

Exact Guarantees on the Absence of Spurious Local Minima for Non-negative Rank-1 Robust Principal Component Analysis

Salar Fattahi,

FATTAHI@BERKELEY.EDU

*Department of Industrial Engineering and Operations Research
University of California, Berkeley*

Somayeh Sojoudi,

SOJOUDI@BERKELEY.EDU

*Departments of Electrical Engineering and Computer Sciences and Mechanical Engineering
University of California, Berkeley*

Editor: Mark Schmidt

Abstract

This work is concerned with the non-negative rank-1 robust principal component analysis (RPCA), where the goal is to recover the dominant non-negative principal components of a data matrix *precisely*, where a number of measurements could be grossly corrupted with sparse and arbitrary large noise. Most of the known techniques for solving the RPCA rely on convex relaxation methods by lifting the problem to a higher dimension, which significantly increase the number of variables. As an alternative, the well-known Burer-Monteiro approach can be used to cast the RPCA as a non-convex and non-smooth ℓ_1 optimization problem with a significantly smaller number of variables. In this work, we show that the low-dimensional formulation of the symmetric and asymmetric positive rank-1 RPCA based on the Burer-Monteiro approach has benign landscape, i.e., 1) it does not have any spurious local solution, 2) has a unique global solution, and 3) its unique global solution coincides with the *true* components. An implication of this result is that simple local search algorithms are guaranteed to achieve a zero global optimality gap when directly applied to the low-dimensional formulation. Furthermore, we provide strong deterministic and probabilistic guarantees for the exact recovery of the true principal components. In particular, it is shown that a constant fraction of the measurements could be grossly corrupted and yet they would not create any spurious local solution.

1. Introduction

The principal component analysis (PCA) is perhaps the most widely-used dimension-reduction method that reveals the components with maximum variability in high-dimensional datasets. In particular, given the data matrix $X \in \mathbb{R}^{m \times n}$, where each row corresponds to a data sample with size n , the goal is to recover its most dominant component under the rank-1 spiked model¹

$$X = \beta \mathbf{u} \mathbf{v}^\top + S \quad (1)$$

1. There are more general models under which the PCA is shown to be useful (see Jolliffe (2011) for more details). We use the rank-1 spiked model since it fits into our framework and is often used as a baseline to evaluate the performance of the PCA.

where β determines the signal-to-noise ratio, S is the additive noise matrix, and \mathbf{u} and \mathbf{v} are two unknown unit norm vectors. If the data matrix X is symmetric (for instance, it corresponds to a sample covariance matrix), then (1) can be modified as

$$X = \beta \mathbf{v}\mathbf{v}^\top + S \quad (2)$$

Depending on the nature of the noise matrix, different methods have been proposed in the literature to recover the principal components from (partial) observations of X . The problem of recovering β , \mathbf{u} , and \mathbf{v} under a Gaussian and sparse noise is conventionally referred to as PCA and robust PCA (or RPCA), respectively.

The properties of both PCA and its robust analog have been heavily studied in the literature and their applications span from quantitative finance to health care and neuroscience (Hull and White 1990; Caprihan et al. 2008; Brenner et al. 2000). Recently, a special focus has been devoted to further exploiting the prior knowledge on the principal components, such as sparsity (Zou et al. 2006) and nonlinearity (Gorban et al. 2008). Accordingly, one such knowledge appearing in different applications is the non-negativity of the principal components (Montanari and Richard 2016). In this scenario, one needs to solve the PCA or the RPCA under the additional constraints $\mathbf{u}, \mathbf{v} \geq 0$. While the non-negative PCA has been recently studied in Montanari and Richard (2016), the main focus of our work is on its robust variant, where the noise matrix is assumed to be sparse and the goal is the *exact* recovery of the non-negative vectors \mathbf{u} and \mathbf{v} . Note that the non-negativity of principal components naturally arises in many real-world problems. In what follows, we will present two classes of real-world applications for which the non-negative RPCA is useful.

1. Non-negative matrix factorization: Extracting the dominant principal component of a symmetric or asymmetric data matrix appears in many applications and the examples are ubiquitous. For instance, an important problem in astronomy is the recovery of non-negative astronomical signals from the covariance matrix of photometric observations (Ren et al. 2018). The measured data samples are prone to sparse and random outliers. Similarly, one can extract moving objects from video frames via non-negative matrix factorization by treating the background as the dominant low-rank component in the video frames and the moving object as sparse noise (the non-negativity of the data is due to the non-negative values of the pixels) (Lee and Seung 1999; Candès et al. 2011). We will conduct a case study on this application later in the paper.

2. Gene networks: Gene activities can be captured by the samples collected from different organs, and are described by multi-spiked models (Lazzeroni and Owen 2002):

$$X = X_0 + \sum_{i=1}^k \mathbf{u}_{(i)} \mathbf{v}_{(i)}^\top \quad (3)$$

where $(i, j)^{\text{th}}$ entry of X measures the strength of the participation of gene i in sample j and X_0 is an offset. Furthermore, k is the number of the gene-blocks, and $\mathbf{u}_{(i)}$ and $\mathbf{v}_{(i)}$ measure the participation of different genes and samples in the i^{th} gene-block. The participation vectors are non-negative and the measurements can be subject to malfunctioning of the measurement tools. Therefore, the problem of obtaining $\mathbf{u}_{(i)}$ and $\mathbf{v}_{(i)}$ can be cast as a non-negative RPCA with multiple principal components.

The seminal work by Candès et al. (2011) proposes a sparsity promoting convex relaxation for the RPCA that is capable of the exact recovery of \mathbf{u} and \mathbf{v} . Upon defining $W = \mathbf{u}\mathbf{v}^\top$, the convex relaxation of the RPCA is defined as

$$\min_{W \in \mathbb{R}^{m \times n}} \|W\|_* + \lambda \|\mathcal{P}_\Omega(X - W)\|_1 \quad (4)$$

where $\|W\|_*$ is the nuclear norm of W , serving as a penalty on the rank of the recovered matrix W , and $\|\cdot\|_1$ is used to denote the element-wise ℓ_1 norm. Furthermore, $\mathcal{P}_\Omega(\cdot)$ is the projection onto the set of matrices with the same support as the measurement set Ω . Therefore, upon defining $S = X - W$ as the corruption or noise matrix, $\|\mathcal{P}_\Omega(X - W)\|_1$ plays the role of promoting sparsity in the estimated noise matrix. After finding an optimal value of W , the matrix can then be decomposed into the desired vectors \mathbf{u} and \mathbf{v} , provided that the relaxation is exact. Notice that the problem is convexified via lifting from $n + m$ variables on (\mathbf{u}, \mathbf{v}) to nm variables on W . Despite the convexity of the lifted problem, its dimension makes it prohibitive to solve in high-dimensional settings. To circumvent this issue, one popular approach is to resort to an alternative formulation, inspired by Burer and Monteiro (2003) (commonly known as the Burer-Monteiro technique):

$$\min_{\mathbf{u} \in \mathbb{R}_+^m, \mathbf{v} \in \mathbb{R}_+^n} \|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^\top)\|_1 \quad (5)$$

Despite the non-convexity of (5), its smooth counterpart (with or without non-negativity constraints) defined as

$$\min_{\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n} \underbrace{\|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^\top)\|_F^2}_{g(\mathbf{u}, \mathbf{v})} \quad (6)$$

has been widely used in matrix completion/sensing and is known to possess *benign global landscape*, i.e., every local solution is also global and every saddle point has a direction with a strictly negative curvature (Bhojanapalli et al. 2016; Ge et al. 2016, 2017). This will be stated below.

Theorem 1 (Informal, Benign Landscape (Ge et al. 2017)) *Under some technical conditions, a regularized version of (6) has benign landscape: every local minimum is global and every saddle point has a direction with a strictly negative curvature.*

In particular, both symmetric and asymmetric matrix completion (or matrix sensing) under dense Gaussian noise can be cast as (6) and in light of the above theorem, they have benign landscape. However, it is well-known that such smooth norms are incapable of correctly identifying and rejecting sparse-but-large noise/outliers in the measurements.

Despite the generality of Theorem 1 within the realm of smooth norms, it does not address the following important question: *Does the non-smooth and non-negative rank-1 RPCA (5) have benign landscape?*

1.1 The Issue with the Known Proof Techniques

To understand the inherent difficulty of examining the landscape of (5), it is essential to explain why the existing proof techniques for the absence of spurious local minima in

matrix sensing/completion cannot naturally be extended to their robust counterparts. In general, the main idea in the literature behind proving the benign landscape of matrix sensing/completion is based on analyzing the gradient and the Hessian of the objective function. More precisely, for every point that satisfies $\nabla g(\mathbf{u}, \mathbf{v}) = 0$ and does not correspond to a globally optimal minimum, it suffices to find a *global* direction of descent \mathbf{d} such that $\text{vec}(\mathbf{d})^\top \nabla^2 g(\mathbf{u}, \mathbf{v}) \text{vec}(\mathbf{d}) < 0$, where $\text{vec}(\mathbf{d})$ is the vectorized version of \mathbf{d} and $\nabla^2 g(\mathbf{u}, \mathbf{v})$ is the Hessian of $g(\mathbf{u}, \mathbf{v})$. Such a direction certifies that every stationary point that is not globally optimal must be either a local maximum or a saddle point with a strictly negative direction. However, this approach cannot be used to prove similar results for (5) mainly because the objective function of (5) is non-differentiable and, hence, the Hessian is not well-defined. This difficulty calls for a new methodology for analyzing the landscape of the robust and non-smooth PCA; a goal that is at the core of this work.

2. Contributions

In this work, we characterize the landscape of both the symmetric non-negative rank-1 RPCA defined as

$$\min_{\mathbf{u} \in \mathbb{R}_+^n} \underbrace{\|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{u}^\top)\|_1 + R_\beta(\mathbf{u})}_{f_{\text{reg}}(\mathbf{u})} \quad (\text{SN-RPCA})$$

and its asymmetric counterpart defined as

$$\min_{\mathbf{u} \in \mathbb{R}_+^m, \mathbf{v} \in \mathbb{R}_+^n} \underbrace{\|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^\top)\|_1 + R_\beta(\mathbf{u}, \mathbf{v})}_{f_{\text{reg}}(\mathbf{u}, \mathbf{v})} \quad (\text{AN-RPCA})$$

In particular, we fully characterize the stationary points of these optimization problems, under both deterministic and probabilistic models for the measurement index Ω and the noise matrix S . The functions $R_\beta(\mathbf{u})$ and $R_\beta(\mathbf{u}, \mathbf{v})$ are regularization functions that prevent the solutions from *blowing up*; roughly speaking, they penalize the points whose norm is greater than β , but do not change the landscape otherwise. The exact definitions of these regularization functions will be presented later in Section 7.

Remark 2 *The focus of this paper is on the symmetric and non-symmetric RPCA under the rank-1 spiked model. A natural extension to this model is its rank- r variant:*

$$X = UV^\top + S \quad (7)$$

where $U := [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r] \in \mathbb{R}_+^{m \times r}$ and $V := [\mathbf{v}_1 \ \cdots \ \mathbf{v}_r] \in \mathbb{R}_+^{n \times r}$ are non-negative matrices encompassing the r principal components of the model (the symmetric version can be defined in a similar manner). Furthermore, similar to the rank-1 case, S is a sparse noise matrix. Under this rank- r spiked model, the aim of the non-negative **rank- r** RPCA is to recover the non-negative matrices U and V given a subset of the elements of the noisy measurement matrix X . In Section 9, we will elaborate on the technical difficulties behind this extension. In addition, we will provide some empirical evidence to support that the developed results may hold for the general non-negative rank- r RPCA with $r \geq 2$.

Definition 3 *Given the set Ω , two graphs are defined below:*

- The sparsity graph $\mathcal{G}(\Omega)$ induced by Ω for an instance of (SN-RPCA) is defined as a graph with the vertex set $V := \{1, 2, \dots, n\}$ that includes an edge (i, j) if $(i, j) \in \Omega$.
- The bipartite sparsity graph $\mathcal{G}_{m,n}(\Omega)$ induced by Ω for an instance of (AN-RPCA) is defined as a graph with the vertex partitions $V_u := \{1, 2, \dots, m\}$ and $V_v := \{m+1, m+2, \dots, m+n\}$ that includes an edge (i, j) if $(i, j-m) \in \Omega$.

Furthermore, define $\Delta(\mathcal{G}(\Omega))$ and $\delta(\mathcal{G}(\Omega))$ as the maximum and minimum degrees of the nodes in $\mathcal{G}(\Omega)$, respectively. Similarly, $\Delta(\mathcal{G}_{m,n}(\Omega))$ and $\delta(\mathcal{G}_{m,n}(\Omega))$ are used to refer to the maximum and minimum degrees of the nodes in $\mathcal{G}_{m,n}(\Omega)$, respectively.

Definition 4 The sets of **bad/corrupted** and **good/correct** measurements are defined as $B = \{(i, j) | (i, j) \in \Omega, S_{ij} \neq 0\}$ and $G = \{(i, j) | (i, j) \in \Omega, S_{ij} = 0\}$, respectively.

Based on the above definitions, the sparsity graph is allowed to include self-loops. For a positive vector \mathbf{x} , we denote its maximum and minimum values with x_{\max} and x_{\min} , respectively. Furthermore, define $\kappa(\mathbf{x}) = \frac{x_{\max}}{x_{\min}}$ as the condition number of the vector \mathbf{x} . The first result of this paper develops deterministic conditions on the measurement set Ω and the sparsity pattern of the noise matrix S to guarantee that the positive rank-1 RPCA has benign landscape. Let \mathbf{u}^* and $(\mathbf{u}^*, \mathbf{v}^*)$ denote the true principal components of (SN-RPCA) and (AN-RPCA), respectively.

Theorem 5 (Informal, Deterministic Guarantee) Assuming that $\mathbf{u}^*, \mathbf{v}^* > 0$, there exist regularization functions $R_\beta(\mathbf{u})$ and $R_\beta(\mathbf{u}, \mathbf{v})$ such that the following statements hold with overwhelming probability:

1. (SN-RPCA) has no spurious local minimum and has a unique global minimum that coincides with the true component, provided that $\mathcal{G}(G)$ has **no bipartite component** and

$$\kappa(\mathbf{u}^*)^4 \Delta(\mathcal{G}(B)) \lesssim \delta(\mathcal{G}(G)) \quad (8)$$

2. (AN-RPCA) has no spurious local minimum and has a unique global minimum that coincides with the true components, provided that $\mathcal{G}_{m,n}(G)$ is **connected** and

$$\max \{\kappa(\mathbf{u}^*)^4, \kappa(\mathbf{v}^*)^4\} \Delta(\mathcal{G}_{m,n}(B)) \lesssim \delta(\mathcal{G}_{m,n}(G)) \quad (9)$$

Theorem 5 puts forward a set of deterministic conditions for the absence of spurious local solutions in (SN-RPCA) and (AN-RPCA) as well as the uniqueness of the global solution. Notice that no upper bound is assumed on the values of the nonzero entries in the noise matrix. The reasoning behind the conditions imposed on the minimum and maximum degrees of the nodes in the sparsity graph of the measurement set is to ensure the identifiability of the problem. We will elaborate more on this subtle point later in Section 7. Furthermore, we will show later in the paper that some of the conditions delineated in Theorem 5—such as the strict positivity of \mathbf{u}^* and \mathbf{v}^* , as well as the absence of bipartite components in $\mathcal{G}(G)$ for (SN-RPCA)—are also necessary for the exact recovery.

The second main result of this paper investigates (SN-RPCA) and (AN-RPCA) under random sampling and noise structures. In particular, suppose that each element (in the

symmetric case, each element of the upper triangular part) of S is nonzero with probability d . Then, for every (i, j) , we have

$$X_{ij} = \begin{cases} u_i^* v_j^* & \text{with probability } 1 - d \\ \text{arbitrary} & \text{with probability } d \end{cases} \quad (10)$$

Furthermore, suppose that every element of X is measured with probability p . In other words, every (i, j) belongs to Ω with probability p . Finally, we assume that the noise and sampling events are independent.

Theorem 6 (Informal, Probabilistic Guarantee) *Assuming that $\mathbf{u}^*, \mathbf{v}^* > 0$, there exist regularization functions $R_\beta(\mathbf{u})$ and $R_\beta(\mathbf{u}, \mathbf{v})$ such that the following statements hold with overwhelming probability:*

1. (SN-RPCA) *has no spurious local minimum and has a unique global minimum that coincides with the true component, provided that*

$$p \gtrsim \frac{\kappa(\mathbf{u}^*)^4 \log n}{n}, \quad d \lesssim \frac{1}{\kappa(\mathbf{u}^*)^4} \quad (11)$$

2. (AN-RPCA) *has no spurious local minimum and has a unique global minimum that coincides with the true components, provided that*

$$p \gtrsim \frac{\kappa(\mathbf{w}^*)^4 n \log n}{m^2}, \quad d \lesssim \frac{r}{\kappa(\mathbf{w}^*)^4} \quad (12)$$

where $\mathbf{w}^* = [\mathbf{u}^{*\top} \quad \mathbf{v}^{*\top}]^\top$, $r = m/n$, and $n \geq m$.

A number of interesting corollaries can be obtained based on Theorem 6. For instance, it can be inferred that the exact recovery is guaranteed even if the number of grossly corrupted measurements is on the same order as the total number of measurements, provided that $\frac{u_{\max}^*}{u_{\min}^*}$ is uniformly bounded from above.

In addition to the absence of spurious local minima and the uniqueness of the global minimum, the next proposition states that the true solution can be recovered via local search algorithms for non-smooth optimization.

Proposition 7 (Informal, Global Convergence) *Under the assumptions of Theorem 5 and 6, local search algorithms converge to the true solutions of (SN-RPCA) and (AN-RPCA) with overwhelming probability.*

Starting from Section 3.3, we will delve into the detailed analysis of the symmetric and asymmetric non-negative RPCA. In particular, we will analyze (SN-RPCA) and (AN-RPCA) under different deterministic and probabilistic settings and provide formal versions of Theorems 5 and 6.

3. Numerical Results

In this section, we demonstrate the efficacy of the above-mentioned results in different experiments. To this goal, first we briefly introduce the recently developed sub-gradient method (Li et al. 2020) that is specifically tailored to non-smooth and non-convex problems, such as those considered in this paper. The main advantage of the sub-gradient algorithm compared to other state-of-the-art methods is its extremely simple implementation; we present a sketch of the algorithm for solving the non-symmetric positive RPCA below¹ (the symmetric version can be solved using a similar algorithm with slight modifications):

Algorithm 1: Sub-gradient algorithm

```

1 Initialization: Strictly positive initial point  $\mathbf{w}_0^\top = [\mathbf{u}_0^\top \ \mathbf{v}_0^\top]^\top$  and step size  $\mu_0$ ;
2 for  $k = 0, 1, \dots$  do
3   | set  $\mathbf{d}_k$  as a sub-gradient of  $f_{\text{reg}}(\mathbf{u}_0, \mathbf{v}_0)$  defined in (AN-RPCA);
4   | set  $\mu_k$  according to a geometrically diminishing rule such that  $\mathbf{w}_k - \mu_k \mathbf{d}_k$  is
   | strictly positive;
5   | set  $\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_k \mathbf{d}_k$ ;
6 end

```

It has been shown in Li et al. (2020) that, under certain conditions on the initial point \mathbf{w}_0 , the initial step size μ_0 , and the update rule for μ_k , the iterates $\mathbf{w}_0, \mathbf{w}_1, \dots$ converge to the globally optimal solution at *linear* rate, provided that \mathbf{w}_0 is sufficiently close to the optimal solution. The closeness of \mathbf{w}_0 to \mathbf{w}^* is required partly to avoid becoming stuck at a spurious local minima. This requirement can be relaxed for the positive RPCA due to the absence of undesired spurious local solutions, as proven in this paper. It is also worthwhile to mention that, even though we use the sub-gradient algorithm to solve the positive RPCA, it will be shown in Section 8 that the results of this paper guarantee that a large class of local-search algorithms converge to the globally optimal solution of (SN-RPCA) or (AN-RPCA).

All of the following simulations are run on a laptop computer with an Intel Core i7 quad-core 2.50 GHz CPU and 16GB RAM. The reported results are for a serial implementation in MATLAB R2017b.

3.1 Exact Recovery:

To demonstrate the strength of the above-mentioned results, we consider thousands of randomly generated instances of the positive rank-1 RPCA with different sizes and noise levels. In particular, the dimension of the instances ranges from 10 to 100. For each instance, the elements of \mathbf{u}^* are uniformly chosen from the interval $[0, 2]$. Note that \mathbf{u}^* will be strictly positive with probability one. Furthermore, each element of the upper triangular part of the symmetric noise matrix S is set to 2 with probability d and 0 with probability $1 - d$. Figure 1a shows the performance of randomly initialized sub-gradient method for the symmetric positive rank-1 RPCA. We declare that a solution is recovered exactly if $\|\mathbf{u}\mathbf{u}^\top - \mathbf{u}^*\mathbf{u}^{*\top}\|_F / \|\mathbf{u}^*\mathbf{u}^{*\top}\|_F \leq 10^{-4}$. For each dimension and noise probability, we consider 100 randomly generated instances of the problem and demonstrate its exact recovery rate.

1. Note that this is a slightly modified version of the sub-gradient algorithm in Li et al. (2020) to ensure the positivity of the iterates.

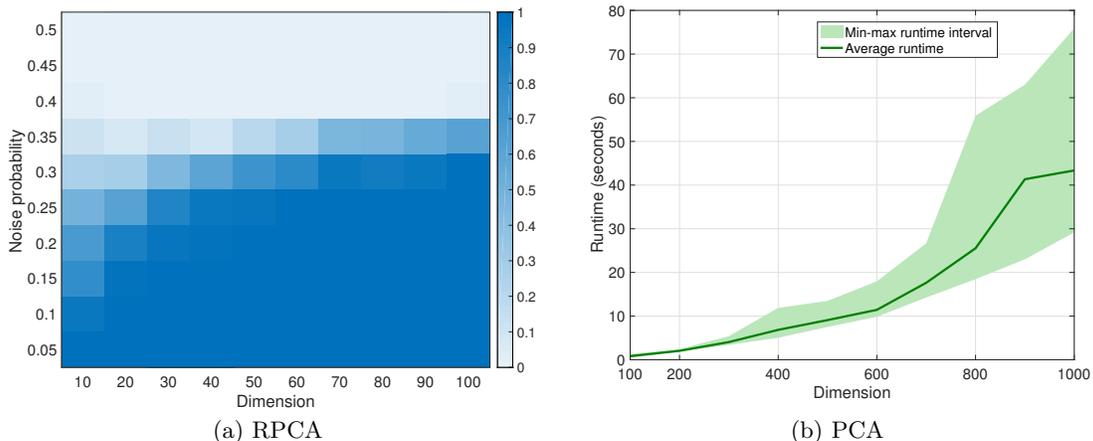


Figure 1: (a) The performance of the randomly initialized sub-gradient method for (SN-RPCA). The intensity of the color is proportional to the exact recovery rate of the true solution (darker blue implies higher recovery rate). (b) The runtime of the sub-gradient method for (SN-RPCA). For each dimension, it shows the average runtime and its min-max interval over 100 independent trials.

The heatmap shows the exact recovery rate of the sub-gradient method, when directly applied to (SN-RPCA). It can be observed that the algorithm has recovered the globally optimal solution even when 35% of the entries in the data matrix were severely corrupted with the noise. In contrast, even a highly sparse additive noise in the data matrix prevents the sub-gradient method from recovering the true solution, when applied to the smooth problem (6). Figure 1b shows the graceful scalability of the sub-gradient algorithm when applied to (SN-RPCA). It can be seen that the algorithm is highly efficient. In particular, its average runtime varies from 0.88 seconds for $n = 100$ to 43.20 seconds for $n = 1000$.

3.2 The Emergence of Local Solutions

Recall that \mathbf{u}^* and \mathbf{v}^* are both assumed to be strictly positive. In what follows, we will illustrate that relaxing these conditions to non-negativity gives rise to spurious local solutions. Consider an instance of the symmetric non-negative rank-1 RPCA with the parameters

$$\mathbf{u}^* = [1 \ 1 \ 0]^T, \quad S = 0, \quad \Omega = \{1, 2, 3\}^2 \setminus \{(3, 3)\} \tag{13}$$

Notice that \mathbf{u}^* consists of two strictly positive and one zero entries. Furthermore, this is a noiseless scenario where Ω consists of all possible measurements except for one. To examine the existence of spurious local solutions in this example, 10000 randomly initialized trials of the sub-gradient method is ran and the normalized distances between the obtained and true solutions are displayed in Figure 2. Based on this histogram, about 20% of the trials converge to spurious local solutions, implying that they are ubiquitous in this instance. This experiment shows why the positivity of the true solution is crucial and cannot be relaxed. We will formalize and prove this statement later in Section 6.

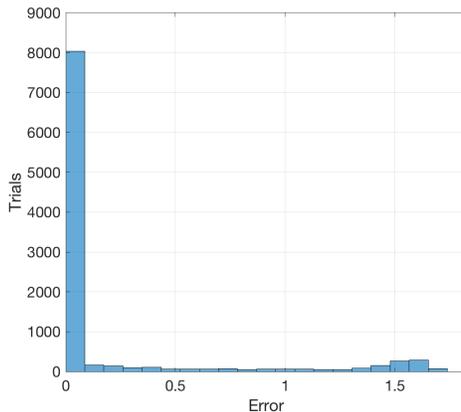


Figure 2: The normalized distance between the obtained solution using randomly initialized sub-gradient method and the true solution.

3.3 Moving Object Detection

In video processing, one of the most important problems is to detect anomaly or moving objects in different frames of a video. In particular, given a video sequence, the goal is to separate the nearly-static or slowly-changing background from the dynamic foreground objects (Cucchiara et al. 2003). Based on this observation, Candès et al. (2011) has proposed to model the background as a low-rank component, and the dynamic foreground as the sparse noise. In particular, suppose that the video sequence consists of d_f gray-scale frames, each with the resolution of $d_m \times d_n$ pixels. The data matrix X is defined as an asymmetric $d_m d_n \times d_f$ matrix whose i^{th} column is the vectorized version of the i^{th} frame. Therefore, the moving object detection problem can be cast as the recovery of the non-negative vectors $\mathbf{u} \in \mathbb{R}_+^{d_m d_n}$ and $\mathbf{v} \in \mathbb{R}_+^{d_f}$, as well as the sparse matrix $S \in \mathbb{R}^{d_m d_n \times d_f}$, such that

$$X \approx \mathbf{u}\mathbf{v}^T + S \quad (14)$$

Note that the background may not always have a rank-1 representation. However, we will show that (14) is sufficiently accurate if the background is relatively static. Furthermore, notice that when the background is completely static, the elements of \mathbf{v} should be equal to one. However, this is not desirable in practice since the background may change due to varying illuminations, which can be captured by the variable vector \mathbf{v} . Each entry of X is an integer between 0 (darkest) and 255 (brightest). To ensure the positivity of the true components, we increase each element of X by 1 without affecting the performance of the method.

The considered test case is borrowed from the work by Toyama et al. (1999)² and is a sequence of video frames taken from a room, where a person walks in, sits on a chair, and uses a phone. We consider 100 gray-scale frames of the sequence, each with the resolution of 120×160 pixels. Therefore, X , \mathbf{u} , and \mathbf{v} belong to $\mathbb{R}_+^{19,200 \times 100}$, $\mathbb{R}_+^{19,200}$, and \mathbb{R}_+^{100} ,

2. The video frames are publicly available at <https://www.microsoft.com/en-us/research/project/test-images-for-wallflower-paper/>.



Figure 3: The performance of the sub-gradient method in the moving object detection problem. The first row shows 3 out of 100 gray-scale frames in the studied test case that contain the moving objects. The second row shows the outcome of (SN-RPCA) solved using randomly initialized sub-gradient method.

respectively. Figure 3 shows that the sub-gradient method with a random initialization can recover the moving object, which is in accordance with the theoretical results of this paper.

4. Related Work

4.1 Non-convex and Low-rank Optimization

A considerable amount of work has been carried out to understand the inherent difficulty of solving low-rank optimization problems both locally and globally.

Convexification: Recently, there has been a pressing need to develop efficient methods for solving large-scale nonconvex optimization problems that naturally arise in data analytics and machine learning (Dumais et al. 1998; Sharif Razavian et al. 2014; Bottou et al. 2018; Zhang et al. 2018a; Olfat and Aswani 2018). One promising approach for making these large-scale problems more tractable is to resort to their convex surrogates; these methods started to receive a great deal of attention after the seminal works by Donoho (2006) and Candes et al. (2006) on the *compressive sensing* and have been extended to emerging problems in machine learning, such as fairness (Olfat and Aswani 2018), robust polynomial regression (Molybog et al. 2018; Madani et al. 2018), and neural networks (Bach 2017), to name a few. Nonetheless, the size of today’s problems has been a major impediment to the tractability of these methods. In practice, the dimension of the real-world problems is overwhelmingly large, often surpassing the ability of these seemingly efficient convex methods to solve the problem in a reasonable amount of time. Due to this so-called *curse of dimensionality*, the common practice is to deploy fast local search algorithms directly applied to the original nonconvex problem with the hope of converging to acceptable solutions. Roughly

speaking, these methods can only guarantee the local optimality, thus exposing themselves to potentially large optimality gaps. However, a recent line of work has shown that a surprisingly large class of nonconvex problems, including matrix completion/sensing (Bhojanapalli et al. 2016; Ge et al. 2016, 2017; Zhu et al. 2017), phase retrieval (Sun et al. 2018), and dictionary recovery (Sun et al. 2017) have *benign global landscape*, i.e., every local solution is also global and every saddle point has a direction with a strictly negative curvature (see Chi et al. (2019) for a comprehensive survey on the related problems). More recently, the work by Zhang et al. (2018c) has introduced a unified framework that shows the benign landscape of nonconvex low-rank optimization problems with general loss functions, provided that they satisfy certain restricted convexity and smoothness properties. This enables most of the saddle-escaping local search algorithms to converge to a global solution, thereby resulting in a zero optimality gap (Ge et al. 2015).

Benign landscape: As mentioned before, it has been recently shown that many low-rank optimization problems can be cast as smooth-but-nonconvex optimization problems that are free of spurious local minima. These methods heavily rely on the notion of *restricted isometry property* (RIP)—a property that was initially introduced by Candes and Tao (2005) and has been used ever since as a metric to measure a norm-preserving property of the objective function. In general, these methods have two major drawbacks: 1) they can only target a narrow set of nearly-isotropic instances (Zhang et al. 2018b), and 2) their proof technique depends on the differentiability of the objective function; a condition that is not satisfied for non-smooth norms, such as ℓ_1 . To the best of our knowledge, the work by Josz et al. (2018) is the only one that studies the landscape of the ℓ_1 minimization problem, where the authors consider the tensor decomposition problem under the full and perfect measurements. Our work is somewhat related to Ma et al. (2018) that derives similar conditions for the absence of spurious local solution of the non-negative rank-1 matrix completion but for the smooth Frobenius norm minimization problem.

PCA with prior information: With an exponential growth in the size and dimensionality of the real-world datasets, it is often required to exploit the additional prior information in the PCA. In many real-world applications, prior knowledge from the underlying physics of the problem—such as non-negativity (Montanari and Richard 2016), sparsity (Zou et al. 2006), robustness (Candès et al. 2011), and nonlinearity (Gorban et al. 2008)—can be taken into account to perform more efficient, consistent, and accurate PCA.

Numerical algorithms for non-smooth optimization: Numerical algorithms for non-smooth optimization problems can be dated back to the work by Clarke on the extended definitions of gradients and directional derivatives, commonly known as generalized derivatives (Clarke 1990). Intuitively, for non-smooth functions, the gradient in the classical sense cease to exist at a subset of the points in the domain. The Clarke generalized derivative is introduced to circumvent this issue by associating a convex differential to these points, even if the original problem is non-convex. In the domain of unconstrained non-smooth optimization, earlier works have introduced simple algorithms that converge to approximate Clarke-stationary points (Goldstein 1977; Chaney and Goldstein 1978). More recent methods take advantage of the fact that many non-smooth optimization problems are smooth in every open dense subset of their domains. This implies that the objective function is smooth with probability one at a randomly drawn point. This observation lays the groundwork for

several gradient-sampling-based algorithms for both unconstrained and constrained non-smooth optimization problems (Burke et al. 2005; Curtis and Overton 2012). As mentioned before, a sub-gradient method has been recently proposed by Li et al. (2020) for solving the RPCA, where the authors prove linear convergence of the algorithm to the true components, provided that the initial point is chosen sufficiently close to the globally optimal solution.

4.2 Comparison to the Existing Results on RPCA

Similar to the non-convex matrix sensing and completion, most of the existing results on the RPCA work on a *lifted* space of the variables via different convex relaxations and they do not incorporate the positivity constraints in the problem. In what follows, we will explain the advantages of our proposed method compared to these results.

Positivity constraints: In the present work, we show that the positivity of the true components is both sufficient and (almost) necessary for the absence of spurious local solutions. We use this prior knowledge to obtain sharp deterministic and probabilistic guarantees on the absence of spurious local minima for the RPCA based on the Burer-Monteiro formulation. For instance, we show that up to a constant factor of the measurements can be grossly corrupted and yet they do not introduce any spurious local solution. Considering the fact that these results heavily rely on the positivity of the true components, it is unclear if similar “no spurious local minima” results hold for the general case without the positivity assumption. The statistical properties of these types of constraints have also been shown to be useful in the classical PCA by Montanari and Richard (2016), where the authors show that by imposing positivity constraints on the principal components, one can guarantee its consistent recovery with smaller signal-to-noise ratio. It is also worthwhile to mention that the incorporation of the non-negativity/positivity constraints in the low-rank matrix recovery can be traced back to some earlier works on the non-negative matrix factorization problem (Lee and Seung 1999; Hoyer 2004).

Computational savings: Similar to the convexification techniques in nonconvex optimization, most of the classical results on the RPCA relax the inherent non-convexity of the problem by lifting it to higher dimensions (Candès et al. 2011; Chandrasekaran et al. 2011; Zhou et al. 2010; Hsu et al. 2011). In particular, by moving from vector to matrix variables, they guarantee the convexity of the problem at the expense of significantly increasing the number of variables. In this work, we show that such lifting is not necessary for the positive rank-1 RPCA since—despite the non-convexity of the problem—it is free of spurious local solutions and, hence, simple local search algorithms converge to the true components when directly applied to its original formulation. More recently, a special attention has been devoted to nonconvex methods for solving robust matrix sensing and robust PCA problems (Gu et al. 2016; Netrapalli et al. 2014; Zhang et al. 2018d). In particular, the work by Netrapalli et al. (2014) proposes an alternating framework for solving robust PCA, where the low-rank component and the sparse noise matrix are obtained by successively projecting their corresponding residuals onto the sets of low-rank and sparse matrices, respectively. The work by Gu et al. (2016) proposes a nonconvex alternating minimization approach for solving robust PCA. More generally, the work by Zhang et al. (2018d) proposes a unified nonconvex framework based on projected gradient descent and a thresholding operator for solving robust matrix sensing and robust PCA problems. Although being cheaper than con-

vexification techniques, the success of the above methods heavily relies on the closeness of their initial points to the optimal solution, and therefore, they fall short of proving general results on the benign landscape of these problems.

Sharp guarantees with mild conditions: In general, most of the existing results on RPCA for guaranteeing the recovery of the true components fall into two categories. First, a large class of methods rely on some deterministic conditions on the spectra of the dominant components and/or the structure of the sparse noise (Hsu et al. 2011; Chandrasekaran et al. 2011; Yi et al. 2016). For instance, the works by Hsu et al. (2011); Chandrasekaran et al. (2011) require the regularization coefficient to be within a specific interval that is defined in terms of the true principal components. Furthermore, the algorithm proposed by Yi et al. (2016) requires prior knowledge on the density of the sparse noise matrix. Although being theoretically significant, these types of conditions cannot be easily verified and met in practice. With the goal of bypassing such stringent conditions, the second category of research has studied the RPCA under probabilistic models. These types of guarantees were popularized by Candès et al. (2011); Wright et al. (2009) and they do not rely on any prior knowledge on the true components or the density of the noise matrix. However, their success is contingent upon specific random models on the sparse noise or the spectra of the true components, neither of which may be satisfied in practice.

In contrast, the method proposed here does not rely on any prior knowledge on the true solution, other than the availability of an upper bound on the maximum absolute value of the elements in the principal components³. Furthermore, unlike the previous works, our results encompass *both deterministic and probabilistic* models under random sampling.

5. Preliminaries

A **directional derivative** of a locally Lipschitz and possibly non-smooth function $h(\mathbf{x})$ at \mathbf{x} in the direction \mathbf{d} is defined as

$$h'(\mathbf{x}, \mathbf{d}) := \lim_{t \downarrow 0} \frac{h(\mathbf{x} + t\mathbf{d}) - h(\mathbf{x})}{t} \quad (15)$$

upon existence. Based on this definition, $\bar{\mathbf{u}}$ is **directional-minimum-stationary** (or D-min-stationary) for (SN-RPCA) if $f'(\bar{\mathbf{u}}, \mathbf{d}) \geq 0$ for every *feasible* direction \mathbf{d} , i.e., a direction that satisfies $d_i \geq 0$ when $u_i = 0$ for every index i . Similarly, $\bar{\mathbf{u}}$ is **directional-maximum-stationary** (or D-max-stationary) for (SN-RPCA) if $f'(\bar{\mathbf{u}}, \mathbf{d}) \leq 0$ for every feasible \mathbf{d} . Finally, $\bar{\mathbf{u}}$ is **directional-stationary** (or D-stationary) for (SN-RPCA) if it is either D-min- or D-max-stationary⁴.

Every local minimum (maximum) $\bar{\mathbf{u}}$ should be D-min (max)-stationary for $f(\mathbf{u})$. On the other hand, $\bar{\mathbf{u}}$ cannot be a D-stationary point if $f(\mathbf{u})$ has strictly positive and negative directional derivatives at that point. In that case, $\bar{\mathbf{u}}$ is neither local maximum nor minimum. A solution to a minimization problem is referred to as **spurious local** (or simply local)

3. Note that in most cases, these types of upper bounds can be immediately inferred by the domain knowledge; see e.g. our discussion on the moving object detection problem.

4. Note that the notion of D-stationary points is often used in lieu of D-min-stationary in the literature. However, we use a slightly more general definition in this paper to account for the local maxima of (SN-RPCA).

if there exists another feasible point with a strictly smaller objective value; a solution is **globally optimal** (or simply global) if no such point exists.

Finally, a **vertex partitioning** of a non-empty bipartite graph is the partition of its vertices into two groups such that there exist no adjacent vertices within each group.

Notation: The upper-case, bold lower-case, and lower-case letters are used to show the matrices, vectors, and scalars, respectively. The space of non-negative and real $n \times 1$ vectors and $m \times n$ matrices are denoted by \mathbb{R}_+^n and $\mathbb{R}_+^{n \times m}$, respectively. The symbols $\|W\|_1$ and $\|W\|_F$ denote the element-wise ℓ_1 norm and Frobenius norm of W , respectively. The $(i, j)^{\text{th}}$ entry of a matrix W is shown as W_{ij} , whereas the i^{th} entry of a vector \mathbf{w} is denoted by w_i . Given the sequences $f_1(n)$ and $f_2(n)$, the notation $f_1(n) \lesssim f_2(n)$ or equivalently $f_1(n) = O(f_2(n))$ means that there exists a number $c_1 \in [0, \infty)$ such that $f_1(n) \leq c_1 f_2(n)$ for all n . Similarly, the notation $f_1(n) \gtrsim f_2(n)$ or $f_1(n) = \Omega(f_2(n))$ means that there exists a number $c_2 > 0$ such that $f_1(n) \geq c_2 f_2(n)$ for all n . The indicator function $\mathbb{I}_{x \geq \alpha}$ takes the value 1 if $x \geq \alpha$ and 0 otherwise. For an event \mathcal{E} , the notation $\mathbb{P}(\mathcal{E})$ is used to show the probability of its occurrence. For a random variable X , the symbol $\mathbb{E}\{X\}$ shows its expected value. For notational simplicity and unless stated otherwise, we will refer to non-negative (or positive) rank-1 RPCA as non-negative (or positive) RPCA in the sequel.

6. Base Case: Noiseless Non-negative RPCA

In this section, we consider the noiseless version of both symmetric and asymmetric non-negative RPCA. While not entirely obvious, the subsequent arguments are at the core of our proofs for the general noisy case. In the noiseless scenario, (SN-RPCA) is reduced to

$$\min_{\mathbf{u} \geq 0} \underbrace{\sum_{(i,j) \in \Omega} |u_i u_j - u_i^* u_j^*|}_{f(\mathbf{u})} \quad (\text{P1-Sym})$$

For the asymmetric problem (AN-RPCA), the solution is invariant to scaling. In other words, if (\mathbf{u}, \mathbf{v}) is a solution to (AN-RPCA), then $(\frac{1}{q}\mathbf{u}, q\mathbf{v})$ is also a valid solution with the same objective value, for every scalar $q > 0$. To circumvent the issue of invariance to scaling, it is common to balance the norms of \mathbf{u} and \mathbf{v} by penalizing their difference. Therefore, similar to the works by Ge et al. (2017); Zheng and Lafferty (2016); Yi et al. (2016), we consider the following regularized variant of (AN-RPCA):

$$\min_{\mathbf{u} \geq 0, \mathbf{v} \geq 0} \underbrace{\|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^\top)\|_1 + \alpha \|\mathbf{u}^\top \mathbf{u} - \mathbf{v}^\top \mathbf{v}\|}_{f_{\text{asym}}(\mathbf{u}, \mathbf{v})} \quad (16)$$

for an arbitrary constant $\alpha > 0$ (note that the positivity of α is the only condition required in this work). To deal with the asymmetric case, we first convert it to a symmetric problem after a simple concatenation of variables. Define $\mathbf{w} = [\mathbf{u}^\top \ \mathbf{v}^\top]^\top$, $\mathbf{w}^* = [\mathbf{u}^{*\top} \ \mathbf{v}^{*\top}]^\top$, and $\bar{\Omega} = \{(i, j) | (i, j - m) \in \Omega\}$. Based on these definitions, one can symmetrize (16) as follows:

$$\min_{\mathbf{w} \geq 0} \underbrace{\sum_{(i,j) \in \bar{\Omega}} |w_i w_j - w_i^* w_j^*| + \alpha \left| \sum_{i=1}^m w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \right|}_{f_{\text{sym}}(\mathbf{w})} \quad (\text{P1-Asym})$$

To simplify the notation, we drop the subscript from $f_{\text{sym}}(\mathbf{w})$ whenever there is no ambiguity in the context.

6.1 Deterministic Guarantees

Symmetric case: First, we introduce deterministic conditions to guarantee a benign landscape for (P1-Sym).

Theorem 8 *Suppose that $\mathbf{u}^* > 0$ and $\mathcal{G}(\Omega)$ has no bipartite component. Then, the following statements hold for (P1-Sym):*

1. *It does not have any spurious local minimum;*
2. *The point $\mathbf{u} = \mathbf{u}^*$ is the unique global minimum;*
3. *In the positive orthant, the point $\mathbf{u} = \mathbf{u}^*$ is the only D-stationary point.*

Additionally, if $\mathcal{G}(\Omega)$ is connected, the following statements hold for (P1-Sym):

4. *The points $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{u} = 0$ are the only D-min-stationary points;*
5. *The point $\mathbf{u} = 0$ is a local maximum.*

The above theorem has a number of important implications for (P1-Sym): 1) it has no spurious local solution, 2) $\mathbf{u} = \mathbf{u}^*$ is its unique global solution, and 3) every feasible point $\mathbf{u} > 0$ such that $\mathbf{u} \neq \mathbf{u}^*$ has at least a strictly negative directional derivative. Additionally, if $\mathcal{G}(\Omega)$ is connected, the feasible points of (P1-Sym) with zero entries either have a strictly negative directional derivative or correspond to the origin that is a local maximum with a strictly negative curvature. Therefore, these points are not local/global minima and can be easily avoided using local search algorithms.

To prove Theorem 8, we first need the following important lemma.

Lemma 9 *Suppose that $\mathcal{G}(\Omega)$ has no bipartite component and $\mathbf{u}^* > 0$. Then, for every D-min-stationary point \mathbf{u} of (P1-Sym), we have $\mathbf{u}[c] > 0$ or $\mathbf{u}[c] = 0$, where $\mathbf{u}[c]$ is a sub-vector of \mathbf{u} induced by the c^{th} component of $\mathcal{G}(\Omega)$.*

Proof See Appendix A. ■

Now, we are ready to present the proof of Theorem 8.

Proof of Theorem 8: We prove the first three statements. Note that Statement 5 can be easily verified and Statement 4 is implied by Lemma 9 and Statement 3.

Suppose that $\mathbf{u} \neq \mathbf{u}^*$ is a local minimum. Note that if $u_i = 0$ for some i , Lemma 9 implies that $\mathbf{u}[c] = 0$ for the c^{th} component that includes node i . However, a strictly positive perturbation of $\mathbf{u}[c]$ decreases the objective function and, therefore, \mathbf{u} cannot be a local minimum. Hence, it is enough to consider the case $\mathbf{u} > 0$. We show that \mathbf{u} cannot be D-stationary. This immediately certifies the validity of the first three statements. First, we prove that

$$\min_{k \in \Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_i}{u_i^*} \leq \max_{k \in \Omega_i} \frac{u_k^*}{u_k} \tag{17}$$

for every $i \in \{1, \dots, n\}$, where $\Omega_i = \{j | (i, j) \in \Omega\}$. By contradiction and without loss of generality, suppose that $u_i/u_i^* > \max_{k \in \Omega_i} u_k^*/u_k$ for some i . This implies that $u_i u_j > u_i^* u_j^*$ for every $j \in \Omega_i$. Therefore, a negative or positive perturbation of u_i results in respective negative or positive directional derivatives, contradicting the D-stationarity of \mathbf{u} . With no loss of generality, assume that the sparsity graph $\mathcal{G}(\Omega)$ is connected (since the arguments made in the sequel can be readily applied to every disjoint component of $\mathcal{G}(\Omega)$) and that the following ordering holds:

$$0 < \frac{u_1^*}{u_1} \leq \frac{u_2^*}{u_2} \leq \dots \leq \frac{u_n^*}{u_n} \quad (18)$$

Therefore, due to (17), we have

$$0 < \frac{u_1^*}{u_1} \leq \min_{k \in \Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_i}{u_i^*} \leq \max_{k \in \Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_n^*}{u_n} \quad (19)$$

for every $i \in \{1, \dots, n\}$.

Since $\mathbf{u} \neq \mathbf{u}^*$, there exists some index t such that $u_t \neq u_t^*$. This implies that $u_n^*/u_n > 1$; otherwise, we should have $u_n^*/u_n \leq 1$. This together with (18), implies that $u_t^*/u_t < 1$ and $u_t/u_t^* > 1$, which contradicts (19). Now, define the sets

$$T_1 = \left\{ i \mid \frac{u_i^*}{u_i} = \frac{u_n^*}{u_n}, 1 \leq i \leq n \right\} \quad (20)$$

$$T_2 = \left\{ j \mid \frac{u_j}{u_j^*} = \frac{u_n^*}{u_n}, 1 \leq j \leq n \right\} \quad (21)$$

Moreover, define the set $N = V \setminus (T_1 \cup T_2)$ and let \mathbf{d} be

$$d_i = \begin{cases} \frac{u_i}{u_n} & \text{if } i \in T_1 \\ -\frac{u_i}{u_n} & \text{if } i \in T_2 \\ 0 & \text{if } i \in N \end{cases} \quad (22)$$

Define a perturbation of \mathbf{u} as $\hat{\mathbf{u}} = \mathbf{u} + \mathbf{d}\epsilon$ where $\epsilon > 0$ is chosen to be sufficiently small. Next, the effect of the above perturbation on different terms of (P1-Sym) will be analyzed. To this goal, we divide Ω into four sets

1. $(i, j) \in \Omega$ and $i, j \in T_1$: In this case, since $u_i < u_i^*$ and $u_j < u_j^*$, one can write

$$\begin{aligned} |\hat{u}_i \hat{u}_j - u_i^* u_j^*| &= u_i^* u_j^* - \hat{u}_i \hat{u}_j = u_i^* u_j^* - \left(u_i + \frac{u_i}{u_n} \epsilon \right) \left(u_j + \frac{u_j}{u_n} \epsilon \right) \\ &= |u_i u_j - u_i^* u_j^*| - \left(\frac{2u_i u_j}{u_n} \right) \epsilon - \left(\frac{u_i u_j}{u_n^2} \right) \epsilon^2 \end{aligned} \quad (23)$$

where we have used the assumption $\mathbf{u}^*, \mathbf{u} > 0$.

2. $(i, j) \in \Omega$ and $i, j \in T_2$: In this case, since $u_i > u_i^*$ and $u_j > u_j^*$, one can write

$$\begin{aligned} |\hat{u}_i \hat{u}_j - u_i^* u_j^*| &= \hat{u}_i \hat{u}_j - u_i^* u_j^* = \left(u_i - \frac{u_i}{u_n} \epsilon \right) \left(u_j - \frac{u_j}{u_n} \epsilon \right) - u_i^* u_j^* \\ &= |u_i u_j - u_i^* u_j^*| - \left(\frac{2u_i u_j}{u_n} \right) \epsilon + \left(\frac{u_i u_j}{u_n^2} \right) \epsilon^2 \end{aligned} \quad (24)$$

where we have used the assumption $\mathbf{u}^*, \mathbf{u} > 0$.

3. $(i, j) \in \Omega$, $i \in N$, and $j \in T_1 \cup T_2$: According to the definitions of T_1 and T_2 , we have

$$\frac{u_i}{u_i^*} < \frac{u_n}{u_n^*}, \quad \frac{u_i^*}{u_i} < \frac{u_n^*}{u_n} \quad (25)$$

Now, if $j \in T_1$, one can write

$$\frac{u_i}{u_i^*} < \frac{u_j^*}{u_j} \implies u_i u_j < u_i^* u_j^* \quad (26)$$

which implies that

$$|\hat{u}_i \hat{u}_j - u_i^* u_j^*| = u_i^* u_j^* - \hat{u}_i \hat{u}_j = u_i^* u_j^* - u_i \left(u_j + \frac{u_j}{u_n} \epsilon \right) = |u_i u_j - u_i^* u_j^*| - \left(\frac{u_i u_j}{u_n} \right) \epsilon \quad (27)$$

Similarly, if $j \in T_2$, one can verify that

$$|\hat{u}_i \hat{u}_j - u_i^* u_j^*| = |u_i u_j - u_i^* u_j^*| - \left(\frac{u_i u_j}{u_n} \right) \epsilon \quad (28)$$

4. $(i, j) \in \Omega$, $i \in T_1$, and $j \in T_2$: In this case, note that

$$|\hat{u}_i \hat{u}_j - u_i^* u_j^*| = \left| \left(u_i + \frac{u_i}{u_n} \epsilon \right) \left(u_j - \frac{u_j}{u_n} \epsilon \right) - u_i^* u_j^* \right| \leq |u_i u_j - u_i^* u_j^*| + \left(\frac{u_i u_j}{u_n^2} \right) \epsilon^2 \quad (29)$$

The above analysis entails that—unless N and the subgraphs of $\mathcal{G}(\Omega)$ induced by the nodes in T_1 or T_2 are empty— $f'(\mathbf{u}, \mathbf{d}) > 0$ and $f'(\mathbf{u}, -\mathbf{d}) < 0$, implying that \mathbf{u} cannot be D-stationary. On the other hand, these conditions enforce $\mathcal{G}(\Omega)$ to be bipartite, which is a contradiction. This completes the proof. \blacksquare

Next, we show that $\mathbf{u}^* > 0$ is *almost* necessary to guarantee the absence of spurious local minima for (P1-Sym).

Proposition 10 *Assume that $\mathbf{u}^* \geq 0$ and that $\mathbf{u}^* \neq 0$ with $u_i^* = 0$ for some i . Then, upon choosing $\Omega = \{1, \dots, n\}^2 \setminus \{(i, i)\}$, (P1-Sym) has a spurious local minimum.*

Proof See Appendix B. \blacksquare

The above corollary shows that if \mathbf{u}^* is non-negative with at least one zero element, even in the almost perfect scenario where the set Ω includes all of the measurements except for one, it may not be free of spurious local minima. The next corollary shows that the assumption on the absence of bipartite components in $\mathcal{G}(\Omega)$ is also necessary for the uniqueness of the global solution.

Proposition 11 *Given any vector $\mathbf{u}^* > 0$ and set Ω , suppose that $\mathcal{G}(\Omega)$ has a bipartite component. Then, the global solution of (P1-Sym) is not unique.*

Proof Without loss of generality, suppose that $\mathcal{G}(\Omega)$ is a connected bipartite graph. For any vector $\mathbf{u}^* > 0$, the solution $\mathbf{u} = \mathbf{u}^*$ is globally optimal for (P1-Sym). Suppose that the bipartite graph $\mathcal{G}(\Omega)$ partitions the entries of \mathbf{u} into two sets V_1 and V_2 such that $u_n \in V_1$. Based on some simple algebra, one can easily verify that, for a sufficiently small $\epsilon > 0$, the solution

$$\hat{u}_i \leftarrow \begin{cases} u_i + \frac{u_i}{u_n} \epsilon & \text{if } i \in V_1 \\ u_i - \frac{u_i}{u_n + \epsilon} \epsilon & \text{if } i \in V_2 \end{cases} \quad (30)$$

is also globally optimal for (P1-Sym). ■

Remark 12 *Suppose that \mathbf{u}^* is a globally optimal solution of (P1-Sym) and that $\mathcal{G}(G)$ includes a bipartite component. Then, according to Proposition 11, the part of \mathbf{u}^* whose elements correspond to the nodes in this bipartite component can be perturbed to attain another globally optimal solution, thereby resulting in the **non-uniqueness of the global solution**. On the other hand, the connectedness assumption is required to eliminate the undesirable stationary points on the boundary of the feasible region. Roughly speaking, the elements of the vector variable \mathbf{u} corresponding to different disconnected components can behave independently from each other, giving rise to spurious D -stationary points in the problem. To elaborate, recall that $\mathbf{u}[c]$ is a sub-vector of \mathbf{u} induced by the c^{th} component of $\mathcal{G}(G)$. Based on Lemma 9, the D -stationary points restricted to each disjoint component of $\mathcal{G}(G)$ are either strictly positive or equal to zero. Therefore, upon having two disconnected components c_1 and c_2 , the points $\mathbf{u}' = [\mathbf{u}^*[c_1]^\top \ 0]^\top$ and $\mathbf{u}'' = [0 \ \mathbf{u}^*[c_2]^\top]^\top$ are indeed D -stationary points of (SN-RPCA), thereby resulting in **spurious stationary points**.*

Asymmetric case: Next, we consider (16) in the noiseless scenario by analyzing its symmetrized counterpart (P1-Asym). Based on the construction of $\bar{\Omega}$, the corresponding sparsity graph $\mathcal{G}(\bar{\Omega})$ is bipartite. On the other hand, according to Proposition 11, the existence of a bipartite component in $\mathcal{G}(\bar{\Omega})$ makes a part of the solution *invariant to scaling*, which subsequently results in the non-uniqueness of the global minimum. The additional regularization term in (P1-Asym) is introduced to circumvent this issue by penalizing the difference in the norms of \mathbf{u} and \mathbf{v} .

Theorem 13 *Suppose that $\mathbf{w}^* > 0$ and $\mathcal{G}(\bar{\Omega})$ is connected. Then, the following statements hold for (P1-Asym):*

1. *The points $\mathbf{w} = 0$ and \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ are the only D -min-stationary points;*
2. *The point $\mathbf{w} = 0$ is a local maximum;*
3. *In the positive orthant, the point \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ is the only D -stationary point.*

Proof See Appendix C. ■

Remark 14 Notice that, unlike the symmetric case, Theorem 13 requires the connectedness of $\mathcal{G}(\bar{\Omega})$. This is due to the additional regularization term in (AN-RPCA). In particular, similar arguments do not necessarily hold for the disjoint components of $\mathcal{G}(\bar{\Omega})$ because of the coupling nature of the regularization term.

6.2 Probabilistic Guarantees

Next, we consider the random sampling regime. Similar to the previous subsection, we first focus on the symmetric case.

Symmetric case: Suppose that every element of the upper triangular part of the matrix $\mathbf{u}^* \mathbf{u}^{*\top}$ is measured independently with probability p . In other words, for every $(i, j) \in \{1, 2, \dots, n\}^2$ and $i \leq j$, the probability of (i, j) belonging to Ω is equal to p .

Theorem 15 Suppose that $n \geq 2$, $\mathbf{u}^* > 0$, and $p \geq \min \left\{ 1, \frac{(2\eta+2) \log n+2}{n-1} \right\}$ for some constant $\eta \geq 1$. Then, the following statements hold for (SN-RPCA) with probability of at least $1 - \frac{3}{2}n^{-\eta}$:

1. The points $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{u} = 0$ are the only D -min-stationary points;
2. The point $\mathbf{u} = 0$ is a local maximum;
3. In the positive orthant, the point $\mathbf{u} = \mathbf{u}^*$ is the only D -stationary point.

Before presenting the proof of Theorem 15, we note that the required lower bound on p is to guarantee that the random graph $\mathcal{G}(\Omega)$ is connected with high probability. This implies that Theorem 8 can be invoked to verify the statements of Theorem 15. It is worthwhile to mention that the classical results on *Erdős-Rényi* graphs characterize the *asymptotic* properties of $\mathcal{G}(\Omega)$ as n approaches infinity. In particular, it is shown by Erdős and Rényi (1959) that with the choice of $p = \frac{\log n+c}{n}$ for some $c > 0$, $\mathcal{G}(\Omega)$ becomes connected with probability of at least $\Omega(e^{-e^{-c}})$ as $n \rightarrow \infty$. In contrast, we introduce the following non-asymptotic result characterizing the probability that $\mathcal{G}(\Omega)$ is connected and non-bipartite for any finite $n \geq 2$, and subsequently use it to prove Theorem 15.

Lemma 16 Given a constant $\eta \geq 1$, suppose that $p \geq \min \left\{ 1, \frac{(2\eta+2) \log n+2}{n-1} \right\}$ and $n \geq 2$. Then, $\mathcal{G}(\Omega)$ is connected and non-bipartite with probability of at least $1 - \frac{3}{2}n^{-\eta}$.

Proof See Appendix D. ■

Proof of Theorem 15: The proof immediately follows from Theorem 8 and Lemma 16. ■

Similar to the deterministic case, we will show that both assumptions $\mathbf{u}^* > 0$ and $p \gtrsim \log n/n$ are *almost* necessary for the successful recovery of the global solution of (P1-Sym). In particular, it will be proven that relaxing $\mathbf{u}^* > 0$ to $\mathbf{u}^* \geq 0$ will result in an instance that possesses a spurious local solution with non-negligible probability. Furthermore, it will be shown that the choice $p \approx \log n/n$ is optimal—modulo $\log n$ -factor—for the unique recovery of the global solution.

Proposition 17 *Assuming that $\mathbf{u}^* \geq 0$ with $u_i^* = 0$ for some $i \in \{1, \dots, n\}$ and that $p < 1$, (P1-Sym) has a spurious local minimum with probability of at least $1 - p > 0$.*

Proof Suppose that $\mathbf{u}^* \geq 0$ and there exists an index i such that $u_i^* = 0$. The proof of Proposition 10 can be used to show that excluding the measurement (i, i) gives rise to a spurious local minimum. This occurs with probability $1 - p$. The details are omitted due to their similarities to the proof of Proposition 10. ■

Proposition 18 *Given any $\mathbf{u}^* > 0$, suppose that $np \rightarrow 0$ as $n \rightarrow \infty$. Then, the global solution of (P1-Sym) is not unique with probability approaching to one.*

Proof See Appendix E. ■

Asymmetric case: Consider (16) under a random sampling regime, where each element of $\mathbf{u}^* \mathbf{v}^{*\top}$ is independently observed with probability p . Next, the analog of Theorem 15 for the asymmetric case is provided.

Theorem 19 *Suppose that $n, m \geq 2$, $\mathbf{w}^* > 0$, and $p \geq \min \left\{ 1, \frac{(m+n)((1+\eta)\log(mn)+1)}{(m-1)(n-1)} \right\}$ for some constant $\eta \geq 1$. Then, the following statements hold for (P1-Asym) with probability of at least $1 - 2(mn)^{-\eta} - 4(mn)^{-2\eta}$:*

1. *The points $\mathbf{w} = 0$ and \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^* \mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ are the only D -min-stationary points;*
2. *The point $\mathbf{w} = 0$ is a local maximum;*
3. *In the positive orthant, the point \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^* \mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ is the only D -stationary point.*

Before presenting the proof of Theorem 19, we note that $\mathcal{G}(\bar{\Omega})$ no longer corresponds to an Erdős-Rényi random graph due to its bipartite structure. Therefore, we present the analog of Lemma 16 for random bipartite graphs.

Lemma 20 *Given a constant $\eta \geq 1$, suppose that $p \geq \min \left\{ 1, \frac{(m+n)((1+\eta)\log(mn)+1)}{(m-1)(n-1)} \right\}$ and $m, n \geq 2$. Then, $\mathcal{G}(\bar{\Omega})$ is connected with probability of at least $1 - 2(mn)^{-\eta} - 4(mn)^{-2\eta}$.*

Proof See Appendix F. ■

Proof of Theorem 19: The proof immediately follows from Theorem 13 and Lemma 20. ■

Before proceeding, we note that, similar to the classical results on the *Erdős-Rényi* graphs, there are asymptotic results guaranteeing the connectedness of a random bipartite graph as a function of p . In particular, Saltykov (1995) shows that $\mathcal{G}(\bar{\Omega})$ is connected with probability approaching to 1 as $m + n \rightarrow \infty$, provided that $p \geq 3 \left(1 + \frac{m}{n}\right)^{-1} \frac{(n+m)\log(n+m)}{nm}$. Lemma 20 offers another lower bound on p that matches this threshold (modulo a constant factor), while being non-asymptotic in nature. In particular, it characterizes the probability that the random bipartite graph is connected for *all* $m, n \geq 2$.

7. Extension to Noisy Positive RPCA

In this section, we will show that an additive sparse noise with arbitrary values does not drastically change the landscape of the RPCA. In other words, a limited number of grossly wrong measurements will not introduce any spurious local solution to the positive RPCA. The key idea is to prove that the direction of descent that was introduced in the previous section is also valid when the measurements are not perfect, i.e., when they are subject to sparse noise. To this goal, consider the following problem in the symmetric case:

$$\min_{\mathbf{u} \geq 0} \underbrace{\sum_{(i,j) \in \Omega} |u_i u_j - X_{ij}|}_{f(\mathbf{u})} \quad (31)$$

where

$$X = \mathbf{u}^* \mathbf{u}^{*\top} + S \quad (32)$$

is the matrix of true measurements perturbed with sparse noise. Similarly, consider the following problem for the asymmetric case:

$$\min_{\mathbf{u} \geq 0, \mathbf{v} \geq 0} \sum_{(i,j) \in \Omega} |u_i v_j - X_{ij}| + \alpha \left| \sum_{i=1}^m u_i^2 - \sum_{j=1}^n v_j^2 \right| \quad (33)$$

where α is an arbitrary positive number. After symmetrization, (33) can be re-written as

$$\min_{\mathbf{w} \geq 0} \underbrace{\sum_{(i,j) \in \bar{\Omega}} |w_i w_j - \bar{X}_{ij}| + \alpha \left| \sum_{i=1}^m w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \right|}_{f(\mathbf{w})} \quad (34)$$

where

$$\bar{X} = \mathbf{w} \mathbf{w}^\top + \bar{S} \quad (35)$$

for $\bar{X} \in \mathbb{R}^{(n+m) \times (n+m)}$ and

$$\bar{S} = \begin{bmatrix} 0 & S \\ S^\top & 0 \end{bmatrix} \quad (36)$$

Furthermore, define $\bar{B} = \{(i, j) : (i, j) \in \bar{\Omega}, \bar{S}_{ij} \neq 0\}$ and $\bar{G} = \{(i, j) : (i, j) \in \bar{\Omega}, \bar{S}_{ij} = 0\}$ as the sets of bad and good measurements for the symmetrized problem, respectively. In this work, we do not impose any assumption on the maximum value of the nonzero elements of S . However, without loss of generality, one may assume that $\mathbf{u}^* \mathbf{u}^{*\top} + S > 0$ and $\mathbf{w}^* \mathbf{w}^{*\top} + \bar{S} > 0$; otherwise, the non-positive elements can be discarded due to the assumptions $\mathbf{u}^* > 0$ and $(\mathbf{u}^*, \mathbf{v}^*) > 0$. In fact, we impose a slightly more stronger condition in this work.

Assumption 1 *There exists a constant $c \in (0, 1]$ such that $S_{ij} + u_i^* u_j^* > c u_{\min}^{*2}$ and $\bar{S}_{ij} + w_i^* w_j^* > c w_{\min}^{*2}$ for (31) and (34), respectively.*

7.1 Identifiability

Intuitively, the non-negative RPCA under the unknown-but-sparse noise is more challenging to solve than its noiseless counterpart. In particular, one may consider (31) as a variant of (P1-Sym) discussed in the previous section, where the locations of the bad measurements are unknown; if these locations were known, they could have been discarded to reduce the problem to (P1-Sym). If the measurements are subject to unknown noise, one of the main issues arises from the identifiability of the solution. To further elaborate, we will offer an example below.

Example 1 Suppose that $X(\epsilon) = (e_1 + \mathbf{1}\epsilon)(e_1 + \mathbf{1}\epsilon)^\top$, where e_1 is the first unit vector and $\mathbf{1}$ is a vector of ones. Assuming that $\Omega = \{1, \dots, n\}^2$, one can decompose $X(\epsilon)$ in two forms

$$X(\epsilon) = \underbrace{(e_1 + \mathbf{1}\epsilon)(e_1 + \mathbf{1}\epsilon)^\top}_{\mathbf{u}_1^* \mathbf{u}_1^{*\top}} + \underbrace{0}_{S_1} \quad (37a)$$

$$X(\epsilon) = \underbrace{\mathbf{1}\mathbf{1}^\top \epsilon^2}_{\mathbf{u}_2^* \mathbf{u}_2^{*\top}} + \underbrace{e_1 e_1^\top + \mathbf{1} e_1^\top \epsilon + e_1 \mathbf{1}^\top \epsilon}_{S_2} \quad (37b)$$

For every $\epsilon > 0$, both S_1 and S_2 can be considered as sparse matrices since the number of nonzero elements in each of these matrices is at most on the order of $O(n)$. However, unless more restrictions on the number of nonzero elements at each row or column of S are imposed, it is impossible to distinguish between these two cases. This implies that the solution is not identifiable.

In order to ensure that the solution is identifiable in the symmetric case, we assume that $\Delta(\mathcal{G}(B)) \leq \eta \cdot \delta(\mathcal{G}(G))$ for some constant $\eta \leq 1$ to be defined later. Roughly speaking, this implies that at each row of the measurement matrix, the number of good measurements should be at least as large as the number of bad ones. Similar to the work by Ge et al. (2016, 2017), we consider the regularized version of the problem, as in

$$\min_{\mathbf{u} \geq 0} \underbrace{\sum_{(i,j) \in \Omega} |u_i u_j - X_{ij}| + R_\beta(\mathbf{u})}_{f_{\text{reg}}(\mathbf{u})} \quad (\text{P2-Sym})$$

where $R_\beta(\mathbf{u})$ is a regularizer defined as

$$R_\beta(\mathbf{u}) = \lambda \sum_{i=1}^n (u_i - \beta)^4 \mathbb{I}_{u_i \geq \beta} \quad (38)$$

for some fixed parameters λ and β to be specified later. Similarly, one can define an analogous regularization for (34) as

$$\min_{\mathbf{w} \geq 0} \underbrace{\sum_{(i,j) \in \bar{\Omega}} |w_i w_j - \bar{X}_{ij}| + \alpha \left| \sum_{i=1}^m w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \right|}_{f_{\text{reg}}(\mathbf{w})} + R(\mathbf{w}) \quad (\text{P2-Asym})$$

with

$$R_\beta(\mathbf{w}) = \lambda \sum_{i=1}^{m+n} (w_i - \beta)^4 \mathbb{I}_{w_i \geq \beta} \quad (39)$$

for some fixed parameters λ and β to be specified later. Note that the defined regularization function is convex in its domain. In particular, it eliminates the candidate solutions that are far from the true solution. Without loss of generality and to streamline the presentation, it is assumed that $u_{\max}^* = w_{\max}^* = 1$ in the sequel.

Lemma 21 *Consider the parameter c defined in Assumption 1. The following statements hold:*

- *By choosing $\beta = 1$ and $\lambda = n/2$, any D -stationary point $\mathbf{u} > 0$ of (P2-Sym) satisfies the inequalities $(c/2)u_{\min}^{*2} \leq u_{\min} \leq u_{\max} \leq 2$.*
- *By choosing $\beta = 1$ and $\lambda = (m+n)/2$, any D -stationary point $\mathbf{w} > 0$ of (P2-Asym) satisfies the inequalities $(c/2)w_{\min}^{*2} \leq w_{\min} \leq w_{\max} \leq 2$.*

Proof See Appendix G. ■

7.2 Deterministic Guarantees

In what follows, the deterministic conditions under which (P2-Sym) and (P2-Asym) have benign landscape will be investigated. The results of this subsection will be the building blocks for the derivation of the main theorems for both symmetric and asymmetric positive RPCA under the random sampling and noise regime. Note that the analysis of the landscape will be more involved in this case since the effect of the regularizer should be taken into account.

Symmetric case: Recall that, for the sparsity graph $\mathcal{G}(\Omega)$, $\Delta(\mathcal{G}(\Omega))$ and $\delta(\mathcal{G}(\Omega))$ correspond to its maximum and minimum degrees, respectively.

Theorem 22 *Suppose that*

- i. $\mathbf{u}^* > 0$;*
- ii. $\delta(\mathcal{G}(G)) > (48/c^2)\kappa(\mathbf{u}^*)^4\Delta(\mathcal{G}(B))$;*
- iii. $\mathcal{G}(\Omega)$ has no bipartite component.*

Then, with the choice of $\beta = 1$ and $\lambda = n/2$ for the parameters of the regularization function $R_\beta(\mathbf{u})$, the following statements hold for (P2-Sym):

- 1. It does not have any spurious local minimum;*
- 2. The point $\mathbf{u} = \mathbf{u}^*$ is the unique global minimum;*
- 3. In the positive orthant, the point $\mathbf{u} = \mathbf{u}^*$ is the only D -stationary point.*

Additionally, if $\mathcal{G}(\Omega)$ is connected, the following statements hold for (P2-Sym):

4. The points $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{u} = 0$ are the only D -min-stationary points;
5. The point $\mathbf{u} = 0$ is a local maximum.

Proof See Appendix H. ■

Asymmetric case: Theorem 22 has the following natural extension to asymmetric problems.

Theorem 23 *Suppose that*

- i. $\mathbf{w}^* > 0$;
- ii. $\delta(\mathcal{G}(\bar{G})) > (48/c^2)\kappa(\mathbf{w}^*)^4\Delta(\mathcal{G}(\bar{B}))$;
- iii. $\mathcal{G}(\bar{G})$ is connected.

Then, with the choice of $\beta = 1$ and $\lambda = (m+n)/2$ for the parameters of the regularization function $R(\mathbf{w})$, the following statements hold for (P2-Asym):

1. The points $\mathbf{w} = 0$ and \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ are the only D -min-stationary points;
2. The point $\mathbf{w} = 0$ is a local maximum;
3. In the positive orthant, the point \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ is the only D -stationary point.

Proof The proof is omitted due to its similarity to that of Theorem 22. ■

7.3 Probabilistic Guarantees

As an extension to our previous results, we analyze the landscape of the noisy non-negative RPCA with randomness both in the location of the samples and in the structure of the noise matrix. Suppose that for the symmetric case, with probability d , each element of the upper triangular part of X is independently corrupted with an arbitrary noise value. In other words, for every (i, j) with $i \leq j$, one can write

$$X_{ij} = \begin{cases} u_i^* u_j^* & \text{with probability } 1 - d \\ \text{arbitrary} & \text{with probability } d \end{cases} \quad (40)$$

Furthermore, similar to the preceding section, suppose that every element of the upper triangular part of $X = \mathbf{u}^*\mathbf{u}^{*\top} + S$ is independently measured with probability p . The randomness in the location of the measurements and noise is naturally extended to the asymmetric case by considering the symmetrized \bar{X} and \bar{S} defined in (35) and (36), respectively.

Symmetric case: First, the main result in the symmetric case is presented below.

Theorem 24 *Suppose that*

- i. $n \geq 2$,*
- ii. $\mathbf{u}^* > 0$,*
- iii. $d < \frac{1}{(144/c^2)\kappa(\mathbf{u}^*)^4+1}$,*
- iv. $p > \frac{(1740/c^2)\kappa(\mathbf{u}^*)^4(1+\eta)\log n}{n}$,*

for some $\eta > 0$. Then, with the choice of $\beta = 1$ and $\lambda = n/2$ for the parameters of the regularization function $R_\beta(\mathbf{u})$, the following statements hold for (P2-Sym) with probability of at least $1 - 3n^{-\eta}$:

- 1. The points $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{u} = 0$ are the only D -min-stationary points;*
- 2. The point $\mathbf{u} = 0$ is a local maximum;*
- 3. In the positive orthant, the point $\mathbf{u} = \mathbf{u}^*$ is the only D -stationary point.*

To prove Theorem 24, first we present the following lemma on the concentration of the minimum and maximum degrees of random graphs.

Lemma 25 *Consider a random graph $\mathcal{G}(n, p)$. Given a constant $\eta > 0$, the inequality:*

$$\mathbb{P}\left(\Delta(\mathcal{G}(n, p)) \geq \max\left\{\frac{3np}{2}, 18(1+\eta)\log n\right\}\right) \leq n^{-\eta} \quad (41)$$

holds for every $0 < p \leq 1$. Furthermore, we have

$$\mathbb{P}\left(\delta(\mathcal{G}(n, p)) \leq \frac{np}{2}\right) \leq n^{-\eta} \quad (42)$$

provided that $p \geq \frac{12(1+\eta)\log n}{n}$.

Proof See Appendix I. ■

Remark 26 *Note that since the degree of each node in $\mathcal{G}(n, p)$ is concentrated around np with high probability, one may speculate that $\Delta(\mathcal{G}(n, p))$ and $\delta(\mathcal{G}(n, p))$ should also concentrate around np for all values of p and hence the inclusion of $18(1+\eta)\log n$ in (41) may seem redundant. Surprisingly, this is not the case in general. In fact, it can be shown that if $p = 1/n$ (and hence $np = 1$), there exists a node whose degree is lower bounded by $\log n / \log \log n$ with high probability. This explains the reasoning behind the inclusion of $18(1+\eta)\log n$ in the lemma.*

Proof of Theorem 24: In light of Lemma 16, the bounds on p and d guarantee that $\mathcal{G}(G)$ is connected and non-bipartite with probability of at least $1 - \frac{3}{2}n^{-430(1+\eta)}$. Therefore, the proof is completed by invoking Theorem 22, provided that the second condition of Theorem 22 holds. Define the events $\mathcal{E}_1 = \left\{\Delta(\mathcal{G}(B)) \leq \max\left\{\frac{3npd}{2}, 18(1+\eta)\log n\right\}\right\}$ and

$\mathcal{E}_2 = \left\{ \delta(\mathcal{G}(G)) \geq \frac{np(1-d)}{2} \right\}$. Observe that Lemma 25 together with the bounds on p and d results in the inequalities

$$\mathbb{P}(\mathcal{E}_1) \geq 1 - n^{-\eta} \tag{43a}$$

$$\mathbb{P}(\mathcal{E}_2) \geq 1 - n^{-144\eta} \tag{43b}$$

This in turn implies that the events \mathcal{E}_1 and \mathcal{E}_2 occur with probability of at least $1 - n^{-\eta} - n^{-144\eta}$. Conditioned on these events, it suffices to show that

$$\frac{np(1-d)}{2} > \frac{48}{c^2} \kappa(\mathbf{u}^*)^4 \max \left\{ \frac{3npd}{2}, 18(1+\eta) \log n \right\} \tag{44}$$

in order to certify the validity of the second condition of Theorem 22. It can be easily verified that the assumed upper and lower bounds on p and d guarantee the validity of (44). Therefore, a simple union bound and the fact that $n^{-\eta} > \frac{3}{2}n^{-430(1+\eta)}$ imply that the conditions of Theorem 22 are satisfied with probability of at least $1 - 3n^{-\eta}$. ■

A number of interesting corollaries can be derived based on Theorem 24.

Corollary 27 *Suppose that p is a positive number independent of n and $d \lesssim \log n/n$. Then, under an appropriate choice of parameters for the regularization function, the statements of Theorem 24 hold with overwhelming probability, provided that $\kappa(\mathbf{u}^*) \lesssim (n/\log n)^{1/4}$.*

Corollary 27 implies that, roughly speaking, if the total number of measurements is sufficiently large (i.e., on the order of n^2), then up to factor of $n \log n$ bad measurements with arbitrary magnitudes will not introduce any spurious local solution to the problem. Under such circumstances, the required upper bound on the ratio between the maximum and the minimum entries of \mathbf{u}^* will be more relaxed as the dimension of the problem grows.

Corollary 28 *Suppose that p is a positive number independent of n and that $d \lesssim n^{\epsilon-1}$ for some $\epsilon \in [0, 1)$. Then, under an appropriate choice of parameters for the regularization function, the statements of Theorem 24 hold with overwhelming probability, provided that $\kappa(\mathbf{u}^*) \lesssim n^{(1-\epsilon)/4}$.*

Corollary 28 describes an interesting trade-off between the sparsity level of the noise and the maximum allowable variation in the entries of \mathbf{u}^* ; roughly speaking, as $\kappa(\mathbf{u}^*)$ decreases, a larger number of noisy elements can be added to the problem without creating any spurious local minimum. The next corollary shows that a constant fraction of the measurements can be grossly corrupted without affecting the landscape of the problem, provided that $\kappa(\mathbf{u}^*)$ is uniformly bounded from above.

Corollary 29 *Suppose that p and d are positive numbers independent of n and that $d < \frac{1}{(144/c^2)+1}$. Then, under an appropriate choice of parameters for the regularization function, the statements of Theorem 24 hold with overwhelming probability, provided that $\kappa(\mathbf{u}^*) \leq \left(\frac{1-d}{(144/c^2)d} \right)^{1/4}$.*

Asymmetric case: The aforementioned results on the symmetric positive RPCA under random sampling and noise will be generalized to the asymmetric case below.

Theorem 30 Define $r = m/n$ and suppose that

- i. $n \geq m \geq 2$,
- ii. $\mathbf{w}^* > 0$,
- iii. $d < \frac{r}{(144/c^2)\kappa(\mathbf{w}^*)^{4+r}}$,
- iv. $p > \frac{(1740/c^2)\kappa(\mathbf{w}^*)^4(1+\eta)n \log n}{m^2}$,

for some $\eta > 0$. Then, with the choice of $\beta = 1$ and $\lambda = (m+n)/2$ for the parameters of the regularization function $R_\beta(\mathbf{w})$, the following statements hold for (P2-Sym) with probability of at least $1 - 10n^{-\eta}$:

1. The points $\mathbf{w} = 0$ and \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ are the only D -min-stationary points;
2. The point $\mathbf{w} = 0$ is a local maximum;
3. In the positive orthant, the point \mathbf{w} with the properties $\mathbf{w}\mathbf{w}^\top = \mathbf{w}^*\mathbf{w}^{*\top}$ and $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ is the only D -stationary point.

To prove Theorem 30, we derive a concentration bound on the minimum and maximum degree of the random bipartite graphs. Define $\mathcal{G}(m, n, p)$ as a bipartite graph with the vertex partitions $V_u = \{1, \dots, m\}$ and $V_v = \{m+1, \dots, m+n\}$ where each edge is independently included in the graph with probability p .

Lemma 31 Consider a random bipartite graph $\mathcal{G}(m, n, p)$. Given a constant $\eta > 0$, the inequality

$$\mathbb{P} \left(\Delta(\mathcal{G}(m, n, p)) \geq \max \left\{ \frac{3np}{2}, \frac{18(1+\eta)n \log n}{m} \right\} \right) \leq 2n^{-\eta} \quad (45)$$

holds for every $0 < p \leq 1$. Furthermore, we have

$$\mathbb{P} \left(\delta(\mathcal{G}(m, n, p)) \leq \frac{mp}{2} \right) \leq 2n^{-\eta} \quad (46)$$

provided that $p \geq 12(1+\eta) \log n/m$.

Proof See Appendix J. ■

Proof of Theorem 30: The bounds on p and d indeed guarantee that $\mathcal{G}(\bar{G})$ is connected with overwhelming probability. Based on this fact, the result of Lemma 31 and the proof of Theorem 24 can be combined to arrive at this theorem. The details are omitted for brevity. ■

Remark 32 *The presented probability guarantees for RPCA share some similarities with those derived for noisy matrix completion in Ge et al. (2017, 2016). In particular, according to Theorems 24 and 30 and similar to the results of Ge et al. (2017, 2016), the probability of having a spurious local solution decreases polynomially with respect to the dimension of the problem. Furthermore, similar to our work, the required lower bound on the sampling probability p in Ge et al. (2017, 2016) scales polynomially with respect to the condition number of the true solution. Finally, for non-symmetric noisy matrix completion problem, Ge et al. (2017) shows that the required lower bound on p scales as $\frac{\log n}{m}$. Comparing this dependency with the one introduced in Theorem 30, it can be inferred that our proposed lower bound is higher by a factor of $\frac{n}{m}$; this is not surprising considering the fundamentally different natures of these problems. Furthermore, Ge et al. (2017) uses this result to solve a modified (smoothed) variant of the robust PCA. Due to this reformulation, their analysis does not directly apply to (SN-RPCA) and AN-RPCA. Moreover, unlike our results, the success of their method is contingent upon the knowledge of an upper bound on the number of nonzero elements in different rows and columns of the noise matrix.*

8. Global Convergence of Local Search Algorithms

So far, it has been shown that the positive RPCA is free of spurious local minima. Furthermore, it has been proven that the global solution is the only D-stationary point in the positive orthant. The question of interest in this section is: How could this unique D-stationary point be obtained? Before answering this question, we will take a detour and revisit the notion of stationarity for smooth optimization problems. Recall that $\bar{\mathbf{x}}$ is a stationary point of a differentiable function $f(\mathbf{x})$ if and only if $\nabla f(\mathbf{x}) = 0$ and, under some mild conditions, basic local search algorithms will converge to a stationary point. Therefore, the uniqueness of the stationary point for a smooth optimization problem immediately implies the convergence to global solution. Extra caution should be taken when dealing with non-smooth optimization. In particular, the convergence of classical local search algorithms may fail to hold since the gradient and/or Hessian of the function may not exist at every iteration. To deal with this issue, different local search algorithms have been introduced to guarantee convergence to generalized notions of stationary points for non-smooth optimization, such as directional-stationary (which is used in this paper) or Clarke-stationary (to be defined next).

For a non-smooth and locally Lipschitz function $h(\mathbf{x})$ over the convex set \mathcal{X} , define the Clarke generalized directional derivative at the point $\bar{\mathbf{x}}$ in the feasible direction \mathbf{d} as

$$h^\circ(\bar{\mathbf{x}}, \mathbf{d}) := \limsup_{\substack{\mathbf{y} \rightarrow \bar{\mathbf{x}} \\ t \downarrow 0}} \frac{h(\mathbf{y} + t\mathbf{d}) - h(\mathbf{y})}{t} \quad (47)$$

Note the difference between the ordinary directional derivative $h'(\bar{\mathbf{x}}, \mathbf{d})$ and its Clarke generalized counterpart: in the latter, the limit is taken with respect to a *variable* vector \mathbf{y} that approaches $\bar{\mathbf{x}}$, rather than taking the limit exactly at $\bar{\mathbf{x}}$. The Clarke differential of $h(\mathbf{x})$ at $\bar{\mathbf{x}}$ is defined as the following set (Clarke 1990):

$$\partial_C h(\bar{\mathbf{x}}) := \{\psi \mid h^\circ(\bar{\mathbf{x}}, \mathbf{d}) \geq \langle \psi, \mathbf{d} \rangle, \forall \mathbf{d} \in \mathbb{R}^n \text{ such that } \bar{\mathbf{x}} + \mathbf{d} \in \mathcal{X}\} \quad (48)$$

where \mathcal{X} is the feasible set of the problem. A point $\bar{\mathbf{x}}$ is Clarke-stationary (or C-stationary) if $0 \in \partial_C(\bar{\mathbf{x}})$, or equivalently, $h^\circ(\bar{\mathbf{x}}, \mathbf{d}) \geq 0$ for every feasible direction \mathbf{d} . It is well known that C-stationary is a weaker condition than the D-min-stationarity. In particular, every D-min-stationary point is C-stationary but not all C-stationary points are D-min-stationary.

On the other hand, although some local search algorithms converge to D-min-stationary points for problems with special structures (Cui et al. 2018), the most well-known numerical algorithms for non-smooth optimization—such as gradient sampling, sequential quadratic programming, and exact penalty algorithms—can only guarantee the C-stationarity of the obtained solutions (Burke et al. 2005; Curtis and Overton 2012; Fasano et al. 2014). Therefore, it remains to study whether the global solution of the positive RPCA is the only C-stationary point. To answer this question, we need the following two lemmas.

Lemma 33 *The following statements hold:*

- If $h : \mathcal{X} \rightarrow \mathbb{R}$ and $g : \mathcal{X} \rightarrow \mathbb{R}$ are continuously differentiable at $\bar{\mathbf{x}} \in \mathcal{X}$, then $(h + g)^\circ(\bar{\mathbf{x}}, \mathbf{d}) = h^\circ(\bar{\mathbf{x}}, \mathbf{d}) + g^\circ(\bar{\mathbf{x}}, \mathbf{d})$ for every feasible direction \mathbf{d} .
- If $h : \mathcal{X} \rightarrow \mathbb{R}$ is continuously differentiable at $\bar{\mathbf{x}} \in \mathcal{X}$, then $h^\circ(\bar{\mathbf{x}}, \mathbf{d}) = h'(\bar{\mathbf{x}}, \mathbf{d})$ for every feasible direction \mathbf{d} .

Proof Refer to the textbook by Clarke (1990). ■

Lemma 34 *Let $h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_m(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$ be continuous and locally Lipschitz functions at $\bar{\mathbf{x}} \in \mathcal{X}$. Define*

$$h(\mathbf{x}) = \max_{1 \leq i \leq m} h_i(\mathbf{x}) \quad (49)$$

and let $I(\bar{\mathbf{x}})$ be the set of indices i such that $h(\bar{\mathbf{x}}) = h_i(\bar{\mathbf{x}})$. Then,

$$h^\circ(\bar{\mathbf{x}}, \mathbf{d}) \leq \max_{i \in I(\bar{\mathbf{x}})} h_i^\circ(\bar{\mathbf{x}}, \mathbf{d}) \quad (50)$$

for every feasible direction \mathbf{d} .

Proof Consider a feasible point $\mathbf{y} \in \mathcal{B}(\bar{\mathbf{x}}, \epsilon) \cap \mathcal{X}$, where $\mathcal{B}(\bar{\mathbf{x}}, \epsilon)$ is the Euclidean ball with the center $\bar{\mathbf{x}}$ and radius ϵ . First, we prove that $I(\mathbf{y}) \subseteq I(\bar{\mathbf{x}})$ for sufficiently small $\epsilon > 0$. Notice that $h_i(\bar{\mathbf{x}}) < h_j(\bar{\mathbf{x}})$ for every $i \in I(\bar{\mathbf{x}})$ and $j \in \{1, \dots, m\} \setminus I(\bar{\mathbf{x}})$. Therefore, due to the continuity of $h_i(\cdot)$ for every $i \in \{1, \dots, m\}$, it follows that there exists $\bar{\epsilon} > 0$ such that $h_i(\mathbf{y}) < h_j(\mathbf{y})$ for every $\mathbf{y} \in \mathcal{B}(\bar{\mathbf{x}}, \epsilon) \cap \mathcal{X}$ with $0 < \epsilon < \bar{\epsilon}$. This implies that $I(\mathbf{y} + t\mathbf{d}) \subseteq I(\mathbf{y}) \subseteq I(\bar{\mathbf{x}})$ for every $\mathbf{y} \in \mathcal{B}(\bar{\mathbf{x}}, \epsilon) \cap \mathcal{X}$ and every feasible direction \mathbf{d} with sufficiently small $\epsilon > 0$ and $t > 0$. Now, note that

$$h(\mathbf{y} + t\mathbf{d}) - h(\mathbf{y}) = \max_{i \in I(\mathbf{y} + t\mathbf{d})} h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y}) \leq \max_{i \in I(\bar{\mathbf{x}})} h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y}) \quad (51)$$

This implies that

$$h^\circ(\bar{\mathbf{x}}, \mathbf{d}) = \limsup_{\substack{\mathbf{y} \rightarrow \bar{\mathbf{x}} \\ t \downarrow 0}} \frac{h(\mathbf{y} + t\mathbf{d}) - h(\mathbf{y})}{t} \leq \max_{i \in I(\bar{\mathbf{x}})} \left\{ \limsup_{\substack{\mathbf{y} \rightarrow \bar{\mathbf{x}} \\ t \downarrow 0}} \frac{h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y})}{t} \right\} = \max_{i \in I(\bar{\mathbf{x}})} h_i^\circ(\bar{\mathbf{x}}, \mathbf{d}) \quad (52)$$

This completes the proof. ■

Based on the above lemmas, we develop the following theorem.

Theorem 35 *Under the conditions of Theorems 22 and assuming that $\mathcal{G}(\Omega)$ is connected, the global solution and the origin are the only C -stationary points of the symmetric positive RPCA. A similar result holds for the asymmetric positive RPCA.*

Proof Without loss of generality, we only consider the symmetric case. At a given point \mathbf{u} , the function $f(\mathbf{u})$ is locally Lipschitz and can be written as

$$f(\mathbf{u}) = \sum_{(i,j) \in \Omega} \max\{u_i u_j - X_{ij}, -u_i u_j + X_{ij}\} = \max_{\sigma \in \mathcal{M}} f_\sigma(\mathbf{u}) \quad (53)$$

where \mathcal{M} is the class of functions from Ω to $\{-1, +1\}$ and $f_\sigma(\mathbf{u})$ is defined as

$$f_\sigma(\mathbf{u}) = \sum_{(i,j) \in \Omega} \sigma(i,j)(u_i u_j - X_{ij}). \quad (54)$$

Hence,

$$f_{\text{reg}}(\mathbf{u}) = R_\beta(\mathbf{u}) + \max_{\sigma \in \mathcal{M}} f_\sigma(\mathbf{u}) \quad (55)$$

Notice that each function $f_\sigma(\mathbf{u})$ is differentiable and locally Lipschitz for every $\sigma \in \mathcal{M}$. By contradiction, suppose that there exists $\mathbf{u} \geq 0$ such that $\mathbf{u} \notin \{\mathbf{u}^*, 0\}$ and $0 \in \partial_C f_{\text{reg}}(\mathbf{u})$. Furthermore, define $I(\mathbf{u})$ as the set of all functions $\sigma \in \mathcal{M}$ for which $f_\sigma(\mathbf{u}) = f(\mathbf{u})$. Using the proof technique developed in Theorem 22, one can easily verify that there exists a feasible direction \mathbf{d} such that $f'_\sigma(\mathbf{u}, \mathbf{d}) + R'(\mathbf{u}, \mathbf{d}) < 0$ for every $\sigma \in I(\mathbf{u})$. By invoking Lemma 33 for every $\sigma \in I(\mathbf{u})$, it can be concluded that $f_\sigma^\circ(\mathbf{u}, \mathbf{d}) + R^\circ(\mathbf{u}, \mathbf{d}) < 0$. This, together with Lemma 34, certifies that $f_{\text{reg}}^\circ(\mathbf{u}, \mathbf{d}) < 0$, hence contradicting the assumption $0 \in \partial_C f_{\text{reg}}(\mathbf{u})$. ■

9. Discussions on Extension to Rank- r

So far, we have characterized the conditions under which the non-negative rank-1 RPCA has no spurious local solution. However, the following question has been left unanswered: *Can these results be extended to the general robust non-negative **rank- r** RPCA?*

As a first step toward answering this question and similar to our analysis in the rank-1 case, we consider the noiseless non-negative rank- r RPCA defined as

$$\min_{U \in \mathbb{R}_+^{n \times r}} f(U) = \|\mathcal{P}_\Omega(UU^\top - U^*U^{*\top})\|_1 \quad (\text{P1-Sym-}r)$$

for the symmetric version, and

$$\min_{U \in \mathbb{R}_+^{m \times r}, V \in \mathbb{R}_+^{n \times r}} f(U, V) = \|\mathcal{P}_\Omega(UV^\top - U^*V^{*\top})\|_1 \quad (\text{P1-Asym-}r)$$

for the asymmetric version. The non-negative rank- r RPCA appears in many applications, such as robust inference in gene networks and motion detection with dynamic background. For instance, according to Section 1 and equation (3), one can define $U^* = [\mathbf{u}_{(1)} \ \mathbf{u}_{(2)} \ \dots \ \mathbf{u}_{(k)}]$ and $V^* = [\mathbf{v}_{(1)} \ \mathbf{v}_{(2)} \ \dots \ \mathbf{v}_{(k)}]$ in gene regulatory networks, and formulate the problem of estimating the non-negative participation values of different genes and samples as an instance of the non-negative rank- k RPCA (P1-Asym- r) (here, k is the number of gene-blocks). Another example is the dynamic variant of the motion detection problem in which the background changes over time. In particular, one can model this problem as a non-negative rank- r RPCA, where the rank r corresponds to the number of *baseline* backgrounds that appear in the video frames. In particular, suppose that in any given video frame at time t , the background B^t can be characterized as $\sum_{i=1}^r w_i^t B_{(i)}$. In other words, the background at time t is a non-negative weighted combination of r baseline backgrounds $\{B_{(i)}\}_{i=1}^r$. Under such circumstances, the motion detection problem with a dynamic background can be written as a non-negative rank- r RPCA (P1-Asym- r), where U^* collects the vectorized version of the baseline backgrounds, and the t^{th} row of V^* corresponds to the non-negative weights $[w_1^t \ w_2^t \ \dots \ w_r^t]$. Other applications of the non-negative rank- r RPCA include topic modeling in text and image processing, as well as other applications of non-negative matrix factorization (Cichocki et al. 2009).

A fundamental roadblock in extending the results of Section 6 to (P1-Sym- r) and (P1-Asym- r) is the implicit *rotational symmetry* in the solution: given a rotation matrix R and a solution \tilde{U} (\tilde{U}, \tilde{V}) to (P1-Sym- r) ((P1-Asym- r) when $m = n$), the matrix $\tilde{U}R$ ($\tilde{U}R, \tilde{V}R$) is another feasible solution with $f(\tilde{U}R) = f(\tilde{U})$ ($f(\tilde{U}R, \tilde{V}R) = f(\tilde{U}, \tilde{V})$), provided that $\tilde{U}R$ ($\tilde{U}R, \tilde{V}R$) is a non-negative matrix. In the rank-1 case, this does not pose any problem since $R = 1$ is the only possible value. However, for the general rank- r case with $r \geq 2$, this rotational symmetry undermines the strict positivity assumption of the true components. In particular, even if the true solution U^* is strictly positive, there exists a rotation matrix R such that U^*R is non-negative with at least one zero entry. Furthermore, it is easy to see that the same rotational symmetry exists in the regularized variant of the problem and in the presence of the noise. An important consequence of this phenomenon is that Lemma 9 and, as a result, the technique used in Theorem 8 may not be readily extended to the rank- r case.

Despite the theoretical difficulties in extending the presented results to the general rank- r instances, we have indeed observed—through thousands of simulations—that in general, the sub-gradient method introduced in Section 3 successfully converges to a solution U that satisfies $UU^\top = U^*U^{*\top}$, even if the measurement matrix is corrupted with a surprisingly dense noise matrix. To illustrate this, we consider randomly generated instances of the problem with the dimension $n = 100$ and the rank $r \in \{2, 3, 4, 5\}$. For each instance, the elements of U^* are uniformly chosen from the interval $[0.5, 2.5]$. Furthermore, each element in the upper triangular part of the noise matrix S is set to 2 and 0 with probabilities d and $1 - d$, respectively. For each rank r and the noise probability d , we consider 500 independent instances of the problem and solve them using the randomly initialized sub-gradient method. Similar to Subsection 3.1, we assume that a solution is recovered exactly if $\|UU^\top - U^*U^{*\top}\|_F / \|U^*U^{*\top}\|_F \leq 10^{-4}$. Figure 4 demonstrates the ratio of the instances for which the sub-gradient method successfully recovers the true solution. As illustrated in

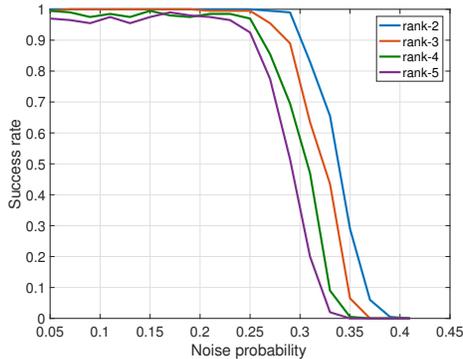


Figure 4: The success rate of the randomly initialized sub-gradient method for the positive rank- r RPCA.

this figure, d can be as large as 0.30, 0.28, 0.26, and 0.25 to guarantee a success rate of at least 90% when r is equal to 2, 3, 4, and 5, respectively.

This empirical study suggests that one of the following statements may hold for the positive rank- r RPCA: (1) it is devoid of spurious local minima, or (2) its spurious local minima can be escaped efficiently using the sub-gradient method. Further investigation of this direction is left as an enticing challenge for future research.

10. Conclusion

This paper deals with the non-negative rank-1 robust principal component analysis (RPCA), where the goal is to recover the true non-negative principal component of the data matrix exactly, using partial and potentially noisy measurements of the data matrix. The main difference between the RPCA and its classical counterpart is the sparse-but-arbitrarily-large values of the additive noise. The most commonly known methods for solving the RPCA are based on convex relaxations, where the problem is *convexified* at the expense of significantly increasing the number of variables. In this work, we show that the original non-convex and non-smooth ℓ_1 formulation of the positive rank-1 RPCA problem based on the well-known Burer-Monteiro approach has benign landscape, i.e., it does not have any spurious local solution and has a unique global solution that coincides with the true components. In particular, we provide strong deterministic and statistical guarantees for the benign landscape of the positive rank-1 RPCA and show that the absence of spurious local solutions is guaranteed to hold with a surprisingly large number of corrupted measurements. While the results on “no spurious local minima” are ubiquitous for smooth problems related to matrix completion and sensing, to the best of our knowledge, the results presented in this paper are the first to prove the absence of local minima when the objective function is non-smooth. Finally, through extensive simulations, we provide strong evidence suggesting that the proposed results may hold for the general non-negative rank- r RPCA. The extension of our theoretical results to this generalized problem is left as a future work.

Acknowledgments

The authors are grateful to Javad Lavaei, Richard Zhang, and Cedric Jozs for insightful discussions on earlier versions of this manuscript. Moreover, the authors thank Richard Zhang for his assistance in providing us with the code for the simulations. This work was supported by grants from ONR, AFOSR and NSF.

References

- Francis Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017.
- Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Global optimality of local search for low rank matrix recovery. In *Advances in Neural Information Processing Systems*, pages 3873–3881, 2016.
- Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.
- Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702, 2000.
- Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
- James V Burke, Adrian S Lewis, and Michael L Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 15(3):751–779, 2005.
- Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.
- Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.
- Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.
- Arvind Caprihan, Godfrey D Pearlson, and Vincent D Calhoun. Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements. *Neuroimage*, 42(2):675–682, 2008.
- Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

- Robin Chaney and Allen Goldstein. An extension of the method of subgradients. *Nonsmooth Optimization*, pages 51–70, 1978.
- Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.
- Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- Frank H Clarke. *Optimization and nonsmooth analysis*, volume 5. Siam, 1990.
- Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 2003.
- Ying Cui, Jong-Shi Pang, and Bodhisattva Sen. Composite difference-max programs for modern statistical estimation problems. *SIAM Journal on Optimization*, 28(4):3344–3374, 2018.
- Frank E Curtis and Michael L Overton. A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization. *SIAM Journal on Optimization*, 22(2):474–500, 2012.
- David L Donoho. For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(6):797–829, 2006.
- Susan Dumais, John Platt, David Heckerman, and Mehran Sahami. Inductive learning algorithms and representations for text categorization. In *Proceedings of the seventh international conference on Information and knowledge management*, pages 148–155. ACM, 1998.
- P Erdős and A Rényi. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.
- Giovanni Fasano, Giampaolo Liuzzi, Stefano Lucidi, and Francesco Rinaldi. A linesearch-based derivative-free approach for nonsmooth constrained optimization. *SIAM Journal on Optimization*, 24(3):959–992, 2014.
- Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points-online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.
- Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*, pages 2973–2981, 2016.
- Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. pages 1233–1242, 2017.

- AA Goldstein. Optimization of lipschitz continuous functions. *Mathematical Programming*, 13(1):14–22, 1977.
- Alexander N Gorban, Balázs Kégl, Donald C Wunsch, Andrei Y Zinovyev, et al. *Principal manifolds for data visualization and dimension reduction*, volume 58. Springer, 2008.
- Quanquan Gu, Zhaoran Wang Wang, and Han Liu. Low-rank and sparse structure pursuit via alternating minimization. In *Artificial Intelligence and Statistics*, pages 600–609, 2016.
- Nora Hartsfield and Gerhard Ringel. *Pearls in graph theory: a comprehensive introduction*. Courier Corporation, 2013.
- Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.
- Daniel Hsu, Sham M Kakade, and Tong Zhang. Robust matrix decomposition with sparse corruptions. *IEEE Transactions on Information Theory*, 57(11):7221–7234, 2011.
- John Hull and Alan White. Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4):573–592, 1990.
- Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.
- Cedric Jozs, Yi Ouyang, Richard Zhang, Javad Lavaei, and Somayeh Sojoudi. A theory on the absence of spurious solutions for nonconvex and nonsmooth optimization. *Advances in neural information processing systems*, 2018.
- Laura Lazzeroni and Art Owen. Plaid models for gene expression data. *Statistica sinica*, pages 61–86, 2002.
- Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.
- Xiao Li, Zhihui Zhu, Anthony Man-Cho So, and Rene Vidal. Nonconvex robust low-rank matrix recovery. *SIAM Journal on Optimization*, 30(1):660–686, 2020.
- Yao Ma, Alexander Olshevsky, Csaba Szepesvari, and Venkatesh Saligrama. Gradient descent for sparse rank-one matrix completion for crowd-sourced aggregation of sparsely interacting workers. In *International Conference on Machine Learning*, pages 3341–3350, 2018.
- Ramtin Madani, Mohsen Kheirandishfard, Javad Lavaei, and Alper Atamtürk. Polynomial optimization via penalized conic relaxation, 2018. URL http://www.uta.edu/faculty/madanir/poly_conic.pdf.
- Igor Molybog, Ramtin Madani, and Javad Lavaei. Conic optimization for robust quadratic regression: Deterministic bounds and statistical analysis. *IEEE 57th Conference on Decision and Control*, 2018.

- Andrea Montanari and Emile Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Transactions on Information Theory*, 62(3):1458–1484, 2016.
- Praneeth Netrapalli, UN Niranjan, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust pca. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014.
- Matt Olfat and Anil Aswani. Spectral algorithms for computing fair support vector machines. *International Conference on Artificial Intelligence and Statistics*, 2018.
- Bin Ren, Laurent Pueyo, Guangtun Ben Zhu, John Debes, and Gaspard Duchêne. Non-negative matrix factorization: Robust extraction of extended structures. *The Astrophysical Journal*, 852(2):104, 2018.
- AI Saltykov. The number of components in a random bipartite graph. *Discrete Mathematics and Applications*, 5(6):515–524, 1995.
- Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.
- Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere i: Overview and the geometric picture. *IEEE Transactions on Information Theory*, 63(2):853–884, 2017.
- Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.
- Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 255–261. IEEE, 1999.
- John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.
- Xinyang Yi, Dohyung Park, Yudong Chen, and Constantine Caramanis. Fast algorithms for robust pca via gradient descent. In *Advances in neural information processing systems*, pages 4152–4160, 2016.
- Richard Zhang, Salar Fattahi, and Somayeh Sojoudi. Large-scale sparse inverse covariance estimation via thresholding and max-det matrix completion. In *International Conference on Machine Learning*, pages 5761–5770, 2018a.
- Richard Y Zhang, Cédric Jozs, Somayeh Sojoudi, and Javad Lavaei. How much restricted isometry is needed in nonconvex matrix recovery? *Advances in neural information processing systems*, 2018b.

Xiao Zhang, Lingxiao Wang, Yaodong Yu, and Quanquan Gu. A primal-dual analysis of global optimality in nonconvex low-rank matrix recovery. In *International conference on machine learning*, pages 5857–5866, 2018c.

Xiao Zhang, Lingxiao Wang Wang, and Quanquan Gu. A unified framework for nonconvex low-rank plus sparse matrix recovery. In *International Conference on Artificial Intelligence and Statistics*, 2018d.

Qinqing Zheng and John Lafferty. Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*, 2016.

Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candes, and Yi Ma. Stable principal component pursuit. In *2010 IEEE international symposium on information theory*, pages 1518–1522. IEEE, 2010.

Zhihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B Wakin. Global optimality in low-rank matrix optimization. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1275–1279. IEEE, 2017.

Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.

Appendix A. Proof of Lemma 9:

Without loss of generality and for simplicity, we will assume that $\mathcal{G}(\Omega)$ is connected since the proof can be readily applied to each disjoint component of $\mathcal{G}(\Omega)$. Consider a point $\mathbf{u} \geq 0$ with $u_k = 0$ for some k . Consider $\Omega_k = \{j | (k, j) \in \Omega\}$ and note that it is non-empty due to the assumption that $\mathcal{G}(\Omega)$ is connected and non-bipartite. Furthermore, if there exists $r \in \Omega_k$ such that $u_r > 0$, a positive perturbation of u_k will result in a feasible and negative directional derivative. Therefore, suppose that $u_r = 0$ for every $r \in \Omega_k$. Similarly, one can show that if $u_t > 0$ for some $t \in \Omega_r$ and $r \in \Omega_k$, then \mathbf{u} has a feasible and strictly negative directional derivative. Invoking the same argument for the neighbors of the nodes with the zero value, one can infer that $\mathbf{u} = 0$. This completes the proof. \blacksquare

Appendix B. Proof of Proposition 10:

Suppose that $\mathbf{u}^* \geq 0$ and there exists an index i such that $u_i^* = 0$. Without loss of generality, assume that $i = 1$ and $u_j^* > 0$ for every $j \geq 2$. Next, we will show that \mathbf{u} defined as $u_1 = \beta > 0$ and $u_j = 0$ for $j \geq 2$ is a local minimum of (P1-Sym). Consider the perturbed version of \mathbf{u} as

$$\hat{u}_1 \leftarrow \beta + \epsilon_1 \tag{56}$$

$$\hat{u}_j \leftarrow \epsilon_j \quad \forall j \in \{2, \dots, n\} \tag{57}$$

for sufficiently small $|\epsilon_1|$ and $\epsilon_2, \dots, \epsilon_n \geq 0$. Upon defining $\Omega = \{1, \dots, n\}^2 \setminus \{(1, 1)\}$, one can write

$$f(\mathbf{u}) = \sum_{j=2}^n u_j^{*2} + \sum_{j,k=2, j \neq k}^n u_j^* u_k^* \quad (58)$$

$$f(\hat{\mathbf{u}}) = \sum_{j=2}^n u_j^{*2} - \epsilon_j^2 + \sum_{j=2}^n (\beta + \epsilon_1) \epsilon_j + \sum_{j,k=2, i \neq j}^n |u_j^* u_k^* - \epsilon_j \epsilon_k| \geq f(\mathbf{u}) + \beta \sum_{j=2}^n \epsilon_j - \left(\sum_{j=1}^n \epsilon_j \right)^2 + \epsilon_1^2 \quad (59)$$

It is easy to verify that there exist constants $\bar{\epsilon}_1 > 0$ and $\bar{\epsilon} > 0$ such that for every $-\bar{\epsilon}_1 \leq \epsilon_1 \leq \bar{\epsilon}_1$ and $0 \leq \sum_{j=2}^n \epsilon_j \leq \bar{\epsilon}$, we have

$$\beta \sum_{j=2}^n \epsilon_j - \left(\sum_{i=1}^n \epsilon_i \right)^2 + \epsilon_1^2 \geq 0 \quad (60)$$

and hence $f(\hat{\mathbf{u}}) \geq f(\mathbf{u})$. This implies that \mathbf{u} is a local minimum for $f(\mathbf{u})$. \blacksquare

Appendix C. Proof of Theorem 13:

First, we present a number of lemmas that are crucial to the proof of this theorem.

Lemma 36 *Suppose that $\mathcal{G}(\bar{\Omega})$ is connected and $\mathbf{w}^* > 0$. Then, for every D-min-stationary point \mathbf{w} , we have $\mathbf{w} > 0$ or $\mathbf{w} = 0$.*

Proof The proof is omitted due to its similarity to that of Lemma 9. \blacksquare

Lemma 37 *Suppose that $\mathcal{G}(\bar{\Omega})$ is connected and $\mathbf{w}^* > 0$. Then, $\sum_{i=1}^m w_i^2 = \sum_{j=m+1}^{m+n} w_j^2$ holds for every D-stationary point $\mathbf{w} > 0$ of (P1-Asym).*

Proof By contradiction, suppose that $\sum_{i=1}^m w_i^2 \neq \sum_{j=m+1}^{m+n} w_j^2$ for a D-stationary point $\mathbf{w} > 0$. Without loss of generality, suppose that $\sum_{i=1}^m w_i^2 > \sum_{j=m+1}^{m+n} w_j^2$ and consider the following perturbation of \mathbf{w}

$$\hat{w}_i \leftarrow \begin{cases} w_i - w_i \epsilon & \text{if } 1 \leq i \leq n \\ w_i + w_i \epsilon & \text{if } n+1 \leq i \leq n+m \end{cases} \quad (61)$$

For $(i, j) \in \bar{\Omega}$, one can write

$$|\hat{w}_i \hat{w}_j - \hat{w}_i^* \hat{w}_j^*| = |(w_i - w_i \epsilon)(w_j + w_j \epsilon) - \hat{w}_i^* \hat{w}_j^*| = |w_i w_j - \hat{w}_i^* \hat{w}_j^*| + w_i w_j \epsilon^2 \quad (62)$$

Therefore, we have

$$f(\hat{\mathbf{w}}) - f(\mathbf{w}) \leq -2\alpha \left(\sum_{i=1}^m w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \right) \epsilon + O(\epsilon^2) \quad (63)$$

This implies the existence of strictly positive and negative directional derivatives, thus resulting in a contradiction. This completes the proof. \blacksquare

Lemma 38 $\mathcal{G}(\bar{\Omega})$ has a unique vertex partitioning.

Proof By contradiction, suppose that there exist two different vertex partitions (S, T) and (\bar{S}, \bar{T}) for $\mathcal{G}(\bar{\Omega})$. Since $\mathcal{G}(\bar{\Omega})$ is a connected bipartite graph, \bar{S} is not equal to S or T , and therefore, $S \cap \bar{S}$ and $T \cap \bar{T}$ are not empty. Now, it is easy to observe that the nodes in $S \cap \bar{S}$ can only be connected to those in $T \cap \bar{T}$ and, similarly, the nodes in $T \cap \bar{T}$ can only be connected to those in $S \cap \bar{S}$. Therefore, unless $(S \cap \bar{S}) \cup (T \cap \bar{T})$ includes all the nodes, the graph will be disconnected, contradicting our assumption. On the other hand, this implies that $S \cap \bar{S} = S$ and $T \cap \bar{T} = T$, contradicting the assumption that (S, T) and (\bar{S}, \bar{T}) are different. \blacksquare

Proof of Theorem 13 For a D-min-stationary point \mathbf{w} , note that if $w_i = 0$ for some index i , then Lemma 36 implies that $\mathbf{w} = 0$, which can be easily verified to be a local maximum. We assume that \mathbf{w}^* satisfies $\sum_{i=1}^m w_i^{*2} = \sum_{j=m+1}^{m+n} w_j^{*2}$, which can be ensured by an appropriate scaling of \mathbf{u}^* and \mathbf{v}^* while keeping $\mathbf{u}^* \mathbf{v}^{*\top}$ intact. Now, it suffices to show that for a D-stationary point $\mathbf{w} > 0$, we have $\mathbf{w} = \mathbf{w}^*$. This proves the validity of the statements of the theorem.

By contradiction, suppose that $\mathbf{w} > 0$ with $\mathbf{w} \neq \mathbf{w}^*$ is a D-stationary point. In what follows, we will construct directions with strictly positive and negative directional derivatives at this point. Similar to the proof of Theorem 8, one can show that

$$0 < \frac{w_1^*}{w_1} \leq \min_{k \in \bar{\Omega}_i} \frac{w_k^*}{w_k} \leq \frac{w_i}{w_i^*} \leq \max_{k \in \bar{\Omega}_i} \frac{w_k^*}{w_k} \leq \frac{w_{m+n}^*}{w_{m+n}} \quad (64)$$

for every $1 \leq i \leq m+n$. By contradiction, suppose that $w_i \neq w_i^*$ for some index i . First, note that $w_{m+n}^*/w_{m+n} > 1$; otherwise, it holds that $w_{m+n}^*/w_{m+n} \leq 1$ and $w_i/w_i^* > 1$, which contradict with (64). Define

$$\begin{aligned} T_1^u &= \left\{ i \mid \frac{w_i^*}{w_i} = \frac{w_{m+n}^*}{w_{m+n}}, 1 \leq i \leq m \right\}, & T_2^u &= \left\{ i \mid \frac{w_i}{w_i^*} = \frac{w_{m+n}^*}{w_{m+n}}, 1 \leq i \leq m \right\} \\ T_1^v &= \left\{ i \mid \frac{w_i^*}{w_i} = \frac{w_{m+n}^*}{w_{m+n}}, m+1 \leq i \leq m+n \right\}, & T_2^v &= \left\{ i \mid \frac{w_i}{w_i^*} = \frac{w_{m+n}^*}{w_{m+n}}, m+1 \leq i \leq m+n \right\} \end{aligned} \quad (65)$$

and

$$N^u = \{1, \dots, m\} \setminus (T_1^u \cup T_2^u) \quad (66a)$$

$$N^v = \{m+1, \dots, m+n\} \setminus (T_1^v \cup T_2^v) \quad (66b)$$

Furthermore, define $\bar{\mathbf{d}}$ as

$$\bar{d}_i = \begin{cases} \frac{w_i}{w_{m+n}} - w_i \gamma & \text{if } i \in T_1^u \\ -w_i \gamma & \text{if } i \in N^u \\ -\frac{w_i}{w_{m+n}} - w_i \gamma & \text{if } i \in T_2^u \\ \frac{w_i}{w_{m+n}} + w_i \gamma & \text{if } i \in T_1^v \\ w_i \gamma & \text{if } i \in N^v \\ -\frac{w_i}{w_{m+n}} + w_i \gamma & \text{if } i \in T_2^v \end{cases} \quad (67)$$

where

$$\gamma = \frac{\sum_{i \in T_1^u} w_i - \sum_{i \in T_2^u} w_i - \sum_{i \in T_1^v} w_i + \sum_{i \in T_2^v} w_i}{w_n \sum_{i=1}^{m+n} w_i} \quad (68)$$

Similar to the symmetric case, we show that if $T_1^u \cup T_1^v$ is non-empty, then $f'(\mathbf{w}, \bar{\mathbf{d}}) < 0$ and $f'(\mathbf{w}, -\bar{\mathbf{d}}) > 0$, which contradicts the D-stationarity of \mathbf{w} . We will only show $f'(\mathbf{w}, \bar{\mathbf{d}}) < 0$ since $f'(\mathbf{w}, -\bar{\mathbf{d}}) > 0$ can be proven in a similar way. Define a perturbation of \mathbf{w} as $\hat{\mathbf{w}} = \mathbf{w} + \mathbf{d}\epsilon$ where $\epsilon > 0$ is chosen to be sufficiently small.

First, we analyze the regularization term in (P1-Asym). One can write

$$\begin{aligned} \left| \sum_{i=1}^m \hat{w}_i^2 - \sum_{j=m+1}^{m+n} \hat{w}_j^2 \right| &\leq \left| \sum_{i=1}^m w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \right| \\ &+ 2 \left(\sum_{i \in T_1^u} \frac{w_i}{w_{m+n}} - \sum_{i \in T_2^u} \frac{w_i}{w_{m+n}} - \sum_{i \in T_1^v} \frac{w_i}{w_{m+n}} + \sum_{i \in T_2^v} \frac{w_i}{w_{m+n}} \right) \epsilon \\ &- 2\gamma \left(\sum_{i=1}^m w_i + \sum_{i=m+1}^{m+n} w_i \right) \epsilon \left| + \left(\frac{1}{w_n} + \gamma \right)^2 \left(\sum_{i=1}^{m+n} w_i \right) \epsilon^2 \right. \end{aligned} \quad (69)$$

Now, according to the definition of γ , one can easily verify that

$$2 \left(\sum_{i \in T_1^u} \frac{w_i}{w_{m+n}} - \sum_{i \in T_2^u} \frac{w_i}{w_{m+n}} - \sum_{i \in T_1^v} \frac{w_i}{w_{m+n}} + \sum_{i \in T_2^v} \frac{w_i}{w_{m+n}} \right) \epsilon - 2\gamma \left(\sum_{i=1}^m w_i + \sum_{i=m+1}^{m+n} w_i \right) \epsilon = 0 \quad (70)$$

This together with Lemma 37, reduces (69) to

$$\left| \sum_{i=1}^m \hat{w}_i^2 - \sum_{j=m+1}^{m+n} \hat{w}_j^2 \right| \leq \left(\frac{1}{w_n} + \gamma \right)^2 \left(\sum_{i=1}^{m+n} w_i \right) \epsilon^2 \quad (71)$$

To analyze the first term of (P1-Asym), similar to our previous proofs, we will divide the set $\bar{\Omega}$ into different cases (4 cases to be precise) and analyze the effect of the defined perturbation in each case. For the sake of simplicity and to streamline the presentation, we only report the final inequalities for these cases:

1. If $(i, j) \in \bar{\Omega}$ and $(i, j) \in (T_1^u \times T_1^v) \cup (T_2^u \times T_2^v)$, then

$$|\hat{w}_i \hat{w}_j - w_i^* w_j^*| \leq |w_i w_j - w_i^* w_j^*| - \frac{2w_i w_j}{w_{m+n}} \epsilon + w_i w_j \left(\frac{1}{w_{m+n}^2} - \gamma^2 \right) \epsilon^2 \quad (72)$$

2. If $(i, j) \in \bar{\Omega}$ and $(i, j) \in (N^u \times (T_1^v \cup T_2^v)) \cup ((T_1^u \cup T_2^u) \times N^v)$, then

$$|\hat{w}_i \hat{w}_j - w_i^* w_j^*| \leq |w_i w_j - w_i^* w_j^*| - \frac{w_i w_j}{w_{m+n}} \epsilon + w_i w_j \left(\frac{\gamma}{w_{m+n}^2} - \gamma^2 \right) \epsilon^2 \quad (73)$$

3. If $(i, j) \in \bar{\Omega}$ and $(i, j) \in (T_1^u \times T_2^v) \cup (T_2^u \times T_1^v)$, then

$$|\hat{w}_i \hat{w}_j - w_i^* w_j^*| \leq |w_i w_j - w_i^* w_j^*| + w_i w_j \left(\frac{\gamma}{w_{m+n}} - \gamma \right)^2 \epsilon^2 \quad (74)$$

4. If $(i, j) \in \bar{\Omega}$ and $(i, j) \in N^u \times N^v$, then

$$|\hat{w}_i \hat{w}_j - w_i^* w_j^*| \leq |w_i w_j - w_i^* w_j^*| + w_i w_j \gamma^2 \epsilon^2 \quad (75)$$

Based on the above inequalities and due to the fact that $\mathcal{G}(\bar{\Omega})$ is connected, one can easily verify that $N^u \cup N^v$ should be empty; otherwise, \mathbf{w} has a strictly negative (and positive) directional derivative. Based on the same reasoning, the graph induced by $T_1^u \cup T_1^v$ or $T_2^u \cup T_2^v$ should be empty. Therefore, \mathcal{G} is bipartite with the components $T_1^u \cup T_1^v$ and $T_2^u \cup T_2^v$. Now, based on Lemma 38, $(T_1^u \cup T_1^v, T_2^u \cup T_2^v)$ induces the same vertex partitioning as (V_u, V_v) (without loss of generality, assume that $T_1^u \cup T_1^v = V_u$ and $T_2^u \cup T_2^v = V_v$). This implies that

$$\frac{w_1}{w_1^*} = \dots = \frac{w_m}{w_m^*} = \frac{w_{m+1}^*}{w_{m+1}} = \dots = \frac{w_{m+n}^*}{w_{m+n}} > 1 \quad (76)$$

Therefore,

$$\sum_{i=1}^m w_i > \sum_{i=1}^m w_i^*, \quad \sum_{i=m+1}^{m+n} w_i^* > \sum_{i=m+1}^{m+n} w_i \quad (77)$$

Together with the assumption $\sum_{i=1}^m w_i^* = \sum_{i=m+1}^{m+n} w_i^*$, this implies that

$$\sum_{i=1}^m w_i > \sum_{i=m+1}^{m+n} w_i \quad (78)$$

which, according to Lemma 37, contradicts the D-stationarity of \mathbf{w} . This completes the proof. \blacksquare

Appendix D. Proof of Lemma 16:

To prove this lemma, first we provide a lower bound on the probability of $\mathcal{G}(\Omega)$ being connected. Define C_k as the number of connected components with exactly k vertices in $\mathcal{G}(\Omega)$. Then, one can write:

$$\mathbb{P}(\mathcal{G}(\Omega) \text{ is connected}) = 1 - \mathbb{P} \left(\bigcup_{k=1}^{\lceil n/2 \rceil} \{C_k > 0\} \right) = 1 - \mathbb{P}(C_1 > 0) - \sum_{k=2}^{\lceil n/2 \rceil} \mathbb{P}(C_k > 0) \quad (79)$$

where $\lceil n/2 \rceil$ denotes the smallest integer that is greater than or equal to $n/2$. Next, we provide an upper bound on $\mathbb{P}(C_k > 0)$ for every $k \in \{2, \dots, \lceil n/2 \rceil\}$. We have

$$\mathbb{P}(C_k > 0) \leq \mathbb{E}\{C_k\} = \sum_{\mathcal{X} \subseteq [1:n], |\mathcal{X}|=k} \mathbb{E}\{I_{\mathcal{X}}\} \quad (80)$$

where $I_{\mathcal{X}}$ is an indicator random variable taking the value 1 if the subgraph $\mathcal{G}_{\mathcal{X}}(\Omega)$ of $\mathcal{G}(\Omega)$ induced by the set of vertices in \mathcal{X} is an isolated connected component of $\mathcal{G}(\Omega)$, and it takes the value 0 otherwise. On the other hand, note that $\mathcal{G}_{\mathcal{X}}(\Omega)$ is connected if and only if it contains a spanning tree. Therefore, one can write

$$\begin{aligned} \mathbb{E}\{I_{\mathcal{X}}\} &= \mathbb{P}(\mathcal{G}_{\mathcal{X}}(\Omega) \text{ has a spanning tree}) \\ &\leq \sum_{\mathcal{T} \subset \mathcal{K}_k} \mathbb{P}(\mathcal{T} \text{ belongs to } \mathcal{G}_{\mathcal{X}}(\Omega)) \\ &\leq k^{k-2} p^{k-1} \end{aligned} \quad (81)$$

where \mathcal{K}_k is a complete graph over k vertices and \mathcal{T} is a spanning tree. The last inequality is due to the fact that the number of different spanning trees in \mathcal{K}_k is equal to k^{k-2} (Hartsfield and Ringel 2013). Combining the above inequality with (80) results in

$$\begin{aligned} \mathbb{P}(C_k > 0) &\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\ &\leq \left(\frac{ne}{k}\right)^k k^{k-2} e^{-pk(n-k)} \\ &\leq \frac{1}{k^2} e^{-pk(n-k) + k \log n + k} \\ &\leq \frac{1}{k^2} e^{-\frac{k(n-1)}{2} \left(p - \frac{2 \log n + 2}{n-1}\right)} \end{aligned} \quad (82)$$

where the second inequality is due to the relations $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$ and $(1-p)^{k(n-k)} \leq e^{-pk(n-k)}$. Furthermore, the last inequality is due to $k \leq (n+1)/2$. Now, upon choosing $p \geq \frac{(2\eta+2) \log n + 2}{n-1}$ for some $\eta > 0$, one can write

$$\mathbb{P}(C_k > 0) \leq \frac{1}{k^2} e^{-\eta k \log n} = \frac{1}{k^2} (n^{-\eta})^k \quad (83)$$

Revisiting (79), one can also verify that

$$\mathbb{P}(C_1 > 0) \leq n(1-p)^{n-1} \leq e^{-p(n-1) + \log n} \leq n^{-\eta} \quad (84)$$

provided that $p \geq \frac{(\eta+1) \log n}{n-1}$, which is implied by $p \geq \frac{(2\eta+2) \log n + 2}{n-1}$. Combining this bound with (79), one can write

$$\begin{aligned} \mathbb{P}(\mathcal{G}(\Omega) \text{ is connected}) &\geq 1 - n^{-\eta} - \sum_{k=2}^{\lceil n/2 \rceil} \frac{1}{k^2} (n^{-\eta})^k \\ &\geq 1 - n^{-\eta} - \frac{1}{4} \frac{n^{-2\eta}}{1 - n^{-\eta}} \\ &\geq 1 - \left(1 + \frac{1}{4(n^\eta - 1)}\right) n^{-\eta} \\ &\geq 1 - \frac{5}{4} n^{-\eta} \end{aligned} \quad (85)$$

where we have used the assumption $n \geq 2$ and $\eta \geq 1$. Finally, given the event that $\mathcal{G}(\Omega)$ is connected, it is non-bipartite if it has at least one self-loop. Therefore, the probability of $\mathcal{G}(\Omega)$ being non-bipartite is lower bounded by $1 - (1 - p)^n$. This implies that

$$\begin{aligned}
 \mathbb{P}(\mathcal{G}(\Omega) \text{ is connected and non-bipartite}) &\geq \left(1 - \frac{5}{4}n^{-\eta}\right) (1 - (1 - p)^n) \\
 &\geq \left(1 - \frac{5}{4}n^{-\eta}\right) (1 - e^{-np}) \\
 &\geq \left(1 - \frac{5}{4}n^{-\eta}\right) \left(1 - e^{-(n-1)p}\right) \\
 &\geq \left(1 - \frac{5}{4}n^{-\eta}\right) \left(1 - e^{-2}n^{-(2\eta+2)}\right) \\
 &\geq 1 - \frac{3}{2}n^{-\eta}
 \end{aligned} \tag{86}$$

This completes the proof. ■

Appendix E. Proof of Proposition 18:

To prove Proposition 18, we present another important result on Erdős-Rényi random graphs.

Lemma 39 (Erdős and Rényi 1959) *Assuming that $np \rightarrow 0$ as $n \rightarrow \infty$, the following properties hold with probability approaching to one:*

- $\mathcal{G}(n, p)$ is acyclic.
- The size of every component of $\mathcal{G}(n, p)$ is $O(\log n)$.

Proof of Proposition 18: Assuming that $np \rightarrow 0$, Lemma 39 implies that $\mathcal{G}(\Omega)$ is the union of disjoint tree components, each with the size of at most $O(\log n)$. In what follows, we will show that, with probability approaching to one, $\mathcal{G}(\Omega)$ has at least a bipartite component without any self loops. This, together with Proposition 11, will immediately conclude the proof. One can write

$$\begin{aligned}
 \mathbb{P}(\mathcal{G}(\Omega) \text{ has a bipartite comp.}) &\stackrel{(a)}{\geq} \mathbb{P}(\mathcal{G}(\Omega) \text{ has a tree comp. without self loops}) \\
 &\geq \mathbb{P}(\text{every comp. is a tree with size } O(\log n)) \\
 &\times \mathbb{P}(\text{no self-loop in at least one comp} \mid \text{every comp. is a tree with size } O(\log n)) \\
 &\stackrel{(b)}{=} \mathbb{P}(\text{every comp. is a tree with size } O(\log n)) \\
 &\times \mathbb{P}(\text{no self-loop in at least one comp} \mid \text{every comp. has the size } O(\log n)) \\
 &\geq \mathbb{P}(\text{every comp. is a tree with size } O(\log n)) \\
 &\times (1 - \mathbb{P}(\text{every comp. has self-loops} \mid \text{every comp. has the size } O(\log n))) \\
 &\geq \underbrace{\mathbb{P}(\text{every comp. is a tree with size } O(\log n))}_A \\
 &\times (1 - \underbrace{\mathbb{P}(\text{there are at least } \Omega(n/\log n) \text{ self-loops})}_B)
 \end{aligned} \tag{87}$$

where (a) is followed by the fact that every tree is bipartite, and (b) is followed by the fact that the self-loops are included in the graph independent of other edges. Based on Lemma 39, we have $\mathbb{P}(\mathcal{A}) \rightarrow 1$ as $n \rightarrow \infty$. Now, we only need to show that $\mathbb{P}(\mathcal{B}) \rightarrow 0$ as $n \rightarrow \infty$. One can easily verify that

$$\mathbb{P}(\mathcal{B}) \leq \binom{n}{\frac{n}{\log n}} p^{\frac{n}{\log n}} \leq (e \log n)^{\frac{n}{\log n}} p^{\frac{n}{\log n}} \quad (88)$$

where the second inequality follows from the relation $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$. Replacing $p = o(1/n)$ gives rise to

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{B}) \leq \lim_{n \rightarrow \infty} (ep \log n)^{\frac{n}{\log n}} = 0 \quad (89)$$

Together with (87), this implies that $\mathcal{G}(\Omega)$ will have a bipartite component without self loops with probability approaching 1. \blacksquare

Appendix F. Proof of Lemma 20

We take an approach similar to the proof of Lemma 16. First, recall that $\{V_u, V_v\}$ with $V_u = \{1, \dots, m\}$ and $V_v = \{m+1, \dots, m+n\}$ is a vertex partitioning of the bipartite graph $\mathcal{G}(\bar{\Omega})$. Define $C_{k,l}$ as the number of connected components with exactly k vertices from V_u and l vertices from V_v . To simplify the presentation and without loss of generality, we assume that m and n are even. One can write:

$$\begin{aligned} \mathbb{P}(\mathcal{G}(\bar{\Omega}) \text{ is connected}) &= 1 - \mathbb{P}\left(\bigcup_{\substack{k=0 \\ k+l \neq 0}}^{\lceil m/2 \rceil} \bigcup_{l=1}^{\lceil n/2 \rceil} \{C_{k,l} > 0\}\right) \\ &\geq 1 - (\mathbb{P}(C_{1,0} > 0) + \mathbb{P}(C_{0,1} > 0)) - \sum_{k=1}^{\lceil m/2 \rceil} \sum_{l=1}^{\lceil n/2 \rceil} \mathbb{P}(C_{k,l} > 0) \quad (90) \end{aligned}$$

First, we provide an upper bound on $\mathbb{P}(C_{k,l} > 0)$ for $k = 1, \dots, \lceil m/2 \rceil$ and $l = 1, \dots, \lceil n/2 \rceil$. Similar to the proof of Lemma 16, one can write

$$\mathbb{P}(C_{k,l} > 0) \leq \mathbb{E}\{C_{k,l}\} = \sum_{\substack{\mathcal{X}_u \subseteq [1:m], |\mathcal{X}_u|=k \\ \mathcal{X}_v \subseteq [m+1:m+n], |\mathcal{X}_v|=l}} \mathbb{E}\{I_{\mathcal{X}_u, \mathcal{X}_v}\} \quad (91)$$

where $I_{\mathcal{X}_u, \mathcal{X}_v}$ is an indicator random variable taking the value 1 if the subgraph $\mathcal{G}_{\mathcal{X}_u, \mathcal{X}_v}(\bar{\Omega})$ of $\mathcal{G}(\bar{\Omega})$ induced by the set of vertices in $\mathcal{X}_u \cup \mathcal{X}_v$ is an isolated connected component of $\mathcal{G}(\bar{\Omega})$, and it takes the value 0 otherwise. On the other hand, we have

$$\begin{aligned} \mathbb{E}\{I_{\mathcal{X}_u, \mathcal{X}_v}\} &= \mathbb{P}(\mathcal{G}_{\mathcal{X}_u, \mathcal{X}_v}(\bar{\Omega}) \text{ has a spanning tree}) \\ &\leq \sum_{\mathcal{T} \subset \mathcal{K}_{k,l}} \mathbb{P}(\mathcal{T} \text{ belongs to } \mathcal{G}_{\mathcal{X}_u, \mathcal{X}_v}(\bar{\Omega})) \\ &\leq k^{l-1} l^{k-1} p^{k+l-1} \quad (92) \end{aligned}$$

where $\mathcal{K}_{k,l}$ is a complete bipartite graph over two sets of vertices with the sizes k and l , and \mathcal{T} is a spanning tree. The last inequality is due to the fact that the number of different spanning trees in $\mathcal{K}_{k,l}$ is equal to $k^{l-1}l^{k-1}$ (Hartsfield and Ringel 2013). Therefore, one can write

$$\begin{aligned} \mathbb{P}(C_{k,l} > 0) &\leq \binom{m}{k} \binom{n}{l} k^{l-1} l^{k-1} p^{k+l-1} (1-p)^{k(n-l)+l(m-k)} \\ &\leq \left(\frac{me}{k}\right)^k \left(\frac{ne}{l}\right)^l k^{l-1} l^{k-1} e^{-p(k(n-l)+l(m-k))} \\ &\leq \frac{1}{kl} \left(\frac{k}{l}\right)^{l-k} e^{-p(k(n-l)+l(m-k))+k \log m + l \log n + (k+l)} \\ &\leq \frac{1}{kl} e^{-p(k(n-l)+l(m-k))+(k+l)(\log(mn)+1)} \end{aligned} \quad (93)$$

where we used the relation $\left(\frac{k}{l}\right)^{l-k} \leq 1$ in the last inequality. Next, we show that the following inequality holds:

$$k(n-l) + l(m-k) \geq (k+l) \frac{(m-1)(n-1)}{m+n} \quad (94)$$

To this goal, note that

$$\begin{aligned} k(n-l) + l(m-k) &\geq (k+l) \frac{(m-1)(n-1)}{m+n} \\ \iff k(m+n)(n-l) + l(m+n)(m-k) &\geq (k+l)(m-1)(n-1) \\ \iff (k+l)mn + kn(n-2l) + lm(m-2k) &\geq (k+l)(m-1)(n-1) \\ \iff kn(n-2l) + lm(m-2k) &\geq -nk - ml - (n-1)l - (m-1)l \end{aligned} \quad (95)$$

where the last inequality holds due to $l \leq (n+1)/2$ and $k \leq (m+1)/2$, which in turn implies that $kn(n-2l) + lm(m-2k) \geq -nk - ml$. Combining (94) and (93) leads to

$$\mathbb{P}(C_{k,l} > 0) \leq \frac{1}{kl} e^{-(k+l) \frac{(m-1)(n-1)}{m+n}} \left(p - \frac{(m+n)(\log(mn)+1)}{(m-1)(n-1)}\right) \quad (96)$$

Upon choosing $p \geq \frac{(m+n)((1+\eta)\log(mn)+1)}{(m-1)(n-1)}$ for some $\eta \geq 1$, one can write

$$\mathbb{P}(C_{k,l} > 0) \leq \frac{1}{kl} ((mn)^{-\eta})^{(k+l)} \quad (97)$$

On the other hand, it is easy to verify that

$$\begin{aligned} \mathbb{P}(C_{0,1} > 0) &\leq n(1-p)^m \leq e^{-pm+\log n} \leq (mn)^{-\eta} \\ \mathbb{P}(C_{1,0} > 0) &\leq m(1-p)^n \leq e^{-pn+\log m} \leq (mn)^{-\eta} \end{aligned} \quad (98)$$

provided that $p \geq \frac{(1+\eta)\log(mn)}{m}$ and $p \geq \frac{(1+\eta)\log(mn)}{n}$, both of which are guaranteed to hold with the choice of $p \geq \frac{(m+n)((1+\eta)\log(mn)+1)}{(m-1)(n-1)}$. Combining (98), (97), and (90) results in

$$\begin{aligned} \mathbb{P}(\mathcal{G}(\bar{\Omega}) \text{ is connected}) &\geq 1 - 2(mn)^{-\eta} - \sum_{k=1}^{\lfloor m/2 \rfloor} \sum_{l=1}^{\lfloor n/2 \rfloor} \frac{1}{kl} ((mn)^{-\eta})^{(k+l)} \\ &\geq 1 - 2(mn)^{-\eta} - 4(mn)^{-2\eta} \end{aligned} \quad (99)$$

where we have used the assumptions $m, n \geq 2$ and $\eta \geq 1$. This completes the proof. \blacksquare

Appendix G. Proof of Lemma 21

We present the proof for the symmetric case (the proof for the asymmetric case follows directly after symmetrization and the fact that the penalty on the norm difference is zero at the positive D-stationary points). First, we prove that $u_{\max} \leq 2$. It suffices to show that $u_{\max} \leq \max\{2\beta, \sqrt{2n/\lambda}\}$. This, together with the choice of β and λ , implies $u_{\max} \leq 2$. To this goal, we only need to verify that $u_{\max} > 2\beta$ implies $u_{\max} \leq \sqrt{2n/\lambda}$. By contradiction, suppose that $u_{\max} > \sqrt{2n/\lambda}$. In what follows, it will be shown that \mathbf{u} has strictly positive and negative directional derivatives, thereby contradicting its D-stationarity. Consider a perturbation of \mathbf{u} as $\hat{\mathbf{u}} = \mathbf{u} - \mathbf{e}_{\max}\epsilon$ for a sufficiently small $\epsilon > 0$, where \mathbf{e}_{\max} is a vector with 1 at the location corresponding to u_{\max} and 0 everywhere else. One can write

$$\begin{aligned}
 f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) &\leq \left(\sum_{i=1}^n u_i \right) \epsilon + \lambda \left((u_{\max} - \epsilon - \beta)^4 - (u_{\max} - \beta)^4 \right) \\
 &= \left(\sum_{i=1}^n u_i \right) \epsilon - 4\lambda(u_{\max} - \beta)^3\epsilon + O(\epsilon^2) \\
 &\stackrel{(a)}{\leq} \left(\sum_{i=1}^n u_i - \frac{\lambda}{2}u_{\max}^3 \right) \epsilon + O(\epsilon^2) \\
 &\leq \left(nu_{\max} - \frac{\lambda}{2}u_{\max}^3 \right) \epsilon + O(\epsilon^2)
 \end{aligned} \tag{100}$$

where (a) is due to the fact that $u_{\max} \geq 2\beta$ implies $u_{\max} - \beta \geq u_{\max}/2$. (100) together with $u_{\max} > \sqrt{2n/\lambda}$, implies that $-\mathbf{e}_{\max}$ is a direction with a negative directional derivative. Similarly, it can be shown that \mathbf{e}_{\max} is a direction with a positive directional derivative. This contradicts the D-stationarity of \mathbf{u} and, hence, $u_{\max} \leq \max\{2\beta, \sqrt{2n/\lambda}\}$.

Next, we aim to show that $(c/2)u_{\min}^{*2} \leq u_{\min}$. By contradiction, suppose that there exists an index i such that $(c/2)u_{\min}^{*2} > u_i$. Now, since $u_i < 1$, we have $\mathbb{I}_{u_i \geq \beta} = 0$ due to the choice of β . Consider the terms in $f_{\text{reg}}(\mathbf{u})$ that involves u_i :

$$\sum_{j \in \Omega_i} |u_i u_j - X_{ij}| = \sum_{j \in G_i} |u_i u_j - u_i^* u_j^*| + \sum_{j \in B_i} |u_i u_j - (u_i^* u_j^* + S_{ij})| \tag{101}$$

Considering the fact that $u_{\max} \leq 2$, one can verify the following inequality for every $(i, j) \in G$:

$$u_i u_j < c u_{\min}^{*2} \leq u_{\min}^{*2} \leq u_i^* u_j^* \tag{102}$$

A similar inequality holds for $(i, j) \in B$:

$$u_i u_j < c u_{\min}^{*2} \stackrel{(a)}{\leq} u_i^* u_j^* + S_{ij} \tag{103}$$

where we have used Assumption 1 for (a). Therefore, a positive and negative perturbation of u_i results in negative and positive directional derivatives at \mathbf{u} , thereby contradicting the D-stationarity of this point. \blacksquare

Appendix H. Proof of Theorem 22:

The next lemma is crucial in proving Theorem 22.

Lemma 40 *Suppose that the assumptions of Theorem 22 hold and define*

$$\begin{aligned}
 s(\mathbf{u}) = & - \underbrace{\sum_{\substack{(i,j) \in \mathcal{B} \\ i,j \in T_1}} \frac{2u_i u_j}{u_n} + \sum_{\substack{(i,j) \in \mathcal{B} \\ i,j \in T_2}} \frac{2u_i u_j}{u_n} + \sum_{\substack{(i,j) \in \mathcal{B} \\ i \in T_1 \cup T_2, j \in \mathcal{N}}} \frac{u_i u_j}{u_n}}_{f_B(\mathbf{u})} \\
 & + \underbrace{\sum_{\substack{(i,j) \in \mathcal{G} \\ i,j \in T_1}} \frac{2u_i u_j}{u_n} + \sum_{\substack{(i,j) \in \mathcal{G} \\ i,j \in T_2}} \frac{2u_i u_j}{u_n} + \sum_{\substack{(i,j) \in \mathcal{G} \\ i \in T_1 \cup T_2, j \in \mathcal{N}}} \frac{u_i u_j}{u_n}}_{f_G(\mathbf{u})} + \underbrace{\sum_{i \in T_2} \frac{4u_i (u_i - 1)^3}{u_n} \mathbb{I}_{u_i \geq 1}}_{f_R(\mathbf{u})} \quad (104)
 \end{aligned}$$

where the sets T_1 and T_2 are defined as (20) and (21), respectively. Then, for every D -stationary point $\mathbf{u} > 0$ such that $\mathbf{u} \neq \mathbf{u}^*$, the following inequalities hold with the choice of $\beta = 1$ and $\lambda = n/2$:

- $f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) \leq -s(\mathbf{u})\epsilon + O(\epsilon^2)$ for $\hat{\mathbf{u}} = \mathbf{u} + \mathbf{d}\epsilon$ and a sufficiently small $\epsilon > 0$.
- $f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) \geq s(\mathbf{u})\epsilon - O(\epsilon^2)$ for $\hat{\mathbf{u}} = \mathbf{u} - \mathbf{d}\epsilon$ and a sufficiently small $\epsilon > 0$.

where \mathbf{d} is defined as (22).

Proof To prove this lemma, first we show the validity of (19). By contradiction, suppose that (19) does not hold. Without loss of generality, assume that there exists an index i such that $u_i/u_i^* > u_n^*/u_n$ (the case with $u_i/u_i^* < u_n^*/u_n$ can be argued in a similar way). This implies that $u_i u_j > u_i^* u_j^*$ for every $(i, j) \in \Omega$. Define $\hat{\mathbf{u}} = \mathbf{u} - \mathbf{e}\epsilon$ for a sufficiently small $\epsilon > 0$, where \mathbf{e} is a vector with $e_k = 1$ if $k = i$ and $e_k = 0$ otherwise. One can write

$$\begin{aligned}
 f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) & \leq - \left(\sum_{j \in G_i} u_j \right) \epsilon + \left(\sum_{j \in B_i} u_j \right) \epsilon + \lambda ((u_i - \epsilon - 1)^4 - (u_i - 1)^4) \mathbb{I}_{u_i \geq 1} \\
 & \leq - \left(\sum_{j \in G_i} u_j \right) \epsilon + \left(\sum_{j \in B_i} u_j \right) \epsilon \\
 & \leq - \frac{cu_{\min}^{*2}}{2} \delta(\mathcal{G}(G)) + 2\Delta(\mathcal{G}(B)) \quad (105)
 \end{aligned}$$

where $G_i = \{j | (i, j) \in G\}$ and $B_i = \{j | (i, j) \in B\}$. The second inequality is due to the fact that $((u_i - \epsilon - 1)^4 - (u_i - 1)^4) \mathbb{I}_{u_i \geq 1}$ is non-negative and the third inequality follows from Lemma 21 and the definitions of $\delta(\mathcal{G}(G))$, $\Delta(\mathcal{G}(B))$. Based on the assumption of Theorem 22, we have

$$\frac{\delta(\mathcal{G}(G))}{\Delta(\mathcal{G}(B))} > \frac{48}{c^2} \kappa(\mathbf{u}^*)^4 = \frac{48}{c^2 u_{\min}^{*4}} > \frac{4}{cu_{\min}^{*2}} \quad (106)$$

which implies $(-cu_{\min}^*/2)\delta(\mathcal{G}(G)) + 2\Delta(\mathcal{G}(B)) < 0$, and hence, $-\mathbf{e}$ is a direction with a negative directional derivative. Similarly, it can be shown that \mathbf{e} is a direction with a positive directional derivative. This contradicts the D-stationarity of \mathbf{u} and hence (19) holds. Now, we will show the correctness of the first statement. Similar to the proof of Theorem 8, one can verify that

$$\sum_{(i,j) \in \Omega} |\hat{u}_i \hat{u}_j - X_{ij}| - \sum_{(i,j) \in \Omega} |u_i u_j - X_{ij}| \leq (f_B(\mathbf{u}) - f_G(\mathbf{u}))\epsilon + O(\epsilon^2) \quad (107)$$

Now, we only need to bound $R(\hat{\mathbf{u}}) - R(\mathbf{u})$. To this goal, notice that if $i \in T_1$, then $u_i < u_i^* \leq 1$ due to the fact that $\mathbf{u} \neq \mathbf{u}^*$ and $u_i^*/u_i = u_n^*/u_n > 1$. Therefore, $\mathbb{I}_{u_i \geq 1} = 0$ for every $i \in T_1$. This implies that

$$\begin{aligned} R(\hat{\mathbf{u}}) - R(\mathbf{u}) &= \sum_{i \in T_2} \left(u_i - \frac{u_i}{u_n} \epsilon - 1 \right)^4 \mathbb{I}_{u_i \geq 1} - \sum_{i \in T_2} (u_i - 1)^4 \mathbb{I}_{u_i \geq 1} \\ &= - \sum_{i \in T_2} \frac{4u_i(u_i - 1)^3}{u_n} \mathbb{I}_{u_i \geq 1} \epsilon + O(\epsilon^2) \end{aligned} \quad (108)$$

A similar approach can be taken to prove the second statement of the lemma. \blacksquare

Lemma 41 *Suppose that $\mathcal{G}(\Omega)$ has no bipartite component and every entry of X is strictly positive. Then, for every D-min-stationary point \mathbf{u} of (P1-Sym), we have $\mathbf{u}[c] > 0$ or $\mathbf{u}[c] = 0$, where $\mathbf{u}[c]$ is a sub-vector of \mathbf{u} induced by the c^{th} component of $\mathcal{G}(\Omega)$.*

Proof The proof is similar to that of Lemma 9. \blacksquare

Proof of Theorem 22: Similar to the proof of Theorem 8, it suffices to show that none of the points $\mathbf{u} > 0$ with $\mathbf{u} \neq \mathbf{u}^*$ can be D-stationary. By contradiction, suppose that this is not the case, i.e., there exists a D-stationary point $\mathbf{u} > 0$ such that $\mathbf{u} \neq \mathbf{u}^*$. Consider the functions $f_B(\mathbf{u})$ and $f_G(\mathbf{u})$ defined in Lemma 40. The main idea behind the proof is to show that the term $f_G(\mathbf{u})$ always dominates $f_B(\mathbf{u})$. This, together with the non-negativity of $f_R(\mathbf{u})$, shows that $s(\mathbf{u}) > 0$ and hence, $f'_{\text{reg}}(\mathbf{u}, \mathbf{d}) < 0$ and $f'_{\text{reg}}(\mathbf{u}, -\mathbf{d}) > 0$, which is a contradiction. One can bound each term in $f_B(\mathbf{u})$ and obtain

$$\begin{aligned} f_B(\mathbf{u}) &\leq \frac{1}{u_n} \left(2 \cdot \frac{\Delta(\mathcal{G}(B))}{2} |T_1| u_{\max}^2 + 2 \cdot \frac{\Delta(\mathcal{G}(B))}{2} |T_2| u_{\max}^2 + \frac{\Delta(\mathcal{G}(B))}{2} (|T_1| + |T_2|) u_{\max}^2 \right) \epsilon + O(\epsilon^2) \\ &\leq \frac{3}{2u_n} \Delta(\mathcal{G}(B)) (|T_1| + |T_2|) u_{\max}^2 \epsilon + O(\epsilon^2) \\ &\leq \frac{6}{u_n} \Delta(\mathcal{G}(B)) (|T_1| + |T_2|) \epsilon + O(\epsilon^2) \end{aligned} \quad (109)$$

where the last inequality follows from the fact that $u_{\max} \leq 2$ due to Lemma 21. Next, we derive a lower bound on $f_G(\mathbf{x})$:

$$\begin{aligned} f_G(\mathbf{x}) &\geq \frac{1}{u_n} \cdot \frac{\delta(\mathcal{G}(G))}{2} (|T_1| + |T_2|) u_{\min}^2 \epsilon + O(\epsilon^2) \\ &\geq \frac{1}{u_n} \cdot \frac{\delta(\mathcal{G}(G))}{2} (|T_1| + |T_2|) \frac{c^2 u_{\min}^{*4}}{4} \epsilon + O(\epsilon^2) \\ &= \frac{c^2 u_{\min}^{*4}}{8u_n} \delta(\mathcal{G}(G)) (|T_1| + |T_2|) \epsilon + O(\epsilon^2) \end{aligned} \quad (110)$$

where the first inequality is due to the fact that the minimum value for $f_G(\mathbf{u})$ happens when the neighbors of $T_1 \cup T_2$ in $\mathcal{G}(G)$ all belong to the set N and their corresponding values in $\mathbf{u}\mathbf{u}^\top$ are all equal to u_{\min}^2 . Furthermore, the second inequality is due to Lemma 16 and the choice of β for $R_\beta(\mathbf{u})$. Therefore, one can write

$$\begin{aligned} f_B(\mathbf{x}) - f_G(\mathbf{x}) &\leq \left(\frac{6}{u_n} \Delta(\mathcal{G}(B)) - \frac{c^2 u_{\min}^{*4}}{8u_n} \delta(\mathcal{G}(G)) \right) (|T_1| + |T_2|) \epsilon + O(\epsilon^2) \\ &= \frac{\Delta(\mathcal{G}(B)) c^2 u_{\min}^{*4}}{8u_n} \left(\frac{48}{c^2} \kappa(\mathbf{u}^*)^4 - \frac{\delta(\mathcal{G}(G))}{\Delta(\mathcal{G}(B))} \right) (|T_1| + |T_2|) \epsilon + O(\epsilon^2). \end{aligned} \quad (111)$$

Therefore, the choice of $(48/c^2) \kappa(\mathbf{u}^*)^4 < \delta(\mathcal{G}(G))/\Delta(\mathcal{G}(B))$ implies that $f_B(\mathbf{x}) - f_G(\mathbf{x}) < 0$, thereby completing the proof. \blacksquare

Appendix I. Proof of Lemma 25

The degree of each node is equal to the summation of n independent Bernoulli random variables, each with parameter p . Therefore, Chernoff bound yields that

$$\mathbb{P}(\deg(v) \geq (1 + \delta)np) \leq e^{-np\delta^2/3} \quad (112a)$$

$$\mathbb{P}(\deg(v) \leq (1 - \delta)np) \leq e^{-np\delta^2/3} \quad (112b)$$

for every vertex v and $0 \leq \delta \leq 1$, where $\deg(v)$ is the degree of vertex v in the graph. Therefore, a simple union bound leads to

$$\mathbb{P}(\Delta(\mathcal{G}(n, p)) \geq (1 + \delta)np) \leq ne^{-np\delta^2/3} = e^{-np\delta^2/3 + \log n} \quad (113a)$$

$$\mathbb{P}(\delta(\mathcal{G}(n, p)) \leq (1 - \delta)np) \leq ne^{-np\delta^2/3} = e^{-np\delta^2/3 + \log n} \quad (113b)$$

Setting $\delta = 1/2$ and assuming that $p \geq 12(1 + \eta) \log n/n$ for some $\eta > 0$, one can write

$$\mathbb{P} \left(\Delta(\mathcal{G}(n, p)) \geq \frac{3np}{2} \right) \leq n^{-\eta} \quad (114a)$$

$$\mathbb{P} \left(\delta(\mathcal{G}(n, p)) \leq \frac{np}{2} \right) \leq n^{-\eta} \quad (114b)$$

Furthermore, $p < 12(1 + \eta) \log n/n$ leads to

$$\begin{aligned} \mathbb{P}(\Delta(\mathcal{G}(n, p)) \geq 18(1 + \eta) \log n) &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{12(1 + \eta) \log n}{n}\right)\right) \geq 18(1 + \eta) \log n\right) \\ &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{12(1 + \eta) \log n}{n}\right)\right) \geq \frac{3np}{2}\right) \\ &\leq n^{-\eta} \end{aligned} \quad (115)$$

Combining (115) with (114a) and (114b) results in the desired inequalities. This completes the proof. \blacksquare

Appendix J. Proof of Lemma 31

Define $S = \{1, \dots, m\}$ and $T = \{m + 1, \dots, m + n\}$. Similar to the proof of Lemma I, one can write the following concentration inequalities:

$$\mathbb{P}(\max_{v \in S} \{\deg(v)\} \geq (1 + \delta)np) \leq me^{-np\delta^2/3} \quad (116a)$$

$$\mathbb{P}(\min_{v \in S} \{\deg(v)\} \leq (1 - \delta)np) \leq me^{-np\delta^2/3} \quad (116b)$$

$$\mathbb{P}(\max_{v \in T} \{\deg(v)\} \geq (1 + \delta)mp) \leq ne^{-mp\delta^2/3} \quad (116c)$$

$$\mathbb{P}(\min_{v \in T} \{\deg(v)\} \leq (1 - \delta)mp) \leq ne^{-mp\delta^2/3} \quad (116d)$$

which imply

$$\mathbb{P}(\Delta(\mathcal{G}(m, n, p)) \geq (1 + \delta)np) \leq me^{-np\delta^2/3} + ne^{-mp\delta^2/3} \leq 2e^{-mp\delta^2/3 + \log n} \quad (117a)$$

$$\mathbb{P}(\delta(\mathcal{G}(m, n, p)) \leq (1 - \delta)mp) \leq me^{-np\delta^2/3} + ne^{-mp\delta^2/3} \leq 2e^{-mp\delta^2/3 + \log n} \quad (117b)$$

Setting $\delta = 1/2$ and assuming that $p \geq 12(1 + \eta) \log n/m$ for some $\eta > 0$ results in

$$\mathbb{P}(\Delta(\mathcal{G}(m, n, p)) \geq \frac{3np}{2}) \leq 2n^{-\eta} \quad (118a)$$

$$\mathbb{P}(\delta(\mathcal{G}(m, n, p)) \leq \frac{mp}{2}) \leq 2n^{-\eta} \quad (118b)$$

Furthermore, if $p < 12(1 + \eta) \log n/m$, one can write

$$\begin{aligned} \mathbb{P}\left(\Delta(\mathcal{G}(n, p)) \geq \frac{18(1 + \eta)n \log n}{m}\right) &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{12(1 + \eta) \log n}{m}\right)\right) \geq \frac{18(1 + \eta)n \log n}{m}\right) \\ &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{12(1 + \eta) \log n}{m}\right)\right) \geq \frac{3np}{2}\right) \\ &\leq 2n^{-\eta} \end{aligned} \quad (119)$$

This completes the proof. \blacksquare