General Low-rank Matrix Optimization: Geometric Analysis and Sharper Bounds

Haixiang Zhang
Department of Mathematics
University of California, Berkeley
Berkeley, CA 94704
haixiang_zhang@berkeley.edu

Yingjie Bi
Department of IEOR
University of California, Berkeley
Berkeley, CA 94704
yingjiebi@berkeley.edu

Javad Lavaei
Department of IEOR
University of California, Berkeley
Berkeley, CA 94704
lavaei@berkeley.edu

Abstract

This paper considers the global geometry of general low-rank minimization problems via the Burer-Monteiro factorization approach. For the rank-1 case, we prove that there is no spurious second-order critical point for both symmetric and asymmetric problems if the rank-2 RIP constant δ is less than $1/2$. Combining with a counterexample with $\delta = 1/2$, we show that the derived bound is the sharpest possible. For the arbitrary rank-r case, the same property is established when the rank-$2r$ RIP constant δ is at most $1/3$. We design a counterexample to show that the non-existence of spurious second-order critical points may not hold if δ is at least $1/2$. In addition, for any problem with δ between $1/3$ and $1/2$, we prove that all second-order critical points have a positive correlation to the ground truth. Finally, the strict saddle property, which can lead to the polynomial-time global convergence of various algorithms, is established for both the symmetric and asymmetric problems when the rank-$2r$ RIP constant δ is less than $1/3$. The results of this paper significantly extend several existing bounds in the literature.

1 Introduction

Given the natural numbers n, m and r, consider the low-rank matrix optimization problems

$$\min_{M\in\mathbb{R}^{n\times n}} \ f_s(M) \quad \text{s.t. } \operatorname{rank}(M)\le r, \quad M^T = M, \quad M\succeq 0 \tag{1}$$

and

$$\min_{M\in\mathbb{R}^{n\times m}} \ f_a(M) \quad \text{s.t. } \operatorname{rank}(M)\le r, \tag{2}$$

where the functions $f_s(\cdot)$ and $f_a(\cdot)$ are twice continuously differentiable. Problems (1)-(2) are referred to as the *symmetric* and the *asymmetric* problems, respectively. In addition, we call these problems *linear* if the objective function is induced by a linear measurement operator, i.e.,

$$f(M) = \tfrac{1}{2}\|\mathcal{A}(M) - b\|_F^2$$

for some vector $b\in\mathbb{R}^p$ and linear operator $\mathcal{A}$ mapping each matrix M to a vector in $\mathbb{R}^p$, where $f(M)$ denotes either $f_s(M)$ or $f_a(M)$. Those problems not fitting into the above model are called *nonlinear*. One common example with non-linearity is the one-bit matrix sensing problem; please see Zhu et al. (2018); Li et al. (2019); Zhu et al. (2021) for more concrete discussions. Low-rank

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

optimization problems arise in a wide range of applications, e.g., matrix completion (Candès & Recht, 2009; Recht et al., 2010), phase synchronization (Singer, 2011; Boumal, 2016), and phase retrieval (Shechtman et al., 2015); see Chen & Chi (2018); Chi et al. (2019) for an overview of the topic. To overcome the non-convex rank constraint, one may resort to convex relaxations. The approach of replacing the rank constraint with a nuclear norm regularization is proven to provide the optimal sample complexity (Candès & Recht, 2009; Recht et al., 2010; Candès & Tao, 2010). However, solving the convexified problems involves computing a Singular Value Decomposition (SVD) in each iteration and results in heavy computational burdens; see the numerical comparison in Zheng & Lafferty (2015). Along with the issue of large space complexities, the convexification approach is impractical for large-scale problems. Therefore, it is important to design efficient alternative methods with similar theoretical guarantees. Another line of works generalizes techniques from Orthogonal Matching Pursuit (OMP) to the low-rank matrix problem (Shalev-Shwartz et al., 2011; Axiotis & Sviridenko, 2020), which also require implementing SVD for r times in their algorithms.

1.1 Burer-Monteiro factorization and basic properties

Instead of directly solving convex relaxations of problems (1)-(2), we consider a computationally efficient approach, namely the Burer-Monteiro factorization (Burer & Monteiro, 2003). The factorization approach is based on the observation that any matrix $M \in \mathbb{R}^{n\times m}$ with rank at most r can be written in the form of UV^T, where $U \in \mathbb{R}^{n\times r}$ and $V \in \mathbb{R}^{m\times r}$. Then, the asymmetric problem (2) is equivalent to

$$\min_{U\in\mathbb{R}^{n\times r}, V\in\mathbb{R}^{m\times r}} h_a(U,V), \tag{3}$$

where $h_a(U,V) := f_a(UV^T)$. Similarly, the symmetric problem (1) is equivalent to

$$\min_{U\in\mathbb{R}^{n\times r}} h_s(U), \tag{4}$$

where $h_s(U) := f_s(UU^T)$. The Burer-Monteiro factorization provides a natural parameterization of the low-rank structure of the unknown solution, and reformulates problems (1)-(2) as unconstrained optimization problems. In addition, the number of variables reduces from $O(n^2)$ or $O(nm)$ to as low as $O(rn)$ or $O(r(n+m))$ when $r \ll \min\{n,m\}$. However, the reformulated problems are highly non-convex, and $\mathcal{NP}$-hard to solve in general. On the other hand, these problems share a specific non-convex structure, which makes it possible to utilize the structure and design efficient algorithms to find a global optimum under some conditions. In addition to the special structure, a regularity condition, named the Restricted Isometry Property, can be used to guarantee the convergence of common iterative algorithms. We state the following two definitions only in the context of the symmetric problem since the corresponding definitions for the asymmetric problem are similar.

Definition 1 (Recht et al. (2010); Zhu et al. (2018))**.** Given natural numbers r and t, the function $f_s(\cdot)$ is said to satisfy the **Restricted Isometry Property** (RIP) of rank $(2r, 2t)$ for a constant $\delta \in [0,1)$, denoted as δ-RIP$_{2r,2t}$, if

$$(1-\delta)\|K\|_F^2 \le \left[\nabla^2 f_s(M)\right](K,K) \le (1+\delta)\|K\|_F^2$$

holds for all matrices $M, K \in \mathbb{R}^{n\times n}$ such that $\mathrm{rank}(M) \le 2r, \mathrm{rank}(K) \le 2t$, where $\left[\nabla^2 f_s(M)\right](\cdot,\cdot)$ is the curvature of the Hessian at point M.

The RIP condition appears in a variety of applications of the low-rank matrix optimization problem. For instance, in the case of linear measurements with a Gaussian model, Candès & Recht (2009) showed that $O\left(nr/\delta^2\right)$ samples are enough to ensure the δ-RIP$_{2r,2r}$ property with high probability. Please see the survey paper by Chi et al. (2019) for more examples. In certain applications, even the RIP condition cannot be established over the whole low-rank manifold, we are able to establish similar strongly convex and smooth conditions on part of the manifold. If the iteration points of algorithms are constrained to or regularized (either explicitly or implicitly) towards those benign regions, the proof techniques in this work may still be applicable. Examples include the phase retrieval problem (Ma et al., 2018) and the matrix completion problem (Chen et al., 2020). However, the analysis of the case when the strong convexity does not hold is usually application-specific and cannot be generalized to general low-rank problems. Moreover, the RIP assumption is standard in the literature of general low-rank matrix optimization problem. Furthermore, if we drop the strong convexity assumption, we are unable to achieve linear convergence in general (Bhojanapalli et al., 2016a). The

work by Zhang et al. (2019) shows that the existence of RIP is enough to obtain guarantees on the local landscape of the problem and the size of this local region depends on the RIP constant that can be anything between 0 and 1 (however, the provided bounds on the RIP constant are not sharp). Although we aimed to obtain sharp bounds on the RIP constant for global landscape of the problem in this paper, we believe that our analysis can be adopted to obtain sharp RIP bounds for local regions. We leave the precise derivation to a future work since it needs a number of lemma and we have space restrictions. We note that the RIP property is equivalent to the restricted strongly convex and smooth property defined in Wang et al. (2017); Park et al. (2018); Zhu et al. (2021) with the condition number $(1+\delta)/(1-\delta)$. Intuitively, the RIP property implies that the Hessian matrix is close to the identity tensor when the perturbation is restricted to be low-rank. This intuition naturally leads to the following definition.

Definition 2 (Bi & Lavaei (2021))**.** Given a natural number r, the function $f_s(\cdot)$ is said to satisfy the **Bounded Difference Property** (BDP) of rank $2r$ for a constant $\kappa \geq 0$, denoted as κ-BDP$_{2r}$, if

$$\left|\left[\nabla^2 f_s(M) - \nabla^2 f_s(M')\right](K, L)\right| \leq \kappa \|K\|_F \|L\|_F$$

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

It has been proven in Bi & Lavaei (2021, Theorem 1) that those functions satisfying the δ-RIP$_{2r,2r}$ property also satisfy the 4δ-BDP$_{2r}$ property. With the RIP property, there are basically two categories of algorithms that can solve the factorized problem in polynomial time. Algorithms in the first category require a careful initialization so that the initial point is already in a small neighbourhood of a global optimum, and a certain local regularity condition in the neighbourhood ensures that local search algorithms will converge linearly to a global optimum; see Tu et al. (2016); Bhojanapalli et al. (2016a); Park et al. (2018) for a detailed discussion. The other class of algorithms is able to converge globally from a random initialization. The convergence of these algorithms is usually established via the geometric analysis of the landscape of the objective function. One of the important geometric properties is the strict saddle property (Sun et al., 2018), which combined with the smoothness properties can guarantee the global polynomial-time convergence for various saddle-escaping algorithms (Jin et al., 2017, 2018; Sun et al., 2018; Huang & Becker, 2019). For the linear case, Ge et al. (2016, 2017) proved the strict saddle property for both problems (3)-(4) when the RIP constant is sufficiently small. More recently, Zhu et al. (2021) extended the results to the nonlinear asymmetric case. Moreover, a weaker geometric property, namely the non-existence of *spurious (non-global) second-order critical points*, has been established for both problems when the RIP constant is small (Li et al., 2019; Ha et al., 2020). We note that second-order critical points are points that satisfy the first-order and the second-order necessary optimality conditions, and thus the result of the non-existence of second-order critical points implies the non-existence of spurious local minima. Under certain regularity conditions, this weaker property is also able to guarantee the global convergence from a random initialization without an explicit convergence rate (Lee et al., 2016; Panageas & Piliouras, 2016). Please refer to Table 1 for a summary of the state-of-the-art results.

Most of the aforementioned papers are based on the following assumption on the low-rank critical points of the functions $f_s(\cdot)$ and $f_a(\cdot)$:

Assumption 1. The function $f_a(\cdot)$ has a first-order critical point M_a^* such that $\operatorname{rank}(M_a^*) \leq r$. Similarly, the function $f_s(\cdot)$ has a first-order critical point M_s^* that is symmetric, positive semi-definite and of rank at most r.

This assumption is inspired by the noiseless matrix sensing problem in the linear case for which the non-negative objective function becomes zero (the lowest value possible) at the true solution. This is a natural property of the matrix sensing problem for nonlinear measurement models as well. Under the above assumption and the RIP property, Zhu et al. (2018) proved that M_s^* and M_a^* are the unique global minima of problems (1)-(2).

Theorem 1 (Zhu et al. (2018))**.** *If the functions $f_s(\cdot)$ and $f_a(\cdot)$ satisfy the δ-RIP$_{2r,2r}$ property, then the critical points M_s^* and M_a^* are the* unique global minima *of problems* (1)-(2)*.*

Given a solution (U^*, V^*) to problem (3), we observe that (U^*P, V^*P^{-T}) is also a solution for any invertible $P \in \mathbb{R}^{r\times r}$. This redundancy may induce an extreme non-convexity on the landscape of the objective function. To reduce this redundancy, Tu et al. (2016) considered the regularized problem

$$\min_{U\in\mathbb{R}^{n\times r}, V\in\mathbb{R}^{m\times r}} \rho(U, V), \tag{5}$$

Table 1: Comparison of the state-of-the-art results and our results. Here $\delta_{2r,2t}$ and κ are the $\text{RIP}_{2r,2t}$ and BDP_{2r} constants of $f_s(\cdot)$ or $f_a(\cdot)$, respectively. Constant $\alpha(M_a^*) \in (0,1)$ only depends on M_a^*.

Problem Setups		**No Spurious Second-order Critical Pts.**		**Strict Saddle Property**	
		Existing	Ours	Existing	Ours
Rank-1 Sym.	**Linear**	$\delta_{2,2} < \frac{1}{2}$ (Zhang et al., 2019)	$\delta_{2,2} < \frac{1}{2}$	-	-
	Nonlinear	$\delta_{2,2} < \frac{2-O(\kappa)}{4+O(\kappa)}$ (Bi & Lavaei, 2021)	$\delta_{2,2} < \frac{1}{2}$	-	-
Rank-1 Asym.	**Linear & Nonlinear**	-	$\delta_{2,2} < \frac{1}{2}$	-	-
Rank-r Sym.	**Linear**	$\delta_{2r,2r} < \frac{1}{5}$ (Ge et al., 2016)	$\delta_{2r,2r} \leq \frac{1}{3}$	$\delta_{2r,2r} < \frac{1}{10}$ (Ge et al., 2017)	$\delta_{2r,2r} < \frac{1}{3}$
	Nonlinear	$\delta_{2r,4r} < \frac{1}{5}$ (Li et al., 2019)	$\delta_{2r,2r} \leq \frac{1}{3}$	-	$\delta_{2r,2r} < \frac{1}{3}$
Rank-r Asym.	**Linear**	$\delta_{2r,2r} < \frac{1}{3}$ (Ha et al., 2020)	$\delta_{2r,2r} \leq \frac{1}{3}$	$\delta_{2r,2r} < \frac{1}{20}$ (Ge et al., 2017)	$\delta_{2r,2r} < \frac{1}{3}$
	Nonlinear	$\delta_{2r,2r} < \frac{1}{3}$ (Ha et al., 2020)	$\delta_{2r,2r} \leq \frac{1}{3}$	$\delta_{2r,4r} < \frac{\alpha(M_a^*)}{100}$ (Zhu et al., 2021)	$\delta_{2r,2r} < \frac{1}{3}$

where

$$\rho(U,V) := h_a(U,V) + \frac{\mu}{4} \cdot g(U,V)$$

with a constant $\mu > 0$ and the regularization term

$$g(U,V) := \|U^T U - V^T V\|_F^2.$$

The regularization term is introduced to balance the magnitudes of U^* and V^*. Zhu et al. (2018) showed that the regularization term does not introduce bias and thus problem (5) is equivalent to the original problem (2) in the sense that any first-order critical point (U,V) of problem (2) corresponds to a first-order critical point of problem (5) with balanced energy, i.e. $U^T U = V^T V$.

Theorem 2 (Zhu et al. (2018))**.** *Any first-order critical point* (U^*, V^*) *of problem* (5) *satisfies* $(U^*)^T U^* = (V^*)^T V^*$. *Moreover, problems* (2) *and* (5) *are equivalent.*

Detailed optimality conditions for problems (1)-(5) are provided in the appendix.

1.2 Contributions

In this work, we analyze the geometric properties of problems (4)-(5). Novel analysis methods are developed to obtain less conservative conditions for guaranteeing benign landscapes for both problems. We note that, unlike the linear measurements case, the RIP constant of nonlinear problems may not concentrate to 0 as the number of samples increases. Therefore, a sharper RIP bound leads to theoretical guarantees that hold under less stringent statistical requirements. In addition, even if the RIP constant concentrates to 0 when more samples are included, there may only be a limited number of samples available, either due to the constraints of specific applications or to the great expense of taking more samples. Hence, obtaining a sharper RIP bound is essential for many applications. We summarize our results in Table 1. More concretely, the contributions of this paper are three-folds.

First, we derive necessary conditions and sufficient conditions for the existence of spurious second-order critical points for both symmetric and asymmetric problems. Using our necessary conditions, we show that the δ-$\text{RIP}_{2r,2r}$ property with $\delta \leq 1/3$ is enough to guarantee the non-existence of such points. This result provides a marginal improvement to the previous work (Ha et al., 2020), which developed the sufficient condition $\delta < 1/3$ for asymmetric problems, and is a major improvement over Ge et al. (2016) and Li et al. (2019), which requires $\delta < 1/5$ for symmetric problems. With this non-existence property and under some common regularity conditions, Lee et al. (2016); Panageas &

Piliouras (2016) showed that the vanilla gradient descent method with a small enough step size and a random initialization almost surely converges to a global minimum. We note that the convergence rate was not studied and could theoretically be exponential in the worst case. In addition, by studying our necessary conditions, we show that every second-order critical point has a positive correlation to the global minimum when $\delta \in (1/3, 1/2)$. When $\delta = 1/2$, a counterexample with spurious second-order critical points is given by utilizing the sufficient conditions. We note that the sufficient conditions can greatly simplify the construction of counterexamples.

Second, we separately study the rank-1 case to further strengthen the bounds. In particular, we utilize the necessary conditions to prove that the δ-RIP$_{2,2}$ property with $\delta < 1/2$ is enough for the non-existence of spurious second-order critical points. Combining with a counterexample in the $\delta = 1/2$ case, we conclude that the bound $\delta < 1/2$ is the sharpest bound for the rank-1 case. Our results significantly extend the bounds in Zhang et al. (2019) derived for the linear symmetric case to the linear asymmetric case and the general nonlinear case. It also improves the bound in Bi & Lavaei (2021) by dropping the BDP constant.

Third, we prove that in the exact parametrization case, problems (4)-(5) both satisfy the strict saddle property (Sun et al., 2018) when the δ-RIP$_{2r,2r}$ property is satisfied with $\delta < 1/3$. This result greatly improves the bounds in Ge et al. (2017); Zhu et al. (2021) and extends the result in Ha et al. (2020) to approximate second-order critical points. With the strict saddle property and certain smoothness properties, a wide range of algorithms guarantee a global polynomial-time convergence with a random initialization; see Jin et al. (2017, 2018); Sun et al. (2018); Huang & Becker (2019). Due to the special non-convex structure of our problems and the RIP property, it is possible to prove the boundedness of the trajectory of the perturbed gradient descent method using a similar method as in Jin et al. (2017). Since the smoothness properties are satisfied over a bounded region, combined with the strict saddle property, it follows that the perturbed gradient descent method (Jin et al., 2017) achieves a polynomial-time global convergence when $\delta < 1/3$.

1.3 Notation and organization

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 $\langle M, N\rangle := \mathrm{tr}(M^T N)$. For any matrix $M \in \mathbb{R}^{n\times m}$, we denote its singular values by $\sigma_1(M) \geq \cdots \geq \sigma_k(M)$, where $k := \min\{n, m\}$. For any symmetric matrix $M \in \mathbb{R}^{n\times n}$, we denote its eigenvalues by $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$. The minimal eigenvalue is denoted as $\lambda_{min}(\cdot)$. For any matrix U, we use $\mathcal{P}_U$ to denote the orthogonal projection onto the column space of U. For any 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 symmetric and positive semi-definite. 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 $[\nabla^2 f(M)](K, L) := \sum_{i,j,k,\ell}[\nabla^2 f(M)]_{i,j,k,\ell}K_{ij}L_{k,\ell}$.

In Section 2, the Singular Value Projection algorithm is analyzed as an enlightening example for our main results. Sections 3 and 4 are devoted to the non-existence of spurious second-order critical points and the strict saddle property of the low-rank optimization problem in both symmetric and asymmetric cases, respectively.

2 Motivating Example: Singular Value Projection Algorithm

Before providing theoretical results for problems (4)-(5), we first consider the Singular Value Projection Method (SVP) algorithm (Algorithm 1) as a motivating example, which is proposed in Jain et al. (2010). The SVP algorithm is basically the projected gradient method of the original low-rank problems (1)-(2) via the truncated SVD. For the asymmetric problem (2), the low-rank manifold is

$$\mathcal{M}_{asym} := \{M \in \mathbb{R}^{n\times m} \mid \mathrm{rank}(M) \leq r\}$$

and the projection is given by only keeping components corresponding to the r largest singular values. For the symmetric problem (1), the low-rank manifold is

$$\mathcal{M}_{sym} := \{M \in \mathbb{R}^{n\times n} \mid \mathrm{rank}(M) \leq r, \quad M^T = M, \quad M \succeq 0\}.$$

We assume without loss of generality that the gradient $\nabla f(\cdot)$ is symmetric; see Appendix A for a discussion. The projection is given by only keeping components corresponding to the r largest

Algorithm 1 Singular Value Projection (SVP) Algorithm

Input: Low-rank manifold $\mathcal{M}$, initial point M_0, number of iterations T, step size η, objective function $f(\cdot)$.
Output: Low-rank solution M_T.
1: **for** $t = 0, \ldots, T-1$ **do**
2: Update $\tilde{M}_{t+1} \leftarrow M_t - \eta \nabla f(M_t)$.
3: Set M_{t+1} to be the projection of $\tilde{M}_{t+1}$ onto $\mathcal{M}$ via truncated SVD.
4: **end for**
5: **return** M_T.

eigenvalues and dropping all components with negative eigenvalues. Since both low-rank manifolds are non-convex, the projection solution may not be unique and we choose an arbitrary solution when it is not unique. We note that the above projections are orthogonal in the sense that

$$\|M_+ - M\|_F = \min_{K \in \mathcal{M}} \|K - M\|_F,$$

where M_+ is the projection of a matrix M. Henceforth, $\mathcal{M}$ stands for $\mathcal{M}_{sym}$ or $\mathcal{M}_{asym}$, which should be clear from the context. Although each truncated SVD operation can be computed within $O(nmr)$ operations, the constant hidden in the $O(\cdot)$ notation is considerably larger than 1. Thus, the truncated SVD operation is significantly slower than matrix multiplication, which makes the SVP algorithm impractical for large-scale problems. However, the analysis of the SVP algorithm, combining with the equivalence property given in Ha et al. (2020), provides some insights into how to develop proof techniques for problems (4)-(5).

We extend the proof in Jain et al. (2010) and show that Algorithm 1 converges linearly to the global minimum under the δ-RIP$_{2r,2r}$ property with $\delta < 1/3$.

Theorem 3. *If function $f_s(\cdot)$ (resp. $f_a(\cdot)$) satisfies the δ-RIP$_{2r,2r}$ property with $\delta < 1/3$ and the step size is chosen to be $\eta = (1+\delta)^{-1}$, then Algorithm 1 applied to problem* (1) *(resp.* (2)*) returns a solution M_T such that $M_T \in \mathcal{M}$ and $f(M_T) - f(M^*) \leq \epsilon$ within*

$$T := \left\lceil \frac{1}{\log[(1-\delta)/(2\delta)]} \cdot \log \left[\frac{f(M_0) - f(M^*)}{\epsilon} \right] \right\rceil$$

iterations, where $f(\cdot) := f_s(\cdot)$ (resp. $f(\cdot) := f_a(\cdot)$), M^* is the global minimum, M_0 is the initial point and $\lceil \cdot \rceil$ is the ceiling function.

The proof is almost identical to that in Jain et al. (2010) except that we have replaced the quadratic function with the RIP bounds. However, the result of the proof provides a key inequality (13) for the subsequent proofs. We note that the above proof can be applied to other low-rank optimization problems with a suitable definition of the orthogonal projection. In Ha et al. (2020), it is proved that the unique global minimum is the only fixed point of the SVP algorithm if the RIP constant δ is less than $1/3$. However, the above paper has not proven the linear convergence (as done in Theorem 3). This difference leads to a strengthened inequality in the following analysis, which further serves as an essential step in proving the strict saddle property. The results in this section provide a hint that the landscape may be benign when the RIP constant is smaller than $1/3$ and we may be able to establish linear convergence under this condition, which is the main topic of the remainder of this paper.

3 No Spurious Second-order Critical Points

In this section, we develop necessary conditions and sufficient conditions for the existence of spurious second-order critical points of problems (4)-(5). Besides the non-existence of spurious local minima, the non-existence of spurious second-order critical points also guarantees the global convergence of many first-order algorithms with random initialization under certain regularity conditions (Lee et al., 2016; Panageas & Piliouras, 2016). More precisely, we require the iterates of the algorithm to converge to a single point and the objective function to have a Lipschitz-continuous gradient. The first condition is satisfied by the gradient descent method applied to a large class of functions known as the KŁ-functions (Attouch et al., 2013). For the second condition, many objective functions that appear in applications, e.g., the ℓ_2-loss function, do not satisfy this condition. However, if the step

size is small enough, the special non-convex structure of the Burer-Monteiro decomposition and the RIP property ensure that the trajectory of the gradient descent method stays in a compact set, where the Lipschitz condition is satisfied due to the second-order continuity of the functions $f_s(\cdot)$ and $f_a(\cdot)$. The proof of this claim is similar to Theorem 8 in Jin et al. (2017) and is omitted here. Therefore, the non-existence of spurious second-order critical points can ensure the global convergence of the gradient descent method for many applications.

The non-existence of spurious second-order critical points has been proved in Ge et al. (2017); Zhu et al. (2018) for problems with linear and nonlinear measurements, respectively. Recently, Ha et al. (2020) proved a relation between the second-order critical points of problem (3) or (5) and the fixed points of the SVP algorithm on problem (2). Using this relation, they showed that problems (3) and (5) have no spurious second-order critical points when the δ-RIP$_{2r,2r}$ property is satisfied with $\delta < 1/3$. In this work, we take a different approach to show that $\delta \leq 1/3$ is enough for the general case in both symmetric and asymmetric scenarios, and that $\delta < 1/2$ is enough for the rank-1 case. Moreover, we prove that there exists a positive correlation between every second-order critical point and the global minimum when $\delta \in (1/3, 1/2)$. We also show that there may exist spurious second-order critical points when $\delta = 1/2$ for both the symmetric and asymmetric problems, which extends the construction of such examples for the linear symmetric rank-1 problem in Zhang et al. (2018) to general cases. We first give necessary conditions and sufficient conditions for the existence of a function that satisfies the δ-RIP$_{2r,2r}$ condition and spurious second-order critical points below.

Theorem 4. *Let $\ell := \min\{m, n, 2r\}$. For a given $\delta \in [0, 1)$, there exists a function $f_a(\cdot)$ with the δ-RIP$_{2r,2r}$ property such that problems* (3) *and* (5) *have a spurious second-order critical point only if $1 - \delta < (1 + \delta)/2$ and there exists a constant $\alpha \in (1 - \delta, (1 + \delta)/2]$, a diagonal matrix $\Sigma \in \mathbb{R}^{r\times r}$, a diagonal matrix $\Lambda \in \mathbb{R}^{(\ell-r)\times(\ell-r)}$ and matrices $A \in \mathbb{R}^{r\times r}, B \in \mathbb{R}^{r\times r}, C \in \mathbb{R}^{(\ell-r)\times r}, D \in \mathbb{R}^{(\ell-r)\times r}$ such that*

$$
\begin{aligned}
&(1+\delta)\min_{1\leq i\leq r}\Sigma_{ii} \geq \max_{1\leq i\leq \ell-r}\Lambda_{ii}, \quad \Sigma \succ 0,\ \Lambda \succeq 0,\\
&\langle \Lambda, CD^T\rangle = \alpha\left[\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, AB^T\rangle + \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2\right], \quad (6)\\
&\mathrm{tr}(\Lambda^2) \leq \alpha^{-1}(2\alpha - 1 + \delta^2)\cdot\langle \Lambda, CD^T\rangle, \quad \langle \Lambda, CD^T\rangle \neq 0.
\end{aligned}
$$

Remark 1. We note that there may exist simpler forms of the above conditions. For instance, we may solve α via the condition in the second line of (6) and substitute into other conditions. In addition, the requirement that $\alpha \in (1 - \delta, (1 + \delta)/2]$ may also be dropped without affecting the conditions. More specifically, the conditions in (6) are equivalent to

$$
\begin{aligned}
&(1+\delta)\min_{1\leq i\leq r}\Sigma_{ii} \geq \max_{1\leq i\leq \ell-r}\Lambda_{ii}, \quad \Sigma \succ 0,\ \Lambda \succeq 0, \quad \langle \Lambda, CD^T\rangle \neq 0,\\
&\mathrm{tr}(\Lambda^2) \leq 2\cdot\langle \Lambda, CD^T\rangle - (1-\delta^2)\Big[\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, AB^T\rangle\\
&\qquad\qquad + \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2\Big].
\end{aligned}
$$

We state Theorem 4 in the current form since it helps with deriving corollaries more directly.

Intuitively, Λ and Σ correspond to the singular values of the second-order critical point and the gradient at the second-order critical point, respectively. Matrices A, B, C, D correspond to the SVD of the global optimum. The original problem of the non-existence of spurious second-order critical points can be viewed as a property of the set of functions satisfying the RIP property, which is a convex set in an infinite-dimensional functional space. The conditions in (6) reduce the infinite-dimensional problem to a finite-dimensional problem by utilizing the optimality conditions and the RIP property, which provides a basis for solving these conditions numerically. We note that the conditions in the third line of (6) are novel and serve as an important step in developing strong theoretical guarantees. Although the conditions in (6) seem complicated, they lead to strong results on the non-existence of spurious second-order critical points. We provide two corollaries below to illustrate the power of the above theorem. The first corollary focuses on the rank-1 case. In this case, we can simplify condition

(6) through suitable relaxations to obtain a sharper bound on δ that ensures the non-existence of spurious second-order critical points.

Corollary 1. *Consider the case $r = 1$, and suppose that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2,2}$ property with $\delta < 1/2$. Then, problems* (3) *and* (5) *have no spurious second-order critical points.*

The following example shows that the counterexample in Zhang et al. (2019) designed for the symmetric case also works for the asymmetric rank-1 case.

Example 1. We note that Example 12 in Zhang et al. (2019) shows that problem (4) may have spurious second-order critical points when $\delta = 1/2$. In general, a second-order critical point for problem (4) is not a second-order critical point for problem (5), since the asymmetric manifold $\mathcal{M}_{asym}$ has a larger second-order critical cone than the symmetric manifold $\mathcal{M}_{sym}$. However, it can be verified that the same example also has a spurious second-order critical point in the asymmetric case. For completeness, we verify the claim in the appendix.

It follows from Corollary 1 and Example 1 that the bound $1/2$ is the *sharpest* bound for the rank-1 asymmetric case. The next corollary provides a marginal improvement to the state-of-the-art result for the general rank case, which derives the RIP bound $\delta < 1/3$ (Ha et al., 2020). In addition, we prove that there exists a positive correlation between every second-order critical point and the global minimum when $\delta < 1/2$.

Corollary 2. *Given an arbitrary r, suppose that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2r,2r}$ property. If $\delta \leq 1/3$, then both problems* (3) *and* (5) *have no spurious second-order critical points. In addition, if $\delta \in [0, 1/2)$, then every second-order critical point $\tilde{M}$ has a positive correlation with the ground truth M_a^*. Namely, there exists a universal function $C(\delta) : (0, 1/2) \mapsto (0, 1]$ such that*

$$\langle \tilde{M}, M_a^* \rangle \geq C(\delta) \cdot \|\tilde{M}\|_F \|M_a^*\|_F.$$

For the general rank-r case, we construct a counterexample with spurious second-order critical points when $\delta = 1/2$.

Example 2. Let $n = m = 2r$. Now, we use the sufficiency part of Theorem 4 to construct a counterexample. We choose

$$\delta := \frac{1}{2}, \quad \alpha := \frac{3}{5}, \quad \Sigma := \frac{1}{2} I_r, \quad \Lambda := \frac{3}{4} I_r, \quad A = B := 0_r, \quad C = D := I_r.$$

It can be verified that the conditions in (6) are satisfied and $CB^T = AD^T = 0$, which means that there exists a function $f_a(\cdot)$ satisfying the δ-RIP$_{2r,2r}$ property for which problems (3) and (5) have spurious second-order critical points. We also give a direct construction with linear measurements in the appendix. This example illustrates that Theorem 4 can be used to systematically design instances of the problem with spurious second-order critical points.

Before closing this section, we note that similar conditions can be obtained for the symmetric problem (4). Although there exists a natural transformation of symmetric problems to asymmetric problems (see the appendix), the approach requires the objective function $f_s(\cdot)$ to have the δ-RIP$_{4r,2r}$ property, which provides sub-optimal RIP bounds compared to a direct analysis. We give the results of the direct analysis below and omit the proof due to the similarity to the asymmetric case.

Theorem 5. *Let $\ell := \min\{n, 2r\}$. For a given $\delta \in [0, 1)$, there exists a function $f_s(\cdot)$ with the δ-RIP$_{2r,2r}$ property such that problem* (4) *has a spurious second-order critical point only if $1 - \delta < (1 + \delta)/2$ and there exists a constant $\alpha \in (1 - \delta, (1 + \delta)/2]$, a diagonal matrix $\Sigma \in \mathbb{R}^{r \times r}$, a diagonal matrix $\Lambda \in \mathbb{R}^{(\ell - r) \times (\ell - r)}$ and matrices $A \in \mathbb{R}^{r \times r}, C \in \mathbb{R}^{(\ell - r) \times r}$ such that*

$$\begin{aligned} &(1 + \delta) \min_{1 \leq i \leq r} \Sigma_{ii} \geq \max_{1 \leq i \leq \ell - r} \Lambda_{ii}, \quad \Sigma \succ 0, \\ &\langle \Lambda, CC^T \rangle = \alpha \left[\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, AA^T \rangle + \|AA^T\|_F^2 + 2\|AC^T\|_F^2 + \|CC^T\|_F^2 \right], \\ &\mathrm{tr}(\Lambda^2) \leq \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot \langle \Lambda, CC^T \rangle, \quad \langle \Lambda, CC^T \rangle \neq 0. \end{aligned} \tag{7}$$

If $AC^T = 0$, then there exists a function $f_s(\cdot)$ with the δ-RIP$_{2r,2r}$ property for which problem (4) *has a spurious second-order critical point.*

Compared to Theorem 4, the diagonal matrix Λ is not enforced to be positive semi-definite. The reason is that the eigenvalue decomposition is used instead of the singular value decomposition in the symmetric case, and therefore some eigenvalues can be negative. Similarly, we can obtain the non-existence and the positive correlation results for the symmetric problem.

Corollary 3. *If function $f_s(\cdot)$ satisfies the δ-RIP$_{2r,2r}$ property, then the following statements hold:*

- *If $\delta \leq 1/3$, then there are no spurious second-order critical points;*
- *If $\delta < 1/2$, then there exists a positive correlation between every second-order critical point and the ground truth;*
- *If $\delta = 1/2$, then there exists a counterexample with spurious second-order critical points;*
- *If $\delta < 1/2$ and $r = 1$, then there are no spurious second-order critical points.*

We note that the last statement serves as a generalization of the results in Zhang et al. (2019) to the nonlinear measurement case, and improves upon the bound in Bi & Lavaei (2021) by dropping the BDP constant.

4 Global Landscape: Strict Saddle Property

Although the non-existence of spurious second-order critical points can ensure the global convergence under certain regularity conditions, it cannot guarantee a fast convergence rate in general. Saddle-point escaping algorithms may become stuck at approximate second-order critical points for exponentially long time. To guarantee the global polynomial-time convergence, the following strict saddle property is commonly considered in the literature:

Definition 3 (Sun et al. (2018))**.** Consider an arbitrary optimization problem $\min_{x\in\mathcal{X}\subset\mathbb{R}^d} F(x)$ and let $\mathcal{X}^*$ denote the set of its global minima. It is said that the problem satisfies the (α,β,γ)-**strict saddle property** for $\alpha,\beta,\gamma > 0$ if at least one of the following conditions is satisfied for every $x \in \mathcal{X}$:

$$\operatorname{dist}(x,\mathcal{X}^*) \leq \alpha; \quad \|\nabla F(x)\|_F \geq \beta; \quad \lambda_{min}[\nabla^2 F(x)] \leq -\gamma.$$

For the low-rank problems, we choose the distance to be the Frobenius norm in the factorization space. This distance is equivalent to the Frobenius norm in the matrix space in the sense that there exist constants $c_1(\mathcal{X}^*) > 0$ and $c_2(\mathcal{X}^*) > 0$ such that

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

holds for all $U \in \mathcal{X}$ as long as $\|U-U^*\|_F$ is small and $\mathcal{X}^*$ is bounded (Tu et al., 2016). A similar relation holds for the asymmetric case.

In Jin et al. (2017), it has been proved that the perturbed gradient descent method can find an ϵ-approximate second-order critical point in $\tilde{O}(\epsilon^{-2})$ iterations with high probability if the Hessian of the objective function is Lipschitz. Namely, the algorithm returns a point $x \in \mathcal{X}$ such that

$$\|\nabla F(x)\|_F \leq O(\epsilon), \quad \lambda_{min}[\nabla^2 F(x)] \geq -O(\sqrt{\epsilon})$$

in $\tilde{O}(\epsilon^{-2})$ iterations with high probability. If we choose $\epsilon > 0$ to be small enough such that $O(\epsilon) < \beta$ and $-O(\sqrt{\epsilon}) > -\gamma$, then the strict saddle property ensures that the returned point satisfies $\operatorname{dist}(x,\mathcal{X}^*) \leq \alpha$ with high probability. We note that the Lipschitz continuity of the Hessian can be similarly guaranteed by the boundedness of trajectories of the perturbed gradient method, which can be proved similarly as Theorem 8 in Jin et al. (2017). Since the smoothness properties are satisfied over a bounded region, we may apply the perturbed gradient descent method (Jin et al., 2017) to achieve the polynomial-time global convergence with random initialization.

In this section, we prove that problems (4) and (5) satisfy the strict saddle property with an arbitrary $\alpha > 0$ in the exact parameterization case, i.e., when the global optimum has rank r.

Assumption 2. The global optimum M_a^* or M_s^* has rank r.

It has been proved in Zhu et al. (2021) that the regularized problem (5) satisfies the strict saddle property if the function $f_a(\cdot)$ has the δ-RIP$_{2r,4r}$ property with

$$\delta < \frac{\sigma_r(M_a^*)^{3/2}}{100\|M_a^*\|_F\|M_a^*\|_2^{1/2}}.$$

Our results improve upon their bounds by allowing a larger problem-free RIP constant and requiring only the RIP$_{2r,2r}$ property (note that there are problems with RIP$_{2r,2r}$ property for which the RIP$_{2r,4r}$ property does not hold (Bi & Lavaei, 2021)). Our result can also be viewed as a robust version of the results in Ha et al. (2020).

Theorem 6. *Suppose that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2r,2r}$ property with $\delta < 1/3$. Given an arbitrary constant $\alpha > 0$, if μ is selected to belong to the interval $[(1-\delta)/3, 1-\delta)$, then there exist positive constants*

$$\epsilon_1 := \epsilon_1(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha), \quad \lambda_1 := \lambda_1(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha)$$

such that for every $\epsilon \in (0, \epsilon_1]$ and $\lambda \in (0, \lambda_1]$, problem (5) *satisfies the (α, β, γ)-strict saddle property with*

$$\beta := \min\left\{\mu(\epsilon/r)^{3/2}, \lambda\right\}, \quad \gamma := \mu\epsilon.$$

We note that the constraint $\mu \in [(1-\delta)/3, 1-\delta)$ is not optimal and it can be similarly proved that $\mu \in (\delta, 1-\delta)$ also guarantees the strict saddle property. The key step in the proof is to show that for every point (U, V) at which the gradient of $f_a(UV^T)$ is small, it holds that

$$\|\nabla f_a(UV^T)\|_2^2 \geq (1+\delta)^2\sigma_r^2(UV^T) + C \cdot (1-3\delta)[f_a(UV^T) - f_a(M_a^*)],$$

where $C > 0$ is a constant independent of (U, V). This inequality can be viewed as a major extension of the non-existence of spurious second-order critical points when $\delta < 1/3$ (Ha et al., 2020), which shows that every spurious second-order critical point (U, V) satisfies

$$\|\nabla f_a(UV^T)\|_2^2 > (1+\delta)\sigma_r^2(UV^T).$$

We emphasize that our proof requires a new framework and is not a standard revision of the existing methods, which is the reason why sharper bounds can be established. By replacing $\|\nabla f_a(M)\|_2$ with $-\lambda_{min}(\nabla f_s(M))$, the analysis for the asymmetric case can be extended to the symmetric case with minor modifications and the same bound follows.

Theorem 7. *Suppose that the function $f_s(\cdot)$ satisfies the δ-RIP$_{2r,2r}$ property with $\delta < 1/3$. Given an arbitrary constant $\alpha > 0$, there exists a positive constant $\lambda_1 := \lambda_1(\delta, r, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha)$ such that for every $\lambda \in (0, \lambda_1]$, problem* (4) *satisfies the (α, β, γ)-strict saddle property with*

$$\beta := \lambda, \quad \gamma := 2\lambda.$$

The above bound is the first theoretical guarantee of the strict saddle property for the nonlinear symmetric problem.

5 Conclusion

In this work, we analyze the geometric properties of low-rank optimization problems via the non-convex factorization approach. We prove novel necessary conditions and sufficient conditions for the non-existence of spurious second-order critical points in both symmetric and asymmetric cases. We show that these conditions lead to sharper bounds and greatly simplify the construction of counterexamples needed to study the sharpness of the bounds. The developed bounds significantly generalize several of the existing results. In the rank-1 case, the bound is proved to be the sharpest possible. In the general rank case, we show that there exists a positive correlation between second-order critical points and the global minimum for problems whose RIP constants are higher than the developed bound but lower than the fundamental limit obtained by the counterexamples. Finally, the strict saddle property is proved with a weaker requirement on the RIP constant for asymmetric problems. The paper develops the first strict saddle property in the literature for nonlinear symmetric problems.

Acknowledgments and Disclosure of Funding

This work was supported by grants from AFOSR, ARO, ONR, NSF and C3.ai Digital Transformation Institute.

References

Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1):91–129, 2013.

Kyriakos Axiotis and Maxim Sviridenko. Sparse convex optimization via adaptively regularized hard thresholding. In *International Conference on Machine Learning*, pp. 452–462. PMLR, 2020.

Srinadh Bhojanapalli, Anastasios Kyrillidis, and Sujay Sanghavi. Dropping convexity for faster semi-definite optimization. In *Conference on Learning Theory*, pp. 530–582. PMLR, 2016a.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 3880–3888, 2016b.

Yingjie Bi and Javad Lavaei. 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.

Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Ji Chen, Dekai Liu, and Xiaodong Li. Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Transactions on Information Theory*, 66(9):5806–5841, 2020.

Yudong Chen and Yuejie Chi. Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization. *IEEE Signal Processing Magazine*, 35(4):14–31, 2018.

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.

Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *Advances in Neural Information Processing Systems*, pp. 2981–2989, 2016.

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*, pp. 1233–1242. PMLR, 2017.

Wooseok Ha, Haoyang Liu, and Rina Foygel Barber. An equivalence between critical points for rank constraints versus low-rank factorizations. *SIAM Journal on Optimization*, 30(4):2927–2955, 2020.

Zhishen Huang and Stephen Becker. Perturbed proximal descent to escape saddle points for non-convex and non-smooth objective functions. In *INNS Big Data and Deep Learning conference*, pp. 58–77. Springer, 2019.

Prateek Jain, Raghu Meka, and Inderjit Dhillon. Guaranteed rank minimization via singular value projection. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems-Volume 1*, pp. 937–945, 2010.

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*, pp. 1724–1732. PMLR, 2017.

Chi Jin, Praneeth Netrapalli, and Michael I Jordan. Accelerated gradient descent escapes saddle points faster than gradient descent. In *Conference On Learning Theory*, pp. 1042–1085. PMLR, 2018.

Jason D Lee, Max Simchowitz, Michael I Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. In *Conference on learning theory*, pp. 1246–1257. PMLR, 2016.

Qiuwei Li, Zhihui Zhu, and Gongguo Tang. The non-convex geometry of low-rank matrix optimization. *Information and Inference: A Journal of the IMA*, 8(1):51–96, 2019.

Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *International Conference on Machine Learning*, pp. 3345–3354. PMLR, 2018.

Ioannis Panageas and Georgios Piliouras. Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions. *arXiv preprint arXiv:1605.00405*, 2016.

Dohyung Park, Anastasios Kyrillidis, Constantine Caramanis, and Sujay Sanghavi. Finding low-rank solutions via nonconvex matrix factorization, efficiently and provably. *SIAM Journal on Imaging Sciences*, 11(4):2165–2204, 2018.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

Shai Shalev-Shwartz, Alon Gonen, and Ohad Shamir. Large-scale convex minimization with a low-rank constraint. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 329–336, 2011.

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

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

Stephen Tu, Ross Boczar, Max Simchowitz, Mahdi Soltanolkotabi, and Ben Recht. Low-rank solutions of linear matrix equations via Procrustes flow. In *International Conference on Machine Learning*, pp. 964–973. PMLR, 2016.

Lingxiao Wang, Xiao Zhang, and Quanquan Gu. A unified computational and statistical framework for nonconvex low-rank matrix estimation. In *Artificial Intelligence and Statistics*, pp. 981–990. PMLR, 2017.

Richard Y Zhang, Cédric Josz, Somayeh Sojoudi, and Javad Lavaei. How much restricted isometry is needed in nonconvex matrix recovery? In *NeurIPS*, 2018.

Richard Y Zhang, Somayeh Sojoudi, and Javad Lavaei. Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery. *Journal of Machine Learning Research*, 20(114):1–34, 2019.

Qinqing Zheng and John Lafferty. 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.

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.

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.

A Optimality Conditions

In this section, we develop the optimality conditions for problems (1)-(5). We assume without loss of generality that $\nabla f_s(M)$ is symmetric for every $M \in \mathbb{R}^{n\times n}$. This is because we can always optimize the equivalent problem

$$\min_{M\in\mathbb{R}^{n\times n}} \frac{1}{2}\left[f_s(M) + f_s(M^T)\right] \quad \text{s.t. } \operatorname{rank}(M) \le r, \quad M^T = M, \quad M \succeq 0.$$

We first consider problems (1) and (2).

Theorem 8 (Li et al. (2019); Ha et al. (2020))**.** *The matrix $\tilde{M} = \tilde{U}\tilde{U}^T$ with $\tilde{U} \in \mathbb{R}^{n\times r}$ is a first-order critical point of the constrained problem* (1) *if and only if*

$$\begin{cases} \nabla f_s(\tilde{M})\tilde{U} = 0 & \text{if } \operatorname{rank}(\tilde{M}) = r \\ \nabla f_s(\tilde{M}) \succeq 0 & \text{if } \operatorname{rank}(\tilde{M}) < r. \end{cases}$$

The matrix $\tilde{M} = \tilde{U}\tilde{V}^T$ with $\tilde{U} \in \mathbb{R}^{n\times r}$ and $\tilde{V} \in \mathbb{R}^{m\times r}$ is a first-order critical point of the constrained problem (2) *if and only if*

$$\begin{cases} [\nabla f_a(\tilde{M})]^T\tilde{U} = 0,\ \nabla f_a(\tilde{M})\tilde{V} = 0 & \text{if } \operatorname{rank}(\tilde{M}) = r \\ \nabla f_a(\tilde{M}) = 0 & \text{if } \operatorname{rank}(\tilde{M}) < r. \end{cases}$$

In Ha et al. (2020), the authors proved that each second-order critical point of problem (3) or (5) is a fixed point of the SVP algorithm run on problem (2). We note that this relation can be extended to the symmetric and positive semi-definite case. This relation plays an important role in the analysis of Section 3.

Theorem 9 (Ha et al. (2020))**.** *The matrix $\tilde{M} = \tilde{U}\tilde{U}^T$ with $\tilde{U} \in \mathbb{R}^{n\times r}$ is a fixed point of the SVP algorithm run on problem* (1) *with the step size $1/(1+\delta)$ if and only if*

$$\nabla f_s(\tilde{M})\tilde{U} = 0, \quad -\lambda_{min}(\nabla f_s(\tilde{M})) \le (1+\delta)\sigma_r(\tilde{U}).$$

The matrix $\tilde{M} = \tilde{U}\tilde{V}^T$ with $\tilde{U} \in \mathbb{R}^{n\times r}$ and $\tilde{V} \in \mathbb{R}^{m\times r}$ is a fixed point of the SVP algorithm run on problem (2) *with the step size $1/(1+\delta)$ if and only if*

$$[\nabla f_a(\tilde{M})]^T\tilde{U} = 0, \quad \nabla f_a(\tilde{M})\tilde{V} = 0, \quad \|\nabla f_a(\tilde{M})\|_2 \le (1+\delta)\sigma_r(\tilde{M}).$$

Next, we consider problems (3)-(5). Since the goal is to study only spurious local minima and saddle points, it is enough to focus on the second-order necessary optimality conditions. The following two theorems follow from basic calculations and we omit the proof.

Theorem 10. *The matrix $\tilde{U} \in \mathbb{R}^{n\times r}$ is a second-order critical point of problem* (4) *if and only if*

$$\nabla f_s(\tilde{U}\tilde{U}^T)\tilde{U} = 0$$

and

$$2\langle \nabla f_s(\tilde{U}\tilde{U}^T), \Delta\Delta^T\rangle + [\nabla^2 f_s(\tilde{U}\tilde{U}^T)](\tilde{U}\Delta^T + \Delta\tilde{U}^T, \tilde{U}\Delta^T + \Delta\tilde{U}^T) \ge 0$$

holds for every $\Delta \in \mathbb{R}^{n\times r}$.

Theorem 11. *The point $(\tilde{U}, \tilde{V})$ with $\tilde{U} \in \mathbb{R}^{n\times r}$ and $\tilde{V} \in \mathbb{R}^{m\times r}$ is a second-order critical point of problem* (3) *if and only if*

$$\nabla[f_a(\tilde{U}\tilde{V}^T)]^T\tilde{U} = 0, \quad \nabla f_a(\tilde{U}\tilde{V}^T)\tilde{V} = 0$$

and

$$2\langle \nabla f_a(\tilde{U}\tilde{V}^T), \Delta_U\Delta_V^T\rangle + [\nabla^2 f_a(\tilde{U}\tilde{V}^T)](\tilde{U}\Delta_V^T + \Delta_U\tilde{V}^T, \tilde{U}\Delta_V^T + \Delta_U\tilde{V}^T) \ge 0$$

holds for every $\Delta_U \in \mathbb{R}^{n\times r}$ and $\Delta_V \in \mathbb{R}^{m\times r}$. Moreover, the given point is a a second-order critical point of problem (5) *if and only if*

$$\nabla[f_a(\tilde{U}\tilde{V}^T)]^T\tilde{U} = 0, \quad \nabla f_a(\tilde{U}\tilde{V}^T)\tilde{V} = 0, \quad \tilde{U}^T\tilde{U} = \tilde{V}^T\tilde{V}$$

and

$$\begin{aligned} 2\langle \nabla f_a(\tilde{U}\tilde{V}^T), \Delta_U\Delta_V^T\rangle + [\nabla^2 f_a(\tilde{U}\tilde{V}^T)](\tilde{U}\Delta_V^T + \Delta_U\tilde{V}^T, \tilde{U}\Delta_V^T + \Delta_U\tilde{V}^T) \\ + \frac{\mu}{2}\|\tilde{U}^T\Delta_U + \Delta_U^T\tilde{U} - \tilde{V}^T\Delta_V - \Delta_V^T\tilde{V}\|_F^2 \ge 0 \end{aligned}$$

holds for every $\Delta_U \in \mathbb{R}^{n\times r}$ and $\Delta_V \in \mathbb{R}^{m\times r}$.

B Relation between the Symmetric and Asymmetric Problems

In this section, we study the relationship between problems (4)-(5). This relationship is more general than the topic of this paper, namely the non-existence of spurious second-order critical points and the strict saddle property, and holds for any property that is characterized by the RIP constant δ and the BDP constant κ. Specifically, we show that any property that holds for the symmetric problems (4) with (δ, κ) also holds for the regularized asymmetric problem (5) with another pair of constants $(\tilde{\delta}, \tilde{\kappa})$ decided by δ, κ, and vice versa.

We first consider the transformation from the asymmetric case to the symmetric case. The transformation to the symmetric case has been established in Ge et al. (2017) for linear problem. Here, we show that the transformation can be revised and extended to the nonlinear measurements case.

Theorem 12. *Suppose that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2r,2s}$ and the κ-BDP$_{2t}$ properties. If we choose $\mu := (1-\delta)/2$, then problem* (5) *is equivalent to a symmetric problem whose objective function satisfies the $2\delta/(1+\delta)$-RIP$_{2r,2s}$ and the $2\kappa/(1+\delta)$-BDP$_{2t}$ properties.*

Proof of Theorem 12. For any matrix $N \in \mathbb{R}^{(n+m)\times(n+m)}$, we divide the matrix into four blocks as

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix},$$

where $N_{11} \in \mathbb{R}^{n\times n}, N_{12} \in \mathbb{R}^{n\times m}, N_{22} \in \mathbb{R}^{m\times m}$. Then, we define a new function

$$\tilde{f}(N) := f_a(N_{12}) + f_a(N_{21}^T).$$

We observe that $\tilde{f}(WW^T) = 2h_a(U,V)$, where

$$W := \begin{bmatrix} U \\ V \end{bmatrix} \in \mathbb{R}^{(n+m)\times r}.$$

For any $K \in \mathbb{R}^{(n+m)\times(n+m)}$, the Hessian of $\tilde{f}(\cdot)$ satisfies

$$[\nabla^2 \tilde{f}(N)](K,K) = [\nabla^2 f_a(N_{12})](K_{12},K_{12}) + [\nabla^2 f_a(N_{21}^T)](K_{21}^T, K_{21}^T). \tag{8}$$

Similarly, we can define

$$\tilde{g}(N) := \|N_{11}\|_F^2 + \|N_{22}\|_F^2 - \|N_{12}\|_F^2 - \|N_{21}\|_F^2.$$

We can also verify that $\tilde{g}(WW^T) = g(U,V)$ and

$$[\nabla^2 \tilde{g}(N)](K,K) = 2\left(\|K_{11}\|_F^2 + \|K_{22}\|_F^2 - \|K_{12}\|_F^2 - \|K_{21}\|_F^2\right). \tag{9}$$

for every $K \in \mathbb{R}^{(n+m)\times(n+m)}$. The minimization problem (5) is then equivalent to

$$\min_{W\in\mathbb{R}^{(n+m)\times r}} F(WW^T) := \tilde{f}(WW^T) + \frac{\mu}{2}\cdot\tilde{g}(WW^T), \tag{10}$$

which is in the symmetric form as problem (4). For every $N, K \in \mathbb{R}^{(n+m)\times(n+m)}$ with $\mathrm{rank}(N) \le 2r$ and $\mathrm{rank}(K) \le 2s$, it results from relations (8) and (9) that

$$\begin{aligned}
&[\nabla^2 F(N)](K,K) \\
&\qquad \ge (1-\delta)\left(\|K_{12}\|_F^2 + \|K_{21}\|_F^2\right) + \mu\left(\|K_{11}\|_F^2 + \|K_{22}\|_F^2 - \|K_{12}\|_F^2 - \|K_{21}\|_F^2\right) \\
&\qquad \ge \min\{1-\delta-\mu, \mu\}\cdot\|K\|_F^2
\end{aligned}$$

and

$$\begin{aligned}
&[\nabla^2 F(N)](K,K) \\
&\qquad \le (1+\delta)\left(\|K_{12}\|_F^2 + \|K_{21}\|_F^2\right) + \mu\left(\|K_{11}\|_F^2 + \|K_{22}\|_F^2 - \|K_{12}\|_F^2 - \|K_{21}\|_F^2\right) \\
&\qquad \le \max\{1+\delta-\mu, \mu\}\cdot\|K\|_F^2.
\end{aligned}$$

Choosing $\mu := (1-\delta)/2$, we obtain

$$\frac{1-\delta}{2}\cdot\|K\|_F^2 \le [\nabla^2 F(N)](K,K) \le \frac{1+3\delta}{2}\cdot\|K\|_F^2.$$

Hence, it follows that the function $2F(\cdot)/(1+\delta)$ satisfies the $2\delta/(1+\delta)$-RIP$_{2r,2s}$ property.

Moreover, for every $N, N', K, L \in \mathbb{R}^{(n+m)\times(n+m)}$ with

$$\operatorname{rank}(N), \operatorname{rank}(N'), \operatorname{rank}(K), \operatorname{rank}(L) \leq 2t,$$

it holds that

$$\begin{aligned}[\nabla^2 \tilde{g}(N)](K,L) &= [\nabla^2 \tilde{g}(N')](K,L) \\ &= 2\left(\langle K_{11}, L_{11}\rangle + \langle K_{22}, L_{22}\rangle - \langle K_{12}, L_{12}\rangle - \langle K_{21}, L_{21}\rangle\right)\end{aligned}$$

and

$$\begin{aligned}&\left|[\nabla^2 F(N) - \nabla^2 F(N')](K,L)\right| \\ &\quad = \left|[\nabla^2 f(N_{12}) - \nabla^2 f(N'_{12})](K_{12}, L_{12}) + [\nabla^2 f(N_{21}^T) - \nabla^2 f((N'_{21})^T)](K_{21}^T, L_{21}^T)\right| \\ &\quad \leq \kappa \|K_{12}\|_F \|L_{12}\|_F + \kappa \|K_{21}\|_F \|L_{21}\|_F \leq \kappa \|K\|_F \|L\|_F,\end{aligned}$$

which implies that the function $\frac{2}{1+\delta} \cdot F(\cdot)$ satisfies the $2\kappa/(1+\delta)$-BDP$_{2r}$ property. Since problem (10) is equivalent to the minimization of $\frac{2}{1+\delta} \cdot F(WW^T)$, it is equivalent to a symmetric problem that satisfies the $2\delta/(1+\delta)$-RIP$_{2r,2s}$ and the $2\kappa/(1+\delta)$-BDP$_{2r}$ properties. □

We can see that both constants δ and κ are approximately doubled in the transformation. As an example, Bhojanapalli et al. (2016b) showed that the symmetric linear problem has no spurious local minima if the δ-RIP$_{2r}$ property is satisfied with $\delta < 1/5$. Using Theorem 12, we know that the asymmetric linear problem has no spurious local minima if the δ-RIP$_{2r}$ property is satisfied with $\delta < 1/9$.

The transformation from a symmetric problem to an asymmetric problem is more straightforward. We can equivalently solve the optimization problem

$$\min_{U,V \in \mathbb{R}^{n\times r}} \; f_s\left[\frac{1}{2}\left(UV^T + VU^T\right)\right] \tag{11}$$

or its regularized version with any parameter $\mu > 0$. It can be easily shown that the above problem has the same RIP and BDP constants as the original symmetric problem. We omit the proof for brevity.

Theorem 13. *Suppose that the function $f_s(\cdot)$ satisfies the δ-RIP$_{4r,2s}$ and the κ-BDP$_{4t}$ properties. For every $\mu > 0$, problem* (4) *is equivalent to an asymmetric problem and its regularized version with the δ-RIP$_{2r,2s}$ and the κ-BDP$_{2t}$ properties.*

Note that the transformation from a symmetric problem to an asymmetric problem will not increase the constants κ and δ but requires stronger RIP and BDP properties. Hence, a direct analysis on the symmetric case may establish the same property under a weaker condition. In addition to problem (11), we can also directly consider the problem $\min_{U,V} f_a(UV^T)$. However, in certain applications, the objective function is only defined for symmetric matrices and we can only use the formulation (11) to construct an asymmetric problem. In more restricted cases when the objective function is only defined for symmetric and positive semi-definite matrices, we can only apply the direct analysis to the symmetric case.

C Proofs for Section 2

C.1 Proof of Theorem 3

Proof of Theorem 3. We denote $f(\cdot) := f_s(\cdot)$ and $f(\cdot) := f_a(\cdot)$ for the symmetric and asymmetric case, respectively. Using the mean value theorem and the δ-RIP$_{2r,2r}$ property, there exists a constant $s \in [0,1]$ such that

$$\begin{aligned}&f(M_{t+1}) - f(M_t) \\ &= \langle \nabla f(M_t), M_{t+1} - M_t\rangle + \frac{1}{2}[\nabla^2 f(M_t + s(M_{t+1} - M_t))](M_{t+1} - M_t, M_{t+1} - M_t)\end{aligned}$$

$$\leq \langle \nabla f(M_t), M_{t+1} - M_t \rangle + \frac{1+\delta}{2}\|M_{t+1} - M_t\|_F^2.$$

We define

$$\phi_t(M) := \langle \nabla f(M_t), M - M_t \rangle + \frac{1+\delta}{2}\|M - M_t\|_F^2 = \frac{1+\delta}{2}\|M - \tilde{M}_{t+1}\|_F^2 + constant,$$

where the last constant term is independent of M. Since the projection is orthogonal, the projected matrix M_{t+1} achieves the minimal value of $\phi_t(M)$ over all matrices on the manifold $\mathcal{M}$. Therefore, we obtain

$$\begin{aligned} f(M_{t+1}) - f(M_t) &\leq \phi_t(M_{t+1}) \leq \phi_t(M^*) \\ &= \langle \nabla f(M_t), M^* - M_t \rangle + \frac{1+\delta}{2}\|M^* - M_t\|_F^2. \end{aligned} \tag{12}$$

On the other hand, we can similarly prove that the δ-RIP$_{2r,2r}$ property ensures

$$\begin{aligned} f(M^*) - f(M_t) &\geq \langle \nabla f(M_t), M^* - M_t \rangle + \frac{1-\delta}{2}\|M^* - M_t\|_F^2, \\ f(M_t) - f(M^*) &\geq \frac{1-\delta}{2}\|M^* - M_t\|_F^2. \end{aligned}$$

Substituting the above two inequalities into (12), it follows that

$$\begin{aligned} f(M_{t+1}) - f(M_t) &\leq f(M^*) - f(M_t) + \delta\|M^* - M_t\|_F^2 \\ &\leq f(M^*) - f(M_t) + \frac{2\delta}{1-\delta}[f(M_t) - f(M^*)]. \end{aligned} \tag{13}$$

Therefore, using the condition that $\delta < 1/3$, we have

$$f(M_{t+1}) - f(M^*) \leq \frac{2\delta}{1-\delta}[f(M_t) - f(M^*)] := \alpha[f(M_t) - f(M^*)],$$

where $\alpha := 2\delta/(1-\delta) < 1$. Combining this single-step bound with the induction method proves the linear convergence of Algorithm 1. □

D Proofs for Section 3

D.1 Proof of Theorem 4

Proof of Theorem 4. We only consider the case when m and n are at least $2r$. In this case, we have $\ell = 2r$. Other cases can be handled similarly. For the notational simplicity, we denote $M^* := M_a^*$ in this proof.

Necessity. We first consider problem (3). Suppose that M^* and $\tilde{M}$ are the optimum and a spurious second-order critical point of problem (3), respectively. It has been proved in Ha et al. (2020) that the spurious second-order critical point $\tilde{M}$ has rank r and is a fixed point of the SVP algorithm with the step size $(1+\delta)^{-1}$. Therefore, the point $\tilde{M}$ should be a minimizer of the projection step of the SVP algorithm. This implies that

$$\|\tilde{M} - [\tilde{M} - (1+\delta)^{-1}\nabla f_a(\tilde{M})]\|_F^2 \leq \|M^* - [\tilde{M} - (1+\delta)^{-1}\nabla f_a(\tilde{M})]\|_F^2,$$

which can be simplified to

$$\langle \nabla f_a(\tilde{M}), \tilde{M} - M^* \rangle \leq \frac{1+\delta}{2}\|\tilde{M} - M^*\|_F^2. \tag{14}$$

Let $\mathcal{U}$ and $\mathcal{V}$ denote the subspaces spanned by the columns and rows of $\tilde{M}$ and M^*, respectively. Namely, we have

$$\mathcal{U} := \{\tilde{M}v_1 + M^*v_2 \mid v_1, v_2 \in \mathbb{R}^m\}, \quad \mathcal{V} := \{\tilde{M}^T u_1 + (M^*)^T u_2 \mid u_1, u_2 \in \mathbb{R}^n\}.$$

Since the ranks of both matrices are bounded by r, the dimensions of $\mathcal{U}$ and $\mathcal{V}$ are bounded by $2r$. Therefore, we can find orthogonal matrices $U \in \mathbb{R}^{n\times 2r}$ and $V \in \mathbb{R}^{m\times 2r}$ such that

$$\mathcal{U} \subset \mathrm{range}(U), \quad \mathcal{V} \subset \mathrm{range}(V)$$

and write $\tilde{M}, M^*$ in the form

$$\tilde{M} = U \begin{bmatrix} \Sigma & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} \end{bmatrix} V^T, \quad M^* = URV^T,$$

where $\Sigma \in \mathbb{R}^{r\times r}$ is a diagonal matrix and $R \in \mathbb{R}^{2r\times 2r}$ has rank at most r. Recalling the first condition in Theorem 11, the column space and the row space of $\nabla f_a(\tilde{M})$ are orthogonal to the column space and the row space of $\tilde{M}$, respectively. Then, the δ-RIP$_{2r,2r}$ property gives

$$\begin{aligned} \exists \alpha \in [1-\delta, 1+\delta] \quad \text{s.t.} \; -\langle \nabla f_a(\tilde{M}), M^* \rangle &= \langle \nabla f_a(\tilde{M}), \tilde{M} - M^* \rangle \\ &= \int_0^1 [\nabla^2 f_a(M^* + s(\tilde{M} - M^*))](\tilde{M} - M^*, \tilde{M} - M^*)\, ds \\ &= \alpha \|\tilde{M} - M^*\|_F^2 > 0. \end{aligned} \tag{15}$$

This means that

$$G := \mathcal{P}_U \nabla f_a(\tilde{M}) \mathcal{P}_V \neq 0,$$

where $\mathcal{P}_U$ and $\mathcal{P}_V$ are the orthogonal projections onto $\mathcal{U}$ and $\mathcal{V}$, respectively. Combining with inequality (14), we obtain $\alpha \leq (1+\delta)/2$. By the definition of G, we have

$$\langle \nabla f_a(\tilde{M}), M^* \rangle = \langle G, M^* \rangle.$$

Since both the column space and the row space of G are orthogonal to $\tilde{M}$, the matrix G has the form

$$G = U \begin{bmatrix} 0_{r\times r} & 0_{r\times r} \\ 0_{r\times r} & -\Lambda \end{bmatrix} V^T, \tag{16}$$

where $\Lambda \in \mathbb{R}^{r\times r}$. We may assume without loss of generality that $\Lambda_{ii} \geq 0$ for all i; otherwise, one can flip the sign of some of the last r columns of U. By another orthogonal transformation, we may assume without loss of generality that Λ is a diagonal matrix. Then, Theorem 9 gives

$$(1+\delta) \min_{1\leq i\leq r} \Sigma_{ii} = (1+\delta)\sigma_r(\tilde{M}) \geq \|\nabla f_a(\tilde{M})\|_2 \geq \|G\|_2 = \max_{1\leq i\leq (\ell - r)} \Lambda_{ii}. \tag{17}$$

In addition, condition (15) is equivalent to

$$\langle \Lambda, R_{r+1:2r, r+1:2r} \rangle = \alpha \|\tilde{M} - M^*\|_F^2 = \alpha \left[\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, R_{1:r,1:r} \rangle + \|R\|_F^2 \right]. \tag{18}$$

By the Taylor expansion, for every $Z \in \mathbb{R}^{n\times m}$, we have

$$\langle \nabla f_a(\tilde{M}), Z \rangle = \int_0^1 [\nabla^2 f_a(M^* + s(\tilde{M} - M^*))](\tilde{M} - M^*, Z)\, ds = (\tilde{M} - M^*) : \mathcal{H} : Z,$$

where the last expression is the tensor multiplication and $\mathcal{H}$ is the tensor such that

$$K : \mathcal{H} : L = \int_0^1 [\nabla^2 f_a(M^* + s(\tilde{M} - M^*))](K, L)\, ds, \quad \forall K, L \in \mathbb{R}^{n\times m}.$$

We define

$$\tilde{G} := G - \alpha(\tilde{M} - M^*).$$

By the definition of α, we know that $\langle \tilde{G}, \tilde{M} - M^* \rangle = 0$. Furthermore, using the definition of $\mathcal{H}$, we obtain

$$\begin{aligned} (\tilde{M} - M^*) : \mathcal{H} : (\tilde{M} - M^*) &= \alpha \|\tilde{M} - M^*\|_F^2, \\ (\tilde{M} - M^*) : \mathcal{H} : \tilde{G} = \tilde{G} : \mathcal{H} : (\tilde{M} - M^*) &= \|\tilde{G}\|_F^2. \end{aligned}$$

Suppose that

$$\tilde{G} : \mathcal{H} : \tilde{G} = \beta \|\tilde{G}\|_F^2$$

for some $\beta \in [1-\delta, 1+\delta]$. We consider matrices of the form

$$K(t) := t(\tilde{M} - M^*) + \tilde{G}, \quad \forall t \in \mathbb{R}.$$

Since $K(t)$ is a linear combination of $\tilde{M} - M^*$ and G, the column space of $K(t)$ is a subspace of $\mathcal{U}$, and thus $K(t)$ has rank at most $2r$ and the δ-RIP$_{2r,2r}$ property implies

$$(1-\delta)\|K(t)\|_F^2 \le K(t) : \mathcal{H} : K(t) \le (1+\delta)\|K(t)\|_F^2. \tag{19}$$

Using the facts that

$$\|K(t)\|_F^2 = \|\tilde{M} - M^*\|_F^2 \cdot t^2 + \|\tilde{G}\|_F^2,$$
$$K(t) : \mathcal{H} : K(t) = \alpha\|\tilde{M} - M^*\|_F^2 \cdot t^2 + 2\|\tilde{G}\|_F^2 \cdot t + \beta\|\tilde{G}\|_F^2,$$

we can write the two inequalities in (19) as quadratic inequalities

$$[\alpha - (1-\delta)]\|\tilde{M} - M^*\|_F^2 \cdot t^2 + 2\|\tilde{G}\|_F^2 \cdot t + [\beta - (1-\delta)]\|\tilde{G}\|_F^2 \ge 0,$$
$$[(1+\delta) - \alpha]\|\tilde{M} - M^*\|_F^2 \cdot t^2 - 2\|\tilde{G}\|_F^2 \cdot t + [(1+\delta) - \beta]\|\tilde{G}\|_F^2 \ge 0. \tag{20}$$

If $\alpha = 1 - \delta$, then we must have $\|\tilde{G}\|_F = 0$ and thus $G = \alpha(\tilde{M} - M^*)$. Equivalently, we have $M^* = \tilde{M} - \alpha^{-1}G$. Since the column and row spaces of $G \neq 0$ are orthogonal to $\tilde{M}$, the rank of M^* is at least $\mathrm{rank}(\tilde{M}) + 1 = r + 1$, which is a contradiction. Since $\alpha \le (1+\delta)/2$, we have $\alpha < 1 + \delta$. Thus, we have proved that

$$1 - \delta < \alpha < 1 + \delta.$$

Checking the condition for quadratic functions to be non-negative, we obtain

$$\|\tilde{G}\|_F^2 \le [\alpha - (1-\delta)][\beta - (1-\delta)] \cdot \|\tilde{M} - M^*\|_F^2,$$
$$\|\tilde{G}\|_F^2 \le [(1+\delta) - \alpha][(1+\delta) - \beta] \cdot \|\tilde{M} - M^*\|_F^2.$$

Since

$$\alpha - (1-\delta) > 0, \quad (1+\delta) - \alpha > 0,$$

the above two inequalities are equivalent to

$$\frac{\|\tilde{G}\|_F^2}{\alpha - (1-\delta)} \le [\beta - (1-\delta)] \cdot \|\tilde{M} - M^*\|_F^2,$$
$$\frac{\|\tilde{G}\|_F^2}{(1+\delta) - \alpha} \le [(1+\delta) - \beta] \cdot \|\tilde{M} - M^*\|_F^2.$$

Summing up the two inequalities and dividing both sides by 2δ gives rise to

$$\frac{\|\tilde{G}\|_F^2}{\delta^2 - (1-\alpha)^2} \le \|\tilde{M} - M^*\|_F^2. \tag{21}$$

We note that the above condition is also sufficient for the inequalities in (20) to hold by choosing $\beta = 2 - \alpha$. Using the relation $\|G\|_F^2 = \|\tilde{G}\|_F^2 + \alpha^2\|\tilde{M} - M^*\|_F^2$, one can write

$$\mathrm{tr}(\Lambda^2) = \|G\|_F^2 \le (2\alpha - 1 + \delta^2)\|\tilde{M} - M^*\|_F^2 = \alpha^{-1}(2\alpha - 1 + \delta^2)\langle \Lambda, R_{r+1:2r,r+1:2r}\rangle. \tag{22}$$

Now, using the fact that $\mathrm{rank}(M^*) \le r$, we can write the matrix R as

$$R = \begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}^T = \begin{bmatrix} AB^T & AD^T \\ CB^T & CD^T \end{bmatrix},$$

where $A, B, C, D \in \mathbb{R}^{r\times r}$. Then, conditions (18) and (22) become

$$\langle \Lambda, CD^T\rangle = \alpha\left[\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, AB^T\rangle + \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2\right] \tag{23}$$

and

$$\mathrm{tr}(\Lambda^2) \le \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot \langle \Lambda, CD^T\rangle. \tag{24}$$

If $\langle \Lambda, CD^T\rangle = 0$, we have

$$\mathrm{tr}(\Sigma^2) - 2\langle \Sigma, AB^T\rangle + \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2 = 0,$$

which implies that

$$AB^T = \Sigma, \quad AD^T = CB^T = CD^T = 0.$$

This contradicts the assumption that $\tilde{M} \neq M^*$. Combining this with conditions (17), (23) and (24), we arrive at the necessity part. For problem (5), Lemma 3 in Ha et al. (2020) ensures that $\tilde{M}$ is still a fixed point of the SVP algorithm. Recalling the necessary conditions in Theorem 11, we know that the same necessary conditions also hold in this case.

Sufficiency. Now, we study the sufficiency part. We first consider problem (3). We choose two orthogonal matrices $U \in \mathbb{R}^{n\times 2r}, V \in \mathbb{R}^{m\times 2r}$ and define

$$\tilde{M} = U \begin{bmatrix} \Sigma & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} \end{bmatrix} V^T, \quad M^* := U \left(\begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}^T \right) V^T, \quad G := U \begin{bmatrix} 0_{r\times r} & 0_{r\times r} \\ 0_{r\times r} & -\Lambda \end{bmatrix} V^T.$$

Since $\langle \Lambda, CD^T \rangle \neq 0$, we have $\tilde{M} \neq M^*$. Then, we know that $\mathrm{rank}(\tilde{M}) \leq r$ and $\mathrm{rank}(M^*) \leq r$. We define

$$\tilde{G} := G - \alpha(\tilde{M} - M^*),$$

which satisfies $\langle \tilde{G}, \tilde{M} - M^* \rangle = 0$ by the condition in the second line of (6). If $\tilde{G} = 0$, then

$$\begin{bmatrix} 0_{r\times r} & 0_{r\times r} \\ 0_{r\times r} & -\Lambda \end{bmatrix} = \alpha \cdot \begin{bmatrix} \Sigma & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} \end{bmatrix} - \alpha \cdot \begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}^T = \alpha \cdot \begin{bmatrix} \Sigma & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} \end{bmatrix} - \alpha \cdot \begin{bmatrix} AB^T & 0 \\ 0 & CD^T \end{bmatrix},$$

where the second step is because of $CB^T = 0$ and $AD^T = 0$. The above relation is equivalent to

$$\Sigma = AB^T, \quad \Lambda = \alpha \cdot CD^T.$$

Since $\Sigma \succ 0$, the matrix AB^T has rank r. Noticing that the decomposition of matrix M^* ensures that the rank of M^* is at most r, we have $CD^T = 0$, which is a contradiction to the condition that $\langle CD^T, \Lambda \rangle \neq 0$. Therefore, we have $\tilde{G} \neq 0$. We consider the rank-2 symmetric tensor

$$\begin{aligned} \mathcal{G}_1 :=& \frac{\alpha}{\|\tilde{M} - M^*\|_F^2} \cdot (\tilde{M} - M^*) \otimes (\tilde{M} - M^*) + \frac{2-\alpha}{\|\tilde{G}\|_F^2} \cdot \tilde{G} \otimes \tilde{G} \\ &+ \frac{1}{\|\tilde{M} - M^*\|_F^2} \left[(\tilde{M} - M^*) \otimes \tilde{G} + \tilde{G} \otimes (\tilde{M} - M^*) \right]. \end{aligned}$$

For every matrix $K \in \mathbb{R}^{n\times m}$, we have the decomposition

$$K = t(\tilde{M} - M^*) + s\tilde{G} + \tilde{K}, \quad \langle \tilde{M} - M^*, \tilde{K} \rangle = \langle \tilde{G}, \tilde{K} \rangle = 0,$$

where $t, s \in \mathbb{R}$ are two suitable constants. Then, using the definition of $\mathcal{G}_1$, we have

$$K : \mathcal{G}_1 : K = \alpha \|\tilde{M} - M^*\|_F^2 \cdot t^2 + 2\|\tilde{G}\|_F^2 \cdot ts + (2-\alpha)\|\tilde{G}\|_F^2 \cdot s^2.$$

By the conditions in the third line of (6), one can write

$$\|\tilde{G}\|_F^2 \leq [\alpha - (1-\delta)][(1+\delta) - \alpha] \cdot \|\tilde{M} - M^*\|_F^2,$$

which leads to

$$\begin{aligned} &[\alpha - (1-\delta)]\|\tilde{M} - M^*\|_F^2 \cdot t^2 + 2\|\tilde{G}\|_F^2 \cdot ts + [(1+\delta) - \alpha]\|\tilde{G}\|_F^2 \cdot s^2 \geq 0, \\ &[(1+\delta) - \alpha]\|\tilde{M} - M^*\|_F^2 \cdot t^2 - 2\|\tilde{G}\|_F^2 \cdot ts + [\alpha - (1-\delta)]\|\tilde{G}\|_F^2 \cdot s^2 \geq 0. \end{aligned}$$

The above two inequalities are equivalent to

$$(1-\delta)[\|\tilde{M} - M^*\|_F^2 \cdot s^2 + \|\tilde{G}\|_F^2 \cdot t^2] \leq K : \mathcal{G}_1 : K \leq (1+\delta)[\|\tilde{M} - M^*\|_F^2 \cdot s^2 + \|\tilde{G}\|_F^2 \cdot t^2]. \tag{25}$$

By restricting to the subspace

$$\mathcal{S} := \mathrm{span}\{\tilde{M} - M^*, \tilde{G}\} = \{s(\tilde{M} - M^*) + t\tilde{G} \mid s, t \in \mathbb{R}\},$$

the tensor $\mathcal{G}_1$ can be viewed as a 2×2 matrix. Then, inequality (25) implies that the matrix has two eigenvalues λ_1 and λ_2 such that

$$1 - \delta \leq \lambda_1, \lambda_2 \leq 1 + \delta.$$

Therefore, we can rewrite the tensor $\mathcal{G}_1$ restricted to $\mathcal{S}$ as

$$[\mathcal{G}_1]_{\mathcal{S}} = \lambda_1 \cdot G_1 \otimes G_1 + \lambda_2 \cdot G_2 \otimes G_2,$$

where G_1, G_2 are linear combinations of $\tilde{M} - M^*, \tilde{G}$ and have the unit norm. Since the orthogonal complementary $\mathcal{S}^\perp$ is in the null space of $\mathcal{G}_1$, we have

$$\mathcal{G}_1 = [\mathcal{G}_1]_{\mathcal{S}} = \lambda_1 \cdot G_1 \otimes G_1 + \lambda_2 \cdot G_2 \otimes G_2.$$

Now, we choose matrices $G_3, \ldots, G_N$ such that $G_1, \ldots, G_N$ form an orthonormal basis of the linear vector space $\mathbb{R}^{n\times m}$, where $N := nm$. We define another symmetric tensor by

$$\mathcal{H} := \mathcal{G}_1 + \sum_{i=3}^{N} (1+\delta) \cdot G_i \otimes G_i.$$

Then, inequality (25) implies that the quadratic form $K : \mathcal{H} : K$ satisfies the δ-RIP$_{2r,2r}$ property.

Therefore, we can choose the Hessian to be the constant tensor $\mathcal{H}$ and define the function $f_a(\cdot)$ as

$$f_a(K) := \frac{1}{2}(K - M^*) : \mathcal{H} : (K - M^*), \quad \forall K \in \mathbb{R}^{n\times m}.$$

Combining with the definition of $\mathcal{H}$, we know

$$\nabla f_a(\tilde{M}) = \mathcal{H} : (\tilde{M} - M^*) = G, \quad \nabla^2 f_a(\tilde{M}) = \mathcal{H}.$$

We choose matrices $\bar{U} \in \mathbb{R}^{n\times r}, \bar{V} \in \mathbb{R}^{m\times r}$ such that $\tilde{M} = \bar{U}\bar{V}^T$ and $\bar{U}^T\bar{U} = \bar{V}^T\bar{V}$. By the definitions of $\tilde{M}$ and G, we know that $\tilde{M}$ and G have orthogonal column and row spaces, i.e.,

$$\bar{U}^T G = 0, \quad G\bar{V} = 0.$$

This means that the first-order optimality conditions are satisfied at the point $(\bar{U}, \bar{V})$. For the second-order necessary optimality conditions, we consider the direction

$$\Delta := \begin{bmatrix} \Delta_U \\ \Delta_V \end{bmatrix} \in \mathbb{R}^{(n+m)\times r}.$$

We consider the decomposition

$$\Delta_U = \mathcal{P}_{\bar{U}}\Delta_U + \mathcal{P}_{\bar{U}}^{\perp}\Delta_U := \Delta_U^1 + \Delta_U^2, \quad \Delta_V = \mathcal{P}_{\bar{V}}\Delta_V + \mathcal{P}_{\bar{V}}^{\perp}\Delta_V := \Delta_V^1 + \Delta_V^2,$$

where $\mathcal{P}_{\bar{U}}, \mathcal{P}_{\bar{V}}$ are the orthogonal projection onto the column space of $\bar{U}, \bar{V}$, respectively. Then, using the conditions in the first line of (6), we have

$$\begin{aligned} \langle \nabla f_a(\tilde{M}), \Delta_U\Delta_V^T \rangle &= \langle G, \Delta_U\Delta_V^T \rangle = \langle G, \Delta_U^2(\Delta_V^2)^T \rangle \geq -\|G^T\Delta_U^2\|_F \|\Delta_V^2\|_F \\ &\geq -(1+\delta)\sigma_r(\tilde{M})\|\Delta_U^2\|_F\|\Delta_V^2\|_F \geq -(1+\delta)\sigma_r(\tilde{M}) \cdot \frac{\|\Delta_U^2\|_F^2 + \|\Delta_V^2\|_F^2}{2}. \end{aligned} \tag{26}$$

We define

$$\Delta_1 := \bar{U}(\Delta_V^1)^T + \Delta_U^1\bar{V}^T, \quad \Delta_2 := \bar{U}(\Delta_V^2)^T + \Delta_U^2\bar{V}^T.$$

Then, we know that $\langle \Delta_1, \Delta_2 \rangle = 0$. Using the assumption that $CB^T = AD^T = 0$, we know that M^* has the form

$$M^* = U \begin{bmatrix} AB^T & 0 \\ 0 & CD^T \end{bmatrix} V^T = \mathcal{P}_{\bar{U}}M^*\mathcal{P}_{\bar{V}} + \mathcal{P}_{\bar{U}}^{\perp}M^*\mathcal{P}_{\bar{V}}^{\perp}. \tag{27}$$

Then, the special form (27) implies that

$$\langle M^*, \Delta_2 \rangle = \langle M^*, \bar{U}(\Delta_V^2)^T + \Delta_U^2\bar{V}^T \rangle = \left\langle M^*, \bar{U}\Delta_V^T\mathcal{P}_{\bar{V}}^{\perp} + \mathcal{P}_{\bar{U}}^{\perp}\Delta_U\bar{V}^T \right\rangle = 0.$$

Using the definitions of $\tilde{M}$ and G, it can be concluded that

$$\langle \tilde{M}, \Delta_2 \rangle = 0, \quad \langle G, \Delta_2 \rangle = \langle G, \bar{U}(\Delta_V^2)^T + \Delta_U^2\bar{V}^T \rangle = 0.$$

Since G_1, G_2 are linear combinations of $\tilde{M} - M^*$ and G, the last three relations lead to

$$\langle G_1, \Delta_2 \rangle = \langle G_2, \Delta_2 \rangle = 0.$$

Therefore, there exist constants $a_3, \ldots, a_N$ such that

$$\Delta_2 = \sum_{i=3}^{N} a_i G_i.$$

Suppose that the constants $b_1, \ldots, b_N$ satisfy

$$\Delta_1 = \sum_{i=1}^{N} b_i G_i.$$

Then, the fact $\langle \Delta_1, \Delta_2 \rangle = 0$ and the orthogonality of $G_1, \ldots, G_N$ imply that

$$\sum_{i=3}^{N} a_i b_i = 0.$$

We can calculate that

$$\begin{aligned}
&[\nabla^2 f_a(\tilde{M})](\bar{U}\Delta_V^T + \Delta_U \bar{V}^T, \bar{U}\Delta_V^T + \Delta_U \bar{V}^T) = (\Delta_1 + \Delta_2) : \mathcal{H} : (\Delta_1 + \Delta_2) \\
=&\lambda_1 \cdot b_1^2 + \lambda_2 \cdot b_2^2 + (1+\delta)\sum_{i=3}^{N}(a_i + b_i)^2 \geq (1+\delta)\sum_{i=3}^{N}(a_i + b_i)^2 \\
=&(1+\delta)\sum_{i=3}^{N}\left(a_i^2 + b_i^2\right) \geq (1+\delta)\sum_{i=3}^{N} a_i^2 = (1+\delta)\|\bar{U}(\Delta_V^2)^T + \Delta_U^2 \bar{V}^T\|_F^2,
\end{aligned}$$

where the third last step is due to $\sum_{i=3}^{N} a_i b_i = 0$. Noticing that $\langle \bar{U}(\Delta_V^2)^T, \Delta_U^2 \bar{V}^T \rangle = 0$, the above inequality gives that

$$\begin{aligned}
&[\nabla^2 f_a(\tilde{M})](\bar{U}\Delta_V^T + \Delta_U \bar{V}^T, \bar{U}\Delta_V^T + \Delta_U \bar{V}^T) \geq (1+\delta)\|\bar{U}(\Delta_V^2)^T\|_F^2 + (1+\delta)\|\Delta_U^2 \bar{V}^T\|_F^2 \\
\geq&(1+\delta)\sigma_r(\bar{U})^2\|\Delta_V^2\|_F^2 + (1+\delta)\sigma_r(\bar{V})^2\|\Delta_U^2\|_F^2 = (1+\delta)\sigma_r(\tilde{M})(\|\Delta_V^2\|_F^2 + \|\Delta_U^2\|_F^2),
\end{aligned}$$

where the last equality is because of $\sigma_r(\bar{U})^2 = \sigma_r(\bar{V})^2 = \sigma_r(\tilde{M})$ when $\bar{U}^T\bar{U} = \bar{V}^T\bar{V}$. Combining with inequality (26), one can write

$$\begin{aligned}
&[\nabla^2 h_a(U,V)](\Delta,\Delta) = 2\langle \nabla f_a(\tilde{M}), \Delta_U \Delta_V^T \rangle + [\nabla^2 f_a(\tilde{M})](\bar{U}\Delta_V^T + \Delta_U \bar{V}^T, \bar{U}\Delta_V^T + \Delta_U \bar{V}^T) \\
\geq&-(1+\delta)\sigma_r(\tilde{M})(\|\Delta_V^2\|_F^2 + \|\Delta_U^2\|_F^2) + (1+\delta)\sigma_r(\tilde{M})(\|\Delta_V^2\|_F^2 + \|\Delta_U^2\|_F^2) = 0.
\end{aligned}$$

This shows that $(\bar{U}, \bar{V})$ satisfies the second-order necessary optimality conditions, and therefore it is a spurious second-order critical point.

Now, we consider problem (5). Since the point $(\bar{U}, \bar{V})$ satisfies $\bar{U}^T\bar{U} = \bar{V}^T\bar{V}$, it is also a local minimum of the regularization term. Hence, the point $(\bar{U}, \bar{V})$ is also a spurious second-order critical point of the regularized problem (5). □

D.2 Proof of Corollary 1

Proof of Corollary 1. We assume that problem (3) has a spurious second-order critical point. By the necessity part of Theorem (6), there exist $\alpha \in (1-\delta, 1+\delta)$ and real numbers $\sigma, \lambda, a, b, c, d$ such that

$$\begin{aligned}
&(1+\delta)\sigma \geq \lambda > 0, \quad \alpha^{-1}(2\alpha - 1 + \delta^2)cd \cdot \lambda \geq \lambda^2 > 0, \\
&cd \cdot \lambda = \alpha[\sigma^2 - 2ab \cdot \sigma + (ab)^2 + (ad)^2 + (cb)^2 + (cd)^2].
\end{aligned} \tag{28}$$

We first relax the second line to

$$cd \cdot \lambda \geq \alpha[\sigma^2 - 2|ab| \cdot \sigma + (ab)^2 + 2|ab| \cdot |cd| + (cd)^2]. \tag{29}$$

Then, we denote $x := |ab|$ and consider the quadratic programming problem

$$\min_{x \geq 0} \; x^2 + 2(|cd| - \sigma) \cdot x,$$

whose optimal value is

$$-(\sigma - |cd|)_+^2,$$

where $(t)_+ := \max\{t, 0\}$. Substituting into inequality (29), we obtain

$$cd \cdot \lambda \geq \alpha[\sigma^2 - (\sigma - |cd|)_+^2 + (cd)^2]. \tag{30}$$

Then, we consider two different cases.

Case I. We first consider the case when $\sigma \geq |cd|$. In this case, the inequality (30) becomes

$$cd \cdot \lambda \geq 2\alpha \cdot \sigma |cd| = 2\alpha \cdot \sigma cd,$$

where the last equality is due to $cd > 0$. Therefore,

$$\lambda \geq 2\alpha \cdot \sigma.$$

The second inequality in (28) implies $\lambda \leq \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot cd$. Combining with the above inequality and the assumption of this case, it follows that

$$\alpha^{-1}(2\alpha - 1 + \delta^2) \cdot \sigma \geq \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot cd \geq 2\alpha \cdot \sigma,$$

which is further equivalent to

$$\alpha^{-1}(2\alpha - 1 + \delta^2) \geq 2\alpha \iff \delta^2 \geq 2\alpha^2 - 2\alpha + 1.$$

Since $2\alpha^2 - 2\alpha + 1 \geq 1/2$, we arrive at $\delta^2 \geq 1/2$, which is a contradiction to $\delta < 1/2$.

Case II. We then consider the case when $\sigma \leq |cd|$. In this case, the inequality (30) becomes

$$cd \cdot \lambda \geq \alpha[\sigma^2 + (cd)^2].$$

Combining with the second inequality in (28), we obtain $\lambda \leq \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot (cd)$. Therefore,

$$\alpha^{-1}(2\alpha - 1 + \delta^2) \cdot (cd)^2 \geq cd \cdot \lambda \geq \alpha[\sigma^2 + (cd)^2].$$

Moreover, the first inequality in (28) gives

$$(1 + \delta)\sigma \cdot cd \geq cd \cdot \lambda \geq \alpha[\sigma^2 + (cd)^2].$$

By denoting $y := cd$, the above two inequalities become

$$\begin{aligned} \alpha^{-1}(2\alpha - 1 + \delta^2) \cdot y^2 &\geq \alpha[\sigma^2 + y^2], \\ (1 + \delta)\sigma \cdot y &\geq \alpha[\sigma^2 + y^2]. \end{aligned} \tag{31}$$

By denoting $z := y/\sigma$, the first inequality in (31) implies

$$z^2 \geq \frac{\alpha^2}{\delta^2 - (1 - \alpha)^2}. \tag{32}$$

Since $\delta < 1/2$, one can write

$$(1 - \alpha)^2 + \alpha^2 \geq \frac{1}{2} > \frac{1}{4} > \delta^2,$$

which is equivalent to $\alpha^2 \geq \delta^2 - (1 - \alpha)^2$. Therefore, inequality (32) implies that $z^2 \geq 1$ and

$$z^2 + \frac{1}{z^2} \geq \frac{\alpha^2}{\delta^2 - (1 - \alpha)^2} + \frac{\delta^2 - (1 - \alpha)^2}{\alpha^2}. \tag{33}$$

On the other hand, the second inequality in (31) implies

$$z + \frac{1}{z} \leq \frac{1 + \delta}{\alpha} \quad \text{and thus} \quad z^2 + \frac{1}{z^2} + 2 \leq \frac{(1 + \delta)^2}{\alpha^2}.$$

Combining with inequality (33), it follows that

$$\frac{\alpha^2}{\delta^2 - (1 - \alpha)^2} + \frac{\delta^2 - (1 - \alpha)^2}{\alpha^2} + 2 \leq \frac{(1 + \delta)^2}{\alpha^2}. \tag{34}$$

By some calculation, the above inequality is equivalent to

$$(\delta^2 + 2\delta + 5) \cdot \alpha^2 + (2\delta^2 - 4\delta - 6) \cdot \alpha + 2(1 + \delta)(1 - \delta^2) \leq 0.$$

Checking the discriminant of the above quadratic function, we obtain

$$(2\delta^2 - 4\delta - 6)^2 - 8(\delta^2 + 2\delta + 5)(1 + \delta)(1 - \delta^2) \geq 0,$$

which is equivalent to

$$4(2\delta - 1)(\delta + 1)^4 \geq 0.$$

However, the above claim contradicts the assumption that $\delta < 1/2$.

In summary, the contradictions in the two cases imply that the condition (28) cannot hold, and therefore there does not exist spurious second-order critical points. □

D.3 Counterexample for the Rank-one Case

Example 3. Let $e_i \in \mathbb{R}^n$ be the i-th standard basis of $\mathbb{R}^n$. We define the tensor

$$\begin{aligned}\mathcal{H} := &\sum_{i,j=1}^{n} (e_i e_j^T) \otimes (e_i e_j^T) + \frac{1}{2}(e_1 e_1^T) \otimes (e_2 e_2^T) + \frac{1}{2}(e_2 e_2^T) \otimes (e_1 e_1^T) \\ &+ \frac{1}{4}\left[(e_1 e_2^T) \otimes (e_1 e_2^T) + (e_2 e_1^T) \otimes (e_2 e_1^T)\right] + \frac{1}{4}(e_1 e_2^T) \otimes (e_2 e_1^T) + \frac{1}{4}(e_2 e_1^T) \otimes (e_1 e_2^T)\end{aligned}$$

and the objective function

$$f_a(M) := (M - e_1 e_1^T) : \mathcal{H} : (M - e_1 e_1^T) \quad \forall M \in \mathbb{R}^{n \times n}.$$

The global minimizer of $f_a(\cdot)$ is the rank-1 matrix $M^* := e_1 e_1^T$. It has been proved in Zhang et al. (2019) that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2,2}$ property with $\delta = 1/2$. Moreover, we define

$$U := \frac{1}{\sqrt{2}} e_2, \quad V := U, \quad \tilde{M} := UU^T \neq M^*.$$

It has been proved in Zhang et al. (2019) that the first-order optimality condition is satisfied. To verify the second-order necessary condition, we can calculate that

$$\begin{aligned}[\nabla^2 h_a(U,U)](\Delta,\Delta) &= 2\langle \nabla f_a(\tilde{M}), \Delta_U \Delta_V^T \rangle + (U\Delta_V^T + \Delta_U U^T) : \mathcal{H} : (U\Delta_V^T + \Delta_U U^T) \\ &= -\frac{3}{2}(\Delta_U)_1(\Delta_V)_1 + \frac{5}{8}\left[(\Delta_U)_1^2 + (\Delta_V)_1^2\right] + \frac{1}{4}(\Delta_U)_1(\Delta_V)_1 \\ &\quad + \frac{1}{2}\left[(\Delta_U)_2 + (\Delta_V)_2\right]^2 + \frac{1}{2}\sum_{i=3}^{n}\left[(\Delta_U)_i^2 + (\Delta_V)_i^2\right] \\ &= \frac{5}{8}\left[(\Delta_U)_1 - (\Delta_V)_1\right]^2 + \frac{1}{2}\left[(\Delta_U)_2 + (\Delta_V)_2\right]^2 + \frac{1}{2}\sum_{i=3}^{n}\left[(\Delta_U)_i^2 + (\Delta_V)_i^2\right],\end{aligned}$$

which is non-negative for every $\Delta \in \mathbb{R}^n$. Hence, we conclude that the point $\tilde{M}$ is a spurious second-order critical point of problem (3). Moreover, since we choose $V = U$, the point $\tilde{M}$ is a global minimizer of the regularizer $\|U^T U - V^T V\|_F^2$ and thus $\tilde{M}$ is also a spurious second-order critical point of problem (5).

D.4 Proof of Corollary 2

Proof of Corollary 2. We first consider the case when $\delta \leq 1/3$. We assume that there exists a spurious second-order critical point $\tilde{M}$. Then, by Theorem 4, we know that there exists a constant $\alpha \in (1-\delta, (1+\delta)/2]$. This means that

$$1 - \delta < \frac{1+\delta}{2},$$

which contradicts the assumption that $\delta \leq 1/3$.

Then, we consider the case when $\delta < 1/2$. With no loss of generality, assume that $\tilde{M} \neq M^*$ and $M^* \neq 0$; otherwise, the inequality in this theorem is trivially true. Define

$$m_{11} := \|\Sigma\|_F^2, \quad m_{12} := \langle \Sigma, AB^T \rangle, \quad m_{22} := \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2.$$

By our construction in Theorem 4, we know that

$$m_{11} = \|\tilde{M}\|_F^2, \quad m_{12} = \langle \tilde{M}, M^* \rangle, \quad m_{22} = \|M^*\|_F^2.$$

Therefore, we only need to prove $m_{12} \geq C(\delta) \cdot \sqrt{m_{11} m_{22}}$ for some constant $C(\delta) > 0$. By the analysis in Ha et al. (2020), we know that the second-order critical point $\tilde{M}$ must have rank r and thus $m_{11} \neq 0$. The remainder of the proof is split into two steps.

Step I. First, we prove that

$$\frac{(m_{11}+m_{22}-2m_{12})^2}{m_{11}m_{22}-m_{12}^2} \le \frac{(1+\delta)^2}{\alpha^2}, \quad \frac{(m_{11}-m_{12})^2}{m_{11}m_{22}-m_{12}^2} \le \frac{\delta^2-(1-\alpha)^2}{\alpha^2}. \tag{35}$$

We first rule out the case when $m_{11}m_{22}-m_{12}^2=0$. In this case, the equality condition of the Cauchy inequality shows that there exists a constant t such that

$$\tilde{M} = tM^*.$$

Since $\tilde{M} \neq 0$, the constant t is not 0. Using the mean value theorem, for any $Z \in \mathbb{R}^{n\times m}$, there exists a constant $c \in [0,1]$ such that

$$\begin{aligned}\langle \nabla f_a(\tilde{M}), Z\rangle &= \nabla^2 f[M^* + c(\tilde{M}-M^*)](\tilde{M}-M^*, Z)\\ &= \nabla^2 f[M^* + c(\tilde{M}-M^*)][(t-1)M^*, Z].\end{aligned}$$

The δ-RIP$_{2r,2r}$ property gives

$$\langle \nabla f_a(\tilde{M}), \tilde{M}\rangle = \nabla^2 f[M^* + c(\tilde{M}-M^*)][(t-1)M^*, tM^*] \ge t(t-1)(1-\delta)\|M^*\|_F^2.$$

If $t=1$, we conclude that $\tilde{M}=M^*$, which contradicts the assumption that $\tilde{M}\neq M^*$. Therefore, it holds that

$$\langle \tilde{M}, \nabla f_a(\tilde{M})\rangle \neq 0.$$

This contradicts the first-order optimality condition, which states that $\langle \tilde{M}, \nabla f_a(\tilde{M})\rangle = 0$. Hence, we have proved that inequality (35) is well defined. We consider the decomposition

$$\begin{bmatrix}0 & 0\\ 0 & \Lambda\end{bmatrix} = c_1\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix} + c_2\begin{bmatrix}A\\ C\end{bmatrix}\begin{bmatrix}B\\ D\end{bmatrix}^T + K, \quad \left\langle K, \begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}\right\rangle = \left\langle K, \begin{bmatrix}A\\ C\end{bmatrix}\begin{bmatrix}B\\ D\end{bmatrix}^T\right\rangle = 0.$$

Using the conditions in Theorem 4, it follows that

$$\left\langle \begin{bmatrix}0 & 0\\ 0 & \Lambda\end{bmatrix}, \begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}\right\rangle = 0, \quad \left\langle \begin{bmatrix}0 & 0\\ 0 & \Lambda\end{bmatrix}, \begin{bmatrix}A\\ C\end{bmatrix}\begin{bmatrix}B\\ D\end{bmatrix}^T\right\rangle = \alpha(m_{11}-2m_{12}+m_{22}).$$

The pair of coefficients (c_1, c_2) can be uniquely solved as

$$c_1 = -\alpha\cdot\frac{m_{11}+m_{22}-2m_{12}}{m_{11}m_{22}-m_{12}^2}\cdot m_{12}, \quad c_2 = \alpha\cdot\frac{m_{11}+m_{22}-2m_{12}}{m_{11}m_{22}-m_{12}^2}\cdot m_{11}.$$

Using the orthogonality of the decomposition, we have

$$\begin{aligned}\|\Lambda\|_F^2 &\ge \left\| c_1\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix} + c_2\begin{bmatrix}A\\ C\end{bmatrix}\begin{bmatrix}B\\ D\end{bmatrix}^T\right\|_F^2 = c_1^2 m_{11} + 2c_1c_2m_{12} + c_2^2 m_{22}\\ &= \alpha^2\cdot\frac{m_{11}(m_{11}+m_{22}-2m_{12})^2}{m_{11}m_{22}-m_{12}^2}.\end{aligned} \tag{36}$$

Using the last two lines of condition (6), one can write

$$\begin{aligned}&\alpha^2\cdot\frac{m_{11}(m_{11}+m_{22}-2m_{12})^2}{m_{11}m_{22}-m_{12}^2} \le \|\Lambda\|_F^2\\ &\le (2\alpha-1+\delta^2)\left[\mathrm{tr}(\Sigma^2) - 2\langle\Sigma, AB^T\rangle + \|AB^T\|_F^2 + \|AD^T\|_F^2 + \|CB^T\|_F^2 + \|CD^T\|_F^2\right]\\ &= (2\alpha-1+\delta^2)(m_{11}-2m_{12}+m_{22}).\end{aligned}$$

Simplifying the above inequality, we arrive at the second inequality in (35). Now, the first inequality in condition (6) implies that

$$\|\Lambda\|_F^2 \le (1+\delta)^2\|\Sigma\|_F^2 = (1+\delta)^2 m_{11}.$$

Substituting inequality (36) into the left-hand side, it follows that

$$\alpha^2\cdot\frac{m_{11}(m_{11}+m_{22}-2m_{12})^2}{m_{11}m_{22}-m_{12}^2} \le (1+\delta)^2 m_{11},$$

which is equivalent to the first inequality in (35).

Step II. Next, we prove the existence of $C(\delta)$. We denote

$$\kappa := \frac{m_{12}}{\sqrt{m_{11}m_{22}}} \in (-1, 1).$$

and

$$C_1 := \frac{\delta^2 - (1-\alpha)^2}{\alpha^2}, \quad C_2 := \frac{(1+\delta)^2}{\alpha^2}, \quad t := \sqrt{\frac{m_{11}}{m_{22}}}.$$

Since $\tilde{M} \neq 0$, we have $t > 0$. The inequalities in (35) can be written as

$$(t-\kappa)^2 \leq (1-\kappa^2)C_1, \quad (t+1/t-2\kappa)^2 \leq (1-\kappa^2)C_2. \tag{37}$$

Using the assumption that $\delta < 1/2$, we can write

$$\delta^2 < \frac{1}{4} < (1-\alpha)^2 + \frac{1}{2}\alpha^2,$$

which leads to

$$C_1 = \frac{\delta^2 - (1-\alpha)^2}{\alpha^2} < \frac{1}{2}.$$

If $\kappa + \sqrt{(1-\kappa^2)C_1} \geq 1$, then

$$|\kappa| \geq \frac{1-C_1}{1+C_1} \geq \frac{1}{3} > 0. \tag{38}$$

If $\kappa < 0$, then it holds that

$$\kappa + \sqrt{(1-\kappa^2)C_1} \leq -\frac{1}{3} + \sqrt{\frac{1}{2}} < 1,$$

which contradicts the assumption. Therefore, we have $\kappa \geq 0$ and inequality (38) gives $\kappa \geq 1/3$.

Now, we assume that $\kappa + \sqrt{(1-\kappa^2)C_1} \leq 1$. Then, the first inequality in (37) gives

$$0 < t \leq \kappa + \sqrt{(1-\kappa^2)C_1} \leq 1,$$

which further leads to

$$t + \frac{1}{t} - 2\kappa \geq -\kappa + \sqrt{(1-\kappa^2)C_1} + \frac{1}{\kappa + \sqrt{(1-\kappa^2)C_1}}.$$

Combining with the second inequality in (37), we obtain

$$-\kappa + \sqrt{(1-\kappa^2)C_1} + \frac{1}{\kappa + \sqrt{(1-\kappa^2)C_1}} \leq \sqrt{(1-\kappa^2)C_2}.$$

The above inequality can be simplified to

$$\sqrt{1-\kappa^2}(1 + C_1 - \sqrt{C_1C_2}) \leq \kappa\sqrt{C_2}.$$

We notice that the inequality $1 + C_1 - \sqrt{C_1C_2} \leq 0$ is equivalent to inequality (34), which cannot hold when $\delta < 1/2$. Therefore, we have $1 + C_1 - \sqrt{C_1C_2} > 0$ and $\kappa > 0$. Then, the above inequality is equivalent to

$$(1-\kappa^2)(1 + C_1 - \sqrt{C_1C_2})^2 \leq \kappa^2 \cdot C_2.$$

Therefore, we have

$$\kappa^2 \geq \frac{(1 + C_1 - \sqrt{C_1C_2})^2}{(1 + C_1 - \sqrt{C_1C_2})^2 + C_2} = 1 - \frac{1}{1+\eta^2},$$

where we define

$$\eta := \frac{1 + C_1 - \sqrt{C_1C_2}}{\sqrt{C_2}}.$$

To prove the existence of $C(\delta)$ such that $\kappa \geq C(\delta) > 0$, we only need to show that η is lower bounded by a positive constant. With δ fixed, η can be viewed as a continuous function of α. Since $\eta = (1-\delta)/(1+\delta) > 0$ when $\alpha = 1 - \delta$, the function/parameter η is defined for all α in the compact set $[1-\delta, (1+\delta)/2]$. Combining with the fact that $1 + C_1 - \sqrt{C_1C_2} > 0$, the function η is positive on a compact set, and thus there exists a positive lower bound $\bar{C}(\delta) > 0$.

In summary, we can define the function

$$C(\delta) := \min\left\{\frac{1}{3}, \bar{C}(\delta)\right\} > 0$$

such that $\kappa \geq C(\delta)$ for every spurious second-order critical point $\tilde{M}$. □

D.5 Counterexample for the General Rank Case with Linear Measurements

Example 4. Using the previous rank-1 example, we design a counterexample with linear measurement for the rank-r case. Let $n \geq 2r$ be an integer and $e_i \in \mathbb{R}^n$ be the i-th standard basis of $\mathbb{R}^n$. We define the tensor

$$\begin{aligned}\mathcal{H} :=& \frac{3}{2}\sum_{i,j=1}^{n}(e_ie_j^T)\otimes(e_ie_j^T)+\sum_{i=1}^{r}\Big\{-\frac{1}{2}\left[(e_{2i-1}e_{2i-1}^T)\otimes(e_{2i-1}e_{2i-1}^T)+(e_{2i}e_{2i}^T)\otimes(e_{2i}e_{2i}^T)\right]\\ &+\frac{1}{2}\left[(e_{2i-1}e_{2i-1}^T)\otimes(e_{2i}e_{2i}^T)+(e_{2i}e_{2i}^T)\otimes(e_{2i-1}e_{2i-1}^T)\right]\\ &-\frac{1}{4}\left[(e_{2i-1}e_{2i}^T)\otimes(e_{2i-1}e_{2i}^T)+(e_{2i}e_{2i-1}^T)\otimes(e_{2i}e_{2i-1}^T)\right]\\ &+\frac{1}{4}\left[(e_{2i-1}e_{2i}^T)\otimes(e_{2i}e_{2i-1}^T)+(e_{2i}e_{2i-1}^T)\otimes(e_{2i-1}e_{2i}^T)\right]\Big\}\end{aligned}$$

and the rank-r global minimum

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

The objective function is defined as

$$f_a(M) := (M-M^*):\mathcal{H}:(M-M^*) \quad \forall M \in \mathbb{R}^{n\times n}.$$

We can similarly prove that the function $f_a(\cdot)$ satisfies the δ-RIP$_{2r,2r}$ property with $\delta = 1/2$. Moreover, we define

$$\tilde{U} := \frac{1}{\sqrt{2}}[e_2 \quad e_4 \quad \cdots \quad e_{2r}], \quad \tilde{M} := \tilde{U}\tilde{U}^T = \frac{1}{2}\sum_{i=1}^{r} e_{2i}e_{2i}^T \neq M^*.$$

The gradient of $f_a(\cdot)$ at point $\tilde{M}$ is

$$\nabla f_a(\tilde{M}) = -\frac{3}{4}\sum_{i=1}^{r} e_{2i-1}e_{2i-1}^T \in \mathbb{R}^{2r\times 2r}.$$

Since the column and row spaces of the gradient are orthogonal to those of $\tilde{M}$, the first-order optimality condition is satisfied. To verify the second-order necessary condition, we can similarly calculate that

$$\begin{aligned}&[\nabla^2 h_a(\tilde{U},\tilde{U})](\Delta,\Delta)\\ &\quad=2\langle\nabla f_a(\tilde{M}),\Delta_U\Delta_V^T\rangle+(\tilde{U}\Delta_V^T+\Delta_V\tilde{U}^T):\mathcal{H}:(\tilde{U}\Delta_V^T+\Delta_U\tilde{U}^T)\\ &\quad=-\frac{3}{2}\sum_{i=1}^{r}\left[\sum_{j=1}^{r}(\Delta_U)_{2i-1,j}\right]\left[\sum_{j=1}^{r}(\Delta_V)_{2i-1,j}\right]+\sum_{i=1}^{r}\Big\{\frac{5}{8}\left[(\Delta_U)_{2i-1,i}^2+(\Delta_V)_{2i-1,i}^2\right]\\ &\qquad+\frac{1}{4}(\Delta_U)_{2i-1,i}(\Delta_V)_{2i-1,i}+\frac{1}{2}\left[(\Delta_U)_{2i,i}+(\Delta_V)_{2i,i}\right]^2\Big\}\\ &\qquad+\sum_{1\leq i,j\leq n,i\neq j}\frac{3}{4}\left[(\Delta_U)_{2j,i}+(\Delta_V)_{2i,j}\right]^2+\sum_{1\leq i,j\leq n,i\neq j}\frac{3}{4}\left[(\Delta_U)_{2j-1,i}^2+(\Delta_V)_{2j-1,i}^2\right]\\ &\quad=\sum_{i=1}^{r}\Big\{\frac{5}{8}\left[(\Delta_U)_{2i-1,i}-(\Delta_V)_{2i-1,i}\right]^2+\frac{1}{2}\left[(\Delta_U)_{2i,i}+(\Delta_V)_{2i,i}\right]^2\Big\}\\ &\qquad+\sum_{1\leq i,j\leq n,i\neq j}\frac{3}{4}\left[(\Delta_U)_{2j,i}+(\Delta_V)_{2i,j}\right]^2+\sum_{1\leq i,j\leq n,i\neq j}\frac{3}{4}\left[(\Delta_U)_{2j-1,i}-(\Delta_V)_{2j-1,i}\right]^2,\end{aligned}$$

which is non-negative for every $\Delta \in \mathbb{R}^{n\times r}$. Hence, the point $\tilde{M}$ is a spurious second-order critical point of problem (3). Moreover, since we choose $\tilde{V} = \tilde{U}$, the point $\tilde{M}$ is a global minimizer of the regularizer $\|\tilde{U}^T\tilde{U} - \tilde{V}^T\tilde{U}\|_F^2$ and thus $\tilde{M}$ is also a spurious second-order critical point of problem (5).

E Proofs for Section 4

E.1 Proof of Theorem 6

In this subsection, we use the following notations:

$$M := UV^T,\ M^* := U^*(V^*)^T, \quad W := \begin{bmatrix} U \\ V \end{bmatrix},\ W^* := \begin{bmatrix} U^* \\ V^* \end{bmatrix}, \quad \hat{W} := \begin{bmatrix} U \\ -V \end{bmatrix},\ \hat{W}^* := \begin{bmatrix} U^* \\ -V^* \end{bmatrix},$$

where $M^* := M_a^*$ is the global optimum. We always assume that U^* and V^* satisfy $(U^*)^T U^* = (V^*)^T V^*$. When there is no ambiguity about W, we use W^* to denote the minimizer of $\min_{X \in \mathcal{X}^*} \|W - X\|_F$, where $\mathcal{X}^*$ is the set of global minima of problem (5). We note that the set $\mathcal{X}^*$ is the trajectory of a global minimum (U^*, V^*) under the orthogonal group:

$$\mathcal{X}^* = \{(U^*R, V^*R) \mid R \in \mathbb{R}^{r \times r}, R^T R = RR^T = I_r\}.$$

Therefore, the set $\mathcal{X}^*$ is a compact set and its minimum can be attained. With this choice, it holds that

$$\operatorname{dist}(W, \mathcal{X}^*) = \|W - W^*\|_F.$$

We first summarize some technical results in the following lemma.

Lemma 1 (Tu et al. (2016); Zhu et al. (2018))**.** *The following statements hold for every $U \in \mathbb{R}^{n \times r}$, $V \in \mathbb{R}^{m \times r}$ and $W \in \mathbb{R}^{(n+m) \times r}$:*

- $4\|M - M^*\|_F^2 \geq \|WW^T - W^*(W^*)^T\|_F^2 - \|U^TU - V^TV\|_F^2$.
- $\|W^*(W^*)^T\|_F^2 = 4\|M^*\|_F^2$.
- *If $\operatorname{rank}(W^*) = r$ and W^* is the minimizer of $\min_{X \in \mathcal{X}^*} \|W - X\|_F$, then $\|WW^T - W^*(W^*)^T\|_F^2 \geq 2(\sqrt{2} - 1)\sigma_r^2(W^*)\|W - W^*\|_F^2$.*
- *If $\operatorname{rank}(U^*) = r$ and U^* is the minimizer of $\min_{X \in \mathcal{X}^*} \|U - X\|_F$, then $\|UU^T - U^*(U^*)^T\|_F^2 \geq 2(\sqrt{2} - 1)\sigma_r^2(U^*)\|U - U^*\|_F^2$.*

The proof of Theorem 6 follows from the following sequence of lemmas. We first identify two cases when the gradient is large. The following lemma proves that an unbalanced solution cannot be a first-order critical point.

Lemma 2. *Given a constant $\epsilon > 0$, if*

$$\|U^TU - V^TV\|_F \geq \epsilon,$$

then

$$\|\nabla \rho(U, V)\|_F \geq \mu(\epsilon/r)^{3/2}.$$

Proof. Using the relationship between the 2-norm and the Frobenius norm, we have

$$\|U^TU - V^TV\|_2 \geq r^{-1}\|U^TU - V^TV\|_F \geq \epsilon/r.$$

Let $q \in \mathbb{R}^r$ be an eigenvector of $U^TU - V^TV$ such that

$$\|q\|_2 = 1, \quad \left|q^T(U^TU - V^TV)q\right| = \|U^TU - V^TV\|_2.$$

We consider the direction

$$\Delta := \hat{W}qq^T.$$

Then, we can calculate that

$$\|\Delta\|_F^2 = \operatorname{tr}\left(\hat{W}qq^Tqq^T\hat{W}^T\right) = \operatorname{tr}\left(q^T\hat{W}^T\hat{W}q\right) = q^T(U^TU + V^TV)q.$$

In addition, we have

$$\begin{aligned}\langle \nabla h_a(U, V), \Delta \rangle &= \left\langle \begin{bmatrix} \nabla f_a(M)V \\ [\nabla f_a(M)]^T U \end{bmatrix}, \begin{bmatrix} Uqq^T \\ -Vqq^T \end{bmatrix} \right\rangle \\ &= \operatorname{tr}\left[V^T[\nabla f_a(M)]^T Uqq^T\right] - \operatorname{tr}\left[U^T \nabla f_a(M) Vqq^T\right]\end{aligned}$$

$$= q^T \left[V^T[\nabla f_a(M)]^T U\right] q - q^T \left[U^T \nabla f_a(M) V\right] q = 0.$$

and

$$\begin{aligned}
\left|\left\langle \frac{\mu}{4}\nabla g(U,V), \Delta \right\rangle\right| &= \mu \left|\left\langle \hat{W}\hat{W}^T W, W q q^T \right\rangle\right| \\
&= \mu \left|\mathrm{tr}\left[(U^TU - V^TV)(U^TU + V^TV) q q^T\right]\right| \\
&= \mu \left|q^T(U^TU - V^TV)(U^TU + V^TV) q\right| \\
&= \mu \|U^TU - V^TV\|_2 \cdot q^T(U^TU + V^TV) q \\
&= \mu \|U^TU - V^TV\|_2 \cdot \sqrt{q^T(U^TU + V^TV) q} \cdot \|\Delta\|_F.
\end{aligned}$$

Hence, Cauchy's inequality implies that

$$\|\nabla \rho(U,V)\|_F \geq \frac{|\langle \nabla \rho(U,V), \Delta\rangle|}{\|\Delta\|_F} = \mu \|U^TU - V^TV\|_2 \cdot \sqrt{q^T(U^TU + V^TV) q}.$$

Using the fact that

$$q^T(U^TU + V^TV) q \geq \left|q^T(U^TU - V^TV) q\right| = \|U^TU - V^TV\|_2,$$

we obtain

$$\|\nabla \rho(U,V)\|_F \geq \mu \|U^TU - V^TV\|_2^{3/2} \geq \mu(\epsilon/r)^{3/2}.$$

□

The next lemma proves that a solution with large norm cannot be a first-order critical point.

Lemma 3. *Given a constant $\epsilon > 0$, if*

$$\frac{1-\delta}{3} \leq \mu < 1 - \delta, \quad \|WW^T\|_F^{3/2} \geq \max\left\{ \left(\frac{1+\delta}{1-\mu-\delta}\right)^2 \|W^*(W^*)^T\|_F^{3/2}, \frac{4\sqrt{r}\lambda}{1-\mu-\delta} \right\},$$

then

$$\|\nabla \rho(U,V)\|_F \geq \lambda.$$

Proof. Choosing the direction $\Delta := W$, we can calculate that

$$\langle \nabla \rho(U,V), \Delta\rangle = 2\langle \nabla f_a(UV^T), UV^T\rangle + \mu \|U^TU - V^TV\|_F^2. \tag{39}$$

Using the δ-RIP$_{2r,2r}$ property, we have

$$[\nabla^2 f_a(N)](M,M) \geq (1-\delta)\|M\|_F^2, \quad [\nabla^2 f_a(N)](M^*, M) \leq (1+\delta)\|M\|_F \|M^*\|_F,$$

where $N \in \mathbb{R}^{n\times m}$ is every matrix with rank at most $2r$. Then, the first term can be estimated as

$$\begin{aligned}
\langle \nabla f_a(UV^T), UV^T\rangle &= \int_0^1 [\nabla^2 f_a(M^* + s(M - M^*))[M - M^*, M]\, ds \\
&\geq (1-\delta)\|M\|_F^2 - (1+\delta)\|M^*\|_F \|M\|_F.
\end{aligned}$$

The second term is

$$\mu \|U^TU - V^TV\|_F^2 = \mu\left(\|UU^T\|_F^2 + \|VV^T\|_F^2\right) - 2\mu\|M\|_F^2.$$

Substituting into equation (39), it follows that

$$\begin{aligned}
\langle \nabla \rho(U,V), \Delta\rangle &\geq \mu\left(\|UU^T\|_F^2 + \|VV^T\|_F^2\right) + 2(1-\delta-\mu)\|M\|_F^2 - 2(1+\delta)\|M^*\|_F\|M\|_F \\
&\geq \mu\left(\|UU^T\|_F^2 + \|VV^T\|_F^2\right) + 2(1-\delta-\mu)\|M\|_F^2 - 2c\|M\|_F^2 - \frac{(1+\delta)^2}{2c}\|M^*\|_F^2 \\
&\geq \min\left\{\mu, 1-\delta-\mu-c\right\}\|WW^T\|_F^2 - \frac{(1+\delta)^2}{2c}\|M^*\|_F^2,
\end{aligned}$$

where $c \in (0, 1-\delta-\mu)$ is a constant to be designed later. Using equality that $(U^*)^TU^* = (V^*)^TV^*$, Lemma 1 gives

$$\|W^*(W^*)^T\|_F^2 = 4\|M^*\|_F^2.$$

As a result,

$$\langle \nabla \rho(U,V), \Delta \rangle \geq \min\{\mu, 1-\delta-\mu-c\} \|WW^T\|_F^2 - \frac{(1+\delta)^2}{8c}\|W^*(W^*)^T\|_F^2.$$

Now, choosing

$$c = \frac{1-\delta-\mu}{2}$$

and noticing that $\mu \geq (1-\delta-\mu)/2$, it yields that

$$\langle \nabla \rho(U,V), \Delta \rangle \geq \frac{1-\delta-\mu}{2}\|WW^T\|_F^2 - \frac{(1+\delta)^2}{4(1-\delta-\mu)}\|W^*(W^*)^T\|_F^2. \tag{40}$$

On the other hand,

$$\|\Delta\|_F = \|W\|_F \leq \sqrt{r}\|WW^T\|_F^{1/2}.$$

Combining with inequality (40) and using the assumption of this lemma, one can write

$$\begin{aligned}
\|\nabla \rho(U,V)\|_F &\geq \frac{\langle \nabla \rho(U,V), \Delta \rangle}{\|\Delta\|_F} \\
&\geq \frac{1-\delta-\mu}{2\sqrt{r}}\|WW^T\|_F^{3/2} - \frac{(1+\delta)^2}{4\sqrt{r}(1-\delta-\mu)}\|W^*(W^*)^T\|_F^2\|WW^T\|_F^{-1/2} \\
&\geq \frac{1-\delta-\mu}{2\sqrt{r}}\|WW^T\|_F^{3/2} - \frac{(1+\delta)^2}{4\sqrt{r}(1-\delta-\mu)}\|W^*(W^*)^T\|_F^{3/2} \\
&\geq \frac{1-\delta-\mu}{4\sqrt{r}}\|WW^T\|_F^{3/2} \geq \lambda.
\end{aligned}$$

□

Using the above two lemmas, we only need to focus on points such that

$$\|U^TU - V^TV\|_F = o(1), \quad \|WW^T\|_F = O(1).$$

The following lemma proves that if (U,V) is an approximate first-order critical point with a small singular value $\sigma_r(W)$, then the Hessian of the objective function at this point has a negative curvature.

Lemma 4. *Consider positive constants $\alpha, C, \epsilon, \lambda$ such that*

$$\epsilon^2 \leq (\sqrt{2}-1)\sigma_r^2(W^*)\cdot\alpha^2, \quad G > \mu\left(\epsilon + \frac{4H^2}{G^2}\right) + \frac{(1+\delta)H^2}{G^2}, \tag{41}$$

where $G := \|\nabla f_a(M)\|_2$ and $H := \lambda + \mu\epsilon C$. If

$$\|U^TU - V^TV\|_F^2 \leq \epsilon^2, \quad \|WW^T\|_F \leq C^2, \quad \|W - W^*\|_F \geq \alpha, \quad \|\nabla\rho(U,V)\|_F \leq \lambda$$

and

$$\sigma_r^2(W) \leq \frac{2}{1+\delta}\left[G - \mu\left(\epsilon + \frac{4H^2}{G^2}\right) - \frac{(1+\delta)H^2}{G^2}\right] - 2\tau \tag{42}$$

for some positive constant τ, then it holds that

$$\lambda_{min}(\nabla^2\rho(U,V)) \leq -(1+\delta)\tau.$$

Proof. We choose a singular vector q of W such that

$$\|q\|_2 = 1, \quad \|Wq\|_2 = \sigma_r(W).$$

Since $\|Wq\|_2 = \sqrt{\|Uq\|_2^2 + \|Vq\|_2^2}$, we have

$$\|Uq\|_2^2 + \|Vq\|_2^2 = \sigma_r^2(W).$$

We choose singular vectors u and v such that

$$\|u\|_2 = \|v\|_2 = 1, \quad \|\nabla f_a(M)\|_2 = u^T\nabla f_a(M)v.$$

We define the direction as

$$\Delta_U := -uq^T, \quad \Delta_V := vq^T, \quad \Delta := \begin{bmatrix}\Delta_U \\ \Delta_V\end{bmatrix}, \quad \hat{\Delta} := \begin{bmatrix}\Delta_U \\ -\Delta_V\end{bmatrix}.$$

For the Hessian of $h_a(\cdot,\cdot)$, we can calculate that

$$\langle \nabla f_a(M), \Delta_U \Delta_V^T \rangle = -\|\nabla f_a(M)\|_2 = -G \tag{43}$$

and the δ-RIP$_{2r,2r}$ property gives

$$\begin{aligned}
[\nabla^2 f_a(M)](\Delta_U V^T + U\Delta_V^T, \Delta_U V^T + U\Delta_V^T) & \\
&\le (1+\delta)\|\Delta_U V^T + U\Delta_V^T\|_F^2 = (1+\delta)\| - u(Vq)^T + (Uq)v^T\|_F^2 \\
&= (1+\delta)\left(\|Vq\|_F^2 + \|Uq\|_F^2\right) - 2(1+\delta)[q^T(U^T u)]\cdot[q^T(V^T v)] \\
&\le (1+\delta)\sigma_r^2(W) + 2(1+\delta)\cdot\|U^T u\|_F\|V^T v\|_F.
\end{aligned} \tag{44}$$

Then, we consider the terms coming from the Hessian of the regularizer. First, we have

$$\begin{aligned}
\langle \hat{\Delta}\hat{W}^T, \Delta W^T\rangle &\le \|U^T U - V^T V\|_F \cdot \|\Delta_U^T\Delta_U - \Delta_V^T\Delta_V\|_F \\
&\le \epsilon\cdot\left[\|\Delta_U^T\Delta_U\|_F + \|\Delta_V^T\Delta_V\|_F\right] = 2\epsilon.
\end{aligned} \tag{45}$$

Next, we can estimate that

$$\begin{aligned}
\langle \hat{W}\hat{\Delta}^T, \Delta W^T\rangle + \langle \hat{W}\hat{W}^T, \Delta\Delta^T\rangle &= \frac{1}{2}\|U^T\Delta_U + \Delta_U^T U - V^T\Delta_V - \Delta_V^T V\|_F^2 \\
&\le 4\left(\|U^T\Delta_U\|_F^2 + \|V^T\Delta_V\|_F^2\right) \\
&= 4\left(\|(U^T u)q^T\|_F^2 + \|(V^T v)q^T\|_F^2\right) \\
&= 4\left(\|U^T u\|_F^2 + \|V^T v\|_F^2\right).
\end{aligned} \tag{46}$$

Using the assumption that $\|WW^T\|_F \le C^2$ and $\|U^TU - V^TV\|_F^2 \le \epsilon^2$, one can write

$$\|\hat{W}\hat{W}^T W\|_F^2 \le \|U^TU - V^TV\|_F^2 \cdot \|U^TU + V^TV\|_F \le \epsilon^2\|WW^T\|_F \le \epsilon^2 C^2$$

and

$$\left\|\begin{bmatrix}\nabla f_a(UV^T)V \\ \nabla f_a(UV^T)^T U\end{bmatrix}\right\|_F = \|\nabla\rho(U,V) - \mu\hat{W}\hat{W}^T W\|_F \le \lambda + \mu\epsilon C = H. \tag{47}$$

The second relation implies that

$$\|\nabla f_a(UV^T)V\|_2 \le \|\nabla f_a(UV^T)V\|_F \le H, \quad \|U^T\nabla f_a(UV^T)\|_2 \le \|U^T\nabla f_a(UV^T)\|_F \le H. \tag{48}$$

By the definition of u and v, it holds that

$$\|v\|_2 = 1, \quad \|\nabla f_a(M)\|_2 u = \nabla f_a(M)v.$$

Therefore,

$$\|U^T u\|_F^2 = \frac{\|U^T\nabla f_a(M)v\|_F^2}{\|\nabla f_a(M)\|_2^2} \le \frac{\|U^T\nabla f_a(M)\|_F^2\|v\|_2^2}{\|\nabla f_a(M)\|_2^2} \le \frac{H^2}{G^2}.$$

Similarly,

$$\|V^T v\|_F^2 \le \frac{H^2}{G^2}.$$

Substituting into (44) and (46) yields that

$$[\nabla^2 f_a(M)](\Delta_U V^T + U\Delta_V^T, \Delta_U V^T + U\Delta_V^T) \le (1+\delta)\sigma_r^2(W) + 2(1+\delta)\cdot\frac{H^2}{G^2} \tag{49}$$

and

$$\langle \hat{W}\hat{\Delta}^T, \Delta W^T\rangle + \langle \hat{W}\hat{W}^T, \Delta\Delta^T\rangle \le 8\cdot\frac{H^2}{G^2}. \tag{50}$$

Combining (43), (45), (49) and (50), it follows that

$$[\nabla^2\rho(U,V)](\Delta,\Delta) \le -2G + (1+\delta)\sigma_r^2(W) + 2\mu\epsilon + [8\mu + 2(1+\delta)] \cdot \frac{H^2}{G^2}.$$

Since $\|\Delta\|_F^2 = 2$, the above relation implies

$$\lambda_{min}(\nabla^2\rho(U,V)) \le -G + \frac{1+\delta}{2}\sigma_r^2(W) + \mu\epsilon + (4\mu + 1 + \delta) \cdot \frac{H^2}{G^2} \le -(1+\delta)\tau.$$

□

Remark 2. The positive constants ϵ and λ in the proof of Lemma 4 can be chosen to be arbitrarily small with α, C fixed. Hence, we may choose small enough ϵ and λ such that the assumptions given in inequality (41) are satisfied. This lemma resolves the case when the minimal singular value $\sigma_r^2(W)$ is on the order of $\|\nabla f_a(M)\|_2/(2+2\delta)$. In the next lemma, we will show that this is the only case when $\delta < 1/3$.

The final step is to prove that condition (42) always holds provided that $\delta < 1/3$ and $\epsilon, \lambda, \tau = o(1)$.

Lemma 5. *Given positive constants $\alpha, C, \epsilon, \lambda$, if*

$$\|U^TU - V^TV\|_F^2 \le \epsilon^2, \quad \max\{\|WW^T\|_F, \|W^*(W^*)^T\|_F\} \le C^2,$$
$$\|W - W^*\|_F \ge \alpha, \quad \|\nabla\rho(U,V)\|_F \le \lambda, \quad \delta < 1/3,$$

then the inequality $G \ge c\alpha$ holds for some constant $c > 0$ independent of $\alpha, \epsilon, \lambda, C$. Furthermore, there exist two positive constants

$$\epsilon_0(\delta, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C), \quad \lambda_0(\delta, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C)$$

such that

$$\sigma_r^2(W) \le \frac{2}{1+\delta}\left[G - \mu\left(2\epsilon + \frac{4H^2}{G^2}\right) - \frac{(1+\delta)H^2}{G^2}\right] \tag{51}$$

whenever

$$0 < \epsilon \le \epsilon_0(\delta, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C),$$
$$0 < \lambda \le \lambda_0(\delta, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C).$$

Here, G and H are defined in Lemma 4.

Proof. We first prove the existence of the constant c. Using Lemma 1, one can write $4\|M - M^*\|_F^2 \ge \|WW^T - W^*(W^*)^T\|_F^2 - \|U^TU - V^TV\|_F^2 \ge \|WW^T - W^*(W^*)^T\|_F^2 - \epsilon^2$. Using Lemma 1 and the assumption that $\|W - W^*\|_F \ge \alpha$, we have

$$\|M - M^*\|_F^2 \ge \frac{\sqrt{2}-1}{2}\sigma_r^2(W^*)\|W - W^*\|_F^2 - \frac{\epsilon^2}{4} \ge \frac{\sqrt{2}-1}{2}\sigma_r^2(W^*) \cdot \alpha^2 - \frac{\epsilon^2}{4}. \tag{52}$$

By the definition of ϵ, it follows that

$$\|M - M^*\|_F^2 \ge \frac{\sqrt{2}-1}{4}\sigma_r^2(W^*) \cdot \alpha^2 > 0.$$

Thus, the δ-RIP$_{2r,2r}$ property gives

$$\|\nabla f_a(M)\|_F \ge \frac{\langle \nabla f_a(M), M - M^*\rangle}{\|M - M^*\|_F} \ge (1-\delta)\|M - M^*\|_F \ge \sqrt{\frac{\sqrt{2}-1}{4}} \cdot \sigma_r(W^*)(1-\delta) \cdot \alpha.$$

Hence, we have

$$G = \|\nabla f_a(M)\|_2 \ge \sqrt{\frac{\sqrt{2}-1}{4r}} \cdot \sigma_r(W^*)(1-\delta) \cdot \alpha = c\alpha,$$

where we define

$$c := \sqrt{\frac{\sqrt{2}-1}{4r}} \cdot \sigma_r(W^*)(1-\delta).$$

Next, we prove inequality (51) by contradiction, i.e., we assume

$$\sigma_r^2(W) > \frac{2}{1+\delta}\left[G - \mu\left(2\epsilon + \frac{4H^2}{G^2}\right) - \frac{(1+\delta)H^2}{G^2}\right] \ge \frac{2c\alpha}{1+\delta} + \text{poly}(\epsilon, \lambda). \tag{53}$$

The remainder of the proof is divided into three steps.

Step I. We first develop a lower bound for $\sigma_r(M)$. We choose a vector $p \in \mathbb{R}^r$ such that

$$\|p\|_F = 1, \quad U^T U p = \sigma_r^2(U) \cdot p.$$

It can be shown that

$$\begin{aligned}\|(Wp)^T W\|_F &= \|p^T U^T U + p^T V^T V\|_F \le 2\|p^T U^T U\|_F + \|p^T (V^T V - U^T U)\|_F \\ &\le 2\sigma_r^2(U) + \|p^T\|_F \|V^T V - U^T U\|_F \le 2\sigma_r^2(U) + \epsilon.\end{aligned}$$

On the other hand, since W has rank r, it holds that

$$\big\|(Wp)^T W\big\|_F \ge \sigma_r^2(W) \cdot \|p\|_F = \sigma_r^2(W).$$

Combining the above two estimates, we arrive at

$$2\sigma_r^2(U) \ge \sigma_r^2(W) - \epsilon > 0,$$

where the last inequality is from the assumption that ϵ, λ are small and $\sigma_r(W)$ is lower bounded by a positive value in (53). Using the inequality that $\sqrt{1-x} \ge 1 - x$ for every $x \in [0, 1]$, the above inequality implies that

$$\sigma_r(U) \ge \frac{1}{\sqrt{2}} \sigma_r(W) \cdot \sqrt{1 - \frac{\epsilon}{\sigma_r^2(W)}} \ge \frac{1}{\sqrt{2}} \sigma_r(W) - \frac{\epsilon}{\sqrt{2}\sigma_r(W)}. \tag{54}$$

Similarly, one can prove that

$$\sigma_r(V) \ge \frac{1}{\sqrt{2}} \sigma_r(W) - \frac{\epsilon}{\sqrt{2}\sigma_r(W)}.$$

When ϵ is small enough, we know that $\sigma_r(U), \sigma_r(V) \neq 0$ and both U, V have rank r. To lower bound the singular value $\sigma_r(M)$, we consider vectors x such that $\|x\|_2 = 1$ and lower bound $x^T V (U^T U) V^T x$. Since the range of $V(U^T U)V^T$ is a subspace of the range of V and the range of V has exactly dimension r, directions x that are in the orthogonal complement of the range of V correspond to exactly $m - r$ zero singular values. Hence, to estimate the r-th largest singular value of M, we only need to consider directions that are in the range of V. Namely, we only consider directions that have the form $x = Vy$ for some vector y. Then, we have

$$\begin{aligned}x^T V (U^T U) V^T x &= y^T (V^T V)(U^T U)(V^T V) y \\ &= y^T (V^T V)^3 y + y^T (V^T V)(U^T U - V^T V)(V^T V) y.\end{aligned}$$

First, we bound the second term by calculating that

$$\begin{aligned}\|V(V^T V - U^T U)V^T\|_2 &\le \|V\|_2^2 \|U^T U - V^T V\|_2 \le \|V^T V\|_F \|U^T U - V^T V\|_F \\ &\le \|W^T W\|_F \|U^T U - V^T V\|_F \le C^2 \epsilon.\end{aligned}$$

This implies that

$$y^T (V^T V)(U^T U - V^T V)(V^T V) y \ge -C^2 \epsilon \cdot \|Vy\|_F^2.$$

Next, we assume that y has the decomposition

$$y = \sum_{i=1}^r c_i v_i,$$

where v_i is an eigenvector of $V^T V$ associated with the eigenvalue $\sigma_i^2(V)$. Then, we can calculate that

$$y^T (V^T V)^3 y = \sum_{i=1}^r c_i^2 \sigma_i^6(V), \quad \|Vy\|_F^2 = \sum_{i=1}^r c_i^2 \sigma_i^2(V) = 1.$$

Combining the above estimates leads to

$$\begin{aligned}x^T V (U^T U) V^T x &\ge \left[\frac{\sum_{i=1}^r c_i^2 \sigma_i^6(V)}{\sum_{i=1}^r c_i^2 \sigma_i^2(V)} - C^2 \epsilon \right] \cdot \|Vy\|_F^2 \\ &= \frac{\sum_{i=1}^r c_i^2 \sigma_i^6(V)}{\sum_{i=1}^r c_i^2 \sigma_i^2(V)} - C^2 \epsilon \ge \sigma_r^4(V) - C^2 \epsilon.\end{aligned}$$

This implies that

$$\begin{aligned}\sigma_r^2(M) &\geq \sigma_r^4(V) - C^2\epsilon \geq \left[\frac{1}{\sqrt{2}}\sigma_r(W) - \frac{\epsilon}{\sqrt{2}\sigma_r(W)}\right]^4 - C^2\epsilon \\ &\geq \frac{1}{4}\sigma_r^4(W) - \sigma_r^2(W)\epsilon - \sigma_r^{-2}(W)\epsilon^3 - C^2\epsilon \\ &\geq \frac{1}{4}\sigma_r^4(W) - \sigma_r^{-2}(W)\epsilon^3 - 2C^2\epsilon \\ &\geq \frac{1}{4}\sigma_r^4(W) - \frac{1+\delta}{G}\cdot\epsilon^3 - 2C^2\epsilon \\ &\geq \frac{1}{4}\sigma_r^4(W) - \frac{1+\delta}{c\alpha}\cdot\epsilon^3 - 2C^2\epsilon. \end{aligned} \tag{55}$$

where the second last inequality is due to (53) and the assumption that ϵ and λ are sufficiently small.

Step II. Next, we derive an upper bound for $\sigma_r(M)$. We define

$$\bar{M} := \mathcal{P}_r\left[M - \frac{1}{1+\delta}\nabla f_a(M)\right],$$

where $\mathcal{P}_r$ is the orthogonal projection onto the low-rank set via SVD. Since $M \neq M^*$ and $\delta < 1/3$, we recall that inequality (13) gives

$$\begin{aligned}-\phi(\bar{M}) &\geq \frac{1-3\delta}{1-\delta}[f_a(M) - f_a(M^*)] \geq \frac{1-3\delta}{2}\|M - M^*\|_F^2 \\ &\geq \frac{1-3\delta}{2}\left[\frac{\sqrt{2}-1}{2}\sigma_r^2(W^*)\alpha^2 - \frac{\epsilon^2}{4}\right] := K,\end{aligned}$$

where the second inequality follows from (52) and

$$-\phi(\bar{M}) = \langle\nabla f_a(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2.$$

Hence,

$$\langle\nabla f_a(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2 \geq K. \tag{56}$$

When we choose ϵ to be small enough, it holds that $K > 0$. For simplicity, we define

$$N := -\frac{1}{1+\delta}\nabla f_a(M).$$

Then, $\bar{M} = \mathcal{P}_r(M+N)$ and the left-hand side of (56) is equal to

$$\begin{aligned}&\langle\nabla f_a(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2 \\ &= (1+\delta)\langle N, \mathcal{P}_r(M+N) - M\rangle - \frac{1+\delta}{2}\|\mathcal{P}_r(M+N) - M\|_F^2 \\ &= \frac{1+\delta}{2}\left[\|N\|_F^2 - \|N + M - \mathcal{P}_r(M+N)\|_F^2\right] \\ &= \frac{1+\delta}{2}\left[\|N\|_F^2 - \|N+M\|_F^2 + \|\mathcal{P}_r(M+N)\|_F^2\right].\end{aligned} \tag{57}$$

Similar to the proof of inequality (48), we can prove that

$$\|NV\|_F \leq \tilde{H} := \frac{H}{1+\delta}, \quad \|U^T N\|_F \leq \tilde{H}.$$

Then, we have

$$-\operatorname{tr}[N^T(UV^T)] \leq \|U^T N\|_F\|V\|_F \leq \tilde{H}\cdot\|W\|_F \leq \tilde{H}\cdot\sqrt{\sqrt{r}\|WW^T\|_F} \leq \sqrt[4]{r}C\cdot\tilde{H}.$$

Using the above relation, we obtain

$$\|N\|_F^2 - \|N+M\|_F^2 = -2\,\mathrm{tr}[N^T(UV^T)] - \|M\|_F^2 \le 2\sqrt[4]{r}C\cdot\tilde{H} - \|M\|_F^2.$$

Suppose that $\mathcal{P}_U$ and $\mathcal{P}_V$ are the orthogonal projections onto the column spaces of U and V, respectively. We define

$$N_1 := \mathcal{P}_U N \mathcal{P}_V,\ N_2 := \mathcal{P}_U N(I-\mathcal{P}_V),\ N_3 := (I-\mathcal{P}_U)N\mathcal{P}_V,\ N_4 := (I-\mathcal{P}_U)N(I-\mathcal{P}_V).$$

Then, recalling the assumption (53) and inequality (54), it follows that

$$\begin{aligned}\|N_1\|_F = \|\mathcal{P}_U N\mathcal{P}_V\|_F &\le \sigma_r^{-1}(U)\|U^T\mathcal{P}_U N\mathcal{P}_V\|_F \le \sigma_r^{-1}(U)\|U^T N\|_F \le \frac{\sqrt{2}\sigma_r(W)}{\sigma_r^2(W)-\epsilon}\cdot\tilde{H}\\ &\le \left[\sqrt{\frac{1+\delta}{G}} + \mathrm{poly}(\epsilon,\lambda)\right]\cdot\tilde{H} \le \left[\sqrt{\frac{1+\delta}{c\alpha}} + \mathrm{poly}(\epsilon,\lambda)\right]\cdot\tilde{H} := \kappa\tilde{H}.\end{aligned}$$

Similarly, we can prove that

$$\|N_1+N_2\|_F = \|\mathcal{P}_U N\|_F \le \kappa\tilde{H}, \quad \|N_1+N_3\|_F = \|N\mathcal{P}_V\|_F \le \kappa\tilde{H},$$

which leads to

$$\|N_2\|_F \le 2\kappa\tilde{H}, \quad \|N_3\|_F \le 2\kappa\tilde{H}.$$

Using Weyl's theorem, the following holds for every $1\le i\le r$:

$$|\sigma_i(M+N) - \sigma_i(M+N_4)| \le \|N_1+N_2+N_3\|_2 \le \|N_1+N_2+N_3\|_F \le 3\kappa\tilde{H}.$$

Therefore, we have

$$\begin{aligned}\|\mathcal{P}_r(M+N)\|_F^2 &= \sum_{i=1}^r \sigma_i^2(M+N)\\ &\ge \sum_{i=1}^r \sigma_i^2(M+N_4) - r\cdot 3\kappa\tilde{H}\cdot(\|M+N\|_2 + \|M+N_4\|_2)\\ &\ge \sum_{i=1}^r \sigma_i^2(M+N_4) - 6r\kappa\tilde{H}\cdot(\|M\|_2 + \|N\|_2)\\ &\ge \sum_{i=1}^r \sigma_i^2(M+N_4) - 6r\kappa\tilde{H}\cdot\left(\|M\|_F + \frac{G}{1+\delta}\right). \end{aligned} \tag{58}$$

Using the assumption (53) and the inequality (55), one can write

$$\frac{G}{1+\delta} \le \frac{\sigma_r^2(W)}{2} + \mathrm{poly}(\sqrt{\epsilon},\lambda) \le \sigma_r(M) + \mathrm{poly}(\sqrt{\epsilon},\lambda) \le \|M\|_F + \mathrm{poly}(\sqrt{\epsilon},\lambda), \tag{59}$$

where $\mathrm{poly}(\sqrt{\epsilon},\lambda)$ means a polynomial of $\sqrt{\epsilon}$ and λ. Therefore, we attain the bound

$$\begin{aligned}\|M\|_F + \|N\|_F \le 2\|M\|_F + \mathrm{poly}(\sqrt{\epsilon},\lambda) &\le 2\cdot\frac{\|WW^T\|_F}{\sqrt{2}} + \mathrm{poly}(\sqrt{\epsilon},\lambda)\\ &\le \sqrt{2}C^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).\end{aligned} \tag{60}$$

Substituting back into the previous estimate (58), it follows that

$$\|\mathcal{P}_r(M+N)\|_F^2 \ge \sum_{i=1}^r \sigma_i^2(M+N_4) - 6\sqrt{2}r\kappa\tilde{H}C^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda) = \sum_{i=1}^r \sigma_i^2(M+N_4) + \mathrm{poly}(\sqrt{\epsilon},\lambda).$$

Now, since M and N_4 have orthogonal column and row spaces, the maximal r singular values of $M+N_4$ are simply the maximal r singular values of the singular values M and N_4, which we assume to be

$$\sigma_i(M),\ i=1,\dots,k \quad \text{and} \quad \sigma_i(N_4),\ i=1,\dots,r-k.$$

Now, it follows from (57) that

$$\frac{2}{1+\delta}\left[\langle\nabla f_a(M), M-\bar{M}\rangle - \frac{1+\delta}{2}\|M-\bar{M}\|_F^2\right]$$

$$
\begin{aligned}
&= \|N\|_F^2 - \|N+M\|_F^2 + \|\mathcal{P}_r(M+N)\|_F^2 \\
&\leq -\sum_{i=1}^{r}\sigma_i^2(M) + \sum_{i=1}^{k}\sigma_i^2(M) + \sum_{i=1}^{r-k}\sigma_i^2(N_4) + \mathrm{poly}(\sqrt{\epsilon},\lambda) + 2\sqrt[4]{r}C\cdot\tilde{H} \\
&= -\sum_{i=k+1}^{r}\sigma_i^2(M) + \sum_{i=1}^{r-k}\sigma_i^2(N_4) + \mathrm{poly}(\sqrt{\epsilon},\lambda) \\
&\leq -(r-k)\sigma_r^2(M) + (r-k)\|N_4\|_2^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda) \\
&\leq -(r-k)\sigma_r^2(M) + (r-k)\|N\|_2^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).
\end{aligned}
$$

If $k = r$, then the above inequality and inequality (56) imply that

$$\mathrm{poly}(\sqrt{\epsilon},\lambda) \geq K = O(\alpha^2),$$

which contradicts the assumption that ϵ and λ are small. Hence, it can be concluded that $r-k \geq 1$. Combining with (56), we obtain the upper bound

$$
\begin{aligned}
\sigma_r^2(M) &\leq -\frac{2}{1+\delta}\cdot\frac{K}{r-k} + \|N\|_2^2 + \frac{1}{r-k}\cdot\mathrm{poly}(\sqrt{\epsilon},\lambda) \\
&= -\frac{2}{1+\delta}\cdot\frac{K}{r} + \|N\|_2^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).
\end{aligned} \tag{61}
$$

Step III. In the last step, we combine the inequalities (55) and (61), which leads to

$$\frac{1}{4}\sigma_r^4(W) - \frac{1+\delta}{c\alpha}\cdot\epsilon^3 - 2C^2\epsilon \leq -\frac{2}{1+\delta}\cdot\frac{K}{r} + \frac{1}{(1+\delta)^2}G^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).$$

This means that

$$\sigma_r^4(W) + \frac{8}{1+\delta}\cdot\frac{K}{r} \leq \frac{4}{(1+\delta)^2}G^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).$$

Since $K > 0$ has lower bounds that are independent of ϵ and λ, we can choose ϵ and λ to be small enough such that

$$\sigma_r^4(W) + \frac{4}{1+\delta}\cdot\frac{K}{r} \leq \frac{4}{(1+\delta)^2}G^2.$$

However, recalling the assumption (53), we have

$$
\begin{aligned}
\sigma_r^4(W) &> \frac{4}{(1+\delta)^2}\left[G - \mu\left(2\epsilon + \frac{4H^2}{G^2}\right) - \frac{(1+\delta)H^2}{G^2}\right]^2 \\
&\geq \frac{4}{(1+\delta)^2}G^2 - \frac{16}{(1+\delta)^2}G\cdot\mu\epsilon + \mathrm{poly}(\sqrt{\epsilon},\lambda) \\
&\geq \frac{4}{(1+\delta)^2}G^2 - \frac{16}{(1+\delta)^2}\mu\epsilon\cdot\frac{1}{\sqrt{2}}(1+\delta)C^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda) \\
&= \frac{4}{(1+\delta)^2}G^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda),
\end{aligned}
$$

where in the third inequality we use inequalities (59)-(60) to conclude that

$$G \leq (1+\delta)\|M\|_F + \mathrm{poly}(\sqrt{\epsilon},\lambda) \leq \frac{1}{\sqrt{2}}(1+\delta)C^2 + \mathrm{poly}(\sqrt{\epsilon},\lambda).$$

The above two inequalities cannot hold simultaneously when λ and ϵ are small enough. This contradiction means that the condition (51) holds by choosing

$$
\begin{aligned}
0 < &\epsilon \leq \epsilon_0(\delta,\mu,\sigma_r(M_a^*),\|M_a^*\|_F,\alpha,C), \\
0 < &\lambda \leq \lambda_0(\delta,\mu,\sigma_r(M_a^*),\|M_a^*\|_F,\alpha,C),
\end{aligned}
$$

for some small enough positive constants

$$\epsilon_0(\delta,\mu,\sigma_r(M_a^*),\|M_a^*\|_F,\alpha,C), \quad \lambda_0(\delta,\mu,\sigma_r(M_a^*),\|M_a^*\|_F,\alpha,C).$$

□

The only thing left is to piecing everything together.

Proof of Theorem 6. We first choose

$$C := \left[\left(\frac{1+\delta}{1-\mu-\delta} \right)^2 \|W^*(W^*)^T\|_F^{3/2} \right]^{1/3}.$$

Then, we select ϵ_1 and λ_1 as

$$\epsilon_1(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha) := \epsilon_0(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C),$$

$$\lambda_1(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha) := \min \left\{ \lambda_0(\delta, r, \mu, \sigma_r(M_a^*), \|M_a^*\|_F, \alpha, C), \right.$$

$$\left. \frac{(1-\mu-\delta)C^3}{4\sqrt{r}} \right\}.$$

Finally, we combine Lemmas 4-5 to get the bounds for the gradient and the Hessian. □

E.2 Proof of Theorem 7

In this subsection, we use similar notations:

$$M := UU^T, \quad M^* := U^*(U^*)^T,$$

where $M^* := M_s^*$ is the global optimum. We also assume that U^* is the minimizer of $\min_{X \in \mathcal{X}^*} \|U - X\|_F$ when there is no ambiguity about U. In this case, the distance is given by

$$\operatorname{dist}(U, \mathcal{X}^*) = \|U - U^*\|_F.$$

The proof of Theorem 7 is similar to that of Theorem 6. We first consider the case when $\|UU^T\|_F$ is large.

Lemma 6. *Given a constant $\epsilon > 0$, if*

$$\|UU^T\|_F^2 \geq \max \left\{ \frac{2(1+\delta)}{1-\delta} \|U^*(U^*)^T\|_F^2, \left(\frac{2\lambda\sqrt{r}}{1-\delta} \right)^{4/3} \right\},$$

then

$$\|\nabla h_s(U)\|_F \geq \lambda.$$

Proof. Choosing the direction $\Delta := U$, we can calculate that

$$\langle \nabla h_s(U), \Delta \rangle = \langle \nabla f_s(UU^T), UU^T \rangle.$$

Using the δ-RIP$_{2r,2r}$ property, we have

$$\begin{aligned} \langle \nabla f_s(UU^T), UU^T \rangle &= \int_0^1 [\nabla^2 f_s(M^* + s(M - M^*)][M - M^*, M] \\ &\geq (1-\delta)\|M\|_F^2 - (1+\delta)\|M^*\|_F\|M\|_F \\ &\geq \frac{1-\delta}{2}\|M\|_F^2. \end{aligned}$$

Moreover,

$$\|\Delta\|_F = \|U\|_F \leq \sqrt{r}\|UU^T\|_F^{1/2}.$$

This leads to

$$\|\nabla h_s(U)\|_F \geq \frac{\langle \nabla h_s(U), \Delta \rangle}{\|\Delta\|_F} = \frac{\langle \nabla f_s(UU^T), UU^T \rangle}{\|U\|_F} \geq \frac{1-\delta}{2\sqrt{r}}\|UU^T\|_F^{3/2} \geq \lambda.$$

□

The next lemma is a counterpart of Lemma 4.

Lemma 7. *Consider positive constants α, C, λ such that*

$$\lambda \leq 2(\sqrt{r}C)^{-1}(\sqrt{2}-1)\sigma_r^2(U^*)\cdot\alpha^2, \quad G > \frac{(1+\delta)\lambda^2}{4G^2},$$

where $G := -\lambda_{min}(\nabla f_s(M))$. If

$$\|UU^T\|_F \leq C^2, \quad \|U-U^*\|_F \geq \alpha, \quad \|\nabla h_s(U)\|_F \leq \lambda,$$

then the inequality $G \geq c\alpha^2$ holds for some constant $c > 0$ independent of α, λ, C. Moreover, if there exists some positive constant τ such that

$$\sigma_r^2(U) \leq \frac{1}{1+\delta}\left[G - \frac{(1+\delta)\lambda^2}{4G^2}\right] - \tau, \tag{62}$$

then

$$\lambda_{min}(\nabla^2 h_s(U)) \leq -2(1+\delta)\tau.$$

Proof. We choose a singular vector q of U such that

$$\|q\|_2 = 1, \quad \|Uq\|_2 = \sigma_r(U).$$

We first prove the existence of the constant c. The δ-RIP$_{2r,2r}$ property gives

$$\langle \nabla f_s(M), M^* - M\rangle \leq -(1-\delta)\|M-M^*\|_F^2.$$

Using the assumption of this lemma, we have

$$\|\nabla f_s(M)U\|_2 \leq \|\nabla f_s(M)U\|_F = \frac{1}{2}\|\nabla h_s(U)\|_F \leq \frac{1}{2}\lambda, \tag{63}$$

which leads to

$$\langle \nabla f_s(M), M\rangle = \langle \nabla f_s(M)U, U\rangle \leq \|\nabla f_s(M)U\|_F\|U\|_F \leq \frac{1}{2}\lambda\cdot\sqrt{r}C.$$

Substituting into (63), it follows that

$$\langle \nabla f_s(M), M^*\rangle \leq -(1-\delta)\|M-M^*\|_F^2 + \frac{1}{2}\lambda\cdot\sqrt{r}C.$$

Using Lemma 1, we have

$$\|M-M^*\|_F^2 \geq 2(\sqrt{2}-1)\sigma_r^2(U^*)\|U-U^*\|_F^2 \geq 2(\sqrt{2}-1)\sigma_r^2(U^*)\cdot\alpha^2.$$

By the condition on λ, it follows that

$$\langle \nabla f_s(M), M^*\rangle \leq -(1-\delta)\|M-M^*\|_F^2 + \frac{1}{2}\lambda\cdot\sqrt{r}C \leq -(\sqrt{2}-1)(1-\delta)\sigma_r^2(U^*)\cdot\alpha^2. \tag{64}$$

The above inequality also indicates that $\lambda_{min}(\nabla f_s(M)) < 0$. Using the relations that

$$\nabla f_s(M) \succeq \lambda_{min}(\nabla f_s(M))\cdot I_n, \quad M^* \succeq 0,$$

we arrive at

$$\langle \nabla f_s(M), M^*\rangle \geq \lambda_{min}(\nabla f_s(M))\,\mathrm{tr}(M^*) \geq \sqrt{r}\|M^*\|_F\cdot\lambda_{min}(\nabla f_s(M)).$$

Combining the last inequality with (64), we obtain

$$\lambda_{min}(\nabla f_s(M)) \leq -(\sqrt{r}\|M^*\|_F)^{-1}(\sqrt{2}-1)(1-\delta)\sigma_r^2(U^*)\cdot\alpha^2 = -c\alpha^2$$

and thus $G \geq c\alpha^2$, where

$$c := (\sqrt{r}\|M^*\|_F)^{-1}(\sqrt{2}-1)(1-\delta)\sigma_r^2(U^*)$$

Next, we prove the upper bound on the minimal eigenvalue. We choose an eigenvector u such that

$$\|u\|_2 = 1, \quad \lambda_{min}(\nabla f_s(M)) = u^T\nabla f_s(M)u.$$

The direction is chosen to be

$$\Delta := uq^T.$$

For the Hessian of $h_s(\cdot,\cdot)$, we can calculate that

$$\langle \nabla f_s(M), \Delta\Delta^T \rangle = \lambda_{min}(\nabla f_s(M)) = -G \tag{65}$$

and the δ-RIP$_{2r,2r}$ property gives

$$\begin{aligned}
[\nabla^2 f_s(M)](\Delta U^T + U\Delta^T, \Delta U^T + U\Delta^T) & \\
&\leq (1+\delta)\|\Delta U^T + U\Delta^T\|_F^2 = (1+\delta)\|u(Uq)^T + (Uq)u^T\|_F^2 \\
&= 2(1+\delta)\|Uq\|_F^2 + 2(1+\delta)[q^T(U^Tu)]^2 \\
&\leq 2(1+\delta)\sigma_r^2(U) + 2(1+\delta)\cdot\|U^Tu\|_F^2.
\end{aligned} \tag{66}$$

By letting the vector $\tilde{v}$ be

$$\|\tilde{v}\|_2 = 1, \quad \lambda_{min}(\nabla f_s(M))u = \nabla f_s(M)\tilde{v},$$

the inequality (63) implies that

$$\|U^Tu\|_F^2 = \frac{\|U^T\nabla f_s(M)\tilde{v}\|_F^2}{\lambda_{min}^2(\nabla f_s(M))} = \frac{\|U^T\nabla f_s(M)\tilde{v}\|_2^2}{\lambda_{min}^2(\nabla f_s(M))} \leq \frac{\|U^T\nabla f_s(M)\|_2^2\|\tilde{v}\|_2^2}{\lambda_{min}^2(\nabla f_s(M))} \leq \frac{\lambda^2}{4G^2}.$$

Substituting into (66), we obtain

$$[\nabla^2 f_s(M)](\Delta U^T + U\Delta^T, \Delta U^T + U\Delta^T) \leq 2(1+\delta)\sigma_r^2(U) + (1+\delta)\cdot\frac{\lambda^2}{2G^2}. \tag{67}$$

Combining (65) and (67), it follows that

$$[\nabla^2 h_s(U)](\Delta,\Delta) \leq -2G + 2(1+\delta)\sigma_r^2(U) + (1+\delta)\cdot\frac{\lambda^2}{2G^2}.$$

Since $\|\Delta\|_F^2 = 1$, the above inequality implies

$$\lambda_{min}(\nabla^2 h_s(U)) \leq -2G + 2(1+\delta)\sigma_r^2(U) + (1+\delta)\cdot\frac{\lambda^2}{2G^2} \leq -(1+\delta)\tau.$$

□

We finally give the counterpart of Lemma 5, which states that the condition (62) always holds when $\delta < 1/3$.

Lemma 8. *Given positive constants $\alpha, C, \epsilon, \lambda$, if*

$$\max\{\|UU^T\|_F, \|U^*(U^*)^T\|_F\} \leq C^2,\ \|U-U^*\|_F \geq \alpha,\ \|\nabla h_s(U)\|_F \leq \lambda,\ \delta < 1/3,$$

then there exists a positive constant $\lambda_0(\delta, W^*, \alpha, C)$ such that

$$\sigma_r^2(U) \leq \frac{1}{1+\delta}\left[G - \frac{(1+\delta)\lambda^2}{4G^2} - \lambda\right] \tag{68}$$

whenever

$$0 < \lambda \leq \lambda_0(\delta, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha, C).$$

Proof. We prove by contradiction, i.e., we assume

$$\sigma_r^2(U) > \frac{1}{1+\delta}\left[G - \frac{(1+\delta)\lambda^2}{4G^2} - \lambda\right] \geq \frac{c\alpha^2}{1+\delta} + \mathrm{poly}(\lambda). \tag{69}$$

To follow the proof of Lemma 5, we also divide the argument into three steps, although the first step is superficial.

Step I. We first give a lower bound for $\lambda_r(M)$. In the symmetric case, this step is straightforward, since we always have

$$\lambda_r^2(M) = \sigma_r^4(U). \tag{70}$$

Step II. Next, we derive an upper bound for $\lambda_r(M)$. We define

$$\bar{M} := \mathcal{P}_r\left[M - \frac{1}{1+\delta}\nabla f_s(M)\right],$$

where $\mathcal{P}_r$ is the orthogonal projection onto the low-rank manifold (we do not drop negative eigenvalues in this proof). Since $M \neq M^*$ and $\delta < 1/3$, we recall that inequality (13) gives

$$\begin{aligned}-\phi(\bar{M}) &\geq \frac{1-3\delta}{1-\delta}[f_s(M) - f_s(M^*)] \geq \frac{1-3\delta}{2}\|M - M^*\|_F^2 \\ &\geq (1-3\delta)\cdot(\sqrt{2}-1)\sigma_r^2(W^*)\alpha^2 := K > 0,\end{aligned}$$

where the second inequality comes from Lemma 1 and

$$-\phi(\bar{M}) = \langle \nabla f_s(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2.$$

Hence,

$$\langle \nabla f_s(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2 \geq K. \tag{71}$$

For simplicity, we define

$$N := -\frac{1}{1+\delta}\nabla f_s(M).$$

Then, $\bar{M} = \mathcal{P}_r(M+N)$ and the left-hand side of (71) is equal to

$$\begin{aligned}&\langle \nabla f_s(M), M - \bar{M}\rangle - \frac{1+\delta}{2}\|M - \bar{M}\|_F^2 \\ &\qquad = (1+\delta)\langle N, \mathcal{P}_r(M+N) - M\rangle - \frac{1+\delta}{2}\|\mathcal{P}_r(M+N) - M\|_F^2 \\ &\qquad = \frac{1+\delta}{2}\left[\|N\|_F^2 - \|N + M - \mathcal{P}_r(M+N)\|_F^2\right] \\ &\qquad = \frac{1+\delta}{2}\left[\|N\|_F^2 - \|N+M\|_F^2 + \|\mathcal{P}_r(M+N)\|_F^2\right].\end{aligned} \tag{72}$$

Similar to the proof of inequality (63), we can prove that

$$\|U^T N\|_F \leq \tilde{H} := \frac{\lambda}{2(1+\delta)}.$$

Then, we have

$$-\operatorname{tr}[N^T(UU^T)] \leq \|U^T N\|_F\|U\|_F \leq \tilde{H}\cdot\|U\|_F \leq \tilde{H}\cdot\sqrt{\sqrt{r}\|UU^T\|_F} \leq \sqrt[4]{r}C\cdot\tilde{H}.$$

Using the above relation, one can write

$$\|N\|_F^2 - \|N+M\|_F^2 = -2\operatorname{tr}[N^T(UU^T)] - \|M\|_F^2 \leq 2\sqrt[4]{r}C\cdot\tilde{H} - \|M\|_F^2.$$

Suppose that $\mathcal{P}_U$ is the orthogonal projections onto the column space of U. We define

$$N_1 := \mathcal{P}_U N\mathcal{P}_U,\ N_2 := \mathcal{P}_U N(I-\mathcal{P}_U),\ N_3 := (I-\mathcal{P}_U)N\mathcal{P}_U,\ N_4 := (I-\mathcal{P}_U)N(I-\mathcal{P}_U).$$

Then, it follows from (69) that

$$\begin{aligned}\|N_1\|_F = \|\mathcal{P}_U N\mathcal{P}_U\|_F &\leq \sigma_r^{-1}(U)\|U^T\mathcal{P}_U N\mathcal{P}_U\|_F \leq \sigma_r^{-1}(U)\|U^T N\|_F \leq \sigma_r^{-1}(U)\cdot\tilde{H} \\ &\leq \left[\sqrt{\frac{1+\delta}{G}} + \operatorname{poly}(\lambda)\right]\cdot\tilde{H} \leq \left[\sqrt{\frac{1+\delta}{c\alpha^2}} + \operatorname{poly}(\lambda)\right]\cdot\tilde{H} := \kappa\tilde{H}.\end{aligned}$$

Similarly, we can prove that

$$\|N_1 + N_2\|_F = \|\mathcal{P}_U N\|_F \leq \kappa\tilde{H}, \quad \|N_1 + N_3\|_F = \|N\mathcal{P}_V\|_F \leq \kappa\tilde{H},$$

which leads to

$$\|N_2\|_F \leq 2\kappa\tilde{H}, \quad \|N_3\|_F \leq 2\kappa\tilde{H}.$$

Using Weyl's theorem, the following holds for every $1 \leq i \leq r$:

$$|\lambda_i(M+N) - \lambda_i(M+N_4)| \leq \|N_1+N_2+N_3\|_2 \leq \|N_1+N_2+N_3\|_F \leq 3\kappa\tilde{H}.$$

Therefore, we have

$$\begin{aligned}
\|\mathcal{P}_r(M+N)\|_F^2 &= \sum_{i=1}^{r} \lambda_i^2(M+N) \\
&\geq \sum_{i=1}^{r} \lambda_i^2(M+N_4) - r\cdot 3\kappa\tilde{H}\cdot(\|M+N\|_2 + \|M+N_4\|_2) \\
&\geq \sum_{i=1}^{r} \lambda_i^2(M+N_4) - 6r\kappa\tilde{H}\cdot(\|M\|_2 + \|N\|_2) \\
&\geq \sum_{i=1}^{r} \lambda_i^2(M+N_4) - 6r\kappa\tilde{H}\cdot\left(\|M\|_F + \frac{G}{1+\delta}\right). \qquad (73)
\end{aligned}$$

Similar to the asymmetric case, we can prove that

$$\frac{G}{1+\delta} \leq \|M\|_F + \mathrm{poly}(\lambda).$$

holds under the assumption (69). Therefore, we obtain the bound

$$\|M\|_F + \|N\|_F \leq 2\|M\|_F + \mathrm{poly}(\lambda) \leq 2C^2 + \mathrm{poly}(\lambda).$$

Substituting back into the previous estimate (73), it follows that

$$\|\mathcal{P}_r(M+N)\|_F^2 \geq \sum_{i=1}^{r} \lambda_i^2(M+N_4) + \mathrm{poly}(\lambda).$$

Now, since M and N_4 have orthogonal column and row spaces, the maximal r eigenvalues of $M+N_4$ are simply the maximal r eigenvalues of the eigenvalues of M and N_4, which we assume to be

$$\lambda_i(M),\ i=1,\ldots,k \quad \text{and} \quad \lambda_i(N_4),\ i=1,\ldots,r-k.$$

Now, it follows from (72) that

$$\begin{aligned}
&\frac{2}{1+\delta}\left[\langle\nabla f_s(M), M-\bar{M}\rangle - \frac{1+\delta}{2}\|M-\bar{M}\|_F^2\right] \\
&\qquad = \|N\|_F^2 - \|N+M\|_F^2 + \|\mathcal{P}_r(M+N)\|_F^2 \\
&\qquad \leq -\sum_{i=1}^{r}\lambda_i^2(M) + \sum_{i=1}^{k}\lambda_i^2(M) + \sum_{i=1}^{r-k}\lambda_i^2(N_4) + \mathrm{poly}(\lambda) + 2\sqrt[4]{r}C\cdot\tilde{H} \\
&\qquad = -\sum_{i=k+1}^{r}\lambda_i^2(M) + \sum_{i=1}^{r-k}\lambda_i^2(N_4) + \mathrm{poly}(\lambda). \qquad (74)
\end{aligned}$$

Using the assumption (69) and the fact that λ is small, we know that $\lambda_i(N_4) > 0$ for all $i \in \{1,\ldots,k\}$. Therefore,

$$-\sum_{i=k+1}^{r}\lambda_i^2(M) + \sum_{i=1}^{r-k}\lambda_i^2(N_4) \leq -(r-k)\lambda_r^2(M) + (r-k)\lambda_{max}(N_4)^2.$$

Substituting into (74) gives rise to

$$\begin{aligned}
&\frac{2}{1+\delta}\left[\langle\nabla f_s(M), M-\bar{M}\rangle - \frac{1+\delta}{2}\|M-\bar{M}\|_F^2\right] \\
&\qquad \leq -(r-k)\lambda_r^2(M) + (r-k)\lambda_{max}(N_4)^2 + \mathrm{poly}(\lambda) \\
&\qquad \leq -(r-k)\lambda_r^2(M) + (r-k)\lambda_{max}(N)^2 + \mathrm{poly}(\lambda).
\end{aligned}$$

If $k = r$, then the above inequality and inequality (71) imply that

$$\text{poly}(\lambda) \geq K = O(\alpha^2),$$

which contradicts the assumption that λ is small. Hence, we conclude that $r - k \geq 1$. Combining with (71), we obtain the upper bound

$$\begin{aligned}\lambda_r^2(M) &\leq -\frac{2}{1+\delta} \cdot \frac{K}{r-k} + \lambda_{max}(N)^2 + \frac{1}{r-k} \cdot \text{poly}(\lambda) \\ &= -\frac{2}{1+\delta} \cdot \frac{K}{r} + \lambda_{max}(N)^2 + \text{poly}(\lambda). \end{aligned} \tag{75}$$

Step III. In the last step, we combine the relations (70) and (75), which leads to

$$\sigma_r^4(U) \leq -\frac{2}{1+\delta} \cdot \frac{K}{r} + \frac{1}{(1+\delta)^2} G^2 + \text{poly}(\lambda).$$

This means that

$$\sigma_r^4(U) + \frac{2}{1+\delta} \cdot \frac{K}{r} \leq \frac{1}{(1+\delta)^2} G^2 + \text{poly}(\lambda).$$

Since $K > 0$ has lower bounds that are independent of λ, we can choose λ to be small enough such that

$$\sigma_r^4(U) + \frac{1}{1+\delta} \cdot \frac{K}{r} \leq \frac{1}{(1+\delta)^2} G^2.$$

However, considering the assumption (69), we have

$$\begin{aligned}\sigma_r^4(U) &\geq \frac{1}{(1+\delta)^2} \left[G - \frac{(1+\delta)\lambda^2}{4G^2} - \lambda \right]^2 = \frac{1}{(1+\delta)^2} G^2 - 2\lambda \cdot G + \text{poly}(\lambda) \\ &\geq \frac{1}{(1+\delta)^2} G^2 - 2\lambda \cdot (1+\delta) C^2 + \text{poly}(\lambda) = \frac{1}{(1+\delta)^2} G^2 + \text{poly}(\lambda),\end{aligned}$$

where the second inequality is due to $G \leq (1+\delta)C^2$, which can be proved similar to the asymmetric case. The above two inequalities cannot hold simultaneously when λ is small enough. This contradiction means that the condition (68) holds by choosing

$$0 < \lambda \leq \lambda_0(\delta, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha, C),$$

for a small enough positive constant $\lambda_0(\delta, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha, C)$.

□

Proof of Theorem 7. We first choose

$$C := \left[\frac{2(1+\delta)}{1-\delta} \|U^*(U^*)^T\|_F^2 \right]^{1/4}.$$

Then, we select λ_1 as

$$\lambda_1(\delta, r, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha) := \min \left\{ \lambda_0(\delta, r, \sigma_r(M_s^*), \|M_s^*\|_F, \alpha, C), \frac{(1-\delta)C^3}{2\sqrt{r}} \right\}.$$

Finally, we combine Lemmas 6-8 to get the bounds for the gradient and the Hessian. □