Geometric Analysis of Matrix Sensing over Graphs

Haixiang Zhang
Department of Mathematics
University of California, Berkeley
Berkeley, CA 94720
haixiang_zhang@berkeley.edu

Ying Chen
Department of IEOR
University of California, Berkeley
Berkeley, CA 94720
ying-chen@berkeley.edu

Javad Lavaei
Department of IEOR
University of California, Berkeley
Berkeley, CA 94720
lavaei@berkeley.edu

Abstract

In this work, we consider the problem of matrix sensing over graphs (MSoG). As a general case of matrix completion and matrix sensing problems, the MSoG problem has not been analyzed in the literature and the existing results cannot be directly applied to the MSoG problem. This work provides the first theoretical results on the optimization landscape of the MSoG problem. More specifically, we propose a new condition, named the Ω-RIP condition, to characterize the optimization complexity of the problem. In addition, with an improved regularizer of the incoherence, we prove that the strict saddle property holds for the MSoG problem with high probability under the incoherence condition and the Ω-RIP condition, which guarantees the polynomial-time global convergence of saddle-avoiding methods. Compared with state-of-the-art results, the bounds in this work are tight up to a constant. Besides the theoretical guarantees, we numerically illustrate the close relation between the Ω-RIP condition and the optimization complexity.

1 Introduction

In a wide range of problems in the fields of machine learning, signal processing and power systems, an unknown low-rank matrix parameter should be estimated from a few measurements of the matrix. To be more specific, given some measurements of the unknown symmetric and positive semi-definite (PSD) matrix $M^* \in \mathbb{R}^{n\times n}$ of rank $r \ll n$, the *low-rank matrix optimization* problem can be formulated as

$$\min_{M\in\mathbb{R}^{n\times n}} f(M;M^*) \quad \mathrm{s.t.} \quad M \succeq 0, \quad \mathrm{rank}(M) \leq r, \tag{1}$$

where $f(\cdot;M^*)$ is a loss function that penalizes the mismatch between M and M^*. The goal is to recover the matrix M^* via finding a global minimizer of problem (1). Applications of this problem include matrix sensing [31, 40, 38], matrix completion [10, 11, 17], phase retrieval [8, 33, 14], and power systems [39, 24]; see the review papers [13, 15] for more applications. Early attempts to deal with the nonconvex rank constraint of the problem focused on solving a convex relaxation of (1); see [10, 31, 11, 7]. However, the convex relaxation approach usually updates the matrix variable via the Singular Value Decomposition (SVD) in each iteration. This will lead to an $O(n^3)$ computational complexity in each iteration and an $O(n^2)$ space complexity, which are prohibitively

Preprint. Under review.

high for large-scale problems; see the numerical comparison in [41]. Similar issues are observed for algorithms based on the Singular Value Projection [20] and Riemannian optimization algorithms [35, 36, 19, 1, 27].

To improve the computation and memory efficiency, the Burer-Monteiro factorization approach was proposed in [6], which is based on the fact that the mapping $U \mapsto UU^T$ is surjective onto the manifold of PSD matrices of rank at most r, where $U \in \mathbb{R}^{n\times r}$. More concretely, problem (1) is equivalent to

$$\min_{U\in\mathbb{R}^{n\times r}} f(UU^T; M^*). \tag{2}$$

Due to the non-convexity of the mapping $U \mapsto UU^T$, problem (2) is an unconstrained non-convex problem and may have spurious second-order critical points (i.e., second-order critical points that do not correspond to the ground truth matrix M^*). In general, saddle-avoiding local search methods are only able to find ϵ-approximate second-order critical points[1]. As a result, local search methods with a random initialization will likely be stuck at spurious second-order critical points and unable to converge to the ground truth solution. However, in a variety of real-world applications, simple algorithms such as perturbed gradient descent methods and alternating minimization methods have achieved empirical success on problem (2). Recently, substantial progress has been made on the theoretical explanation of the benign behavior of these algorithms. For example, the alternating minimization algorithm was studied in [21, 28, 29]. The (stochastic) gradient descent algorithm, which is in general easier to implement than the alternating minimization algorithm, was analyzed in [8, 34, 37, 14, 13]. Moreover, the gradient descent algorithm is proved to have the implicit regularization phenomenon in the over-parameterization case [26, 16, 32].

Besides the algorithmic analysis, a large amount of literature [17, 33, 42, 38] focused on the geometric analysis of the landscape of problem (2), which usually depends on the *strict-saddle property* [33]. Intuitively, the strict-saddle property states that at any feasible point of problem (2), at least one of the three properties will hold: (i) the point is close to a global solution; (ii) the norm of the gradient is large; (iii) the Hessian matrix has a negative eigenvalue. In the later two cases, saddle-escaping algorithms [12, 23, 2] are able to find a descent direction and thus, these algorithms will converge globally in polynomial time. The formal definition of the strict-saddle property is provided in Section 3.1. In the following, we review the state-of-the-art conditions for two special classes of problem (2) that guarantee the strict-saddle property; see the survey [15] for other problem classes.

1.1 Matrix sensing and the Restricted Isometry Property (RIP) condition

In the matrix sensing problem, the information of the ground truth matrix M^* is gathered via the measurement operator $\mathcal{A}$. The loss function f is usually chosen to be the negative log-likelihood function. For example, in the classic linear matrix sensing problem, the operator $\mathcal{A}$ is a linear operator defined as

$$\mathcal{A}(M) := [\langle A_1, M\rangle, \dots, \langle A_m, M\rangle], \quad \forall M \in \mathbb{R}^{n\times n}, \tag{3}$$

where m is the number of measurements and $A_i \in \mathbb{R}^{n\times n}$ contains independently identically distributed (i.i.d) Gaussian random entries and A_i's are independent of each other. If the measurement noise is Gaussian, the maximum likelihood estimation is equivalent to the following minimization problem

$$\min_{U\in\mathbb{R}^{n\times r}} \|\mathcal{A}(UU^T) - \mathcal{A}(M^*)\|_F^2. \tag{4}$$

More examples of the (non-linear) matrix sensing problem are discussed in [42]. One of the most important conditions that guarantee the benign landscapes is the RIP condition:

Definition 1 (RIP Condition[31, 42])**.** *Given natural numbers r and s, the function $f(\cdot; M^*)$ is said to satisfy the **Restricted Isometry Property** (RIP) of rank $(2r, 2s)$ for a constant $\delta \in [0, 1)$, denoted as δ-RIP$_{2r,2s}$, if*

$$(1-\delta)\|K\|_F^2 \le \left[\nabla^2 f(M; M^*)\right](K,K) \le (1+\delta)\|K\|_F^2 \tag{5}$$

holds for all matrices $M, K \in \mathbb{R}^{n\times n}$ such that $\operatorname{rank}(M) \le 2r$ and $\operatorname{rank}(K) \le 2s$, where $\left[\nabla^2 f(M; M^*)\right](K,K)$ is the curvature of the Hessian at point M along direction K.

[1] A point x_0 is called an ϵ-approximate second-order critical point to the optimization problem $\min_x F(x)$ if $\|\nabla F(x_0)\|_F \le \epsilon$ and $\lambda_{min}[\nabla^2 F(x_0)] \ge -\epsilon$.

For the linear matrix sensing problem (4), it is proved in [9] that the δ-RIP$_{2r,2s}$ condition holds with high probability if $m = \Theta(nr\delta^{-2})$. The RIP condition is also established in many other applications of the matrix sensing problem [42, 4] and is of independent research interest. The constant δ plays a critical role in characterizing the optimization landscape of problem (2). More specifically, in [5], the authors showed that the strict-saddle property holds for problem (2) if the δ-RIP$_{2r,2r}$ condition holds with $\delta < 1/2$. Counterexamples have been constructed in [40, 38] to illustrate that the strict-saddle property may fail under the δ-RIP$_{2r,2r}$ condition with $\delta \geq 1/2$.

1.2 Matrix completion and the incoherence condition

In spite of the strong theoretical results under the RIP condition, there exist a large number of applications that do not satisfy the RIP condition. A well-studied class of problems that does not satisfy the RIP condition is the matrix completion problem. For the matrix completion problem, a subset $\Omega \subset [n] \times [n]$ of entries of M^* are observed and the goal is to recover the low-rank ground truth matrix from the observed entries. For every matrix $M \in \mathbb{R}^{n\times n}$, we denote the projection of M onto Ω as M_Ω, namely,

$$(M_\Omega)_{ij} = \begin{cases} M_{ij} & \text{if } (i,j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

Using the ℓ_2-loss function, the matrix completion problem is defined as

$$\min_{U\in\mathbb{R}^{n\times r}} \|M_\Omega - M^*_\Omega\|_F^2. \tag{6}$$

The matrix completion problem (6) is a special case of the linear matrix sensing problem (4), where the sample size m is equal to $|\Omega|$ and each measurement matrix A_i contains exactly one nonzero entry. We note that the δ-RIP$_{2r,2r}$ condition does not hold for problem (6) unless we observe all entries of M^*, i.e., when $\Omega = [n] \times [n]$. As an alternative to the RIP condition, the incoherence of M^* is useful in characterizing the complexity of problem (6).

Definition 2 (Incoherence Condition [10])**.** *Given $\mu \in [1, n]$, the ground truth matrix M^* is said to be **μ-incoherent** if*

$$\|e_i^T V^*\|_F \leq \sqrt{\mu r/n}, \quad \forall i \in [n], \tag{7}$$

where $V^*\Lambda^*(V^*)^T$ is the truncated SVD of M^* and e_i is the i-th standard basis of $\mathbb{R}^n$.

Intuitively, the incoherence of M^* measures the sparsity of the low-rank ground truth matrix. If the incoherence is large, the matrix M^* is highly sparse and it is necessary to measure considerably many entries of M^* to observe nonzero entries. On the other hand, a relatively small incoherence of M^* is able to avoid the extreme case. Except the limited literature on the deterministic matrix completion problem [3, 25, 30], the majority of the matrix completion literature considered the following Bernoulli sampling model:

Definition 3 (Bernoulli Sampling Model)**.** *Given a sampling rate $p \in (0, 1]$, each index $(i, j) \in [n] \times [n]$ belongs to the set Ω independently with probability p.*

Under the Bernoulli sampling model, the objective function of problem (6) is well-behaved over the set of matrices with a small incoherence. Therefore, a regularizer that penalizes the incoherence of UU^T is included and we instead solve the following regularized matrix completion problem:

$$\min_{U\in\mathbb{R}^{n\times r}} \frac{1}{p} \left\|(UU^T)_\Omega - M^*_\Omega\right\|_F^2 + \lambda \sum_{i\in[n]} \left(\|e_i^T U\|_F - \alpha\right)_+^4, \tag{8}$$

where $(x)_+ := \max\{x, 0\}$ for all $x \in \mathbb{R}$ and $\alpha, \lambda > 0$ are constants. Intuitively, the coefficient $1/p$ is used to "normalize" the ℓ_2-norm. We note that there exists an algorithm that can solve problem (6) without the incoherence regularizer [13]. However, the algorithm relies on the spectral initialization, which requires more computational effort than the factorization approach (2). It is proved in [18, 17] that if the sampling rate satisfies

$$np \geq \Theta[\mu^4 r^6 (\kappa^*)^6 \log n], \tag{9}$$

the problem (8) satisfies the strict-saddle property with high probability. On the other hand, it is proved in [11] that the information-theoretical lower bound $np \geq \Theta(\mu r \log n)$ is necessary for the exact completion with high probability.

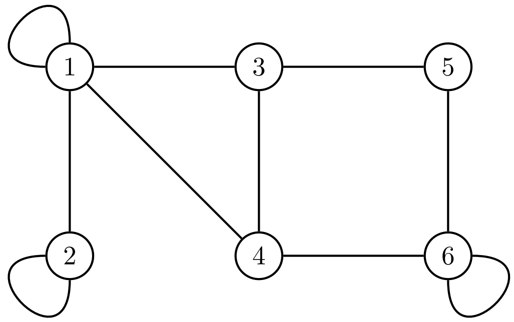

Figure 1: Example of the problem of matrix sensing over graphs. The vertex set is $\mathcal{V} = [6]$ and the edge set is $\mathcal{E} = \{(1,1),(1,2),(1,3),(1,4),(2,2),(3,4),(3,5),(4,6),(5,6),(6,6)\}$.

1.3 Motivating example: Matrix sensing over graphs

Although the matrix sensing problem and the matrix completion problem are well-studied in literature, we show that there exist important applications of (2) that belong to a much broader class of problems than the special classes previously studied by the existing works. Hence, the current theoretical results cannot be directly applied to these applications and it remains unknown whether saddle-avoiding algorithms can find the ground truth matrix M^* in polynomial time.

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the set of vertices and the set of edges, respectively. For the notational simplicity, we assume that $\mathcal{V} = [n]$. Then, the edge set $\mathcal{E}$ is a subset of $[n] \times [n]$. The goal of this problem is to recover the ground truth matrix M^* via measurements of its entries $\{M^*_{ij} \mid (i,j) \in \mathcal{E}\}$. Instead of directly observing each entry M^*_{ij}, for each node $i \in \mathcal{V}$, we observe a "mixture" (i.e., the output of a function) of the entries in the set

$$\mathcal{E}_i := \{(i,j) \mid j \text{ is incident to } i\}.$$

Denote the loss function for node i as $f_i\left(M_{\mathcal{E}_i}; M^*_{\mathcal{E}_i}\right)$. The total loss function is the sum of the loss function for all nodes; namely, we consider the problem

$$\min_{U \in \mathbb{R}^{n\times r}} \sum_{i\in[n]} f_i\left[(UU^T)_{\mathcal{E}_i}; M^*_{\mathcal{E}_i}\right]. \tag{10}$$

Since $\mathcal{E} = \cup_{i\in[n]}\mathcal{E}_i$, the above problem is a special case of the general problem

$$\min_{U \in \mathbb{R}^{n\times r}} f\left[(UU^T)_{\mathcal{E}}; M^*_{\mathcal{E}}\right], \tag{11}$$

where we define $f\left[(UU^T)_{\mathcal{E}}; M^*_{\mathcal{E}}\right] := \sum_{i\in[n]} f_i\left[(UU^T)_{\mathcal{E}_i}; M^*_{\mathcal{E}_i}\right]$. We name the problem (11) as the *matrix sensing over graphs* (MSoG) problem. The MSoG problem has a number of applications, including the state estimation problem in power systems [39, 24]. We note that in those applications, the loss function f_i is a quadratic function of entries of $M_{\mathcal{E}_i}$ and $M^*_{\mathcal{E}_i}$; see, for example, the graph-structured quadratic sensing problem [24].

To better understand (10) as a special type of MSoG, we consider a toy example where the undirected graph $\mathcal{G}$ has $n = 6$ vertices and is plotted in Figure 1. In this example, the objective function of the MSoG problem is

$$\begin{aligned} f(M_{\mathcal{E}}; M^*_{\mathcal{E}}) = {} & f_1(M_{11}, M_{12}, M_{13}; M^*_{11}, M^*_{12}, M^*_{13}) + f_2(M_{21}, M_{22}; M^*_{21}, M^*_{22}) \\ & + f_3(M_{31}, M_{34}, M_{35}; M^*_{31}, M^*_{34}, M^*_{35}) + f_4(M_{41}, M_{43}, M_{46}; M^*_{41}, M^*_{43}, M^*_{46}) \\ & + f_5(M_{53}, M_{56}; M^*_{53}, M^*_{56}) + f_6(M_{64}, M_{65}, M_{66}; M^*_{64}, M^*_{65}, M^*_{66}), \end{aligned}$$

where we recall that we factorize M into UU^T in the Burer-Monteiro factorization approach. Hence, only 13 entries of the matrix M^* with $n^2 = 36$ entries appear in the measurements. For example, the entry M^*_{16} does not appear in any measurements. If we directly observe these 13 entries, we can define the loss function f_1 by the ℓ_2-loss function:

$$f_1(M_{11}, M_{12}, M_{13}; M^*_{11}, M^*_{12}, M^*_{13}) = (M_{11} - M^*_{11})^2 + (M_{12} - M^*_{12})^2 + (M_{13} - M^*_{13})^2.$$

We can define other loss functions $f_2, \dots, f_6$ in a similar way. Then, the MSoG problem reduces to the matrix completion problem with $\Omega = \mathcal{E}$, namely, the objective function becomes

$$f(M_{\mathcal{E}}; M^*_{\mathcal{E}}) = \|M_{\mathcal{E}} - M^*_{\mathcal{E}}\|_F^2.$$

Therefore, the matrix completion problem (6) is a special case of the MSoG problem (11) and the existing results for the matrix completion problem cannot be directly applied to problem (11). Similarly, if Ω is a complete graph, the matrix sensing problem (10) becomes a special case of the MSOG problem. Moreover, since some entries of M^* do not appear in any of the measurements in general, it is easy to verify that the RIP condition (5) does not hold for problem (11). In summary, the existing results based on the RIP condition and the incoherence condition, as well as the results for other applications of the low-rank matrix optimization problem, cannot be directly applied to the MSoG problem.

1.4 Problem formulation and contributions

In this work, we propose the first sufficient condition that guarantees the benign landscape of the MSoG problem. More specifically, we consider the case when the graph $\mathcal{G}$ is a random graph obeying the Erdös–Rényi model. Namely, each pair $(i, j) \in \mathcal{V} \times \mathcal{V}$ belongs to the edge set independently with probability p. Under this random graph model and the assumption that the vertex set $\mathcal{V}$ is $[n]$, each entry M^*_{ij} is indirectly observed (i.e., is involved in some measurements) independently with probability p. For comparison with the existing results of the matrix completion problem (6), we denote the edge set, which is equivalent to the set of indices of observed entries, as $\Omega \subset [n] \times [n]$. Then, the set Ω follows the Bernoulli sampling model (Definition 3). Similar to the matrix completion problem, we aim at recovering the ground truth matrix M^* by solving the following problem with an improved regularizer:

$$\min_{U \in \mathbb{R}^{n \times r}} \frac{1}{p} f\left[\left(UU^T\right)_\Omega ; M^*_\Omega\right] + \sum_{i \in [n]} r\left(\|e_i^T U\|_F\right), \tag{12}$$

where we define the regularizer

$$r(x) := \lambda \int_{-1}^{1} \left[(x + \alpha y - 10\alpha)_+ + 9\alpha\right]^4 (1 - |y|)\ dy, \quad \forall x \in \mathbb{R}.$$

Note that in problem (12), we use a novel incoherence regularizer that is different from those in the prior literature. The regularizer is the convolution between $\left[(x + \alpha y - 10\alpha)_+ + 9\alpha\right]^4$ and the probability density function $1 - |y|$ on $[-1, 1]$, which is twice continuously differentiable in x (the set of discontinuous points of the derivatives is a zero measure set). Hence, the regularizer $r\left(\|e_i^T U\|_F\right)$ is twice continuously differentiable in U. In addition, we can exchange the convolution and the differentiation with respect to x. Since the integrand is a quartic polynomial,tThe objective value and the derivatives of the regularizer can be exactly evaluated by numerical integration schemes with $O(1)$ computations. Besides the randomness model of Ω, we make the following three assumptions about problem (12).

Assumption 1. *The loss function f is twice continuously differentiable.*

Assumption 2. *The ground truth matrix M^* is PSD and rank-r.*

Assumption 3. *The ground truth matrix M^* is a global minimizer of $f[(\cdot)_\Omega; M^*_\Omega]$.*

We note that these assumptions are standard in the low-rank optimization literature. The first assumption is a mild regularity assumption on the loss function and is satisfied by a wide range of loss functions, including the ℓ_2-loss and the negative log-likelihood function of various probability distributions. The second assumption is based on the prior knowledge about the specific application. The third assumption is necessary for the exact recovery of the ground truth M^*.

We give an informal statement of the main result in the following theorem. Here, we consider the (δ, Ω)-RIP$_{2r,2r}$ condition, which is an extension of the classic δ-RIP$_{2r,2r}$ condition. Intuitively, the constant $\delta \in [0, 1)$ measures the similarity between f and the ℓ_2-loss function. The rigorous definition is provided in Definition 4.

Theorem 1 (Informal)**.** *Suppose that the loss function f satisfies the $(1/16, \Omega)$-RIP$_{2r,2r}$ condition with a high probability over Ω and the sampling rate p satisfies*

$$np \geq \Theta[\mu^2 r^3 (\kappa^*)^2 \log n],$$

where κ^* is the condition number of M^*. Then, with a suitable choice of the parameters α and λ, problem (12) satisfies the strict saddle property with high probability. Furthermore, there exist *algorithms that can find a solution U_0 such that $\|U_0U_0^T - M^*\|_F \leq \epsilon$ in polynomial time with the same probability.*

The above theorem provides the first theoretical result for the MSoG problem. Basically, Theorem 1 says that a combination of the incoherence condition and the Ω-RIP condition is sufficient to guarantee that problem (12) can be solved in polynomial time with high probability. In addition, the lower bound on the sampling rate is better than the bound in [17] (i.e., the bound in (9)). This is a result of our improved regularizer. Finally, the upper bound on the Ω-RIP constant is an absolute constant (namely, $1/16$).

Remark 1. *We note that since the result of Theorem 1 only holds with high probability, it is not necessary for the Ω-RIP condition to hold for all subsets Ω. Instead, we only require the Ω-RIP condition to hold for a set of Ω that correspond to a high probability over the distribution of Ω. In the special case when Ω is a deterministic subset that satisfies similar "benign" properties as in the random graph case, we only need the Ω-RIP condition for this deterministic subset Ω.*

The next informal theorem shows that our upper bound is optimal up to a constant.

Theorem 2 (Informal)**.** *There exists a loss function f such that: (i) f satisfies the $(1/2, \Omega)$-RIP$_{2r,2r}$ condition for all $\Omega \subset [n] \times [n]$, (ii) for all $p \in [0, 1]$, problem (12) has a spurious second-order critical point[2] with probability at least $(3 - \sqrt{5})/2 \approx 0.38$.*

Intuitively, Theorem 2 provides a negative result saying that the tightest upper bound on the RIP constant cannot be better than $1/2$ if the problem (12) has a benign landscape with high probability. This is also the first negative result on the problem (12).

1.5 Notation

For every natural number n, we denote $[n] := \{1, \ldots, n\}$. The operator 2-norm and the Frobenius norm of a matrix M are denoted as $\|M\|_2$ and $\|M\|_F$, respectively. The trace of matrix M is denoted as $\mathrm{tr}(M)$. The inner product between two matrices is defined as $\langle M, N\rangle := \mathrm{tr}(M^TN)$. For any matrix $M \in \mathbb{R}^{n\times n}$, we denote its singular values by $\sigma_1(M) \geq \cdots \geq \sigma_k(M)$. Let σ_i^* be the singular value $\sigma_i(M^*)$. The condition number of M^* is $\kappa^* = \sigma_1^*/\sigma_r^*$. The minimum eigenvalue of matrix M is denoted as $\lambda_{min}(M)$. For any two matrices $A, B \in \mathbb{R}^{n\times m}$, we use $A \otimes B$ to denote the fourth-order tensor whose (i, j, k, ℓ) element is $A_{i,j}B_{k,\ell}$. The identity tensor is denoted as $\mathcal{I}$. The notation $M \succeq 0$ means that the matrix M is PSD. The sub-matrix $R_{i:j,k:\ell}$ consists of the i-th to the j-th rows and the k-th to the ℓ-th columns of matrix R. The action of the Hessian $\nabla^2 f(M)$ on any two matrices K and L is given by $[\nabla^2 f(M)](K, L) := \sum_{i,j,k,\ell}[\nabla^2 f(M)]_{i,j,k,\ell}K_{ij}L_{k,\ell}$. The notation $f = O(g)$ means that there exists an absolute constant $C > 0$ such that $f \leq C \cdot g$. The notation $f = \Theta(g)$ means that there exist absolute constants $C_1, C_2, > 0$ such that $C_1 \cdot g \leq f \leq C_2 \cdot g$.

2 Ω-RIP Condition and Numerical Illustration

Since the objective function $f(M_\Omega; M_\Omega^*)$ does not satisfy the classic RIP condition unless all elements of M^* are observed (i.e., when $\Omega = [n] \times [n]$), it is necessary to consider a generalization of the RIP condition to the partial observation case.

Definition 4 (Ω-RIP Condition)**.** *Given a subset $\Omega \subset [n] \times [n]$ and natural numbers r, s, the function $f(\cdot; M^*)$ is said to satisfy the Ω-**Restricted Isometry Property** (Ω-RIP) of rank $(2r, 2s)$ for a constant $\delta \in [0, 1)$, denoted as (δ, Ω)-RIP$_{2r,2s}$, if*

(13) $$(1 - \delta)\|K_\Omega\|_F^2 \leq \left[\nabla^2 f[M_\Omega; (M^*)_\Omega]\right](K_\Omega, K_\Omega) \leq (1 + \delta)\|K_\Omega\|_F^2$$

holds for all matrices $M, K \in \mathbb{R}^{n\times n}$ such that $\mathrm{rank}(M) \leq 2r, \mathrm{rank}(K) \leq 2s$.

Note that the matrix completion problem satisfies the $(0, \Omega)$-RIP$_{2r,2s}$ condition. In the following, we demonstrate another example for which the Ω-RIP condition holds with a non-zero constant δ.

[2] A point U_0 is called a spurious second-order critical point of problem (12) if U_0 satisfies the first-order and the second-order necessary optimality conditions but $U_0U_0^T \neq M^*$.

Example 1 (Linear matrix sensing over graphs)**.** *In the linear matrix sensing over graphs problem, the loss function is defined as*

$$f(M_\Omega; M^*_\Omega) := \|\mathcal{A}(M_\Omega) - \mathcal{A}(M^*_\Omega)\|_F^2,$$

where the linear operator $\mathcal{A} : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^m$ *(defined in* (3)*) is generated by Gaussian measurements and* m *is the number of measurements. For all subsets* $\Omega \subset [n] \times [n]$*, a similar proof to that of Theorem 2.3 of [9] implies that the* (δ, Ω)*-RIP*$_{2r,2r}$ *condition holds with high probability when* $m \geq cnr/\delta^2$ *for some constant* $c > 0$*. The intuition behind the proof is that the constructed* ϵ*-net for the linear matrix sensing problem is also an* ϵ*-net for the linear MSoG problem since* $\|M_\Omega - M'_\Omega\|_F \leq \|M - M'\|_F$ *for all* $M, M' \in \mathbb{R}^{n\times n}$*.*

Remark 2. *In the above example and other examples in practice, the* Ω*-RIP condition holds with high probability for a fixed subset* Ω*. Therefore, we can focus on the event when the* Ω*-RIP condition holds since otherwise the results will only differ by a sufficiently small probability.*

Next, we show how the optimization complexity is related to the Ω-RIP constant δ and the sampling rate p via a numerical example. Here, the optimization complexity refers to the probability that the randomly initialized gradient descent algorithm can find the ground truth matrix M^*. In this example, we choose a random orthogonal matrix $V \in \mathbb{R}^{n\times n}$ and define the loss function to be

$$f_c[M_\Omega; (VM^*V^T)_\Omega] := \frac{1}{2}[M - (VM^*V^T)]_\Omega : (c\cdot\mathcal{I} + \mathcal{H}) : [M - (VM^*V^T)]_\Omega, \ \forall M \in \mathbb{R}^{n\times n},$$

where $c \in \mathbb{R}$ is a hyper-parameter and the tensor $\mathcal{H}$ and the ground truth M^* are defined in Section 3.2. In addition, by a similar analysis as in Section 3.2, we can prove that the function f_c satisfies the $(1/2, \Omega)$-RIP$_{2r,2r}$ condition if we choose $c = 3/2$. We also numerically verify this conclusion by checking the curvature $[\nabla^2 f_{3/2}(M_\Omega; M^*_\Omega)](K, K)$ along 10^4 random directions $K \in \mathbb{R}^{n\times n}$. Since $f_{3/2}$ is a quadratic function, the $(1/2, \Omega)$-RIP$_{2r,2r}$ condition is given by

$$\frac{1}{2}\|K_\Omega\|_F^2 \leq K_\Omega : \left(\frac{3}{2}\cdot\mathcal{I} + \mathcal{H}\right) : K_\Omega \leq \frac{3}{2}\|K_\Omega\|_F^2, \quad \forall K \in \mathbb{R}^{n\times n},$$

where we define the tensor multiplication $K : \mathcal{H}' : K = \sum_{i,j,k,\ell\in[n]} \mathcal{H}'_{ijk\ell} K_{ij} K_{k\ell}$ for all $K \in \mathbb{R}^{n\times n}$ and fourth-order tensor $\mathcal{H}' \in \mathbb{R}^{n\times n\times n\times n}$. In addition, since the identity tensor $\mathcal{I}$ satisfies the $(0, \Omega)$-RIP$_{2r,2r}$ condition, it is straightforward to prove that the function f_c satisfies the (δ, Ω)-RIP$_{2r,2r}$ condition with $\delta = 1/(2c - 1)$ for all $c \geq 1$.

We choose the problem size to be $n = 10$ and $r = 5$. The regularization parameters are $\alpha = 10$ and $\lambda = 100$. The set of sampling rates and Ω-RIP$_{2r,2r}$ constants are

$$p \in \{0.7, 0.75, \dots, 0.95, 1.0\}, \quad \delta \in \{0.2, 0.25, \dots, 0.75, 0.8\}.$$

We solve each problem instance by the Burer-Monterio factorization and the perturbed accelerated gradient descent algorithm [23], where the constant step size is $0.007/c$. We generate 100 independent problem instances and compute the success rate of the gradient descent algorithm with random initialization. We say that the algorithm successfully solves the instance if the generalization error $\|UU^T - M^*\|_F$ is less than $10^{-3} \times c$. If this condition fails, it means that the algorithm is stuck at a spurious local solution.

The results are plotted in Figure 2. We can see from the figure that the optimization complexity grows when δ becomes smaller and when p becomes larger. This result shows that the Ω-RIP condition plays an important role in characterizing the optimization complexity of problem (12). To be more concrete, we expect that an upper bound on the Ω-RIP constant will be able to guarantee the benign optimization landscape of problem (12). Moreover, the result is consistent with the results for the matrix sensing problem and the matrix completion problem. Furthermore, we can see that when the Ω-RIP constant is smaller than $1/2$, the success rate has a stronger correlation with p than δ. This observation is also reflected in Theorem 1, where the lower bound on the sampling rate is on the same order as those in the existing works [18, 17] if δ is upper bounded. For the cases when p is close to 1 and δ is larger than 0.5, the success rate is better than the case when $p = 1$. One possible explanation for this phenomenon is that the loss function f_c has multiple global minima when the sampling rate is smaller than 1. Instead of converging to a spurious local solution, the implicit regularization [26] makes the perturbed gradient descent algorithm more likely to converge to the global solution with the "minimal complexity", which is likely the ground truth VM^*V^T. If the sampling rate is much smaller than 1, the number of spurious local minima will increase and the perturbed gradient descent algorithm may stuck at spurious local minima and fail to converge to the ground truth.

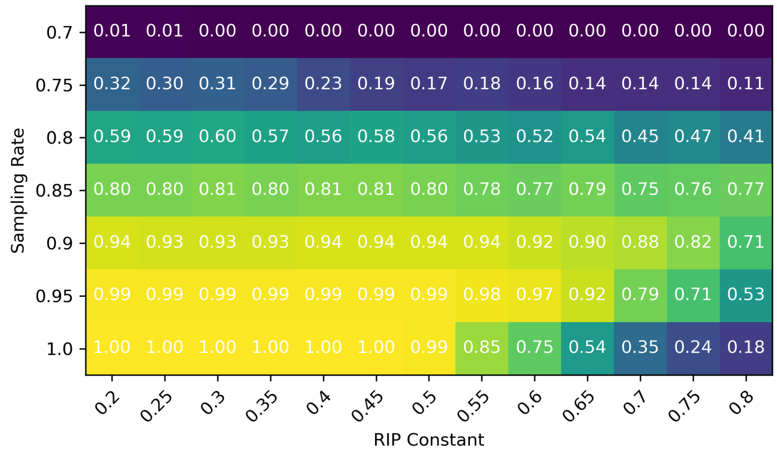

Figure 2: Success rate of the algorithm with different sampling rate p and RIP constant δ.

3 Theoretical Results

In this section, we provide strong theoretical results on the MSoG problem (12). We first develop a sufficient condition on the benign landscape of problem (12) and then study the tightness of our developed condition.

3.1 Global landscape: Strict saddle property

First, we develop conditions under which problem (12) does not have any spurious second-order critical points and therefore saddle-escaping methods (e.g., [22, 23]) can find an approximate global minimum in polynomial time. To guarantee the global convergence, the strict saddle property is commonly considered in the literature:

Definition 5 (Strict Saddle Property [33])**.** *Consider an optimization problem* $\min_{x\in\mathcal{X}\subset\mathbb{R}^d} F(x)$ *and let* $\mathcal{X}^*$ *denote the set of its global minima. We say that the problem satisfies the* (θ,β,γ)*-**strict saddle property** for* $\theta,\beta,\gamma>0$ *if at least one of the following conditions is satisfied for every* $x\in\mathcal{X}$*:*

$$\mathrm{dist}(x,\mathcal{X}^*)\le\theta;\quad \|\nabla F(x)\|_F\ge\beta;\quad \lambda_{min}[\nabla^2F(x)]\le-\gamma,$$

where $\mathrm{dist}(x,\mathcal{X}^*):=\inf_{x^*\in\mathcal{X}^*}\|x-x^*\|_F$ *is the distance between* x *and* $\mathcal{X}^*$*.*

For the low-rank optimization problem, the distance in the factorization space is equivalent to the distance in the matrix space in the sense that there exist constants $c_1(\mathcal{X}^*),c_2(\mathcal{X}^*)>0$ such that

$$c_1(\mathcal{X}^*)\cdot\|U-U^*\|_F\le\|UU^T-U^*(U^*)^T\|_F\le c_2(\mathcal{X}^*)\cdot\|U-U^*\|_F$$

holds for all $U\in\mathcal{X}$ as long as $\|U-U^*\|_F$ is small and $\mathcal{X}^*$ is bounded [34]. Denote the objective function of problem (12) as $\ell(U)$. As an example of saddle-avoiding methods, the accelerated perturbed gradient descent algorithm [23] can find a point U_0 such that

$$\|\nabla\ell(U_0)\|_F=O(\epsilon),\quad \lambda_{min}[\nabla^2\ell(U_0)]=-O(\sqrt{\epsilon})$$

in $O(\epsilon^{-1.75})$ iterations with high probability. If we choose $\epsilon>0$ to be small enough, the strict saddle property ensures that the point U_0 satisfies $\|U_0U_0^T-M^*\|_F=O(\theta)$. We note that the Lipschitz continuity of the Hessian of ℓ can be guaranteed by the regularity assumption 1 and the boundedness of trajectories of the algorithm, which can be proved in a similar way as Theorem 8 in [22]. In summary, if the strict saddle property holds, we can apply saddle-avoiding methods to achieve the polynomial-time global convergence for problem (12).

Now, we prove that the problem (12) satisfies the strict saddle property with high probability under the Ω-RIP$_{2r,2r}$ condition and the incoherence condition. Compared with Theorem 1, we assume, for the simplicity of the statement of the theorem, that the Ω-RIP condition holds for all subset Ω since otherwise, the result will only differ by a small probability.

Theorem 3. *Suppose that the loss function f satisfies the (δ, Ω)-RIP$_{2r,2r}$ condition for all $\Omega \subset [n] \times [n]$ and*

$$\alpha^2 = \Theta\left(\frac{\mu r \sigma_1^*}{n}\right), \quad \lambda = \Theta\left[\frac{(\sqrt{\mu} + \sqrt{n}\delta)n}{\sqrt{\mu} r}\right], \quad np \geq \Theta[\mu^2 r^3 (\kappa^*)^2 \log n], \quad \delta < \frac{1}{16}.$$

Then, there exists a small constant $\epsilon > 0$ such that with probability at least $1 - 1/\mathrm{poly}(n)$, the MSoG problem (12) satisfies the (θ, β, γ)-strict saddle property with

$$\theta = \Theta(\epsilon/\sigma_r^*), \quad \beta = \epsilon, \quad \gamma = \Theta(\epsilon^2/\sigma_r^*).$$

In Theorem 3, we provide the first theoretical results on the MSoG problem (12). Basically, the theorem shows that if the Ω-RIP$_{2r,2r}$ constant is smaller than an absolute constant, then the same sampling rate as for the matrix completion problem is sufficient to guarantee the benign landscape of problem (12). Therefore, the result is a generalization of the results in [17], which considered the case when the Ω-RIP constant δ is zero and the sampling rate p satisfies a stricter condition (9). On the other hand, if the sampling rate p is equal to 1, Theorem 3 guarantees that the strict saddle property holds when the (regular) RIP condition holds with $\delta < 1/16$. Compared with the state-of-the-art results in [38, 5], the upper bound on δ in our work may not be optimal but is only worse by an absolute constant. We leave the improvement of the upper bound as future work.

3.2 Tightness of the Ω-RIP condition

To study the tightness of our Ω-RIP condition for problem (12), we construct an instance of problem (12) that satisfies the $(1/2, \Omega)$-RIP condition but has spurious second-order critical points. Note that the existence of spurious second-order critical points negates the strict saddle property. The counterexample is based on a similar idea as those in [38]. More specifically, we assume that $n \geq 2r$ and consider the tensor:

$$\begin{aligned}\mathcal{H} := \sum_{i \in [r]} \Big\{ &-\frac{1}{2}\left[(e_{2i-1}e_{2i-1}^T) \otimes (e_{2i-1}e_{2i-1}^T) + (e_{2i}e_{2i}^T) \otimes (e_{2i}e_{2i}^T)\right] \\ &+ \frac{1}{2}\left[(e_{2i-1}e_{2i-1}^T) \otimes (e_{2i}e_{2i}^T) + (e_{2i}e_{2i}^T) \otimes (e_{2i-1}e_{2i-1}^T)\right] \\ &- \frac{1}{4}\left[(e_{2i-1}e_{2i}^T) \otimes (e_{2i-1}e_{2i}^T) + (e_{2i}e_{2i-1}^T) \otimes (e_{2i}e_{2i-1}^T)\right] \\ &+ \frac{1}{4}\left[(e_{2i-1}e_{2i}^T) \otimes (e_{2i}e_{2i-1}^T) + (e_{2i}e_{2i-1}^T) \otimes (e_{2i-1}e_{2i}^T)\right] \Big\},\end{aligned}$$

where $e_i \in \mathbb{R}^n$ is the i-th standard basis of $\mathbb{R}^n$. The rank-r ground truth matrix M^* is constructed as

$$U^* := [e_1 \quad e_3 \quad \cdots \quad e_{2r-1}], \quad M^* := U^*(U^*)^T = \sum_{i \in [r]} e_{2i-1}e_{2i-1}^T.$$

Then, the loss function is given by

$$f_{3/2}(M_\Omega; M_\Omega^*) := \frac{1}{2}(M_\Omega - M_\Omega^*) : \left(\frac{3}{2} \cdot \mathcal{I} + \mathcal{H}\right) : (M_\Omega - M_\Omega^*), \quad \forall M \in \mathbb{R}^{n \times n}.$$

It is proved in [38] that if $\Omega = [n] \times [n]$, the function $f_{3/2}$ satisfies the $1/2$-RIP$_{2r,2r}$ condition and has a spurious second-order critical point. In this work, we generalize the results to the case when the set Ω is random and the objective function contains a regularizer.

Theorem 4. *Suppose that the loss function in problem (12) is chosen to be $f_{3/2}$. Then, it holds that:*

1. *The loss function $f_{3/2}$ satisfies the $(1/2, \Omega)$-RIP$_{2r,2r}$ condition for all $\Omega \subset [n] \times [n]$;*
2. *For all $p \in [0, 1]$, problem (12) has a spurious second-order critical point with probability at least $\max\{p^{r(r+1)}, 1 - p^{r(r+1)/2}\} \geq (3 - \sqrt{5})/2$.*

We note that the results in Theorem 4 holds for all $p \in [0, 1]$. From Theorem 4, we cannot improve the upper bound of δ in Theorem 3 to be better than $1/2$. In addition, this result shows that our bound in Theorem 3 is optimal up to a constant. This tightness result is consistent with that of the matrix sensing problem in [40, 38].

4 Conclusion

In this work, we provide the first theoretical analysis of the MSoG problem. A new notion, dubbed as the Ω-RIP condition, is proposed and shown to be useful in characterizing the optimization complexity of the MSoG problem. Using an improved incoherence regularizer, we proved the polynomial-time global convergence of saddle-avoiding methods under the Ω-RIP condition and the incoherence condition. The bounds on the sampling rate and the Ω-RIP constant are state-of-the-art up to a constant. Moreover, we showed that our bound on the Ω-RIP condition is tight (up to a constant). Future works include improving the upper bound on the Ω-RIP constant and the lower bound on the sampling rate.

References

[1] Ahn, K., Suarez, F.: Riemannian perspective on matrix factorization. arXiv preprint arXiv:2102.00937 (2021)

[2] Allen-Zhu, Z., Li, Y.: Neon2: Finding local minima via first-order oracles. Advances in Neural Information Processing Systems **31** (2018)

[3] Bhojanapalli, S., Jain, P.: Universal matrix completion. In: International Conference on Machine Learning, pp. 1881–1889. PMLR (2014)

[4] Bi, Y., Lavaei, J.: On the absence of spurious local minima in nonlinear low-rank matrix recovery problems. In: International Conference on Artificial Intelligence and Statistics, pp. 379–387. PMLR (2021)

[5] Bi, Y., Zhang, H., Lavaei, J.: Local and global linear convergence of general low-rank matrix recovery problems. In: Proceedings of 36th AAAI Conference on Artificial Intelligence (AAAI), Vancouver, Canada, pp. 1–9 (2022)

[6] Burer, S., Monteiro, R.D.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Programming **95**(2), 329–357 (2003)

[7] Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM (JACM) **58**(3), 1–37 (2011)

[8] Candes, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via wirtinger flow: Theory and algorithms. IEEE Transactions on Information Theory **61**(4), 1985–2007 (2015)

[9] Candes, E.J., Plan, Y.: 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)

[10] Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Foundations of Computational mathematics **9**(6), 717–772 (2009)

[11] Candès, E.J., Tao, T.: The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory **56**(5), 2053–2080 (2010)

[12] Cartis, C., Gould, N.I., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization. part i: motivation, convergence and numerical results. Mathematical Programming **127**(2), 245–295 (2011)

[13] Chen, J., Liu, D., Li, X.: Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. IEEE Transactions on Information Theory **66**(9), 5806–5841 (2020)

[14] Chen, Y., Chi, Y., Fan, J., Ma, C.: Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. Mathematical Programming **176**(1), 5–37 (2019)

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

[16] Chou, H.H., Gieshoff, C., Maly, J., Rauhut, H.: Gradient descent for deep matrix factorization: Dynamics and implicit bias towards low rank. arXiv preprint arXiv:2011.13772 (2020)

[17] Ge, R., Jin, C., Zheng, Y.: No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In: International Conference on Machine Learning, pp. 1233–1242. PMLR (2017)

[18] Ge, R., Lee, J.D., Ma, T.: Matrix completion has no spurious local minimum. Advances in Neural Information Processing Systems pp. 2981–2989 (2016)

[19] Hou, T.Y., Li, Z., Zhang, Z.: Fast global convergence for low-rank matrix recovery via riemannian gradient descent with random initialization. arXiv preprint arXiv:2012.15467 (2020)

[20] Jain, P., Meka, R., Dhillon, I.: Guaranteed rank minimization via singular value projection. Advances in Neural Information Processing Systems **23** (2010)

[21] Jain, P., Netrapalli, P., Sanghavi, S.: Low-rank matrix completion using alternating minimization. In: Proceedings of the forty-fifth annual ACM symposium on Theory of computing, pp. 665–674 (2013)

[22] Jin, C., Ge, R., Netrapalli, P., Kakade, S.M., Jordan, M.I.: How to escape saddle points efficiently. In: International Conference on Machine Learning, pp. 1724–1732. PMLR (2017)

[23] Jin, C., Netrapalli, P., Jordan, M.I.: Accelerated gradient descent escapes saddle points faster than gradient descent. In: Conference On Learning Theory, pp. 1042–1085. PMLR (2018)

[24] Jin, M., Lavaei, J., Sojoudi, S., Baldick, R.: Boundary defense against cyber threat for power system state estimation. IEEE Transactions on Information Forensics and Security **16**, 1752–1767 (2020)

[25] Király, F.J., Theran, L., Tomioka, R.: The algebraic combinatorial approach for low-rank matrix completion. J. Mach. Learn. Res. **16**(1), 1391–1436 (2015)

[26] Li, Y., Ma, T., Zhang, H.: Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In: Conference On Learning Theory, pp. 2–47. PMLR (2018)

[27] Luo, Y., Li, X., Zhang, A.R.: Nonconvex factorization and manifold formulations are almost equivalent in low-rank matrix optimization. arXiv preprint arXiv:2108.01772 (2021)

[28] Netrapalli, P., Jain, P., Sanghavi, S.: Phase retrieval using alternating minimization. Advances in Neural Information Processing Systems **26** (2013)

[29] Netrapalli, P., UN, N., Sanghavi, S., Anandkumar, A., Jain, P.: Non-convex robust pca. Advances in Neural Information Processing Systems **27** (2014)

[30] Pimentel-Alarcón, D.L., Boston, N., Nowak, R.D.: A characterization of deterministic sampling patterns for low-rank matrix completion. IEEE Journal of Selected Topics in Signal Processing **10**(4), 623–636 (2016)

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

[32] Stöger, D., Soltanolkotabi, M.: 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** (2021)

[33] Sun, J., Qu, Q., Wright, J.: A geometric analysis of phase retrieval. Foundations of Computational Mathematics **18**(5), 1131–1198 (2018)

[34] Tu, S., Boczar, R., Simchowitz, M., Soltanolkotabi, M., Recht, B.: Low-rank solutions of linear matrix equations via procrustes flow. In: International Conference on Machine Learning, pp. 964–973. PMLR (2016)

[35] Wei, K., Cai, J.F., Chan, T.F., Leung, S.: Guarantees of riemannian optimization for low rank matrix recovery. SIAM Journal on Matrix Analysis and Applications **37**(3), 1198–1222 (2016)

[36] Wei, K., Cai, J.F., Chan, T.F., Leung, S.: Guarantees of riemannian optimization for low rank matrix completion. Inverse Problems & Imaging **14**(2) (2020)

[37] Yi, X., Park, D., Chen, Y., Caramanis, C.: Fast algorithms for robust pca via gradient descent. Advances in neural information processing systems **29** (2016)

[38] Zhang, H., Bi, Y., Lavaei, J.: General low-rank matrix optimization: Geometric analysis and sharper bounds. Advances in Neural Information Processing Systems **34** (2021)

[39] Zhang, R.Y., Lavaei, J., Baldick, R.: Spurious local minima in power system state estimation. IEEE transactions on control of network systems **6**(3), 1086–1096 (2019)

[40] Zhang, R.Y., Sojoudi, S., Lavaei, J.: Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery. J. Mach. Learn. Res. **20**(114), 1–34 (2019)

[41] Zheng, Q., Lafferty, J.: A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. In: Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1, pp. 109–117 (2015)

[42] Zhu, Z., Li, Q., Tang, G., Wakin, M.B.: Global optimality in low-rank matrix optimization. IEEE Transactions on Signal Processing **66**(13), 3614–3628 (2018)

Appendix

Without loss of generality, we assume that the loss function is symmetric in M, namely,

$$f(M; M^*) = f(M^T; M^*), \quad \forall M \in \mathbb{R}^{n\times n}.$$

Otherwise, we can replace the loss function by $\bar{f}(M; M^*) := [f(M; M^*) + f(M^T; M^*)]/2$ and this will not change the values of the objective function in problem (1) in the feasible set.

Denote the objective function of problem (12) as $\frac{1}{p}h(U) + g(U)$, where

$$h(U) := f[(UU^T)_\Omega; M^*_\Omega], \quad g(U) := \sum_{i\in[n]} r(\|e_i^T U\|_F).$$

For every $\Delta \in \mathbb{R}^{n\times r}$, the gradient and the Hessian matrix of $h(U)$ satisfy

$$\begin{aligned}\langle \nabla h(U), \Delta\rangle &= 2\left\langle \nabla f[(UU^T)_\Omega; M^*_\Omega], \left(U\Delta^T\right)_\Omega\right\rangle,\\ \left[\nabla^2 h(U)\right](\Delta,\Delta) &= 2\left\langle \nabla f[(UU^T)_\Omega; M^*_\Omega], \left(\Delta\Delta^T\right)_\Omega\right\rangle\\ &\quad + \left[\nabla^2 f[(UU^T)_\Omega; M^*_\Omega]\right]\left[\left(\Delta U^T + U\Delta^T\right)_\Omega, \left(\Delta U^T + U\Delta^T\right)_\Omega\right].\end{aligned}$$

Since we can exchange the convolution and the derivatives, the gradient and the Hessian matrix of $g(U)$ satisfy

$$\langle \nabla g(U), \Delta\rangle = 4\lambda \int_{-1}^{1} \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F}(1-|y|)\,dy,$$

$$\begin{aligned}&\left[\nabla^2 g(U)\right](\Delta,\Delta)\\ =&4\lambda \int_{-1}^{1} \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \frac{\|e_i^T U\|_F\|e_i^T\Delta\|_F - e_i^T U\Delta^T e_i}{\|e_i^T U\|_F^3}(1-|y|)\,dy\\ &+ 12\lambda \int_{-1}^{1} \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^2 \left(\frac{e_i^T U\Delta^T e_i}{\|e_i^T U\|_F}\right)^2 (1-|y|)\,dy.\end{aligned}$$

For all $\epsilon > 0$, we say that a point $U \in \mathbb{R}^{n\times r}$ is an *ϵ-approximate first-order critical point* of problem (12) if

$$\left\|\frac{1}{p}\nabla h(U) + \nabla g(U)\right\|_F \le \epsilon.$$

A Proof of Theorem 3

We first bound the norm of an approximate first-order critical point.

Lemma 1. *Suppose that U is an ϵ-approximate first-order critical point of problem* (12) *for a sufficiently small $\epsilon > 0$ and*

$$\delta < 1, \quad \max_{i\in[n]} \|e_i^T U\|_F \ge 11\alpha.$$

Then, it holds that

$$\|(UU^T)_\Omega\|_F \le \frac{1+\delta}{1-\delta}\|(M^*)_\Omega\|_F.$$

Proof. Assume that U is an ϵ-approximate first-order critical point such that

$$\|(UU^T)_\Omega\|_F > \frac{1+\delta}{1-\delta}\|(M^*)_\Omega\|_F. \tag{14}$$

Using the approximate first-order stationarity, we have

$$\epsilon\|U\|_F \ge \left\langle \left[\frac{1}{p}\nabla h(U) + \nabla r(U)\right], U\right\rangle = \left\langle \frac{1}{p}\nabla h(U), U\right\rangle + \langle \nabla g(U), U\rangle.$$

For the first term, we can calculate that

$$
\begin{aligned}
&\left\langle \frac{1}{p}\nabla h(U), U \right\rangle = \frac{2}{p}\left\langle \nabla f[(UU^T)_\Omega], (UU^T)_\Omega \right\rangle \\
=&\frac{2}{p}\int_0^1 \left[\nabla^2 f[(M^*)_\Omega + t(UU^T - M^*)_\Omega; (M^*)_\Omega]\right] [(UU^T)_\Omega, (UU^T - M^*)_\Omega]\, dt \\
\geq&\frac{2(1-\delta)}{p}\left\|(UU^T)_\Omega\right\|_F^2 - \frac{2(1+\delta)}{p}\|(UU^T)_\Omega\|_F\|(M^*)_\Omega\|_F \geq 0,
\end{aligned}
$$

where the first inequality is from the (δ, Ω)-RIP$_{2r,2r}$ condition and Lemma 11 of [4], and the last inequality is from the assumption (14). Define the set

$$
\mathcal{I} := \{i \in [n] \mid \|e_i^T U\|_F \geq 11\alpha\}.
$$

Then, for the second term, we have

$$
\begin{aligned}
\langle \nabla g(U), U\rangle &= 4\lambda \int_{-1}^1 \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \frac{e_i^T UU^T e_i}{\|e_i^T U\|_F}(1-|y|)\, dy \\
&= 4\lambda \int_{-1}^1 \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \|e_i^T U\|_F(1-|y|)\, dy \\
&\geq 4\lambda \int_{-1}^1 \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F - 11\alpha\right)_+ + 8\alpha\right]^3 \|e_i^T U\|_F(1-|y|)\, dy \\
&\geq 4\lambda \int_{-1}^1 \sum_{i\in\mathcal{I}} \left(\|e_i^T U\|_F - 3\alpha\right)^3 \|e_i^T U\|_F(1-|y|)\, dy \\
&= 4\lambda \sum_{i\in\mathcal{I}} \left(\|e_i^T U\|_F - 3\alpha\right)^3 \|e_i^T U\|_F.
\end{aligned}
$$

Combining the last two inequalities, we obtain

$$
\epsilon\|U\|_F \geq 4\lambda \sum_{i\in\mathcal{I}} \left(\|e_i^T U\|_F - 3\alpha\right)^3 \|e_i^T U\|_F.
$$

Define

$$
i^* := \arg\max_{i\in[n]} \|e_i^T U\|.
$$

Then, we obtain

$$
\epsilon\sqrt{n}\|e_{i^*}^T U\|_F \geq \epsilon\|U\|_F \geq 4\lambda \sum_{i\in\mathcal{I}} \left(\|e_i^T U\|_F - 3\alpha\right)^3 \|e_i^T U\|_F \geq 4\lambda \left(\|e_{i^*}^T U\|_F - 3\alpha\right)^3 \|e_{i^*}^T U\|_F,
$$

which further leads to

$$
\left(\|e_{i^*}^T U\|_F - 3\alpha\right)^3 \leq \frac{\epsilon\sqrt{n}}{4\lambda}.
$$

By choosing ϵ to be sufficiently small, the above inequality implies that

$$
\|e_{i^*}^T U\|_F < 11\alpha.
$$

This is a contradiction to the condition in the lemma. □

Next, we provide a generalization to Lemma 9 of [17].

Lemma 2. *Suppose that U is an ϵ-approximate first-order critical point of problem (12) for a sufficiently small $\epsilon > 0$ and*

$$
\alpha^2 = \Theta\left(\frac{\mu r \sigma_1^*}{n}\right), \quad \lambda = \Theta\left[\frac{(1-\delta)(\sqrt{\mu} + \sqrt{n}\delta)n}{\sqrt{\mu} r}\right], \quad np \geq \Theta\left(\mu r \log n\right), \quad \delta < 1.
$$

Then, there exists a constant $c > 0$ such that it holds with probability at least $1 - 1/\mathrm{poly}(n)$ that

$$
\max_{i\in[n]} \|e_i^T U\|_F^2 = O\left(\frac{\mu r \sigma_1^*}{n}\right).
$$

Proof. Let

$$i^* := \arg\max_{i\in[n]} \|e_i^T U\|.$$

We only need to consider the case when $\|e_{i^*}^T U\|_F \geq 11\alpha$. Since U is an ϵ-approximate first-order critical point, it holds that

$$\epsilon\|e_{i^*}^T U\|_F \geq \left\langle e_{i^*}^T \left[\frac{1}{p}\nabla h(U) + \nabla g(U)\right], e_{i^*}^T U\right\rangle. \tag{15}$$

Using Taylor's expansion, it holds that

$$\begin{aligned}&\left\langle e_{i^*}^T \cdot \frac{1}{p}\nabla h(U), e_{i^*}^T U\right\rangle\\ =&\frac{2}{p}\int_0^1 \left[\nabla^2 f[(M^*)_\Omega + t(UU^T - M^*)_\Omega; (M^*)_\Omega]\right] \left[(UU^T e_{i^*} e_{i^*}^T)_\Omega, (UU^T - M^*)_\Omega\right] dt.\end{aligned}$$

By the (δ,Ω)-RIP$_{2r,2r}$ condition and Lemma 11 of [4], one can write

$$\begin{aligned}&\left\langle e_{i^*}^T \cdot \frac{1}{p}\nabla h(U), e_{i^*}^T U\right\rangle \\ \geq&\frac{2}{p}\left\langle (UU^T e_{i^*} e_{i^*}^T)_\Omega, (UU^T - M^*)_\Omega\right\rangle - \frac{2\delta}{p}\|(UU^T e_{i^*} e_{i^*}^T)_\Omega\|_F \left(\|(UU^T)_\Omega\|_F + \|(M^*)_\Omega\|_F\right)\\ =&\frac{2}{p}\left\langle e_{i^*}^T (UU^T)_\Omega, e_{i^*}^T (UU^T - M^*)_\Omega\right\rangle - \frac{2\delta}{p}\|e_{i^*}^T (UU^T)_\Omega\|_F \left(\|(UU^T)_\Omega\|_F + \|(M^*)_\Omega\|_F\right)\\ \geq& -\frac{2}{p}\|e_{i^*}^T (UU^T)_\Omega\|_F \|e_{i^*}^T (M^*)_\Omega\|_F - \frac{2\delta}{p}\|e_{i^*}^T (UU^T)_\Omega\|_F \left(\|(UU^T)_\Omega\|_F + \|(M^*)_\Omega\|_F\right).\end{aligned} \tag{16}$$

By Lemmas 35 and 39 of [17] and Lemma 1, it holds with probability at least $1 - 1/\mathrm{poly}(n)$ that

$$\begin{aligned}\frac{1}{\sqrt{p}}\|e_{i^*}^T (M^*)_\Omega\|_F &\leq \sqrt{1+\nu}\|e_{i^*}^T M^*\|_F \leq \sqrt{\frac{(1+\nu)\mu r}{n}}\sigma_1^*,\\ \frac{1}{\sqrt{p}}\|e_{i^*}^T (UU^T)_\Omega\|_F &\leq O(\sqrt{n})\|UU^T\|_\infty,\\ \frac{1}{\sqrt{p}}\left(\|(UU^T)_\Omega\|_F + \|(M^*)_\Omega\|_F\right) &\leq \frac{2}{(1-\delta)\sqrt{p}}\|(M^*)_\Omega\|_F \leq \frac{2\sqrt{(1+\nu)r}\sigma_1^*}{1-\delta},\end{aligned}$$

where the constant $\nu > 0$ can be made sufficiently small by choosing a large constant in the condition $np \geq \Theta(\mu r \log n)$. Substituting into the inequality (16), it follows that

$$\begin{aligned}\left\langle e_{i^*}^T \cdot \frac{1}{p}\nabla h(U), e_{i^*}^T U\right\rangle &\geq -O(\sqrt{\mu r}\sigma_1^*)\|UU^T\|_\infty - O[\sqrt{nr}\sigma_1^*\delta/(1-\delta)] \cdot \|UU^T\|_\infty\\ &= -O[(\sqrt{\mu} + \sqrt{n}\delta)\sqrt{r}\sigma_1^*/(1-\delta)] \cdot \|e_{i^*}^T U\|_F^2,\end{aligned}$$

where the last equality is from $\|UU^T\|_\infty = \|e_{i^*}^T U\|_F^2$. Additionally, since $\|e_{i^*}^T U\|_F \geq 9\alpha$, we have

$$\begin{aligned}&\left\langle e_{i^*}^T \nabla h(U), e_{i^*}^T U\right\rangle\\ =&4\lambda\int_{-1}^1 \left\langle e_{i^*}^T \cdot \sum_{i\in[n]} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \frac{e_i e_i^T U}{\|e_i^T U\|_F}, e_{i^*}^T U\right\rangle (1-|y|)\, dy\\ =&4\lambda\int_{-1}^1 \left(\|e_{i^*}^T U\|_F + \alpha y - 2\alpha\right)^3 \|e_{i^*}^T U\|_F (1-|y|)\, dy\\ \geq&4\lambda\left(\|e_{i^*}^T U\|_F - 3\alpha\right)^3 \|e_{i^*}^T U\|_F \geq 4\lambda \cdot \left(\frac{8}{11}\right)^3 \|e_{i^*}^T U\|_F^4 > \lambda\|e_{i^*}^T U\|_F^4.\end{aligned}$$

Substituting the last two bounds into inequality (15), we know that

$$\epsilon\|e_{i^*}^T U\|_F \geq \lambda\|e_{i^*}^T U\|_F^4 - O[(\sqrt{\mu} + \sqrt{n}\delta)\sqrt{r}\sigma_1^*/(1-\delta)] \cdot \|e_{i^*}^T U\|_F^2$$

holds with high probability. We can rewrite the above bound as

$$\lambda\|e_{i^*}^T U\|_F^3 \le \epsilon + \frac{c(\sqrt{\mu}+\sqrt{n}\delta)\sqrt{r}\sigma_1^*}{1-\delta}\cdot\|e_{i^*}^T U\|_F,$$

where $c > 0$ is a constant. If we choose ϵ sufficiently small such that

$$\epsilon^{2/3} \le \frac{c(\sqrt{\mu}+\sqrt{n}\delta)\sqrt{r}\sigma_1^*}{\lambda^{1/3}(1-\delta)},$$

the inequality leads to

$$\|e_{i^*}^T U\|_F^2 \le \frac{c(\sqrt{\mu}+\sqrt{n}\delta)\sqrt{r}\sigma_1^*}{\lambda(1-\delta)}.$$

By our choice of λ, we conclude that

$$\|e_{i^*}^T U\|_F^2 \le O\left(\frac{\mu r\sigma_1^*}{n}\right).$$

This concludes our proof. □

Now, we consider the second-order necessary optimality condition for an approximate first-order critical point U. The next lemma bounds the curvature along the direction Δ.

Lemma 3. *For all $U \in \mathbb{R}^{n\times r}$, we define the direction $\Delta := U - U^*R$, where $U^* \in \mathbb{R}^{n\times r}$ satisfies $U^*(U^*)^T = M^*$ and*

$$R \in \underset{S\in\mathbb{R}^{r\times r}, SS^T=I_r}{\arg\min} \|U - U^*S\|_F.$$

Then, it holds that

$$\begin{aligned}\left[\nabla^2 h(U)\right](\Delta,\Delta) \le 4\langle\nabla h(U),\Delta\rangle - [3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 \\ + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2,\end{aligned}$$

where the constant $t = \sqrt{2}$.

Proof. Define $M := UU^T$. By Taylor's expansion, we have

$$\begin{aligned}&\left[\nabla^2 h(U)\right](\Delta,\Delta)\\ =&2\left\langle\nabla f(M_\Omega), \left(\Delta\Delta^T\right)_\Omega\right\rangle + \left[\nabla^2 f(M_\Omega;M_\Omega^*)\right]\left[(\Delta U^T+U\Delta^T)_\Omega, (\Delta U^T+U\Delta^T)_\Omega\right]\\ =&4\langle\nabla h(U),\Delta\rangle - 4\left\langle\nabla f(M_\Omega), (M-M^*)_\Omega\right\rangle - 2\left\langle\nabla f(M_\Omega), \left(\Delta\Delta^T\right)_\Omega\right\rangle\\ &+\left[\nabla^2 f(M_\Omega;M_\Omega^*)\right]\left[(\Delta U^T+U\Delta^T)_\Omega, (\Delta U^T+U\Delta^T)_\Omega\right]\\ =&4\langle\nabla h(U),\Delta\rangle - 4\int_0^1 \left[\nabla^2 f(M_\Omega + t(M-M^*)_\Omega;M_\Omega^*)\right]\left[(M-M^*)_\Omega,(M-M^*)_\Omega\right]\,dt\\ &-2\int_0^1 \left[\nabla^2 f(M_\Omega + t(M-M^*)_\Omega;M_\Omega^*)\right]\left[(M-M^*)_\Omega,(\Delta\Delta^T)_\Omega\right]\,dt\\ &+\left[\nabla^2 f(M_\Omega;M_\Omega^*)\right]\left[(\Delta U^T+U\Delta^T)_\Omega, (\Delta U^T+U\Delta^T)_\Omega\right].\end{aligned}$$

Using the (δ,Ω)-RIP$_{2r,2r}$ condition, it follows that

(17)
$$\begin{aligned}\left[\nabla^2 h(U)\right](\Delta,\Delta) \le&4\langle\nabla h(U),\Delta\rangle - 4(1-\delta)\|(M-M^*)_\Omega\|_F^2 - 2\left\langle(M-M^*)_\Omega,(\Delta\Delta^T)_\Omega\right\rangle\\ &+4\delta\|(M-M^*)_\Omega\|_F\|(\Delta\Delta^T)_\Omega\|_F + (1+\delta)\|(\Delta U^T+U\Delta^T)_\Omega\|_F^2\\ =&4\langle\nabla h(U),\Delta\rangle - (3-5\delta)\|(M-M^*)_\Omega\|_F^2 + (1+\delta)\|(\Delta\Delta^T)_\Omega\|_F^2\\ &+4\delta\|(M-M^*)_\Omega\|_F\|(\Delta\Delta^T)_\Omega\|_F + 2\delta\left\langle(M-M^*)_\Omega,(\Delta\Delta^T)_\Omega\right\rangle\\ \le&4\langle\nabla h(U),\Delta\rangle - (3-5\delta)\|(M-M^*)_\Omega\|_F^2 + (1+\delta)\|(\Delta\Delta^T)_\Omega\|_F^2\\ &+6\delta\|(M-M^*)_\Omega\|_F\|(\Delta\Delta^T)_\Omega\|_F,\end{aligned}$$

where we have used the relation $\Delta U^T + U\Delta^T = M - M^* + \Delta\Delta^T$. Using Hölder's inequality, we have

$$2\|(M-M^*)_\Omega\|_F\|(\Delta\Delta^T)_\Omega\|_F \leq t\|(M-M^*)_\Omega\|_F^2 + t^{-1}\|(\Delta\Delta^T)_\Omega\|_F^2.$$

Substituting the above inequality into (17), we obtain

$$\begin{aligned}\left[\nabla^2 h(U)\right](\Delta,\Delta) \leq 4\langle\nabla h(U),\Delta\rangle - [3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 \\ + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2.\end{aligned}$$

This is the desired result. □

The following lemma is a generalization of Lemma 10 in [17].

Lemma 4. *Suppose that U is an ϵ-approximate first-order critical point of problem (12) for a sufficiently small ϵ and*

$$\alpha^2 = \Theta\left(\frac{\mu r\sigma_1^*}{n}\right), \quad \lambda = \Theta\left[\frac{(\sqrt{\mu}+\sqrt{n}\delta)n}{\sqrt{\mu}r}\right], \quad np \geq C\mu^2 r^3(\kappa^*)^2\log n, \quad \delta < \frac{1}{16},$$

where $C > 0$ is a sufficiently large constant. Then, it holds with probability at least $1 - 1/\mathrm{poly}(n)$ that

$$\frac{1}{p}\left[-[3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2\right] \leq -0.03\sigma_r^*\|\Delta\|_F^2,$$

where the constant $t = \sqrt{2}$ and $\Delta \in \mathbb{R}^{n\times r}$ is defined in Lemma 3.

Proof. By Lemma 2, we can bound the norm of each row of U by

$$\max_i \|e_i^T U\|_F^2 = O\left(\frac{\mu r\sigma_1^*}{n}\right)$$

with probability at least $1 - 1/\mathrm{poly}(n)$. Then, we split the proof into two different cases.

Case I. We first consider the case when $\|\Delta\|_F^2 \leq \sigma_r^*/4$. We can calculate that

$$\begin{aligned}&\frac{1}{p}\left[-[3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2\right] \\ &= -\frac{4[3-(5+3t)\delta]}{p}\left[\langle(U^*\Delta^T)_\Omega,(\Delta\Delta^T)_\Omega\rangle + \|(U^*\Delta^T)_\Omega\|_F^2\right] \\ &\qquad + [1+(1+3t^{-1})\delta - 3 + (5+3t)\delta]\|(\Delta\Delta^T)_\Omega\|_F^2 \\ &\leq -\frac{9}{p}\left[\langle(U^*\Delta^T)_\Omega,(\Delta\Delta^T)_\Omega\rangle + \|(U^*\Delta^T)_\Omega\|_F^2\right] \\ &\leq -\frac{9}{p}\cdot\|(U^*\Delta^T)_\Omega\|_F\left[\|(U^*\Delta^T)_\Omega\|_F - \|(\Delta\Delta^T)_\Omega\|_F\right],\end{aligned} \tag{18}$$

where the first inequality is from the condition that $\delta < 1/16$ and $t = \sqrt{2}$. Using a similar analysis as Lemma 10 of [17] and the condition $\delta < 1/16$, it holds with probability at least $1 - 1/\mathrm{poly}(n)$ that

$$\frac{1}{p}\|(U^*\Delta^T)_\Omega\|_F^2 \geq (1-\nu)\sigma_r^*\|\Delta\|_F^2, \quad \frac{1}{p}\|(\Delta\Delta^T)_\Omega\|_F^2 \leq \frac{\sigma_r^*}{2}\|\Delta\|_F^2,$$

where constant $\nu > 0$ can be made sufficiently small by choosing a large enough C. Substituting the above bounds into inequality (18), it holds with the same probability that

$$\begin{aligned}&\frac{1}{p}\left[-[3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2\right] \\ &\leq -9\sqrt{1-\nu}\left(\sqrt{1-\nu} - 1/\sqrt{2}\right)\sigma_r^*\|\Delta\|_F^2 < -0.03\sigma_r^*\|\Delta\|_F^2,\end{aligned}$$

where the last inequality is by choosing a sufficiently small ν.

Case II. Now, we consider the case when $\|\Delta\|_F^2 \geq \sigma_r^*/4$. By a similar analysis as Lemma 10 of [17], it holds with probability at least $1 - 1/\text{poly}(n)$ that

$$\frac{1}{p}\|(\Delta\Delta^T)_\Omega\|_F^2 \leq \|\Delta\Delta^T\|_F^2 + \nu\sigma_r^*\|\Delta\|_F^2,$$
$$\frac{1}{p}\|(M - M^*)_\Omega\|_F^2 \geq (1-\nu)\|M - M^*\|_F^2 - \nu\sigma_r^*\|\Delta\|_F^2,$$

where constant $\nu > 0$ can be made sufficiently small by choosing a large enough C. Therefore, the condition $\delta < 1/16$ implies that with the same probability, we have

$$\begin{aligned}
&\frac{1}{p}\left[-[3-(5+3t)\delta]\cdot\|(M-M^*)_\Omega\|_F^2 + [1+(1+3t^{-1})\delta]\cdot\|(\Delta\Delta^T)_\Omega\|_F^2\right] \\
\leq&[1+(1+3t^{-1})\delta]\cdot\left(\|\Delta\Delta^T\|_F^2 + \nu\sigma_r^*\|\Delta\|_F^2\right) \\
&\qquad\qquad - [3-(5+3t)\delta]\cdot\left[(1-\nu)\|M-M^*\|_F^2 - \nu\sigma_r^*\|\Delta\|_F^2\right] \\
\leq&\left[2 + 2(1+3t^{-1})\delta - (1-\nu)[3-(5+3t)\delta]\right]\cdot\|M-M^*\|_F^2 \\
&\qquad\qquad + [1+(1+3t^{-1})\delta - 3 + (5+3t)\delta]\cdot\nu\sigma_r^*\|\Delta\|_F^2 \\
\leq&\left[-1 + (7+3t+6t^{-1})\delta + O(\nu)\right]\cdot 2(\sqrt{2}-1)\sigma_r^*\|\Delta\|_F^2 \\
=&2(\sqrt{2}-1)\left[-1 + (7+6\sqrt{2})\delta + O(\nu)\right]\cdot\sigma_r^*\|\Delta\|_F^2 < -0.03\sigma_r^*\|\Delta\|_F^2,
\end{aligned}$$

where the second inequality is from $\|\Delta\Delta^T\|_F^2 \leq 2\|M - M^*\|_F^2$, the second last inequality is from $2(\sqrt{2}-1)\|\Delta\|_F^2 \leq 2\|M - M^*\|_F^2$ and the last inequality is by choosing a sufficiently small ν.

Combining the two cases completes the proof. □

By the same proof as that of Lemma 11 of [17], we can bound the curvature of the regularizer.

Lemma 5. *Suppose that U is an ϵ-approximate first-order critical point of problem (12) for a sufficiently small ϵ and*

$$\alpha^2 = \Theta\left(\frac{\mu r\sigma_1^*}{n}\right), \quad \lambda \geq 0,$$

Then, it holds with probability at least $1 - 1/\text{poly}(n)$ that

$$\left[\nabla^2 g(U)\right](\Delta, \Delta) - 4\langle\nabla g(U), \Delta\rangle \leq 0,$$

where $\Delta \in \mathbb{R}^{n\times r}$ is defined in Lemma 3.

Proof. For each $i \in [n]$, since the regularizer is non-zero only if $\|e_i^T U\| \geq 9\alpha$, we only need to consider the index set

$$\mathcal{I} := \{i \in [n] \mid \|e_i^T U\|_F \geq 9\alpha\}.$$

We can calculate that

$$\begin{aligned}
&\left[\nabla^2 g(U)\right](\Delta, \Delta) - 4\langle\nabla g(U), \Delta\rangle \\
=&4\lambda\int_{-1}^{1}\sum_{i\in\mathcal{I}}\left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3 \frac{\|e_i^T U\|_F\|e_i^T\Delta\|_F - e_i^T U\Delta^T e_i}{\|e_i^T U\|_F^3}(1-|y|)\,dy \\
&+ 12\lambda\int_{-1}^{1}\sum_{i\in\mathcal{I}}\left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^2\left(\frac{e_i^T U\Delta^T e_i}{\|e_i^T U\|_F}\right)^2(1-|y|)\,dy \\
&- 16\lambda\int_{-1}^{1}\sum_{i\in\mathcal{I}}\left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha\right)_+ + 8\alpha\right]^3\frac{e_i^T U\Delta^T e_i}{\|e_i^T U\|_F}(1-|y|)\,dy \\
=&I_1 + I_2,
\end{aligned}$$

where we define

$$I_1 := 4\lambda \int_{-1}^{1} \sum_{i \in \mathcal{I}} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha \right)_+ + 8\alpha \right]^3 \cdot \left(\frac{\|e_i^T U\|_F \|e_i^T \Delta\|_F - e_i^T U \Delta^T e_i}{\|e_i^T U\|_F^3} - 0.4 \cdot \frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} \right) (1 - |y|) \, dy$$

$$I_2 := 12\lambda \int_{-1}^{1} \sum_{i \in \mathcal{I}} \left[\left(\|e_i^T U\|_F + \alpha y - 10\alpha \right)_+ + 8\alpha \right]^2 \cdot \left(\frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} - 1.2 \cdot \frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} \right)^2 (1 - |y|) \, dy.$$

It is proved in Lemma 11 of [17] that

$$\frac{\|e_i^T U\|_F \|e_i^T \Delta\|_F - e_i^T U \Delta^T e_i}{\|e_i^T U\|_F^3} - 0.4 \cdot \frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} < 0,$$

which implies that

$$I_1 < 0.$$

Similarly, since we assume $\|e_i^T U\|_F \geq 9\alpha$, the second case of Lemma 11 of [17] implies that

$$\frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} - 1.2 \cdot \frac{e_i^T U \Delta^T e_i}{\|e_i^T U\|_F} \leq 0,$$

which leads to

$$I_2 \leq 0.$$

Hence, we get $I_1 + I_2 \leq 0$ and

$$\left[\nabla^2 g(U) \right] (\Delta, \Delta) - 4 \langle \nabla g(U), \Delta \rangle \leq 0,$$

which is the desired result. ☐

The next lemma establishes the bound on the curvature along Δ for an ϵ-approximate first-order critical point.

Lemma 6. *Suppose that U is an ϵ-approximate first-order critical point of problem (12) for a sufficiently small ϵ and*

$$\alpha^2 = \Theta\left(\frac{\mu r \sigma_1^*}{n} \right), \quad \lambda = \Theta\left[\frac{(\sqrt{\mu} + \sqrt{n}\delta) n}{\sqrt{\mu} r} \right], \quad np \geq C \mu^2 r^3 (\kappa^*)^2 \log n, \quad \delta < \frac{1}{16},$$

Then, it holds with probability at least $1 - 1/\mathrm{poly}(n)$ that

$$\left[\frac{1}{p} \nabla^2 h(U) + \nabla^2 g(U) \right] (\Delta, \Delta) \leq -0.03 \sigma_r^* \|\Delta\|_F^2 + 4\epsilon \|\Delta\|_F,$$

where $\Delta \in \mathbb{R}^{n \times r}$ is defined in Lemma 3.

Proof. With probability at least $1 - 1/\mathrm{poly}(n)$, we have

$$\begin{aligned}
&\left[\frac{1}{p} \nabla^2 h(U) + \nabla^2 g(U) \right] (\Delta, \Delta) \\
\leq & 4 \left\langle \frac{1}{p} \nabla h(U) + \nabla g(U), \Delta \right\rangle + \left[\nabla^2 r(U) \right] (\Delta, \Delta) - 4 \langle \nabla g(U), \Delta \rangle \\
&+ \frac{1}{p} \left[-[3 - (5 + 3t)\delta] \cdot \|(M - M^*)_\Omega\|_F^2 + [1 + (1 + 3t^{-1})\delta] \cdot \|(\Delta\Delta^T)_\Omega\|_F^2 \right] \\
\leq & \epsilon \|\Delta\|_F - 0.03 \sigma_r^* \|\Delta\|_F^2,
\end{aligned}$$

where the first inequality is by Lemma 3 and the second inequality is by Lemmas 4 and 5. This finishes the proof of the lemma. ☐

Using Lemmas 1-6, we are ready to prove the main theorem.

Proof of Theorem 3. Define $\Delta \in \mathbb{R}^{n\times r}$ in the same way as in Lemma 3. Suppose that U is an ϵ-approximate first-order critical point of problem (12) for a sufficiently small ϵ and

$$\mathrm{dist}(U, U^*) = \|\Delta\|_F \geq \frac{100\epsilon}{\sigma_r^*}.$$

Then, Lemma 6 implies that

$$\left[\frac{1}{p}\nabla^2 h(U) + \nabla^2 g(U)\right](\Delta,\Delta) \leq \epsilon\|\Delta\|_F - 0.03\sigma_r^*\|\Delta\|_F^2 \leq -\frac{200\epsilon^2}{\sigma_r^*}$$

holds with probability at least $1 - 1/\mathrm{poly}(n)$. Hence, with the same probability, the MSoG problem (12) satisfies the (θ,β,γ)-strict saddle property with

$$\theta = 100\epsilon/\sigma_r^*, \quad \beta = \epsilon, \quad \gamma = 200\epsilon^2/\sigma_r^*.$$

This completes the proof. □

B Proof of Theorem 4

We split the proof into two parts. We first prove that the $(1/2,\Omega)$-RIP$_{2r,2r}$ condition holds for all non-empty Ω and then prove the existence of spurious second-order critical points.

Proof of the Ω-RIP condition. Since the loss function $f_{3/2}$ is a quadratic function, it holds that

$$[\nabla^2 f_{3/2}(M_\Omega; M_\Omega^*)](K,K) = \frac{3}{2}\|K_\Omega\|_F^2 + K_\Omega : \mathcal{H} : K_\Omega, \quad \forall K, M \in \mathbb{R}^{n\times n}.$$

For the notational simplicity, we define

$$\tilde{K} := K_\Omega.$$

By the definition of tensor $\mathcal{H}$, we can calculate that

$$\begin{aligned} K_\Omega : \mathcal{H} : K_\Omega = &\sum_{i\in[r]}\Big[-\frac{1}{2}\left(\tilde{K}_{2i-1,2i-1}^2 + \tilde{K}_{2i,2i}^2\right) + \tilde{K}_{2i-1,2i-1}\tilde{K}_{2i,2i} \\ &\quad - \frac{1}{4}\left(\tilde{K}_{2i-1,2i}^2 + \tilde{K}_{2i,2i-1}^2 - 2\tilde{K}_{2i-1,2i}\tilde{K}_{2i,2i-1}\right)\Big] \\ = &-\frac{1}{2}\sum_{i\in[r]}\left(\tilde{K}_{2i-1,2i-1} - \tilde{K}_{2i,2i}\right)^2 - \frac{1}{4}\sum_{i\in[r]}\left(\tilde{K}_{2i-1,2i} - \tilde{K}_{2i,2i-1}\right)^2. \end{aligned}$$

It is straightforward to see that

$$K_\Omega : \mathcal{H} : K_\Omega \leq 0. \tag{19}$$

For all real numbers a, b, we have the inequality $(a-b)^2 \leq 2(a^2+b^2)$. This inequality leads to

$$\begin{aligned} K_\Omega : \mathcal{H} : K_\Omega &\geq -\sum_{i\in[r]}\left(\tilde{K}_{2i-1,2i-1}^2 + \tilde{K}_{2i,2i}^2\right) - \frac{1}{2}\sum_{i\in[r]}\left(\tilde{K}_{2i-1,2i}^2 + \tilde{K}_{2i,2i-1}^2\right) \\ &\geq -\|\tilde{K}\|_F^2 = -\|K_\Omega\|_F^2, \end{aligned}$$

where the last equality is from the definition of $\tilde{K}$. Combining with inequality (19), it follows that

$$-\|K_\Omega\|_F^2 \leq K_\Omega : \mathcal{H} : K_\Omega \leq 0.$$

Hence, we have

$$\frac{1}{2}\|K_\Omega\|_F^2 \leq [\nabla^2 f_{3/2}(M_\Omega; M_\Omega^*)](K,K) \leq \frac{3}{2}\|K_\Omega\|_F^2, \quad \forall K, M \in \mathbb{R}^{n\times n},$$

which further implies the $(1/2,\Omega)$-RIP$_{2r,2r}$ condition of $f_{3/2}$.

Existence of spurious second-order critical points. Now, we prove the existence of a spurious second-order critical point by explicit construction. For all $i, j \in [n]$, we define

$$\omega_{i,j} := \begin{cases} 1, & \text{if } (i,j) \in \Omega \\ 0, & \text{otherwise.} \end{cases}$$

Note that we can choose α to be large enough so that

$$\alpha \geq 2 \max_{i \in [n]} \|e_i^T U^*\|_F = 2.$$

Otherwise, the ground truth U^* is not a global minimum of problem (12) since the regularizer has a non-zero gradient at U^*. This is consistent with our choice of α in Theorem 3, because the incoherence of M^* is $\mu = \sqrt{n/r}$ and α^2 is on the order of $\Theta(1)$. We consider two different cases.

Case I. We first consider the case when

$$p^{\frac{r(r+1)}{2}} \leq \frac{\sqrt{5}-1}{2}. \tag{20}$$

In this case, we show that the loss function $f_{3/2}$ has multiple global minima with probability at least $(3-\sqrt{5})/2$. By the condition (20), we can estimate the probability

$$\mathbb{P}\left(\omega_{2i,2j} = 1,\ \forall i, j \in [r]\right) = p^{\frac{r(r+1)}{2}} \leq \frac{\sqrt{5}-1}{2}. \tag{21}$$

This is because $\omega_{2i,2j} = \omega_{2j,2i}$ for all $i, j \in [r]$ and thus, there are $r(r+1)/2$ independent Bernoulli random variables with parameter p. Suppose that the event in (21) does not happen (this event has probability $(3-\sqrt{5})/2$) and $\omega_{2i,2j} = 0$ for some $i, j \in [r]$. Then, we consider the matrix

$$\tilde{M}^* := M^* + \epsilon \cdot e_{2i} e_{2j}^T + \epsilon \cdot e_{2j} e_{2i}^T,$$

where $\epsilon > 0$ is sufficiently small. We can verify that $\tilde{M}^*$ is a PSD matrix and

$$\tilde{M}^*_\Omega = M^*_\Omega, \quad \text{which further leads to } f_{3/2}(\tilde{M}^*_\Omega; M^*_\Omega) = 0.$$

This implies that when the event in (21) does not happen, there exists a global minimum of $f_{3/2}$ that is different from M^*. Therefore, function $h(U)$ also has a global minimum $\tilde{U}^*$ such that $\tilde{U}^*(\tilde{U}^*)^T = \tilde{M}^* \neq M^*$. Note that we can choose ϵ to be small enough such that

$$\alpha \geq \max_{i \in [n]} \|e_i^T \tilde{U}^*\|_F.$$

Then, we know that the regularizer $r(U)$ is zero at $\tilde{U}^*$ and $\tilde{U}^*$ is a spurious second-order critical point of problem (12).

Case II. Now, we consider the case when (20) does not hold, namely,

$$p^{\frac{r(r+1)}{2}} \geq \frac{\sqrt{5}-1}{2}.$$

In this case, a similar calculation to (20) leads to

$$\mathbb{P}\left(\omega_{2i,2i} = \omega_{2i-1,j} = 1,\ \forall i, j \in [r]\right) \geq p^{r(r+1)} \geq \frac{3-\sqrt{5}}{2}. \tag{22}$$

We focus on the case when the event in (22) holds. Define

$$U_0 := \frac{1}{\sqrt{2}} \begin{bmatrix} e_2 & e_4 & \cdots & e_{2r} \end{bmatrix}.$$

It is straightforward that

$$U_0 U_0^T = \frac{1}{2} \sum_{i \in [r]} e_{2i} e_{2i}^T \neq M^*.$$

We prove that U_0 is a second-order critical point of problem (12). By the construction of U_0, we have

$$\|e_i^T U_0\|_F \leq \frac{1}{\sqrt{2}} \leq \alpha, \quad \forall i \in [n].$$

Hence, the regularizer $r(U)$ does not contribute to the local landscape of problem (12) around point U_0 and we only need to prove that U_0 is a second-order critical point of $h(U)$.

For the first-order optimality condition, we can calculate that

$$\nabla f_{3/2}\left[\left(U_0 U_0^T\right)_\Omega; M_\Omega^*\right] = \sum_{i\in[r]} \left(-\omega_{2i-1,2i-1} + \frac{\omega_{2i,2i}}{4}\right) e_{2i-1}e_{2i-1}^T + \sum_{i\in[r]} \frac{1}{2}\left(\omega_{2i,2i} - \omega_{2i-1,2i-1}\right) e_{2i}e_{2i}^T. \tag{23}$$

Therefore, the i-th column of the gradient of $h(U)$ at U_0 is

$$[\nabla h(U_0)]_i = 2\nabla f_{3/2}\left[\left(U_0 U_0^T\right)_\Omega; M_\Omega^*\right](U_0)_i = \sum_{i\in[r]} \left(\omega_{2i,2i} - \omega_{2i-1,2i-1}\right)\omega_{2i,2i} \cdot e_{2i}.$$

If $\omega_{2i-1,2i-1}$ is 0 for some $i \in [r]$, a similar construction as **Case I** shows that function $f_{3/2}$ has multiple global minima. Thus, we only need to consider the case when $\omega_{2i-1,2i-1} = 1$ for all $i \in [r]$ and under this condition, it holds that

$$[\nabla h(U_0)]_i = \sum_{i\in[r]} \left(\omega_{2i,2i} - 1\right)\omega_{2i,2i} \cdot e_{2i} = 0,$$

where the last equality is from the property $\omega_{2i,2i} \in \{0, 1\}$. This verifies the first-order optimality condition of U_0.

Next, we check the second-order necessary optimality condition for U_0. For all direction $K \in \mathbb{R}^{n\times r}$, the curvature of $h(U)$ at U_0 along K is

$$\begin{aligned}
\left[\nabla^2 h(U_0)\right](K,K) &= 2\left\langle \nabla f_{3/2}[(U_0U_0^T)_\Omega; M_\Omega^*], \left(KK^T\right)_\Omega \right\rangle \\
&\quad + \left[\nabla^2 f_{3/2}[(U_0U_0^T)_\Omega; M_\Omega^*]\right]\left[\left(KU_0^T + U_0K^T\right)_\Omega, \left(KU_0^T + U_0K^T\right)_\Omega\right] \\
&= 2\left\langle \nabla f_{3/2}[(U_0U_0^T)_\Omega; M_\Omega^*], \left(KK^T\right)_\Omega \right\rangle \\
&\quad + \left(KU_0^T + U_0K^T\right)_\Omega : \left(\frac{3}{2}\cdot\mathcal{I} + \mathcal{H}\right) : \left(KU_0^T + U_0K^T\right)_\Omega.
\end{aligned}$$

By the event in (22), equation (23) and the condition $\omega_{2i-1,2i-1} = 1$ for all $i \in [r]$, we have

$$\begin{aligned}
&2\left\langle \nabla f_{3/2}[(U_0U_0^T)_\Omega; M_\Omega^*], \left(KK^T\right)_\Omega \right\rangle \\
=&\left\langle \sum_{i\in[r]} \left(-2\omega_{2i-1,2i-1} + \frac{\omega_{2i,2i}}{2}\right) e_{2i-1}e_{2i-1}^T + \sum_{i\in[r]} \left(\omega_{2i,2i} - \omega_{2i-1,2i-1}\right) e_{2i}e_{2i}^T, \left(KK^T\right)_\Omega \right\rangle \\
=&\sum_{i\in[r]} \left(-2\omega_{2i-1,2i-1} + \frac{\omega_{2i,2i}}{2}\right)\omega_{2i-1,2i-1}\|K_{2i-1}\|_F^2 + \sum_{i\in[r]} \left(\omega_{2i,2i} - \omega_{2i-1,2i-1}\right)\omega_{2i,2i}\|K_{2i}\|_F^2 \\
=&\left(-2 + \frac{\omega_{2i,2i}}{2}\right)\sum_{i\in[r]} \|K_{2i-1,:}\|_F^2,
\end{aligned} \tag{24}$$

where $K_{i,:}$ is the i-th row of K for all $i \in [n]$. By the definition of U_0, we can calculate that

$$U_0 K^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & K_{:,1}^T & 0 & K_{:,2}^T & \cdots & K_{:,r}^T & 0 & \cdots \end{bmatrix}, \quad K U_0^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ K_{:,1} \\ 0 \\ K_{:,2} \\ \vdots \\ K_{:,r} \\ 0 \\ \vdots \end{bmatrix}.$$

Therefore, it holds that

$$\begin{aligned} (25) \qquad \left(KU_0^T + U_0K^T\right)_\Omega : \frac{3}{2} \cdot \mathcal{I} : \left(KU_0^T + U_0K^T\right)_\Omega &= \frac{3}{2} \left\| \left(KU_0^T + U_0K^T\right)_\Omega \right\|_F^2 \\ &\geq 3 \sum_{i \in [r]} \omega_{2i,2i} K_{2i,i}^2 + \frac{3}{2} \sum_{i \in [r]} \sum_{j \in [r]} \omega_{2i-1,j} K_{2i-1,j}^2 \end{aligned}$$

and

$$(26) \qquad \left(KU_0^T + U_0K^T\right)_\Omega : \mathcal{H} : \left(KU_0^T + U_0K^T\right)_\Omega = -\sum_{i \in [r]} \omega_{2i,2i} K_{2i,i}^2.$$

Combining the relations in (24)-(26), it follows that

$$\begin{aligned} \left[\nabla^2 h(U_0)\right](K,K) \geq & \left(-2 + \frac{\omega_{2i,2i}}{2}\right) \sum_{i \in [r]} \|K_{2i-1,:}\|_F^2 + \frac{3}{2} \sum_{i \in [r]} \sum_{j \in [r]} \omega_{2i-1,j} K_{2i-1,j}^2 \\ & + 2 \sum_{i \in [r]} \omega_{2i,2i} K_{2i,i}^2 \\ = & \sum_{i \in [r]} \sum_{j \in [r]} \left(-2 + \frac{\omega_{2i,2i}}{2} + \frac{3\omega_{2i-1,j}}{2}\right) K_{2i-1,j}^2 + 2 \sum_{i \in [r]} \omega_{2i,2i} K_{2i,i}^2. \end{aligned}$$

Now, when the event in (22) happens, we have

$$\omega_{2i,2i} = \omega_{2i-1,j} = 1, \quad \forall i, j \in [r].$$

Therefore, we have

$$\left[\nabla^2 h(U_0)\right](K,K) \geq 2 \sum_{i \in [r]} \omega_{2i,2i} K_{2i,i}^2 \geq 0, \quad \forall K \in \mathbb{R}^{n \times r},$$

which is the second-order necessary optimality condition for $h(U)$. In summary, the point U_0 is a spurious second-order critical point of problem (12) with probability at least $(3 - \sqrt{5})/2$.