Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

On critical points of quadratic low-rank matrix optimization problems

(revised version: March 2020)

by

André Uschmajew and Bart Vandereycken

Preprint no.: 58 2018

On critical points of quadratic low-rank matrix optimization problems

André Uschmajew* Bart Vandereycken†

Abstract

The absence of spurious local minima in certain nonconvex low-rank matrix recovery problems has been of recent interest in computer science, machine learning and compressed sensing since it explains the convergence of some low-rank optimization methods to global optima. One such example is low-rank matrix sensing under restricted isometry properties (RIP). It can be formulated as a minimization problem for a quadratic function on the Riemannian manifold of low-rank matrices, with a positive semidefinite Riemannian Hessian that acts almost like an identity on low-rank matrices. In this work, new estimates for singular values of local minima for such problems are given which lead to improved bounds on RIP constants to ensure absence of non-optimal local minima and sufficiently negative curvature at all other critical points. A geometric viewpoint is taken which is inspired by the fact that the Euclidean distance function to a rank-k matrix possesses no critical points on the corresponding embedded submanifold of rank-k matrices except for the single global minimum.

1 Introduction

On the space $\mathbb{R}^{m\times n}$ of real $m \times n$ matrices we consider a quadratic function

$$f_{\mathcal{A},B}(X) = \tfrac{1}{2}\langle \mathcal{A}[X], X\rangle_F - \langle B, X\rangle_F \tag{1.1}$$

with a given symmetric linear operator $\mathcal{A}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ and matrix B. The gradient of this function equals

$$\nabla f_{\mathcal{A},B}(X) = \mathcal{A}[X] - B$$

and critical (or stationary) points of the function $f_{\mathcal{A},B}$ thus correspond to solutions of the linear matrix equation

$$\mathcal{A}[X] = B. \tag{1.2}$$

Thus, critical points only exist if B is in the range of $\mathcal{A}$. This is, for instance, the case when $\mathcal{A}$ is positive definite with respect to the Frobenius inner product, in which case the solution to the matrix equation (1.2) is also unique.

For $m = n$ and L a symmetric positive definite $n \times n$ matrix a well-known example of the type described above is the Lyapunov matrix equation

$$LX + XL = B, \tag{1.3}$$

which has a unique solution X^* for any right-hand side B, since the operator $\mathcal{A}[X] = LX + XL$ is symmetric positive definite on $\mathbb{R}^{m\times n}$.

*Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany (uschmajew@mis.mpg.de)

†University of Geneva, Section of Mathematics, 1211 Geneva, Switzerland (bart.vandereycken@unige.ch)

1.1 Rank constrained quadratic problems

In certain applications the aim is to solve (1.1) or (1.2) with the additional requirement that the solution (or its approximation) is of sufficiently low rank. For instance, when B in the Lyapunov equation (1.3) itself is low rank, it can be proven [30, Thm. 1] that X^* has exponentially decaying singular values and, hence, can be approximated well by a low-rank matrix. To obtain such approximations there exist a few methods that are built from classical solvers in numerical linear algebra, like the ADI and Krylov methods; see [35] for a recent overview.

Let $k \leq \min(m, n)$ and denote by

$$\mathcal{M}_k = \{X \in \mathbb{R}^{m \times n} \colon \operatorname{rank}(X) = k\}$$

the smooth manifold of fixed rank-k matrices, and by

$$\mathcal{M}_{\leq k} = \{X \in \mathbb{R}^{m \times n} \colon \operatorname{rank}(X) \leq k\}$$

its closure in $\mathbb{R}^{m \times n}$. A natural alternative approach for obtaining low-rank (approximate) solutions to the matrix equation (1.2) is to minimize the quadratic function (1.1) on the set $\mathcal{M}_{\leq k}$:

$$\min_{X \in \mathcal{M}_{\leq k}} f_{\mathcal{A},B}(X). \tag{1.4}$$

Since the set $\mathcal{M}_{\leq k}$ is closed, this problem admits at least one solution if $\mathcal{A}$ is positive definite on the cone $\mathcal{M}_{\leq 2k}$, that is, $\langle X, \mathcal{A}[X] \rangle_F > 0$ for all $X \in \mathcal{M}_{\leq 2k}$ (see Proposition 2.5 further below). In case that $B = \mathcal{A}[X^*]$ for some $X^* \in \mathbb{R}^{m \times n}$ it holds that

$$f_{\mathcal{A},B}(X) = \tfrac{1}{2} \langle X - X^*, \mathcal{A}[X - X^*] \rangle_F + f_{\mathcal{A},B}(X^*), \tag{1.5}$$

and thus, if, say, $\mathcal{A}$ is positive semidefinite, the minimizers of $f_{\mathcal{A},B}$ on $\mathcal{M}_{\leq k}$ admit an interpretation as best rank-k approximations of X^* in energy (semi)norm.

The constrained optimization problem (1.4) is nonconvex and can be tackled by various methods. While our forthcoming theoretical results do not depend on the particular method used we point out two popular and efficient approaches. The first is based on the bilinear representation $X = UV^T$ for low-rank matrices which allows to optimize the $m \times k$ and $n \times k$ matrices U and V using local search algorithms. This is also known as the Burer–Monteiro factorization in the more general context of low-rank approximations for SDPs [7] and can be very efficient when $k \ll mn$ since it avoids constructing large matrices of size $m \times n$ explicitly. It is advisable to break the nonuniqueness of the factorization $X = UV^T$ by adding penalty terms or applying alternating least squares; see, e.g., [46, 29, 48, 24].

Another family of methods is based on exploiting that the set $\mathcal{M}_k$ is a smooth Riemannian manifold. This allows again the use of local search algorithms but now using techniques from Riemannian optimization [1]. As these methods directly optimize over $\mathcal{M}_k$ they do not require regularization since non-uniqueness of representations is not an issue. In addition, careful implementations using retractions and low-rank factorization have similar cost per iteration as algorithms based on bilinear factorizations; see, e.g., [44, 34, 43].

There has been considerable interest in the case of positive semidefinite operators $\mathcal{A}$. This concerns, for instance, the non-convex formulation of low-rank matrix completion problems [9]

$$\min_{X \in \mathcal{M}_k} \tfrac{1}{2} \langle X, \mathsf{P}_\Omega[X] \rangle_F - \langle X, \mathsf{P}_\Omega[B] \rangle_F,$$

which corresponds to solutions of

$$\mathsf{P}_\Omega[X] = \mathsf{P}_\Omega[B].$$

Here, $\mathcal{A} = \mathsf{P}_\Omega$ is the orthogonal projection on a subset Ω of known entries, which means that $\mathcal{A}$ is not invertible. Solvers for these problems with the good theoretical guarantees are based on convex relaxation [9, 10], but they can also be treated well by non-convex local optimization techniques [22, 46, 21, 43, 16] which are much less costly per iteration; see also [12] for a recent overview. More generally, the problems with semidefinite operator are instances of a matrix sensing problem [31], that is, the recovery of a matrix from few linear measurement. Here in general one is faced with the problem

$$\min \tfrac{1}{2}\|\mathsf{F}[X] - b\|_F^2 \tag{1.6}$$

where F is a linear operator from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^d$ and $d \leq mn$. Up to an inconsequential constant, this problem fits in our symmetric and semidefinite framework (1.4) using $\mathcal{A} = \mathsf{F}^T\mathsf{F}$ and $B = \mathsf{F}^T[b]$.

1.2 Contributions and existing results

In this paper we focus on problems of the form (1.1) where $\mathcal{A}$ is positive semidefinite. In many applications it can be observed in numerical experiments that if one true solution X^* of the corresponding matrix equation has exactly low rank, that is, the global minimizer X^* of the function (1.1) is an element of $\mathcal{M}_k$ for some small rank k, and if this rank is known, then local optimization methods for (1.4) do typically recover this global minimizer. This is somewhat surprising since the problem is nonconvex. Furthermore, in the so-called noisy case, when X^* is only close to but not in $\mathcal{M}_k$, such algorithms typically return different local minima on $\mathcal{M}_k$ that are, however, all close to X^*.

As explained in many works the reason for this fortunate behavior seems to be a relatively benign optimization landscape: given an objective function f that is sufficiently well-conditioned and convex when restricted to cones $\mathcal{M}_{\leq k}$, the local minima always appear to be global minima after restricting f to $\mathcal{M}_k$. In other words other critical points are either saddle points or local maxima, and are hence unlikely to attract sequences generated by local optimization algorithms that impose monotonic reduction of the objective. Moreover, the saddle points have directions with sufficiently large negative curvature (called strict saddle property in [14]), so that algorithms can escape them sufficiently fast. Such remarkable properties have been rigorously proven under suitable assumptions for different low-rank optimization problems like matrix completion [40, 16], matrix sensing [3, 29], more general convex functions on $\mathcal{M}_{\leq k}$ [48, 24], SDPs [5, 6] and also for some other problems in the context of compressed sensing such as phase retrieval [39] and sparse dictionary recovery [36, 37, 38, 39]. See also [15] for an overview.

Our aim is to provide here similar results by studying the critical points of quadratic functions $f_{\mathcal{A},B}$ as in (1.1) on manifolds $\mathcal{M}_k$ for semidefinite operators $\mathcal{A}$. We will show that when the restriction of $\mathcal{A}$ to the cone $\mathcal{M}_{\leq 2k}$ behaves like a sufficiently small perturbation of identity, and if a solution of the matrix equation (1.2) lies on $\mathcal{M}_k$, then $f_{\mathcal{A},B}$ has no local minima on $\mathcal{M}_k$ except the global one. Additionally, bounds on the negative eigenvalues of the Riemannian Hessian at other critical points are given. This is important for escaping such critical points in local search methods. These results are in Theorem 3.5 and Corollary 3.6, which, to our knowledge, provide improved and simple conditions on the restricted isometry constants δ_k (see Definition 3.1 later) when applied to matrix sensing as compared to those we could find in the literature. For example we obtain that $\delta_{3k} \leq 0.3446$ or $\delta_{2k} \leq 0.2807$ are each sufficient for absence of local minima in noiseless matrix sensing with nonsymmetric matrices. This can be compared to the condition $\delta_{4k} \leq 0.0363$ in [29], $\delta_{4k} \leq 1/5$ in [48, 24] and $\delta_{2k} < 1/5$ in [3] for symmetric positive semidefinite matrices (observe that $\delta_k \leq \delta_\ell$ for all integers $k \leq \ell$ hence our bounds are less restrictive). On the other hand there exist examples showing that with $\delta_2 \geq 1/2$ quadratic functions may exhibit non-global local minima on the set of positive semidefinite rank-one matrices, even in the noiseless case [24, 47]. Other sufficient conditions in the literature for guaranteed recovery of rank-k matrices using different approaches include $\delta_{2k} < 1/3$ [20] for the singular value projection (iterative hard thresholding) algorithm, and even $\delta_{2k} \leq 1/2$ [8] for nuclear norm minimization, which has the additional theoretical advantage of not requiring the rank k as an input parameter. These last two approaches, however, can become very expensive to implement for large matrices compared to local methods that operate on rank-k matrices directly.

The most general version of our analysis is Theorem 3.9 which also deals with the inexact (or noisy) case, where the matrix equation (1.2) admits only an approximate solution on $\mathcal{M}_{\leq k}$ of some accuracy $\varepsilon \geq 0$. In this case all critical points whose Riemannian Hessians have small or no negative eigenvalues (e.g. local minima) are optimal up to constant. While the statements are in principal easy to use it might be difficult to get an intuition on the actual values. We therefore provide some concrete examples on the interplay of restricted spectral bounds, negative eigenvalues of the (Riemannian) Hessian and ε in section 3.4.

Our strategy to obtain our results is motivated by an interesting observation on the Euclidean distance function $f(X) = \|X - B\|_F^2$, namely that for $B \in \mathcal{M}_k$ it has no critical points on $\mathcal{M}_k$ at all except for the global minimum $X = B$. In order to generalize this rather peculiar behavior of the operator $\mathcal{A} = \mathsf{Id}$ to more general ones, we introduce a certain norm in which we measure the distance of $X - (\mathcal{A}[X] - B)$ from the cone $\mathcal{M}_{\leq k}$. By comparing upper and lower bounds for this distance we obtain our results.

Currently, the main results in this paper do not cover the important cases of matrix completion or matrix equations with badly conditioned operators (as they arise in numerical linear algebra), but some of the observations obtained alongside still provide general insight into the problem.

Finally, we hope to make a contribution to the subject by taking a geometric viewpoint on the problem that focuses on the critical points of the constrained problem (1.4), regardless of the method or parametrization used for representing the low-rank matrices. This is in contrast to [29, 23, 48, 24] where an explicit regularization has to be used to cope with the nonuniqueness of the $X = UV^T$ factorization. Also, thanks to the manifold setup, we believe our analysis has potential implications to most local search methods for (1.4). This is illustrated in the last section on numerical examples where we solve matrix sensing problems with a few popular nonconvex algorithms. The methods are not novel but serve their purpose in confirming our theoretical result on nonexistence of suprious local minima.

2 Properties of critical points

In this work we study the critical points of the function $f_{\mathcal{A},B}$ defined in (1.1) on the smooth manifold $\mathcal{M}_k$ of fixed rank k matrices only. We shortly justify this restriction to the smooth part of $\mathcal{M}_{\leq k}$ in subsection 2.2. In general, we neither assume that B is in the range of $\mathcal{A}$, nor that $\mathcal{A}$ is positive semidefinite. Instead, a so-called restricted positive definiteness on the cones $\mathcal{M}_{\leq k}$ will play the crucial role for the main results on absence of local minima of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ in Section 3.

2.1 Tangent space and critical points

A point $X \in \mathcal{M}_k$ is called a *critical point of $f_{\mathcal{A},X}$ on $\mathcal{M}_k$*, if $\nabla f_{\mathcal{A},B}(X) = \mathcal{A}[X] - B$ is orthogonal to the tangent space $T_X\mathcal{M}_k$ at X. This tangent space is known to be the set (see, e.g., [18, Prop. 4.1])

$$T_X\mathcal{M}_k = \{CX + XD \colon C \in \mathbb{R}^{m\times m},\ D \in \mathbb{R}^{n\times n}\}. \tag{2.1}$$

Note that all matrices in $T_X\mathcal{M}_k$ have rank at most $2k$, that is, $T_X\mathcal{M}_k \subseteq \mathcal{M}_{\le 2k}$.

Let P_X^{col} and P_X^{row} denote the respective orthogonal projections on the column and row space of a matrix X. From (2.1) we see that a matrix Z is orthogonal to $T_X\mathcal{M}_k$ if $P_X^{\mathrm{col}}Z = 0$ and $ZP_X^{\mathrm{row}} = 0$, or, in other words,

$$Z = (I - P_X^{\mathrm{col}})Z(I - P_X^{\mathrm{row}}).$$

Hence, with $X = U\Sigma V^T$ and $Z = \tilde{U}\tilde{\Sigma}\tilde{V}^T$ two SVDs, we obtain that

$$X + Z = \begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \tilde{\Sigma} \end{bmatrix} \begin{bmatrix} V & \tilde{V} \end{bmatrix}^T \tag{2.2}$$

is also an SVD. A main consequence of this is that

$$\operatorname{rank}(X+Z) = \operatorname{rank}(X) + \operatorname{rank}(Z) \quad \text{for } Z \text{ orthogonal to } T_X\mathcal{M}_k. \tag{2.3}$$

The seemingly simple observation (2.2) turns out to be quite useful and is the main argument for Lemma 2.3 below. In fact its immediate consequence (2.3) already has some surprising implication on the critical points of the Euclidean distance function, which arises, up to a constant, from $f_{\mathcal{A},B}$ by taking $\mathcal{A} = \mathsf{Id}$ to be the identity operator.

Proposition 2.1. *For any $B \in \mathcal{M}_k$ the function $f(X) = \frac{1}{2}\|X - B\|_F^2$ has only one critical point on $\mathcal{M}_k$, namely the global minimizer $X^* = B$.*

Proof. Since $\nabla f(X) = X - B$ the condition for a critical point is that $X - B$ is orthogonal to $T_X\mathcal{M}_k$. Then, by (2.3),

$$k = \operatorname{rank}(X - (X - B)) = \operatorname{rank}(X) + \operatorname{rank}(X - B) = k + \operatorname{rank}(X - B),$$

which implies $X = B$. □

The following equivalent statement is even more interesting from a geometric point of view. It follows directly from $X - Y$ being the gradient of the function $f(X) = \frac{1}{2}\|X - Y\|_F^2$.

Proposition 2.2. *Let X and Y be two distinct points on $\mathcal{M}_k$. Then $X - Y$ is not orthogonal to $T_X\mathcal{M}_k$.*

Our aim in this paper is to study how far the observation in Proposition 2.1 for the identity operator carries over to those functions $f_{\mathcal{A},B}$ in which $\mathcal{A}$ is a perturbation of the identity, at least in a restricted sense. For this we will have to quantify the 'rank increase' property (2.3). The starting point will be the inequality stated in Lemma 2.3 below, which first require some definitions.

By $\sigma_1(Z) \ge \sigma_2(Z) \ge \dots$ we denote the singular values of a matrix Z, with the agreement $\sigma_i(Z) = 0$ for $i \ge \min(m,n)$. We then consider the norm

$$\|Z\|_{\sigma,k} \coloneqq \sqrt{\sigma_1^2(Z) + \cdots + \sigma_k^2(Z)} = \max_{\substack{Y \in \mathcal{M}_{\le k} \\ \|Y\|_F = 1}} \langle Y, Z\rangle_F. \tag{2.4}$$

Here the equality of both expressions is a consequence of the fact that truncated singular value decomposition yields best approximations in Frobenius norm on the cone $\mathcal{M}_{\le k}$, and hence maximizes the orthogonal projection on it. The norm properties of $\|Z\|_{\sigma,k}$ then follow easily from the expression on the right hand side of (2.4). Note that $\|X\|_{\sigma,k} \le \|X\|_F$ for every matrix X. The norm (2.4) is a unitarily invariant norm, that is, $\|UZV\|_{\sigma,k} = \|Z\|_{\sigma,k}$ for all orthogonal U and V.

Hence, the truncated singular value decomposition also provides best rank-k approximations in this norm; see, e.g., [19, Section 7.4.9]. Therefore, for any fixed Z it holds

$$\operatorname{dist}_{\|\cdot\|_{\sigma,k}}(Z, \mathcal{M}_{\leq k}) \coloneqq \min_{Y \in \mathcal{M}_{\leq k}} \|Z - Y\|_{\sigma,k} = \sqrt{\sigma_{k+1}^2(Z) + \cdots + \sigma_{2k}^2(Z)}. \tag{2.5}$$

For the case of the identity operator $\mathcal{A} = \mathsf{Id}$ we have obtained a contradiction to the existence of two critical points $X \neq X^* = B$ on $\mathcal{M}_k$ from two facts: on the one hand, the matrix $X - (\mathcal{A}[X] - B) = X - \nabla f_{\mathcal{A},B}(X)$ should have a higher rank than X, that is a positive distance to $\mathcal{M}_{\leq k}$, while on the other hand it cannot, since it equals X^*. For $\mathcal{A}$ close to Id we expect a similar contradiction, but to obtain it, we need both upper and lower bounds for the distance of $X - (\mathcal{A}[X] - B)$ from $\mathcal{M}_{\leq k}$. Our key idea is to obtain such bounds for the distance in $\|\cdot\|_{\sigma,k}$-norm.

Lemma 2.3. *Let X be a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ and*

$$\alpha = \operatorname{dist}_{\|\cdot\|_{\sigma,k}}(X - (\mathcal{A}[X] - B), \mathcal{M}_{\leq k}).$$

(i) *For any $Y \in \mathcal{M}_{\leq k}$ it holds that*

$$\alpha \leq \|(\mathsf{Id} - \mathcal{A})[X - Y]\|_{\sigma,k} + \|B - \mathcal{A}[Y]\|_{\sigma,k}.$$

(ii) *Let $0 \leq j \leq k$ be the largest integer such that $\sigma_i(\mathcal{A}[X] - B) > \sigma_{k-i+1}(X)$ for $1 \leq i \leq j$. Then*

$$\alpha^2 \geq \sum_{i=1}^{j} \sigma_{k-i+1}^2(X) + \sum_{i=j+1}^{k} \sigma_i^2(\mathcal{A}[X] - B).$$

Proof. Item (i) is immediate from the definition of α and the triangle inequality:

$$\alpha \leq \|X - (\mathcal{A}[X] - B) - Y\|_{\sigma,k} \leq \|X - \mathcal{A}[X] + \mathcal{A}[Y] - Y\|_{\sigma,k} + \|B - \mathcal{A}[Y]\|_{\sigma,k}.$$

To show (ii) we use the characterization

$$\alpha = \sqrt{\sigma_{k+1}^2(X - \mathcal{A}[X] + B) + \cdots + \sigma_{2k}^2(X - \mathcal{A}[X] + B)},$$

which holds by (2.5). Let $\varsigma_i = \sigma_i(X)$ and $s_i = \sigma_i(\mathcal{A}[X] - B)$ for abbreviation. By (2.2) (with $Z = -\mathcal{A}[X] + B$) the largest $2k$ singular values of the matrix $X - \mathcal{A}[X] + B$ are among the $3k$ numbers $\varsigma_1, \dots, \varsigma_k, s_1, \dots, s_{2k}$. By definition of j the largest k of these numbers are $\varsigma_1, \dots, \varsigma_{k-j}, s_1, \dots, s_j$ (the notation is slightly abusive when $j = k$), and hence α^2 is the sum of squares of the largest k remaining ones. In particular α^2 is larger or equal to any sum of squares of k of the remaining singular values, which implies the asserted lower bound. ☐

In agreement to what was pointed out above, the inequalities (i) and (ii) in Lemma 2.3 are contradictory in the case $\mathcal{A} = \mathsf{Id}$ and $Y = B \in \mathcal{M}_k$ unless $X = B$. Our strategy is to show that they remain contradictory when $\mathcal{A}$ acts like a perturbation of identity on low-rank matrices. However, different from the case $\mathcal{A} = \mathsf{Id}$, we will have to confine ourselves to local minima on $\mathcal{M}_k$, or at least critical points with almost positive semidefinite Riemannian Hessian (see Section 2.4), in order to deal with the *a priori* unknown singular values of X in the lower bound for α.

2.2 Restriction to the smooth part $\mathcal{M}_k$

We justify why we are ignoring potential local minima of $f_{\mathcal{A},B}$ on $\mathcal{M}_{\le k}$ of rank less than k. By the textbook definition (e.g., [32, Theorem 6.12]) $X \in \mathcal{M}_{\le k}$ is a critical point of the nonsmooth problem (1.4), if $-\nabla f_{\mathcal{A},B}(X) = -\mathcal{A}[X] - B$ belongs to the polar cone of the Bouligand tangent cone at X. In particular local minima on $\mathcal{M}_{\le k}$ are critical points. If $\operatorname{rank}(X) = s \le k$ the Bouligand tangent cone can be shown to be [17, 11, 33]

$$T_X^B \mathcal{M}_{\le k} = T_X \mathcal{M}_s + \{Z \in \mathbb{R}^{m\times n} \colon \operatorname{rank}(Z) \le k - s\}.$$

As it then follows, when $s < k$, the polar cone $(T_X^B \mathcal{M}_{\le k})^\circ$ is just the point $\{0\}$, and hence a critical point satisfies $\nabla f_{\mathcal{A},B}(X) = 0$ that is, solves the equation $\mathcal{A}[X] = B$. Let us repeat this as a proposition.

Proposition 2.4. *Let $X \in \mathcal{M}_{\le k}$ be a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_{\le k}$ in the sense $-\nabla f_{\mathcal{A},B}(X) \in (T_X^B \mathcal{M}_{\le k})^\circ$. Then either* $\operatorname{rank}(X) = k$ *and X is a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$, or $\mathcal{A}[X] = B$. In the latter case, if $\mathcal{A}$ is a positive semidefinite operator, then X is a global minimizer of $f_{\mathcal{A},B}$ on $\mathbb{R}^{m\times n}$.*

Now we can make different assumptions regarding the critical points X satisfying $\operatorname{rank}(X) < k$. If we assume $\mathcal{A}$ is positive semidefinite, then by the above proposition such X are necessarily unconstrained global minimizers of $f_{\mathcal{A},B}$. If we assume instead that there exist at least one solution $\mathcal{A}[X^*] = B$ with $\operatorname{rank}(X^*) \le k$ and $\mathcal{A}$ satisfies the condition $\lambda(\mathcal{A}, 2k) > 0$ (see (2.6) below for the definition) then it follows $X = X^*$. Finally, if we simply assume that the equation $\mathcal{A}[X] = B$ does not admit solutions of rank strictly less than k at all, such a critical point X cannot exist and therefore all critical points of $f_{\mathcal{A},B}$ on $\mathcal{M}_{\le k}$ in this broader sense in fact lie in $\mathcal{M}_k$. This is for instance the case if $\mathcal{A}$ is positive definite and $\operatorname{rank}(X^*) \ge k$, where X^* is the unique solution $\mathcal{A}[X^*] = B$, that is, the global unconstrained minimizer of $f_{\mathcal{A},B}$ [33].

Based on these facts all subsequent theorems will be formulated for critical points on the smooth manifold $\mathcal{M}_k$ only. A key challenge, however, is to bound the distance of critical points $X \in \mathcal{M}_k$ to $\mathcal{M}_{\le k-1}$, that is, the smallest singular value $\sigma_k(X)$, from below, see subsection 2.4.

2.3 Restricted spectral bounds

The central tool to analyze the local minima of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ is the 'restricted spectral bounds', that is, the minima and maxima of the Rayleigh quotient of the symmetric operator $\mathcal{A}$ on cones of low-rank matrices. We use the following definitions:

$$\lambda(\mathcal{A}, k) = \min_{\substack{X \in \mathcal{M}_{\le k} \\ \|X\|_F = 1}} \langle X, \mathcal{A}[X] \rangle_F, \tag{2.6}$$

and

$$\Lambda(\mathcal{A}, k) = \max_{\substack{X \in \mathcal{M}_{\le k} \\ \|X\|_F = 1}} \langle X, \mathcal{A}[X] \rangle_F. \tag{2.7}$$

Note that both the minimum and maximum are attained since $\mathcal{M}_{\le k}$ is closed.

Obviously, whenever $k' \ge k$ it holds

$$\lambda(\mathcal{A}, k') \le \lambda(\mathcal{A}, k) \le \Lambda(\mathcal{A}, k) \le \Lambda(\mathcal{A}, k').$$

In particular,

$$\lambda(\mathcal{A}, k) \le \Lambda(\mathcal{A}, \ell)$$

for all combinations of k and ℓ.

If $\mathcal{A} \neq 0$ is positive semidefinite it holds $\Lambda(\mathcal{A}, 1) > 0$, since the space $\mathbb{R}^{m\times n}$ possesses an orthonormal basis of rank-one matrices. Furthermore, one can then show that

$$\Lambda(\mathcal{A}, k+\ell) \leq \Lambda(\mathcal{A}, k) + \Lambda(\mathcal{A}, \ell).$$

For this inequality to hold it is sufficient that $\lambda(\mathcal{A}, k+\ell) \geq 0$.[1]

We note that the lower spectral bounds provide conditions for the existence of minimizers as follows.

Proposition 2.5. *Assume* $\lambda(\mathcal{A}, 2k) > 0$*. Then the function* $f_{\mathcal{A},B}$ *has at least one minimizer on* $\mathcal{M}_{\leq k}$*, that is, the problem* (1.4) *admits at least one solution.*

Proof. Fix $Y \in \mathcal{M}_{\leq k}$. Since $\lambda(\mathcal{A}, 2k) > 0$ the representation

$$f_{\mathcal{A},B}(X) = f_{\mathcal{A},B}(Y) + \langle A[Y] - B, X - Y\rangle_F + \frac{1}{2}\langle X - Y, \mathcal{A}[X - Y]\rangle_F$$

easily shows that f is coercive on $\mathcal{M}_{\leq k}$, that is, $f_{\mathcal{A},B}(X) \to \infty$ for $\|X\|_F \to \infty$ on $\mathcal{M}_{\leq k}$. It means that the restriction of $f_{\mathcal{A},B}$ to $\mathcal{M}_{\leq k}$ has bounded sublevel sets, and so the existence of a minimizer follows from the fact that $\mathcal{M}_{\leq k}$ is closed. □

We will also need upper estimates for mixed products $\langle Y, \mathcal{A}[Z]\rangle_F$ in terms of the restricted spectral bounds. They can be derived using the 'parallelogram identity', similar to [10, Lemma 3.3].

Lemma 2.6. *Let* $\lambda_i \leq \lambda(\mathcal{A}, i) \leq \Lambda(\mathcal{A}, i) \leq \Lambda_i$ *for all* i*. Then for any* $Y \in \mathcal{M}_{\leq k}$ *and* $Z \in \mathcal{M}_{\leq \ell}$ *it holds that*

$$\langle Y, \mathcal{A}[Z]\rangle_F \leq \frac{1}{4}(\Lambda_{k+\ell} - \lambda_{k+\ell})(\|Y\|_F^2 + \|Z\|_F^2) + \frac{1}{2}(\Lambda_{k+\ell} + \lambda_{k+\ell})\langle Y, Z\rangle_F. \tag{2.8}$$

Proof. Since $\mathcal{A}$ is symmetric we have

$$\begin{aligned} 4\langle Y, \mathcal{A}[Z]\rangle_F &= \langle Y + Z, \mathcal{A}[Y + Z]\rangle_F - \langle Y - Z, \mathcal{A}[Y - Z]\rangle_F \\ &\leq \Lambda_{k+\ell}\|Y + Z\|_F^2 - \lambda_{k+\ell}\|Y - Z\|_F^2, \end{aligned}$$

which easily yields the asserted bound. □

The upper bound (2.8) will be required for the shifted operator $\mathsf{Id} - \mathcal{A}$. Under the assumptions of the lemma, it follows from the definitions that

$$\Lambda(\mathsf{Id} - \mathcal{A}, k) = 1 - \lambda(\mathcal{A}, k) \leq 1 - \lambda_k$$

and

$$\lambda(\mathsf{Id} - \mathcal{A}, k) = 1 - \Lambda(\mathcal{A}, k) \geq 1 - \Lambda_k$$

for all k. Therefore, by applying (2.8),

$$\langle Y, (\mathsf{Id} - \mathcal{A})[Z]\rangle_F \leq \frac{1}{4}(\Lambda_{k+\ell} - \lambda_{k+\ell})(\|Y\|_F^2 + \|Z\|_F^2) + \frac{1}{2}(2 - \Lambda_{k+\ell} - \lambda_{k+\ell})\langle Y, Z\rangle_F. \tag{2.9}$$

[1] Using SVD, every matrix Z of rank at most $k+\ell$ can be written as $Z = sX + tY$, where $s, t \in \mathbb{R}$ and X and Y are of rank at most k and ℓ, respectively, and orthonormal with respect to the Frobenius inner product. Consider then the 2×2 symmetric matrix $G = \begin{bmatrix} \langle X, \mathcal{A}[X]\rangle_F & \langle Y, \mathcal{A}[X]\rangle_F \\ \langle X, \mathcal{A}[Y]\rangle_F & \langle Y, \mathcal{A}[Y]\rangle_F \end{bmatrix}$. With $a = [s,\ t]^T$ it follows $\langle Z, \mathcal{A}[Z]\rangle_F = a^T G a$. Hence, the matrix G is positive semidefinite since $\lambda(\mathcal{A}, k+\ell) \geq 0$. From $a^T G a \leq \operatorname{trace}(G) = (\langle X, \mathcal{A}[X]\rangle_F + \langle Y, \mathcal{A}[Y]\rangle_F)\|a\|_2^2 \leq (\Lambda_k + \Lambda_\ell)\|Z\|_F^2$ one obtains the result.

Let us further introduce the constants

$$\Gamma(\mathcal{A},k,\ell) = \max_{\substack{Y\in\mathcal{M}_{\le k},\ Z\in\mathcal{M}_{\le \ell} \\ \|Y\|_F=\|Z\|_F=1}} \langle Y, \mathcal{A}[Z]\rangle_F.$$

They can be related to the $\|\cdot\|_{\sigma,k}$-norms introduced in (2.4) in the following way.

Lemma 2.7. *Let Z have rank ℓ, then $\|\mathcal{A}[Z]\|_{\sigma,k} \le \Gamma(\mathcal{A},k,\ell)\|Z\|_F$.*

The proof is immediate from the right hand side of (2.4).

The scaling behavior of $\Gamma(\mathcal{A},k,\ell)$ with respect to the ranks k and ℓ will turn out to be useful later to relate our results to existing ones.

Lemma 2.8. *For the positive integers p,q it holds that*

$$\Gamma(\mathcal{A},pk,q\ell) \le \sqrt{pq}\,\Gamma(\mathcal{A},k,\ell).$$

Proof. Let Y and Z be the maximizers in $\Gamma(\mathcal{A},pk,q\ell)$. Using SVD, we can write $Y = a_1Y_1 + \cdots + a_pY_p$, where the matrices $Y_1,\dots,Y_p \in \mathcal{M}_{\le k}$ are pairwise orthogonal and have Frobenius norm one, and the scalars $a_1,\dots,a_p$ are not negative. Observe that $a_1^2+\cdots+a_p^2 = 1$. We decompose $Z = b_1Z_1+\cdots+b_qZ_q$ similarly. Hence,

$$\Gamma(\mathcal{A},pk,q\ell) = \sum_{i=1}^{p}\sum_{j=1}^{q} a_ib_j\langle Y_i,\mathcal{A}[Z_j]\rangle_F \le \left(\sum_{i=1}^{p} a_i\right)\left(\sum_{j=1}^{q} b_j\right)\Gamma(\mathcal{A},k,\ell)$$

and the result follows from the Cauchy–Schwarz inequality. $\square$

2.4 Estimates related to the Riemannian Hessian

Here we provide lower estimates on the smallest singular values of a critical point $X\in\mathcal{M}_k$ of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$. These estimates are expressed in terms of the restricted spectral bounds (2.6)–(2.7) of the operator $\mathcal{A}$, the singular values of the residual $\mathcal{A}[X]-B$, as well as a lower bound on the eigenvalues of the Riemannian Hessian of $f_{\mathcal{A},B}$ at X. We refer to [1, Ch. 5] for the concept of the Riemannian Hessian.

Denote by $\mathcal{H}_X$ the Riemannian Hessian of $f_{\mathcal{A},B}$ (restricted to $\mathcal{M}_k$) at $X\in\mathcal{M}_k$. As metric on the submanifold $\mathcal{M}_k$ we choose the restriction of the Frobenius inner product from the ambient space $\mathbb{R}^{m\times n}$.[2] Let $X = U\Sigma V^T$ an SVD of X, and set

$$\bar{G} = U\Sigma^{1/2}, \quad \bar{H} = V\Sigma^{1/2}.$$

By (2.1) every tangent vector $Z\in T_X\mathcal{M}_k$ can be written as

$$Z = \Delta G\cdot\bar{H}^T + \bar{G}\cdot\Delta H^T, \tag{2.10}$$

for some matrices $\Delta G\in\mathbb{R}^{m\times k}$ and $\Delta H\in\mathbb{R}^{n\times k}$. With this representation of tangent vectors we prove in Appendix A that

$$\mathcal{H}_X[Z,Z] = \langle Z,\mathcal{A}[Z]\rangle_F + 2\langle \Delta G\cdot\Delta H^T, (I-P_X^{\mathrm{col}})(\mathcal{A}[X]-B)(I-P_X^{\mathrm{row}})\rangle_F, \tag{2.11}$$

[2]This is arguably the most simple metric one can take. It is used in the Riemannian algorithms in [44, 34, 43, 45] for low-rank optimization. Other metrics also exist in [25, 27, 26] but they do not always lead to improved bounds in the current context.

with P_X^{col} and P_X^{row} being the orthogonal projections onto the column and row space of X, respectively.

When X is a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$, then $(I - P_X^{\mathrm{col}})(\mathcal{A}[X] - B)(I - P_X^{\mathrm{row}}) = \mathcal{A}[X] - B$ (see section 2.1), and so the Riemannian Hessian at such points reads

$$\mathcal{H}_X[Z, Z] = \langle Z, \mathcal{A}[Z] \rangle_F + 2\langle \Delta G \cdot \Delta H^T, \mathcal{A}[X] - B \rangle_F. \tag{2.12}$$

In case that X is a local minimum the Riemannian Hessian is positive semidefinite, that is, $\mathcal{H}_X[Z, Z] \geq 0$ for all $Z \in T_X\mathcal{M}_k$. In the following proposition we consider arbitrary critical points X for which the Riemannian Hessian satisfies a nonpositive lower spectral bound.

Proposition 2.9. *Let X be a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ and let $\varsigma_1 \geq \cdots \geq \varsigma_k > 0$ denote its singular values. Further, let $s_1 \geq \cdots \geq s_k \geq 0$ denote some k largest singular values (some might be zero) of $\mathcal{A}[X] - B$. Assume for some $\mu \geq 0$ that the Riemannian Hessian satisfies*

$$\mathcal{H}_X[Z, Z] \geq -\mu \|Z\|_F^2 \quad \textit{for all } Z \in T_X\mathcal{M}_k.$$

Then for any $j = 1, \ldots, k$ and $\Lambda_{2j} > 0$ with $\Lambda(\mathcal{A}, 2j) \leq \Lambda_{2j}$ it holds

$$\sqrt{\varsigma_k^2 + \cdots + \varsigma_{k-j+1}^2} \geq \frac{\sqrt{s_1^2 + \cdots + s_j^2}}{\Lambda_{2j} + \mu} = \frac{\|\mathcal{A}[X] - B\|_{\sigma,j}}{\Lambda_{2j} + \mu}.$$

Proof. Let $p_1, \ldots, p_j$ be the normalized dominant j left singular vectors of $\mathcal{A}[X] - B$, and $q_1, \ldots, q_j$ the dominant right singular vectors (or some of them if there are equal singular values). We consider the matrices

$$\Delta G = \begin{bmatrix} 0 & \cdots & 0 & s_1^{1/2} p_1 & \cdots & s_j^{1/2} p_j \end{bmatrix},$$

and

$$\Delta H = \begin{bmatrix} 0 & \cdots & 0 & -s_1^{1/2} q_1 & \cdots & -s_j^{1/2} q_j \end{bmatrix}.$$

Then

$$\langle \Delta G \cdot \Delta H^T, \mathcal{A}[X] - B \rangle_F = -(s_1^2 + \cdots + s_j^2).$$

We now consider the tangent vector

$$Z = \Delta G \cdot \bar{H}^T + \bar{G} \cdot \Delta H^T = \sum_{i=1}^{j} s_i^{1/2} \varsigma_{k-j+i}^{1/2} p_i v_{k-j+i}^T - s_i^{1/2} \varsigma_{k-j+i}^{1/2} u_{k-j+i} q_i^T$$

of the form (2.10) for this choice of ΔG and ΔH. Since X is a critical point, that is, $\mathcal{A}[X] - B$ is orthogonal to $T_X\mathcal{M}_k$, each vector p_i is orthogonal to all columns $u_1, \ldots, u_k$ of U, and each q_i is orthogonal to the columns $v_1, \ldots, v_k$ of V (see section 2.1), or $s_i = 0$. Therefore, Z is a sum of $2j$ rank-one (or zero) matrices that are pairwise orthogonal with respect to the Frobenius inner product. Thus,

$$\|Z\|_F^2 = 2(s_1 \varsigma_{k-j+1} + \cdots + s_j \varsigma_k).$$

In light of (2.12) one obtains the inequalities

$$-2\mu(s_1 \varsigma_{k-j+1} + \cdots + s_j \varsigma_k) \leq \mathcal{H}_X[Z, Z] \leq 2\Lambda_{2j}(s_1 \varsigma_{k-j+1} + \cdots + s_j \varsigma_k) - 2(s_1^2 + \cdots + s_j^2).$$

Rearranging and applying the Cauchy–Schwarz inequality leads to

$$s_1^2 + \cdots + s_j^2 \leq (\Lambda_{2j} + \mu)\sqrt{\varsigma_k^2 + \cdots + \varsigma_{k-j+1}^2}\sqrt{s_1^2 + \cdots + s_j^2},$$

which is equivalent to the asserted inequality. $\square$

We present two corollaries of Proposition 2.9 that will not be used later, but are of independent interest. They concern the positive semidefinite case $\mu = 0$, which includes local minima, for the case that $B = \mathcal{A}[X^*]$ for some $X^* \in \mathbb{R}^{m \times n}$, that is, B is in the range of $\mathcal{A}$.

Corollary 2.10. *Let $B = \mathcal{A}[X^*]$ and X be a critical point of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ at which the Riemannian Hessian $\mathcal{H}_X$ is positive semidefinite. Assume $\operatorname{rank}(X - X^*) \leq \ell$ and $\lambda(\mathcal{A}, \ell) > 0$. Then the k-th singular value of X satisfies the inequality*

$$\sigma_k(X) \geq \frac{1}{\sqrt{\ell}} \cdot \frac{\lambda(\mathcal{A}, \ell)}{\Lambda(\mathcal{A}, 2)} \cdot \|X - X^*\|_F.$$

Proof. Let $s_1 \geq \cdots \geq s_\ell$ denote the ℓ largest singular values (some might be zero) of $\mathcal{A}[X] - B = \mathcal{A}[X - X^*]$. Then, since $X - X^* \in \mathcal{M}_{\leq \ell}$ and by (2.4), we have the lower bounds

$$s_1 \geq \sqrt{\frac{s_1^2 + \cdots + s_\ell^2}{\ell}} = \frac{\|\mathcal{A}[X - X^*]\|_{\sigma,\ell}}{\sqrt{\ell}} \geq \frac{\langle X - X^*, \mathcal{A}[X - X^*]\rangle_F}{\sqrt{\ell}\|X - X^*\|_F} \geq \frac{\lambda(\mathcal{A}, \ell)}{\sqrt{\ell}}\|X - X^*\|_F.$$

The assertion now follows from the previous proposition with $\mu = 0$, $j = 1$, and $\Lambda_2 = \Lambda(\mathcal{A}, 2)$, which is positive by assumption. ☐

Since the kth singular value of a rank-k matrix equals its distance to $\mathcal{M}_{\leq k-1}$ in Frobenius norm, the previous corollary can be rephrased in the following way.

Corollary 2.11. *Under the same assumptions as in Corollary 2.10 it holds that*

$$\operatorname{dist}_F(X, \mathcal{M}_{\leq k-1}) \geq \frac{1}{2}\left(\frac{1}{\sqrt{\ell}} \cdot \frac{\lambda(\mathcal{A}, \ell)}{\Lambda(\mathcal{A}, 2)}\right)\sigma_k(X^*).$$

Proof. If $\|X - X^*\|_F \leq \sigma_k(X^*)/2$, then

$$\sigma_k(X^*) \leq \operatorname{dist}_F(X^*, \mathcal{M}_{\leq k-1}) \leq \operatorname{dist}_F(X, \mathcal{M}_{\leq k-1}) + \|X - X^*\|_F$$

implies $\operatorname{dist}_F(X, \mathcal{M}_{\leq k-1}) \geq \sigma_k(X^*)/2$, which is stronger than the asserted bound (since $\lambda(\mathcal{A}, \ell) \leq \Lambda(\mathcal{A}, 2)$). If on the other hand $\|X - X^*\|_F > \sigma_k(X^*)/2$ the previous corollary provides the asserted bound since $\operatorname{dist}_F(X, \mathcal{M}_{\leq k-1}) = \sigma_k(X)$. ☐

Note that in the case that $X^* \in \mathcal{M}_k$ we can choose $\ell = 2k$ and the lower bounds in both corollaries become independent of the size of considered matrices.

3 RPD property and its implications for critical points

We now come to the main results of the paper on the critical points of the function $f_{\mathcal{A},B}$ for the case that $\mathcal{A}$ almost acts as an identity operator on cones of low-rank matrices. This property is quantified by the *restricted positive definiteness* (RPD) constants below, which are equivalent to the restricted isometry property (RIP) constants in matrix sensing. The most notable result then is on the case that $B = \mathcal{A}[X^*]$ for some $X^* \in \mathcal{M}_k$. In this so-called 'noiseless case', and under the RPD assumptions, one can show that $f_{\mathcal{A},B}$ has no local minima on $\mathcal{M}_k$ except the single global minimum X^*. Moreover, at all other critical points the Riemannian Hessian has sufficiently negative eigenvalues, which is important in optimization methods in order to 'escape' such saddle points. The required bounds for the RPD constants for obtaining this conclusion are, to our knowledge, considerably weaker than the ones available in the literature. The results on the noiseless case are stated in Section 3.2. In Section 3.3 the most general version of our analysis is stated which deals with the case that the equation $\mathcal{A}[X] = B$ admits only an approximate solution X_ε on the set $\mathcal{M}_{\leq k}$. Then local minima or saddle points with small negative curvature may exist, but their distance to X_ε will be bounded.

3.1 RPD property

We still consider the family (1.1) of quadratic functions $f_{\mathcal{A},B}$, and make some assumptions on the restricted spectral bounds of $\mathcal{A}$ that quantify deviation from identity.

Definition 3.1 (RPD property)**.** Let $k \geq 1$. We say that the symmetric operator $\mathcal{A}$ satisfies a (k, δ_k)-*RPD property*, if there exist $0 \leq \delta_k < 1$ such that

$$1 - \delta_k \leq \lambda(\mathcal{A}, k) \leq \Lambda(\mathcal{A}, k) \leq 1 + \delta_k, \tag{3.1}$$

with $\lambda(\mathcal{A}, k)$ and $\Lambda(\mathcal{A}, k)$ defined in (2.6) and (2.7).

Remark 3.2. In the context of the matrix sensing problem (1.6) the RIP condition,

$$(1 - \delta_k)\|X\|_F^2 \leq \|\mathsf{F}[X]\|_F^2 \leq (1 + \delta_k)\|X\|_F^2 \quad \text{for all } X \in \mathcal{M}_{\leq k},$$

has been introduced in [31] and used in many subsequent works. We have already mentioned that the matrix sensing problem fits our framework using the operator $\mathcal{A} = \mathsf{F}^T\mathsf{F}$. So the RPD property above is equal to the RIP in this model. In fact, if we assume $\mathcal{A}$ to be positive semidefinite, we can always find a decomposition $\mathcal{A} = \mathsf{F}^T\mathsf{F}$ where $\mathsf{F}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^d$ with $d = \operatorname{rank}(\mathcal{A})$. Then we can write

$$f_{\mathsf{A},B}(X) = \tfrac{1}{2}(\|\mathsf{F}[X] - b\|_F^2 - \|b\|_F^2)$$

with $B = \mathsf{F}^T[b]$, so that there is no essential difference between the matrix sensing problem and our setup of finding critical points of (then convex) quadratic functions $f_{\mathcal{A},B}$ on $\mathcal{M}_k$. Regarding the existence of operators $\mathcal{A}$ obeying suitable RPD bounds we can therefore rely on the well known results on operators F satisfying RIP conditions; see [31]. A typical result is, for example, that specifically scaled random Gaussian matrices give rise to F that satisfy a (k, δ_k)-RIP with high probability if $d \geq \delta_k^{-2}k(m + n)$; see [10]. We refer to [13] for more references on the RIP.

Remark 3.3. Let us comment on operators whose restricted spectral bounds are not centered around one. We may say that the symmetric operator $\mathcal{A}$ is positive definite on the cone $\mathcal{M}_{\leq k}$ if $\lambda(\mathcal{A}, k) > 0$. We can then define the restricted condition number as

$$\kappa(\mathcal{A}, k) = \frac{\Lambda(\mathcal{A}, k)}{\lambda(\mathcal{A}, k)},$$

and consider the scaled operator $\omega_k\mathcal{A}$ with

$$\omega_k = \frac{2}{\lambda(\mathcal{A}, k) + \Lambda(\mathcal{A}, k)}.$$

This operator has the spectral bounds

$$\lambda(\omega_k\mathcal{A}, k) = \omega_k\lambda(\mathcal{A}, k) = 1 - \delta_k, \quad \Lambda(\omega_k\mathcal{A}, k) = \omega_k\Lambda(\mathcal{A}, k) = 1 + \delta_k,$$

where

$$\delta_k = \frac{\kappa(\mathcal{A}, k) - 1}{\kappa(\mathcal{A}, k) + 1}.$$

Obviously, a matrix $X \in \mathcal{M}_k$ is a critical point (local minimum) of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ if and only if it is a critical point (local minimum) of $\omega f_{\mathcal{A},B} = f_{\omega\mathcal{A},\omega B}$ for any $\omega > 0$. Therefore, the results below on those operators that satisfy RPD conditions translate to more general operators $\mathcal{A}$ if the properly scaled operator $\omega_k\mathcal{A}$ satisfies the assumptions. Yet this will mean that $\mathcal{A}$ must have a rather small restricted condition number $\kappa(\mathcal{A}, 2k)$ or $\kappa(\mathcal{A}, 3k)$. We will comment on this issue at the end of Section 3.2, and in Remark 3.11.

We note that with the RPD bounds (3.1) the estimate (2.9) leads in a straightforward way into the following upper bound (see also [29, Proposition 2.1] for essentially the same result).

Lemma 3.4. *Under the (k, δ_k)-RPD conditions it holds*

$$\Gamma(\mathsf{Id} - \mathcal{A}, k, \ell) = \max_{\substack{Y \in \mathcal{M}_{\leq k},\ Z \in \mathcal{M}_{\leq \ell} \\ \|Y\|_F = \|Z\|_F = 1}} \langle Y, (\mathcal{A} - \mathsf{Id})[Z] \rangle_F \leq \delta_{k+\ell}.$$

3.2 Noiseless case

In the noiseless case we assume that there exist $X^* \in \mathcal{M}_k$ such that $B = \mathcal{A}[X^*]$. In other words it is assumed that a desired low-rank solution X^* to the matrix equation $\mathcal{A}[X] = B$ can be found among the critical points of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ (provided we know k), which allows, for example, using Riemannian optimization methods for finding it.

Theorem 3.5. *Let $X^* \in \mathcal{M}_k$ such that $B = \mathcal{A}[X^*]$ and $\mu \geq 0$. Assume $\mathcal{A}$ satisfies RPD properties such that*

$$\delta_{3k} < -\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right)^2 + \frac{1}{\sqrt{2}}}.$$

Then on $\mathcal{M}_k$, X^* is the unique solution of $\mathcal{A}[X] = B$ and the unique global minimum of $f_{\mathcal{A},B}$. At all other critical points $X \neq X^*$ of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ the Riemannian Hessian satisfies

$$\mathcal{H}_X[Z, Z] < -\mu \|Z\|_F^2$$

for some tangent vector $Z \in T_X\mathcal{M}_k$. Alternatively, the same statements hold in the case that

$$\delta_{2k} < -\left(\frac{3}{4} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{3}{4} + \frac{\mu}{2}\right)^2 + \frac{1}{2}}.$$

Proof. The uniqueness statements follow from the representation (1.5) together with the RPD assumptions. The statement on the other critical points follows from considering the special case $\varepsilon = 0$ in the more general Theorem 3.9 proved further below. □

As an example consider the value $\mu = 1$. Then under the conditions

$$\delta_{3k} \leq 0.2399 \quad \text{or} \quad \delta_{2k} \leq 0.1861$$

the Riemannian Hessian at all critical points except the global minimum X^* has a negative eigenvalue smaller than -1. More examples on the relation between μ and δ are presented in Tables 1 and 2 below.

We highlight the case $\mu = 0$ separately as it implies the absence of local minima.

Corollary 3.6. *Let $X^* \in \mathcal{M}_k$ and $\mathcal{A}[X^*] = B$. Assume $\mathcal{A}$ satisfies RPD properties such that*

$$\delta_{3k} < -\frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right) + \sqrt{\frac{1}{4}\left(1 + \frac{1}{\sqrt{2}}\right)^2 + \frac{1}{\sqrt{2}}} \approx 0.3446.$$

Then on $\mathcal{M}_k$, X^* is the unique solution of $\mathcal{A}[X] = B$ and the unique global minimum of $f_{\mathcal{A},B}$. There exist no other local minima of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$. Alternatively, the same statements hold in the case that

$$\delta_{2k} < -\frac{1}{2}\left(1 + \frac{1}{2}\right) + \sqrt{\frac{1}{4}\left(1 + \frac{1}{2}\right)^2 + \frac{1}{2}} = \frac{\sqrt{17} - 3}{4} \approx 0.2807.$$

For reference we also generalize Theorem 3.5 to operators whose restricted spectral bounds are not centered around one. The proof follows according to Remark 3.3 by considering the scaled operators $\omega_{3k}\mathcal{A}$ and $\omega_{2k}\mathcal{A}$, respectively.

Corollary 3.7. *Let $X^* \in \mathcal{M}_k$ and $B = \mathcal{A}[X^*]$ and $\mu \geq 0$. Assume $\lambda(\mathcal{A}, 3k) > 0$ and that*

$$\frac{\kappa(\mathcal{A},3k)-1}{\kappa(\mathcal{A},3k)+1} < -\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right)^2 + \frac{1}{\sqrt{2}}},$$

where $\kappa(\mathcal{A}, 3k) = \Lambda(\mathcal{A}, 3k)/\lambda(\mathcal{A}, 3k)$. Then X^* is the unique solution of $\mathcal{A}[X] = B$ on $\mathcal{M}_k$ and the unique global minimum of $f_{\mathcal{A},B}$. At all other critical points $X \neq X^*$ of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ the Riemannian Hessian satisfies

$$\mathcal{H}_X[Z, Z] < -\frac{1}{2}(\lambda(\mathcal{A}, 3k) + \Lambda(\mathcal{A}, 3k))\mu\|Z\|_F^2$$

for some tangent vector $Z \in T_X\mathcal{M}_k$. Alternatively, the same conclusions hold (with $\lambda(\mathcal{A}, 3k), \Lambda(\mathcal{A}, 3k)$ replaced by $\lambda(\mathcal{A}, 2k), \Lambda(\mathcal{A}, 2k)$) in the case that $\lambda(\mathcal{A}, 2k) > 0$ and

$$\frac{\kappa(\mathcal{A},2k)-1}{\kappa(\mathcal{A},2k)+1} < -\left(\frac{3}{4} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{3}{4} + \frac{\mu}{2}\right)^2 + \frac{1}{2}}.$$

Taking $\mu = 0$ we get the following sufficient conditions for the absence of local minima $X \neq X^*$:

$$\kappa(\mathcal{A}, 3k) \leq \frac{2}{1 - 0.3446} - 1 \leq 2.0515, \quad \kappa(\mathcal{A}, 2k) \leq \frac{2}{1 - 0.2807} - 1 \leq 1.7804.$$

3.3 General case

In the general case the exact solution X^* of the matrix equation $\mathcal{A}[X] = B$ may have rank larger than k but still admits good low-rank approximations, that is, it will be close to $\mathcal{M}_{\leq k}$. Or, the given matrix B may only satisfy $\mathcal{A}[X^*] \approx B$ approximately and may not even belong to the range of $\mathcal{A}$. In the matrix sensing problem (1.6) this may correspond to noisy observations b. For linear matrix equations (1.2) this corresponds to a perturbation of the right hand side.

In the main result of this paper we focus on points $X_\varepsilon \in \mathcal{M}_{\leq k}$ that satisfy $\|\mathcal{A}[X_\varepsilon]-B\|_{\sigma,2k} \leq \varepsilon$ and estimate the distance of certain other critical points on $\mathcal{M}_k$ (including local minima) to X_ε. After giving the proof we calculate some concrete values and make some comments how to interpret this result in the context of the strict saddle point property. Regarding potential critical points of $f_{\mathcal{A},B}$ on $\mathcal{M}_{\leq k}$ with rank strictly less than k we refer once more to Proposition 2.4.

We first present a lemma that will allow to state our conditions on the RPD constants in the main result more conveniently.

Lemma 3.8. *Let $c, \mu > 0$. Then the restriction of the function*

$$\delta \mapsto K(c, \mu, \delta) = \left(\frac{1-\delta}{1+\delta+\mu} - c\delta\right)^{-1} = \frac{1+\delta+\mu}{1-(1+c+c\mu)\delta - c\delta^2}$$

to the positive axis possesses a single pole and is positive if and only if

$$0 < \delta < -\left(\frac{1+c}{2c} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{1+c}{2c} + \frac{\mu}{2}\right)^2 + \frac{1}{c}}.$$

For fixed c and μ it holds $K(c, \mu, \delta) \to \infty$ when δ approaches the upper bound. On the other hand the bound is monotonically decreasing with respect to c and μ.

Proof. The statement is equivalent to

$$\delta^2 + \big(1 + \tfrac{1}{c} + \mu\big)\delta - \tfrac{1}{c} < 0$$

under the restriction $\delta > 0$. This open parabola is negative at zero and therefore δ must lie between zero and the positive root, which is the asserted condition. $\square$

We state the main result. Note that we could replace the $\|\cdot\|_{\sigma,2k}$ norm in the assumptions with the Frobenius norm, since $\|\mathcal{A}[X_\varepsilon] - B\|_F \le \varepsilon$ would be a stronger condition.

Theorem 3.9. *Let $\mathcal{A}$, B and $\varepsilon > 0$ be given such that there exists $X_\varepsilon \in \mathcal{M}_{\le k}$ satisfying*

$$\|\mathcal{A}[X_\varepsilon] - B\|_{\sigma,2k} \le \varepsilon. \tag{3.2}$$

Let $\mu \ge 0$. Consider a critical point $X \in \mathcal{M}_k$ of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ satisfying

$$\mathcal{H}_X[Z,Z] \ge -\mu\|Z\|_F^2 \quad \text{for all } Z \in T_X\mathcal{M}_k.$$

The following two statements hold.

(i) *If $\mathcal{A}$ satisfies RPD properties such that*

$$\delta_2 \le \delta_{2k} \le \delta_{3k} < -\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{1+\sqrt{2}}{2\sqrt{2}} + \frac{\mu}{2}\right)^2 + \frac{1}{\sqrt{2}}} \tag{3.3}$$

(the first two inequalities pose no restriction), then X satisfies the estimate

$$\|X - X_\varepsilon\|_F \le \left(\sqrt{2} + \frac{1}{1+\delta_2+\mu}\right)\left[\frac{1+\delta_2+\mu}{1-\delta_{2k}-\sqrt{2}(1+\mu)\delta_{3k}-\sqrt{2}\delta_2\delta_{3k}}\right]\cdot\varepsilon. \tag{3.4}$$

(ii) *Alternatively, if*

$$\delta_2 \le \delta_{2k} < -\left(\frac{3}{4} + \frac{\mu}{2}\right) + \sqrt{\left(\frac{3}{4} + \frac{\mu}{2}\right)^2 + \frac{1}{2}} \tag{3.5}$$

(the first inequality is no restriction), then X satisfies the estimate

$$\|X - X_\varepsilon\|_F \le \left(\sqrt{2} + \frac{1}{1+\delta_2+\mu}\right)\left[\frac{1+\delta_2+\mu}{1-(3+2\mu)\delta_{2k}-2\delta_2\delta_{2k}}\right]\cdot\varepsilon.$$

Proof. We assume $X \ne X_\varepsilon$, otherwise there is nothing to show. We consider upper and lower bounds for

$$\alpha = \operatorname{dist}_{\|\cdot\|_{\sigma,k}}(X - (\mathcal{A}[X] - B), \mathcal{M}_{\le k}).$$

Obviously,

$$\alpha \le \|X - \mathcal{A}[X] + B - X_\varepsilon\|_{\sigma,k},$$

and by triangle inequality,

$$\alpha \le \|(\mathsf{Id} - \mathcal{A})[X - X_\varepsilon]\|_{\sigma,k} + \|\mathcal{A}[X_\varepsilon] - B\|_{\sigma,k} \le \|(\mathsf{Id} - \mathcal{A})[X - X_\varepsilon]\|_{\sigma,k} + \varepsilon.$$

Lemmata 2.7 and 2.8 give

$$\|(\mathsf{Id} - \mathcal{A})[X - X_\varepsilon]\|_{\sigma,k} \le \Gamma(\mathsf{Id} - \mathcal{A}, k, 2k)\|X - X_\varepsilon\|_F \le \sqrt{2}\Gamma(\mathsf{Id} - \mathcal{A}, k, k)\|X - X_\varepsilon\|_F.$$

Applying Lemma 3.4 to either of these bounds then results in the two estimates

$$\alpha \leq \delta_{3k}\|X - X_\varepsilon\|_F + \varepsilon \tag{3.6}$$

and

$$\alpha \leq \sqrt{2}\delta_{2k}\|X - X_\varepsilon\|_F + \varepsilon. \tag{3.7}$$

We turn to the lower bound on α provided by Lemma 2.3. Let $\varsigma_1 \geq \cdots \geq \varsigma_k > 0$ denote the singular values of X, and $s_1 \geq \cdots \geq s_{2k}$ the $2k$ largest singular values (some might be zero) of $\mathcal{A}[X] - B$. By the lemma, there exists some $0 \leq j \leq k$ for which

$$\alpha^2 \geq \varsigma_k^2 + \cdots + \varsigma_{k-j+1}^2 + s_{j+1}^2 + \cdots + s_k^2. \tag{3.8}$$

(If $j = 0$ there are no ς_i, while if $j = k$ there are no s_i.) By Proposition 2.9 (with $j = 1$)

$$\varsigma_k \geq \frac{s_1}{1 + \delta_2 + \mu}.$$

Due to $\varsigma_{k-j+1} \geq \cdots \geq \varsigma_k$ and $s_1 \geq \cdots \geq s_j$ (3.8) then entails

$$\alpha \geq \left(\frac{1}{1 + \delta_2 + \mu}\right)\sqrt{s_1^2 + \cdots + s_k^2} \geq \frac{1}{\sqrt{2}}\left(\frac{1}{1 + \delta_2 + \mu}\right)\sqrt{s_1^2 + \cdots + s_{2k}^2}. \tag{3.9}$$

Using (2.4) we can estimate

$$\begin{aligned}
\sqrt{s_1^2 + \cdots + s_{2k}^2} &\geq \left\langle \frac{X - X_\varepsilon}{\|X - X_\varepsilon\|_F}, \mathcal{A}[X] - B \right\rangle_F \\
&= \left\langle \frac{X - X_\varepsilon}{\|X - X_\varepsilon\|_F}, \mathcal{A}[X - X_\varepsilon] \right\rangle_F + \left\langle \frac{X - X_\varepsilon}{\|X - X_\varepsilon\|_F}, \mathcal{A}[X_\varepsilon] - B \right\rangle_F \\
&\geq (1 - \delta_{2k})\|X - X_\varepsilon\|_F - \varepsilon.
\end{aligned}$$

With (3.9) we arrive at the lower bound

$$\alpha \geq \frac{1}{\sqrt{2}}\left(\frac{1}{1 + \delta_2 + \mu}\right)(1 - \delta_{2k})\|X - X_\varepsilon\|_F - \frac{1}{\sqrt{2}}\left(\frac{1}{1 + \delta_2 + \mu}\right)\varepsilon. \tag{3.10}$$

Taken together and multiplying by $\sqrt{2}$ the bounds (3.6) and (3.10) yield the inequality

$$\left[\frac{1 - \delta_{2k}}{1 + \delta_2 + \mu} - \sqrt{2}\delta_{3k}\right]\|X - X_\varepsilon\|_F \leq \left(\sqrt{2} + \frac{1}{1 + \delta_2 + \mu}\right)\varepsilon.$$

Since $\delta_2 \leq \delta_{2k} \leq \delta_{3k}$ the term in brackets on the left side is positive under the given condition (3.3) on δ_{3k} by Lemma 3.8 (here $c = \sqrt{2}$). This leads to the assertion in item (i). Item (ii) is obtained by combining (3.7) and (3.10) instead. $\square$

Remark 3.10. The theorem is formulated for the distances $\|X - X_\varepsilon\|_F$ in Frobenius norm, but in applications the difference in function values $|f_{\mathcal{A},B}(X) - f_{\mathcal{A},B}(X_\varepsilon)|$ may be more relevant, in particular if $f_{\mathcal{A},B}$ takes the form (1.5) of a shifted squared (semi)-norm for the operator $\mathcal{A}$ up to a constant (one may additionally assume $f_{\mathcal{A},B}(X_\varepsilon) = 0$). Using Taylor expansion at X_ε it holds that

$$f_{\mathcal{A},B}(X) - f_{\mathcal{A},B}(X_\varepsilon) = \langle \mathcal{A}[X_\varepsilon] - B, X - X_\varepsilon \rangle_F + \frac{1}{2}\langle X - X_\varepsilon, \mathcal{A}[X - X_\varepsilon] \rangle_F,$$

and thus, under the assumptions of the theorem,

$$|f_{\mathcal{A},B}(X) - f_{\mathcal{A},B}(X_\varepsilon)| \leq \varepsilon\|X - X_\varepsilon\|_F + \frac{1 + \delta_{2k}}{2}\|X - X_\varepsilon\|_F^2.$$

Now the estimates for $\|X - X_\varepsilon\|_F$ from the theorem can be used.

Remark 3.11. Similar to Corollary 3.7, and based on Remark 3.3, the theorem can be generalized to operators whose restricted spectral bounds are not centered around one, but are otherwise well conditioned on $\mathcal{M}_{\leq 2k}$ or $\mathcal{M}_{\leq 3k}$. If the conditions on δ_{2k} or δ_{3k} in Theorem 3.9 are fulfilled for a scaled operator $\omega\mathcal{A}$ where $\omega > 0$ the statement of the theorem remains true for the initial operator $\mathcal{A}$ if one replaces μ and ε by μ/ω and ε/ω, respectively.

3.4 Some concrete bounds

The constants in the estimates on $\|X - X_\varepsilon\|_F$ provided by Theorem 3.9 become arbitrarily large when δ_{3k} and δ_{2k} approach the required upper bounds. Therefore, in order to obtain reasonable estimates one needs smaller values for δ_{3k} and δ_{2k}. To get some intuition on the actual numbers we computed for several values of μ the guaranteed error bounds for $\|X - X_\varepsilon\|_F$ when $\mathcal{A}$ satisfies an RPD property with δ_{3k} or δ_{2k} is 90% or 50% of the critical upper bound (3.3) and (3.5), respectively. These values are presented in Table 1 (for δ_{3k}) and Table 2 (for δ_{2k}), where ε and X_ε are as in the theorem. Clearly, when μ is fixed, smaller RPD constants lead to better estimates.

In the context of so-called *strict saddle point properties* that have been discussed in related work one can spell out these results as follows: for given $\mu > 0$ and assuming δ_{3k} (or δ_{2k}) satisfies the bound asserted in the table, all critical points X of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ either have the asserted distance $\|X - X_\varepsilon\|_F$ to a point $X_\varepsilon \in \mathcal{M}_{\leq k}$ satisfying $\|\mathcal{A}[X_\varepsilon] - B\|_{\sigma,2k} < \varepsilon$, or the Riemannian Hessian at X has at least one negative eigenvalue strictly less than $-\mu$. In particular the rows for $\mu = 0$ in the tables provide bounds on the distance of local minima to the set of all such X_ε.

4 Numerical experiments

We now report on numerical experiments that verify our main result in Theorem 3.9. The experiments confirm that different algorithms indeed find ε-close solutions of the noisy matrix sensing problem, as predicted by theory. Note that the conditions on the RPD constants δ_{2k} or δ_{3k} obtained in this work are only sufficient, and perhaps still rather loose. The influence of these constants is not explored here in detail.

We consider $m = n$ and construct matrix sensing problems involving two types of RPD operators $\mathcal{A}$ on $\mathbb{R}^{n\times n}$. The first construction, called *deterministic*, is of the vectorized form

μ	$\delta_{3k} \leq 0.9 \cdot \delta_{3k}^{\text{crit}}$	$\|X - X_\varepsilon\|_F$	$\delta_{3k} \leq 0.5 \cdot \delta_{3k}^{\text{crit}}$	$\|X - X_\varepsilon\|_F$
0	0.3101	$24.73 \cdot \varepsilon$	0.1723	$4.91 \cdot \varepsilon$
0.2	0.2860	$27.49 \cdot \varepsilon$	0.1589	$5.46 \cdot \varepsilon$
0.4	0.2649	$30.23 \cdot \varepsilon$	0.1472	$6.01 \cdot \varepsilon$
0.6	0.2465	$32.96 \cdot \varepsilon$	0.1369	$6.57 \cdot \varepsilon$
0.8	0.2303	$35.74 \cdot \varepsilon$	0.1279	$7.13 \cdot \varepsilon$
1	0.2159	$38.52 \cdot \varepsilon$	0.1199	$7.69 \cdot \varepsilon$

Table 1: Error bounds for different values of μ. The second column states a sufficient condition on δ_{3k}, taken as nine tenth of the upper bound (3.3), to obtain the estimate on $\|X - X_\varepsilon\|_F$ in the third column (choosing $\delta_2 = \delta_{2k} = \delta_{3k}$ in (3.4)). The fourth and fifth column display the results when δ_{3k} is less than one half of the upper bound.

μ	$\delta_{2k} \leq 0.9 \cdot \delta_{2k}^{\text{crit}}$	$\|X - X_\varepsilon\|_F$	$\delta_{2k} \leq 0.5 \cdot \delta_{2k}^{\text{crit}}$	$\|X - X_\varepsilon\|_F$
0	0.2526	$24.28 \cdot \varepsilon$	0.1403	$4.85 \cdot \varepsilon$
0.2	0.2301	$27.05 \cdot \varepsilon$	0.1278	$5.41 \cdot \varepsilon$
0.4	0.2108	$29.84 \cdot \varepsilon$	0.1171	$5.97 \cdot \varepsilon$
0.6	0.1943	$32.64 \cdot \varepsilon$	0.1079	$6.53 \cdot \varepsilon$
0.8	0.18	$35.45 \cdot \varepsilon$	0.1	$7.1 \cdot \varepsilon$
1	0.1675	$38.27 \cdot \varepsilon$	0.0930	$7.66 \cdot \varepsilon$

Table 2: Same as Table 1 but with conditions on δ_{2k}.

$\mathcal{A} = \mathsf{Id} + \delta \cdot \mathsf{QDQ}^T$ with Q a random orthogonal matrix and D a diagonal matrix with random ± 1 on its diagonal. Such $\mathcal{A}$ will be RPD with constant $\delta_k = \delta$ for all k. The other construction, called *random*, uses the nearly isometric random matrices from [31]. In particular $\mathcal{A} = \mathsf{F}^T\mathsf{F}$ with $\mathsf{F} \in \mathbb{R}^{d \times n^2}$ and F_{ij} random Gaussian $\mathcal{N}(0, 1/d)$. For certain large enough choices of n and d this $\mathcal{A}$ will satisfy any desired RPD property with high probability; see [31]. Due to size limitation, however, we took $n = 50$, $k = 5$, $d = 5nk$ which does not yet correspond to $\delta_{3k} \leq \delta_{3k}^{\text{crit}}$ but still allows us to verify convergence of the algorithms.[3]

In all cases, we generate an 'exact' solution $X^* = GH^T \in \mathcal{M}_k$ with G, H random Gaussian matrices of size $n \times k$. We then compute $B = \mathcal{A}(X^*) + \varepsilon \cdot N$ with N a random Gaussian matrix, scaled so that $\|N\|_F = 1$, and $\varepsilon \geq 0$ a noise factor. This guarantees in particular that $\|\mathcal{A}[X_\varepsilon] - B\|_{\sigma,2k} \leq \varepsilon$ as required in Theorem 3.9. When $\varepsilon > 0$,the global minimizer of $f_{\mathcal{A},B}$ is unknown and we therefore take $X_\varepsilon = X^*$.

The methods that minimize $f_{\mathcal{A},B}$ for rank k matrices are as follows:

- *Embedded SD*: Riemannian steepest descent on $\mathcal{M}_k$ with the embedded submanifold geometry and Euclidean restricted metric. This is the same geometry as in [44, 34, 43, 45].
- *Embedded CG*: Same as Embedded SD but now with nonlinear conjugate gradients. This corresponds to the solver GeomCG from [43] but applied to sensing instead of completion.
- *Quotient SD*: Same as the embedded solver but using the quotient geometry from [26].
- *ALS*: Alternating least squares with QR stabilization of the iterates to avoid ill-conditioning. This appears, for example, in [46, §2.1] where it is called the Nonlinear Gauss–Seidel method.

All methods except ALS were implemented using Manopt [4] with standard options except we used exact line-search to verify theoretical convergence rates. For ALS the exact solution of the least-square system was computed using Cholesky decomposition. Computations were done in MATLAB v2017b using 34 decimal digits[4] to better judge whether the iterates have converged to the global optimum. In the figures we will also display estimated asymptotic convergence rates ρ^ℓ for the iterations $\ell = 1, 2, \ldots$. These were computed from the (Riemannian) Hessian $\mathcal{H}_X$ at the limit point X. In particular, with κ the condition number of $\mathcal{H}_X$ as computed by Manopt, we used

$$\rho_{\text{SD}} = \left(\frac{\kappa - 1}{\kappa + 1}\right)^2, \qquad \rho_{\text{CG}} = \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^2.$$

[3] The Riemannian Hessian at convergence has condition number about 18 which is too large when $\delta_{3k} < \delta_{3k}^{\text{crit}}$.
[4] We used the Advanpix multiprecision toolbox.

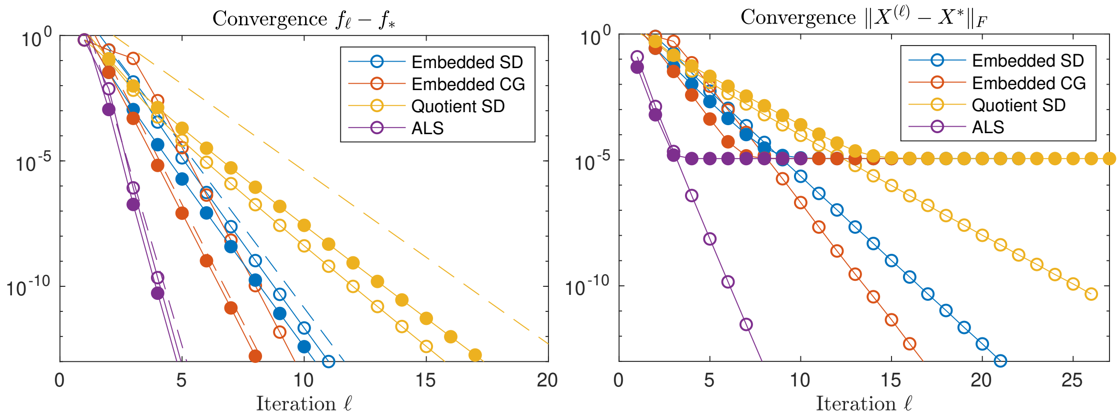

Figure 1: Deterministic $\mathcal{A}$ with $n = 50$, $k = 5$ and $\delta = 0.95 \cdot \delta_{3k}^{\mathrm{crit}}$. Both panels show in open circles for zero noise ($\varepsilon = 0$), and in closed circles for $\varepsilon = 10^{-5}$. Left panel shows in line the asymptotic convergence rate based on spectrum of Hessian.

ε	$\min_{\ell=1}^{500} \|X^{(\ell)} - X_\varepsilon\|_F$		
	max	mean	min
0	$4.021 \cdot 10^{-12}$	$1.058 \cdot 10^{-12}$	$1.190 \cdot 10^{-14}$
10^{-10}	$1.165 \cdot 10^{-10}$	$6.600 \cdot 10^{-11}$	$2.603 \cdot 10^{-11}$
10^{-08}	$1.135 \cdot 10^{-08}$	$9.103 \cdot 10^{-09}$	$2.739 \cdot 10^{-09}$
10^{-06}	$1.131 \cdot 10^{-06}$	$9.269 \cdot 10^{-07}$	$7.029 \cdot 10^{-07}$
10^{-04}	$1.153 \cdot 10^{-04}$	$9.062 \cdot 10^{-05}$	$6.752 \cdot 10^{-05}$
10^{-02}	$1.148 \cdot 10^{-02}$	$9.288 \cdot 10^{-03}$	$6.435 \cdot 10^{-03}$

Table 3: Statistics for 20 random realizations of the deterministic $\mathcal{A}$ operator with $n = 50$, $k = 5$ and $\delta = 0.9 \cdot \delta_{3k}^{\mathrm{crit}}$.

The rate ρ_{SD} of SD corresponds to the Euclidean case with exact line search and is well known. For SD on Riemannian manifolds this rate has been suggested to hold as well; see [41, p. 270].[5] The rate ρ_{CG} is intended for numerical verification and is rigorous for the unconstrained CG method. For ALS the asymptotic rate is computed as the largest eigenvalue of modulus less than one of the linearized iteration matrix at the limit point, which is obtained from a block triangular decomposition of the Hessian of $F(G, H) = f(GH^T)$ as in the general nonlinear Gauss–Seidel method; see [28, §10.3.4-5].

We could have compared to many different solvers, but for simplicity, we have restricted ourselves to mainly Riemannian algorithms since they are cheap per iteration and typically perform very well. In addition many other low-rank optimization methods, like iterative hard thresholding, have the same asymptotic behavior. Another reason was to verify that the Riemannian Hessian used in Theorem 3.9 indeed captures the correct behavior in the convergence plots.

The convergence plots are displayed in Figs 1 and 2. The left panels show convergence of the

[5]The rate without the square is easy to prove; see, e.g., [41, Ch. 7, Thm. 4.3].

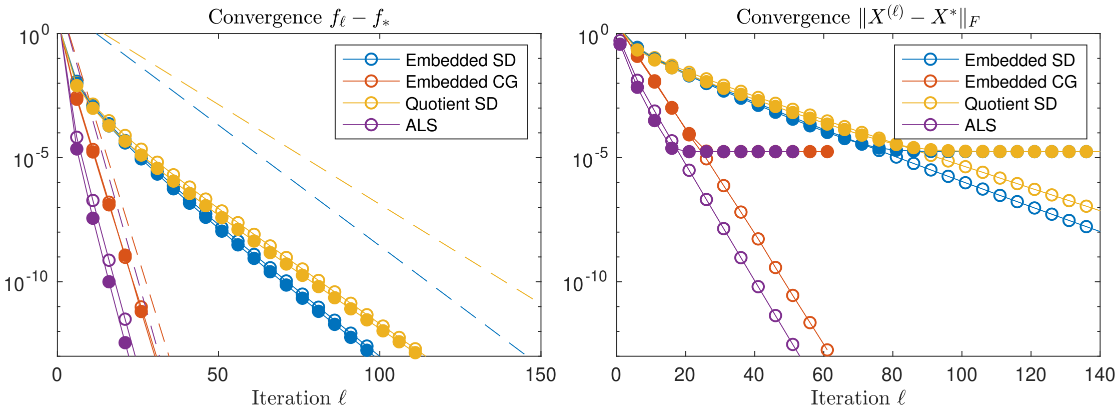

Figure 2: Random Gaussian $\mathcal{A}$ with $n = 50$, $k = 5$, $d = 5nk$. Both panels show in open circles for zero noise ($\varepsilon = 0$), and in closed circles for $\varepsilon = 10^{-5}$ (every five iterations shown). Left panel shows in line the asymptotic convergence rate based on spectrum of the Hessian.

function values and clearly indicate a good correspondence of the theoretical asymptotic rates. The right panel is to verify that the error of the local minima obtained for the noisy problem are on the order of ε, as predicted by Theorem 3.9. This bound is explored in more detail in Table 3 for the deterministic case. Compared to Fig. 1, we continued the iteration for 500 iterations and took the minimal value of $\|X^{(\ell)} - X_\varepsilon\|_F$ as approximation of the limit point of each iteration. Note that Table 1 predicts that the error should be bounded by $24.28 \cdot \varepsilon$, which is always achieved. Indeed, (for $\varepsilon > 0$) we observe a much better constant $1.2 \cdot \varepsilon$ for the final error in this example.

5 Conclusion

We have studied some properties of critical points of quadratic functions on manifolds of fixed-rank matrices. In particular estimates for singular values of local minima have been derived that relate them to the singular values of the gradient at local minima. Then, under certain assumptions on bounds for the Rayleigh quotient of the Hessian on the cones of bounded rank matrices, which generalize the popular RIP conditions for matrix sensing, our estimates imply that there cannot be spurious local minima far away from the global one. In particular local minima are absent in the noiseless case. The required restricted spectral bounds to obtain these results are considerably weaker than in related previous publications.

So far our approach does not cover the important cases of matrix completion or the typical matrix equations in numerical linear algebra, as they do not meet the restricted spectral bounds. However, some of the presented techniques may still be useful when studying these cases as well.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[2] P.-A. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM J. Optim.*, 22(1):135–158, 2012.

[3] S. Bhojanapalli, B. Neyshabur, and N. Srebro. Global optimality of local search for low rank matrix recovery. In *Advances in Neural Information Processing Systems 29*, pages 3873–3881. Curran Associates, Inc., 2016.

[4] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.*, 15:1455–1459, 2014.

[5] N. Boumal, V. Voroninski, and A. S. Bandeira. The non-convex Burer–Monteiro approach works on smooth semidefinite programs. In *Advances in Neural Information Processing Systems 29*, pages 2757–2765. Curran Associates, Inc., 2016.

[6] N. Boumal, V. Voroninski, and A.S. Bandeira. Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs. *Comm. Pure Appl. Math.*, 2019. In press, doi:10.1002/cpa.21830.

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

[8] T. T. Cai and A. Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Appl. Comput. Harmon. Anal.*, 35(1):74–93, 2013.

[9] E. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

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

[11] T. P. Cason, P.-A. Absil, and P. Van Dooren. Iterative methods for low rank approximation of graph similarity matrices. *Linear Algebra Appl.*, 438(4):1863–1882, 2013.

[12] Y. Chi, Y. M. Lu, and Y. Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.*, 67(20):5239–5269, 2019.

[13] A. M. Davenport and J. Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE J. Sel. Topics Signal Process.*, 10(4):608–622, 2016.

[14] R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping from saddle points–online stochastic gradient for tensor decomposition. In *Proceedings of The 28th Conference on Learning Theory*, pages 797–842. PMLR, 2015.

[15] R. Ge, C. Jin, and Y. Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1233–1242. PMLR, 2017.

[16] R. Ge, J. D. Lee, and T. Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems 29*, pages 2973–2981. Curran Associates, Inc., 2016.

[17] J. Harris. *Algebraic geometry. A first course.* Springer-Verlag, New York, 1995. Corrected reprint of the 1992 original.

[18] U. Helmke and M. A. Shayman. Critical points of matrix least squares distance functions. *Linear Algebra Appl.*, 215:1–19, 1995.

[19] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[20] P. Jain, R. Meka, and I. S. Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems 23*, pages 937–945. Curran Associates, Inc., 2010.

[21] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization (extended abstract). In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, pages 665–674. ACM, 2013.

[22] R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *J. Mach. Learn. Res.*, 11:2057–2078, 2010.

[23] Q. Li, Z. Zhu, and G. Tang. Geometry of factored nuclear norm regularization. *arXiv preprint arXiv:1704.01265*, 2017.

[24] Q. Li, Z. Zhu, and G. Tang. The non-convex geometry of low-rank matrix optimization. *Inf. Inference*, 8(1):51–96, 2019.

[25] G. Meyer, S. Bonnabel, and R. Sepulchre. Linear regression under fixed-rank constraints: a Riemannian approach. In *Proceedings of the 28th International Conference on Machine Learning*, pages 545–552. ACM, 2011.

[26] B. Mishra, K. Adithya Apuroop, and R. Sepulchre. A Riemannian geometry for low-rank matrix completion. *arXiv preprint arXiv:1211.1550*, 2012.

[27] B. Mishra, G. Meyer, F. Bach, and R. Sepulchre. Low-rank optimization with trace norm penalty. *SIAM J. Optim.*, 23(4):2124–2149, 2013.

[28] J. M. Ortega and W. C. Reinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[29] D. Park, A. Kyrillidis, C. Carmanis, and S. Sanghavi. Non-square matrix sensing without spurious local minima via the Burer–Monteiro approach. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 65–74. PMLR, 2017.

[30] T. Penzl. Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case. *Systems Control Lett.*, 40(2):139–144, 2000.

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

[32] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[33] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM J. Optim.*, 25(1):622–646, 2015.

[34] U. Shalit, D. Weinshall, and G. Chechik. Online learning in the embedded manifold of low-rank matrices. *J. Mach. Learn. Res.*, 13:429–458, 2012.

[35] V. Simoncini. Computational methods for linear matrix equations. *SIAM Rev.*, 58(3):377–441, 2016.

[36] J. Sun, Q. Qu, and J. Wright. When are nonconvex problems not scary? *arXiv preprint arXiv:1510.06096*, 2015.

[37] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere I: Overview and the geometric picture. *IEEE Trans. Inform. Theory*, 63(2):853–884, 2017.

[38] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere II: Recovery by Riemannian trust-region method. *IEEE Trans. Inform. Theory*, 63(2):885–914, 2017.

[39] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. *Found. Comput. Math.*, 18(5):1131–1198, 2018.

[40] R. Sun and Z.-Q. Luo. Guaranteed matrix completion via nonconvex factorization. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 270–289. IEEE, 2015.

[41] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds.* Kluwer Academic Publishers, 1994.

[42] B. Vandereycken. Low-rank matrix completion by Riemannian optimization (extended version). *arXiv preprint arXiv:1209.3834*, 2012.

[43] B. Vandereycken. Low-rank matrix completion by Riemannian optimization. *SIAM J. Optim.*, 23(2):1214—1236, 2013.

[44] B. Vandereycken and S. Vandewalle. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. *SIAM J. Matrix Anal. Appl.*, 31(5):2553–2579, 2010.

[45] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung. Guarantees of Riemannian optimization for low rank matrix recovery. *SIAM J. Matrix Anal. Appl.*, 37(3):1198–1222, 2016.

[46] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a non-linear successive over-relaxation algorithm. *Math. Program. Comput.*, 4(4):333–361, 2012.

[47] R. Zhang, C. Josz, S. Sojoudi, and J. Lavaei. How much restricted isometry is needed in nonconvex matrix recovery? In *Advances in Neural Information Processing Systems 31*, pages 5586–5597. Curran Associates, Inc., 2018.

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

A Proof of the Riemannian Hessian

Take $X = U\Sigma V^T \in \mathcal{M}_k$ and let $\mathrm{Exp}_X : \mathcal{B} \to \mathcal{M}_k$ be the exponential map defined on a sufficiently small ball $\mathcal{B} \subset T_X\mathcal{M}_k$ around zero. It has been shown in [42, Proposition A1] (see also [2, Proposition 24] for a different proof) that

$$\mathrm{Exp}_X(Z) = X + Z + (I - P_X^{\mathrm{col}})ZV\Sigma^{-1}U^T Z(I - P_X^{\mathrm{row}}) + O(\|Z\|^3). \tag{A.1}$$

The Riemannian Hessian $\mathcal{H}_X \colon T_X\mathcal{M}_k \to T_X\mathcal{M}_k$ of $f_{\mathcal{A},B}$ on $\mathcal{M}_k$ is obtained as the standard (Euclidean) Hessian of the pullback $f \circ \mathrm{Exp}_X$ at 0; see [1, Proposition 5.5.4]. Substituting (A.1) into

$$f_{\mathcal{A},B}(X) = \frac{1}{2}\langle \mathcal{A}[X], X\rangle_F - \langle B, X\rangle_F$$

gives the expansion

$$\begin{aligned} f_{\mathcal{A},B}(\mathrm{Exp}_X(Z)) = f_{\mathcal{A},B}(X) + \langle \mathcal{A}[X] - B, Z\rangle_F + \frac{1}{2}\langle \mathcal{A}[Z], Z\rangle_F \\ + \langle \mathcal{A}[X] - B, (I - P_X^{\mathrm{col}})ZV\Sigma^{-1}U^T Z(I - P_X^{\mathrm{row}})\rangle_F + O(\|Z\|^3). \end{aligned}$$

The third and fourth term on the right-hand side of this expansion are second order in Z. The Riemannian Hessian is therefore given as

$$\mathcal{H}_X[Z, Z] = \langle \mathcal{A}[Z], Z\rangle_F + \langle \mathcal{A}[X] - B, (I - P_X^{\mathrm{col}})ZVS^{-1}U^T Z(I - P_X^{\mathrm{row}})\rangle_F.$$

With the particular choices

$$\bar{G} = U\Sigma^{1/2}, \quad \bar{H} = V\Sigma^{1/2}, \quad Z = \Delta G \cdot \bar{H}^T + \bar{G} \cdot \Delta H^T,$$

we have $(I - P_X^{\mathrm{col}})ZV\Sigma^{-1}U^T Z(I - P_X^{\mathrm{row}}) = (I - P_X^{\mathrm{col}})\Delta G \cdot \Delta H^T(I - P_X^{\mathrm{row}})$. This establishes our expression (2.11) for the Riemannian Hessian.