

# Global Optimality in Low-rank Matrix Optimization

Zihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B. Wakin

**Abstract**—This paper considers the minimization of a general objective function  $f(\mathbf{X})$  over the set of rectangular  $n \times m$  matrices that have rank at most  $r$ . To reduce the computational burden, we factorize the variable  $\mathbf{X}$  into a product of two smaller matrices and optimize over these two matrices instead of  $\mathbf{X}$ . Despite the resulting nonconvexity, recent studies in matrix completion and sensing have shown that the factored problem has no spurious local minima and obeys the so-called strict saddle property (the function has a directional negative curvature at all critical points but local minima). We analyze the global geometry for a general and yet well-conditioned objective function  $f(\mathbf{X})$  whose restricted strong convexity and restricted strong smoothness constants are comparable. In particular, we show that the reformulated objective function has no spurious local minima and obeys the strict saddle property. These geometric properties imply that a number of iterative optimization algorithms (such as gradient descent) can provably solve the factored problem with global convergence.

**Index Terms**—Low-rank matrix optimization, matrix sensing, nonconvex optimization, optimization geometry, strict saddle

## I. INTRODUCTION

Consider the minimization of a general objective function  $f(\mathbf{X})$  over all low-rank  $n \times m$  matrices:

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad f(\mathbf{X}) \\ & \text{subject to } \text{rank}(\mathbf{X}) \leq r, \end{aligned} \quad (1)$$

where the objective function  $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$  is smooth. Low-rank matrix optimizations of the form (1) appear in a wide variety of applications, including quantum tomography [2], [3], collaborative filtering [4], [5], sensor localization [6], low-rank matrix recovery from compressive measurements [7], [8], and matrix completion [9], [10]. Due to the rank constraint, however, low-rank matrix optimizations of the form (1) are highly nonconvex and computationally NP-hard in general [11] even if  $f$  itself is convex. In order to deal with the rank constraint and to find a low-rank solution, the nuclear norm is widely used in matrix inverse problems [8], [12] arising in machine learning [13], signal processing [14], and control [15]. Although nuclear norm minimization enjoys strong statistical guarantees [9], its computational complexity is very high (as most algorithms require performing an expensive singular value decomposition (SVD) in each iteration), prohibiting it from scaling to practical problems.

This work was supported by NSF grant CCF-1409261, NSF grant CCF-1704204, NSF CAREER grant CCF-1149225, and the DARPA Lagrange Program under ONR/SPAWAR contract N660011824020. A preliminary version of this paper highlighting some of the key results also appeared in [1], compared to which the current paper includes the formal proofs, two more stylized applications and extensive experimental validation.

Z. Zhu, Q. Li, G. Tang, and M. B. Wakin are with the Department of Electrical Engineering, Colorado School of Mines, Golden, CO 80401 USA. Email: {zzhu, quli, gtang, mwakin}@mines.edu.

To relieve the computational bottleneck and provide an alternative way of dealing with the rank constraint, recent studies propose to factorize the variable into the Burer-Monteiro type decomposition [16], [17] with  $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ , and optimize over the  $n \times r$  and  $m \times r$  matrices  $\mathbf{U}$  and  $\mathbf{V}$ . With this parameterization of  $\mathbf{X}$ , we can recast (1) into the following program:

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}}{\text{minimize}} \quad h(\mathbf{U}, \mathbf{V}) := f(\mathbf{U}\mathbf{V}^T). \quad (2)$$

The bilinear nature of the parameterization renders the objective function of (2) nonconvex even when  $f(\mathbf{X})$  is a convex function. Hence, the objective function in (2) can potentially have spurious local minima (i.e., local minimizers that are not global minimizers) or “bad” saddle points that prevent a number of iterative algorithms from converging to the global solution. By analyzing the landscape of nonconvex functions, several recent works have shown that the factored objective function  $h(\mathbf{U}, \mathbf{V})$  in certain matrix inverse problems has no spurious local minima [18]–[20].

We generalize this line of work by focusing on a general objective function  $f(\mathbf{X})$  in the optimization (1), not necessarily a quadratic loss function coming from a matrix inverse problem. By focusing on a general objective function, we attempt to provide a unifying framework for low-rank matrix optimizations with the factorization approach. We provide a geometric analysis for the factored program (2) and show that, under certain conditions on  $f(\mathbf{X})$ , all critical points of the objective function  $h(\mathbf{U}, \mathbf{V})$  are well-behaved. Our characterization of the geometry of the objective function ensures that a number of iterative optimization algorithms converge to a global minimum.

## A. Summary of Results

The purpose of this paper is to analyze the geometry of the factored problem  $h(\mathbf{U}, \mathbf{V})$  in (2). In particular, we attempt to understand the behavior of all of the critical points of the objective function in the reformulated problem (2).

Before presenting our main results, we lay out the necessary assumptions on the objective function  $f(\mathbf{X})$ . As is known, without any assumptions on the problem, even minimizing traditional quadratic objective functions is challenging. For this purpose, we focus on the model where  $f(\mathbf{X})$  is  $(2r, 4r)$ -restricted strongly convex and smooth, i.e., for any  $n \times m$  matrices  $\mathbf{X}, \mathbf{G}$  with  $\text{rank}(\mathbf{X}) \leq 2r$  and  $\text{rank}(\mathbf{G}) \leq 4r$ , the Hessian of  $f(\mathbf{X})$  satisfies

$$\alpha \|\mathbf{G}\|_F^2 \leq [\nabla^2 f(\mathbf{X})](\mathbf{G}, \mathbf{G}) \leq \beta \|\mathbf{G}\|_F^2 \quad (3)$$

for some positive  $\alpha$  and  $\beta$ . A similar assumption is also utilized in [21, Conditions 5.3 and 5.4]. With this assumption

on  $f(\mathbf{X})$ , we summarize our main results in the following informal theorem.

**Theorem 1.** (informal) *Suppose the function  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) and has a critical point  $\mathbf{X}^* \in \mathbb{R}^{n \times m}$  with  $\text{rank}(\mathbf{X}^*) = r^* \leq r$ . Then the factored objective function  $h(\mathbf{U}, \mathbf{V})$  (with an additional regularizer; see Theorem 3) in (2) has no spurious local minima and obeys the strict saddle property (see Definition 3 in Section II).*

*Remark 1.* As guaranteed by Proposition 1 (in Section III), the  $(2r, 4r)$ -restricted strong convexity and smoothness property (3) ensures that  $\mathbf{X}^*$  is the unique global minimum of (1). Theorem 1 then implies that we can recover the rank- $r^*$  global minimizer  $\mathbf{X}^*$  of (1) by many iterative algorithms (such as the trust region method [22] and stochastic gradient descent [23]) even from a random initialization. This is because 1) as guaranteed by Theorem 2, the strict saddle property ensures local search algorithms converge to a local minimum, and 2) there are no spurious local minima.

*Remark 2.* Since our main result only requires the  $(2r, 4r)$ -restricted strong convexity and smoothness property (3), aside from low-rank matrix recovery [12], it can also be applied to many other low-rank matrix optimization problems [24] which do not necessarily involve quadratic loss functions. Typical examples include robust PCA [25], [26], 1-bit matrix completion [27], [28] and Poisson principal component analysis (PCA) [29].

*Remark 3.* Similar results on positive semi-definite (PSD) matrix optimization problems (but without the rank constraint) with generic objective functions were obtained in [30]. We note that one cannot directly apply the results in [30] to the optimization (1) when the matrices under consideration are nonsymmetric or rectangular, even if we ignore the rank constraint. One could attempt to convert minimizing  $f(\mathbf{X})$  over general  $n \times m$  matrices into minimizing  $q(\mathbf{Z})$  over the cone of PSD matrices of size  $(m+n) \times (m+n)$ , where  $\mathbf{X}$  and  $\mathbf{X}^T$  form the upper right and lower left blocks of  $\mathbf{Z}$ . The problem with this transformation, however, is that  $q(\mathbf{Z})$  will no longer satisfy the same properties as  $f(\mathbf{X})$ , in particular the restricted strong convexity and smoothness condition (3) which is a key assumption utilized in [30]. For this reason, one cannot apply the results for the PSD optimization in [30] directly to our problem. In terms of the proof techniques, although the generalization from the PSD case might not seem technically challenging at first sight, quite a few technical difficulties had to be overcome to develop the theory for the general case in this paper. In fact, the non-triviality of extending to the nonsymmetric case is also highlighted in [20], [31].

## B. Related Works

Compared with the original program (1), the factored form (2) typically involves many fewer variables (or variables with much smaller size) and can be efficiently solved by simple but powerful methods (such as gradient descent [23], [32], the trust region method [33], and alternating methods [34]) for

large-scale settings, though it is nonconvex. In recent years, tremendous effort has been devoted to analyzing nonconvex optimizations by exploiting the geometry of the corresponding objective functions. These works can be separated into two types based on whether the geometry is analysed locally or globally. One type of work analyzes the behavior of the objective function in a small neighborhood containing the global optimum and requires a good initialization that is close enough to a global minimum. Problems such as phase retrieval [35], matrix sensing [31], and semi-definite optimization [36] have been studied.

Another type of work attempts to analyze the landscape of the objective function and show that it obeys the strict saddle property. If this particular property holds, then simple algorithms such as gradient descent and the trust region method are guaranteed to converge to a local minimum from a random initialization [23], [32], [37] rather than requiring a good guess. We approach low-rank matrix optimization with general objective functions (1) via a similar geometric characterization. Similar geometric results are known for a number of problems including complete dictionary learning [37], phase retrieval [22], orthogonal tensor decomposition [23], and matrix inverse problems [18], [19], [30]. Empirical evidence also supports using the factorization approach for estimating a low-rank PSD matrix from a set of rank-one measurements corrupted by arbitrary outliers [38] and for recovering a dynamically evolving low-rank matrix from incomplete observations [39].

Our work is most closely related to certain recent works in low-rank matrix optimization. Bhojanapalli et al. [18] showed that the low-rank, PSD matrix sensing problem has no spurious local minima and obeys the strict saddle property. Similar results were exploited for PSD matrix completion [19], PSD matrix factorization [40] and low-rank, PSD matrix optimization problems with generic objective functions [30]. Our work extends this line of analysis to general low-rank matrix (not necessary PSD or even square) optimization problems. Another closely related work considers the low-rank, non-square matrix sensing problem and matrix completion with the factorization approach [20], [41], [42]. We note that our general objective function framework includes the low-rank matrix sensing problem as a special case (see Section III-C). Furthermore, our result covers both over-parameterization where  $r > r^*$  and exact parameterization where  $r = r^*$ . Wang et al. [21] also considered the factored low-rank matrix minimization problem with a general objective function which satisfies the restricted strong convexity and smoothness condition. Their algorithms require good initializations for global convergence since they characterized only the local landscapes around the global optima. By categorizing the behavior of all the critical points, our work differs from [21] in that we instead characterize the global landscape of the factored objective function.

This paper continues in Section II with formal definitions for strict saddles and the strict saddle property. We present the main results and their implications in matrix sensing, weighted low-rank approximation, and 1-bit matrix completion in Section III. The proof of our main results is given in

Section IV. We conclude the paper in Section VI.

## II. PRELIMINARIES

### A. Notation

To begin, we first briefly introduce some notation used throughout the paper. The symbols  $\mathbf{I}$  and  $\mathbf{0}$  respectively represent the identity matrix and zero matrix with appropriate sizes. The set of  $r \times r$  orthonormal matrices is denoted by  $\mathcal{O}_r := \{\mathbf{R} \in \mathbb{R}^{r \times r} : \mathbf{R}^T \mathbf{R} = \mathbf{I}\}$ . If a function  $h(\mathbf{U}, \mathbf{V})$  has two arguments,  $\mathbf{U} \in \mathbb{R}^{n \times r}$  and  $\mathbf{V} \in \mathbb{R}^{m \times r}$ , we occasionally use the notation  $h(\mathbf{W})$  when we put these two arguments into a new one as  $\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$ . For a scalar function  $f(\mathbf{Z})$  with a matrix variable  $\mathbf{Z} \in \mathbb{R}^{n \times m}$ , its gradient is an  $n \times m$  matrix whose  $(i, j)$ -th entry is  $[\nabla f(\mathbf{Z})]_{ij} = \frac{\partial f(\mathbf{Z})}{\partial Z_{ij}}$  for all  $i \in [n], j \in [m]$ . Here  $[n] = \{1, 2, \dots, n\}$  for any  $n \in \mathbb{N}$  and  $Z_{ij}$  is the  $(i, j)$ -th entry of the matrix  $\mathbf{Z}$ . The Hessian of  $f(\mathbf{Z})$  can be viewed as an  $nm \times nm$  matrix  $[\nabla^2 f(\mathbf{Z})]_{ij} = \frac{\partial^2 f(\mathbf{Z})}{\partial z_i \partial z_j}$  for all  $i, j \in [nm]$ , where  $z_i$  is the  $i$ -th entry of the vectorization of  $\mathbf{Z}$ . An alternative way to represent the Hessian is by a bilinear form defined via  $[\nabla^2 f(\mathbf{Z})](\mathbf{A}, \mathbf{B}) = \sum_{i,j,k,l} \frac{\partial^2 f(\mathbf{Z})}{\partial Z_{ij} \partial Z_{kl}} A_{ij} B_{kl}$  for any  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ . The bilinear form for the Hessian is widely utilized through the paper.

### B. Strict Saddle Property

Suppose  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  is a twice continuously differentiable objective function. We begin with the notion of strict saddles and the strict saddle property.

**Definition 1** (Critical points). *We say  $\mathbf{x}$  a critical point if the gradient at  $\mathbf{x}$  vanishes, i.e.,  $\nabla h(\mathbf{x}) = \mathbf{0}$ .*

**Definition 2** (Strict saddles). *A critical point  $\mathbf{x}$  is a strict saddle if the Hessian matrix evaluated at this point has a strictly negative eigenvalue, i.e.,  $\lambda_{\min}(\nabla^2 h(\mathbf{x})) < 0$ .*

**Definition 3** (Strict saddle property [23]). *A twice differentiable function satisfies the strict saddle property if each critical point either corresponds to a local minimum or is a strict saddle.*

Intuitively, the strict saddle property requires a function to have a directional negative curvature at all critical points but local minima. This property allows a number of iterative algorithms such as noisy gradient descent [23] and the trust region method [43] to further decrease the function value at all the strict saddles and thus converge to a local minimum.

**Theorem 2.** [23], [32], [33] (informal) *For a twice continuously differentiable objective function satisfying the strict saddle property, a number of iterative optimization algorithms (such as gradient descent and the the trust region method) can find a local minimum.*

## III. PROBLEM FORMULATION AND MAIN RESULTS

### A. Problem Formulation

This paper considers the problem (1) of minimizing a general function  $f(\mathbf{X})$  (over the set of low-rank matrices)

which is assumed to have a low-rank critical point  $\mathbf{X}^*$  with  $\text{rank}(\mathbf{X}^*) = r^* \leq r$  such that  $\nabla f(\mathbf{X}^*) = \mathbf{0}$ . Because of the restricted strong convexity and smoothness condition (3), the following result establishes that if  $f(\mathbf{X})$  has a critical point  $\mathbf{X}^*$  with  $\text{rank}(\mathbf{X}^*) \leq r$ , then it is the unique global minimum of (1).

**Proposition 1.** *Suppose  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) with positive  $\alpha$  and  $\beta$ . Assume  $\mathbf{X}^*$  is a critical point of  $f(\mathbf{X})$  with  $\text{rank}(\mathbf{X}^*) = r^* \leq r$ . Then  $\mathbf{X}^*$  is the global minimum of (1), i.e.,*

$$f(\mathbf{X}^*) \leq f(\mathbf{X}), \forall \mathbf{X} \in \mathbb{R}^{n \times m}, \text{rank}(\mathbf{X}) \leq r$$

and the equality holds only at  $\mathbf{X} = \mathbf{X}^*$ .

*Proof of Proposition 1.* First note that if  $\mathbf{X}^*$  is a critical point of  $f(\mathbf{X})$ , then

$$\nabla f(\mathbf{X}^*) = \mathbf{0}.$$

Now for any  $\mathbf{X} \in \mathbb{R}^{n \times m}$  with  $\text{rank}(\mathbf{X}) \leq r$ , the second order Taylor expansion gives

$$\begin{aligned} f(\mathbf{X}) &= f(\mathbf{X}^*) + \langle \nabla f(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^* \rangle \\ &\quad + \frac{1}{2} [\nabla^2 f(\widetilde{\mathbf{X}})](\mathbf{X} - \mathbf{X}^*, \mathbf{X} - \mathbf{X}^*), \end{aligned}$$

where  $\widetilde{\mathbf{X}} = t\mathbf{X}^* + (1-t)\mathbf{X}$  for some  $t \in [0, 1]$ . This Taylor expansion together with  $\nabla f(\mathbf{X}^*) = \mathbf{0}$  and (3) (both  $\widetilde{\mathbf{X}}$  and  $\mathbf{X} - \mathbf{X}^*$  have rank at most  $2r$ ) gives

$$\begin{aligned} f(\mathbf{X}) - f(\mathbf{X}^*) &= \frac{1}{2} [\nabla^2 f(\widetilde{\mathbf{X}})](\mathbf{X} - \mathbf{X}^*, \mathbf{X} - \mathbf{X}^*) \\ &\geq \frac{\alpha}{2} \|\mathbf{X} - \mathbf{X}^*\|_F^2. \end{aligned}$$

□

With this, in the sequel, we use  $\mathbf{X}^*$  to denote the global minimum of (1) (i.e., the low-rank critical point of  $f(\mathbf{X})$ ), unless stated otherwise. We note that the assumption of the existence of a low-rank critical point  $\mathbf{X}^*$  is very mild and holds in many matrix inverse problems [8], [9], where the unknown matrix to be recovered is a critical point of  $f$ . We factorize the variable  $\mathbf{X} = \mathbf{U}\mathbf{V}^T$  with  $\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}$  and transform (1) into its factored counterpart (2). Throughout the paper,  $\mathbf{X}, \mathbf{W}$  and  $\widehat{\mathbf{W}}$  are matrices depending on  $\mathbf{U}$  and  $\mathbf{V}$ :

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}, \quad \widehat{\mathbf{W}} = \begin{bmatrix} \mathbf{U} \\ -\mathbf{V} \end{bmatrix}, \quad \mathbf{X} = \mathbf{U}\mathbf{V}^T.$$

Although the new variable  $\mathbf{W}$  has much smaller size than  $\mathbf{X}$  when  $r \ll \min\{n, m\}$ , the objective function in the factored problem (2) may have a much more complicated landscape due to the bilinear form about  $\mathbf{U}$  and  $\mathbf{V}$ . The reformulated objective function  $h(\mathbf{U}, \mathbf{V})$  could introduce spurious local minima or degenerate saddle points even when  $f(\mathbf{X})$  is convex. Our goal is to guarantee that this does not happen.

Let  $\mathbf{X}^* = \mathbf{Q}_{\mathbf{U}^*} \Sigma^* \mathbf{Q}_{\mathbf{V}^*}^T$  denote an SVD of  $\mathbf{X}^*$ , where  $\mathbf{Q}_{\mathbf{U}^*} \in \mathbb{R}^{n \times r}$  and  $\mathbf{Q}_{\mathbf{V}^*} \in \mathbb{R}^{m \times r}$  are orthonormal matrices of appropriate sizes, and  $\Sigma^* \in \mathbb{R}^{r \times r}$  is a diagonal matrix with

non-negative diagonal (but with some zeros on the diagonal if  $r > r^* = \text{rank}(\mathbf{X}^*)$ ). We denote

$$\mathbf{U}^* = \mathbf{Q}_{\mathbf{U}^*} \boldsymbol{\Sigma}^{*1/2}, \quad \mathbf{V}^* = \mathbf{Q}_{\mathbf{V}^*} \boldsymbol{\Sigma}^{*1/2},$$

where  $\mathbf{X}^* = \mathbf{U}^* \mathbf{V}^{*\top}$  forms a balanced factorization of  $\mathbf{X}^*$  since  $\mathbf{U}^*$  and  $\mathbf{V}^*$  have the same singular values. Throughout the paper, we utilize the following two ways to stack  $\mathbf{U}^*$  and  $\mathbf{V}^*$  together:

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{U}^* \\ \mathbf{V}^* \end{bmatrix}, \quad \widehat{\mathbf{W}}^* = \begin{bmatrix} \mathbf{U}^* \\ -\mathbf{V}^* \end{bmatrix}.$$

Before moving on, we note that for any solution  $(\mathbf{U}, \mathbf{V})$  to (2),  $(\mathbf{U}\mathbf{\Psi}, \mathbf{V}\mathbf{\Phi})$  is also a solution to (2) for any  $\mathbf{\Psi}, \mathbf{\Phi} \in \mathbb{R}^{r \times r}$  such that  $\mathbf{U}\mathbf{\Psi}\mathbf{\Phi}^{\top}\mathbf{V}^{\top} = \mathbf{U}\mathbf{V}^{\top}$ . In order to address this ambiguity (i.e., to reduce the search space of  $\mathbf{W}$  for (2)), we utilize the trick in [20], [21], [31] by introducing a regularizer

$$g(\mathbf{U}, \mathbf{V}) = \frac{\mu}{4} \left\| \mathbf{U}^{\top}\mathbf{U} - \mathbf{V}^{\top}\mathbf{V} \right\|_F^2 \quad (4)$$

and solving the following problem

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{m \times r}}{\text{minimize}} \rho(\mathbf{U}, \mathbf{V}) := f(\mathbf{U}\mathbf{V}^{\top}) + g(\mathbf{U}, \mathbf{V}), \quad (5)$$

where  $\mu > 0$  controls the weight for the term  $\left\| \mathbf{U}^{\top}\mathbf{U} - \mathbf{V}^{\top}\mathbf{V} \right\|_F^2$ , which will be discussed soon.

We remark that  $\mathbf{W}^*$  is still a global minimizer of the factored problem (5) since  $f(\mathbf{X})$  achieves its global minimum over the low-rank set of matrices at  $\mathbf{X}^*$  and  $g(\mathbf{W})$  also achieves its global minimum at  $\mathbf{W}^*$ . The regularizer  $g(\mathbf{W})$  is applied to force the difference between the two Gram matrices of  $\mathbf{U}$  and  $\mathbf{V}$  to be as small as possible. The global minimum of  $g(\mathbf{W})$  is 0, which is achieved when  $\mathbf{U}$  and  $\mathbf{V}$  have the same Gram matrices, i.e., when  $\mathbf{W}$  belongs to

$$\mathcal{E} := \left\{ \mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} : \mathbf{U}^{\top}\mathbf{U} - \mathbf{V}^{\top}\mathbf{V} = \mathbf{0} \right\}. \quad (6)$$

Informally, we can view (5) as finding a point from  $\mathcal{E}$  that also minimizes  $f(\mathbf{U}\mathbf{V}^{\top})$ . This is formally established in Theorem 3.

## B. Main Results

Our main argument is that, under certain conditions on  $f(\mathbf{X})$ , the objective function  $\rho(\mathbf{W})$  has no spurious local minima and satisfies the strict saddle property. This is equivalent to categorizing all the critical points into two types: 1) the global minima which correspond to the global solution of the original convex problem (1) and 2) strict saddles such that the Hessian matrix  $\nabla^2 \rho(\mathbf{W})$  evaluated at these points has a strictly negative eigenvalue. We formally establish this in the following theorem, whose proof is given in the next section.

**Theorem 3.** For any  $\mu > 0$ , each critical point  $\mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$  of  $\rho(\mathbf{W})$  defined in (5) satisfies

$$\mathbf{U}^{\top}\mathbf{U} - \mathbf{V}^{\top}\mathbf{V} = \mathbf{0}. \quad (7)$$

Furthermore, suppose that the function  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition

(3) with positive constants  $\alpha$  and  $\beta$  satisfying  $\frac{\beta}{\alpha} \leq 1.5$  and that the function  $f(\mathbf{X})$  has a critical point  $\mathbf{X}^* \in \mathbb{R}^{n \times m}$  with  $\text{rank}(\mathbf{X}^*) = r^* \leq r$ . Set  $\mu \leq \frac{\alpha}{16}$  for the factored problem (5). Then  $\rho(\mathbf{W})$  has no spurious local minima, i.e., any local minimum of  $\rho(\mathbf{W})$  is a global minimum corresponding to the global solution of the original problem (1):  $\mathbf{U}\mathbf{V}^{\top} = \mathbf{X}^*$ . In addition,  $\rho(\mathbf{W})$  obeys the strict saddle property that any critical point not being a local minimum is a strict saddle with

$$\begin{aligned} \lambda_{\min}(\nabla^2(\rho(\mathbf{W}))) \leq & \\ & -0.08\alpha\sigma_r(\mathbf{X}^*), & r = r^* \\ & -0.05\alpha \cdot \min\{\sigma_{r^c}^2(\mathbf{W}), 2\sigma_{r^*}(\mathbf{X}^*)\}, & r > r^* \\ & -0.1\alpha\sigma_{r^*}(\mathbf{X}^*), & r^c = 0, \end{aligned} \quad (8)$$

where  $r^c \leq r$  is the rank of  $\mathbf{W}$ ,  $\lambda_{\min}(\cdot)$  represents the smallest eigenvalue, and  $\sigma_{\ell}(\cdot)$  denotes the  $\ell$ -th largest singular value.

**Remark 4.** Equation (7) shows that any critical point  $\mathbf{W}$  belongs to  $\mathcal{E}$  for the objective function in the factored problem (5) with any positive  $\mu$ . This demonstrates the reason for adding the regularizer  $g(\mathbf{U}, \mathbf{V})$ . Thus, any iterative optimization algorithm converging to some critical point of  $\rho(\mathbf{W})$  results in a solution within  $\mathcal{E}$ . Furthermore, the strict saddle property along with the lack of spurious local minima ensures that a number of iterative optimization algorithms find the global minimum.

**Remark 5.** For any critical point  $\mathbf{W} \in \mathbb{R}^{(n+m) \times r}$  that is not a local minimum, the right hand side of (8) is strictly negative, implying that  $\mathbf{W}$  is a strict saddle. We also note that Theorem 3 not only covers exact parameterization where  $r = r^*$ , but also includes the over-parameterization case where  $r > r^*$ .

**Remark 6.** The constants appearing in Theorem 3 are not optimized. We use  $\mu \leq \frac{1}{16}\alpha$  simply to include  $\mu = \frac{1}{16}$  which is utilized for the matrix sensing problem in [31]. If the ratio between the restricted strong convexity and smoothness constants  $\frac{\beta}{\alpha} \leq 1.4$ , then we can show that  $\rho(\mathbf{W})$  has no spurious local minima and obeys the strict saddle property for any  $\mu \leq \frac{1}{4}\alpha$  (where  $\mu = \frac{1}{4}$  is utilized for the matrix sensing problem in [20]). In all cases, a smaller  $\mu$  yields a more negative constant in (8); see Section IV for more discussion on this. This implies that when the restricted strong convexity constant  $\alpha$  is not provided a priori, one can always choose a small  $\mu$  to ensure the strict saddle property holds, and hence guarantee the global convergence of many iterative optimization algorithms.

The constant 1.5 for the dynamic range  $\frac{\beta}{\alpha}$  in Theorem 3 is also not optimized and it is possible to slightly relax this constraint with more sophisticated analysis. However, the following example involving weighted symmetric matrix factorization implies that the room for improving this constant is rather limited. Let

$$\mathbf{\Omega} = \begin{bmatrix} \sqrt{1+a} & 1 \\ 1 & \sqrt{1+a} \end{bmatrix}$$

for some  $a \geq 0$ ,

$$\mathbf{X}^* = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \text{ and } \mathbf{U} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Now consider the following weighted low-rank matrix factorization:

$$\begin{aligned} h(\mathbf{U}) &= \frac{1}{2} \|\Omega \odot (\mathbf{U}\mathbf{U}^T - \mathbf{X}^*)\|_F^2 \\ &= \frac{1+a}{2} (x^2 - 1)^2 + \frac{1+a}{2} (y^2 - 1)^2 + (xy - 1)^2, \end{aligned} \quad (9)$$

whose gradient  $\nabla h(\mathbf{U})$  and Hessian  $\nabla^2 h(\mathbf{U})$  are given by:

$$\nabla h(\mathbf{U}) = 2 \begin{bmatrix} (a+1)(x^2-1)x + y(xy-1) \\ (a+1)(y^2-1)y + x(xy-1) \end{bmatrix},$$

and

$$\begin{aligned} \nabla^2 h(\mathbf{U}) &= \\ &2 \begin{bmatrix} y^2 + (3x^2 - 1)(a+1) & 2xy - 1 \\ 2xy - 1 & x^2 + (3y^2 - 1)(a+1) \end{bmatrix}. \end{aligned}$$

Then,

$$\mathbf{U} = \begin{bmatrix} \sqrt{\frac{a}{a+2}} \\ -\sqrt{\frac{a}{a+2}} \end{bmatrix}$$

is a critical point with

$$\nabla^2 h(\mathbf{U}) = \begin{bmatrix} 4a + \frac{8}{a+2} - 6 & \frac{8}{a+2} - 6 \\ \frac{8}{a+2} - 6 & 4a + \frac{8}{a+2} - 6 \end{bmatrix},$$

which has eigenvalues

$$\lambda_1 = \frac{4(a-2)(a+1)}{a+2} \begin{cases} < 0, & a \in [0, 2), \\ > 0, & a > 2, \end{cases}$$

and  $\lambda_2 = 4a > 0$ . We conclude that this  $\mathbf{U}$  is a strict saddle point when  $a < 2$  and a spurious local minimum when  $a > 2$ . This weighted symmetric matrix factorization problem (9) satisfies the restricted strong convexity and smoothness condition (3) with constants  $\alpha = \|\Omega\|_{\min}^2 = 1$  and  $\beta = \|\Omega\|_{\max}^2 = 1+a$  (where  $\|\Omega\|_{\min}$  and  $\|\Omega\|_{\max}$  represent the smallest and largest entries in  $\Omega$ ; see Section III-C). Thus, we have a counter example which demonstrates the existence of spurious local minima when  $\frac{\beta}{\alpha} > 3$ .

*Remark 7.* We finally remark that although Theorem 3 requires the additional regularizer (4), empirical evidence (see experiments in Section V) shows we can get rid of this regularizer for many iterative algorithms with random initialization.

We prove Theorem 3 in Section IV. Before proceeding, we present two stylized applications of Theorem 3 in matrix sensing and weighted low-rank approximation.

### C. Stylized Applications

1) *Matrix Sensing:* We first consider the implication of Theorem 3 in the matrix sensing problem where

$$f(\mathbf{X}) = \frac{1}{2} \|\mathcal{A}(\mathbf{X} - \mathbf{X}^*)\|_F^2.$$

Here  $\mathcal{A} : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^p$  is a known measurement operator satisfying the following restricted isometry property.

**Definition 4.** (Restricted Isometry Property (RIP) [8]) The map  $\mathcal{A} : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^p$  satisfies the  $r$ -RIP with constant  $\delta_r$  if

$$(1 - \delta_r) \|\mathbf{X}\|_F^2 \leq \|\mathcal{A}(\mathbf{X})\|^2 \leq (1 + \delta_r) \|\mathbf{X}\|_F^2 \quad (10)$$

holds for any  $n \times m$  matrix  $\mathbf{X}$  with  $\text{rank}(\mathbf{X}) \leq r$ .

Note that, in this case, the gradient of  $f(\mathbf{X})$  at  $\mathbf{X}^*$  is

$$\nabla f(\mathbf{X}^*) = \mathcal{A}^* \mathcal{A}(\mathbf{X}^* - \mathbf{X}^*) = \mathbf{0},$$

which implies that  $\mathbf{X}^*$  is a critical point of  $f(\mathbf{X})$ . The Hessian quadrature form  $\nabla^2 f(\mathbf{X})[\mathbf{Y}, \mathbf{Y}]$  for any  $n \times m$  matrices  $\mathbf{X}$  and  $\mathbf{Y}$  is given by

$$\nabla^2 f(\mathbf{X})[\mathbf{Y}, \mathbf{Y}] = \|\mathcal{A}(\mathbf{Y})\|^2.$$

If  $\mathcal{A}$  satisfies the  $4r$ -restricted isometry property with constant  $\delta_{4r}$ , then  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) with constants  $\alpha = 1 - \delta_{4r}$  and  $\beta = 1 - \delta_{4r}$  since

$$(1 - \delta_{4r}) \|\mathbf{Y}\|_F^2 \leq \|\mathcal{A}(\mathbf{Y})\|^2 \leq (1 + \delta_{4r}) \|\mathbf{Y}\|_F^2$$

for any rank- $4r$  matrix  $\mathbf{Y}$ . Now, applying Theorem 3, we can characterize the geometry for the following matrix sensing problem with the factorization approach:

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{n \times r}}{\text{minimize}} \frac{1}{2} \|\mathcal{A}(\mathbf{U}\mathbf{V}^T - \mathbf{X}^*)\|_F^2 + g(\mathbf{U}, \mathbf{V}), \quad (11)$$

where  $g(\mathbf{U}, \mathbf{V})$  is the added regularizer defined in (4).

**Corollary 1.** Suppose  $\mathcal{A}$  satisfies the  $4r$ -RIP with constant  $\delta_{4r} \leq \frac{1}{5}$ , and set  $\mu \leq \frac{1-\delta_{4r}}{16}$ . Then the objective function in (11) has no spurious local minima and satisfies the strict saddle property.

This result follows directly from Theorem 3 by noting that  $\frac{\beta}{\alpha} = \frac{1+\delta_{4r}}{1-\delta_{4r}} \leq 1.5$  if  $\delta_{4r} \leq \frac{1}{5}$ . We remark that Park et al. [20, Theorem 4.3] provided a similar geometric result for (11). Compared to their result which requires  $\delta_{4r} \leq \frac{1}{100}$ , our result has a much weaker requirement on the RIP of the measurement operator.

2) *Weighted Low-Rank Matrix Factorization:* We now consider the implication of Theorem 3 in the weighted matrix factorization problem [44], where

$$f(\mathbf{X}) := \frac{1}{2} \|\Omega \odot (\mathbf{X} - \mathbf{X}^*)\|_F^2.$$

Here  $\Omega$  is an  $n \times m$  weight matrix consisting of positive elements and  $\odot$  denotes the point-wise product between two matrices. In this case, the gradient of  $f(\mathbf{X})$  at  $\mathbf{X}^*$  is

$$\nabla f(\mathbf{X}^*) = \Omega \odot \Omega \odot (\mathbf{X}^* - \mathbf{X}^*) = \mathbf{0},$$

which implies that  $\mathbf{X}^*$  is a critical point of  $f(\mathbf{X})$ . The Hessian quadrature form  $\nabla^2 f(\mathbf{X})[\mathbf{Y}, \mathbf{Y}]$  for any  $n \times m$  matrices  $\mathbf{X}$  and  $\mathbf{Y}$  is given by

$$\nabla^2 f(\mathbf{X})[\mathbf{Y}, \mathbf{Y}] = \|\Omega \odot \mathbf{Y}\|_F^2.$$

Thus  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) with constants  $\alpha = \|\Omega\|_{\min}^2$  and  $\beta = \|\Omega\|_{\max}^2$  since

$$\|\Omega\|_{\min}^2 \|\mathbf{Y}\|_F^2 \leq \|\Omega \odot \mathbf{Y}\|_F^2 \leq \|\Omega\|_{\max}^2 \|\mathbf{Y}\|_F^2,$$

where  $\|\Omega\|_{\min}$  and  $\|\Omega\|_{\max}$  represent the smallest and largest entries in  $\Omega$ , respectively. Now we consider the following weighted matrix factorization problem:

$$\underset{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{n \times r}}{\text{minimize}} \frac{1}{2} \left\| \Omega \circ (\mathbf{U} \mathbf{V}^T - \mathbf{X}^*) \right\|_F^2 + g(\mathbf{U}, \mathbf{V}), \quad (12)$$

where  $g(\mathbf{U}, \mathbf{V})$  is the added regularizer defined in (4). For an arbitrary weight matrix  $\Omega$ , it is proven that the weighted low-rank factorization can be NP-hard [45] and has spurious local minima. When the elements in the weight matrix  $\Omega$  are concentrated, it is expected that (12) can be efficiently solved by a number of iterative optimization algorithms as it is close to an (unweighted) matrix factorization problem (where  $\Omega$  is a matrix of ones) which obeys the strict saddle property [40]. The following result characterizes the geometric structure in the objection function of (12) by directly applying Theorem 3.

**Corollary 2.** Suppose  $\Omega$  satisfies  $\frac{\|\Omega\|_{\max}^2}{\|\Omega\|_{\min}^2} \leq 1.5$ . Set  $\mu \leq \frac{\|\Omega\|_{\min}^2}{16}$ . Then the objective function in (12) has no spurious local minima and satisfies the strict saddle property.

3) *1-bit Matrix Completion:* Finally, we consider the problem of completing a low-rank matrix from a subset of 1-bit measurements [27]. Given  $\mathbf{X}^\diamond \in \mathbb{R}^{n \times m}$ , a subset of indices  $\Omega \subset [m] \times [n]$ , and a differentiable function  $q : \mathbb{R} \rightarrow [0, 1]$ , we observe

$$Y_{i,j} = \begin{cases} +1 & \text{with probability } q(X_{i,j}^\diamond), \\ -1 & \text{with probability } 1 - q(X_{i,j}^\diamond), \end{cases} \quad (13)$$

for all  $(i, j) \in \Omega$ . Typical choices for  $q$  include the logistic regression model where  $q(x) = \frac{e^x}{1+e^x}$  and the probit regression model where  $q(x) = 1 - \Phi(-x/\sigma) = \Phi(x/\sigma)$ . Here  $\Phi$  is the cumulative distribution function (CDF) of a mean-zero Gaussian distribution with variance  $\sigma^2$ . In [27], the authors attempt to recover  $\mathbf{X}^\diamond$  from the incomplete nonlinear measurements  $\{Y_{i,j}\}_{(i,j) \in \Omega}$  by minimizing the negative log-likelihood function

$$\begin{aligned} F_{\Omega, \mathbf{Y}}(\mathbf{X}) := & - \sum_{(i,j) \in \Omega} (\mathbb{1}_{(Y_{i,j}=1)} \log(q(X_{i,j})) \\ & + \mathbb{1}_{(Y_{i,j}=-1)} \log(1 - q(X_{i,j}))) \end{aligned}$$

which results in a maximum likelihood (ML) estimate.

We note that  $F_{\Omega, \mathbf{Y}}$  is a convex function for both the logistic model and the probit model. The following result also establishes that  $F_{\Omega, \mathbf{Y}}$  satisfies the restricted strong convexity and smoothness condition if we observe full 1-bit measurements, i.e.,  $\Omega = [n] \times [m]$ .

**Lemma 1.** Suppose  $\Omega = [n] \times [m]$ . Let

$$\alpha_{q,\gamma} = \min_{|x| \leq \gamma} \min \left( \frac{(q'(x))^2 - q(x)q''(x)}{q^2(x)}, \frac{(q'(x))^2 + (1 - q(x))q''(x)}{(1 - q(x))^2} \right)$$

and

$$\beta_{q,\gamma} = \max_{|x| \leq \gamma} \max \left( \frac{(q'(x))^2 - q(x)q''(x)}{q^2(x)}, \frac{(q'(x))^2 + (1 - q(x))q''(x)}{(1 - q(x))^2} \right).$$

Then  $F_{\Omega, \mathbf{Y}}$  satisfies the restricted strong convexity and smoothness condition:

$$\alpha_{q,\gamma} \|\mathbf{G}\|_F^2 \leq [\nabla^2 F_{\Omega, \mathbf{Y}}(\mathbf{X})](\mathbf{G}, \mathbf{G}) \leq \beta_{q,\gamma} \|\mathbf{G}\|_F^2$$

for any  $\mathbf{G} \in \mathbb{R}^{n \times m}$  and  $\|\mathbf{X}\|_\infty \leq \gamma$ .

The proof of Lemma 1 is given in Appendix A. Now we consider the logistic regression model where  $q(x) = \frac{e^x}{1+e^x}$ .

**Corollary 3.** Suppose  $\Omega = [n] \times [m]$  and  $\gamma \leq 1.3$ . Consider the logistic regression model where  $q(x) = \frac{e^x}{1+e^x}$ . Then  $F_{\Omega, \mathbf{Y}}$  satisfies the restricted strong convexity and smoothness condition with

$$\frac{\beta_{q,\gamma}}{\alpha_{q,\gamma}} \leq 1.5.$$

*Proof of Corollary 3.* Applying Lemma 1 with direct calculation gives

$$\begin{aligned} \alpha_{q,\gamma} &= q'(\gamma) = \frac{e^\gamma}{(1+e^\gamma)^2}, \\ \beta_{q,\gamma} &= q'(0) = \frac{e^0}{(1+e^0)^2} = \frac{1}{4}, \end{aligned}$$

where  $q'(x) = \frac{e^x}{(1+e^x)^2}$ . Now if we restrict  $\|\mathbf{X}\|_\infty \leq 1.3$ , we have

$$\frac{\beta_{q,\gamma}}{\alpha_{q,\gamma}} = 4 \frac{e^{1.3}}{(1+e^{1.3})^2} \leq 1.5.$$

□

Under the assumption that  $\mathbf{X}^\diamond$  is low-rank, a nuclear norm constraint is utilized in [27] to force a low-rank solution. Corollary 3 implies that we can apply matrix factorization for 1-bit matrix recovery given that the elements of  $\mathbf{X}$  are bounded. For the setting where  $\Omega$  is only a subset of  $[n] \times [m]$ , [46] considered the 1-bit matrix *completion* problem with the rank constraint and established a stronger statistical recovery guarantee than that in [27]. Empirical evidence (see [46] and Section V-C) supports that matrix factorization also works for 1-bit matrix completion.

#### IV. PROOF OF THEOREM 3

In this section, we provide a formal proof of Theorem 3. The main argument involves showing that each critical point of  $\rho(\mathbf{W})$  either corresponds to the global solution of (1) or is a strict saddle whose Hessian  $\nabla^2 \rho(\mathbf{W})$  has a strictly negative eigenvalue. Specifically, we show that  $\mathbf{W}$  is a strict saddle by arguing that the Hessian  $\nabla^2 \rho(\mathbf{W})$  has a strictly negative curvature along  $\Delta := \mathbf{W} - \mathbf{W}^* \mathbf{R}$ , i.e.,  $[\nabla^2 \rho(\mathbf{W})](\Delta, \Delta) \leq -\tau \|\Delta\|_F^2$  for some  $\tau > 0$ . Here  $\mathbf{R}$  is an  $r \times r$  orthonormal matrix such that the distance between  $\mathbf{W}$  and  $\mathbf{W}^*$  rotated through  $\mathbf{R}$  is as small as possible.

#### A. Supporting Results

We first present some useful results. The  $(2r, 4r)$ -restricted strong convexity and smoothness assumption (3) implies the following isometry property, whose proof is given in Appendix B.

**Proposition 2.** Suppose the function  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) with positive  $\alpha$  and  $\beta$ . Then for any  $n \times m$  matrices  $\mathbf{Z}, \mathbf{G}, \mathbf{H}$  of rank at most  $2r$ , we have

$$\left| \frac{2}{\alpha + \beta} [\nabla^2 f(\mathbf{Z})](\mathbf{G}, \mathbf{H}) - \langle \mathbf{G}, \mathbf{H} \rangle \right| \leq \frac{\beta - \alpha}{\beta + \alpha} \|\mathbf{G}\|_F \|\mathbf{H}\|_F.$$

The following result provides an upper bound on the energy of the difference  $\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}$  when projected onto the column space of  $\mathbf{W}$ . Its proof is given in Appendix C.

**Lemma 2.** Suppose  $f(\mathbf{X})$  satisfies the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3). For any critical point  $\mathbf{W}$  of (5), let  $\mathbf{P}_\mathbf{W} \in \mathbb{R}^{(m+n) \times (m+n)}$  be the orthogonal projector onto the column space of  $\mathbf{W}$ . Then

$$\|(\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T})\mathbf{P}_\mathbf{W}\|_F \leq 2 \frac{\beta - \alpha}{\beta + \alpha} \|\mathbf{X} - \mathbf{X}^*\|_F.$$

We remark that Lemma 2 is a variant of [20, Lemma 3.2]. While the result there requires the  $4r$ -RIP condition of the objective function, our result depends on the  $(2r, 4r)$ -restricted strong convexity and smoothness condition. Our result is also slightly tighter than [20, Lemma 3.2].

In addition, for any matrices  $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times r}$ , the following result relates the distance between  $\mathbf{C}\mathbf{C}^T$  and  $\mathbf{D}\mathbf{D}^T$  to the distance between  $\mathbf{C}$  and  $\mathbf{D}$ .

**Lemma 3.** For any matrices  $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times r}$  with ranks  $r_1$  and  $r_2$ , respectively, let  $\mathbf{R} = \arg \min_{\mathbf{R}' \in \mathcal{O}_r} \|\mathbf{C} - \mathbf{D}\mathbf{R}'\|_F$ . Then

$$\begin{aligned} & \|\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T\|_F^2 / \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F^2 \\ & \geq \max \left\{ 2(\sqrt{2} - 1)\sigma_r^2(\mathbf{D}), \min \left\{ \sigma_{r_1}^2(\mathbf{C}), \sigma_{r_2}^2(\mathbf{D}) \right\} \right\}. \end{aligned}$$

If  $\mathbf{C} = \mathbf{0}$ , then we have

$$\|\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T\|_F^2 \geq \sigma_{r_2}^2(\mathbf{D}) \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F^2.$$

We present one more useful result in the following Lemma.

**Lemma 4.** [30, Lemma 3.6] For any matrices  $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times r}$ , let  $\mathbf{P}_\mathbf{C}$  be the orthogonal projector onto the range of  $\mathbf{C}$ . Let  $\mathbf{R} = \arg \min_{\mathbf{R}' \in \mathcal{O}_r} \|\mathbf{C} - \mathbf{D}\mathbf{R}'\|_F$ . Then

$$\begin{aligned} \|\mathbf{C}(\mathbf{C} - \mathbf{D}\mathbf{R})^T\|_F^2 & \leq \frac{1}{8} \|\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T\|_F^2 \\ & + (3 + \frac{1}{2(\sqrt{2} - 1)}) \|(\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T)\mathbf{P}_\mathbf{C}\|_F^2. \end{aligned}$$

Finally, we provide the gradient and Hessian expressions for  $\rho(\mathbf{W})$ . The gradient of  $\rho(\mathbf{W})$  is given by

$$\begin{aligned} \nabla_{\mathbf{U}} \rho(\mathbf{U}, \mathbf{V}) &= \nabla f(\mathbf{X})\mathbf{V} + \mu \mathbf{U}(\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}), \\ \nabla_{\mathbf{V}} \rho(\mathbf{U}, \mathbf{V}) &= \nabla f(\mathbf{X})^T \mathbf{U} - \mu \mathbf{V}(\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}). \end{aligned}$$

Standard computations give the the Hessian quadrature form  $[\nabla^2 \rho(\mathbf{W})](\Delta, \Delta)$  for any  $\Delta = \begin{bmatrix} \Delta_{\mathbf{U}} \\ \Delta_{\mathbf{V}} \end{bmatrix}$  where  $\Delta_{\mathbf{U}} \in \mathbb{R}^{n \times r}, \Delta_{\mathbf{V}} \in \mathbb{R}^{m \times r}$ :

$$\begin{aligned} & [\nabla^2 \rho(\mathbf{W})](\Delta, \Delta) \\ &= [\nabla^2 f(\mathbf{X})](\Delta_{\mathbf{U}} \mathbf{V}^T + \mathbf{U} \Delta_{\mathbf{V}}^T, \Delta_{\mathbf{U}} \mathbf{V}^T + \mathbf{U} \Delta_{\mathbf{V}}^T) \\ & + 2 \langle \nabla f(\mathbf{X}), \Delta_{\mathbf{U}} \Delta_{\mathbf{V}}^T \rangle + [\nabla^2 g(\mathbf{W})](\Delta, \Delta), \end{aligned}$$

where

$$\begin{aligned} & [\nabla^2 g(\mathbf{W})](\Delta, \Delta) = \mu \langle \widehat{\mathbf{W}}^T \mathbf{W}, \widehat{\Delta}^T \Delta \rangle \\ & + \mu \langle \widehat{\mathbf{W}} \widehat{\Delta}^T, \Delta \mathbf{W}^T \rangle + \mu \langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, \Delta \Delta^T \rangle. \end{aligned}$$

### B. The Formal Proof

*Proof of Theorem 3.* Any critical point  $\mathbf{W}$  of  $\rho(\mathbf{W})$  satisfies  $\nabla \rho(\mathbf{W}) = \mathbf{0}$ , i.e.,

$$\nabla f(\mathbf{X})\mathbf{V} + \mu \mathbf{U}(\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}) = \mathbf{0}, \quad (14)$$

$$\nabla f(\mathbf{X})^T \mathbf{U} - \mu \mathbf{V}(\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}) = \mathbf{0}. \quad (15)$$

By (15), we obtain

$$\mathbf{U}^T \nabla f(\mathbf{X}) = \mu (\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}) \mathbf{V}^T.$$

Multiplying (14) by  $\mathbf{U}^T$  and plugging in the expression for  $\mathbf{U}^T \nabla f(\mathbf{X})$  from the above equation  $\mathbf{V}^T$  gives

$$(\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}) \mathbf{V}^T \mathbf{V} + \mathbf{U}^T \mathbf{U} (\mathbf{U}^T \mathbf{U} - \mathbf{V}^T \mathbf{V}) = \mathbf{0},$$

which further implies

$$\mathbf{U}^T \mathbf{U} \mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} \mathbf{V}^T \mathbf{V}.$$

Note that  $\mathbf{U}^T \mathbf{U}$  and  $\mathbf{V}^T \mathbf{V}$  are the principal square roots (i.e., PSD square roots) of  $\mathbf{U}^T \mathbf{U} \mathbf{U}^T \mathbf{U}$  and  $\mathbf{V}^T \mathbf{V} \mathbf{V}^T \mathbf{V}$ , respectively. Utilizing the result that a PSD matrix has a unique principal square root [47], we obtain

$$\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V}. \quad (16)$$

Thus, we can simplify (14) and (15) by

$$\nabla_{\mathbf{U}} \rho(\mathbf{U}, \mathbf{V}) = \nabla f(\mathbf{X})\mathbf{V} = \mathbf{0}, \quad (17)$$

$$\nabla_{\mathbf{V}} \rho(\mathbf{U}, \mathbf{V}) = \nabla f(\mathbf{X})^T \mathbf{U} = \mathbf{0}. \quad (18)$$

Now we turn to prove the strict saddle property and that there are no spurious local minima.

First, note that as guaranteed by Proposition 1,  $\mathbf{X}^*$  is the unique  $n \times m$  matrix with rank at most  $r$ . Also the gradient of  $f(\mathbf{X})$  vanishes at  $\mathbf{X}^*$  since (1) is an unconstraint optimization problem. Denote the set of critical points of  $\rho(\mathbf{W})$  by

$$\mathcal{C} := \left\{ \mathbf{W} \in \mathbb{R}^{(n+m) \times r} : \nabla \rho(\mathbf{W}) = \mathbf{0} \right\}.$$

We separate  $\mathcal{C}$  into two subsets:

$$\mathcal{C}_1 := \mathcal{C} \cap \left\{ \mathbf{W} \in \mathbb{R}^{(n+m) \times r} : \mathbf{U} \mathbf{V}^T = \mathbf{X}^* \right\},$$

$$\mathcal{C}_2 := \mathcal{C} \cap \left\{ \mathbf{W} \in \mathbb{R}^{(n+m) \times r} : \mathbf{U} \mathbf{V}^T \neq \mathbf{X}^* \right\},$$

satisfying  $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ . Since any critical point  $\mathbf{W}$  satisfies (16),  $g(\mathbf{W})$  achieves its global minimum at  $\mathbf{W}$ . Also  $f(\mathbf{X})$  achieves its global minimum at  $\mathbf{X}^*$ . We conclude that  $\mathbf{W}$  is the globally optimal solution of  $\rho$  for any  $\mathbf{W} \in \mathcal{C}_1$ . If we show that any  $\mathbf{W} \in \mathcal{C}_2$  is a strict saddle, then we prove that there are no spurious local minima as well as the strict saddle property. Thus, the remaining part is to show that  $\mathcal{C}_2$  is the set of strict saddles.

To show that  $\mathcal{C}_2$  is the set of strict saddles, it is sufficient to find a direction  $\Delta$  along which the Hessian has a strictly

negative curvature for each of these points. We construct  $\Delta = \mathbf{W} - \mathbf{W}^* \mathbf{R}$ , the difference from  $\mathbf{W}$  to its nearest global factor  $\mathbf{W}^*$ , where

$$\mathbf{R} = \arg \min_{\mathbf{R}' \in \mathcal{O}_r} \|\mathbf{W} - \mathbf{W}^* \mathbf{R}'\|_F.$$

Such  $\Delta$  satisfies  $\Delta \neq \mathbf{0}$  since  $\mathbf{X} \neq \mathbf{X}^*$  implying  $\mathbf{W}\mathbf{W}^T \neq \mathbf{W}^*\mathbf{W}^{*T}$ . Then we evaluate the Hessian bilinear form along the direction  $\Delta$ :

$$\begin{aligned} [\nabla^2 \rho(\mathbf{W})](\Delta, \Delta) &= 2 \underbrace{\langle \nabla f(\mathbf{X}), \Delta_U \Delta_V^T \rangle}_{\Pi_1} \\ &+ \underbrace{[\nabla^2 f(\mathbf{X})](\Delta_U \mathbf{V}^T + \mathbf{U} \Delta_V^T, \Delta_U \mathbf{V}^T + \mathbf{U} \Delta_V^T)}_{\Pi_2} \\ &+ \mu \underbrace{\langle \widehat{\mathbf{W}} \widehat{\Delta}^T, \Delta \mathbf{W}^T \rangle}_{\Pi_3} + \mu \underbrace{\langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, \Delta \Delta^T \rangle}_{\Pi_4}. \end{aligned} \quad (19)$$

The following result (which is proved in Appendix E) states that  $\Pi_1$  is strictly negative, while the remaining terms are relatively small, though they may be nonnegative:

$$\begin{aligned} \Pi_1 &\leq -\alpha \|\mathbf{X} - \mathbf{X}^*\|_F^2, \quad \Pi_2 \leq \beta \|\mathbf{W} \Delta^T\|_F^2, \\ \Pi_3 &\leq \|\mathbf{W} \Delta^T\|_F^2, \quad \Pi_4 \leq 2 \|\mathbf{X} - \mathbf{X}^*\|_F^2. \end{aligned} \quad (20)$$

Now, substituting (20) into (19) gives

$$\begin{aligned} &[\nabla^2 \rho(\mathbf{W})](\Delta, \Delta) \\ &= 2\Pi_1 + \Pi_2 + \mu\Pi_3 + \mu\Pi_4 \\ &\leq -2\alpha \|\mathbf{X} - \mathbf{X}^*\|_F^2 + (\beta + \mu) \cdot \|\mathbf{W} \Delta^T\|_F^2 \\ &\quad + 2\mu \|\mathbf{X} - \mathbf{X}^*\|_F^2 \\ &\stackrel{(i)}{\leq} (-2\alpha + 2\mu) \|\mathbf{X} - \mathbf{X}^*\|_F^2 \\ &\quad + (\beta + \mu) \left( \frac{1}{2} + (12 + \frac{2}{\sqrt{2}-1}) \left( \frac{\beta-\alpha}{\beta+\alpha} \right)^2 \right) \|\mathbf{X} - \mathbf{X}^*\|_F^2 \\ &\stackrel{(ii)}{\leq} -0.2\alpha \|\mathbf{X} - \mathbf{X}^*\|_F^2, \end{aligned} \quad (21)$$

where (i) utilizes Lemmas 2 and 4, (ii) utilizes the following inequality (which is proved in Appendix F)

$$\|\mathbf{W}\mathbf{W}^T - \mathbf{W}^*\mathbf{W}^*\|_F^2 \leq 4 \|\mathbf{X} - \mathbf{X}^*\|_F^2, \quad (22)$$

and (ii) holds because  $\frac{\beta}{\alpha} \leq 1.5$  and  $\mu \leq \frac{1}{16}\alpha$ . Thus, if  $\mathbf{X} \neq \mathbf{X}^*$ ,  $[\nabla^2 \rho(\mathbf{X})](\Delta, \Delta)$  is always negative. This implies that  $\mathbf{W}$  is a strict saddle.

To complete the proof, we utilize Lemma 3 to further bound the last term in (21):

$$\begin{aligned} &[\nabla^2 \rho(\mathbf{W})](\Delta, \Delta) \leq -0.05\alpha \|\mathbf{W}\mathbf{W}^T - \mathbf{W}^*\mathbf{W}^{*T}\|_F^2 \\ &\leq -0.05\alpha \|\Delta\|_F^2 \begin{cases} 2(\sqrt{2}-1)\sigma_r^2(\mathbf{W}^*), & r = r^*, \\ \min \{\sigma_{r^c}^2(\mathbf{W}), \sigma_{r^*}^2(\mathbf{W}^*)\}, & r > r^*, \\ \sigma_{r^*}^2(\mathbf{W}^*), & r_c = 0, \end{cases} \end{aligned}$$

where  $r^c$  is the rank of  $\mathbf{W}$ , the fist inequality utilizes (22), and the second inequality follows from Lemma 3. We complete the proof of Theorem 3 by noting that  $\sigma_\ell^2(\mathbf{W}^*) = 2\sigma_\ell(\mathbf{X}^*)$  for all  $\ell \in \{1, \dots, r^*\}$  since

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{Q}_{\mathbf{U}^*} \Sigma^{*1/2} \\ \mathbf{Q}_{\mathbf{V}^*} \Sigma^{*1/2} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\mathbf{U}^*}/\sqrt{2} \\ \mathbf{Q}_{\mathbf{V}^*}/\sqrt{2} \end{bmatrix} \left( \sqrt{2}\Sigma^{*1/2} \right) \mathbf{I}$$

is an SVD of  $\mathbf{W}^*$ , where we recall that  $\mathbf{X}^* = \mathbf{Q}_{\mathbf{U}^*} \Sigma^* \mathbf{Q}_{\mathbf{V}^*}^T$  is an SVD of  $\mathbf{X}^*$ .  $\square$

*Remark 8.* From (21), we observe that a smaller  $\mu$  yields a more negative bound on  $[\nabla^2 \rho(\mathbf{X})](\Delta, \Delta)$ . This can be explained intuitively as follows. First note that any critical point  $\mathbf{W}$  satisfies (16) provided  $\mu > 0$ , no matter how large or small  $\mu$  is. The Hessian information about  $g(\mathbf{W})$  is represented by the terms  $\Pi_3$  and  $\Pi_4$ . We have

$$\begin{aligned} \Pi_3 + \Pi_4 &= \langle \widehat{\mathbf{W}} \widehat{\Delta}^T, \Delta \mathbf{W}^T \rangle + \langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, \Delta \Delta^T \rangle \\ &= \langle \widehat{\mathbf{W}}^T \Delta, \Delta^T \widehat{\mathbf{W}} \rangle + \langle \widehat{\mathbf{W}}^T \Delta, \widehat{\mathbf{W}}^T \Delta \rangle \\ &= \langle \widehat{\mathbf{W}}^T \Delta, \widehat{\mathbf{W}}^T \Delta + \Delta^T \widehat{\mathbf{W}} \rangle \\ &\geq 0, \end{aligned}$$

where the last line holds since for any  $r \times r$  matrix  $\mathbf{A}$ ,

$$\begin{aligned} &\langle \mathbf{A}, \mathbf{A} + \mathbf{A}^T \rangle \\ &= \frac{1}{2} \langle \mathbf{A} + \mathbf{A}^T, \mathbf{A} + \mathbf{A}^T \rangle + \frac{1}{2} \langle \mathbf{A} - \mathbf{A}^T, \mathbf{A} + \mathbf{A}^T \rangle \\ &= \frac{1}{2} \|\mathbf{A} + \mathbf{A}^T\|_F^2 \geq 0. \end{aligned}$$

Thus the Hessian of  $\rho$  evaluated at any critical point  $\mathbf{W}$  is a PSD matrix<sup>1</sup> instead of having a negative eigenvalue. In low-rank, PSD matrix optimization problems, the corresponding objective function (without any regularizer such as  $g(\mathbf{W})$ ) is proved to have the strict saddle property [18], [30]. Therefore,  $h(\mathbf{W})$  is also expected to have the strict saddle property, and so is  $\rho(\mathbf{W})$  when  $\mu$  is small, i.e., the Hessian of  $g(\mathbf{W})$  has little influence on the Hessian of  $\rho(\mathbf{W})$  when  $\mu$  is small. Our results also indicate that when the restricted strict convexity constant  $\alpha$  is not provided a priori, we can always choose a small  $\mu$  to ensure the strict saddle property of  $\rho(\mathbf{W})$  is met, and hence we are guaranteed the global convergence of a number of local search algorithms applied to (5).

## V. EXPERIMENTS

In this section, we present a set of experiments on matrix sensing, matrix completion, and 1-bit matrix completion to demonstrate the performance of iterative algorithms for low-rank matrix optimization. Unless noted otherwise, we denote the matrix factorization approach by NVX and use the minFunc package<sup>2</sup> to perform the local search algorithms for the factored problem.

### A. Matrix Sensing

We first present some experiments to illustrate the performance of local search algorithms for the matrix sensing problem with the factorization approach (11). In these experiments, we set  $n = 50$ ,  $m = 50$  and vary the rank  $r$  from 1 to 19. We generate a rank- $r$   $n \times m$  random matrix  $\mathbf{X}^*$  by setting  $\mathbf{X}^* = \tilde{\mathbf{U}} \tilde{\mathbf{V}}^T$  where  $\tilde{\mathbf{U}}$  and  $\tilde{\mathbf{V}}$  are respectively  $n \times r$  and

<sup>1</sup>This can also be observed since any critical point  $\mathbf{W}$  is a global minimum point of  $\rho(\mathbf{W})$ , which directly indicates that  $\nabla^2 \rho(\mathbf{W}) \succeq \mathbf{0}$ .

<sup>2</sup>Software available at <https://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>

$m \times r$  matrixes of normally distributed random numbers. We then obtain  $p$  random measurements  $\mathbf{y} = \mathcal{A}(\mathbf{X}^*)$  with

$$y_i = \langle \mathbf{X}^*, \mathbf{Y}_i \rangle,$$

where the entries of each  $n \times m$  matrix  $\mathbf{Y}_i$  are independent and identically distributed (i.i.d.) normal random variables with zero mean and variance  $\frac{1}{p}$  for  $i \in \{1, 2, \dots, p\}$ . For each pair of  $r$  and the number of measurements, 10 Monte Carlo trials are carried out and for each trial, and we claim matrix recovery to be successful if the relative reconstruction error satisfies

$$\frac{\|\mathbf{X}^* - \widehat{\mathbf{X}}\|_F}{\|\mathbf{X}^*\|_F} \leq 10^{-4},$$

where we denote by  $\widehat{\mathbf{X}}$  the reconstructed matrix. Figure 1 displays the phase transition for factorized gradient descent starting from a random initialization, the singular value projection (SVP) method proposed in [48] which requires a SVD in each iteration, and the convex approach which solves

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} \|\mathbf{X}\|_* \\ & \text{subject to } \mathbf{y} = \mathcal{A}(\mathbf{X}). \end{aligned} \quad (23)$$

We see that there are only negligible differences between the different approaches for matrix sensing; these approaches also have very similar performance guarantees when the Gaussian sensing operator  $\mathcal{A}$  satisfies the RIP [12]. We note that with or without the regularizer  $g$  as defined in (4), local search algorithms have similar performance with random initialization. Hence, throughout all of the experiments, we simply discard the regularizer  $g$ , but we stress that identical performance is observed if we have this regularizer  $g$ .

The previous experiments suppose that  $r$  is known for SVP and the matrix factorization approach. We note, however, that our result in Theorem 3 also covers the over-parameterization case where  $r > r^*$ . To illustrate the possible influence of over-parameterization, we generate a rank- $r^*$  random matrix  $\mathbf{X}^* \in \mathbb{R}^{n \times m}$  with  $r^* = 4$  and  $n = m = 50$  and obtain  $p = 4Rn$  random measurements (so that the measurement operator  $\mathcal{A}$  satisfies the RIP of rank  $R$ ), where  $R = 7$ . We then solve the matrix factorization problem<sup>3</sup> with  $r = 4, 5, 6, 7$  and display the corresponding convergence results in Figure 2. As can be seen, the matrix factorization approach converges to the target matrix  $\mathbf{X}^*$  in both the exact-parameterization and over-parameterization cases. However, we also observe that it converges slower in the over-parameterization case (i.e.,  $r > r^*$ ) than in the exact-parameterization case (i.e.,  $r = r^*$ ).

### B. Matrix Completion

We compare the performance of the matrix factorization approach with SVP [48], the convex approach, and singular value thresholding<sup>4</sup> (SVT) [50] for matrix completion where we want to recover a low-rank matrix  $\mathbf{X}^*$  from incomplete measurements  $\{X_{ij}^*\}_{(i,j) \in \Omega}$ , where  $\Omega \subset [n] \times [m]$ . Let  $\mathcal{P}_\Omega$  denote the projection onto the index set  $\Omega$ . The convex

<sup>3</sup>To avoid tuning the parameters (such as step-size) for different  $r$ , we use the minFunc package with the default setting, which solves the factored problem by the “LBFGS” algorithm [49].

<sup>4</sup>Software available at <http://svt.stanford.edu/>

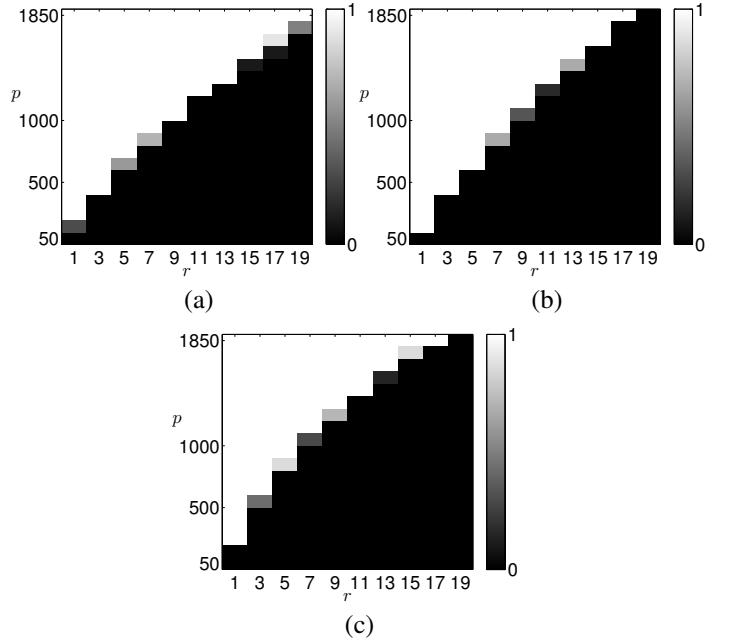


Figure 1. Rate of success for matrix sensing by (a) solving the factorized problem (11) with gradient descent; (b) SVP [48]; (c) solving the convex problem (23).

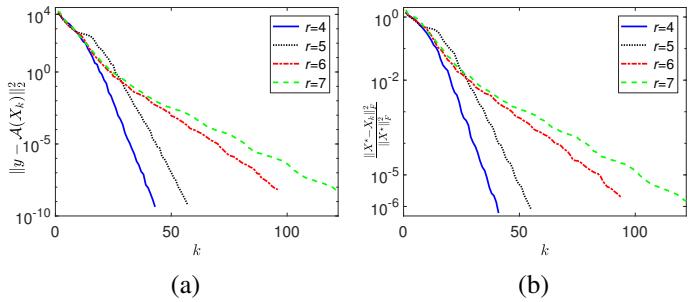


Figure 2. The performance in terms of (a) objective value and (b) the relative Frobenius norm of the error versus the iteration  $k$  for the matrix factorization approach solving matrix sensing with  $r^* = 4, n = m = 50, p = 4Rn, R = 7$  and  $r$  varying from  $r^*$  to  $R$ .

approach (denoted by CVX) attempts to use the nuclear norm as a convex relaxation of the rankness and solves

$$\begin{aligned} & \underset{\mathbf{X}}{\text{minimize}} \|\mathbf{X}\|_* \\ & \text{subject to } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{X}^*). \end{aligned} \quad (24)$$

To make the recovery of  $\mathbf{X}^*$  well-posed, we require  $\mathbf{X}^*$  to be incoherent such that the information in  $\mathbf{X}$  is not concentrated in a small number of entries [9]. A matrix  $\mathbf{X} \in \mathbb{R}^{n \times m}$  with singular value decomposition  $\mathbf{X} = \mathbf{L}\mathbf{\Sigma}\mathbf{Q}^T$  is  $u$ -incoherent if [48, Definition 2.1]

$$\max_{ij} |L_{ij}| \leq \sqrt{\frac{u}{n}}, \quad \max_{ij} |Q_{ij}| \leq \sqrt{\frac{u}{m}}.$$

Though  $\mathcal{P}_\Omega$  does not satisfy the  $r$ -RIP (10) for all low-rank matrices  $\mathbf{X}$ , it satisfies the RIP when restricted to low-rank incoherent matrices.

**Theorem 4.** [48, Theorem 4.2] *Without loss of generality, assume  $n \geq m$ . There exists a constant  $C \geq 0$  such that for*

$\Omega \in [n] \times [m]$  chosen according to the Bernouli model with density greater than  $Cu^2r^2 \log n / \delta^2 m$ , with probability at least  $1 - e^{-n \log n}$ , the RIP holds for all  $\mu$ -incoherent matrices  $\mathbf{X}$  of rank at most  $r$ .

Thus, if local search algorithms (such as gradient descent) start with a random initialization and the iterates remain incoherent, then Theorem 3 guarantees the global convergence of the matrix factorization approach with these algorithms. We note that this hypothesis is also required for SVP [48]. Though we can add a regularizer for incoherence as in [19], empirical evidence supports this hypothesis that the iterates in gradient descent are incoherent.

In the first set of experiments, we set  $n = m = 100$  and vary the rank  $r$  from 1 to 30. Similar to the setup for matrix sensing in Section V-A, we generate a rank- $r$  random matrix and randomly obtain  $p$  entries, i.e.,  $|\Omega| = p$ . Figure 3 displays the phase transition for gradient descent with a random initialization, SVP [48], singular value thresholding (SVT) [50], and the convex approach. As can be seen, the matrix factorization approach has similar phase transition to SVP, and is slightly better than SVT and the convex approach in terms of the number of measurements needed for successful recovery.

In the second set of experiments, we set  $r = 5$  and  $p = 3r(2n - r)$  (3 times the number of degrees of freedom within a rank- $r$   $n \times n$  matrix), and vary  $n$  from 40 to 5120. We compare the time needed for the four approaches in Figure 4; our matrix factorization approach is much faster than the other methods. The time savings for the matrix factorization approach comes from avoiding performing the SVD, which is needed both for SVT and SVP in each iteration. We also observe that convex approach has the highest computational complexity and is not scalable (which is the reason that we only present its time for  $n$  up to 640).

### C. 1-bit Matrix Completion

In the last set of experiments, we compare the performance of the matrix factorization approach with the convex approach<sup>5</sup> in [27] for 1-bit matrix completion. We first note that to make the recovery problem well-posed, a constraint on  $\|\mathbf{X}\|_\infty$  (the entry-wise maximum of the matrix  $\mathbf{X}$ ) is applied in [27] to require that the matrix is not too “spiky”. Instead of using the constraint on  $\|\mathbf{X}\|_\infty$ , we add a smooth regularizer  $\|\mathbf{X}\|_F^2$  and turn to minimize the following objective function

$$f_{\Omega, \mathbf{Y}}(\mathbf{X}) = F_{\Omega, \mathbf{Y}}(\mathbf{X}) + \frac{\eta}{2} \|\mathbf{X}\|_F^2,$$

which is also a convex function over  $\mathbf{X}$  and satisfies a similar restricted strong convexity and smoothness condition to  $F_{\Omega, \mathbf{Y}}$  in Lemma 1. In the case where we only observe part of the entries, then in light of Theorem 4, the corresponding objective function is expected to satisfy the strong convexity and smoothness condition for all incoherent matrices. Thus, we

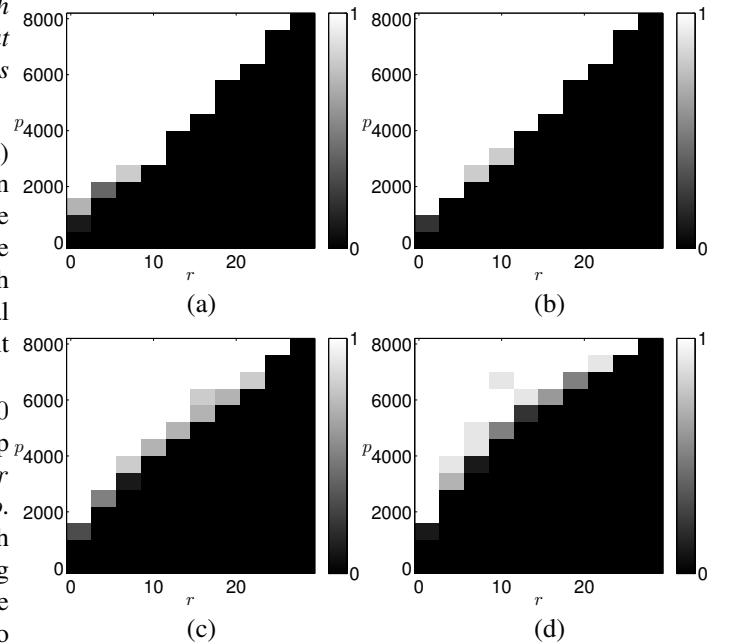


Figure 3. Rate of success for matrix sensing by (a) the matrix factorization approach with gradient descent; (b) SVP [48]; (c) solving the convex problem (24); (d) SVT [48].

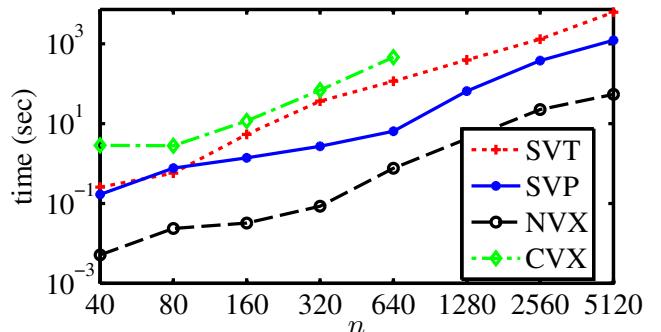


Figure 4. Average computation time needed for different algorithms solving matrix completion.

factorize  $\mathbf{X}$  into  $\mathbf{U}\mathbf{V}^\top$  and solve the following optimization problem over the  $n \times r$  and  $m \times r$  matrices  $\mathbf{U}$  and  $\mathbf{V}$ :

$$\underset{\mathbf{U}, \mathbf{V}}{\text{minimize}} \rho_{\Omega, \mathbf{Y}}(\mathbf{U}, \mathbf{V}) = f_{\Omega, \mathbf{Y}}(\mathbf{U}\mathbf{V}^\top). \quad (25)$$

To evaluate the performance of this factorization approach on 1-bit matrix completion, we generate  $n \times r$  matrices  $\mathbf{U}^\diamond$  and  $\mathbf{V}^\diamond$  with entries drawn i.i.d. from a uniform distribution on  $[-\frac{1}{2}, \frac{1}{2}]$  and construct a random  $n \times n$  matrix  $\mathbf{X}^\diamond$  with rank  $r$ . Similar to the setup in [27], the matrix is then scaled so that  $\|\mathbf{X}^\diamond\| = 1$ . We obtain 1-bit observations  $\{Y_{i,j}\}_{(i,j) \in \Omega}$  by adding Gaussian noise of variance  $\sigma^2$  and recording the sign of the resulting value (13), where the subset of indices  $\Omega$  is chosen at random with  $E|\Omega| = p$ . We compare the performance of the factorization approach and the convex approach [27] over a range of different values of  $n$ ,  $p$ ,  $r$  or  $\sigma$ . Figures 5(a)-(d) show the normalized squared Frobenius norm of the error  $\frac{\|\widehat{\mathbf{X}} - \mathbf{X}^\diamond\|}{\|\mathbf{X}^\diamond\|_F^2}$  (where  $\widehat{\mathbf{X}}$  denotes the reconstructed matrix) and average the results over 10 draws of Monte

<sup>5</sup>Software available at <http://mdav.ece.gatech.edu/software/>

Carlo trials. We observe that matrix factorization approach has slightly better performance than the convex approach for 1-bit matrix completion [27]. Note that this phenomenon (the factorization approach having better performance) is also observed in [46]. We repeat these experiments but obtaining 1-bit observations with the logistic regression model where  $g(x) = \frac{e^x}{1+e^x}$  for (13) and display the results in Figure 6.

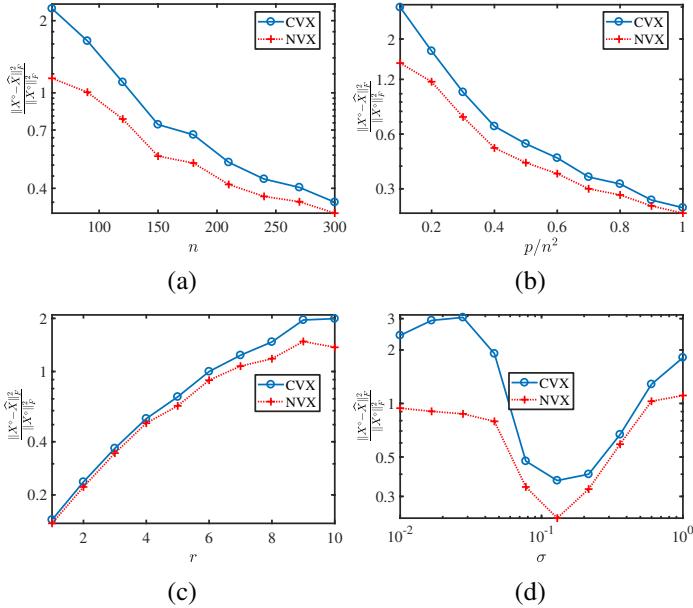


Figure 5. The performance in terms of the relative Frobenius norm of the error for the matrix factorization approach (denoted by NVX) and the convex approach in [27] (denoted by CVX) for solving the 1-bit matrix completion with probit regression model and (a) varying  $n$  and  $\sigma = 0.3$ ,  $r = 7$ ,  $p = 0.5n^2$ ; (b) varying  $p$  and  $\sigma = 0.3$ ,  $n = 200$ ,  $r = 7$ ; (c) varying  $r$  and  $\sigma = 0.3$ ,  $n = 200$ ,  $p = 0.25n^2$ ; (d) varying  $\sigma$  and  $n = 200$ ,  $r = 4$ ,  $p = 0.25n^2$ . The results are plotted in the log scale.

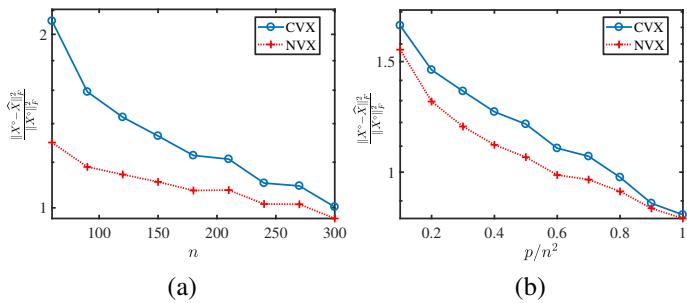


Figure 6. The performance in terms of the relative Frobenius norm of the error for the matrix factorization approach (denoted by NVX) and the convex approach in [27] (denoted by CVX) for solving the 1-bit matrix completion with logistic regression model and (a) varying  $n$  and  $r = 2$ ,  $p = 0.5n^2$ ; (b) varying  $p$  and  $n = 200$ ,  $r = 2$ . The results are plotted in the log scale.

## VI. CONCLUSION

This paper considers low-rank matrix optimization on general (nonsymmetric and rectangular) matrices with general objective functions. By focusing on general objective functions, we provide a unifying framework for low-rank matrix optimizations with the factorization approach. Although the resulting optimization problem is not convex, we show that

the reformulated objection function has a simple landscape: there are no spurious local minima and any critical point not being a local minimum is a strict saddle such that the Hessian evaluated at this point has a strictly negative eigenvalue. These properties guarantee that a number of iterative optimization algorithms (such as gradient descent and the trust region method) will converge to the global optimum from a random initialization.

## APPENDIX A PROOF OF LEMMA 1

*Proof of Lemma 1.* We compute the partial derivative of  $F_{\Omega, \mathbf{Y}}$  in terms of  $X_{i,j}$  as

$$\frac{\partial F_{\Omega, \mathbf{Y}}}{\partial X_{i,j}} = -\mathbb{1}_{(Y_{i,j}=1)} \frac{q'(X_{i,j})}{q(X_{i,j})} + \mathbb{1}_{(Y_{i,j}=-1)} \frac{q'(X_{i,j})}{1-q(X_{i,j})},$$

which implies

$$\begin{aligned} \frac{\partial^2 F_{\Omega, \mathbf{Y}}}{\partial X_{i,j} \partial X_{i,j}} &= \mathbb{1}_{(Y_{i,j}=1)} \frac{(q'(X_{i,j}))^2 - q(X_{i,j})q''(X_{i,j})}{q^2(X_{i,j})} \\ &\quad + \mathbb{1}_{(Y_{i,j}=-1)} \frac{(q'(X_{i,j}))^2 + (1-q(X_{i,j}))q''(X_{i,j})}{(1-q(X_{i,j}))^2} \end{aligned}$$

and

$$\frac{\partial^2 F_{\Omega, \mathbf{Y}}}{\partial X_{i,j} \partial X_{k,\ell}} = 0$$

for all  $(k, \ell) \neq (i, j)$ . Thus, the bilinear form for the Hessian of  $\nabla^2 F_{\Omega, \mathbf{Y}}(\mathbf{X})$  can be computed as

$$[\nabla^2 F_{\Omega, \mathbf{Y}}(\mathbf{X})](\mathbf{G}, \mathbf{G}) = \sum_i \sum_j \frac{\partial^2 F_{\Omega, \mathbf{Y}}}{\partial X_{i,j} \partial X_{i,j}} G_{i,j}^2$$

for any  $\mathbf{G} \in \mathbb{R}^{n \times m}$ . Now since by assumption  $\|\mathbf{X}\|_\infty \leq \gamma$ , we have

$$\alpha_{q, \gamma} \|\mathbf{G}\|_F^2 \leq [\nabla^2 F_{\Omega, \mathbf{Y}}(\mathbf{X})](\mathbf{G}, \mathbf{G}) \leq \beta_{q, \gamma} \|\mathbf{G}\|_F^2.$$

□

## APPENDIX B PROOF OF PROPOSITION 2

*Proof of Proposition 2.* This proof follows similar steps to the proof of [51, Lemma 2.1]. First note that the bilinear form  $[\nabla^2 f(\mathbf{Z})](\mathbf{G}, \mathbf{H}) = \sum_{i,j,k,l} \frac{\partial^2 f(\mathbf{Z})}{\partial Z_{ij} \partial Z_{kl}} G_{ij} H_{kl}$  implies  $[\nabla^2 f(\mathbf{Z})](\mathbf{G}, \mathbf{H})$  is invariant under all scalings for both  $\mathbf{G}$  and  $\mathbf{H}$ , i.e.,

$$[\nabla^2 f(\mathbf{Z})](a\mathbf{G}, b\mathbf{H}) = ab[\nabla^2 f(\mathbf{Z})](\mathbf{G}, \mathbf{H})$$

for any  $a, b \in \mathbb{R}$ . If either  $\mathbf{G}$  or  $\mathbf{H}$  is zero, (3) holds since both sides are 0.

Now suppose both  $\mathbf{G}$  or  $\mathbf{H}$  are nonzero. By the scaling invariance property of both sides in (3), we assume  $\|\mathbf{G}\|_F = \|\mathbf{H}\|_F = 1$  without loss of generality. Note that the  $(2r, 4r)$ -restricted strong convexity and smoothness condition (3) implies

$$\begin{aligned} \alpha \|\mathbf{G} \pm \mathbf{H}\|_F^2 &\leq [\nabla^2 f(\mathbf{X})](\mathbf{G} \pm \mathbf{H}, \mathbf{G} \pm \mathbf{H}) \\ &\leq \beta \|\mathbf{G} \pm \mathbf{H}\|_F^2. \end{aligned}$$

Thus we have

$$\begin{aligned} & -\frac{\beta-\alpha}{2} \left( \|\mathbf{G}\|_F^2 + \|\mathbf{H}\|_F^2 \right) \\ & \leq 2 [\nabla^2 f(\mathbf{Z})] (\mathbf{G}, \mathbf{H}) - (\alpha + \beta) \langle \mathbf{G}, \mathbf{H} \rangle \\ & \leq \frac{\beta-\alpha}{2} \left( \|\mathbf{G}\|_F^2 + \|\mathbf{H}\|_F^2 \right), \end{aligned}$$

which further implies

$$\begin{aligned} & |2 [\nabla^2 f(\mathbf{Z})] (\mathbf{G}, \mathbf{H}) - (\alpha + \beta) \langle \mathbf{G}, \mathbf{H} \rangle| \\ & \leq \beta - \alpha = (\beta - \alpha) \|\mathbf{G}\|_F \|\mathbf{H}\|_F. \end{aligned}$$

Now let  $\mathbf{Z} = (\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{W}^{T\dagger}$ , which gives  $\mathbf{Z}\mathbf{W}^T = (\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W$ . Here  $\dagger$  denotes the pseudoinverse of a matrix and  $\mathbf{P}_W$  is the orthogonal projector onto the range of  $\mathbf{W}$ . Utilizing the fact  $\widehat{\mathbf{W}}^T \mathbf{W} = \mathbf{0}$  from (7), we further connect the left hand side of (27) with  $\|(\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W\|_F^2$  by

$$\begin{aligned} \frac{\alpha+\beta}{2} \mathfrak{T}_2 &= \frac{\alpha+\beta}{2} \mathfrak{T}_2 + \frac{\alpha+\beta}{4} \langle \widehat{\mathbf{W}}\widehat{\mathbf{W}}^T, \mathbf{Z}\mathbf{W}^T \rangle \\ &= \frac{\alpha+\beta}{4} \langle \mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}, (\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W \rangle \\ &\quad + \frac{\alpha+\beta}{4} \langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, (\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W \rangle \\ &\geq \frac{\alpha+\beta}{4} \langle \mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}, (\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W \rangle \\ &= \frac{\alpha+\beta}{4} \|(\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W\|_F^2, \end{aligned} \quad (28)$$

where the inequality follows because  $\langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W}^* \mathbf{W}^{*T} \mathbf{P}_W \rangle = 0$  (noting that  $\widehat{\mathbf{W}}^{*T} \widehat{\mathbf{W}}^* = \mathbf{0}$ ) and  $\langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W}\mathbf{W}^T \mathbf{P}_W \rangle = \langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W}\mathbf{W}^T \rangle \geq 0$  since it is the inner product between two PSD matrices.

On the other hand, we give an upper bound on the right hand side of (27):

$$\begin{aligned} & \|\mathbf{X} - \mathbf{X}^*\|_F \|\mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T\|_F \\ & \leq \|\mathbf{X} - \mathbf{X}^*\|_F \sqrt{2 \|\mathbf{Z}_U \mathbf{V}^T\|_F^2 + 2 \|\mathbf{U} \mathbf{Z}_V^T\|_F^2} \\ & \leq \|\mathbf{X} - \mathbf{X}^*\|_F \|(\mathbf{W}\mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T}) \mathbf{P}_W\|_F, \end{aligned}$$

where the last line follows because  $\|\mathbf{Z}_U \mathbf{V}^T\|_F^2 + \|\mathbf{Z}_V \mathbf{U}^T\|_F^2 = \|\mathbf{Z}_U \mathbf{U}^T\|_F^2 + \|\mathbf{Z}_V \mathbf{V}^T\|_F^2$  (since  $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V}$ ), implying  $2 \|\mathbf{Z}_U \mathbf{V}^T\|_F^2 + 2 \|\mathbf{U} \mathbf{Z}_V^T\|_F^2 = \|\mathbf{Z}\mathbf{W}^T\|_F^2$ . This together with (27) and (28) completes the proof.  $\square$

## APPENDIX D PROOF OF LEMMA 3

*Proof of Lemma 3.* When  $\mathbf{C} \neq \mathbf{0}$ , the proof follows directly from the following results.

**Lemma 5.** [30, Lemma 3.4] For any matrices  $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times r}$  with rank  $r_1$  and  $r_2$ , respectively, let  $\mathbf{R} = \arg \min_{\tilde{\mathbf{R}} \in \mathcal{O}_r} \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F$ . Then

$$\|\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T\|_F \geq \min \{\sigma_{r_1}(\mathbf{C}), \sigma_{r_2}(\mathbf{D})\} \cdot \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F.$$

**Lemma 6.** [31, Lemma 5.4] For any matrices  $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times r}$  with rank( $\mathbf{D}$ ) =  $r$ , let  $\mathbf{R} = \arg \min_{\tilde{\mathbf{R}} \in \mathcal{O}_r} \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F$ . Then

$$\|\mathbf{C}\mathbf{C}^T - \mathbf{D}\mathbf{D}^T\|_F^2 \geq 2(\sqrt{2} - 1) \sigma_r^2(\mathbf{D}) \|\mathbf{C} - \mathbf{D}\mathbf{R}\|_F^2.$$

$$\begin{aligned} \mathfrak{T}_1 &= \\ & \int_0^1 [\nabla^2 f(t\mathbf{X} + (1-t)\mathbf{X}^*)] (\mathbf{X} - \mathbf{X}^*, \mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T) dt \\ &\quad - \frac{\alpha+\beta}{2} \langle \mathbf{X} - \mathbf{X}^*, \mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T \rangle. \end{aligned} \quad (26)$$

Noting that all the three matrices  $t\mathbf{X} + (1-t)\mathbf{X}^*$ ,  $\mathbf{X} - \mathbf{X}^*$  and  $\mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T$  have rank at most  $2r$ , it follows from Proposition 2 that

$$|\mathfrak{T}_1| \leq \frac{\beta-\alpha}{2} \|\mathbf{X} - \mathbf{X}^*\|_F \|\mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T\|_F,$$

which when plugged into (26) gives

$$\begin{aligned} \frac{\alpha+\beta}{2} \mathfrak{T}_2 &= -\mathfrak{T}_1 \\ &\leq \frac{\beta-\alpha}{2} \|\mathbf{X} - \mathbf{X}^*\|_F \|\mathbf{Z}_U \mathbf{V}^T + \mathbf{U} \mathbf{Z}_V^T\|_F. \end{aligned} \quad (27)$$

If  $C = 0$ , then we have

$$\begin{aligned} \|CC^T - DD^T\|_F^2 &= \|DD^T\|_F^2 = \sum_{i=1}^{r_2} \sigma_i^4(D) \\ &\geq \sigma_{r_2}^2(D) \sum_{i=1}^{r_2} \sigma_i^2(D) = \sigma_{r_2}^2(D) \|C - DR\|_F^2. \end{aligned}$$

$\square$

## APPENDIX E PROOF OF (20)

*Proof of (20).* We prove the upper bounds for the four terms as follows.

**Bounding term  $\Pi_1$ :** Utilizing the fact that  $\Delta_U = U - U^*R$  and  $\Delta_V = V - V^*R$ , we have

$$\begin{aligned} \Pi_1 &= \langle \nabla f(\mathbf{X}), \Delta_U \Delta_V^T \rangle \\ &= \langle \nabla f(\mathbf{X}), (U - U^*R)(V - V^*R)^T \rangle \\ &= \langle \nabla f(\mathbf{X}), \mathbf{X} + \mathbf{X}^* - U^*R^T V^T - U R^T V^{*T} \rangle \\ &\stackrel{(i)}{=} -\langle \nabla f(\mathbf{X}), \mathbf{X} - \mathbf{X}^* \rangle \\ &\stackrel{(ii)}{=} -\langle \nabla f(\mathbf{X}) - \nabla f(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^* \rangle \\ &\stackrel{(iii)}{\leq} -\alpha \|\mathbf{X} - \mathbf{X}^*\|_F^2, \end{aligned}$$

where (i) follows from (17) and (18), (ii) utilizes  $\nabla f(\mathbf{X}^*) = 0$ , and (iii) follows by using the  $(2r, 4r)$ -restricted strict convexity property (3):

$$\begin{aligned} &\langle \nabla f(\mathbf{X}) - \nabla f(\mathbf{X}^*), \mathbf{X} - \mathbf{X}^* \rangle \\ &= \int_0^1 [\nabla^2 f(t\mathbf{X} + (1-t)\mathbf{X}^*)] (\mathbf{X} - \mathbf{X}^*, \mathbf{X} - \mathbf{X}^*) dt \\ &\geq \int_0^1 \alpha \langle \mathbf{X} - \mathbf{X}^*, \mathbf{X} - \mathbf{X}^* \rangle dt \\ &= \alpha \|\mathbf{X} - \mathbf{X}^*\|_F^2, \end{aligned}$$

where the first line follows from the integral form of the mean value theorem for vector-valued functions, and the second line uses the fact that both  $t\mathbf{X} + (1-t)\mathbf{X}^*$  and  $\mathbf{X} - \mathbf{X}^*$  have rank at most  $2r$ , and the  $(2r, 4r)$ -restricted strong convexity of the Hessian  $\nabla^2 f(\cdot)$ .

**Bounding term  $\Pi_2$ :** By the smoothness condition (3), we have

$$\begin{aligned} \Pi_2 &= [\nabla^2 f(\mathbf{X})] (\Delta_U V^T + U \Delta_V^T, \Delta_U V^T + U \Delta_V^T) \\ &\leq \beta \|\Delta_U V^T + U \Delta_V^T\|_F^2 \\ &\leq 2\beta \left( \|\Delta_U V^T\|_F^2 + \|U \Delta_V^T\|_F^2 \right) \\ &= \beta \|\mathbf{W} \Delta^T\|_F^2, \end{aligned}$$

where the last line holds because  $\|DU^T\|_F^2 = \|DV^T\|_F^2$  for any  $D \in \mathbb{R}^{p \times r}$  with arbitrary  $p \geq 1$  since any critical point  $\mathbf{W}$  satisfies  $U^T U = V^T V$ .

**Bounding term  $\Pi_3$ :**

$$\begin{aligned} \Pi_3 &= \langle U \Delta_U^T, \Delta_U U^T \rangle + \langle V \Delta_V^T, \Delta_V V^T \rangle - 2 \langle U \Delta_V^T, \Delta_U V^T \rangle \\ &\leq \|\Delta_U^T\|_F^2 + \|\Delta_V^T\|_F^2 + \|\Delta_U^T\|_F^2 + \|\Delta_V^T\|_F^2 \\ &= \|\mathbf{W} \Delta^T\|_F^2. \end{aligned}$$

**Bounding term  $\Pi_4$ :**

$$\begin{aligned} \Pi_4 &= \left\langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, (\mathbf{W} - \mathbf{W}^* R) (\mathbf{W} - \mathbf{W}^* R)^T \right\rangle \\ &\stackrel{(i)}{=} -\left\langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, \mathbf{W} \mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T} \right\rangle \\ &\stackrel{(ii)}{\leq} -\left\langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T, \mathbf{W} \mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T} \right\rangle \\ &\quad + \left\langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W} \mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T} \right\rangle \\ &= -\left\langle \widehat{\mathbf{W}} \widehat{\mathbf{W}}^T - \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W} \mathbf{W}^T - \mathbf{W}^* \mathbf{W}^{*T} \right\rangle \\ &\leq 2 \|\mathbf{X} - \mathbf{X}^*\|_F^2, \end{aligned}$$

where (i) holds because  $\widehat{\mathbf{W}}^T \mathbf{W} = \mathbf{0}$ , and (ii) follows because  $\widehat{\mathbf{W}}^{*T} \mathbf{W}^* = \mathbf{0}$  and  $\langle \widehat{\mathbf{W}}^* \widehat{\mathbf{W}}^{*T}, \mathbf{W} \mathbf{W}^T \rangle \geq 0$  since it is the inner product between two PSD matrices.  $\square$

## APPENDIX F

### PROOF OF (22)

*Proof of (22).* To show (22), expanding the left hand side of (22), it is equivalent to show

$$\|UU^T - U^*U^{*T}\|_F^2 + \|VV^T - V^*V^{*T}\|_F^2 \leq 2 \|\mathbf{X} - \mathbf{X}^*\|_F^2.$$

Expanding both sides of the above equation and utilizing the fact  $U^T U = V^T V$  and  $U^{*T} U^* = V^{*T} V^*$ , the remaining step is to show

$$\begin{aligned} &\text{trace} (UU^T U^* U^{*T}) + (VV^T V^* V^{*T}) \\ &\geq 2 \text{trace} (UV^T V^* U^{*T}). \end{aligned}$$

Thus, we obtain (22) by noting that the above equation is equivalent to

$$\text{trace} \left( (U^{*T} U - V^{*T} V)^2 \right) \geq 0.$$

$\square$

## REFERENCES

- [1] Z. Zhu, Q. Li, G. Tang, and M. B. Wakin, “Global optimality in low-rank matrix optimization,” in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1275–1279, Nov 2017.
- [2] S. Aaronson, “The learnability of quantum states,” in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 463, pp. 3089–3114, The Royal Society, 2007.
- [3] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, “Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators,” *New Journal of Physics*, vol. 14, no. 9, p. 095022, 2012.
- [4] N. Srebro, J. Rennie, and T. S. Jaakkola, “Maximum-margin matrix factorization,” in *Advances in Neural Information Processing Systems*, pp. 1329–1336, 2004.

[5] D. DeCoste, “Collaborative prediction using ensembles of maximum margin matrix factorizations,” in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 249–256, ACM, 2006.

[6] P. Biswas and Y. Ye, “Semidefinite programming for ad hoc wireless sensor network localization,” in *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pp. 46–54, ACM, 2004.

[7] G. Tang and A. Nehorai, “Lower bounds on the mean-squared error of low-rank matrix reconstruction,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4559–4571, 2011.

[8] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[9] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[10] G. Liu and P. Li, “Low-rank matrix completion in the presence of high coherence,” *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5623–5633, 2016.

[11] M. Fazel, H. Hindi, and S. Boyd, “Rank minimization and applications in system theory,” in *American Control Conference*, vol. 4, pp. 3273–3278, IEEE, 2004.

[12] E. J. Candès and Y. Plan, “Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[13] Z. Harchaoui, M. Douze, M. Paulin, M. Dudik, and J. Malick, “Large-scale image classification with trace-norm regularization,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3386–3393, 2012.

[14] M. A. Davenport and J. Romberg, “An overview of low-rank matrix recovery from incomplete observations,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 608–622, 2016.

[15] K. Mohan and M. Fazel, “Reweighted nuclear norm minimization with application to system identification,” in *Proceedings of the 2010 American Control Conference*, pp. 2953–2959, IEEE, 2010.

[16] S. Burer and R. D. Monteiro, “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization,” *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[17] S. Burer and R. D. Monteiro, “Local minima and convergence in low-rank semidefinite programming,” *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, 2005.

[18] S. Bhojanapalli, B. Neyshabur, and N. Srebro, “Global optimality of local search for low rank matrix recovery,” in *Advances in Neural Information Processing Systems*, pp. 3873–3881, 2016.

[19] R. Ge, J. D. Lee, and T. Ma, “Matrix completion has no spurious local minimum,” in *Advances in Neural Information Processing Systems*, pp. 2973–2981, 2016.

[20] D. Park, A. Kyrillidis, C. Carmanis, and S. Sanghavi, “Non-square matrix sensing without spurious local minima via the burer-monteiro approach,” in *Artificial Intelligence and Statistics*, pp. 65–74, 2017.

[21] L. Wang, X. Zhang, and Q. Gu, “A unified computational and statistical framework for nonconvex low-rank matrix estimation,” in *Artificial Intelligence and Statistics*, pp. 981–990, 2017.

[22] J. Sun, Q. Qu, and J. Wright, “A geometric analysis of phase retrieval,” *Foundations of Computational Mathematics*, pp. 1–68.

[23] 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*, pp. 797–842, 2015.

[24] M. Udell, C. Horn, R. Zadeh, S. Boyd, *et al.*, “Generalized low rank models,” *Foundations and Trends® in Machine Learning*, vol. 9, no. 1, pp. 1–118, 2016.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[26] T. Bouwmans, N. S. Aybat, and E.-h. Zahzah, *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, 2016.

[27] M. A. Davenport, Y. Plan, E. van den Berg, and M. Wootters, “1-bit matrix completion,” *Information and Inference*, vol. 3, no. 3, pp. 189–223, 2014.

[28] T. Cai and W.-X. Zhou, “A max-norm constrained minimization approach to 1-bit matrix completion.,” *Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3619–3647, 2013.

[29] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, “Poisson noise reduction with non-local PCA,” *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.

[30] Q. Li, Z. Zhu, and G. Tang, “The non-convex geometry of low-rank matrix optimization,” *Information and Inference: A Journal of the IMA*, p. iay003, 2018.

[31] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, “Low-rank solutions of linear matrix equations via procrustes flow,” in *International Conference on International Conference on Machine Learning (ICML)*, pp. 964–973, 2016.

[32] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, “Gradient descent converges to minimizers,” *University of California, Berkeley*, vol. 1050, p. 16, 2016.

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

[34] P. Jain, P. Netrapalli, and S. Sanghavi, “Low-rank matrix completion using alternating minimization,” in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 665–674, ACM, 2013.

[35] E. J. Candès, X. Li, and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[36] S. Bhojanapalli, A. Kyrillidis, and S. Sanghavi, “Dropping convexity for faster semi-definite optimization,” in *Conference on Learning Theory*, pp. 530–582, 2016.

[37] J. Sun, Q. Qu, and J. Wright, “Complete dictionary recovery over the sphere I: Overview and the geometric picture,” *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 853–884, 2017.

[38] Y. Li, Y. Sun, and Y. Chi, “Low-rank positive semidefinite matrix recovery from corrupted rank-one measurements,” *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 397–408, 2017.

[39] L. Xu and M. Davenport, “Dynamic matrix recovery from incomplete observations under an exact low-rank constraint,” in *Advances in Neural Information Processing Systems*, pp. 3585–3593, 2016.

[40] X. Li, Z. Wang, J. Lu, R. Arora, J. Haupt, H. Liu, and T. Zhao, “Symmetry, saddle points, and global geometry of nonconvex matrix factorization,” *arXiv preprint arXiv:1612.09296*, 2016.

[41] Z. Zhu, Q. Li, G. Tang, and M. B. Wakin, “The global optimization geometry of nonsymmetric matrix factorization and sensing,” *arXiv preprint arXiv:1703.01256*, 2017.

[42] R. Ge, C. Jin, and Y. Zheng, “No spurious local minima in nonconvex low rank problems: A unified geometric analysis,” *arXiv preprint arXiv:1704.00708*, 2017.

[43] A. R. Conn, N. I. Gould, and P. L. Toint, *Trust region methods*. SIAM, 2000.

[44] N. Srebro, T. Jaakkola, *et al.*, “Weighted low-rank approximations,” in *International Conference on Machine Learning*, vol. 3, pp. 720–727, 2003.

[45] N. Gillis and F. Glineur, “Low-rank matrix approximation with weights or missing data is NP-hard,” *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1149–1165, 2011.

[46] S. A. Bhaskar and A. Javanmard, “1-bit matrix completion under exact low-rank constraint,” in *49th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2015.

[47] C. R. Johnson, K. Okubo, and R. Reams, “Uniqueness of matrix square roots and an application,” *Linear Algebra and Its Applications*, vol. 323, no. 1, pp. 51–60, 2001.

[48] P. Jain, R. Meka, and I. S. Dhillon, “Guaranteed rank minimization via singular value projection,” in *Advances in Neural Information Processing Systems*, pp. 937–945, 2010.

[49] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Math. Program.*, vol. 45, no. 1, pp. 503–528, 1989.

[50] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

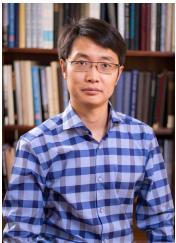
[51] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9–10, pp. 589–592, 2008.



**Zhihui Zhu** (S'17) is a Postdoctoral Fellow in the Center for Imaging Science (CIS) at the Johns Hopkins University. He obtained a B.Eng. in communications engineering in 2012 from Zhejiang University of Technology and a Ph.D. in electrical engineering in 2017 from Colorado School of Mines. He was the Meritorious winner of MCM/ICM in 2011. His research interests include exploiting inherent structures and applying optimization methods with guaranteed performance for signal processing, machine learning and data analysis.



**Qiuwei Li** (S'17) is currently a PhD student at the Department of Electrical Engineering, Colorado School of Mines, CO, USA. His research interests include using convex and nonconvex optimization methods to model and solve problems in signal and information processing. He is especially interested in designing optimization procedures that come with theoretical performance guarantees and that are scalable to large data sets.



**Gongguo Tang** (S'09-M'11) received his Ph.D. degree in Electrical Engineering from Washington University in St. Louis in 2011, and was a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Wisconsin-Madison in 2011-2013.

He is an Assistant Professor at Colorado School of Mines. His research focuses on using convex and non-convex optimization methods to model and solve problems in signal and information processing, physical sciences, and machine learning. He is designing optimization procedures that come with theoretical performance guarantees and are scalable to large data sets.

especially interested in theoretical performance guarantees and are scalable to large data sets.



**Michael B. Wakin** (S'01-M'07-SM'13) is an Associate Professor in the Department of Electrical Engineering at the Colorado School of Mines. Dr. Wakin received a B.S. in electrical engineering and a B.A. in mathematics in 2000 (summa cum laude), an M.S. in electrical engineering in 2002, and a Ph.D. in electrical engineering in 2007, all from Rice University. He was an NSF Mathematical Sciences Postdoctoral Research Fellow at Caltech from 2006-2007, an Assistant Professor at the University of Michigan from 2007-2008, and a Ben L. Fryrear

Associate Professor at Mines from 2015-2017. His research interests include sparse, geometric, and manifold-based models for signal processing and compressive sensing.

In 2007, Dr. Wakin shared the Hershel M. Rich Invention Award from Rice University for the design of a single-pixel camera based on compressive sensing. In 2008, Dr. Wakin received the DARPA Young Faculty Award for his research in compressive multi-signal processing for environments such as sensor and camera networks. In 2012, Dr. Wakin received the NSF CAREER Award for research into dimensionality reduction techniques for structured data sets. In 2014, Dr. Wakin received the Excellence in Research Award for his research as a junior faculty member at Mines. Dr. Wakin is a recipient of the Best Paper Award from the IEEE Signal Processing Society. He has served as an Associate Editor for IEEE Signal Processing Letters and is currently an Associate Editor for IEEE Transactions on Signal Processing.