2 2(n-p)p^{3/2} \log\left(\frac{c_1}{\epsilon}\right) d\epsilon \\
&\leq c_4 np^{3/2}.
\end{aligned} \tag{37}$$

Here $c_4 = 2\int_0^2 \log\left(\frac{c_1}{\epsilon}\right) d\epsilon$ is a universal constant. Then, upon choosing $\delta = c_5 np^{3/2}$ with $c_5 = Cc_4$ we have with probability at least $1 - e^{-\Omega(np)}$

$$\sup_{X\in\mathcal{G}(n,p)} \frac{1}{m}\sum_{i=1}^m w_i \langle A_i, X\rangle^2 \leq \frac{2c_5 np^2}{m}\sum_{i=1}^m w_i. \tag{38}$$

This completes the proof. □

A.3 Other Useful Bounds

Lemma 19. *Suppose that $A_1, \cdots, A_m$ are i.i.d. standard Gaussian matrices. For any sequence of scalars $w_1, \cdots, w_m$, the matrix $B = (\sum_{i=1}^m w_i^2)^{-1/2}\sum_{i=1}^m w_i A_i$ is standard Gaussian.*

Proof. It is easy to see that each element of $\sum_{i=1}^m w_i A_i$ is i.i.d. and Gaussian with mean zero and variance $\sum_{i=1}^m w_i^2$. Therefore, the elements of B are all i.i.d. with standard Gaussian distribution. □

Lemma 20. *Suppose that $A_1, \cdots, A_m$ are i.i.d. standard Gaussian matrices and $w_1, \cdots, w_m$ are nonnegative scalars. Then, for any $\delta > 0$ and a fixed matrix $X \in \mathbb{R}^{n\times n}$ with $\|X\|_F = 1$, we have*

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m w_i |\langle A_i, X\rangle| - \sqrt{\frac{2}{\pi}}\frac{1}{m}\sum_{i=1}^m w_i\right| \geq \delta\right) \leq 2\exp\left(-\frac{m^2\delta^2}{C\sum_{i=1}^m w_i^2}\right). \tag{39}$$

Proof. Note that $\langle A_i, X\rangle \sim \mathcal{N}(0,1)$ and $\mathbb{E}[|\langle A_i, X\rangle|] = \sqrt{\frac{2}{\pi}}$. Due to Lemma 13, we have $\||\langle A_i, X\rangle| - \mathbb{E}[|\langle A_i, X\rangle|]\|_{\psi_2} \leq \||\langle A_i, X\rangle|\|_{\psi_2}$. Then final result follows from Lemma 12. □

Lemma 21. *Suppose that $A_1, \cdots, A_m$ are i.i.d. standard Gaussian matrices. We have*

$$\mathbb{P}\left(\max_{1\leq i\leq m} \geq \sqrt{2}d\right) \leq \exp\left(\log m - \Omega(d^2)\right).$$

Proof. For any fixed i and $1 \leq k, l \leq$, $(A_i)_{kl}^2 - 1$ is a sub-exponential random variable with zero mean and parameter 1. Therefore, an application of Lemma 12 implies that $\mathbb{P}(\|A_i\|_F \geq \sqrt{1+\delta}d) \leq \exp(-\Omega(d^2))$. Setting $\delta = 1$ followed by a union bound leads to the final result. □

Lemma 22 (Example 6.2 in [Wai19]). *Consider a random matrix $X \in \mathbb{R}^{n\times d}$ with i.i.d. entries drawn from $\mathcal{N}(0,1)$. The following inequalities hold with probability at least $1 - 2\exp(-n/288)$, provided that $n \geq 16d$:*

$$\|X\| \leq (4/3)\sqrt{n}, \qquad (1/6)n \leq \left\|X^\top X\right\| \leq (7/6)n.$$

Lemma 23 (Concentration of binomial random variable, [CL06]). *Suppose that X has a binomial distribution with parameters (m,p). Then, we have*

$$\mathbb{P}\left(|X - pm| \geq \delta\right) \leq 2\exp(-\delta^2/2pm). \tag{40}$$

In particular, when choosing $\delta = \frac{pm}{2}$, we have that with probability at least $1 - 2\exp(pm/4)$,

$$X \geq \frac{pm}{2}, \qquad X \leq \frac{3pm}{2}. \tag{41}$$

Lemma 24. *Suppose that X has a binomial distribution with parameters (m,p) and $0 \leq p < 1/2$. Then, we have*

$$X \leq \max\left\{\frac{3}{8}, \frac{1-(1-2p)^2}{2}\right\} \cdot m \qquad \text{with probability at least } 1 - \exp(-\Omega((1-2p)^2 m))$$

Proof. We consider two cases. First, suppose that $1/4 \leq p < 1/2$. Define $\eta = 1 - 2p$. According to Lemma 23, we have

$$\mathbb{P}\left(X \leq pm(1+\eta)\right) \geq 1 - \exp(-\Omega(\eta^2 m)) \implies \mathbb{P}\left(X \leq \frac{1-(1-2p)^2}{2}\right) \geq 1 - \exp(-\Omega((1-2p)^2 m)),$$

where the last inequality follows from the fact that $pm(1+\eta) = \frac{1-(1-2p)^2}{2}$. On the other hand, for $0 \leq p \leq 1/4$, Lemma 23 implies that

$$\mathbb{P}\left(X \leq 3m/8\right) \geq 1 - \exp(-\Omega(m)).$$

Combining the above two cases leads to the desired result. □

B Proof of Lemma 9

Define $r_0 = \min\{k-r, d/2\}$ and let $\mathcal{C} = (\gamma^2/\sqrt{r_0})\mathcal{G}(d, k-r)$. It is easy to see that if $X \in \mathcal{C}$, then $X = UU^\top$ for some $U \in \mathcal{B}_F^{k-r}(\gamma)$. Consider $\bar{X} = \arg\min_{X\in\mathcal{C}} A(X)$ and let $\bar{X} = \bar{U}\bar{U}^\top$. Note that $\bar{U} \in \mathcal{B}_F^{k-r}(\gamma)$. We prove the desired inequality is attained at $\bar{U}$.

Bounding $C(\bar{X})$: According its definition, $\bar{X}$ is independent of $\{A_i\}_{i\notin\mathcal{S}_t}$. Let $\{B_i\}_{i\in\mathcal{S}_t}$ be independent copies of the measurement matrices that are generated according to the matrix sensing model (Assumption 1). We have

$$\begin{aligned}
C(\bar{X}) &= \frac{1}{m}\sum_{i\notin\mathcal{S}_t} q(2t_0)^{q-1}|\left\langle A_i, \bar{X}\right\rangle| \\
&\le q(2t_0)^{q-1}\left(\frac{1}{m}\sum_{i\notin\mathcal{S}_t}|\left\langle A_i, \bar{X}\right\rangle| + \frac{1}{m}\sum_{i\in\mathcal{S}_t}|\left\langle B_i, \bar{X}\right\rangle|\right) \\
&\le q(2t_0)^{q-1}\gamma^2
\end{aligned}$$

with probablity at least $1-\exp(-\Omega(m))$, where in the last inequality we invoked Lemma 20 with $\delta = 1-\sqrt{\frac{2}{\pi}}$ and $w_i = 1$.

Bounding $A(\bar{X})$. Recall that

$$A(\bar{X}) = \inf_{X\in\mathcal{C}}\left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t} q\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}\left\langle A_i, X\right\rangle\right\}.$$

Note that, unlike $C(\bar{X})$, the choice of $\bar{X}$ depends on the measurement matrices involved in $A(\bar{X})$. To control this dependency, consider $\tilde{A} = -\left(\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}\right)^{-1/2}\sum_{i\in\mathcal{S}_t}\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}A_i$. Invoking Lemma 19 with $\omega_i = -\mathrm{Sign}(\epsilon_i)^q\epsilon_i^{q-1}$, we know that $\tilde{A}$ is a standard Gaussian matrix. Therefore, we have

$$\begin{aligned}
A(\bar{X}) &= \frac{q}{m}\left(\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}\right)^{1/2}\inf_{X\in\mathcal{C}}\left\langle -\tilde{A}, X\right\rangle \\
&= -\frac{q}{m}\left(\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}\right)^{1/2}\sup_{X\in\mathcal{C}}\left\langle \tilde{A}, X\right\rangle \\
&= -\frac{q}{m}\left(\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}\right)^{1/2}\frac{\gamma^2}{\sqrt{r_0}}\sup_{X\in\mathcal{G}(d,r_0)}\left\langle \tilde{A}, X\right\rangle \\
&\le -cq\left(\frac{dr_0\sum_{i\in\mathcal{S}_t}\epsilon_i^{2(q-1)}}{m^2}\right)^{1/2}\cdot\gamma^2,
\end{aligned}$$

with probability at least $1-\exp -\Omega(dr_0)$, where the last inequality follows from Lemma 17.

Bounding $B(\bar{X})$: Note that $B(\bar{X}) = 0$ if $q < 2$. Therefore, we assume $q \ge 2$. We have

$$\begin{aligned}
B(\bar{X}) &\le 2^{q-2}\binom{q}{2}\cdot\sup_{X\in\mathcal{C}}\left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{q-2}\left\langle A_i, X\right\rangle^2\right\} \\
&\le 2^{q-2}\binom{q}{2}\frac{\gamma^4}{r_0}\cdot\sup_{X\in\mathcal{G}(d,r_0)}\left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{q-2}\left\langle A_i, X\right\rangle^2\right\}.
\end{aligned}$$

To control $\sup_{X\in\mathcal{G}(d,r_0)}\left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{q-2}\langle A_i,X\rangle^2\right\}$, we apply Lemma 18 with $w_i=|\epsilon_i|^{q-2}$ and $n=|\mathcal{S}_t|$ to obtain

$$\begin{aligned}
\sup_{X\in\mathcal{G}(d,r_0)}\left\{\frac{1}{m}\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{q-2}\langle A_i,X\rangle^2\right\} &\le c'\cdot\frac{dr_0^2}{m}\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{q-2}\\
&\overset{(a)}{\le} c'\cdot\frac{dr_0^2}{m}|\mathcal{S}_t|^{\frac{q}{2(q-1)}}\left(\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{2(q-1)}\right)^{\frac{q-2}{2(q-1)}}\\
&\le c'\cdot\frac{dr_0^2}{m^{\frac{q-2}{2(q-1)}}}\cdot t_0^{q-2}\left(\sum_{i\in\mathcal{S}_t}|\epsilon_i/t_0|^{2(q-1)}\right)^{\frac{q-2}{2(q-1)}}\\
&\le c'\cdot\frac{dr_0^2}{m^{\frac{q-2}{2(q-1)}}}\cdot t_0^{q-2}\left(\sum_{i\in\mathcal{S}_t}|\epsilon_i/t_0|^{2(q-1)}\right)^{1/2}\\
&= c'\cdot\frac{dr_0^2}{m^{\frac{q-2}{2(q-1)}}}\cdot t_0^{-1}\left(\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{2(q-1)}\right)^{1/2}
\end{aligned}$$

with probability at least $1-\exp(-\Omega(dr_0))$. Here, (a) holds with strict equality for $q=2$, and it follows from the auxiliary Lemma 25 with $a_i=|\epsilon_i|^{q-2}$ and $\alpha=\frac{2(q-1)}{q-2}$ for $q>2$. Combining the above two inequalities yields

$$\begin{aligned}
B(\bar{X}) &\le c'\cdot 2^{q-2}\binom{q}{2}\frac{\gamma^4}{r_0}\cdot\frac{dr_0^2}{m^{\frac{q-2}{2(q-1)}}}\cdot t_0^{-1}\left(\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{2(q-1)}\right)^{1/2}\\
&\le c(q/2)\left(\frac{dr_0\sum_{i\in\mathcal{S}_t}|\epsilon_i|^{2(q-1)}}{m^2}\right)^{1/2}\cdot\gamma^2
\end{aligned}$$

where $c>0$ is the same constant appeared in the upper bound of $A(\bar{X})$. Furthermore, the last inequality follows from $\gamma^2\le\left(\frac{4c}{c'}\right)\frac{t_0}{q\cdot 2^q\sqrt{dr_0}m}$. This completes the proof of this lemma. □

C Auxiliary Lemmas

Lemma 25. *For any $a_1,\ldots,a_n\ge 0$ and $\alpha\ge 1$, we have $\sum_{i=1}^n a_i\le n^{\frac{\alpha-1}{\alpha}}\left(\sum_{i=1}^n a_i^\alpha\right)^{\frac{1}{\alpha}}$.*

Proof. The proof follows readily follows from the convexity of $f(a)=|a|^\alpha$. The details are omitted for brevity. □

Lemma 26. *For any $a\ge b\ge 0$ and $q\ge 1$, we have $a^q-(a-b)^q\ge\frac{q}{q+1}a^{q-1}b$.*

Proof. We have

$$a^q-(a-b)^q=a^q\left(1-(1-b/a)^q\right)\ge a^q\left(1-\frac{1}{1+qb/a}\right)=a^q\cdot\frac{qb}{a+qb}\ge\frac{q}{q+1}a^{q-1}b \tag{42}$$

where in the first inequality we used Bernoulli and in the second inequality we used the fact that $a\ge b$. □

Lemma 27. *For any $a, b \geq 0$, we have*

$$\sum_{k=2}^{q} \binom{q}{k} a^{q-k} b^k \leq \binom{q}{2} (a+b)^{q-2} b^2.$$

Proof. We have

$$\begin{aligned}
\binom{q}{2}(a+b)^{q-2}b^2 &= \binom{q}{2}\left(\sum_{k=0}^{q-2}\binom{q-2}{k}a^{q-2-k}b^k\right)b^2 \\
&= \sum_{k=0}^{q-2}\binom{q}{2}\binom{q-2}{k}a^{q-2-k}b^{k+2} \\
&= \sum_{k=2}^{q}\binom{q}{2}\binom{q-2}{k-2}a^{q-k}b^k \\
&\geq \sum_{k=2}^{q}\binom{q}{k}a^{q-k}b^k.
\end{aligned}$$

□