UC Berkeley
UC Berkeley Previously Published Works

Title

Semidefinite Programming versus Burer-Monteiro Factorization for Matrix Sensing

Permalink

https://escholarship.org/uc/item/03g103hq

Authors

Yalcin, Baturalp

Ma, Ziye

Lavaei, Javad

et al.

Publication Date

2023

Copyright Information

Peer reviewed

Semidefinite Programming versus Burer-Monteiro Factorization for Matrix Sensing

Baturalp Yalçın[*1], **Ziye Ma**[*2], **Javad Lavaei**[1], **Somayeh Sojoudi**[2]

[1] UC Berkeley, Industrial Engineering and Operations Research
[2] UC Berkeley, Electrical Engineering and Computer Science
{baturalp_yalcin, ziyema, lavaei, sojoudi}@berkeley.edu

Abstract

Many fundamental low-rank optimization problems, such as matrix completion, phase retrieval, and robust PCA, can be formulated as the matrix sensing problem. Two main approaches for solving matrix sensing are based on semidefinite programming (SDP) and Burer-Monteiro (B-M) factorization. The former suffers from high computational and space complexities, whereas the latter may return a spurious solution due to the non-convexity of the problem. The existing theoretical guarantees for the success of these methods have led to similar conservative conditions, which may wrongly imply that these methods have comparable performances. In this paper, we shed light on some major differences between these two methods. First, we present a class of *structured* matrix completion problems for which the B-M methods fail with an overwhelming probability, while the SDP method works correctly. Second, we identify a class of *highly sparse* matrix completion problems for which the B-M method works and the SDP method fails. Third, we prove that although the B-M method exhibits the same performance independent of the rank of the unknown solution, the success of the SDP method is correlated to the rank of the solution and improves as the rank increases. Unlike the existing literature that has mainly focused on those instances of matrix sensing for which both SDP and B-M work, this paper offers the first result on the unique merit of each method over the alternative approach.

1 Introduction

Low-rank matrix recovery problems have ubiquitous applications in machine learning and data analytics, including collaborative filtering (Koren, Bell, and Volinsky 2009), phase retrieval (Candes et al. 2015; Singer 2011; Boumal 2016; Shechtman et al. 2015), motion detection (Fattahi and Sojoudi 2020), and power system state estimation (Jin et al. 2020; Zhang, Madani, and Lavaei 2017; Jin et al. 2019). This problem is formally defined as follows: Given a measurement operator $\mathcal{A}(\cdot) : \mathbb{R}^{m \times n} \mapsto \mathbb{R}^d$ returning a d-dimensional measurement vector $\mathcal{A}(\mathbf{M}^*)$ from a low-rank ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{m \times n}$ with rank r, the goal is to obtain a matrix with rank less than equal to r that conforms with the measurements, preferably the ground truth matrix $\mathbf{M}^*$. This

problem can be stated as the feasibility problem

$$
\begin{aligned}
\text{find} \quad & \mathbf{M} \in \mathbb{R}^{m \times n} \qquad\qquad\qquad (1)\\
\text{s.t.} \quad & \mathcal{A}(\mathbf{M}) = \mathcal{A}(\mathbf{M}^*)\\
& \text{rank}(\mathbf{M}) \leq r.
\end{aligned}
$$

While the measurement operator $\mathcal{A}$ can be nonlinear as in the case of one-bit matrix sensing (Davenport et al. 2014) and phase retrieval (Shechtman et al. 2015), matrix sensing and matrix completion that are widely studied have linear measurement operators (Candès and Recht 2009; Recht, Fazel, and Parrilo 2010). We focus on the matrix sensing and matrix completion problems throughout this paper. Despite the linearity of $\mathcal{A}$, there are two types of problems depending on the structure of the ground truth matrix $\mathbf{M}^*$. The first type, symmetric problem, consists of a low-rank positive semidefinite ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n \times n}$, whereas the second type, asymmetric problem, consists of a ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{m \times n}$ that is possibly sign indefinite and non-square. Since each asymmetric problem can be converted to an equivalent symmetric problem (Zhang, Bi, and Lavaei 2021a), we study only the symmetric problem in this paper.

The matrix sensing and completion problems have linear measurements; hence, the first constraint in problem (1) is linear. Therefore, the only nonconvexity of the problem arises from the nonconvex rank constraint. Earlier works on these problems focused on their convex relaxations by penalizing high-rank solutions (Candès and Recht 2009; Recht, Fazel, and Parrilo 2010; Candès and Tao 2010). They utilized the nuclear norm of a matrix as the convex surrogate of the rank function. This led to semidefinite programming (SDP) relaxations, which solve the original non-convex problems exactly with high probability based on some assumptions on the linear measurement operator and the ground truth matrix, such as the Restricted Isometry Property (RIP) and incoherence conditions. High computational time and storage requirements of the SDP algorithms incentivized the implementation of the B-M factorization approach (Burer and Monteiro 2003). This approach factorizes the symmetric matrix variable $\mathbf{M} \in \mathbb{R}^{n \times n}$ as $\mathbf{M} = \mathbf{X}\mathbf{X}^T$ for some matrix $\mathbf{X} \in \mathbb{R}^{n \times r}$, which obviates imposing the positive semidefiniteness and rank constraints. Although the dimension of the decision variable reduces dramatically when r is small, the problem is still nonconvex since its objective function is nonconvex in terms of the factorized $\mathbf{X}$.

Problem Formulation

Formally, the SDP formulation of the matrix sensing problem uses the nuclear norm of the variable, $\|\mathbf{M}\|_*$, to serve as a surrogate of the rank, and replaces the rank constraint in (1) with an objective to minimize $\|\mathbf{M}\|_*$. Due to the symmetricity and positive semidefiniteness of the variable, the nuclear norm is equivalent to the trace of the matrix variable $\mathbf{M}$. Hence, the SDP formulation can be written as

$$\min_{\mathbf{M} \in \mathbb{R}^{n \times n}} \quad \operatorname{tr}(\mathbf{M}) \qquad \text{s.t. } \mathcal{A}(\mathbf{M}) = b, \ \mathbf{M} \succeq 0, \quad (2)$$

where $b = \mathcal{A}(\mathbf{M}^*) = [\langle \mathbf{A}_1, \mathbf{M}^* \rangle, \ldots, \langle \mathbf{A}_d, \mathbf{M}^* \rangle]^T$ is given and $\{\mathbf{A}_i\}_{i=1}^d \in \mathbb{R}^{n \times n}$ are called sensing matrices. Moreover, the matrix completion problem is a special case of the matrix sensing problem with each sensing matrix measuring only one entry of $\mathbf{M}^*$. We can represent the measurement operator $\mathcal{A}$ as $\mathcal{A}_\Omega : \mathbb{R}^{n \times n} \mapsto \mathbb{R}^{n \times n}$ for this special case, which is defined as follows:

$$\mathcal{A}_\Omega(\mathbf{M})_{ij} := \begin{cases} \mathbf{M}_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise,} \end{cases}$$

where Ω is the set of indices of observed entries. We denote the measurement operator as $\mathbf{M}_\Omega := \mathcal{A}_\Omega(\mathbf{M})$ for simplicity. Besides the SDP formulation, the B-M factorization formulation of the matrix sensing (MS) and matrix completion (MC) problems can be stated as

$$(\text{MS}) \quad \min_{\mathbf{X} \in \mathbb{R}^{n \times r}} g\left[\mathcal{A}(\mathbf{X}\mathbf{X}^T) - b\right], \qquad (3a)$$

$$(\text{MC}) \quad \min_{\mathbf{X} \in \mathbb{R}^{n \times r}} g\left[(\mathbf{X}\mathbf{X}^T - \mathbf{M}^*)_\Omega\right], \qquad (3b)$$

where $g(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}$ is some twice continuously differentiable function such that $\mathbf{0}_{n \times n}$ is its unique minimizer and the Hessian of $g(\cdot)$ is positive definite at $\mathbf{0}_{n \times n}$. These assumptions are satisfied by the common loss functions considered in the literature. The main objective of this paper is to compare the SDP and B-M methods for the MC and MS problems.

Background and Related Work

SDP formulation (2) can be used to solve the matrix sensing problem if the sensing matrices are sampled independently from a sub-Gaussian distribution and the number of measurements d is large enough (Recht, Fazel, and Parrilo 2010; Recht, Xu, and Hassibi 2008). This is also a sufficient condition for the sensing matrices to satisfy the RIP condition with high probability, which is defined below:

Definition 1 (RIP). *(Candès and Recht 2009) The linear map $\mathcal{A} : \mathbb{R}^{n \times n} \mapsto \mathbb{R}^m$ is said to satisfy δ_p-RIP if there is a constant $\delta_p \in [0, 1)$ such that*

$$(1 - \delta_p)\|\mathbf{M}\|_F^2 \leq \|\mathcal{A}(\mathbf{M})\|^2 \leq (1 + \delta_p)\|\mathbf{M}\|_F^2$$

holds for all matrices $\mathbf{M} \in \mathbb{R}^{n \times n}$ satisfying $\operatorname{rank}(\mathbf{M}) \leq p$.

The RIP constant δ_p represents how similar the linear operator $\mathcal{A}$ is to an isometry, and various upper bounds on δ_p have been proposed to serve as sufficient conditions for the exact recovery (meaning that one can recover the ground truth $\mathbf{M}^*$ by solving the SDP problem). A few notable ones include $\delta_{4r} < \sqrt{2} - 1$ in (Candes and Plan 2010),

$\delta_{5r} < 0.607, \delta_{3r} < 0.472$ in (Mohan and Fazel 2010), and $\delta_{2r} < 1/2, \delta_{3r} < 1/3$ in (Cai and Zhang 2013). On the other hand, when the sensing matrices are not sampled independently from a sub-Gaussian distribution or when the RIP condition is not met, the SDP formulation may still recover the ground truth matrix with a high probability. This is the case for MC problems for which RIP fails to hold while SDP works as long as entries of observation follow a Bernoulli model (Candès and Recht 2009; Candès and Tao 2010).

However, recent works have shown that if we use the B-M method instead of the SDP approach, we can still recover the ground truth matrix via first-order methods under similar RIP or coherence assumptions in both the matrix sensing and matrix completion cases (Ge, Jin, and Zheng 2017; Bhojanapalli, Neyshabur, and Srebro 2016; Park et al. 2017; Zhang et al. 2018; Zhu et al. 2018; Zhang, Sojoudi, and Lavaei 2019; Zhang and Zhang 2020; Bi and Lavaei 2021; Ha, Liu, and Barber 2020; Zhu et al. 2021; Zhang 2021; Zhang, Bi, and Lavaei 2021b; Ma et al. 2022; Ma and Sojoudi 2022). Namely, the state-of-the-art result states that as long as $\delta_{\tilde{r}+r} < 1/2$ for the matrix sensing problem, there exists no spurious local minima for an over-parametrized B-M formulation and the gradient descent algorithm can recover $\mathbf{M}^*$ exactly (Zhang 2021). Here, $\tilde{r} \geq r$ is the search rank that we choose manually in the B-M formulation. If we know the value of r, we can set $\tilde{r}$ to r, making the B-M approach enjoy the same RIP guarantee as the SDP approach. Since the B-M approach enjoys far better scalability, it has become an increasingly popular tool for solving the matrix sensing problem.

Nevertheless, the B-M approach cannot be routinely used without careful consideration since it could fail on easy (from an information-theoretic perspective) instances of the problem as demonstrated in (Yalçın et al. 2022), especially in cases when the RIP condition is not satisfied.

Thus, it is important to compare and contrast both the SDP and B-M approaches to discover which method is superior to the other one. This comparison is timely since specialized sparse SDP algorithms have become more efficient in recent years, making the SDP method more practical than before (Zhang and Lavaei 2021; Yurtsever et al. 2017, 2021). In this paper, we show that the SDP approach is more powerful than the B-M method as far as the RIP measure is concerned. We also discover that the B-M method is able to solve certain instances for which the SDP approach fails. This means that none of these techniques is universally better than the other one and the best technique should be chosen based on the nature of the problem. This work provides the first step towards understanding the trade-off between a well-known convex relaxation and first-order descent algorithms applied to the B-M factorization formulation.

Our Motivation and Contributions

This research problem is motivated by the preliminary simulations conducted. Our goal is to understand when the B-M factorization and SDP formulation methods can successfully recover the ground-truth matrix and delineate the instances for which one would succeed whereas the other would fail. The results are presented in Tables 1 and 2. We generated synthetic matrix completion problems with symmetric pos-

p	%B-M Works SDP Fails	%SDP Works B-M Fails	% Both Work
0.1	0 %	0 %	0 %
0.2	0 %	9 %	1 %
0.3	1 %	24 %	16 %
0.4	1 %	27 %	41 %
0.5	1 %	21 %	63 %
0.6	1 %	16 %	77 %
0.7	1 %	11 %	86 %
0.8	5 %	4 %	78 %
0.9	13 %	2 %	85 %

Table 1: Success Rate of the B-M and SDP Methods for Random Observations for Rank-1 Ground Truth Matrices

p	%B-M Works SDP Fails	%SDP Works B-M Fails	% Both Work
0.1	0 %	0 %	0 %
0.2	0 %	0 %	0 %
0.3	0 %	6 %	0 %
0.4	0 %	22 %	9 %
0.5	0 %	23%	30 %
0.6	1 %	25 %	46 %
0.7	1 %	20 %	64 %
0.8	4 %	14 %	76 %
0.9	11 %	5 %	82 %

Table 2: Success Rate of the B-M and SDP Methods for Random Observations for Rank-2 Ground Truth Matrices

itive semidefinite ground truth matrices with dimensions $n = 5, n = 10$, and $n = 20$. An entry of the ground truth matrix is observed independently with probability p. 1000 different instances are constructed for each p and n. The B-M factorization problem is solved with a gradient descent algorithm and SDP formulation is solved by using the CVX solver. We report the percentage of times each method successfully recovers the ground truth matrix and we take the average of the success rate for various problem dimensions n. As seen from the tables, there are various instances where one method would work successfully while the other one fails. Therefore, our goal is to understand why and when these situations arise. We provide a comparative analysis between the SDP approach and the B-M method. We first present the advantages of the SDP approach over the B-M method:

1. First, we focus on an important class of MC problems recently studied in (Yalçın et al. 2022). That paper has shown that even though this class has low information-theoretic complexity, the B-M method would utterly fail and the probability of success via first-order methods is almost zero. We prove that the SDP method successfully solves this class and, therefore, SDP may not suffer from the unusual behavior of B-M with regard to easy instances of MC. This also implies that the information-theoretic and optimization complexities are expected to be more aligned for SDP than B-M.

2. We investigate a class of MS problems found in the recent paper (Zhang et al. 2022). Each MS instance belonging to this class satisfies δ_2-RIP with $r = 1$ for some $\delta > 1/2$

such that the B-M formulation leads to $\mathcal{O}((1 - \delta)^{-1})$ spurious solutions and this number goes to infinity as δ approaches 1. We show that although each instance is extremely non-convex based on the number of spurious solutions, the SDP method successfully solves all of the problems in this class. This implies that, unlike the B-M method, the success of the SDP approach is not directly correlated to the presence of many spurious solutions.

3. The recent paper (Zhang, Bi, and Lavaei 2021a) has shown that the sharpest RIP bound for the success of the B-M method on the MS problem is $1/2$ and this is independent of the rank r. This is an undesirable result since high-rank problems have lower information-theoretic complexity than low-rank problems. We derive a sufficient RIP bound for the SDP method and show that it can increase from $1/2$ to 1 as the rank r becomes larger. This implies that the SDP approach does not suffer from a major shortcoming of the B-M method.

Despite the above advantages, we show that the SDP approach is not universally better than the B-M method. To prove this, we identify a class of MC problems with $\mathcal{O}(n)$ observations in the rank-1 case for which B-M works while SDP fails. It is clear from these comparisons that although the B-M approach is known to be more powerful due to its scalability property, the SDP approach enjoys some unique merits and deserves to be revisited, especially in light of the advancements of fast SDP solvers (Zhang and Lavaei 2021; Yurtsever et al. 2017, 2021)

2 Notations

$[n]$ represents the set of integers from 1 to n. Lower-case bold letters, namely $\mathbf{x}$, represent vectors and capital bold letters, namely $\mathbf{X}$, represent matrices. $\mathbf{I}_n$ and $\mathbf{0}_{n \times n}$ refer to the identity matrix and zero matrix of size $n \times n$, respectively. $\|\mathbf{x}\|$ denotes the Euclidean norm of $\mathbf{x}$, $\|\mathbf{X}\|$ and $\|\mathbf{X}\|_F$ are the 2-norm and the Frobenius norm of $\mathbf{X}$, respectively. For $\mathbf{x}$, $[\mathbf{x}]_{\mathbf{i}}$ denotes the i-th entry and $[\mathbf{x}]_{\mathbf{i:j}}$ denotes the subvector of entries from index i to index j for $i < j$. For $\mathbf{X}$, $[\mathbf{X}]_{\mathbf{i:j,k:l}}$ denotes the submatrix with rows between i and j and columns between k and l with $i < j$ and $k < l$. Let $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^T \mathbf{B})$ be the inner product between matrices. For a matrix $\mathbf{X}$, $\text{vec}(\mathbf{X})$ is the usual vectorization operation by stacking the columns of the matrix $\mathbf{X}$ into a vector. For a vector $\mathbf{x} \in \mathbb{R}^{n^2}$, $\text{mat}(\mathbf{x})$ converts $\mathbf{x}$ to a square matrix and $\text{mat}_S(\mathbf{x})$ converts $\mathbf{x}$ to a symmetric matrix, i.e., $\text{mat}(\mathbf{x}) = \mathbf{X}$ and $\text{mat}_S(\mathbf{x}) = (\mathbf{X} + \mathbf{X^T})/\mathbf{2}$, where $\mathbf{X} \in \mathbb{R}^{n \times n}$ is the unique matrix satisfying $\mathbf{x} = \text{vec}(\mathbf{X})$. The notations $\mathbf{X} \succeq 0$ and $\mathbf{X} \succ 0$ mean that the matrix $\mathbf{X}$ is positive semidefinite (PSD) and positive definite (PD), respectively. We use $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ to denote the ceiling and floor operators, respectively. The cardinality of a set $\mathcal{S}$ is shown as $|\mathcal{S}|$.

3 Advantages of the SDP Approach

B-M Fails While SDP Succeeds

In this section, we focus on a class of MC instances that was first proposed in (Yalçın et al. 2022) for which the B-M factorization fails. We focus on the matrix completion problem since it is the most common special case of the matrix

sensing problem that does not satisfy the RIP condition. We show that while the B-M approach fails to recover $\mathbf{M}^*$, the SDP approach can provably find $\mathbf{M}^*$.

We first give an introduction to this class of MC instances. Consider a rank-r ground truth matrix $\mathbf{M}^* \in \mathbb{S}_+^n$ with $r \geq 1$ and $n \geq 2r$. Let $m := n/r$ and assume without the loss of generality that n is divisible by r. We decompose the ground truth matrix into blocks of dimension $r \times r$; thus, $\mathbf{M}^*$ is an $m \times m$ block matrix whose block element at the position (i,j) is denoted as $\mathbf{M}_{i,j}^*$ for $i,j \in [m]$. We require some graph-theoretic notions before introducing the underlying class of MC instances.

Definition 2 (Induced Measurements). *Let $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2) = (\mathcal{V}, \mathcal{E}_1, \mathcal{E}_2)$ be a pair of undirected graphs with the node set $\mathcal{V} = [m]$ and the disjoint edge sets $\mathcal{E}_1, \mathcal{E}_2 \subset [m] \times [m]$, respectively. The induced measurement set $\Omega(\mathcal{G})$ is defined as follows: if $(i,j) \in \mathcal{E}_1$, then the entire block $\mathbf{M}_{i,j}^*$ is observed; if $(i,j) \in \mathcal{E}_2$, then all nondiagonal entries of the block $\mathbf{M}_{i,j}^*$ are observed; otherwise, none of the entries of the block is observed. $\mathcal{G}$ is referred to as the block sparsity graph.*

We represent the general problem (1) with the linear measurement operator $\mathcal{A}$ and rank-r ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n \times n}$ as $\mathcal{P}_{\mathbf{M}^*, \mathcal{A}, n, r}$. If this is a matrix completion problem with the measurement set Ω, then this special case of the same problem is denoted as $\mathcal{P}_{\mathbf{M}^*, \Omega, n, r}$. Based on this and Definition 2, a low-complexity class of MC instances will be introduced. These instances have a low complexity because graph-theoretical algorithms can solve them in polynomial time in terms of n and r.

Definition 3 (Low-complexity class of MC instances). *Define $\mathcal{L}(\mathcal{G}, n, r)$ to be the class of low-complexity MC instances $\mathcal{P}_{\mathbf{M}^*, \Omega, n, r}$ with the following properties:*
i) *The ground truth $\mathbf{M}^* \in \mathbb{S}_+^n$ is rank-r.*
ii) *The matrix $\mathbf{M}_{i,j}^* \in \mathbb{R}^{r \times r}$ is rank-r for all $i, j \in [m]$.*
iii) *The measurement set $\Omega = \Omega(\mathcal{G})$ is induced by $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2)$, where $\mathcal{G}_1$ is connected, non-bipartite, and its vertices have self-loops.*

The next theorem borrowed from (Yalçın et al. 2022) illustrates the failure of the B-M factorization method.

Theorem 1. *Consider a maximal independent set $\mathcal{S}(\mathcal{G}_1)$ of $\mathcal{G}_1$ such that the induced subgraph by vertices in $\mathcal{S}$, $\mathcal{G}_2[\mathcal{S}]$, is connected. There exists an instance in $\mathcal{L}(\mathcal{G}, n, r)$ for which the problem (3b) has at least $2^{r|\mathcal{S}(\mathcal{G}_1)|} - 2^r$ spurious local minima. In addition, the randomly initialized gradient descent algorithm converges to a global minimum with probability at most $\mathcal{O}(2^{-r|\mathcal{S}(\mathcal{G}_1)|})$, while there is a graph-theoretical algorithm that can solve the problem in $\mathcal{O}(n^2/r^2 + nr^2)$ time.*

The proof of Theorem 1 utilizes the Implicit Function Theorem (IFT). Specifically, their work has generated ground truth matrices $\mathbf{M}^*$ for which the B-M method has $2^{r|\mathcal{S}(\mathcal{G}_1)|}$ global solutions and only 2^r of them correspond to the correct completion of the $\mathbf{M}^*$. A generic small perturbation of the problem results in a new instance of an MC problem that belongs to the low-complexity class of MC instances. The conditions on $\mathcal{G}_1$ guarantee that the perturbed problem belongs to the low-complexity class, while the conditions

on $\mathcal{G}_2$ guarantee that the Hessian of the objective function of the unperturbed problem is PD at the global solutions. Since the instances in the low-complexity class are well defined, the new perturbed problem has a unique completion with 2^r possible global solutions for the B-M method. On the other hand, the other stationary points that correspond to global solutions of the unperturbed problem must be spurious local minima of the new instance. This is concluded by using the IFT. The perturbation that yields a new instance in the low-complexity class of the MC problem is achieved by perturbing the ground truth matrix $\mathbf{M}^* = \mathbf{X}^*(\mathbf{X}^*)^T$ by a small and generic perturbation $\epsilon \in \mathbb{R}^{n \times r}$. The new ground truth matrix is $\mathbf{M}^*(\epsilon) = \mathbf{X}^*(\epsilon)(\mathbf{X}^*(\epsilon))^T$, where $\mathbf{X}^*(\epsilon)_i = \mathbf{X}_i^* + \epsilon_i$ if $i \in \mathcal{S}(\mathcal{G}_1)$ and $\mathbf{X}^*(\epsilon)_i = \epsilon_i$ otherwise and $\text{rank}(\mathbf{X}_i^*) = \text{rank}(\mathbf{X}_i^* + \epsilon_i) = r, \forall i \in [m]$. A generic perturbation ϵ does not belong to a measure zero set in $\mathbb{R}^{n \times r}$.

It is desirable to study how the SDP method performs on this low-complexity class of MC instances. We will present the result for a larger class of problems that contains all instances discussed in Theorem 1.

Theorem 2. *Given $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2) = (\mathcal{V}, \mathcal{E}_1, \mathcal{E}_2)$, consider any maximal independent set $\mathcal{S}(\mathcal{G}_1)$. Consider also $\mathbf{M}^*(\epsilon) = \mathbf{X}^*(\epsilon)(\mathbf{X}^*(\epsilon))^T$ for any arbitrary $\epsilon \in \mathbb{R}^{n \times r}$, where $\mathbf{X}^*(\epsilon)_i = \mathbf{X}_i^* + \epsilon_i$ if $i \in \mathcal{S}(\mathcal{G}_1)$ and $\mathbf{X}^*(\epsilon)_i = \epsilon_i$ otherwise and $\text{rank}(\mathbf{X}_i^*) = \text{rank}(\mathbf{X}_i^* + \epsilon_i) = r, \forall i \in [m]$. The SDP formulation (2) with the observation set Ω induced by $\mathcal{G}_1$ uniquely recovers the ground truth matrix $\mathbf{M}^*(\epsilon)$.*

Note that we do not require ϵ to be small or have access to partial observations of blocks induced by edges in $\mathcal{G}_2$. Hence, Theorem 2 shows that SDP solves all MC instances introduced in Theorem 1 and beyond. As a result of Theorem 2, the SDP approach is a viable choice for those MC instances for which the preferable and faster B-M factorization method fails to recover the ground truth matrix. Similar to perturbing the ground truth matrix, one can perturb the linear measurement operator of the matrix completion problem $\mathcal{A}_\Omega$ as

$$\mathcal{A}_{\Omega(\epsilon)}(\mathbf{M})_{ij} := \begin{cases} \mathbf{M}_{ij}, & \text{if } (i,j) \in \Omega \\ \epsilon \mathbf{M}_{ij}, & \text{otherwise} \end{cases}, \quad (4)$$

where $\epsilon > 0$ is a sufficiently small real number (Zhang et al. 2022). Note that $\mathcal{A}_{\Omega(\epsilon)}$ satisfies the RIP condition with $\delta = (1 - \epsilon)/(1 + \epsilon)$.

Theorem 3. *Suppose that g is the squared loss function, i.e $g(\mathbf{x}) = \|\mathbf{x}\|^2$. Consider the measurement set Ω defined in Theorem 1. For every sufficiently small $\epsilon > 0$, there exists a low-complexity instance of the MS problem $\mathcal{P}_{\mathbf{M}^*, \mathcal{A}_\Omega(\epsilon), n, r}$ with $\mathcal{O}(2^{r|\mathcal{S}(\mathcal{G}_1)|})$ spurious local minima.*

The proof of the above theorem is similar to the proof of Theorem 1. Hence, the proof is omitted. The above theorem states that there are not only MC instances but also MS instances that suffer from this undesirable behavior of the B-M factorization approach. The ground truth matrix $\mathbf{M}^*$ is generated as in Theorem 1 to have $2^{r|\mathcal{S}(\mathcal{G}_1)|}$ global solutions for the unperturbed problem. Furthermore, the number of spurious solutions for this scheme can be quantified as $\mathcal{O}((1 - \delta)^{-1})$ for $\delta \geq 1/2$ in the rank-1 case (Zhang et al. 2022). Nevertheless, the SDP formulation approach trivially solves all these

undesirable MS instances because the perturbed measurement operator $\mathcal{A}_{\Omega(\epsilon)}$ corresponds to observing all the entries. Hence, the feasible set only contains the ground truth matrix.

Proposition 1. *Given a measurement set Ω, the SDP formulation (2) uniquely recovers the rank-r ground truth matrix $\mathbf{M}^*$ for the MS instance $\mathcal{P}_{\mathbf{M}^*,\mathcal{A}_{\Omega(\epsilon)},n,r}$, where $\mathcal{A}_{\Omega}(\epsilon)$ is defined in (4) and ϵ is an arbitrary nonzero number.*

Hence, the SDP approach successfully solves all the instances in Theorem 3 for which the RIP constant exists (while greater than $1/2$), unlike the B-M method. Consequently, SDP could be the preferred method when sufficient conditions on RIP for exact recovery by the B-M factorization are not met. We will provide sharper sufficiency bounds for the SDP approach next, which further corroborates its strength.

Sharper RIP bound for SDP

Since the SDP method is more powerful than the B-M factorization for certain classes of MC and MS problems as shown in the previous section and since specialized SDP algorithms can solve large-scale MC and MS problems, it is useful to further study the SDP method through the lens of the well-known RIP notion. We will derive a strong lower bound δ_{lb} on the RIP constant δ to guarantee convergence to the ground truth solution by using a proof technique called the inexistence of incorrect solution (Zhang, Sojoudi, and Lavaei 2019). We aim to find a linear measurement operator $\mathcal{A}$ with the smallest RIP constant such that the SDP formulation converges to a wrong solution. To do so, we need to solve the optimization problem

$$\min_{\delta,\mathcal{A}} \quad \delta$$
$$\text{s.t.} \quad \mathcal{A}(\mathbf{M}) = \mathcal{A}(\mathbf{M}^*)$$
$$\text{tr}(\mathbf{M}) \leq \text{tr}(\mathbf{M}^*) \qquad (5)$$
$$\mathcal{A} \text{ satisfies the } \delta_{2r}\text{-RIP property,}$$

where $\mathbf{M} \neq \mathbf{M}^*$. The condition $\text{tr}(\mathbf{M}) \leq \text{tr}(\mathbf{M}^*)$ guarantees that SDP cannot uniquely recover $\mathbf{M}^*$. Checking the RIP constant for a linear measurement operator is proven to be NP-hard (Tillmann and Pfetsch 2013). Therefore, it is difficult to solve the problem (5) analytically. To simplify the problem, we will introduce some notations. We use a matrix representation of the measurement operator $\mathcal{A}$ as follows:

$$\mathbf{A} = [\text{vec}(\mathbf{A}_1), \text{vec}(\mathbf{A}_2), \dots, \text{vec}(\mathbf{A}_d)]^T \in \mathbb{R}^{d \times n^2}.$$

Then, $\mathbf{A}\,\text{vec}(\mathbf{M}) = \mathcal{A}(\mathbf{M})$ for every matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$. We define $\mathbf{H} = \mathbf{A}^T\mathbf{A}$, which is the matrix representation of the kernel operator $\mathcal{H} = \mathcal{A}^T\mathcal{A}$ to simplify the last constraint of the problem (5). To derive a RIP bound, we consider the next optimization problem given $\mathbf{M}$ and $\mathbf{M}^*$, where $\mathbf{M}$ is the global solution of (2) and $\mathbf{M}^*$ is the ground truth solution:

$$\min_{\delta,\mathbf{H}} \quad \delta$$
$$\text{s.t.} \quad \mathbf{e}^T\mathbf{He} = 0 \qquad (6)$$
$$\mathbf{H} \text{ is symmetric and satisfies the } \delta_{2r}\text{-RIP,}$$

where $\mathbf{e} = \text{vec}(\mathbf{M}^* - \mathbf{M})$. For this fixed $\mathbf{M}$ and $\mathbf{M}^*$, we assume that $\mathbf{M} \neq \mathbf{M}^*$ and that $\text{rank}(\mathbf{M}^* - \mathbf{M}) > 2r$, since

if $\text{rank}(\mathbf{M}^* - \mathbf{M}) \leq 2r$, the relation $\mathbf{M} = \mathbf{M}^*$ holds automatically by definition of δ_{2r}-RIP for any δ since it implies strong convexity. Denote the optimal value to (6) as $\delta(\mathbf{e})$, which is a function of $\mathbf{e}$. It is desirable to find

$$\delta^* := \min_{\mathbf{e}:\text{tr}(\mathbf{M}) \leq \text{tr}(\mathbf{M}^*)} \delta(\mathbf{e}).$$

By the logic of the in-existence of counterexample, we know that if a problem $\mathbf{H} = \mathbf{A}^T\mathbf{A}$ has δ_{2r}-RIP with $\delta < \delta^*$, then the solution to (2) will be $\mathbf{M}^*$, which is the ground truth solution. However, since the last constraint of (6) is non-convex, it is useful to replace it with a surrogate condition that allows solving the problem analytically. The following problem helps to achieve this goal:

$$\min_{\delta,\mathbf{H}} \quad \delta$$
$$\text{s.t.} \quad \mathbf{e}^T\mathbf{He} \leq 2\|\mathbf{e}_c\|^2 + 2(l-3)\delta\|\mathbf{e}_c\|^2 \qquad (7)$$
$$(1-\delta)\mathbf{I}_{n^2} \preceq \mathbf{H} \preceq (1+\delta)\mathbf{I}_{n^2}.$$

Here, $l = \lceil n/r \rceil$ and we define $\{\mathbf{e}_i\}_{i=1}^l$ and $\mathbf{e}_c$ in the following fashion. First, consider the eigendecomposition of $\mathbf{M}^* - \mathbf{M}$ and assume that the eigenvalues are ordered in terms of their absolute values, namely, $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Let $\mathbf{u}_k$'s denote the corresponding orthonormal eigenvectors:

$$\text{mat}_S(\mathbf{e}) = \mathbf{M}^* - \mathbf{M} = \sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{u}_k^T.$$

Then, we define:

$$\mathbf{e}_i = \text{vec}\left(\sum_{k=(i-1)*r+1}^{\min\{i*r,n\}} \lambda_k \mathbf{u}_k \mathbf{u}_k^T\right),$$

$\mathbf{e}_{2r} = \mathbf{e}_1 + \mathbf{e}_2$, and $\mathbf{e}_c = \sum_{i=3}^l \mathbf{e}_i$. The next proposition allows us to replace (6) with (7) because the optimal value of the (7), $\delta_{lb}(\mathbf{e})$, gives a lower bound on $\delta(\mathbf{e})$.

Proposition 2. *The optimal objective value of the problem (7), $\delta_{lb}(\mathbf{e})$, is always less than or equal to the optimal objective value of the problem (6), i.e., $\delta_{lb}(\mathbf{e}) \leq \delta(\mathbf{e})$.*

The proof of this proposition is central to the construction of the sufficiency bound, which is based on using a convex program to serve as an estimate of the non-convex problem. After we extend the RIP_{2r} constraint in (7) to be RIP_n(thus making it convex), it is necessary to somehow preserve the information that the near isometric property of $\mathbf{H}$ should only apply to low-rank matrices. This is achieved by changing the first constraint so that $\mathbf{e}$ does not need to be completely in the null space of $\mathbf{H}$. (7) approximately requires that $\mathbf{H}$ only maps a certain low-rank sub-manifold to 0. As a result of Proposition 2, it immediately follows that

$$\delta_{\text{lb}} = \min_{\mathbf{e}:\text{tr}(\mathbf{M}) \leq \text{tr}(\mathbf{M}^*)} \delta_{\text{lb}}(\mathbf{e}) \leq \delta^*.$$

In fact, we can obtain a lower bound on the value δ_{lb} by solving the problem (7) analytically. The following lemma quantifies a lower bound on δ_{lb}.

Lemma 1. *It holds that*

$$\delta_{lb} \geq \frac{2r}{n + (n-2r)(2l-5)}.$$

The best-known sufficiency bound presented in (Cai and Zhang 2013) is independent of n and r. This sufficiency lower bound presented in Lemma 1 can be tighter than $1/2$ depending on the size of the problem n and the rank of the ground truth matrix r. The SDP formulation converges to ground truth solution whenever RIP constant δ is close to 1 as $r \to n/2$. Conversely, whenever r/n is ratio is small, e.g. rank-1 matrix sensing problem with large n, $\delta < 1/2$ is stronger for recovery of the ground truth. Combined with the $1/2$ sufficiency bound that works for both the symmetric and asymmetric cases, we obtain the following result:

Theorem 4. *Let $l = \lceil n/r \rceil$. The global solution of the SDP formulation (2) will be the ground truth matrix $\mathbf{M}^*$ if the sensing matrix $\mathcal{A}$ satisfies the RIP condition with the RIP constant δ_{2r} satisfying the inequality:*

$$\delta_{2r} < \max\left\{ 1/2, \frac{2r}{n + (n-2r)(2l-5)} \right\}.$$

Compared with the existing sufficiency RIP bounds, this new result has a striking advantage. The bound $\delta_{2r} < 1/2$ has already been proven to be the sharpest for the B-M formulation, which is independent of the search rank. In contrast, Theorem 4 shows that the RIP bound for SDP exceeds this bound and approaches 1 as the rank r increases. As opposed to the popular belief that B-M enjoys very similar RIP guarantees as the SDP approach, there are real benefits to switching to the SDP formulation, making it a more competitive option since specialized SDP solvers are becoming more efficient in recent years. However, we will next provide some problem instances for which the SDP method fails to solve the problem while the B-M method contains no spurious solutions, which balances the desirable properties of the SDP method.

4 Advantages of the B-M Method

We provide two classes of rank-1 matrix completion problems for which the B-M factorization does not contain any spurious solution while SDP fails to recover its ground truth matrix. Throughout this section, the rank-1 PSD ground truth matrix $\mathbf{M}^* = \mathbf{x}^*(\mathbf{x}^*)^T$ is assumed not to contain any zero entries, meaning that $x_i^* \neq 0$ for all $i \in [n]$. Before proceeding with the results, we provide two small examples to highlight the underlying ideas behind the main results.

Example 1. *Consider a block sparsity graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = 3$ nodes and the edge set $\mathcal{E} = \{(1,1),(1,2),(2,3)\}$. It is a chain graph with 3 nodes and a self-loop at the first node. We aim to show that only second-order critical points are the global solutions of the B-M factorization method. The objective with the squared loss function can be explicitly written as $\min_{\mathbf{x} \in \mathbb{R}^3} f(\mathbf{x})$, where*

$$f(\mathbf{x}) = \frac{1}{4} \sum_{\substack{(i,j)\in\mathcal{E} \\ i=j}} (x_i^2 - (x_i^*)^2)^2 + \frac{1}{2} \sum_{\substack{(i,j)\in\mathcal{E} \\ i\neq j}} (x_i x_j - x_i^* x_j^*)^2.$$

The corresponding gradient and Hessian are:

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \sum_{\substack{i,j\in\mathcal{E} \\ i=j}} (x_i^2 - (x_i^*)^2)x_i + \sum_{\substack{i,j\in\mathcal{E} \\ i\neq j}} (x_i x_j - x_i^* x_j^*)x_j,$$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i^2} = 1[(i,i) \in \mathcal{E}](3x_i^2 - (x_i^*)^2) + \sum_{i,j\in\mathcal{E}} x_j^2,$$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \begin{cases} 2x_i x_j - x_i^* x_j^*, & \text{if } i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}.$$

Each second-order critical point $\hat{\mathbf{x}}$ must satisfy the conditions $\nabla f(\hat{\mathbf{x}}) = 0$ and $\nabla^2 f(\hat{\mathbf{x}}) \succeq 0$. The third entry of the gradient implies either $\hat{x}_2 = 0$ or $\hat{x}_2 \hat{x}_3 = x_2^* x_3^*$. Whenever $\hat{x}_2 = 0$, the Hessian is not PSD since $[\nabla^2 f(\hat{\mathbf{x}})]_{2:3,2:3} \not\succeq 0$. Thus, $\hat{x}_2 \hat{x}_3 = x_2^* x_3^*$ must hold. Following this, $\partial f(\hat{\mathbf{x}})/\partial x_2$ implies either $\hat{x}_1 = 0$ or $\hat{x}_1 \hat{x}_2 = x_1^* x_2^*$. However, if $\hat{x}_1 = 0$, then $\partial f(\hat{\mathbf{x}})/\partial x_1$ gives $-x_1^* x_2^* \hat{x}_2 = 0$, which implies $\hat{x}_2 = 0$, a contradiction. Thus, each second-order critical point must have the following properties:

$$\hat{x}_1^2 = (x_1^*)^2, \quad \hat{x}_1 \hat{x}_2 = x_1^* x_2^*, \quad \hat{x}_2 \hat{x}_3 = x_2^* x_3^*.$$

The solution to this system of equations proves the exact recovery of the ground truth matrix $\mathbf{M}^*$. Hence, the only second-order critical points are the valid factors of the ground truth solution, i.e $\pm\mathbf{x}^*$. The next step is to demonstrate the failure of the SDP formulation (2) for some instances of the MC problem with this given block sparsity matrix $\mathcal{G}$. The problem (2) is equivalent to the optimization

$$\min_{\mathbf{M}\in\mathbb{R}^{3\times 3}} \mathbf{M}_{2,2} + \mathbf{M}_{3,3}$$

$$s.t \begin{bmatrix} (x_1^*)^2 & x_1^* x_2^* & \mathbf{M}_{1,3} \\ x_1^* x_2^* & \mathbf{M}_{2,2} & x_2^* x_3^* \\ \mathbf{M}_{3,1} & x_2^* x_3^* & \mathbf{M}_{3,3} \end{bmatrix} \succeq 0.$$

Consider a feasible solution $\hat{\mathbf{M}}$ with $\hat{\mathbf{M}}_{2,2} = \hat{\mathbf{M}}_{3,3} = |x_2^* x_3^*|$ and $\hat{\mathbf{M}}_{1,3} = \hat{\mathbf{M}}_{3,1} = |x_1^* x_2^*|$. $\hat{\mathbf{M}}$ is feasible whenever $|x_3^*| \geq |x_2^*|$. One can show that the feasible solution $\hat{\mathbf{M}}$ is strictly better than the ground truth solution whenever $|x_3^*| > |x_2^*|$,. Thus, SDP fails to recover the ground truth matrix.

This example clearly demonstrates the existence of MC instances for which the B-M method successfully converges to the ground truth solution while the SDP fails to find the solution. One reason is that the number of measurements is $\mathcal{O}(n)$, which is the minimum threshold for exact completion. However, the statistical guarantees on SDP often need more observations. Example 1 is generalized to any chain graph with n nodes and a single self-loop at one of the ends.

Theorem 5. *Consider the MC problem with a rank-1 PD ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n\times n}$ that can be factorized as $\mathbf{M}^* = \mathbf{x}^*(\mathbf{x}^*)^T$ with $x_i^* \neq 0, \forall i \in [n]$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a block sparsity graph with $|\mathcal{V}| = n$ and $\mathcal{E} = \{(1,1),(1,2),(2,3),\cdots,(n-1,n)\}$. Then, the B-M method (3b) does not contain any spurious solutions.*

Theorem 6. *Consider the ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n\times n}$ satisfying the conditions in Theorem 5. Suppose that there exist two indices j, k such that $x_k^* > x_j^*$ and $j, k > 2$. Then, the SDP problem (2) fails to recover the ground truth matrix.*

In addition to the success of the B-M factorization, the previous result establishes the failure of the SDP for the instances described in the above theorem. As mentioned before, SDP fails due to a lack of observations on the diagonal entries

of the ground truth matrix. Note that the RIP condition is not satisfied since these are MC problems. As a result, whenever we do not have sufficient guarantees on linear measurement operator, none of the methods are superior to the other one in terms of exact recovery. The next example identifies another class of problem instances that corroborates these findings.

Example 2. *Consider a block sparsity graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = 3$ nodes and the edge set $\mathcal{E} = \{(1,2), (2,3), (3,1)\}$. Namely, it is a simple cycle with 3 nodes. The B-M factorization formulation (3b) with the squared loss function can be written the same as in Example 1. We can show that each second-order critical point $\hat{\mathbf{x}}$ only has nonzero entries, i.e., $\hat{x}_i \neq 0$ for all $i \in [n]$. Without loss of generality, suppose by contradiction that $\hat{x}_1 = 0$. For stationarity, either $\hat{x}_2 \hat{x}_3 = x_2^* x_3^*$ or $\hat{x}_2 = \hat{x}_3 = 0$ should be satisfied. The latter implies $\hat{\mathbf{x}} = 0$ and $\nabla^2 f(\hat{\mathbf{x}}) \not\succeq 0$ in that case. Thus, $\hat{x}_2 \hat{x}_3 = x_2^* x_3^*$ must hold. Following this, $\partial f(\hat{\mathbf{x}})/\partial x_1$ yields $-x_1^* x_2^* \hat{x}_2 - x_1^* x_3^* \hat{x}_3 = 0$. Combining these two yields $(\hat{x}_3)^2 = -(x_2^*)^2$, which does not have any real solution. Hence, each second-order critical point $\hat{\mathbf{x}}$ must have only nonzero entries. By the condition $\nabla f(\hat{\mathbf{x}}) = 0$, whenever $\hat{x}_i \hat{x}_j = x_i^* x_j^*$ for some $(i,j) \in \mathcal{E}$, then $\hat{x}_i \hat{x}_j = x_i^* x_j^*$ holds for every $(i,j) \in \mathcal{E}$. This system of equations yields the ground truth solution. Accordingly, a spurious solution $\hat{\mathbf{x}}$ must have the following characteristics: $\hat{x}_i \hat{x}_j \neq x_i^* x_j^*, \forall (i,j) \in \mathcal{E}$ and $\hat{x}_i \neq 0, \forall i \in \{1,2,3\}$. Define $a_{i,j} = x_i x_j - x_i^* x_j^*$. Then, the stationarity condition becomes*

$$\nabla f(\hat{\mathbf{x}}) = \begin{bmatrix} \hat{a}_{1,2} \hat{x}_2 + \hat{a}_{1,3} \hat{x}_3 \\ \hat{a}_{1,2} \hat{x}_1 + \hat{a}_{2,3} \hat{x}_3 \\ \hat{a}_{1,3} \hat{x}_1 + \hat{a}_{2,3} \hat{x}_2 \end{bmatrix} = 0.$$

Multiplying the first entry of the gradient by $\hat{x}_1/\hat{x}_2$ and substituting with the second entry gives $\hat{a}_{1,2} \hat{x}_1 = -\hat{a}_{2,3} \hat{x}_3$. Successively, substituting this to the third entry of the gradient results in

$$\hat{x}_3 (\hat{a}_{1,3} \hat{x}_1 - \hat{a}_{2,3} \hat{x}_2) = 0.$$

Because we search for a solution with $\hat{x}_i \neq 0$, we must have $\hat{a}_{1,3} \hat{x}_1 - \hat{a}_{2,3} \hat{x}_2 = 0$. This condition combined with the last entry of the gradient results in the condition $\hat{a}_{1,3} \hat{x}_1 = 0$, which is a contradiction. As a result, all the second-order critical points are global solutions that yield the ground truth matrix completion. Our next goal is to show that the SDP formulation (2) fails for this class of instances of MC instances. Note that the SDP formulation of the matrix completion problem considered in this example is equivalent to the formulation:

$$\min_{\mathbf{M} \in \mathbb{R}^{3 \times 3}} \mathbf{M}_{1,1} + \mathbf{M}_{2,2} + \mathbf{M}_{3,3}$$

$$s.t \begin{bmatrix} \mathbf{M}_{1,1} & x_1^* x_2^* & x_1^* x_3^* \\ x_1^* x_2^* & \mathbf{M}_{2,2} & x_2^* x_3^* \\ x_1^* x_3^* & x_2^* x_3^* & \mathbf{M}_{3,3} \end{bmatrix} \succeq 0.$$

Without loss of generality, assume that $x_1^* \leq x_2^* \leq x_3^*$ by the symmetry of the problem. Consider a feasible rank-2 solution $\hat{\mathbf{M}}$ given as $\hat{\mathbf{M}}_{1,1} = x_1^*(x_3^* - x_2^*), \hat{\mathbf{M}}_{2,2} = x_2^*(x_3^* - x_1^*)$ and $\mathbf{M}_{3,3} = x_3^*(x_1^* + x_2^*)$. One can show that the feasible solution $\hat{\mathbf{M}}$ is strictly better than the ground truth solution if $x_3^* > x_1^* + x_2^*$. Hence, the SDP cannot recover the ground

truth solution for all the instances with a simple cycle block sparsity graph.

Similar to Example 1, SDP fails in this example due to a lack of diagonal observations. Next, we can generalize this instance to any simple cycle block sparsity graph with an odd number of vertices.

Theorem 7. *Consider the MC problem with a rank-1 PD ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n \times n}$ with factorization $\mathbf{M}^* = \mathbf{x}^*(\mathbf{x}^*)^T$, $x_i^* \neq 0, \forall i \in [n]$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a block sparsity graph with $|\mathcal{V}| = |\mathcal{E}| = n = 2k+1$ and $\mathcal{E} = \{(0,1), (1,2), \dots (2k-1, 2k), (2k, 0)\}$. Then, the B-M factorization problem (3b) does not contain any spurious solutions.*

Theorem 8. *Consider the ground truth matrix $\mathbf{M}^* \in \mathbb{R}^{n \times n}$ satisfying the conditions in Theorem 7. Suppose that the condition $\sum_{t=1}^{k} (x_{2t-1}^*)^2 > \sum_{t=0}^{k} (x_{2t}^*)^2$ holds for arbitrary node 0. Then, the SDP problem (2) fails to recover the ground truth matrix.*

Note that we can choose any node as node 0 due to the symmetry of the problem. Therefore, the condition stated in Theorem 8 is not restrictive because this condition suffices to hold for a chosen node 0 among $2k+1$ ones. Moreover, we highlight that we proved that every stationary point with a PSD Hessian must be the global solution, thereby also implying that any stationary point that is not a global solution (saddle points) has at least one negative eigenvalue in its Hessian, which means that this optimization problem satisfies the strict saddle property. This property is widely known in the literature to lead to polynomial time convergence. As a result, the B-M factorization can outperform the convex relaxation approach.

5 Conclusions

We conducted a comparison between two main approaches to the matrix completion and matrix sensing problems: a convex relaxation that gives an SDP formulation and the B-M factorization method. Both of these methods enjoy mathematical guarantees for the recovery of the ground truth matrix whenever the RIP assumption is satisfied with a sufficiently small δ. We offered the first result in the literature that compares these two methods whenever the RIP condition is not satisfied or only satisfied with a large constant. We discovered classes of problems for which B-M factorization fails while the SDP recovers the ground truth matrix. This inspired us to investigate sharper bounds on sufficient conditions for the SDP formulation. We provided RIP bounds for the SDP formulation that depend on the rank of the solution and are automatically satisfied for high-rank problems, unlike the B-M method. Conversely, when the number of measurements from the ground truth matrix is not high, we showed that SDP fails drastically while the B-M method does not contain any spurious solutions. As a result, we conclude that none of the methods outperforms the other one whenever the sufficiency guarantees are not met. The parameters of the problem, such as dimension, rank, and linear measurement operator, determine which solution method performs better. Consequently, it is prudent to apply both solution methods in case the RIP and incoherence are not satisfied.

Acknowledgements

References

Bhojanapalli, S.; Neyshabur, B.; and Srebro, N. 2016. Global Optimality of Local Search for Low Rank Matrix Recovery. In *Advances in Neural Information Processing Systems*, volume 29.

Bi, Y.; and Lavaei, J. 2021. On the absence of spurious local minima in nonlinear low-rank matrix recovery problems. In *International Conference on Artificial Intelligence and Statistics*, 379–387. PMLR.

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

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

Cai, T. T.; and Zhang, A. 2013. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1): 74–93.

Candes, E. J.; Eldar, Y. C.; Strohmer, T.; and Voroninski, V. 2015. Phase retrieval via matrix completion. *SIAM review*, 57(2): 225–251.

Candes, E. J.; and Plan, Y. 2010. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. *arXiv preprint arXiv:1001.0339*.

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

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

Davenport, M. A.; Plan, Y.; Van Den Berg, E.; and Wootters, M. 2014. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3): 189–223.

Fattahi, S.; and Sojoudi, S. 2020. Exact guarantees on the absence of spurious local minima for non-negative rank-1 robust principal component analysis. *Journal of machine learning research*.

Ge, R.; Jin, C.; and Zheng, Y. 2017. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *International Conference on Machine Learning*, 1233–1242. PMLR.

Ha, W.; Liu, H.; and Barber, R. F. 2020. An Equivalence Between Critical Points for Rank Constraints Versus Low-Rank Factorizations. *SIAM Journal on Optimization*, 30(4): 2927–2955.

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

Jin, M.; Molybog, I.; Mohammadi-Ghazi, R.; and Lavaei, J. 2019. Towards robust and scalable power system state estimation. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 3245–3252. IEEE.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

Ma, Z.; Bi, Y.; Lavaei, J.; and Sojoudi, S. 2022. Sharp Restricted Isometry Property Bounds for Low-rank Matrix Recovery Problems with Corrupted Measurements. *AAAI-22*.

Ma, Z.; and Sojoudi, S. 2022. Noisy Low-rank Matrix Optimization: Geometry of Local Minima and Convergence Rate. *arXiv preprint arXiv:2203.03899*.

Mohan, K.; and Fazel, M. 2010. New restricted isometry results for noisy low-rank recovery. In *2010 IEEE International Symposium on Information Theory*, 1573–1577. IEEE.

Park, D.; Kyrillidis, A.; Carmanis, C.; and Sanghavi, S. 2017. Non-square Matrix Sensing Without Spurious Local Minima via the Burer–Monteiro Approach. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, 65–74.

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

Recht, B.; Xu, W.; and Hassibi, B. 2008. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *2008 47th IEEE Conference on Decision and Control*, 3065–3070. IEEE.

Shechtman, Y.; Eldar, Y. C.; Cohen, O.; Chapman, H. N.; Miao, J.; and Segev, M. 2015. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3): 87–109.

Singer, A. 2011. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1): 20–36.

Tillmann, A. M.; and Pfetsch, M. E. 2013. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2): 1248–1259.

Yalçın, B.; Zhang, H.; Lavaei, J.; and Sojoudi, S. 2022. Factorization approach for low-complexity matrix completion problems: Exponential number of spurious solutions and failure of gradient methods. In *International Conference on Artificial Intelligence and Statistics*, 319–341. PMLR.

Yurtsever, A.; Tropp, J. A.; Fercoq, O.; Udell, M.; and Cevher, V. 2021. Scalable Semidefinite Programming. *SIAM Journal on Mathematics of Data Science*, 3(1): 171–200.

Yurtsever, A.; Udell, M.; Tropp, J.; and Cevher, V. 2017. Sketchy Decisions: Convex Low-Rank Matrix Optimization with Optimal Storage. In Singh, A.; and Zhu, J., eds., *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, 1188–1196. PMLR.

Zhang, G.; and Zhang, R. Y. 2020. How Many Samples Is a Good Initial Point Worth in Low-Rank Matrix Recovery? In *Advances in Neural Information Processing Systems*, volume 33, 12583–12592.

Zhang, H.; Bi, Y.; and Lavaei, J. 2021a. General low-rank matrix optimization: Geometric analysis and sharper bounds. *arXiv preprint arXiv:2104.10356*.

Zhang, H.; Bi, Y.; and Lavaei, J. 2021b. General Low-Rank Matrix Optimization: Geometric Analysis and Sharper Bounds. In *Advances in Neural Information Processing Systems*.

Zhang, H.; Yalcin, B.; Lavaei, J.; and Sojoudi, S. 2022. A Unified Complexity Metric for Nonconvex Matrix Completion and Matrix Sensing in the Rank-one Case. *arXiv preprint arXiv:2204.02364*.

Zhang, R. Y. 2021. Sharp Global Guarantees for Nonconvex Low-Rank Matrix Recovery in the Overparameterized Regime. ArXiv:2104.10790.

Zhang, R. Y.; and Lavaei, J. 2021. Sparse semidefinite programs with guaranteed near-linear time complexity via dualized clique tree conversion. *Mathematical programming*, 188(1): 351–393.

Zhang, R. Y.; Sojoudi, S.; and Lavaei, J. 2019. Sharp Restricted Isometry Bounds for the Inexistence of Spurious Local Minima in Nonconvex Matrix Recovery. *Journal of Machine Learning Research*, 20(114): 1–34.

Zhang, X.; Wang, L.; Yu, Y.; and Gu, Q. 2018. A Primal-Dual Analysis of Global Optimality in Nonconvex Low-Rank Matrix Recovery. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 5862–5871.

Zhang, Y.; Madani, R.; and Lavaei, J. 2017. Conic relaxations for power system state estimation with line measurements. *IEEE Transactions on Control of Network Systems*, 5(3): 1193–1205.

Zhu, Z.; Li, Q.; Tang, G.; and Wakin, M. B. 2018. Global Optimality in Low-Rank Matrix Optimization. *IEEE Transactions on Signal Processing*, 66(13): 3614–3628.

Zhu, Z.; Li, Q.; Tang, G.; and Wakin, M. B. 2021. The global optimization geometry of low-rank matrix optimization. *IEEE Transactions on Information Theory*, 67(2): 1308–1331.