Exact Guarantees on the Absence of Spurious Local Minima for Non-negative 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:

Abstract

This work is concerned with the non-negative robust principal component analysis (PCA), where the goal is to recover the dominant non-negative principal component 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 robust PCA 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 robust PCA 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 robust PCA 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 statistical guarantees for the exact recovery of the true principal component. 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 spiked model[1]

$$X = \beta \mathbf{u}\mathbf{v}^T + S \tag{1}$$

[1] There are more general models under which the PCA is shown to be useful (see Jolliffe (2011) for more details). We use the 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

$$X = \beta \mathbf{v}\mathbf{v}^T + S \tag{2}$$

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, 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 robust PCA 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 robust PCA 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)):

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

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-block, 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 robust PCA with multiple principal components.

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

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

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):

$$\min_{\mathbf{u}\in\mathbb{R}^m_+, \mathbf{v}\in\mathbb{R}^n_+} \|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^T)\|_1 \tag{5}$$

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

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

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 robust PCA** (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 $\nabla g(\mathbf{u},\mathbf{v}) = 0$, it suffices to find a *global* direction of descent $\mathbf{d}$ such that $\text{vec}(\mathbf{d})^T\nabla^2 g(\mathbf{u},\mathbf{v})\text{vec}(\mathbf{d}) < 0$, 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.

1.2 Contributions

In this work, we characterize the landscape of both the symmetric non-negative robust PCA defined as

$$\min_{\mathbf{u}\in\mathbb{R}^n_+} f(\mathbf{u}) = \|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{u}^T)\|_1 \tag{SNR-PCA}$$

and its asymmetric counterpart defined as

$$\min_{\mathbf{u}\in\mathbb{R}^m_+,\mathbf{v}\in\mathbb{R}^n_+} f(\mathbf{u},\mathbf{v}) = \|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^T)\|_1 \tag{ANR-PCA}$$

In particular, we fully characterize the stationary points of these optimization problems, under both deterministic and probabilistic scenarios for the measurement index Ω and the noise matrix S.

Definition 2 *Given the set Ω, two graphs are defined below:*

- *The sparsity graph $\mathcal{G}(\Omega)$ induced by Ω for an instance of* (SNR-PCA) *is defined as a graph with the vertex set $V := \{1, 2, ..., n\}$ 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* (ANR-PCA) *is defined as a graph with the vertex partitions $V_u := \{1, 2, ..., m\}$ and $V_v := \{m + 1, 2, ..., m + n\}$ 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 3 *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 robust PCA has benign landscape. Let $\mathbf{u}^*$ and $(\mathbf{u}^*, \mathbf{v}^*)$ denote the true principal components of (SNR-PCA) and (ANR-PCA), respectively.

Theorem 4 (Informal, Deterministic Guarantee) *Assuming that* $\mathbf{u}^*, \mathbf{v}^* > 0$ *and under an appropriate regularization, the following statements hold:*

1. *The* (SNR-PCA) *problem 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*
$$\kappa(\mathbf{u}^*)^4 \Delta(\mathcal{G}(B)) \lesssim \delta(\mathcal{G}(G)) \tag{7}$$

2. *The* (ANR-PCA) *problem 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*
$$\max\left\{\kappa(\mathbf{u}^*)^4, \kappa(\mathbf{v}^*)^4\right\} \Delta(\mathcal{G}_{m,n}(B)) \lesssim \delta(\mathcal{G}_{m,n}(G)) \tag{8}$$

Theorem 4 puts forward a set of deterministic conditions for the absence of spurious local solutions in (SNR-PCA) and (ANR-PCA) 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 4. Furthermore, we will show later in the paper that some of the conditions delineated in Theorem 4—such as the strict positivity of $\mathbf{u}^*$ and $\mathbf{v}^*$, as well as the absence of bipartite components in $\mathcal{G}(G)$ for (SNR-PCA)—are also necessary for the exact recovery.

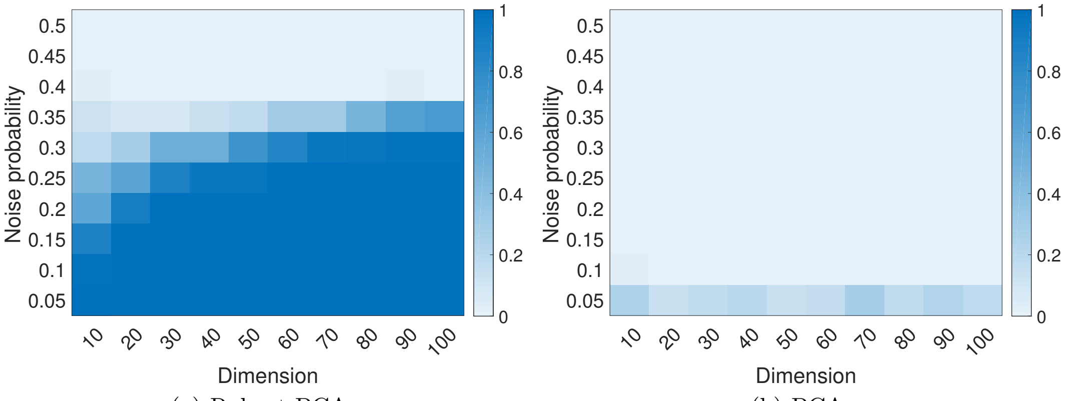
The second main result of this paper investigates (SNR-PCA) and (ANR-PCA) under random sampling and noise structures. In particular, suppose that each element (in the symmetric case, each element of the upper triangular part) of S is nonzero with probability d. Then, for every (i,j), we have

$$X_{ij} = \begin{cases} u_i^* v_j^* & \text{with probability } 1-d \\ \text{arbitrary} & \text{with probability } d \end{cases} \tag{9}$$

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 5 (Informal, Probabilistic Guarantee) *Assuming that* $\mathbf{u}^*, \mathbf{v}^* > 0$ *and under an appropriate regularization, the following statements hold with probability approaching to one:*

1. *The* (SNR-PCA) *problem has no spurious local minimum and has a unique global minimum that coincides with the true component, provided that*
$$p \gtrsim \frac{\kappa(\mathbf{u}^*)^4 \log n}{n}, \qquad d \lesssim \frac{1}{\kappa(\mathbf{u}^*)^4} \tag{10}$$



(a) Robust PCA

(b) PCA

Figure 1: The performance of the randomly initialized sub-gradient method when applied to (SNR-PCA) and the symmetric version of (6). The intensity of the color is proportional to the exact recovery rate of the true solution (darker blue implies higher recovery rate).

2. *The* (ANR-PCA) *problem has no spurious local minimum and has a unique global minimum that coincides with the true components, provided that*

$$p \gtrsim \frac{\max\left\{\kappa(\mathbf{u}^*)^4, \kappa(\mathbf{v}^*)^4\right\}(m+n)\log(m+n)}{m^2}, \qquad d \lesssim \frac{r}{\max\left\{\kappa(\mathbf{u}^*)^4, \kappa(\mathbf{v}^*)^4\right\}} \tag{11}$$

where $r = m/n$ and $n \geq m$.

A number of interesting corollaries can be obtained based on Theorem 5. 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 6 (Informal, Global Convergence) *Under the assumptions of Theorem 4 and 5, local search algorithms converge to the true solutions of* (SNR-PCA) *and* (ANR-PCA) *with probability approaching to one.*

Starting from Section 2, we will delve into the detailed analysis of the symmetric and asymmetric non-negative robust PCA. In particular, we will analyze (SNR-PCA) and (ANR-PCA) under different deterministic and probabilistic settings and provide formal versions of Theorems 4 and 5.

1.3 Numerical results

Exact recovery: To demonstrate the strength of the above-mentioned results, we consider thousands of randomly generated instances of the positive robust PCA 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 1 shows the performance of randomly initialized sub-gradient method (see Li et al. (2018) for more details on the algorithm) for the symmetric positive robust PCA compared to its conventional counterpart. More specifically, we compare the accuracy of the solutions obtained by solving (SNR-PCA) and the symmetric version of 6 using a local search algorithm. For each dimension and noise probability, we consider 100 randomly generated instances of the problem and demonstrate its exact recovery rate[1]. The left heatmap shows the exact recovery rate of the sub-gradient method, when directly applied to (SNR-PCA). It can be observed that the algorithm has recovered the globally optimal solution even when 30% of the entries in the data matrix were severely corrupted with the noise. In contrast, 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).

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 robust PCA with the parameters

$$\mathbf{u}^* = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T, \qquad S = 0, \qquad \Omega = \{1, 2, 3\}^2 \backslash \{(3, 3)\} \tag{12}$$

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 3.

Case study on 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

$$X \approx \mathbf{u}\mathbf{v}^T + S \tag{13}$$

[1] We assume that a solution is recovered exactly if $\|\mathbf{u} - \mathbf{u}^*\|_2 / \|\mathbf{u}^*\|_2 \leq 0.05$.

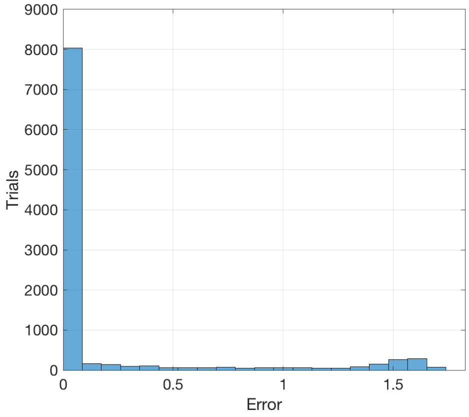


Figure 2: The normalized distance between the obtained solution using randomly initialized sub-gradient method and the true solution.

Note that the background may not always have a rank-1 representation. However, we will show that (13) 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 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 $\mathbb{R}_+^{100}$, 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.

1.4 Related work

A considerable amount of work has been carried out to understand the inherent difficulty of solving low-rank optimization problem 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 polyno-

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

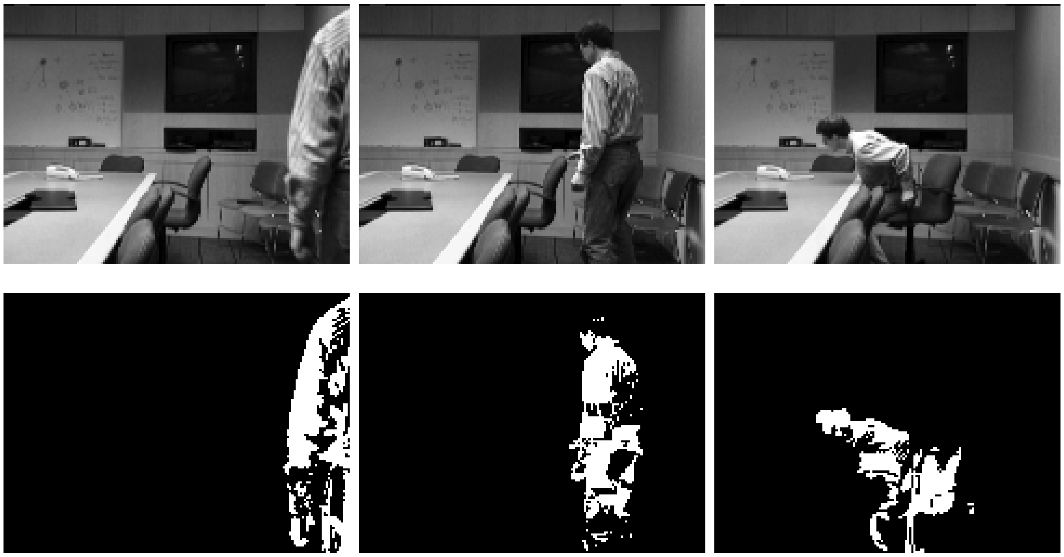


Figure 3: The performance of the sub-gradient method in the moving object detection problem. The first row shows 3 out of 100 gray-scale frames in the studied test case that contain the moving objects. The second row shows the outcome of (SNR-PCA) solved using randomly initialized sub-gradient method.

mial 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. (2018) for a comprehensive survey on the related problems). 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. The statistical properties of non-negative PCA have been recently studied by Montanari and Richard (2016), where it is shown that by considering the non-negativity of the principal component, one can guarantee its consistent recovery with a significantly smaller signal-to-noise ratio. Note that the non-negativity assumption on the true solution can be traced back to some earlier works on the non-negative matrix factorization problem (Lee and Seung (1999); Hoyer (2004)).

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 seize 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)). More recently, a sub-gradient method is proposed in Li et al. (2018) for solving the robust PCA, 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.

2. 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

$$h'(\mathbf{x},\mathbf{d}) := \lim_{t\downarrow 0}\frac{h(\mathbf{x}+t\mathbf{d})-h(\mathbf{x})}{t} \tag{14}$$

upon existence. Based on this definition, $\bar{\mathbf{u}}$ is **directional-minimum-stationary** (or D-min-stationary) for (SNR-PCA) 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 (SNR-PCA) 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 (SNR-PCA) if it is either D-min- or D-max-stationary[3].

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) 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.

3. Base Case: Noiseless Non-negative Robust PCA

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

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

For the asymmetric problem (ANR-PCA), the solution is invariant to scaling. In other words, if $(\mathbf{u}, \mathbf{v})$ is a solution to (ANR-PCA), 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 (ANR-PCA):

$$\min_{\mathbf{u}\geq 0, \mathbf{v}\geq 0} f_{\text{asym}}(\mathbf{u}, \mathbf{v}) = \|\mathcal{P}_\Omega(X - \mathbf{u}\mathbf{v}^T)\|_1 + \alpha|\mathbf{u}^T\mathbf{u} - \mathbf{v}^T\mathbf{v}| \tag{15}$$

[3] 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 (SNR-PCA).

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 $\mathbf{w} = [\mathbf{u}^T \;\; \mathbf{v}^T]^T$, $\mathbf{w}^* = [\mathbf{u}^{*T} \;\; \mathbf{v}^{*T}]^T$, and $\bar{\Omega} = \{(i,j) | (i, j-m) \in \Omega\}$. Based on these definitions, one can symmetrize (15) as follows:

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

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

3.1 Deterministic Guarantees

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

Theorem 7 *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 7, we first need the following important lemma.

Lemma 8 *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 7.

Proof of Theorem 7: We prove the first three statements. Note that Statement 5 can be easily verified and Statement 4 is implied by Lemma 8 and Statement 3.

Suppose that $\mathbf{u} \neq \mathbf{u}^*$ is a local minimum. Note that if $u_i = 0$ for some i, Lemma 8 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

$$\min_{k\in\Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_i}{u_i^*} \leq \max_{k\in\Omega_i} \frac{u_k^*}{u_k} \tag{16}$$

for every $i \in \{1, \cdots, n\}$, 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 $u_iu_j > u_i^*u_j^*$ 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:

$$0 < \frac{u_1^*}{u_1} \leq \frac{u_2^*}{u_2} \leq \cdots \leq \frac{u_n^*}{u_n} \tag{17}$$

Therefore, due to (16), we have

$$0 < \frac{u_1^*}{u_1} \leq \min_{k\in\Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_i}{u_i^*} \leq \max_{k\in\Omega_i} \frac{u_k^*}{u_k} \leq \frac{u_n^*}{u_n} \tag{18}$$

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

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 (17), implies that $u_t^*/u_t < 1$ and $u_t/u_t^* > 1$, which contradicts (18). Now, define the sets

$$T_1 = \left\{i|\frac{u_i^*}{u_i} = \frac{u_n^*}{u_n}, 1 \leq i \leq n\right\} \tag{19}$$

$$T_2 = \left\{j|\frac{u_j}{u_j^*} = \frac{u_n^*}{u_n}, 1 \leq j \leq n\right\} \tag{20}$$

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

$$d_i = \begin{cases} \frac{u_i}{u_n} & \text{if } i \in T_1 \\ -\frac{u_i}{u_n} & \text{if } i \in T_2 \\ 0 & \text{if } i \in N \end{cases} \tag{21}$$

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

$$\begin{aligned}|\hat{u}_i\hat{u}_j-u_i^*u_j^*| = u_i^*u_j^*-\hat{u}_i\hat{u}_j &= u_i^*u_j^*-\left(u_i+\frac{u_i}{u_n}\epsilon\right)\left(u_j+\frac{u_j}{u_n}\epsilon\right)\\ &= |u_iu_j-u_i^*u_j^*|-\left(\frac{2u_iu_j}{u_n}\right)\epsilon-\left(\frac{u_iu_j}{u_n^2}\right)\epsilon^2 \qquad (22)\end{aligned}$$

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

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

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

$$\frac{u_i}{u_i^*}<\frac{u_n^*}{u_n},\qquad \frac{u_i^*}{u_i}<\frac{u_n^*}{u_n} \qquad (24)$$

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

$$\frac{u_i}{u_i^*}<\frac{u_j^*}{u_j}\implies u_iu_j<u_i^*u_j^* \qquad (25)$$

which implies that

$$|\hat{u}_i\hat{u}_j-u_i^*u_j^*| = u_i^*u_j^*-\hat{u}_i\hat{u}_j = u_i^*u_j^*-u_i\left(u_j+\frac{u_j}{u_n}\epsilon\right) = |u_iu_j-u_i^*u_j^*|-\left(\frac{u_iu_j}{u_n}\right)\epsilon \qquad (26)$$

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

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

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

$$|\hat{u}_i\hat{u}_j-u_i^*u_j^*| = \left|\left(u_i+\frac{u_i}{u_n}\epsilon\right)\left(u_j-\frac{u_j}{u_n}\epsilon\right)-u_i^*u_j^*\right| \leq |u_iu_j-u_i^*u_j^*|+\left(\frac{u_iu_j}{u_n^2}\right)\epsilon^2 \qquad (28)$$

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. ■

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

Proposition 9 *Assume that $\mathbf{u}^* \geq 0$ and $\mathbf{u}^* \neq 0$ with at least one zero element. Then, there exists a set Ω such that*

1. *Ω includes all possible measurements except for one;*
2. (P1-Sym) *has a spurious local minimum.*

Proof See Appendix B. ■

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 10 *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

$$\hat{u}_i \leftarrow \begin{cases} u_i + \frac{u_i}{u_n}\epsilon & \text{if } i \in V_1 \\ u_i - \frac{u_i}{u_n+\epsilon}\epsilon & \text{if } i \in V_2 \end{cases} \tag{29}$$

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

Asymmetric case: Next, we consider (15) 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 10, 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 11 *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 $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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 $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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 12 *Notice that, unlike the symmetric case, Theorem 11 requires the connectedness of $\mathcal{G}(\bar{\Omega})$. This is due to the additional regularization term in* (ANR-PCA)*. 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.*

3.2 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 $\mathbf{u}^*\mathbf{u}^{*T}$ is measured independently with probability p. In other words, for every $(i,j) \in \{1,2,...,n\}^2$ and $i \leq j$, the probability of (i,j) belonging to Ω is equal to p.

Theorem 13 *Suppose that $\mathbf{u}^* > 0$ and $p \geq 2\log n/n$. Then, the following statements hold for* (SNR-PCA) *with probability approaching to one:*

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 13, it is useful to review a classical result on *Erdös-Rényi* random graphs. Recall that a random graph (without self-loops) is *Erdös-Rényi* if every edge is included in the graph independently with probability p. Therefore, excluding the random self-loops, the sparsity graph $\mathcal{G}(\Omega)$ is *Erdös-Rényi.* The following classical result is borrowed from Erdös and Rényi's seminal paper (Erdös and Rényi (1959)).

Theorem 14 (Erdös and Rényi (1959)) *Consider a random graph $\mathcal{G}(n,p)$ with n nodes where each edge is independently included in the graph with probability p. Upon choosing*

$$p = \frac{\log n + c}{n}, \tag{30}$$

for some $c > 0$, $\mathcal{G}(n,p)$ becomes connected with probability of at least $\Omega(e^{-e^{-c}})$.

Our next lemma is based on Theorem 14 and will be crucial in proving Theorem 13.

Lemma 15 *Suppose that $p \geq 2\log n/n$. Then, $\mathcal{G}(\Omega)$ is connected and non-bipartite with probability approaching to one.*

Proof Replacing c with $\log n$ in (30) yields that $\mathcal{G}(\Omega)$ is connected with probability approaching to one (to be precise, with probability of at least $e^{-1/n}$). Furthermore, conditioned on the connectedness of $\mathcal{G}(\Omega)$, the probability of the event that it contains at least one self-loop is $1-(1-p)^n \geq 1-(1-\log n/n)^n = 1 - O(1/n)$, which implies that $\mathcal{G}(\Omega)$ is non-bipartite with probability approaching to one. ■

Proof of Theorem 13: The proof immediately follows from Theorem 7 and Lemma 15. ■

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 16 *Assuming that* $\mathbf{u}^* \geq 0$ *and* $p < 1$, (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 9 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 9. ■

Proposition 17 *Given any* $\mathbf{u}^* > 0$, *suppose that* $np \to 0$ *as* $n \to \infty$. *Then, the global solution of* (P1-Sym) *is not unique with probability approaching to one.*

Proof See Appendix D. ■

Remark 18 *The statements based on "probability approaching to one" can be made more rigorous. This requires a more detailed analysis of the properties of Erdös-Rényi random graphs. This is fairly straightforward for Theorem 13; in this case, the statement "probability approaching to one" can be replaced by "probability of at least* $\Omega\left(e^{-1/n}(1-1/n)\right)$*". However, similar arguments for Propositions 16 and 17 require a more involved analysis of the non-asymptotic characteristics of random graphs and is outside of the scope of this paper.*

Asymmetric case: Consider (15) under a random sampling regime, where each element of $\mathbf{u}^*\mathbf{v}^{*T}$ is independently observed with probability p. Next, the analog of Theorem 13 for the asymmetric case is provided.

Theorem 19 *Suppose that* $\mathbf{w}^* > 0$ *and* $p \geq \frac{3(n+m)\log(n+m)}{nm}$. *Then, the following statements hold for* (P1-Asym) *with probability approaching to one as* $n + m \to \infty$*:*

1. *The points* $\mathbf{w} = 0$ *and* $\mathbf{w}$ *with the properties* $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ *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* $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ *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 Erdös-Rényi random graph due to its bipartite structure. Instead, we present the analog of Theorem 14 for random bipartite graphs. In particular, the following Theorem is a slightly modified version of the main result by Saltykov (1995) which guarantees the connectedness of a random bipartite graph, provided that the sampling probability exceeds a threshold. Equipped with this result and Theorem 13, one can easily verify the correctness of Theorem 19.

Theorem 20 (Saltykov (1995)) *Consider a random bipartite graph $\mathcal{G}(m,n,p)$ with the vertex partitions $V_u = \{1,\cdots,m\}$ and $V_v = \{m+1,\cdots,m+n\}$ where each edge is independently included with probability p. Without loss of generality, assume that $n \geq m$. Upon choosing*

$$p \geq 3\left(1+\frac{m}{n}\right)^{-1}\frac{(n+m)\log(n+m)}{nm} \tag{31}$$

$\mathcal{G}(m,n,p)$ becomes connected with probability approaching to one as $n+m \to \infty$.

Proof of Theorem 19: Based on Theorem 20, $p \geq \frac{3(n+m)\log(n+m)}{nm}$ guarantees the connectedness of $\mathcal{G}(\bar{\Omega})$ with probability approaching to one. Invoking Theorem 13 will complete the proof. ■

4. Extension to Noisy Positive Robust PCA

In this section, we will show that an additive sparse noise with arbitrary values does not drastically change the landscape of the robust PCA. In other words, a limited number of grossly wrong measurements will not introduce any spurious local solution to the positive robust PCA. 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:

$$\min_{\mathbf{u}\geq 0} \quad f(\mathbf{u}) = \sum_{(i,j)\in\Omega}|u_iu_j - X_{ij}| \tag{32}$$

where

$$X = \mathbf{u}^*\mathbf{u}^{*T} + S \tag{33}$$

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

$$\min_{\mathbf{u}\geq 0,\mathbf{v}\geq 0} \quad \sum_{(i,j)\in\Omega}|u_iv_j - X_{ij}| + \alpha\left|\sum_{i=1}^{m}u_i^2 - \sum_{j=1}^{n}v_j^2\right| \tag{34}$$

where α is an arbitrary positive number. After symmetrization, (34) can be re-written as

$$\min_{\mathbf{w}\geq 0} \quad f(\mathbf{w}) = \sum_{(i,j)\in\bar{\Omega}}|w_iw_j - \bar{X}_{ij}| + \alpha\left|\sum_{i=1}^{m}w_i^2 - \sum_{j=m+1}^{m+n}w_j^2\right| \tag{35}$$

where

$$\bar{X} = \mathbf{w}\mathbf{w}^T + \bar{S} \tag{36}$$

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

$$\bar{S} = \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix} \tag{37}$$

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 $\mathbf{u}^*\mathbf{u}^{*T}+S>0$ and $\mathbf{w}^*\mathbf{w}^{*T}+\bar{S}>0$; 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 $S_{ij}+u_i^*u_j^*>cu_{\min}^{*^2}$ and $\bar{S}_{ij}+w_i^*w_j^*>cw_{\min}^{*^2}$ for (32) and (35), respectively.*

4.1 Identifiability

Intuitively, the non-negative robust PCA under the unknown-but-sparse noise is more challenging to solve than its noiseless counterpart. In particular, one may consider (32) 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 $X(\epsilon)=(e_1+\mathbf{1}\epsilon)(e_1+\mathbf{1}\epsilon)^T$, where e_1 is the first unit vector and $\mathbf{1}$ is a vector of ones. Assuming that $\Omega=\{1,...,n\}^2$, one can decompose $X(\epsilon)$ in two forms*

$$X(\epsilon)=\underbrace{(e_1+\mathbf{1}\epsilon)(e_1+\mathbf{1}\epsilon)^T}_{\mathbf{u}_1^*\mathbf{u}_1^{*T}}+\underbrace{0}_{S_1} \tag{38a}$$

$$X(\epsilon)=\underbrace{\mathbf{1}\mathbf{1}^T\epsilon^2}_{\mathbf{u}_2^*\mathbf{u}_2^{*T}}+\underbrace{e_1e_1^T+\mathbf{1}e_1^T\epsilon+e_1\mathbf{1}^T\epsilon}_{S_2} \tag{38b}$$

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

$$f_{\text{reg}}(\mathbf{u})=\min_{\mathbf{u}\geq 0}\quad\sum_{(i,j)\in\Omega}|u_iu_j-X_{ij}|+R(\mathbf{u}) \tag{P2-Sym}$$

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

$$R(\mathbf{u})=\lambda\sum_{i=1}^{n}(u_i-\beta)^4\,\mathbb{I}_{u_i\geq\beta} \tag{39}$$

for some fixed parameters λ and β to be specified later. Similarly, one can define an analogous regularization for (35) as

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

with

$$R(\mathbf{u}) = \lambda\sum_{i=1}^{m+n}(w_i-\beta)^4\,\mathbb{I}_{w_i\geq\beta} \tag{40}$$

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 E. ■

4.2 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 robust PCA 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 $R(\mathbf{u})$, 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 F. ■

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.* $\mathcal{G}(\bar{\Omega})$ *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 $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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 $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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. ■

4.3 Probabilistic Guarantees

As an extension to our previous results, we analyze the landscape of the noisy robust PCA 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 $i \le j$, one can write

$$X_{ij} = \begin{cases} u_i^* u_j^* & \text{with probability } 1-d \\ \text{arbitrary} & \text{with probability } d \end{cases} \tag{41}$$

Furthermore, similar to the preceding section, suppose that every element of the upper triangular part of $X = \mathbf{u}^*\mathbf{u}^{*T} + S$ 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 (36) and (37), respectively.

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

Theorem 24 *Suppose that*

i. $\mathbf{u}^* > 0$*;*

ii. $d < \frac{1}{(144/c^2)k(\mathbf{u}^*)^4+1}$*;*

iii. $p > \frac{(3480/c^2)\kappa(\mathbf{u}^*)^4 \log n}{n}$.

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

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* $\mathcal{G}(n,p)$*. The following inequality holds for every* $p \in (0,1]$*:*

$$\mathbb{P}\left(\Delta(\mathcal{G}(n,p)) \geq \max\left\{\frac{3np}{2}, 36\log n\right\}\right) \leq \frac{1}{n} \tag{42}$$

Furthermore, the following inequality holds if $p \geq 24\log n/n$*:*

$$\mathbb{P}\left(\delta(\mathcal{G}(n,p)) \leq \frac{np}{2}\right) \leq \frac{1}{n} \tag{43}$$

Proof See Appendix G. ■

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* $36\log n$ *in* (42) *may seem redundant. Surprisingly, this is not the case in general. In fact, it can be shown that if* $p = 1/n$*, 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* $36\log n$ *in the lemma.*

Proof of Theorem 24: Notice that the bounds on p and d guarantee that $\mathcal{G}(G)$ is connected and non-bipartite with probability approaching to one. Therefore, the proof is completed by invoking Theorem 22, provided that the second condition of Theorem 22 holds. To verify this condition, observe that Lemma 25 implies the validity of the inequalities

$$\Delta(\mathcal{G}(B)) \leq \max\left\{\frac{3npd}{2}, 36\log n\right\} \tag{44a}$$

$$\delta(\mathcal{G}(G)) \geq \frac{np(1-d)}{2} \tag{44b}$$

with high probability. Therefore, it suffices to show that

$$\frac{np(1-d)}{2} > \frac{48}{c^2}\kappa(\mathbf{u}^*)^4 \max\left\{\frac{3npd}{2}, 36\log n\right\} \tag{45}$$

It can be easily verified that the assumed upper and lower bounds on p and d guarantee the validity of (45). ■

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 probability approaching to one, 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 probability approaching to one, 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 probability approaching to one, 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 robust PCA under random sampling and noise will be generalized to the asymmetric case below.

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

i. $n \geq m$,

ii. $\mathbf{w}^* > 0$,

iii. $d < \frac{r}{(144/c^2)\kappa(\mathbf{u}^*)^4 + r}$

iv. $p > \frac{(3480/c^2)\kappa(\mathbf{w}^*)^4 (n+m)\log(m+n)}{m^2}$.

Then, with the choice of $\beta = 1$ and $\lambda = (m+n)/2$ for the parameters of the regularization function $R(\mathbf{u})$, the following statements hold for (P2-Sym) *with probability approaching to one:*

1. *The points $\mathbf{w} = 0$ and $\mathbf{w}$ with the properties $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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 $\mathbf{w}\mathbf{w}^T = \mathbf{w}^*\mathbf{w}^{*T}$ 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. Recall that $\mathcal{G}(m,n,p)$ is defined as a bipartite graph with the vertex partitions $V_u = \{1, \cdots, m\}$ and $V_v = \{m+1, \cdots, m+n\}$ where each edge is independently included in the graph with probability p.

Lemma 31 *Consider a random bipartite graph $\mathcal{G}(m,n,p)$ and suppose that $n \geq m$. The following inequality holds for every $p \in (0,1]$.*

$$\mathbb{P}\left(\Delta(\mathcal{G}(m,n,p)) \geq \max\left\{\frac{3np}{2}, \frac{36n\log n}{m}\right\}\right) \leq \frac{2}{n} \tag{46}$$

Furthermore, the following inequality holds if $p \geq 24n\log n/m$:

$$\mathbb{P}\left(\delta(\mathcal{G}(m,n,p)) \leq \frac{mp}{2}\right) \leq \frac{2}{n} \tag{47}$$

Proof See Appendix H. ■

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

5. Global convergence of local search algorithms

So far, it has been shown that the positive robust PCA 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

$$h^{\circ}(\mathbf{x},\mathbf{d}) := \limsup_{\substack{\mathbf{y}\to\mathbf{x}\\ t\downarrow 0}} \frac{h(\mathbf{y}+t\mathbf{d})-h(\mathbf{y})}{t} \tag{48}$$

Note the difference between the ordinary directional derivative $h'(\mathbf{x},\mathbf{d})$ 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)):

$$\partial_C h(\bar{\mathbf{x}}) := \{\psi | h^{\circ}(\mathbf{x},\mathbf{d}) \geq \langle \psi, \mathbf{d}\rangle, \forall \mathbf{d}\in\mathbb{R}^n \text{ such that } \mathbf{x}+\mathbf{d}\in\mathcal{X}\} \tag{49}$$

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 robust PCA is the only C-stationary point. To answer this question, we need the following two lemmas.

Lemma 32 *The following statements hold:*

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

Proof Refer to the textbook Clarke (1990). ■

Lemma 33 *Let $h_1(\mathbf{x}), h_1(\mathbf{x}), ..., h_m(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$ be continuous and locally Lipschitz functions at $\bar{\mathbf{x}} \in \mathcal{X}$. Define*

$$h(\mathbf{x}) = \max_{1\leq i\leq m} h_i(\mathbf{x}) \tag{50}$$

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

$$h^\circ(\bar{\mathbf{x}}, \mathbf{d}) \leq \max_{i\in I(\bar{\mathbf{x}})} h_i^\circ(\bar{\mathbf{x}}, \mathbf{d}) \tag{51}$$

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 $j \in \{1, ..., m\}\backslash I(\bar{\mathbf{x}})$. Therefore, due to the continuity of $h_i(\cdot)$ for every $i \in \{1, ..., m\}$, 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

$$h(\mathbf{y} + t\mathbf{d}) - h(\mathbf{y}) = \max_{i\in I(\mathbf{y}+t\mathbf{d})} h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y}) \leq \max_{i\in I(\bar{\mathbf{x}})} h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y}) \tag{52}$$

This implies that

$$h^\circ(\bar{\mathbf{x}}, \mathbf{d}) = \limsup_{\substack{\mathbf{y}\to\mathbf{x} \\ t\downarrow 0}} \frac{h(\mathbf{y} + t\mathbf{d}) - h(\mathbf{y})}{t} \leq \max_{i\in I(\bar{\mathbf{x}})} \left\{ \limsup_{\substack{\mathbf{y}\to\mathbf{x} \\ t\downarrow 0}} \frac{h_i(\mathbf{y} + t\mathbf{d}) - h_i(\mathbf{y})}{t} \right\} = \max_{i\in I(\bar{\mathbf{x}})} h_i^\circ(\bar{\mathbf{x}}, \mathbf{d}) \tag{53}$$

This completes the proof. ■

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

Theorem 34 *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 robust PCA. A similar result holds for the asymmetric positive robust PCA.*

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

$$f(\mathbf{u}) = \sum_{(i,j)\in\Omega} \max\{u_iu_j - X_{ij}, -u_iu_j + X_{ij}\} = \max_{\sigma\in\mathcal{M}} f_\sigma(\mathbf{u}) \tag{54}$$

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

$$f_\sigma(\mathbf{u}) = \sum_{(i,j)\in\Omega} \sigma(i,j)(u_iu_j - X_{ij}). \tag{55}$$

Hence,

$$f_{\text{reg}}(\mathbf{u}) = R(\mathbf{u}) + \max_{\sigma\in\mathcal{M}} f_\sigma(\mathbf{u}) \tag{56}$$

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 32 for every $\sigma\in I(\mathbf{u})$, it can be concluded that $f^\circ_\sigma(\mathbf{u},\mathbf{d}) + R^\circ(\mathbf{u},\mathbf{d}) < 0$. This, together with Lemma 33, certifies that $f^\circ_{\text{reg}}(\mathbf{u},\mathbf{d}) < 0$, hence contradicting the assumption $0\in\partial_C f_{\text{reg}}(\mathbf{u})$. ■

6. Conclusion

This paper deals with the non-negative robust principal component analysis (PCA), 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 robust PCA and its classical counterpart is the sparse-but-arbitrarily large values of the additive noise. The most commonly known methods for solving the robust PCA are based on convex relaxations, where the problem is *convexified* at the expense 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 robust PCA 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 robust PCA 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.

Acknowledgments

The authors are grateful to Javad Lavaei, Richard Zhang, and Cedric Josz 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.

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. *arXiv preprint arXiv:1809.09573*, 2018.

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. *arXiv preprint arXiv:1803.00205*, 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 Erdös 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. *arXiv preprint arXiv:1704.00708*, 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.

Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

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 Josz, 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. *arXiv preprint arXiv:1809.09237*, 2018.

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.

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.

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 Josz, Somayeh Sojoudi, and Javad Lavaei. How much restricted isometry is needed in nonconvex matrix recovery? *Advances in neural information processing systems*, 2018b.

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

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 8:

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 Ω_k 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. ■

Appendix B. Proof of Proposition 9:

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

$$\hat{u}_1 \leftarrow \beta + \epsilon_1 \tag{57}$$

$$\hat{u}_i \leftarrow \epsilon_i \qquad \forall i \in \{2, ..., n\} \tag{58}$$

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

$$f(\mathbf{u}) = \sum_{i=2}^{n} u_i^{*2} + \sum_{i,j=2,i\neq j}^{n} u_i^* u_j^* \tag{59}$$

$$f(\hat{\mathbf{u}}) = \sum_{i=2}^{n} u_i^{*2} - \epsilon_i^2 + \sum_{j=2}^{n}(\beta + \epsilon_1)\epsilon_i + \sum_{i,j=2,i\neq j}^{n} |u_i^* u_j^* - \epsilon_i \epsilon_j| \geq f(\mathbf{u}) + \beta \sum_{j=2}^{n} \epsilon_i - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 + \epsilon_1^2 \tag{60}$$

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 $0 \leq \sum_{j=2}^{n} \epsilon_i \leq \bar{\epsilon}$, we have

$$\beta \sum_{j=2}^{n} \epsilon_i - \left(\sum_{i=1}^{n} \epsilon_i\right)^2 + \epsilon_1^2 \geq 0 \tag{61}$$

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

Appendix C. Proof of Theorem 11:

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

Lemma 35 *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 8. ■

Lemma 36 *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 $\mathbf{w}$

$$\hat{w}_i \leftarrow \begin{cases} w_i - w_i\epsilon & \text{if } 1 \leq i \leq n \\ w_i + w_i\epsilon & \text{if } n+1 \leq i \leq n+m \end{cases} \tag{62}$$

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

$$|\hat{w}_i \hat{w}_j - \hat{w}_i^* \hat{w}_j^*| = |(w_i - w_i\epsilon)(w_j + w_j\epsilon) - \hat{w}_i^* \hat{w}_j^*| = |w_i w_j - \hat{w}_i^* \hat{w}_j^*| + w_i w_j \epsilon^2 \tag{63}$$

Therefore, we have

$$f(\hat{\mathbf{w}}) - f(\mathbf{w}) \leq -2\alpha \left(\sum_{i=1}^{m} w_i^2 - \sum_{j=m+1}^{m+n} w_j^2\right)\epsilon + O(\epsilon^2) \tag{64}$$

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

Lemma 37 *$\mathcal{G}(\bar{\Omega})$ 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. ■

Proof of Theorem 11 For a D-min-stationary point $\mathbf{w}$, note that if $w_i = 0$ for some index i, then Lemma 35 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 $\mathbf{u}^*{\mathbf{v}^*}^T$ 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 7, one can show that

$$0 < \frac{w_1^*}{w_1} \leq \min_{k\in\bar{\Omega}_i} \frac{w_k^*}{w_k} \leq \frac{w_i}{w_i^*} \leq \max_{k\in\bar{\Omega}_i} \frac{w_k^*}{w_k} \leq \frac{w_{m+n}^*}{w_{m+n}} \tag{65}$$

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 (65). Define

$$\begin{aligned} T_1^u &= \left\{ i | \frac{w_i^*}{w_i} = \frac{w_{m+n}^*}{w_{m+n}}, 1\leq i\leq m \right\}, \qquad & T_2^u &= \left\{ i | \frac{w_i}{w_i^*} = \frac{w_{m+n}^*}{w_{m+n}}, 1\leq i\leq m \right\} \\ T_1^v &= \left\{ i | \frac{w_i^*}{w_i} = \frac{w_{m+n}^*}{w_{m+n}}, m+1\leq i\leq m+n \right\}, & T_2^v &= \left\{ i | \frac{w_i}{w_i^*} = \frac{w_{m+n}^*}{w_{m+n}}, m+1\leq i\leq m+n \right\} \end{aligned} \tag{66}$$

and

$$N^u = \{1, \ldots, m\}\backslash(T_1^u \cup T_2^u) \tag{67a}$$

$$N^v = \{m+1, \ldots, m+n\}\backslash(T_1^u \cup T_2^u) \tag{67b}$$

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

$$\bar{d}_i = \begin{cases} \frac{w_i}{w_{m+n}} - w_i\gamma & \text{if } i\in T_1^u \\ -w_i\gamma & \text{if } i\in N^u \\ -\frac{w_i}{w_{m+n}} - w_i\gamma & \text{if } i\in T_2^u \\ \frac{w_i}{w_{m+n}} + w_i\gamma & \text{if } i\in T_1^v \\ w_i\gamma & \text{if } i\in N^v \\ -\frac{w_i}{w_{m+n}} + w_i\gamma & \text{if } i\in T_2^v \end{cases} \tag{68}$$

where

$$\gamma = \frac{\sum_{i\in T_1^u} w_i - \sum_{i\in T_2^u} w_i - \sum_{i\in T_1^v} w_i + \sum_{i\in T_2^v} w_i}{w_n \sum_{i=1}^{m+n} w_i} \tag{69}$$

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

$$\begin{aligned}\left|\sum_{i=1}^{m}\hat{w}_i^2 - \sum_{j=m+1}^{m+n}\hat{w}_j^2\right| \leq \Bigg| &\sum_{i=1}^{m} w_i^2 - \sum_{j=m+1}^{m+n} w_j^2 \\ &+ 2\left(\sum_{i\in T_1^u}\frac{w_i}{w_{m+n}} - \sum_{i\in T_2^u}\frac{w_i}{w_{m+n}} - \sum_{i\in T_1^v}\frac{w_i}{w_{m+n}} + \sum_{i\in T_2^v}\frac{w_i}{w_{m+n}}\right)\epsilon \\ &- 2\gamma\left(\sum_{i=1}^{m} w_i + \sum_{i=m+1}^{m+n} w_i\right)\epsilon\Bigg| + (\frac{1}{w_n}+\gamma)^2\left(\sum_{i=1}^{m+n} w_i\right)\epsilon^2\end{aligned} \tag{70}$$

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

$$2\left(\sum_{i\in T_1^u}\frac{w_i}{w_{m+n}} - \sum_{i\in T_2^u}\frac{w_i}{w_{m+n}} - \sum_{i\in T_1^v}\frac{w_i}{w_{m+n}} + \sum_{i\in T_2^v}\frac{w_i}{w_{m+n}}\right)\epsilon - 2\gamma\left(\sum_{i=1}^{m} w_i + \sum_{i=m+1}^{m+n} w_i\right)\epsilon = 0 \tag{71}$$

This together with Lemma 36, reduces (70) to

$$\left|\sum_{i=1}^{m}\hat{w}_i^2 - \sum_{j=m+1}^{m+n}\hat{w}_j^2\right| \leq (\frac{1}{w_n}+\gamma)^2\left(\sum_{i=1}^{m+n} w_i\right)\epsilon^2 \tag{72}$$

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

$$|\hat{w}_i\hat{w}_j - w_i^*w_j^*| \leq |w_iw_j - w_i^*w_j^*| - \frac{2w_iw_j}{w_{m+n}}\epsilon + w_iw_j\left(\frac{1}{w_{m+n}^2} - \gamma^2\right)\epsilon^2 \tag{73}$$

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

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

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

$$|\hat{w}_i\hat{w}_j-w_i^*w_j^*|\leq|w_iw_j-w_i^*w_j^*|+w_iw_j\left(\frac{\gamma}{w_{m+n}}-\gamma\right)^2\epsilon^2 \tag{75}$$

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

$$|\hat{w}_i\hat{w}_j-w_i^*w_j^*|\leq|w_iw_j-w_i^*w_j^*|+w_iw_j\gamma^2\epsilon^2 \tag{76}$$

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 37, $(T_1^u\cup T_1^v, T_2^u\cup T_2^v)$ 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

$$\frac{w_1}{w_1^*}=\cdots=\frac{w_m}{w_m^*}=\frac{w_{m+1}^*}{w_{m+1}}=\cdots=\frac{w_{m+n}^*}{w_{m+n}}>1 \tag{77}$$

Therefore,

$$\sum_{i=1}^{m}w_i>\sum_{i=1}^{m}w_i^*,\qquad\sum_{i=m+1}^{m+n}w_i^*>\sum_{i=m+1}^{m+n}w_i \tag{78}$$

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

$$\sum_{i=1}^{m}w_i>\sum_{i=m+1}^{m+n}w_i \tag{79}$$

which, according to Lemma 36, contradicts the D-stationarity of $\mathbf{w}$. This completes the proof. ■

Appendix D. Proof of Proposition 17:

To prove Proposition 17, we present another important result on Erdös-Rényi random graphs.

Lemma 38 (Erdös and Rényi (1959)) *Assuming that $np\to 0$ as $n\to\infty$, 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 17: Assuming that $np\to 0$, Lemma 38 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 10, immediately will conclude the proof. One can write

$$\begin{aligned}\mathbb{P}(\mathcal{G}(\Omega) \text{ has a bipartite comp.}) &\overset{(a)}{\geq} \mathbb{P}(\mathcal{G}(\Omega) \text{ has a tree comp. without self loops})\\ &\geq \mathbb{P}(\text{all comp.'s are tree with size } O(\log n))\\ \times\, \mathbb{P}(\text{no self-loop in at least one comp}&|\text{all comp.'s are tree with size } O(\log n))\\ &\overset{(b)}{=} \mathbb{P}(\text{all comp.'s are tree with size } O(\log n))\\ \times\, \mathbb{P}(\text{no self-loop in at least one comp}&|\text{all comp.'s have the size } O(\log n))\\ &\geq \mathbb{P}(\text{all comp.'s are tree with size } O(\log n))\\ \times\, (1-\mathbb{P}(\text{all comp.'s have self-loops}&|\text{all comp.'s have the size } O(\log n)))\\ &\geq \mathbb{P}(\underbrace{\text{all comp.'s are tree with size } O(\log n)}_{\mathcal{A}})\\ \times\, (1-\mathbb{P}(\underbrace{\text{there are at least } \Omega(n/\log n) \text{ self-loops}}_{\mathcal{B}})) & \qquad (80)\end{aligned}$$

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 15, we have $\mathbb{P}(\mathcal{A}) \to 1$ as $n \to \infty$. Now, we only need to show that $\mathbb{P}(\mathcal{B}) \to 0$ as $n \to \infty$. One can easily verify that

$$\mathbb{P}(\mathcal{B}) \leq \binom{n}{\frac{n}{\log n}} p^{\frac{n}{\log n}} \tag{81}$$

Now, based on the Stirling's formula, one can write

$$n! \to \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{82}$$

$$\left(\frac{n}{\log n}\right)! \to \sqrt{2\pi\frac{n}{\log n}}\left(\frac{n}{e\log n}\right)^{\frac{n}{\log n}} \tag{83}$$

$$\left(n-\frac{n}{\log n}\right)! \to \sqrt{2\pi\left(n-\frac{n}{\log n}\right)}\left(\frac{n\log n - n}{e\log n}\right)^{n-\frac{n}{\log n}} \tag{84}$$

as $n \to \infty$. Plugging this back into (81) and following some simple algebra, one can verify that

$$\begin{aligned}\mathbb{P}(\mathcal{B}) &\leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\log^2 n}{n(\log n - 1)}}\cdot\frac{\log^n n}{(\log n - 1)^n}\cdot(\log n - 1)^{\frac{n}{\log n}} p^{\frac{n}{\log n}}\\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\log^2 n}{n(\log n - 1)}}\cdot(\log n - 1)^{\frac{n}{\log n}} p^{\frac{n}{\log n}}\end{aligned} \tag{85}$$

as $n \to \infty$. Replacing $p = o(1/n)$ gives rise to

$$\mathbb{P}(\mathcal{B}) \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\log^2 n}{n(\log n - 1)}}\cdot o\left(\left(\frac{\log n}{n}\right)^{\frac{n}{\log n}}\right) \to 0 \tag{86}$$

as $n \to \infty$. Together with (80), this implies that $\mathcal{G}(\Omega)$ will have a bipartite component without self loops with probability approaching to one. ■

Appendix E. 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

$$\begin{aligned}
f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) &\leq \left(\sum_{i=1}^n u_i\right)\epsilon + \lambda\left((u_{\max} - \epsilon - \beta)^4 - (u_{\max} - \beta)^4\right) \\
&= \left(\sum_{i=1}^n u_i\right)\epsilon - 4\lambda(u_{\max} - \beta)^3\epsilon + O(\epsilon^2) \\
&\overset{(a)}{\leq} \left(\sum_{i=1}^n u_i - \frac{\lambda}{2}u_{\max}^3\right)\epsilon + O(\epsilon^2) \\
&\leq \left(nu_{\max} - \frac{\lambda}{2}u_{\max}^3\right)\epsilon + O(\epsilon^2)
\end{aligned} \tag{87}$$

where (a) is due to the fact that $u_{\max} \geq 2\beta$ implies $u_{\max} - \beta \geq u_{\max}/2$. (87) 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 $(c/2)u^{*^2}_{\min} \leq u_{\min}$. By contradiction, suppose that there exists an index i such that $(c/2)u^{*^2}_{\min} > u_i$. 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:

$$\sum_{j\in\Omega_i}|u_iu_j - X_{ij}| = \sum_{j\in G_i}|u_iu_j - u_i^*u_j^*| + \sum_{j\in B_i}|u_iu_j - (u_i^*u_j^* + S_{ij})| \tag{88}$$

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

$$u_iu_j < cu^{*^2}_{\min} \leq u^{*^2}_{\min} \leq u_i^*u_j^* \tag{89}$$

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

$$u_iu_j < cu^{*^2}_{\min} \overset{(a)}{\leq} u_i^*u_j^* + S_{ij} \tag{90}$$

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. ■

Appendix F. Proof of Theorem 22:

The next lemma is crucial in proving Theorem 22.

Lemma 39 *Suppose that the assumptions of Theorem 22 hold and define*

$$s(\mathbf{u}) = \underbrace{-\sum_{\substack{(i,j)\in\mathcal{B}\\ i,j\in T_1}}\frac{2u_iu_j}{u_n}+\sum_{\substack{(i,j)\in\mathcal{B}\\ i,j\in T_2}}\frac{2u_iu_j}{u_n}+\sum_{\substack{(i,j)\in\mathcal{B}\\ i\in T_1\cup T_2, j\in N}}\frac{u_iu_j}{u_n}}_{f_B(\mathbf{u})}$$
$$+\underbrace{\sum_{\substack{(i,j)\in\mathcal{G}\\ i,j\in T_1}}\frac{2u_iu_j}{u_n}+\sum_{\substack{(i,j)\in\mathcal{G}\\ i,j\in T_2}}\frac{2u_iu_j}{u_n}+\sum_{\substack{(i,j)\in\mathcal{G}\\ i\in T_1\cup T_2, j\in N}}\frac{u_iu_j}{u_n}}_{f_G(\mathbf{u})}+\underbrace{\sum_{i\in T_2}\frac{4u_i(u_i-1)^3}{u_n}\mathbb{I}_{u_i\geq 1}}_{f_R(\mathbf{u})} \tag{91}$$

where the sets T_1 and T_2 are defined as (19) *and* (20), *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 (21).

Proof To prove this lemma, first we show the validity of (18). By contradiction, suppose that (18) 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 $u_i/u_i^* < u_1^*/u_1$ can be argued in a similar way). This implies that $u_iu_j > u_i^*u_j^*$ 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

$$\begin{aligned} f_{\text{reg}}(\hat{\mathbf{u}}) - f_{\text{reg}}(\mathbf{u}) &\leq -\left(\sum_{j\in G_i}u_j\right)\epsilon + \left(\sum_{j\in B_i}u_j\right)\epsilon + \lambda\left((u_i-\epsilon-1)^4-(u_i-1)^4\right)\mathbb{I}_{u_i\geq 1} \\ &\leq -\left(\sum_{j\in G_i}u_j\right)\epsilon + \left(\sum_{j\in B_i}u_j\right)\epsilon \\ &\leq -\frac{cu_{\min}^{*^2}}{2}\delta(\mathcal{G}(G)) + 2\Delta(\mathcal{G}(B)) \end{aligned} \tag{92}$$

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 $\left((u_i-\epsilon-1)^4-(u_i-1)^4\right)\mathbb{I}_{u_i\geq 1}$ 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

$$\frac{\delta(\mathcal{G}(G))}{\Delta(\mathcal{G}(B))} > \frac{48}{c^2}\kappa(\mathbf{u}^*)^4 = \frac{48}{c^2u_{\min}^{*^4}} > \frac{4}{cu_{\min}^{*^2}} \tag{93}$$

which implies $(-cu_{\min}^{*^2}/2)\delta(\mathcal{G}(G))+2\Delta(\mathcal{G}(B))<0$, 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 (18) holds. Now, we will show the correctness of the first statement. Similar to the proof of Theorem 7, one can verify that

$$\sum_{(i,j)\in\Omega}|\hat{u}_i\hat{u}_j-X_{ij}|-\sum_{(i,j)\in\Omega}|u_iu_j-X_{ij}|\leq(f_B(\mathbf{u})-f_G(\mathbf{u}))\epsilon+O(\epsilon^2) \tag{94}$$

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

$$\begin{aligned}R(\hat{\mathbf{u}})-R(\mathbf{u})&=\sum_{i\in T_2}\left(u_i-\frac{u_i}{u_n}\epsilon-1\right)^4\mathbb{I}_{u_i\geq 1}-\sum_{i\in T_2}(u_i-1)^4\,\mathbb{I}_{u_i\geq 1}\\&=-\sum_{i\in T_2}\frac{4u_i(u_i-1)^3}{u_n}\mathbb{I}_{u_i\geq 1}\epsilon+O(\epsilon^2)\end{aligned} \tag{95}$$

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

Lemma 40 *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 8. ■

Proof of Theorem 22: Similar to the proof of Theorem 7, 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 39. 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

$$\begin{aligned}f_B(\mathbf{u})&\leq\frac{1}{u_n}\left(2\cdot\frac{\Delta(\mathcal{G}(B))}{2}|T_1|u_{\max}^2+2\cdot\frac{\Delta(\mathcal{G}(B))}{2}|T_2|u_{\max}^2+\frac{\Delta(\mathcal{G}(B))}{2}(|T_1|+|T_2|)u_{\max}^2\right)\epsilon+O(\epsilon^2)\\&\leq\frac{3}{2u_n}\Delta(\mathcal{G}(B))(|T_1|+|T_2|)u_{\max}^2\epsilon+O(\epsilon^2)\\&\leq\frac{6}{u_n}\Delta(\mathcal{G}(B))(|T_1|+|T_2|)\epsilon+O(\epsilon^2)\end{aligned} \tag{96}$$

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

$$\begin{aligned} f_G(\mathbf{x}) \geq & \frac{1}{u_n} \cdot \frac{\delta(\mathcal{G}(G))}{2}(|T_1|+|T_2|)u_{\min}^2\epsilon + O(\epsilon^2) \\ \geq & \frac{1}{u_n} \cdot \frac{\delta(\mathcal{G}(G))}{2}(|T_1|+|T_2|)\frac{c^2 {u^*_{\min}}^4}{4}\epsilon + O(\epsilon^2) \\ = & \frac{c^2 {u^*_{\min}}^4}{8u_n}\delta(\mathcal{G}(G))(|T_1|+|T_2|)\epsilon + O(\epsilon^2) \end{aligned} \tag{97}$$

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 $\mathbf{u}\mathbf{u}^T$ are all equal to $u_{\min}^2$. Furthermore, the second inequality is due to Lemma 15 and the choice of β for $R(\mathbf{u})$. Therefore, one can write

$$\begin{aligned} f_B(\mathbf{x}) - f_G(\mathbf{x}) &\leq \left(\frac{6}{u_n}\Delta(\mathcal{G}(B)) - \frac{c^2 {u^*_{\min}}^4}{8u_n}\delta(\mathcal{G}(G))\right)(|T_1|+|T_2|)\epsilon + O(\epsilon^2) \\ &= \frac{\Delta(\mathcal{G}(B))c^2 {u^*_{\min}}^4}{8u_n}\left(\frac{48}{c^2}\kappa(\mathbf{u}^*)^4 - \frac{\delta(\mathcal{G}(G))}{\Delta(\mathcal{G}(B))}\right)(|T_1|+|T_2|)\epsilon + O(\epsilon^2). \end{aligned} \tag{98}$$

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. ■

Appendix G. Proof of Lemma 25

The degree of each node is the summation of n independent Bernoulli random variables, each with parameter p. Therefore, standard concentration bounds yields that

$$\mathbb{P}(\deg(v) \geq (1+\delta)np) \leq e^{-np\delta^2/3} \tag{99a}$$

$$\mathbb{P}(\deg(v) \leq (1-\delta)np) \leq e^{-np\delta^2/3} \tag{99b}$$

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

$$\mathbb{P}(\Delta(\mathcal{G}(n,p)) \geq (1+\delta)np) \leq ne^{-np\delta^2/3} \tag{100a}$$

$$\mathbb{P}(\delta(\mathcal{G}(n,p)) \leq (1-\delta)np) \leq ne^{-np\delta^2/3} \tag{100b}$$

Setting $\delta = 1/2$ and assuming that $p \geq 24\log n/n$, one can write

$$\mathbb{P}\left(\Delta(\mathcal{G}(n,p)) \geq \frac{3np}{2}\right) \leq \frac{1}{n} \tag{101a}$$

$$\mathbb{P}\left(\delta(\mathcal{G}(n,p)) \leq \frac{np}{2}\right) \leq \frac{1}{n} \tag{101b}$$

Furthermore, $p < 24 \log n / n$ leads to

$$\begin{aligned}\mathbb{P}\left(\Delta(\mathcal{G}(n,p)) \geq 36 \log n\right) &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{24 \log n}{n}\right)\right) \geq 36 \log n\right)\\ &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{24 \log n}{n}\right)\right) \geq \frac{3np}{2}\right)\\ &\leq \frac{1}{n} \qquad (102)\end{aligned}$$

Combining (102) with (101a) and (101b) results in the desired inequalities. ■

Appendix H. Proof of Lemma 31

Define $S = \{1, ..., m\}$ and $T = \{m+1, ..., m+n\}$. Similar to the proof of Lemma 21, one can write the following concentration inequalities:

$$\mathbb{P}(\max_{v \in S}\{\deg(v)\} \geq (1+\delta)np) \leq me^{-np\delta^2/3} \qquad (103a)$$

$$\mathbb{P}(\min_{v \in S}\{\deg(v)\} \leq (1-\delta)np) \leq me^{-np\delta^2/3} \qquad (103b)$$

$$\mathbb{P}(\max_{v \in T}\{\deg(v)\} \geq (1+\delta)mp) \leq ne^{-mp\delta^2/3} \qquad (103c)$$

$$\mathbb{P}(\min_{v \in T}\{\deg(v)\} \leq (1-\delta)mp) \leq ne^{-mp\delta^2/3} \qquad (103d)$$

which imply

$$\mathbb{P}(\Delta(\mathcal{G}(m,n,p)) \geq (1+\delta)np) \leq me^{-np\delta^2/3} + ne^{-mp\delta^2/3} \leq 2ne^{-mp\delta^2/3} \qquad (104a)$$

$$\mathbb{P}(\delta(\mathcal{G}(m,n,p)) \leq (1-\delta)mp) \leq me^{-np\delta^2/3} + ne^{-mp\delta^2/3} \leq 2ne^{-mp\delta^2/3} \qquad (104b)$$

Setting $\delta = 1/2$ and assuming that $p \geq 24 \log n / m$ results in

$$\mathbb{P}(\Delta(\mathcal{G}(m,n,p)) \geq \frac{3np}{2}) \leq \frac{2}{n} \qquad (105a)$$

$$\mathbb{P}(\delta(\mathcal{G}(m,n,p)) \leq \frac{mp}{2}) \leq \frac{2}{n} \qquad (105b)$$

Furthermore, if $p < 24 \log n / m$, one can write

$$\begin{aligned}\mathbb{P}\left(\Delta(\mathcal{G}(m,n,p)) \geq \frac{36n \log n}{m}\right) &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{24 \log n}{m}\right)\right) \geq \frac{36n \log n}{m}\right)\\ &\leq \mathbb{P}\left(\Delta\left(\mathcal{G}\left(n, \frac{24 \log n}{m}\right)\right) \geq \frac{3np}{2}\right)\\ &\leq \frac{2}{n} \qquad (106)\end{aligned}$$

This completes the proof. ■