Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals

FANGZHENG XIE[1,a],

[1] *Department of Statistics, Indiana University, Bloomington, United States.* [a] *fxie@iu.edu*

We establish a finite-sample Berry-Esseen theorem for the entrywise limits of the eigenvectors for a broad collection of signal-plus-noise random matrix models under challenging weak signal regimes. The signal strength is characterized by a scaling factor ρ_n through $n\rho_n$, where n is the dimension of the random matrix, and we allow $n\rho_n$ to grow at the rate of $\log n$. The key technical contribution is a sharp finite-sample entrywise eigenvector perturbation bound. The existing error bounds on the two-to-infinity norms of the higher-order remainders are not sufficient when $n\rho_n$ is proportional to $\log n$. We apply the general entrywise eigenvector analysis results to the symmetric noisy matrix completion problem, random dot product graphs, and two subsequent inference tasks for random graphs: the estimation of pure nodes in mixed membership stochastic block models and the hypothesis testing of the equality of latent positions in random graphs.

Keywords: Berry-Esseen theorem; Entrywise eigenvector analysis; Random dot product graphs; Signal-plus-noise matrix model; Symmetric noisy matrix completion

1. Introduction

In the contemporary world of data science, many statistical problems involve random matrix models with low-rank structures. Random matrices with low expected rank, also referred to as the signal-plus-noise matrix models, are pervasive in many applications, including social networks [39,58,84], compressed sensing [26,28], and recommendation systems [14,36]. A broad range of statistical models also fall into the category of signal-plus-noise matrix models, such as the low-rank matrix denoising model [23,27,65], matrix completion problems [17,18,46], principal component analysis [7,44], and stochastic block models [1,39].

In signal-plus-noise matrix models, spectral estimators and eigenvectors of random matrices have been extensively explored. These estimators can either be applied to obtain the desired inference results [61,66,69] or serve as ideal initial guesses of certain iterative algorithms [33,46,82]. The theoretical support of spectral estimators is fundamentally backboned by the matrix perturbation theory [16,25,67,78] and the recent progress in random matrix theory [11,13,59,83]. From the practical perspective, the implementation of these spectral-based estimators typically only requires the truncated spectral/singular value decomposition of the data matrix, which is computationally cheap. In contrast, the maximum likelihood estimators for low-rank matrix models are less preferred because they are intractable to compute in general due to the nonconvex optimization problems involved [3].

1.1. Overview

This paper investigates the entrywise behavior of the leading eigenvectors of a symmetric random matrix $\mathbf{A}$ whose expected value $\mathbf{P} = \mathbb{E}\mathbf{A}$ has a low rank. This class of random matrix models is referred

to as the signal-plus-noise matrix models (see Section 2.1 for the formal description). We establish a generic finite-sample Berry-Esseen theorem for the rows of the leading eigenvectors under challenging weak signal regimes. The resulting Berry-Esseen bound is quite general and allows for a possibly increasing rank($\mathbf{P}$).

As a special case of the entrywise eigenvector limit theorem for the signal-plus-noise matrix models, we obtain the Berry-Esseen bounds for the entrywise estimates for the symmetric noisy matrix completion model (see Section 4.1 for the formal definition). Our analysis is sharper than the entrywise maximum norm error bounds for $\mathbf{P}$ obtained by [3].

Our generic entrywise Berry-Esseen theorem leads to the limit results of the rows of the adjacency spectral embedding of the random dot product graph model (see Section 4.2 for the formal definition) under the sparse regime that the graph average expected degree is at the order of $\Omega(\log n)$, where n is the number of vertices. The sparsity assumption is minimal because the graph adjacency matrix $\mathbf{A}$ no longer concentrates around its expected value $\mathbf{P}$ when the average expected degree is $o(\log n)$. Our result also relaxes the sparsity assumptions posited in [21,72,82].

Leveraging the generic entrywise eigenvector concentration bound for the signal-plus-noise matrix models, we further study the entrywise limit theorem of the one-step refinement of the eigenvectors for random dot product graphs proposed in [82]. The corresponding covariance matrix of the rows of the one-step estimator is no greater than that of the rows of the eigenvectors. We then investigate the impact of the one-step estimator for two subsequent inference tasks. Specifically, the one-step estimator has smaller asymptotic variances than the eigenvectors for estimating the pure nodes in mixed membership stochastic block models; It also leads to a more powerful test than the eigenvectors for testing the equality of latent positions in random dot product graphs.

1.2. A motivating example

Let us take a glimpse into a simple yet popular random graph model that has attracted much attention in the recent decade: the stochastic block model. Consider a graph with n vertices that are labeled as $[n] := \{1,2,\ldots,n\}$. These vertices are partitioned into two communities by a community assignment rule $\tau : [n] \to \{1,2\}$, where $\tau(i) = 1$ indicates that vertex i lies in the first community, and $\tau(i) = 2$ otherwise. Let $\mathbf{A} = [A_{ij}]_{n\times n}$ be the adjacency matrix of the stochastic block model, $\rho_n \in (0,1]$ be the sparsity factor, and $a, b \in (0,1)$ be constants. For each vertex pair (i,j) with $i \le j$, $(A_{ij})_{i\le j}$ are independent, $A_{ij} \sim \text{Bernoulli}(\rho_n a)$ if $\tau(i) = \tau(j)$, $A_{ij} \sim \text{Bernoulli}(\rho_n b)$ if $\tau(i) \neq \tau(j)$, and $A_{ij} = A_{ji}$ for all $i > j$. Here, $\rho_n a$ and $\rho_n b$ represent the within-community probability and between-community probability, respectively, and $n\rho_n$ governs the growing rate of the graph average expected degree as a function of n.

The stochastic block models were first introduced in [39] and have motivated the development of network science and analysis substantially in recent years. There have also been countless papers addressing statistical analyses of stochastic block models and their fundamental limits. The readers are referred to [1] for a survey.

A fundamental inference task for stochastic block models is the community detection, namely, the recovery of the cluster assignment rule τ. In the context of the aforementioned two-block stochastic block model, we are particularly interested in the case where $n\rho_n = \alpha \log n$ for some constant $\alpha > 0$. There are, however, other fundamental aspects of the behavior of the leading eigenvectors of the adjacency matrix $\mathbf{A}$ beyond the community detection. In this work, we focus on the asymptotic distribution of the rows of the leading eigenvector matrix of $\mathbf{A}$. We begin the analysis with the population eigenvectors. For simplicity, we assume that n is an even integer, $\tau(i) = 1$ if $i = 1,\ldots,n/2$, and $\tau(i) = 2$ if $i = n/2+1,\ldots,n$. Namely, the first $n/2$ vertices are in the first community, and the rest of the $n/2$ vertices fall into the second community. The non-zero eigenvalues of $\mathbb{E}\mathbf{A}$ are $\lambda_1 = n\rho_n(a+b)/2$ and $\lambda_2 = n\rho_n(a-b)/2$,

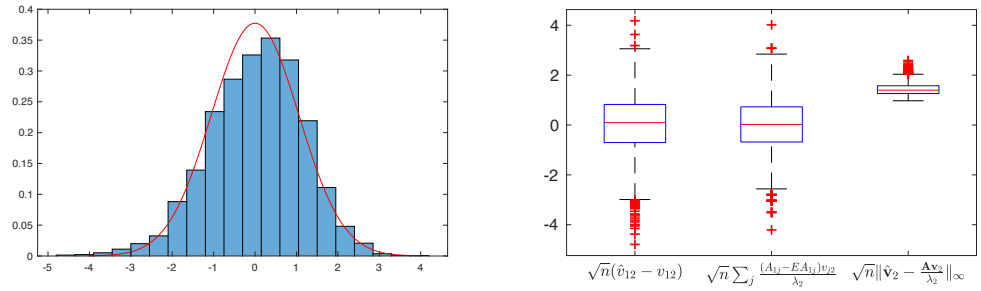

Figure 1. Left panel: The histogram of $\sqrt{n}(\widehat{v}_{12} - v_{12})$ over the 3000 Monte Carlo replicates with the density of $\mathrm{N}(0,(a+b)/(a-b))$ highlighted in the red curve. Right panel: The boxplots of $\sqrt{n}(\widehat{v}_{12} - v_{12})$, its linear approximation $\sqrt{n}\sum_j(A_{1j} - \mathbb{E}A_{1j})v_{j2}/\lambda_2$, and the infinity norm of the remainder $\sqrt{n}\|\widehat{\mathbf{v}}_2 - \mathbf{A}\mathbf{v}_2/\lambda_2\|_\infty$ across the 3000 Monte Carlo replicates.

and the associated eigenvectors are $\mathbf{u}_1 = n^{-1/2}[1,\ldots,1]^{\mathrm{T}}$ and $\mathbf{u}_2 = n^{-1/2}[1,\ldots,1,-1,\ldots,-1]^{\mathrm{T}}$. We also consider the scaled eigenvectors $\mathbf{v}_1 = \lambda_1^{1/2}\mathbf{u}_1$ and $\mathbf{v}_2 = \lambda_2^{1/2}\mathbf{u}_2$. Because $\mathbf{v}_1$ and $\mathbf{u}_1$ are non-informative for the community structure whereas the signs of $\mathbf{v}_2$ and $\mathbf{u}_2$ encode the community assignment, we focus on $\mathbf{v}_2$ and $\mathbf{u}_2$. Let $\widehat{\mathbf{v}}_2 = [\widehat{v}_{12},\ldots,\widehat{v}_{n2}]^{\mathrm{T}}$ be the eigenvector of $\mathbf{A}$ associated with the second largest eigenvalue $\widehat{\lambda}_2$ of $\mathbf{A}$ and $\widehat{\mathbf{u}}_2 = [\widehat{u}_{12},\ldots,\widehat{u}_{n2}]^{\mathrm{T}} = \widehat{\mathbf{v}}_2/\|\widehat{\mathbf{v}}_2\|_2$. We scale $\widehat{\mathbf{v}}_2$ such that $\|\widehat{\mathbf{v}}_2\|_2 = |\widehat{\lambda}_2|^{1/2}$ to keep the scaling consistent.

To explore the entrywise asymptotic distributions of $\widehat{\mathbf{v}}_2$ and $\widehat{\mathbf{u}}_2$, we consider the following decompositions motivated by [3] and [21] for each fixed $i \in [n]$:

$$\sqrt{n}(\widehat{v}_{i2} - v_{i2}) = \sqrt{n}\sum_{j=1}^{n}\frac{(A_{ij} - \mathbb{E}A_{ij})v_{2j}}{\lambda_2} + \sqrt{n}\left(\widehat{v}_{i2} - \sum_{j=1}^{n}\frac{A_{ij}v_{j2}}{\lambda_2}\right), \tag{1}$$

$$n\rho_n^{1/2}(\widehat{u}_{i2} - u_{i2}) = n\rho_n^{1/2}\sum_{j=1}^{n}\frac{(A_{ij} - \mathbb{E}A_{ij})u_{2j}}{\lambda_2} + n\rho_n^{1/2}\left(\widehat{u}_{i2} - \sum_{j=1}^{n}\frac{A_{ij}u_{j2}}{\lambda_2}\right). \tag{2}$$

The key observation is that the first terms on the right-hand sides of (1) and (2) are two sums of independent mean-zero random variables. These two terms converge to $\mathrm{N}(0,(a+b)/(a-b))$ and $\mathrm{N}(0,2(a+b)/(a-b)^2)$ in distribution, respectively, by Lyapunov's central limit theorem (see, e.g., Theorem 7.1.2. in [24]). The technical challenge lies in sharp controls of the second terms arising in these equations.

We pause the theoretical discussion for a moment and turn to a simulation study. The parameters for the simulation are set as follows: $n = 5000$, $\alpha = 5$, $a = 0.9$, $b = 0.05$, and $n\rho_n = \alpha\log n$. We then generate 3000 independent Monte Carlo replicates of $\mathbf{A}$ and compute the corresponding eigenvectors $\widehat{\mathbf{v}}_2$ and $\widehat{\mathbf{u}}_2$. The left panels of Figures 1 and 2 visualize the histograms of $\sqrt{n}(\widehat{v}_{12} - v_{12})$ and $n\rho_n^{1/2}(\widehat{u}_{12} - u_{12})$ (for the vertex $i = 1$), respectively. The shapes of the two histograms are closely aligned with the corresponding asymptotic normal densities. This observation leads to the conjecture that $\sqrt{n}(\widehat{v}_{i2} - v_{i2})$ and $n\rho_n^{1/2}(\widehat{u}_{i2} - u_{i2})$ are asymptotically normal.

Continuing the theoretical investigation of $\sqrt{n}(\widehat{v}_{i2} - v_{i2})$ and $n\rho_n^{1/2}(\widehat{u}_{i2} - u_{i2})$, we can write (1) and (2) alternatively as

$$\sqrt{n}(\widehat{\mathbf{v}}_2 - \mathbf{v}_2) = \sqrt{n}\frac{(\mathbf{A} - \mathbb{E}\mathbf{A})\mathbf{v}_2}{\lambda_2} + \sqrt{n}\left(\widehat{\mathbf{v}}_2 - \frac{\mathbf{A}\mathbf{v}_2}{\lambda_2}\right),$$

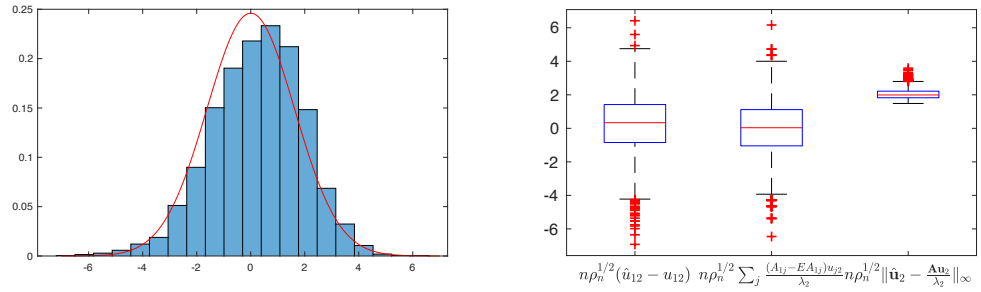

Figure 2. Left panel: The histogram of $n\rho_n^{1/2}(\widehat{u}_{12}-u_{12})$ over the 3000 Monte Carlo replicates with the density of $\mathrm{N}(0,2(a+b)/(a-b)^2)$ highlighted in the red curve. Right panel: The boxplots of $n\rho_n^{1/2}(\widehat{u}_{12}-u_{12})$, its linear approximation $n\rho_n^{1/2}\sum_j(A_{1j}-\mathbb{E}A_{1j})u_{j2}/\lambda_2$, and the infinity norm of the remainder $n\rho_n^{1/2}\|\widehat{\mathbf{u}}_2-\mathbf{A}\mathbf{u}_2/\lambda_2\|_\infty$ across the 3000 Monte Carlo replicates.

$$n\rho_n^{1/2}(\widehat{\mathbf{u}}_2-\mathbf{u}_2)=n\rho_n^{1/2}\frac{(\mathbf{A}-\mathbb{E}\mathbf{A})\mathbf{u}_2}{\lambda_2}+n\rho_n^{1/2}\left(\widehat{\mathbf{u}}_2-\frac{\mathbf{A}\mathbf{u}_2}{\lambda_2}\right).$$

One seemingly plausible approach is to show that $\sqrt{n}\|\widehat{\mathbf{v}}_2-\mathbf{A}\mathbf{v}_2/\lambda_2\|_\infty$ and $n\rho_n^{1/2}\|\widehat{\mathbf{u}}_2-\mathbf{A}\mathbf{u}_2/\lambda_2\|_\infty$ are $o_{\mathbb{P}}(1)$ using the recently developed tools in [3,21,22,29,32,50,56]. However, the right panels of Figures 1 and 2 suggest that this strategy may fail. Taking the unscaled eigenvectors for example, we present the boxplots of $n\rho_n^{1/2}(\widehat{u}_{12}-u_{12})$, $n\rho_n^{1/2}\sum_{j=1}^n(A_{1j}-\mathbb{E}A_{1j})u_{j2}/\lambda_2$, and $n\rho_n^{1/2}\|\widehat{\mathbf{u}}_2-\mathbf{A}\mathbf{u}_2/\lambda_2\|_\infty$ over the aforementioned 3000 Monte Carlo replicates in the right panel of Figure 2. The boxplots suggest that $n\rho_n^{1/2}\|\widehat{\mathbf{u}}_2-\mathbf{A}\mathbf{u}_2/\lambda_2\|_\infty\neq o_{\mathbb{P}}(1)$. A similar phenomenon for the scaled eigenvectors can also be observed from the right panel of Figure 1. These numerical results motivate us to explore the entrywise limits of the eigenvectors for signal-plus-noise matrices beyond the two-to-infinity error bounds.

1.3. Related work

Entrywise limit theorems for the eigenvectors of random matrices first appeared in the context of network models. Based on the random dot product graph model [58,84], the authors of [10] explored the asymptotic distributions of the rows of the eigenvectors of the random adjacency matrix for dense graphs. Generalizations of [10] to sparse graphs were later explored in [72] and [82] under a weaker condition that the average expected degree scales at $\omega((\log n)^4)$. The authors of [21] established a general entrywise limit theorem for the eigenvectors of random matrices with low expected rank by exploiting the von-Neumann matrix series expansion of the solution to a matrix Sylvester equation [15,29]. Recently, a general framework for studying the asymptotic theory of eigenvectors for generalized spiked Wigner models has been developed in [30].

Another line of the related research is on the two-to-infinity norm error bounds for eigenvectors of random matrices. Previously, the authors of [53] have explored the exact community detection of stochastic block models by studying the eigenvector error bound with respect to the two-to-infinity norm. Recently, the authors of [22] established a general framework for studying the two-to-infinity norm eigenvector perturbation bounds. However, the deterministic nature of their approach may lead to sub-optimal results in challenging low signal-to-noise ratio regimes [3]. Since then, several related papers have emerged to address the entrywise eigenvector estimation problems under various contexts [3,4,21,50,79].

The literature on the specific applications considered in this paper is quite rich. The symmetric noisy matrix completion models can be viewed as a special case of the general noisy matrix completion models for rectangular random matrices, which have seen enormous progress in the past decades. For an incomplete list of reference, see [14,17–19,23,38,41,46,47]. The random dot product graphs, which were originally developed for social networks [58,84], have been extensively studied in recent years, including the theoretical properties [10,69,71–73,81,82] and the involved applications [60,74]. We refer to the survey paper [9] for a review of random dot product graphs.

1.4. Organization

The rest of the paper is structured as follows. Section 2 sets the stage for the generic signal-plus-noise matrix models and introduces the corresponding entrywise eigenvector analysis framework. Section 3, which is the main technical contribution of this paper, elaborates on the Berry-Esseen theorem for the rows of the eigenvectors of the signal-plus-noise matrix models. We apply the main results to the symmetric noisy matrix completion models and random dot product graphs in Section 4. Section 5 provides illustrative numerical examples, and we conclude the paper with some discussions concerning future extensions in Section 6.

1.5. Notations

The symbol := is used to assign mathematical definitions. For any positive integer n, let $[n] := \{1,2,\ldots,n\}$. The set of all positive integers is denoted by $\mathbb{N}_+$. For any $a,b \in \mathbb{R}$, we denote $a \wedge b := \min(a,b)$ and $a \vee b := \max(a,b)$. For any two non-negative sequences $(a_n)_{n\in\mathbb{N}_+}$, $(b_n)_{n\in\mathbb{N}_+}$, we write $a_n \lesssim b_n$ ($a_n \gtrsim b_n$, resp.), if there exists some absolute constant $C > 0$, such that $a_n \le Cb_n$ ($a_n \ge Cb_n$, resp.) for all $n \in \mathbb{N}_+$. If the constant C also depends on another parameter c that is independent of $n \in \mathbb{N}_+$, then we write $a_n \lesssim_c b_n$ ($a_n \gtrsim_c b_n$, resp.). We use $K_c, N_c, \ldots$ to denote constants that may depend on another parameter c but is independent of the varying index $n \in \mathbb{N}_+$. Absolute constants are usually hidden using notations $\lesssim$ and $\gtrsim$, and, when necessary, we use C_0 and c_0 to denote generic absolute constants that may vary from line to line. We use the notation $a_n \asymp b_n$ to indicate that $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. If a_n/b_n stays bounded away from $+\infty$, we write $a_n = O(b_n)$ and $b_n = \Omega(a_n)$, and if $a_n/b_n \to 0$, we denote $a_n = o(b_n)$ and $b_n = \omega(a_n)$. For any symmetric positive semidefinite matrices $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$, we denote $\mathbf{\Sigma} \ge \mathbf{\Gamma}$ ($\mathbf{\Sigma} \le \mathbf{\Gamma}$, resp.), if $\mathbf{\Sigma} - \mathbf{\Gamma}$ ($\mathbf{\Gamma} - \mathbf{\Sigma}$, resp.) is positive semidefinite. When $\mathbf{\Sigma} - \mathbf{\Gamma}$ ($\mathbf{\Gamma} - \mathbf{\Sigma}$, resp.) is strictly positive definite, we use the notation $\mathbf{\Sigma} > \mathbf{\Gamma}$ ($\mathbf{\Sigma} < \mathbf{\Gamma}$, resp.). For any $d \in \mathbb{N}_+$, we use $\mathbf{I}_d$ to denote the $d \times d$ identity matrix and $\mathbf{0}_d$ to denote the zero vector in $\mathbb{R}^d$. For $n,d \in \mathbb{N}_+$, $n \ge d$, let $\mathbb{O}(n,d) := \{\mathbf{U} \in \mathbb{R}^{n\times d} : \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_d\}$ denote the set of all orthonormal d-frames in $\mathbb{R}^n$. When $n = d$, we simply write $\mathbb{O}(d) = \mathbb{O}(d,d)$. For an $n \times d$ matrix $\mathbf{M}$, we denote $\sigma_k(\mathbf{M})$ the kth largest singular value of $\mathbf{M}$, $k \in [\min(n,d)]$. For any $j \in [n]$ and $k \in [d]$, we use $[\mathbf{M}]_{j*}$ to denote its jth row, $[\mathbf{M}]_{*k}$ to denote its kth column, and $[\mathbf{M}]_{jk}$ to denote its (j,k)th entry. When $\mathbf{M} \in \mathbb{R}^{n\times n}$ is a square symmetric matrix, we use $\lambda_k(\mathbf{M})$ to denote the kth largest eigenvalue of $\mathbf{M}$, namely, $\lambda_1(\mathbf{M}) \ge \lambda_2(\mathbf{M}) \ge \ldots \ge \lambda_n(\mathbf{M})$, and $\mathrm{tr}(\mathbf{M})$ the trace of $\mathbf{M}$ (the sum of its diagonal elements). If $\mathbf{M} \in \mathbb{R}^{d\times d}$ is positive definite, then we let $\kappa(\mathbf{M}) := \lambda_1(\mathbf{M})/\lambda_d(\mathbf{M})$ denote the condition number of $\mathbf{M}$. The spectral norm of a rectangular matrix $\mathbf{M}$, denoted by $\|\mathbf{M}\|_2$, is defined as the largest singular value of $\mathbf{M}$. The Frobenius norm of a rectangular matrix $\mathbf{M}$, denoted by $\|\mathbf{M}\|_{\mathrm{F}}$, is defined as $\|\mathbf{M}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\mathbf{M}^{\mathrm{T}}\mathbf{M})}$. We use $\|\mathbf{M}\|_{2\to\infty}$ to denote the two-to-infinity norm of a matrix $\mathbf{M} = [M_{jk}]_{n\times d}$, defined as $\|\mathbf{M}\|_{2\to\infty} = \max_{j\in[n]} \sqrt{\sum_{k=1}^d M_{jk}^2}$, and $\|\mathbf{M}\|_\infty$ to denote the matrix infinity norm $\|\mathbf{M}\|_\infty = \max_{j\in[n]} \sum_{k=1}^d |M_{jk}|$. Given d real numbers $a_1,\ldots,a_d \in \mathbb{R}$, we let $\mathrm{diag}(a_1,\ldots,a_d)$ to denote the $d \times d$ diagonal matrix whose (k,k)th element is a_k

for $k \in [d]$. For a Euclidean vector $\mathbf{x} = [x_1, \ldots, x_d]^{\mathrm{T}} \in \mathbb{R}^d$, $\|\mathbf{x}\|_2$ denotes the Euclidean norm of $\mathbf{x}$ given by $\|\mathbf{x}\|_2 = \sqrt{\sum_{k=1}^{d} x_k^2}$ and $\|\mathbf{x}\|_\infty$ denotes the infinity norm of $\mathbf{x}$ defined as $\|\mathbf{x}\|_\infty = \max_{k\in[d]} |x_k|$. When the dimension of the underlying Euclidean space is clear, with use $\mathbf{e}_i$ to denote the unit basis vector whose ith coordinate is one and the rest of the coordinates are zeros.

2. Preliminaries

2.1. Setup

Consider an $n \times n$ symmetric observable data matrix $\mathbf{A}$ that can be viewed as a noisy version of an unobserved low-rank signal matrix $\mathbf{P}$ through the following signal-plus-noise matrix model:

$$\mathbf{A} = \mathbf{P} + \mathbf{E}, \tag{3}$$

where $\mathbf{E}$ is an $n \times n$ symmetric noise matrix that is unobserved. Suppose $\mathrm{rank}(\mathbf{P}) = d$ and $d \ll n$. Let $p \in \{1, 2, \ldots, d\}$ be the number of positive eigenvalues of $\mathbf{P}$ and $q{=}d - p$ be the number of negative eigenvalues of $\mathbf{P}$. Namely, $\lambda_1(\mathbf{P}) \geq \ldots \geq \lambda_p(\mathbf{P}) > 0 > \lambda_{n-q+1}(\mathbf{P}) \geq \ldots \geq \lambda_n(\mathbf{P})$. Let $\mathbf{U}_{\mathbf{P}_+} \in \mathbb{O}(n,p)$ be the eigenvector matrix of $\mathbf{P}$ corresponding to $\lambda_1(\mathbf{P}), \ldots, \lambda_p(\mathbf{P})$, $\mathbf{U}_{\mathbf{P}_-} \in \mathbb{O}(n,q)$ be the eigenvector matrix of $\mathbf{P}$ corresponding to $\lambda_{n-q+1}(\mathbf{P}), \ldots, \lambda_n(\mathbf{P})$, $\mathbf{U}_{\mathbf{P}} := [\mathbf{U}_{\mathbf{P}_+}, \mathbf{U}_{\mathbf{P}_-}]$, $\mathbf{S}_{\mathbf{P}_+} := \mathrm{diag}\{\lambda_1(\mathbf{P}), \ldots, \lambda_p(\mathbf{P})\}$, $\mathbf{S}_{\mathbf{P}_-} := \mathrm{diag}\{\lambda_{n-q+1}(\mathbf{P}), \ldots, \lambda_n(\mathbf{P})\}$, and $\mathbf{S}_{\mathbf{P}} := \mathrm{diag}(\mathbf{S}_{\mathbf{P}_+}, \mathbf{S}_{\mathbf{P}_-})$. For convenience, we introduce the inertia matrix $\mathbf{I}_{p,q} = \mathrm{diag}(\mathbf{I}_p, -\mathbf{I}_q)$ encoding the numbers of positive and negative eigenvalues of $\mathbf{P}$.

The signal matrix $\mathbf{P}$ is associated with a scaling factor $\rho_n \in (0, 1]$ that governs the signal strength of the model (3) through $n\rho_n$. For example, in the context of network models, $n\rho_n$ controls the average expected degree of the resulting random graphs. Note that the spectral decomposition of $\mathbf{P}$ can be written as $\mathbf{P} = \mathbf{U}_{\mathbf{P}_+}|\mathbf{S}_{\mathbf{P}_+}|\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}} - \mathbf{U}_{\mathbf{P}_-}|\mathbf{S}_{\mathbf{P}_-}|\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}$, where the absolute value $|\cdot|$ is applied entrywise on the eigenvalues. We define $\mathbf{X}_+ := \rho_n^{-1/2}\mathbf{U}_{\mathbf{P}_+}|\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_{\mathbf{X}_+}$ and $\mathbf{X}_- := \rho_n^{-1/2}\mathbf{U}_{\mathbf{P}_-}|\mathbf{S}_{\mathbf{P}_-}|^{1/2}\mathbf{W}_{\mathbf{X}_-}$, where $\mathbf{W}_{\mathbf{X}_+} \in \mathbb{O}(p)$ and $\mathbf{W}_{\mathbf{X}_-} \in \mathbb{O}(q)$ are deterministic orthogonal matrices. This allows us to write $\mathbf{P}$ alternatively as $\mathbf{P} = \rho_n\mathbf{X}_+\mathbf{X}_+^{\mathrm{T}} - \rho_n\mathbf{X}_-\mathbf{X}_-^{\mathrm{T}}$. Denote $\mathbf{X} := [\mathbf{X}_+, \mathbf{X}_-]$, $\mathbf{\Delta}_{n+} := (1/n)\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+$, $\mathbf{\Delta}_{n-} := (1/n)\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-$, $\mathbf{\Delta}_n := (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$, and $\mathbf{W}_{\mathbf{X}} := \mathrm{diag}(\mathbf{W}_{\mathbf{X}_+}, \mathbf{X}_{\mathbf{X}_-})$. Clearly, $\mathbf{\Delta}_n = \mathrm{diag}(\mathbf{\Delta}_{n+}, \mathbf{\Delta}_{n-})$ because of the orthogonality between $\mathbf{X}_+$ and $\mathbf{X}_-$.

The focus of this work is to characterize the entrywise limit behavior of the eigenvector matrices of the data matrix $\mathbf{A}$ as the sample versions of their population counterparts $\mathbf{U}_{\mathbf{P}_\pm}$ and $\mathbf{X}_\pm$. To this end, we let $\mathbf{U}_{\mathbf{A}_+} \in \mathbb{O}(n,p)$ be the eigenvector matrix of $\mathbf{A}$ corresponding to the sample eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$, $\mathbf{U}_{\mathbf{A}_-} \in \mathbb{O}(n,q)$ be the eigenvector matrix of $\mathbf{A}$ corresponding to the sample eigenvalues $\lambda_{n-q+1}(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$, $\mathbf{U}_{\mathbf{A}} := [\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{A}_-}]$, $\mathbf{S}_{\mathbf{A}_+} = \mathrm{diag}\{\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})\}$, $\mathbf{S}_{\mathbf{A}_-} = \mathrm{diag}\{\lambda_{n-q+1}(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})\}$, and $\mathbf{S}_{\mathbf{A}} := \mathrm{diag}(\mathbf{S}_{\mathbf{A}_+}, \mathbf{S}_{\mathbf{A}_-})$. Let $\widetilde{\mathbf{X}}_+ := \mathbf{U}_{\mathbf{A}_+}|\mathbf{S}_{\mathbf{A}_+}|^{1/2}$, $\widetilde{\mathbf{X}}_- := \mathbf{U}_{\mathbf{A}_-}|\mathbf{S}_{\mathbf{A}_-}|^{1/2}$ and $\widetilde{\mathbf{X}} := [\widetilde{\mathbf{X}}_+, \widetilde{\mathbf{X}}_-]$. To reiterate, $\mathbf{U}_{\mathbf{A}_+}$, $\mathbf{U}_{\mathbf{A}_-}$, $\widetilde{\mathbf{X}}_+$, and $\widetilde{\mathbf{X}}_-$ play the roles of the population counterparts of $\mathbf{U}_{\mathbf{P}_+}$, $\mathbf{U}_{\mathbf{P}_-}$, $\rho_n^{1/2}\mathbf{X}_+$, and $\rho_n^{1/2}\mathbf{X}_-$, respectively.

2.2. Entrywise eigenvector analysis framework

We now briefly discuss the entrywise eigenvector analysis framework for the signal-plus-noise matrix model (3). Unlike the case in Section 1.2, $\mathbf{X}_+$, $\mathbf{X}_-$, $\mathbf{U}_{\mathbf{P}_+}$, and $\mathbf{U}_{\mathbf{P}_-}$ are only identifiable up to an orthogonal matrix due to the potential multiplicity of the non-zero eigenvalues of $\mathbf{P}$. To find the suitable orthogonal alignment matrix, we follow the Procrustes analysis idea in [2,20,38,64]. Let $\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}$

yield the singular value decompositions $\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} = \mathbf{W}_{1+}\mathrm{diag}\{\sigma_1(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}),\ldots,\sigma_d(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})\}\mathbf{W}_{2+}^{\mathrm{T}}$ and $\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} = \mathbf{W}_{1-}\mathrm{diag}\{\sigma_1(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}),\ldots,\sigma_d(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-})\}\mathbf{W}_{2-}^{\mathrm{T}}$, where $\mathbf{W}_{1+},\mathbf{W}_{2+} \in \mathbb{O}(p)$ and $\mathbf{W}_{1-},\mathbf{W}_{2-} \in \mathbb{O}(q)$. Denote $\mathbf{W}_+^* = \mathrm{sgn}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+})$ and $\mathbf{W}_-^* = \mathrm{sgn}(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-})$ the matrix signs of $\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}$ defined as $\mathbf{W}_+^* = \mathbf{W}_{1+}\mathbf{W}_{2+}^{\mathrm{T}}$ and $\mathbf{W}_-^* = \mathbf{W}_{1-}\mathbf{W}_{2-}^{\mathrm{T}}$, respectively [3,38], and let $\mathbf{W}^* := \mathrm{diag}(\mathbf{W}_+^*,\mathbf{W}_-^*)$. Then the orthogonal alignment matrix between $\widetilde{\mathbf{X}}_+$ and $\mathbf{X}_+$ is selected as $\mathbf{W}_+ = (\mathbf{W}_+^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_+}$. Similarly, the orthogonal alignment matrix between $\widetilde{\mathbf{X}}_-$ and $\mathbf{X}_-$ is selected as $\mathbf{W}_- = (\mathbf{W}_-^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_-}$ It is believable that $\widetilde{\mathbf{X}}_+\mathbf{W}_+$, $\widetilde{\mathbf{X}}_-\mathbf{W}_-$, $\mathbf{U}_{\mathbf{A}_+}$, and $\mathbf{U}_{\mathbf{A}_-}$ are reasonable approximations to $\rho_n^{1/2}\mathbf{X}_+$, $\rho_n^{1/2}\mathbf{X}_-$, $\mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*$, and $\mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*$, respectively. For convenience, we denote $\mathbf{W} := \mathrm{diag}(\mathbf{W}_+,\mathbf{W}_-)$.

The keystone observation of the framework lies in the following decompositions:

$$\begin{aligned}\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{1/2}\mathbf{X}_+ &= \rho_n^{-1/2}\mathbf{E}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1} + \left\{\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{-1/2}\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}\right\},\\ \widetilde{\mathbf{X}}_-\mathbf{W}_- - \rho_n^{1/2}\mathbf{X}_- &= -\rho_n^{-1/2}\mathbf{E}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1} + \left\{\widetilde{\mathbf{X}}_-\mathbf{W}_- + \rho_n^{-1/2}\mathbf{A}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1}\right\},\end{aligned} \tag{4}$$

and

$$\begin{aligned}\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^* &= \mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* + (\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*),\\ \mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^* &= \mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^* + (\mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*).\end{aligned} \tag{5}$$

To see why equation (4) holds, we first observe that $\mathbf{P} = \rho_n\mathbf{X}_+\mathbf{X}_+^{\mathrm{T}} - \rho_n\mathbf{X}_-\mathbf{X}_-^{\mathrm{T}}$, so that $\mathbf{P}\mathbf{X}_+ = \rho_n\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)$ and $\mathbf{P}\mathbf{X}_- = -\rho_n\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)$, implying that

$$\begin{aligned}\rho_n^{-1/2}\mathbf{E}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1} &= \rho_n^{-1/2}\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1} - \rho_n^{1/2}\mathbf{X}_+,\\ -\rho_n^{-1/2}\mathbf{E}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1} &= -\rho_n^{-1/2}\mathbf{A}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1} - \rho_n^{1/2}\mathbf{X}_-\end{aligned}$$

since we assume that $\lambda_p(\mathbf{P}) > 0$ and $\lambda_{n-q+1}(\mathbf{P}) < 0$. Substituting $\rho_n^{1/2}\mathbf{X}_+$ and $\rho_n^{1/2}\mathbf{X}_-$ above to the left-hand side of (4) leads to the right-hand side of (4). The argument for (5) is similar. As observed in [3] and [21], viewing $(\widetilde{\mathbf{X}}_+,\widetilde{\mathbf{X}}_-,\mathbf{U}_{\mathbf{A}_+},\mathbf{U}_{\mathbf{A}_-})$ and $(\rho_n^{1/2}\mathbf{X}_+,\rho_n^{1/2}\mathbf{X}_-,\mathbf{U}_{\mathbf{P}_+},\mathbf{U}_{\mathbf{P}_-})$ as functionals of $\mathbf{A}$ and $\mathbf{P}$, we see that the first terms on the right-hand sides of (4) and (5) are linear approximations to $\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \rho_n^{1/2}\mathbf{X}_+$, $\widetilde{\mathbf{X}}_-\mathbf{W}_- - \rho_n^{1/2}\mathbf{X}_-$, $\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*$, and $\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*$, respectively, whereas the second terms are the higher-order remainders.

To shed some light on the entrywise limits of $\widetilde{\mathbf{X}}_+$ and $\widetilde{\mathbf{X}}_-$, we fix the vertex $i \in [n]$ and re-write (4) as

$$\begin{aligned}\mathbf{W}_+^{\mathrm{T}}(\widetilde{\mathbf{x}}_i)_+ - \rho_n^{1/2}(\mathbf{x}_i)_+ &= \sum_{j=1}^n \frac{[\mathbf{E}]_{ij}(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}(\mathbf{x}_j)_+}{\rho_n^{1/2}} + \left\{\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \frac{\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}}{\rho_n^{1/2}}\right\}^{\mathrm{T}}\mathbf{e}_i,\\ \mathbf{W}_-^{\mathrm{T}}(\widetilde{\mathbf{x}}_i)_- - \rho_n^{1/2}(\mathbf{x}_i)_- &= -\sum_{j=1}^n \frac{[\mathbf{E}]_{ij}(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1}(\mathbf{x}_j)_-}{\rho_n^{1/2}} + \left\{\widetilde{\mathbf{X}}_-\mathbf{W}_- + \frac{\mathbf{A}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1}}{\rho_n^{1/2}}\right\}^{\mathrm{T}}\mathbf{e}_i,\end{aligned}$$

where $(\widetilde{\mathbf{x}}_i)_+$, $(\widetilde{\mathbf{x}}_i)_-$, $(\mathbf{x}_i)_+$, and $(\mathbf{x}_i)_-$ denote the ith row of $\widetilde{\mathbf{X}}_+$, $\widetilde{\mathbf{X}}_-$, $\mathbf{X}_+$, and $\mathbf{X}_-$, respectively. An immediate observation is that the first terms above on the right-hand sides are sums of independent mean-zero random variables, which are quite accessible for the analysis. The non-trivial part is sharp controls of the second terms above. Using the facts that $\rho_n\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+ = \mathbf{W}_{\mathbf{X}_+}^{\mathrm{T}}|\mathbf{S}_{\mathbf{P}_+}|\mathbf{W}_{\mathbf{X}_+}$ and

$\rho_n \mathbf{X}_-^{\mathrm{T}}\mathbf{X}_- = \mathbf{W}_{\mathbf{X}_-}^{\mathrm{T}}|\mathbf{S}_{\mathbf{P}_-}|\mathbf{W}_{\mathbf{X}_-}$, we further write

$$\begin{aligned}
\widetilde{\mathbf{X}}_+\mathbf{W}_+ - \frac{\mathbf{A}\mathbf{X}_+(\mathbf{X}_+^{\mathrm{T}}\mathbf{X}_+)^{-1}}{\rho_n^{1/2}} &= \mathbf{U}_{\mathbf{A}_+}(\mathbf{W}_+^*|\mathbf{S}_{\mathbf{A}_+}|^{1/2} - |\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_+^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_+} \\
&\quad + (\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*)(\mathbf{W}_+^*)^{\mathrm{T}}|\mathbf{S}_{\mathbf{P}_+}|^{1/2}\mathbf{W}_{\mathbf{X}_+}, \\
\widetilde{\mathbf{X}}_-\mathbf{W}_- + \frac{\mathbf{A}\mathbf{X}_-(\mathbf{X}_-^{\mathrm{T}}\mathbf{X}_-)^{-1}}{\rho_n^{1/2}} &= \mathbf{U}_{\mathbf{A}_-}(\mathbf{W}_-^*|\mathbf{S}_{\mathbf{A}_-}|^{1/2} - |\mathbf{S}_{\mathbf{P}_-}|^{1/2}\mathbf{W}_-^*)^{\mathrm{T}}\mathbf{W}_{\mathbf{X}_-} \\
&\quad + (\mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*)(\mathbf{W}_-^*)^{\mathrm{T}}|\mathbf{S}_{\mathbf{P}_-}|^{1/2}\mathbf{W}_{\mathbf{X}_-}.
\end{aligned} \tag{6}$$

Because the analysis of the first terms on the right-hand sides of (6) are relatively easy (see, for example, Lemma 49 in [9]), we focus on the entrywise control of $\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$ and $\mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*$ in the second line in (6), which are also related to the entrywise limit of $\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{U}_{\mathbf{A}_-}$ through (5). Although there has been some recent progress on the uniform control of $\|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty}$ (see [3,21,50]), the numerical experiment in Section 1.2 suggests that the uniform error bound may not be sufficient for studying the entrywise limits of $\widetilde{\mathbf{X}}_+$, $\widetilde{\mathbf{X}}_-$, $\mathbf{U}_{\mathbf{A}_+}$, and $\mathbf{U}_{\mathbf{A}_-}$. This motivates us to develop a sharp control of $\|\mathbf{e}_i^{\mathrm{T}}(\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*)\|_2$ for each fixed $i \in [n]$.

3. Entrywise limit theorem for the eigenvectors

3.1. Main results

This section establishes the entrywise limit results for the eigenvectors $\mathbf{U}_{\mathbf{A}}$ and $\widetilde{\mathbf{X}}$. We first present several necessary assumptions for the generic signal-plus-noise matrix model (3).

Assumption 1. $\|\mathbf{X}\|_{2\to\infty}$ is upper bounded by a constant.

Assumption 2. $\rho_n \in (0,1]$, $\rho := \lim_{n\to\infty}\rho_n$ exists, and $n\rho_n = \Omega(\log n)$.

Assumption 3. The upper diagonal entries of $\mathbf{E}$, $([\mathbf{E}]_{ij} : 1 \le i \le j \le n)$, are independent mean-zero random variables; There exists mean-zero random variables $([\mathbf{E}_1]_{ij}, [\mathbf{E}_2]_{ij} : 1 \le i \le j \le n)$, such that $[\mathbf{E}]_{ij} = [\mathbf{E}_1]_{ij} + [\mathbf{E}_2]_{ij}$, and they satisfy the following conditions:

(i) There exists constants $B, \sigma^2 > 0$ independent of n such that $\max_{i,j\in[n]}|[\mathbf{E}_1]_{ij}| \le B$ with probability one and $\max_{i,j\in[n]}\mathrm{var}([\mathbf{E}_1]_{ij}) \le \sigma^2\rho_n$.

(ii) The random variables $([\mathbf{E}_2]_{ij} : 1 \le i \le j \le n)$ are uniformly sub-Gaussian in the following sense: $\max_{i,j\in[n]}\|[\mathbf{E}_2]_{ij}\|_{\psi_2} \le \sigma\rho_n^{1/2}$ for some constant $\sigma > 0$ independent of n, where $\|\cdot\|_{\psi_2}$ is the sub-Gaussian norm of a random variable (see, for example, [48,76]).

Assumption 4. There exist absolute constants $c_0 > 0$, $\xi \ge 1$, and a non-decreasing function $\varphi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ with $\varphi(0) = 0$, $\varphi(x)/x$ non-increasing in $\mathbb{R}_+$, such that for all $i, m \in [n]$ and any deterministic $n \times d$ matrix $\mathbf{V}$, with probability at least $1 - c_0 n^{-(1+\xi)}$,

$$\max\{\|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}^{(m)}\mathbf{V}\|_2, \|\mathbf{e}_i^{\mathrm{T}}\mathbf{E}\mathbf{V}\|_2\} \le n\rho_n\lambda_d(\mathbf{\Delta}_n)\|\mathbf{V}\|_{2\to\infty}\varphi\left(\frac{\|\mathbf{V}\|_{\mathrm{F}}}{\sqrt{n}\|\mathbf{V}\|_{2\to\infty}}\right),$$

where $\mathbf{E}^{(m)}$ is obtained by replacing the mth row and mth column of $\mathbf{E}$ by zeros.

Assumption 5. There exist absolute constants $K, c_0 > 0$, $\zeta \geq 1$, such that $\|\mathbf{E}\|_2 \leq K(n\rho_n)^{1/2}$ with probability at least $1 - c_0 n^{-\zeta}$ and $32\kappa(\mathbf{\Delta}_n)\max\{\gamma, \varphi(\gamma)\} \leq 1$, where $\varphi(\cdot)$ is the function in Assumption 4 and $\gamma := \max\{3K, \|\mathbf{X}\|_{2\to\infty}^2\}/\{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)\} \to 0$.

Several remarks regarding Assumptions 1-5 are in order. Assumption 1 is related to the notion of bounded coherence in random matrix theory and matrix recovery [17,18]. Indeed, observe that $\|\mathbf{U_P}\|_{2\to\infty} \leq \|\mathbf{X}\|_{2\to\infty}/\sqrt{n\lambda_d(\mathbf{\Delta}_n)}$. Therefore, Assumption 1 implies the bounded coherence of $\mathbf{U_P}$ (i.e., $\|\mathbf{U_P}\|_{2\to\infty} \leq C_\mu\sqrt{d/n}$ for some constant $C_\mu \geq 1$) as long as $d\lambda_d(\mathbf{\Delta}_n) = \Omega(1)$, which is a mild condition. Assumption 2 requires that $n\rho_n = \Omega(\log n)$. In the context of the two-block stochastic block model illustrated in Section 1.2, this amounts to requiring that the average graph expected degree is $\Omega(\log n)$. Assumption 3 is a general requirement for the tail of the distributions of the noise $\mathbf{E}$. It includes a variety of popular random matrix models such as random dot product graphs, the low-rank matrix denoising model, and the matrix completion model. Assumption 4 is motivated by the row-wise concentration assumption in [3]. The row-wise concentration behavior of $\mathbf{E}$ is characterized by a function $\varphi(\cdot)$ that depends on the distributions of $\mathbf{E}$ fundamentally. Assumption 5 is a standard assumption on the spectral concentration of the noise matrix $\mathbf{E}$ and is satisfied under the binary random graph model by [49] and the matrix completion model by [46].

Theorem 3.1 below is the main result of this section. It asserts that when $n\rho_n = \Omega(\log n)$, the distributions of the rows of $\widehat{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}$ and $\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*$ are approximately Gaussians.

Theorem 3.1. *Suppose Assumptions 1-5 hold. For each $i \in [n]$, let*

$$\mathbf{\Sigma}_{ni} = \mathbf{I}_{p,q}\mathbf{\Delta}_n^{-1}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}([\mathbf{E}]_{ij}^2)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\mathbf{\Delta}_n^{-1}\mathbf{I}_{p,q},$$

$$\mathbf{\Gamma}_{ni} = \mathbf{I}_{p,q}\mathbf{\Delta}_n^{-3/2}\left\{\frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}([\mathbf{E}]_{ij}^2)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\mathbf{\Delta}_n^{-3/2}\mathbf{I}_{p,q}$$

be invertible. Define $\chi = \varphi(1) + (\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)/\lambda_d(\mathbf{\Delta}_n)$. Then for each fixed index $i \in [n]$ and for any sufficiently large n,

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\mathbf{\Sigma}_{ni}^{-1/2}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) \in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right| \\
&\lesssim_\sigma \frac{d^{1/2}\chi\|\mathbf{\Sigma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}^2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\} \qquad (7)\\
&\quad + \frac{d^{1/2}\|\mathbf{\Sigma}_n^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)}\sum_{j=1}^n \mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\left\{\frac{1}{n\rho_n}\sum_{l=1}^n \mathbb{E}([\mathbf{E}]_{il}^2)\mathbf{x}_l\mathbf{x}_l^{\mathrm{T}}\right\}^{-1}\mathbf{x}_j,
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{n\rho_n^{1/2}\mathbf{\Gamma}_{ni}^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U_A}]_{i*}-[\mathbf{U_P}]_{i*})\in A\right\}-\mathbb{P}(\mathbf{z}\in A)\right| \\
&\lesssim_\sigma \frac{d^{1/2}\chi\|\mathbf{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\rho_n\right\} \\
&\quad+\frac{d^{1/2}\|\mathbf{\Gamma}_{ni}^{-1/2}\|_2\|\mathbf{X}\|_{2\to\infty}}{(n\rho_n)^{3/2}\lambda_d(\mathbf{\Delta}_n)^{3/2}}\sum_{j=1}^n\mathbb{E}|[\mathbf{E}]_{ij}|^3\mathbf{x}_j^{\mathrm{T}}\left\{\frac{1}{n\rho_n}\sum_{l=1}^n\mathbb{E}([\mathbf{E}]_{il}^2)\mathbf{x}_l\mathbf{x}_l^{\mathrm{T}}\right\}^{-1}\mathbf{x}_j,
\end{aligned}\tag{8}$$

where $\mathcal{A}$ is the collection of all convex measurable sets in $\mathbb{R}^d$ and $\mathbf{z}\sim\mathrm{N}_d(\mathbf{0}_d,\mathbf{I}_d)$.

Remark 1 (Generality of Theorem 3.1). Theorem 3.1 is stated in terms of Berry-Esseen type bounds for $\sqrt{n}\mathbf{\Sigma}_{ni}^{-1/2}\{\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i\}$ and $\sqrt{n}\mathbf{\Gamma}_{ni}^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U_A}]_{i*}-[\mathbf{U_P}]_{i*})$. The upper bounds only depend on $n\rho_n$, the rank of $\mathbf{P}$, the eigenvalues of $\mathbf{\Delta}_n$, a constant depending on σ, $\|\mathbf{X}\|_{2\to\infty}$, and the third absolute moments of $[\mathbf{E}]_{ij}$'s. Compared to the limit theorems in [21,72,82], Theorem 3.1 allows the rank d and the eigenvalues of $\mathbf{P}$ to vary with the number of vertices n. Consequently, as long as the right-hand sides of (7) and (8) converge to 0 as $n\to\infty$, the asymptotic shapes of the distributions of $\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i-\rho_n^{1/2}\mathbf{x}_i$ and $\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U_A}]_{i*}-[\mathbf{U_P}]_{i*})$ can be approximated by multivariate Gaussians.

The key to the proof of Theorem 3.1 is Theorem 3.2 below. It provides the second-order entrywise perturbation bounds for the eigenvectors $\mathbf{U_A}$ and $\widetilde{\mathbf{X}}$.

Theorem 3.2. *Suppose the conditions of Theorem 3.1 hold. Then there exists an absolute constant $c_0>0$, such that for each fixed $m\in[n]$, for all $t\geq 1$, $t\lesssim n\rho_n$, for sufficiently large n, with probability at least $1-c_0n^{-\zeta\wedge\xi}-c_0de^{-t}$,*

$$\begin{aligned}
&\|\mathbf{e}_m^{\mathrm{T}}(\mathbf{U_A}-\mathbf{AU_PS_P}^{-1}\mathbf{W}^*)\|_2\lesssim_\sigma\frac{\chi\|\mathbf{U_P}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}, \\
&\|\mathbf{e}_i^{\mathrm{T}}\{\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{-1/2}\mathbf{AX}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{I}_{p,q}\}\|_2 \\
&\quad\lesssim_\sigma\frac{\chi\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)t^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},t\right\}.
\end{aligned}$$

Furthermore, if $\zeta\wedge\xi>1$, then for sufficiently large n, with probability at least $1-c_0n^{-\zeta\wedge\xi}$,

$$\begin{aligned}
&\|\mathbf{U_A}-\mathbf{AU_PS_P}^{-1}\mathbf{W}^*\|_{2\to\infty}\lesssim_\sigma\frac{\chi\|\mathbf{U_P}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)},\frac{1}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\right\}, \\
&\|\widetilde{\mathbf{X}}\mathbf{W}-\rho_n^{-1/2}\mathbf{AX}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{I}_{p,q}\|_{2\to\infty} \\
&\quad\lesssim_\sigma\frac{\chi\|\mathbf{X}\|_{2\to\infty}\|\mathbf{U_P}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2\vee 1)(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2},\frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2},\log n\right\}.
\end{aligned}$$

3.2. Proof sketch for Theorem 3.2

In this section, we discuss the basic idea of the proof of Theorem 3.2. We begin with a warm-up matrix decomposition motivated by [3] and [21]. Denote $\mathbf{E} = \mathbf{A} - \mathbf{P}$. Recall that $\mathbf{U}_{\mathbf{P}_+} = \mathbf{P}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}$ and $\mathbf{U}_{\mathbf{P}_-} = \mathbf{P}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}$ by the definitions of $(\mathbf{U}_{\mathbf{P}_+}, \mathbf{S}_{\mathbf{P}_+})$ and $(\mathbf{U}_{\mathbf{P}_-}, \mathbf{S}_{\mathbf{P}_-})$. This leads to the following observation

$$\begin{aligned}\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* &= -\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* + (\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*),\\ \mathbf{U}_{\mathbf{A}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^* &= -\mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^* + (\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)\end{aligned} \tag{9}$$

because $\mathbf{U}_{\mathbf{P}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1} = (\mathbf{P} - \mathbf{A})\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}$ and $\mathbf{U}_{\mathbf{P}_-} - \mathbf{A}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1} = (\mathbf{P} - \mathbf{A})\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}$. For the second term on the right-hand side of (9), we recall that $\mathbf{U}_{\mathbf{A}_+}$ ($\mathbf{U}_{\mathbf{A}_-}$, resp.) is the eigenvector matrix of $\mathbf{A}$ corresponding to the eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$ ($\lambda_{n-q+1}(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$, resp.). Therefore,

$$\begin{aligned}\mathbf{U}_{\mathbf{A}_+} &= \mathbf{A}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} = \mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} + \mathbf{P}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1},\\ \mathbf{U}_{\mathbf{A}_-} &= \mathbf{A}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} = \mathbf{E}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} + \mathbf{P}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}.\end{aligned} \tag{10}$$

We first focus on the second terms $\mathbf{P}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}$ and $\mathbf{P}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}$ on the right-hand sides of (10) above. By the spectral decomposition $\mathbf{P} = \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}} + \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}$, we can write

$$\begin{aligned}\mathbf{P}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} &= \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} + \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1},\\ \mathbf{P}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} &= \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} + \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}.\end{aligned}$$

Recall that $\mathbf{W}_+^*$ and $\mathbf{W}_-^*$ are the matrix signs of $\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}$, suggesting that $\mathbf{W}_+^* \approx \mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{W}_-^* \approx \mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}$. It is then conceivable that $\mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^* \approx \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} = \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}$ and $\mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^* \approx \mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} = \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}$. This motivates us to write $\mathbf{P}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}$ and $\mathbf{P}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}$ as

$$\begin{aligned}\mathbf{P}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} &= \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} + \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}\\ &= \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}) + \mathbf{U}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} + \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1},\\ \mathbf{P}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} &= \mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}(\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} - \mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}) + \mathbf{U}_{\mathbf{P}_-}\mathbf{U}_{\mathbf{P}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-} + \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}.\end{aligned} \tag{11}$$

We next turn our attention to the first terms $\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}$ and $\mathbf{E}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1}$ on the right-hand sides of (10). Intuitively, these two terms should be closed to $\mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$, which leads to the following decompositions

$$\begin{aligned}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} &= \mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* + (\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*),\\ \mathbf{E}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} &= \mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^* + (\mathbf{E}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} - \mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*).\end{aligned} \tag{12}$$

Because $\mathbf{A}$ concentrates around $\mathbf{P}$ in spectral norm and $\mathbf{U}_{\mathbf{A}}, \mathbf{U}_{\mathbf{P}}$ are their eigenvector matrices, the matrix perturbation theory suggests that $\mathbf{U}_{\mathbf{A}_+} \approx \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*$ and $\mathbf{U}_{\mathbf{A}_-} \approx \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*$. Hence, we can write the second terms in (12) above as

$$\begin{aligned}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{E}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* &= \mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1} + \mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*),\\ \mathbf{E}\mathbf{U}_{\mathbf{A}_-}\mathbf{S}_{\mathbf{A}_-}^{-1} - \mathbf{E}\mathbf{U}_{\mathbf{P}_-}\mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^* &= \mathbf{E}(\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)\mathbf{S}_{\mathbf{A}_-}^{-1} + \mathbf{E}\mathbf{U}_{\mathbf{P}_-}(\mathbf{W}_-^*\mathbf{S}_{\mathbf{A}_-}^{-1} - \mathbf{S}_{\mathbf{P}_-}^{-1}\mathbf{W}_-^*).\end{aligned} \tag{13}$$

We now combine equations (9), (10), (11), (12), and (13) to obtain the following decomposition of $\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*$:

$$\begin{aligned}
\mathbf{U}_{\mathbf{A}_+} - \mathbf{A}\mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^* &= \mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\mathbf{S}_{\mathbf{A}_+}^{-1} && (14)\\
&\quad + \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+}) && (15)\\
&\quad + \mathbf{U}_{\mathbf{P}_+}(\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{A}_+} - \mathbf{W}_+^*) && (16)\\
&\quad + \mathbf{E}\mathbf{U}_{\mathbf{P}_+}(\mathbf{W}_+^*\mathbf{S}_{\mathbf{A}_+}^{-1} - \mathbf{S}_{\mathbf{P}_+}^{-1}\mathbf{W}_+^*) && (17)\\
&\quad + (\mathbf{P} - \mathbf{U}_{\mathbf{P}_+}\mathbf{S}_{\mathbf{P}_+}\mathbf{U}_{\mathbf{P}_+}^{\mathrm{T}})\mathbf{U}_{\mathbf{A}_+}\mathbf{S}_{\mathbf{A}_+}^{-1}. && (18)
\end{aligned}$$

By the same reasoning, the above decomposition also holds when $(\mathbf{U}_{\mathbf{A}_+}, \mathbf{S}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+}, \mathbf{S}_{\mathbf{P}_+}, \mathbf{W}_+^*)$ are replaced by $(\mathbf{U}_{\mathbf{A}_-}, \mathbf{S}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-}, \mathbf{S}_{\mathbf{P}_-}, \mathbf{W}_-^*)$. Among the five terms above, lines (15), (16), (17), and (18) are relatively easy to control using classical matrix perturbation tools and the concentration of $\|\mathbf{E}\|_2$ due to [49]. The formal concentration bounds of these remainders are given in the Supplementary Material [80]. The challenging part is a delicate analysis of the row-wise behavior of $\mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)$ and $\mathbf{E}(\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)$, which we sketch below. We borrow the decoupling strategy and a "leave-one-out" analysis that appeared in [3,12,42,50,85]. Consider the following collection of auxiliary matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(n)}$. For each row index $m \in [n]$, the matrix $\mathbf{A}^{(m)} = [A_{ij}]_{n\times n}$ is a function of $\mathbf{A}$ defined by

$$\mathbf{A}^{(m)} = \begin{cases} A_{ij}, & \text{if } i \neq m \text{ and } j \neq m, \\ \mathbb{E}A_{ij}, & \text{if } i = m \text{ or } j = m. \end{cases} \tag{19}$$

Namely, the matrix $\mathbf{A}^{(m)}$ is constructed by replacing the mth row and mth column of $\mathbf{A}$ by their expected values. Now let $\mathbf{U}_{\mathbf{A}_+}^{(m)}$ and $\mathbf{U}_{\mathbf{A}_-}^{(m)}$ be the leading eigenvector matrices of $\mathbf{A}^{(m)}$ ($\mathbf{U}_{\mathbf{A}_+} \in \mathbb{O}(n,p)$ and $\mathbf{U}_{\mathbf{A}_-} \in \mathbb{O}(n,q)$) such that $\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{S}_{\mathbf{A}_+}^{(m)} = \mathbf{A}^{(m)}\mathbf{U}_{\mathbf{A}_+}^{(m)}$ and $\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{S}_{\mathbf{A}_-}^{(m)} = \mathbf{A}^{(m)}\mathbf{U}_{\mathbf{A}_-}^{(m)}$, where

$$\mathbf{S}_{\mathbf{A}_+}^{(m)} = \mathrm{diag}\{\lambda_1(\mathbf{A}^{(m)}), \ldots, \lambda_d(\mathbf{A}^{(m)})\} \quad \text{and} \quad \mathbf{S}_{\mathbf{A}_-}^{(m)} = \mathrm{diag}\{\lambda_{n-q+1}(\mathbf{A}^{(m)}), \ldots, \lambda_n(\mathbf{A}^{(m)})\}.$$

Denote $\mathbf{H}_+ = \mathbf{U}_{\mathbf{A}_+}^{\mathrm{T}}\mathbf{U}_{\mathbf{P}_+}$, $\mathbf{H}_- = \mathbf{U}_{\mathbf{A}_-}^{\mathrm{T}}\mathbf{U}_{\mathbf{P}_-}$, $\mathbf{H}_+^{(m)} = (\mathbf{U}_{\mathbf{A}_+}^{(m)})^{\mathrm{T}}\mathbf{U}_{\mathbf{P}_+}$, and $\mathbf{H}_-^{(m)} = (\mathbf{U}_{\mathbf{A}_-}^{(m)})^{\mathrm{T}}\mathbf{U}_{\mathbf{P}_-}$. The smartness of introducing $\mathbf{A}^{(m)}$ lies in the striking fact that $\mathbf{e}_m^{\mathrm{T}}\mathbf{E}$ and $\mathbf{A}^{(m)}$ are independent. With this in mind, we can focus on the mth row of $\mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)$ and $\mathbf{E}(\mathbf{U}_{\mathbf{A}_-} - \mathbf{U}_{\mathbf{P}_-}\mathbf{W}_-^*)$ by inserting $\mathbf{U}_{\mathbf{A}}^{(m)}$ as follows:

$$\begin{aligned}
\|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+} - \mathbf{U}_{\mathbf{P}_+}\mathbf{W}_+^*)\|_2 &\le \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}\mathbf{U}_{\mathbf{A}_+}\{\mathrm{sgn}(\mathbf{H}_+) - \mathbf{H}_+\}\|_2 \\
&\quad + \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+ - \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)})\|_2 \\
&\quad + \|\mathbf{e}_m^{\mathrm{T}}\mathbf{E}(\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+})\|_2.
\end{aligned} \tag{20}$$

Note that by the same reasoning, the above inequality also holds with $(\mathbf{U}_{\mathbf{A}_+}, \mathbf{U}_{\mathbf{P}_+}, \mathbf{U}_{\mathbf{A}_+}^{(m)}, \mathbf{H}_+, \mathbf{H}_+^{(m)}, \mathbf{W}_+^*)$ replaced by $(\mathbf{U}_{\mathbf{A}_-}, \mathbf{U}_{\mathbf{P}_-}, \mathbf{U}_{\mathbf{A}_-}^{(m)}, \mathbf{H}_-, \mathbf{H}_-^{(m)}, \mathbf{W}_-^*)$. Here, we have used the facts that $\mathbf{W}_+^* \in \mathbb{O}(p)$ and $\mathbf{W}_-^* \in \mathbb{O}(q)$. Since $\mathbf{e}_m^{\mathrm{T}}\mathbf{E}$ and $\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+}$ (or $\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)} - \mathbf{U}_{\mathbf{P}_-}$) are independent, we can apply Bernstein's or Hoeffding's inequality to the third term above. The success of this decoupling strategy critically depends on the following sharp concentration bounds on $\|\mathbf{U}_{\mathbf{A}_+}\mathbf{H}_+ - \mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)}\|_2$, $\|\mathbf{U}_{\mathbf{A}_-}\mathbf{H}_- - \mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)}\|_2$, $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+ - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_- - \mathbf{U}_{\mathbf{P}_-}\|_{2\to\infty}$, $\|\mathbf{U}_{\mathbf{A}_+}^{(m)}\mathbf{H}_+^{(m)} - \mathbf{U}_{\mathbf{P}_+}\|_{\mathrm{F}}$, and $\|\mathbf{U}_{\mathbf{A}_-}^{(m)}\mathbf{H}_-^{(m)} - \mathbf{U}_{\mathbf{P}_-}\|_{\mathrm{F}}$.

Lemma 3.3. *Suppose Assumptions 1-5 hold. Denote* $\mathbf{\Delta}_n = (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$. *Let* $m \in [n]$ *be any fixed row index and* $\mathbf{A}^{(m)}$, $\mathbf{U}^{(m)}_{\mathbf{A}_+}$, $\mathbf{U}^{(m)}_{\mathbf{A}_-}$, $\mathbf{H}^{(m)}_{+}$, *and* $\mathbf{H}^{(m)}_{-}$ *be defined as above. Then there exists an absolute constant* $c_0 > 0$, *such that for sufficiently large* n, *with probability at least* $1 - c_0 n^{-\zeta \wedge \xi}$,

$$\|\mathbf{U}^{(m)}_{\mathbf{A}_+}\mathbf{H}^{(m)}_{+} - \mathbf{U}_{\mathbf{P}_+}\|_2 \lesssim \frac{1}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

$$\|\mathbf{U}^{(m)}_{\mathbf{A}_+}\|_{2\to\infty} \lesssim \{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\|\mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty},$$

$$\|\mathbf{U}^{(m)}_{\mathbf{A}_+}\mathrm{sgn}(\mathbf{H}^{(m)}_{+}) - \mathbf{U}_{\mathbf{P}_+}\|_{2\to\infty} \lesssim [\kappa(\mathbf{\Delta}_n)\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}\{\gamma + \varphi(\gamma)\} + \kappa(\mathbf{\Delta}_n) + \varphi(1)]\,\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty},$$

and for all $t \geq 1$ *and* $t \lesssim n\rho_n$, *with probability at least* $1 - c_0 n^{-\zeta\wedge\xi} - c_0 d e^{-t}$,

$$\|\mathbf{U}^{(m)}_{\mathbf{A}_+}\mathbf{H}^{(m)}_{+} - \mathbf{U}_{\mathbf{A}_+}\mathbf{H}_{+}\|_2 \lesssim_{\sigma} \frac{\{\kappa(\mathbf{\Delta}_n) + \varphi(1)\}t^{1/2}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}.$$

We remark that the above concentration bounds hold with $(\mathbf{U}_{\mathbf{P}_+}, \mathbf{U}^{(m)}_{\mathbf{A}_+}, \mathbf{H}_{+}, \mathbf{H}^{(m)}_{+}, \mathbf{W}^{*}_{+})$ *replaced by* $(\mathbf{U}_{\mathbf{P}_-}, \mathbf{U}^{(m)}_{\mathbf{A}_-}, \mathbf{H}_{-}, \mathbf{H}^{(m)}_{-}, \mathbf{W}^{*}_{-})$.

3.3. Comparison with existing results

We first briefly compare Theorem 3.1 with some existing entrywise limit theorems for the eigenvectors of signal-plus-noise matrix models. For simplicity, we assume that the non-zero eigenvalues of $\mathbf{P}$ are positive. In [21], the authors established the asymptotic normality of $\mathbf{W}^{\mathrm{T}}_{\mathbf{X}}(\mathbf{W}^{*}[\mathbf{U}_{\mathbf{A}}]_{i*} - [\mathbf{U}_{\mathbf{P}}]_{i*})$ when $n\rho_n = \omega((\log n)^{4\xi})$ for some constant $\xi > 1$, provided that d is fixed across all n and $\mathbf{\Delta}_n$ converges to some positive definite $\mathbf{\Delta}$. Later, the requirement for $n\rho_n$ is relaxed to $n\rho_n = \omega((\log n)^4)$ in [82] for the rows of the scaled eigenvectors in the context of random dot product graphs (see Section 4.2 for the formal definition). The same sparsity requirement for $n\rho_n$ was required in [72] when the rows of $\mathbf{X}$ are independent and identically distributed latent random vectors, and their limit result is stated as multivariate normal mixtures. In contrast, Theorem 3.1 only requires that $n\rho_n = \Omega(\log n)$ when d is fixed and $\lambda_d(\mathbf{\Delta}_n)$ is bounded away from 0.

We next provide several remarks regarding Theorem 3.2 and compare it with some results in the literature. Again, for simplicity, we assume that the non-zero eigenvalues of $\mathbf{P}$ are positive and $\lambda_d(\mathbf{\Delta}_n)$ stays bounded away from 0. Then the asymptotic normality of (8) holds only if

$$\|\mathbf{e}^{\mathrm{T}}_m(\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}^{-1}_{\mathbf{P}}\mathbf{W}^{*})\|_2 = o_{\mathbb{P}}\left(\frac{1}{n\rho_n^{1/2}}\right), \tag{21}$$

which can be obtained from Theorem 3.2 with $t = \log n\rho_n$. We argue that the concentration bound (21) is sharper than the recently developed concentration bounds for $\|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}^{-1}_{\mathbf{P}}\mathbf{W}^{*}\|_{2\to\infty}$ in [3,21,50]. In [21], the authors assumed that $n\rho_n = \omega((\log n)^{2\xi})$ for some $\xi > 1$ and showed that

$$\|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}^{-1}_{\mathbf{P}}\mathbf{W}^{*}\|_{2\to\infty} = O_{\mathbb{P}}\left\{\frac{1}{n\rho_n^{1/2}} \times \frac{(\log n)^{2\xi}}{(n\rho_n)^{1/2}}\right\}. \tag{22}$$

The bound (22) is not sufficient for (21) to occur unless $n\rho_n = \omega((\log n)^{4\xi})$. Under the most challenging regime that $n\rho_n = \Omega(\log n)$, in the context of random graph models, the authors of [3] and [50] have

established that

$$\|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}\log\log n}\right). \tag{23}$$

This bound leads to a sharp analysis of the community detection using the signs of the second leading eigenvector of $\mathbf{A}$ for a two-block stochastic block model but does not imply (21) either. The underlying reason is that these two-to-infinity norm error bounds are obtained using a union bound, leading to sub-optimal entrywise concentration bounds for $\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*$.

4. Applications

4.1. Symmetric noisy matrix completion

The matrix completion problem has been extensively explored in recent decades, and the literature review included here is by no means complete and exhaustive. It refers to a large class of random matrix problems where the observed data matrix contains partial observations, and the task of interest is to predict the missing entries. A canonical real-world application is the "Netflix problem" [14], where the data matrix consists of multiple users' ratings of multiple movies. The missingness is intrinsic to the nature of the problem because it is unlikely to have the users watch all movies available in the database. Predicting the missing entries is worthwhile because accurate predictions allow the system to make appropriate individual-wise recommendations to the users. Theoretical properties of the matrix completion model have also been well studied. For example, the theory of noiseless matrix completion has been explored in [18,19,38], whereas the extensions for more general noisy matrix completion problems have been developed in [17,23,41,46,47].

This subsection considers a special case of the noisy matrix completion problem where the data matrix is a symmetric random matrix with missing observations, also referred to as the symmetric noisy matrix completion (SNMC) model. It also appears in the context of network cross-validation by edge sampling [52]. We follow the definition from [3] and assume that $\mathbf{P}$ has rank d.

Definition 4.1. Let $\mathbf{P} = [p_{ij}]_{n\times n}$ be a rank-d $n \times n$ symmetric matrix. The symmetric noisy matrix completion model, denoted by SNMC($\mathbf{P}, \rho_n, \sigma$), is the distribution of a symmetric random matrix $\mathbf{A} = [A_{ij}]_{n\times n}$ given by $A_{ij} = (p_{ij} + \epsilon_{ij})I_{ij}/\rho_n$, where $(I_{ij}, \epsilon_{ij} : 1 \le i \le j \le n)$ are jointly independent, $I_{ij} \sim$ Bernoulli(ρ_n), $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$, and $A_{ij} = A_{ji}$ for all $i > j$.

We follow the same notations and definitions in Sections 2 and 3. For example, recall that $\mathbf{P}$ can be written as $\mathbf{P} = \rho_n\mathbf{X}_+\mathbf{X}_+^{\mathrm{T}} - \rho_n\mathbf{X}_-\mathbf{X}_-^{\mathrm{T}}$ for some $n\times p$ matrix $\mathbf{X}_+$ and $n\times q$ matrix $\mathbf{X}_-$, p, q are the numbers of positive and negative eigenvalues of $\mathbf{P}$, respectively, $\mathbf{X} = [\mathbf{X}_+, \mathbf{X}_-]$, and $\mathbf{\Delta}_n = (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Below, we leverage the entrywise eigenvector analysis results in Section 3 to establish the entrywise Berry-Esseen bounds for $\mathbf{P}$ in the context of symmetric noisy matrix completion problems. For simplicity, we assume that (p, q) are known. In the case where they are unknown, one can use the numbers of positive and negative eigenvalues among the d-largest eigenvalues of $\mathbf{A}$ in magnitude as estimates for them.

Theorem 4.2. *Suppose* $\mathbf{A} \sim$ SNMC($\mathbf{P}, \rho_n, \sigma$). *With the same notations and definitions in Sections 2 and 3, assume that* $n\rho_n \ge 6\log n$, $\sigma = \tau\rho_n^2$ *for some constant* $\tau > 0$, $\|\mathbf{X}\|_{2\to\infty}$ *is upper bounded by an absolute constant, and* $n\rho_n\lambda_d(\mathbf{\Delta}_n)^2 = \omega(\kappa(\mathbf{\Delta}_n)^2\log n)$. *Let*

$$\sigma_{nii}^2 = \frac{4}{n}\sum_{l=1}^{n}\left\{(1-\rho_n)(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2 + \tau^2\rho_n^2\right\}(\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_n^{-1}\mathbf{x}_l)^2$$

and

$$\sigma_{nij}^2 = \frac{1}{n}\sum_{l=1}^{n}\left\{(1-\rho_n)(\mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2 + \tau^2\rho_n^2\right\}\{\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_l + \mathbb{1}(l=j)\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_i\}^2$$
$$+ \frac{1}{n}\sum_{l\neq i}^{n}\left\{(1-\rho_n)(\mathbf{x}_j^{\mathrm{T}}\mathbf{I}_{p,q}\mathbf{x}_l)^2 + \tau^2\rho_n^2\right\}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Delta}_n^{-1}\mathbf{x}_l)^2$$

if $i \neq j$. For each $(i,j) \in [n]\times[n]$, define $\widetilde{p}_{ij} = \widetilde{\mathbf{x}}_i^{\mathrm{T}}\mathbf{I}_{p,q}\widetilde{\mathbf{x}}_j$. Then for each fixed index pair $(i,j) \in [n]\times[n]$ and for any sufficiently large n,

$$\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{\frac{n}{\rho_n}}\frac{(\widetilde{p}_{ij}-p_{ij})}{\sigma_{nij}}\in A\right\} - \mathbb{P}(z\in A)\right|$$
$$\lesssim_\tau \frac{\beta_n(\mathbf{X})\|\mathbf{X}\|_{2\to\infty}}{\sigma_{nij}\sqrt{n\rho_n}} + \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)\log n\rho_n}{\sigma_{nij}\lambda_d(\boldsymbol{\Delta}_n)^2\sqrt{n\rho_n}} + \frac{\beta_n(\mathbf{X})^2}{\sigma_{nij}(n\rho_n)^{3/2}},$$

where $\mathcal{A}$ is the collection of all convex measurable sets in $\mathbb{R}$, $z \sim \mathrm{N}(0,1)$, and

$$\beta_n(\mathbf{X}) := \frac{(\|\mathbf{X}\|_{2\to\infty}^6 \vee 1)}{\lambda_d(\boldsymbol{\Delta}_n)^{5/2}}\max\left\{\frac{(\|\mathbf{X}\|_{2\to\infty}^2 \vee 1)\sqrt{\log n\rho_n}}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \frac{\kappa(\boldsymbol{\Delta}_n)}{\lambda_d(\boldsymbol{\Delta}_n)^2}, \log n\rho_n\right\}.$$

We now compare Theorem 4.2 with the entrywise recovery results for the noisy matrix completion problem in [3]. Again, we assume that d is fixed and $\lambda_d(\boldsymbol{\Delta}_n), \lambda_1(\boldsymbol{\Delta}_n)$ are bounded away from 0 and ∞ for simplicity. Under the condition that $n\rho_n = \omega(\log n)$, the asymptotic normality in Theorem 4.2 implies that $\widetilde{p}_{ij} - p_{ij} = O_{\mathbb{P}}(\sqrt{\rho_n/n})$. In contrast, Theorem 3.4 in [3] implies that $\max_{i,j\in[n]}|\widetilde{p}_{ij} - p_{ij}| = O_{\mathbb{P}}\{\sqrt{(\rho_n \log n)/n}\}$ under the same conditions. Similar to the reasoning in Section 3.3, the above entrywise-maximum norm error bound does not imply the asymptotic normality for $\widetilde{p}_{ij} - p_{ij}$ established in Theorem 4.2. Therefore, Theorem 4.2 provides a sharper entrywise recovery result and a refined analysis compared to [3] for the symmetric noisy matrix completion model.

4.2. Eigenvectors of random dot product graphs

In recent years, statistical network analysis has attracted much attention and has gained substantial progress in theoretical foundations and methodological development. Network data are also pervasive in numerous application domains, including social networks [35,77,84], neuroscience [60,74], and computer networks [57,62]. In the statistical analyses of network data, spectral methods and eigenvector analysis of random adjacency matrices are of fundamental interest because the eigenvectors not only contain the underlying network latent structure but also provide gateways to various subsequent inference tasks, such as community detection [61,69], vertex classification [70,73], and nonparametric graph testing [71].

In this subsection, we focus on the random dot product graph model [84] and study the behavior of its eigenvectors. It is a class of random graphs in which each vertex is assigned a latent position vector encoding the vertex-wise information. The random dot product graph model is easy to interpret (especially in social networks) and rich enough to include a variety of popular network models, including stochastic block models [39] and their offspring [5,45,54]. Below, we first provide the formal definition of the random dot product graph model.

Definition 4.3. Consider a graph with n vertices that are labeled as $[n] = \{1,2,\ldots,n\}$. Let $\mathcal{X}$ be a subset of $\mathbb{R}^d$ such that $\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 \in [0,1]$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$, where $d \leq n$, and let $\rho_n \in (0,1]$ be a sparsity factor. Each vertex $i \in [n]$ is associated with a vector $\mathbf{x}_i \in \mathcal{X}$, referred to as the latent position for vetex i. We say that a symmetric random matrix $\mathbf{A} = [A_{ij}]_{n\times n} \in \{0,1\}^{n\times n}$ is the adjacency matrix of a random dot product graph with latent position matrix $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_n]^{\mathrm{T}}$ and sparsity factor ρ_n, denoted by $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$, if the random variables $A_{ij} \sim \mathrm{Bernoulli}(\rho_n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)$ independently for all $i,j \in [n]$, $i \leq j$, and $A_{ij} = A_{ji}$ for all $i > j$.

The sparsity factor ρ_n in a random dot product graph model $\mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ fundamentally controls the graph average expected degree through $n\rho_n$ as a function of the number of vertices, provided that $\sum_{i,j}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j = \Omega(n^2)$. When $\rho_n \equiv 1$, the resulting graph is dense, and the average expected degree scales as $\Omega(n)$. The more interesting scenario happens when $\rho_n \to 0$ as $n \to \infty$, which gives rise to a sparse random graph whose average expected degree is a vanishing proportion of the number of vertices. A fast decaying ρ_n corresponds to a challenging weak signal regime, which is one of the focuses of this subsection.

We now present the Berry-Esseen theorem for the rows of the leading eigenvectors for random dot product graphs. The scaled eigenvector matrix $\widetilde{\mathbf{X}}$ is also referred to as the adjacency spectral embedding of $\mathbf{A}$ into $\mathbb{R}^d$ [69].

Theorem 4.4. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *with* $n\rho_n \gtrsim \log n$. *Denote* $\mathbf{\Delta}_n = (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$ *and suppose there exists a constant* $\delta > 0$ *such that* $\min_{i\in[n]}(1/n)\sum_{j=1}^n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \geq \delta$. *For each* $i \in [n]$, *let*

$$\mathbf{\Sigma}_n(\mathbf{x}_i) = \mathbf{\Delta}_n^{-1}\left\{\frac{1}{n}\sum_{j=1}^n \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\right\}\mathbf{\Delta}^{-1}, \quad \mathbf{\Gamma}_n(\mathbf{x}_i) = \mathbf{\Delta}_n^{-1/2}\mathbf{\Sigma}_n(\mathbf{x}_i)\mathbf{\Delta}_n^{-1/2}.$$

If $\mathbf{\Sigma}_n(\mathbf{x}_i)$ *and* $\mathbf{\Gamma}_n(\mathbf{x}_i)$ *are invertible and* $\kappa(\mathbf{\Delta}_n)/\lambda_d(\mathbf{\Delta}_n)\{(n\rho_n)^{-1/2} \vee \log(n\rho_n\lambda_d(\mathbf{\Delta}_n)^2)^{-1}\} \to 0$, *then for each fixed index* $i \in [n]$ *and for any sufficiently large* n,

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\mathbf{\Sigma}_n(\mathbf{x}_i)^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) \in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right| \\
&\quad\lesssim \frac{d^{1/2}\|\mathbf{\Sigma}_n(\mathbf{x}_i)^{-1/2}\|_2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\}, \\
&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{n\rho_n^{1/2}\mathbf{\Gamma}_n(\mathbf{x}_i)^{-1/2}\mathbf{W}_{\mathbf{X}}^{\mathrm{T}}(\mathbf{W}^*[\mathbf{U}_{\mathbf{A}}]_{i*} - [\mathbf{U}_{\mathbf{P}}]_{i*}) \in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right| \\
&\quad\lesssim \frac{d^{1/2}\|\mathbf{\Gamma}_n(\mathbf{x}_i)^{-1/2}\|_2}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\log n\rho_n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\rho_n\right\},
\end{aligned}$$

where $\mathcal{A}$ *is the collection of all convex measurable sets in* $\mathbb{R}^d$ *and* $\mathbf{z} \sim \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d)$.

Compared to the eigenvector limit theorems for random dot product graphs in [10,21,72,82], Theorem 4.4 requires a much weaker sparsity condition on ρ_n. Specifically, the authors of [10] explored the entrywise eigenvector limits by assuming that $\rho_n \equiv 1$ and the minimal sparsity condition in [21,72,82] is $n\rho_n = \omega((\log n)^4)$. In contrast, in Theorem 4.4, we only require that $n\rho_n = \Omega(\log n)$ if the eigenvalues of $\mathbf{\Delta}_n$ are bounded away from 0 and ∞. As mentioned in Section 1.1, our sparsity assumption $n\rho_n = \Omega(\log n)$ is minimal because $\mathbf{A}$ no longer concentrates around $\mathbf{P}$ in spectral norm when $n\rho_n = o(\log n)$ [72].

Next, we establish the two-to-infinity norm perturbation bounds for the eigenvectors of random dot product graphs in Corollary 4.1 below.

Corollary 4.1. *Suppose* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and the conditions of Theorem 3.1 hold. Denote* $\mathbf{\Delta}_n = (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$. *Then there exists an absolute constant* $c_0 > 0$, *such that given any fixed* $c > 0$,

$$\|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty} \lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{n\rho_n\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)}, \frac{1}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\right\},$$

$$\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty} \lesssim_c \|\mathbf{U}_{\mathbf{A}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{W}^*\|_{2\to\infty} + \frac{(\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)},$$

$$\left\|\widetilde{\mathbf{X}}\mathbf{W} - \frac{\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}}{\rho_n^{1/2}}\right\|_{2\to\infty} \lesssim_c \frac{\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{(n\rho_n)^{1/2}\lambda_d(\mathbf{\Delta}_n)^2}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\right\},$$

$$\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{2\to\infty} \lesssim_c \left\|\widetilde{\mathbf{X}}\mathbf{W} - \frac{\mathbf{A}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}}{\rho_n^{1/2}}\right\|_{2\to\infty} + \frac{(\log n)^{1/2}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}}{\lambda_d(\mathbf{\Delta}_n)^{1/2}}$$

with probability at least $1 - c_0 n^{-c}$ *for sufficiently large* n.

Corollary 4.1 provides a sharp concentration bound for $\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty}$ compared to some recently obtained results. Assuming that $\lambda_d(\mathbf{\Delta}_n)$ is bounded away from 0 for simplicity, we see that Corollary 4.1 leads to $\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty} \lesssim_{\lambda_d(\mathbf{\Delta}_n)} \sqrt{(\log n)/(n\rho_n)}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}$ with high probability. This also coincides with the concentration bound obtained in [50]. In [21] and [56], it has been shown that $\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty} \lesssim_{\lambda_d(\mathbf{\Delta}_n)} \sqrt{(\log n)^{2\xi}/(n\rho_n)}\|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}$ with high probability under a stronger assumption that $n\rho_n = \omega((\log n)^{2\xi})$ for some $\xi > 1$. Our result is tighter than the above large probability bound by a $(\log n)^{\xi-1/2}$ factor. In [3], the authors proved that $\|\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}}\mathbf{W}^*\|_{2\to\infty} \lesssim_{\lambda_d(\mathbf{\Delta}_n)} \|\mathbf{U}_{\mathbf{P}}\|_{2\to\infty}$ with high probability, which coincides with Corollary 4.1 when $n\rho_n \asymp \log n$ but deteriorates when $n\rho_n = \omega(\log n)$.

We also remark that the concentration bound on $\|\widetilde{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}\|_{2\to\infty}$ plays a fundamental role in establishing the entrywise limit theorem for the one-step estimator in Section 4.3 next.

4.3. One-step estimator for random dot product graphs

We continue the investigation of the entrywise estimation of the eigenvectors of random dot product graphs. As observed in [82], the adjacency spectral embedding (the scaled eigenvector matrix $\widetilde{\mathbf{X}}$) can be further refined by a one-step procedure implemented in the following vertex-wise fashion.

Definition 4.5. Let $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ and $\widetilde{\mathbf{X}} = [\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_n]^{\mathrm{T}} \in \mathbb{R}^{n\times d}$ be the adjacency spectral embedding of $\mathbf{A}$ into $\mathbb{R}^d$. Then the one-step refinement of $\widetilde{\mathbf{X}}$ is the $n \times d$ matrix $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_n]^{\mathrm{T}}$, whose ith row $\widehat{\mathbf{x}}_i$ is given by

$$\widehat{\mathbf{x}}_i = \widetilde{\mathbf{x}}_i + \left\{\sum_{j=1}^{n} \frac{\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}}{\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1 - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}\right\}^{-1}\left\{\sum_{j=1}^{n} \frac{(A_{ij} - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)\widetilde{\mathbf{x}}_j}{\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1 - \widetilde{\mathbf{x}}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)}\right\}, \quad i = 1, 2, \ldots, n. \tag{24}$$

The one-step refinement above is motivated by the one-step estimator in the classical M-estimation theory for parametric models (see, for example, Section 5.7 in [75]). In short, under mild conditions,

given a root-n consistent initial estimator, the one-step refinement achieves the information lower bound in a parametric model asymptotically. The same idea also applies to the random dot product graph model. Denote $\ell_{\mathbf{A}}(\mathbf{X})$ the log-likelihood function of RDPG($\rho_n^{1/2}\mathbf{X}$). Then a straightforward computation shows that the score function and the Fisher information matrix with regard to $\mathbf{x}_i$ are

$$\nabla_{\mathbf{x}_i}\ell_{\mathbf{A}}(\mathbf{X}) = \sum_{j=1}^{n}\frac{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)} \quad \text{and} \quad \mathcal{I}_i(\mathbf{X}) = \rho_n\sum_{j=1}^{n}\frac{\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}.$$

Given the adjacency spectral embedding $\widetilde{\mathbf{X}}$ as an initial guess, the right-hand side of (24) is precisely the updating rule of the Newton-Raphson algorithm for $\rho_n^{1/2}\mathbf{x}_i$ initialized at $\widetilde{\mathbf{x}}_i$, with the Hessian replaced by the negative Fisher information matrix.

Below, Theorem 4.6 presents the Berry-Esseen bound for the rows of the one-step refinement $\widehat{\mathbf{X}}$ of the adjacency spectral embedding (the scaled eigenvectors $\widetilde{\mathbf{X}}$).

Theorem 4.6. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and suppose the conditions of Theorem 3.1 hold. Further assume that there exists a constant* $\delta > 0$ *such that* $\min_{i,j\in[n]}\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \wedge (1-\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\} \geq \delta$. *Denote* $\mathbf{\Delta}_n = (1/n)\mathbf{X}^{\mathrm{T}}\mathbf{X}$ *and* $\mathbf{G}_n(\mathbf{x}_i) = (1/n)\sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)\}^{-1}$ *for each* $i \in [n]$. *If*

$$\frac{1}{(n\rho_n)\lambda_d(\mathbf{\Delta}_n)^{5/2}}\max\left\{\frac{(\log n)^{1/2}}{\lambda_d(\mathbf{\Delta}_n)^2}, \frac{\kappa(\mathbf{\Delta}_n)}{\lambda_d(\mathbf{\Delta}_n)^2}, \log n\right\} \to 0,$$

then for each fixed index $i \in [n]$ *and for all sufficiently large* n,

$$\begin{aligned}&\sup_{A\in\mathcal{A}}\left|\mathbb{P}\left\{\sqrt{n}\mathbf{G}_n(\mathbf{x}_i)^{1/2}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) \in A\right\} - \mathbb{P}(\mathbf{z}\in A)\right| \\ &\lesssim \frac{d^{1/2}}{n\rho_n^{1/2}\delta^8\lambda_d(\mathbf{\Delta}_n)^{9/2}}\max\left\{\frac{\log n\rho_n}{\lambda_d(\mathbf{\Delta}_n)^4}, \frac{\kappa(\mathbf{\Delta}_n)^2}{\lambda_d(\mathbf{\Delta}_n)^4}, (\log n\rho_n)^2\right\},\end{aligned} \tag{25}$$

where $\mathcal{A}$ *is the set of all convex measurable sets in* $\mathbb{R}^d$ *and* $\mathbf{z} \sim \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d)$.

Remark 2. Theorem 4.6 generalizes Theorem 5 in [82] in the following aspects: First, we allow $n\rho_n$ to grow at $\omega(\log n)$ when $\lambda_d(\mathbf{\Delta}_n) = \Omega(1)$, which is significantly weaker than the assumption $n\rho_n^5 = \omega((\log n)^2)$ in [82]; Secondly, we have the least requirement on the embedding dimension d and the latent position matrix $\mathbf{X}$, whereas the authors of [82] assumed that d is fixed and $\mathbf{X}$ satisfies a Glivenko-Cantelli type condition. In addition, Theorem 4.6 is also stated in terms of a Berry-Esseen type bound that only depends on $n\rho_n$, the embedding dimension d, the eigenvalues of $\mathbf{\Delta}_n$, and a constant δ governing the entries of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Hence, the rows of $\widehat{\mathbf{X}}\mathbf{W} - \rho_n^{1/2}\mathbf{X}$ can be approximated by a multivariate Gaussian as long as the right-hand side of (25) converges to 0.

The authors of [82] have shown that the covariance matrix $\mathbf{G}_n(\mathbf{x}_i)^{-1}$ for the rows of $\widehat{\mathbf{X}}$ satisfies $\mathbf{G}_n(\mathbf{x}_i)^{-1} \preceq \mathbf{\Sigma}_n(\mathbf{x}_i)$. Consequently, the one-step refinement of $\widetilde{\mathbf{X}}$ reduces the asymptotic variance of the rows of the scaled eigenvectors $\widetilde{\mathbf{X}}$ in spectra. This result is particularly useful in stochastic block models whose block probability matrix is rank-deficient (see Section 5 below for a numerical example).

Theorem 4.7 below provides a row-wise concentration bound for the one-step refinement $\widehat{\mathbf{X}}$ and is instrumental towards establishing Theorem 4.7. It also generalizes Theorem 4 in [82].

Theorem 4.7. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.6 hold. Then*

$$\mathbf{G}_n(\mathbf{x}_i)^{1/2}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_i - \rho_n^{1/2}\mathbf{x}_i) = \frac{1}{n\rho_n^{1/2}}\sum_{j=1}^{n}\frac{(A_{ij}-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j(1-\rho_n\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j)}\mathbf{G}_n(\mathbf{x}_i)^{-1/2}\mathbf{x}_j + \widehat{\mathbf{r}}_i, \tag{26}$$

where, given any fixed $c > 0$*, for all* $t \geq 1$*,* $t \lesssim \log n$*, and sufficiently large* n*, the remainder* $\widehat{\mathbf{r}}_i$ *satisfies*

$$\|\widehat{\mathbf{r}}_i\|_2 \lesssim_c \frac{1}{n\rho_n^{1/2}\delta^8\lambda_d(\mathbf{\Delta}_n)^{9/2}}\max\left\{\frac{t}{\lambda_d(\mathbf{\Delta}_n)^4}, \frac{\kappa(\mathbf{\Delta}_n)^2}{\lambda_d(\mathbf{\Delta}_n)^4}, t^2\right\}$$

with probability at least $1 - c_0 n^{-c} - c_0 e^{-t}$ *for some absolute constant* $c_0 > 0$*.*

4.4. Eigenvector-based subsequent inference for random graphs

In this subsection, we apply the theory in Sections 4.2 and 4.3 to two subsequent random graph inference problems: the estimation of pure nodes in mixed membership stochastic block models and the hypothesis testing of the equality of latent positions in random dot product graphs.

Pure node estimation in mixed membership stochastic block models

The mixed membership stochastic block model [5] generalizes the stochastic block model [39] in which the community memberships are continuously relaxed. Each vertex can have multiple community memberships governed by a probability vector called the community membership profile. There have been several works that explore the computation algorithms for mixed membership stochastic block models [5,37]. There have also been some recent progress in exploring the theoretical aspects of mixed membership stochastic block models (see, for example, [6,40,43,55,56]).

We first introduce the formal definition of the mixed membership stochastic block models.

Definition 4.8. Let $\mathbf{\Theta} = [\theta_{jk}]_{n\times d} \in [0,1]^{n\times d}$ be the membership profile matrix with $\sum_{k=1}^{d}\theta_{jk} = 1$ for all $j \in [n]$, $\mathbf{B} \in (0,1)^{d\times d}$ be the block probability matrix, and $\rho_n \in (0,1]$ be the sparsity factor. We say that a symmetric random matrix $\mathbf{A} \in \{0,1\}^{n\times n}$ is the adjacency matrix of a mixed membership stochastic block model $\mathrm{MMSBM}(\mathbf{\Theta},\mathbf{B},\rho_n)$, if $A_{ij} \sim \mathrm{Bernoulli}(\rho_n\boldsymbol{\theta}_i^{\mathrm{T}}\mathbf{B}\boldsymbol{\theta}_j)$ independently for all $i,j \in [n]$, $i \leq j$, and $A_{ij} = A_{ji}$ for all $i > j$, where $\boldsymbol{\theta}_i = [\theta_{i1},\ldots,\theta_{id}]^{\mathrm{T}}$.

For simplicity, we assume that the block probability matrix $\mathbf{B}$ is positive definite and there exist $\mathbf{X}^* = [\mathbf{x}_1^*,\ldots,\mathbf{x}_d^*]^{\mathrm{T}} \in \mathbb{R}^{d\times d}$ such that $\mathbf{B} = (\mathbf{X}^*)(\mathbf{X}^*)^{\mathrm{T}}$. Namely, $\mathbf{A} \sim \mathrm{MMSBM}(\mathbf{\Theta},(\mathbf{X}^*)(\mathbf{X}^*)^{\mathrm{T}},\rho_n)$ implies that $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{\Theta}\mathbf{X}^*)$. Geometrically, the latent positions (i.e., the rows of $\mathbf{\Theta}\mathbf{X}^*$) can be viewed as scatter points taken from a simplex whose corners are the rows of $\mathbf{X}^*$, and the rows of $\mathbf{X}^*$ are referred to as the pure nodes [56]. A standard condition for estimating the membership profile matrix $\mathbf{\Theta}$ is the existence of a pure node for each community [56]. Formally, we say that each of the d communities contains at least one pure node, if the vertex set $\{i \in [n] : \boldsymbol{\theta}_i = \mathbf{e}_k\}$ is non-empty for each $k \in [d]$. Then there exists d distinct row indices $i_1,\ldots,i_d \in [n]$ such that $i_k = \min\{i \in [n] : \boldsymbol{\theta}_i = \mathbf{e}_k\}$, where $\boldsymbol{\theta}_i$ is the ith row of $\mathbf{\Theta}$, $i \in [n]$. Namely, $\{i_1,\ldots,i_d\}$ are the vertices in the graph whose latent positions are exactly given by one of the pure nodes.

Given $\mathbf{A} \sim \mathrm{MMSBM}(\mathbf{\Theta},(\mathbf{X}^*)(\mathbf{X}^*)^{\mathrm{T}},\rho_n)$, an important inference task is to detect and estimate the pure nodes $\mathbf{x}_1^*,\ldots,\mathbf{x}_d^*$. There are several earlier attempts in detecting the row indices corresponding to the pure nodes [34,43,55,56]. These algorithms are based on the finding that the corners of a simplex have

the highest norm (see Lemma 2.1 in [56]). Here, we adopt the successive projection algorithm proposed in [34]. The detailed algorithm is provided in Appendix for completeness.

We now construct two estimators for the pure nodes in a mixed membership stochastic block model based on the adjacency spectral embedding $\widetilde{\mathbf{X}} = [\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_n]^{\mathrm{T}}$ (i.e., the scaled eigenvectors) and its one-step refinement $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_n]^{\mathrm{T}}$. Let $J := \mathtt{SPA}(\mathbf{A}, d)$ be the output row indices of the successive projection algorithm (see Algorithm 1 in Appendix) and $\mathbf{V}_{\mathbf{A}} \in \mathbb{R}^{d\times d}$ the sub-matrix of $\mathbf{U}_{\mathbf{A}}$ corresponding to the row indices in J. We then estimate the membership profile matrix $\boldsymbol{\Theta}$ by $\widehat{\boldsymbol{\Theta}} := \mathbf{U}_{\mathbf{A}}\mathbf{V}_{\mathbf{A}}^{-1}$. Define

$$\iota_k := \min\left\{i \in [n] : \|\mathbf{e}_i^{\mathrm{T}}\widehat{\boldsymbol{\Theta}} - \mathbf{e}_k^{\mathrm{T}}\|_2 \le \eta\right\}, \quad k \in [d], \tag{27}$$

where $\eta > 0$ is a tuning parameter taken to be sufficiently small. Note that the membership profile matrix is only identifiable up to a permutation. The two estimators for $\mathbf{x}_k^*$ (modulus a permutation) are then given by $\widetilde{\mathbf{x}}_{\iota_k}$ and $\widehat{\mathbf{x}}_{\iota_k}$, which are based on $\widetilde{\mathbf{X}}$ and $\widehat{\mathbf{X}}$, respectively. Leveraging Corollary 4.1, Theorem 4.4, and Theorem 4.7, we establish the two-to-infinity norm error bound for $\widehat{\boldsymbol{\Theta}}$ and the asymptotic normality of $\widetilde{\mathbf{x}}_{\iota_k}$ and $\widehat{\mathbf{x}}_{\iota_k}$ in Theorem 4.9 below.

Theorem 4.9. *Suppose* $\mathbf{A} \sim \mathrm{MMSBM}(\boldsymbol{\Theta}, (\mathbf{X}^*)(\mathbf{X}^*)^{\mathrm{T}}, \rho_n)$ *and the following conditions hold:*

(i) *There exists at least one pure node for each of the* d *communities.*
(ii) $n\rho_n = \omega(\log n)$ *and* d *is fixed.*
(iii) *There exists a positive constant* $c_1 > 0$ *such that* $\min\{n^{-1/2}\sigma_d(\boldsymbol{\Theta}), \sigma_d(\mathbf{X}^*)\} \ge c_1$.
(iv) *There exists a positive constant* $\delta > 0$ *such that* $\min_{k,l\in[d]}[(\mathbf{x}_k^*)^{\mathrm{T}}(\mathbf{x}_l^*) \wedge \{1 - (\mathbf{x}_k^*)^{\mathrm{T}}(\mathbf{x}_l^*)\}] \ge \delta$.

Then for each sufficiently large n*, there exists a permutation matrix* $\boldsymbol{\Pi}_n \in \{0,1\}^{d\times d}$*, such that with probability at least* $1 - c_0 n^{-2}$*,* $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\boldsymbol{\Pi}_n\|_{2\to\infty} \le K\sqrt{(\log n)/(n\rho_n)}$ *for some constants* $K, c_0 > 0$*. Furthermore, if* $\min_{i\in[n],\boldsymbol{\theta}_i\neq\mathbf{e}_k} \|\boldsymbol{\theta}_i - \mathbf{e}_k\|_2 \ge c_2$ *for a constant* $c_2 > 0$ *for all* $k \in [d]$ *and* $\eta \le c_2/2$*, then there exists a sequence of permutations* $(\pi_n)_n$ *over* $[d]$*, such that for each* $k \in [d]$*,*

$$\sqrt{n}\boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})^{-1/2}(\mathbf{W}^{\mathrm{T}}\widetilde{\mathbf{x}}_{\iota_k} - \rho_n^{1/2}\mathbf{x}^*_{\pi_n(k)}) \xrightarrow{\mathcal{L}} \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d),$$

$$\sqrt{n}\mathbf{G}_n(\mathbf{x}^*_{\pi_n(k)})^{1/2}(\mathbf{W}^{\mathrm{T}}\widehat{\mathbf{x}}_{\iota_k} - \rho_n^{1/2}\mathbf{x}^*_{\pi_n(k)}) \xrightarrow{\mathcal{L}} \mathrm{N}_d(\mathbf{0}_d, \mathbf{I}_d),$$

where $\boldsymbol{\Sigma}_n(\cdot)$ *and* $\mathbf{G}_n(\cdot)$ *are defined in Theorem 4.4 and Theorem 4.6, respectively, with* $\mathbf{X} := \boldsymbol{\Theta}\mathbf{X}^*$.

The implication of Theorem 4.9 is two-fold. Firstly, we establish the following uniform error bound for the membership profile estimator $\widehat{\boldsymbol{\Theta}}$: $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\boldsymbol{\Pi}_n\|_{2\to\infty} = O\{\sqrt{(\log n)(n\rho_n)}\}$ with probability at least $1 - O(n^{-2})$, where $\boldsymbol{\Pi}_n$ is a $d \times d$ permutation matrix. This concentration bound is sharper than that in [56] by a poly-log n factor and our sparsity assumption is weaker: we only assume that $n\rho_n = \omega(\log n)$, whereas the authors of [56] required that $n\rho_n = \Omega((\log n)^{2\xi})$ for some constant $\xi > 1$. Secondly, we show the asymptotic normality for the pure node estimators $\widetilde{\mathbf{x}}_{\iota_k}$ based on the adjacency spectral embedding and $\widehat{\mathbf{x}}_{\iota_k}$ based on the one-step estimator, with the asymptotic covariance matrices being $\boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})$ and $\mathbf{G}_n(\mathbf{x}^*_{\pi_n(k)})^{-1}$, respectively. By Theorem 2 in [82], we have $\mathbf{G}_n(\mathbf{x}^*_{\pi_n(k)})^{-1} \preceq \boldsymbol{\Sigma}_n(\mathbf{x}^*_{\pi_n(k)})$. Therefore, the estimator $\widehat{\mathbf{x}}_{\iota_k}$, which is derived from the one-step estimator $\widehat{\mathbf{X}}$, improves upon the eigenvector-based estimator $\widetilde{\mathbf{x}}_{\iota_k}$ with a smaller asymptotic covariance matrix in spectra.

Hypothesis testing for equality of latent positions

The second subsequent inference problem is to test whether the latent positions of two given vertices are the same or not in a random dot product graph. This subsequent network inference task is inspired by the

hypothesis testing of the membership profiles in degree-corrected mixed membership stochastic block models proposed in [31]. The testing procedure could be useful in, e.g., diversifying the portfolios in the stock market investment and maximizing the expected returns [31]. Formally, given $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ and fixed vertex indices $i, j \in [n]$, $i \neq j$, we consider testing the null hypothesis $H_0 : \mathbf{x}_i = \mathbf{x}_j$ against the alternative hypothesis $H_A : \mathbf{x}_i \neq \mathbf{x}_j$. Motivated by the asymptotic normality in Theorems 4.4 and 4.6, we consider the following two test statistics associated with the adjacency spectral embedding $\widetilde{\mathbf{X}}$ and its one-step refinement $\widehat{\mathbf{X}}$, respectively:

$$T_{ij}^{(\mathrm{ASE})} = n(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_{ij}^{-1}(\widetilde{\mathbf{x}}_i - \widetilde{\mathbf{x}}_j) \quad \text{and} \quad T_{ij}^{(\mathrm{OSE})} = n(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)^{\mathrm{T}}\widetilde{\mathbf{G}}_{ij}^{-1}(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j),$$

where $\widetilde{\boldsymbol{\Sigma}}_{ij} = \widetilde{\boldsymbol{\Sigma}}_n(\widetilde{\mathbf{x}}_i) + \widetilde{\boldsymbol{\Sigma}}_n(\widetilde{\mathbf{x}}_j)$, $\widetilde{\mathbf{G}}_{ij} = \widetilde{\mathbf{G}}_n(\widetilde{\mathbf{x}}_i)^{-1} + \widetilde{\mathbf{G}}_n(\widetilde{\mathbf{x}}_j)^{-1}$,

$$\widetilde{\boldsymbol{\Sigma}}_n(\mathbf{x}) = \widetilde{\boldsymbol{\Delta}}_n^{-1}\left\{\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\mathbf{x}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}\right\}\widetilde{\boldsymbol{\Delta}}_n, \quad \widetilde{\mathbf{G}}_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n}\frac{\widetilde{\mathbf{x}}_j\widetilde{\mathbf{x}}_j^{\mathrm{T}}}{\mathbf{x}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j(1-\mathbf{x}_i^{\mathrm{T}}\widetilde{\mathbf{x}}_j)},$$

and $\widetilde{\boldsymbol{\Delta}}_n = (1/n)\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}$. In what follows, we establish the asymptotic distributions of the test statistics $T_{ij}^{(\mathrm{ASE})}$ and $T_{ij}^{(\mathrm{OSE})}$ under the null and alternative hypotheses.

Theorem 4.10. *Let* $\mathbf{A} \sim \mathrm{RDPG}(\rho_n^{1/2}\mathbf{X})$ *and assume the conditions of Theorem 4.6 hold. Further assume that* d *is fixed and* $\lambda_d(\boldsymbol{\Delta}_n)$ *is bounded away from* 0.

(i) *Under the null hypothesis* $H_0 : \mathbf{x}_i = \mathbf{x}_j$, *we have* $T_{ij}^{(\mathrm{ASE})} \xrightarrow{\mathcal{L}} \chi_d^2$ *and* $T_{ij}^{(\mathrm{OSE})} \xrightarrow{\mathcal{L}} \chi_d^2$.

(ii) *Under the alternative hypothesis* $H_A : \mathbf{x}_i \neq \mathbf{x}_j$, *if* $(n\rho_n)^{1/2}(\mathbf{x}_i - \mathbf{x}_j) \to \boldsymbol{\mu}$ *for some non-zero vector* $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}_n(\mathbf{x}_i) \to \boldsymbol{\Sigma}_i$, *and* $\mathbf{G}_n(\mathbf{x}_i) \to \mathbf{G}_i$ *for some fixed positive definite* $\boldsymbol{\Sigma}_i$ *and* $\mathbf{G}_i$ *as* $n \to \infty$, *then* $T_{ij}^{(\mathrm{ASE})} \xrightarrow{\mathcal{L}} \chi_d^2(\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j)^{-1}\boldsymbol{\mu})$ *and* $T_{ij}^{(\mathrm{OSE})} \xrightarrow{\mathcal{L}} \chi_d^2(\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{G}_i^{-1} + \mathbf{G}_j^{-1})^{-1}\boldsymbol{\mu})$, *where, for any* $a > 0$, $\chi_d^2(a)$ *is the noncentral chi-squared distribution with noncentral parameter* a *and degree of freedom* d.

An important consequence of Theorem 4.10 is the power comparison between the two test statistics. It turns out that the test based on $T_{ij}^{(\mathrm{OSE})}$ is more powerful than the test based on $T_{ij}^{(\mathrm{ASE})}$ under the conditions of Theorem 4.10 (ii). Given a significance level $\alpha \in (0,1)$, we can construct the following test functions:

$$\phi_{ij}^{(\mathrm{ASE})} = \mathbb{1}\left\{T_{ij}^{(\mathrm{ASE})} > q_{\chi_d^2}(1-\alpha)\right\} \quad \text{and} \quad \phi_{ij}^{(\mathrm{OSE})} = \mathbb{1}\left\{T_{ij}^{(\mathrm{OSE})} > q_{\chi_d^2}(1-\alpha)\right\},$$

where $q_{\chi_d^2}(1-\alpha)$ is the $(1-\alpha)$ quantile of the chi-squared distribution with degree of freedom d. Then under the conditions of Theorem 4.10 (i), we see that the two tests are asymptotically valid level-α tests, i.e., $\mathbb{E}_{H_0}\phi_{ij}^{(\mathrm{ASE})} \to \alpha$ and $\mathbb{E}_{H_0}\phi_{ij}^{(\mathrm{OSE})} \to \alpha$ as $n \to \infty$. To compare the powers of the two tests under the alternative $H_A : \mathbf{x}_i \neq \mathbf{x}_j$ under the conditions of Theorem 4.10 (ii), we first observe that the non-central chi-squared distribution is stochastic increasing in its non-central parameter [68]. By Theorem 2 in [82], the non-central parameters for $T_{ij}^{(\mathrm{ASE})}$ and $T_{ij}^{(\mathrm{OSE})}$ satisfy the inequality $\boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j)^{-1}\boldsymbol{\mu} \leq \boldsymbol{\mu}^{\mathrm{T}}(\mathbf{G}_i^{-1} + \mathbf{G}_j^{-1})^{-1}\boldsymbol{\mu}$. Therefore, under the alternative hypothesis $H_A : \mathbf{x}_i \neq \mathbf{x}_j$ and the conditions of Theorem 4.10 (ii), we conclude that $\lim_{n\to\infty}\mathbb{E}_{H_A}\phi_{ij}^{(\mathrm{ASE})} \leq \lim_{n\to\infty}\mathbb{E}_{H_A}\phi_{ij}^{(\mathrm{OSE})}$. Namely, the test based on $T_{ij}^{(\mathrm{OSE})}$ is asymptotically more powerful than the test based on $T_{ij}^{(\mathrm{ASE})}$.

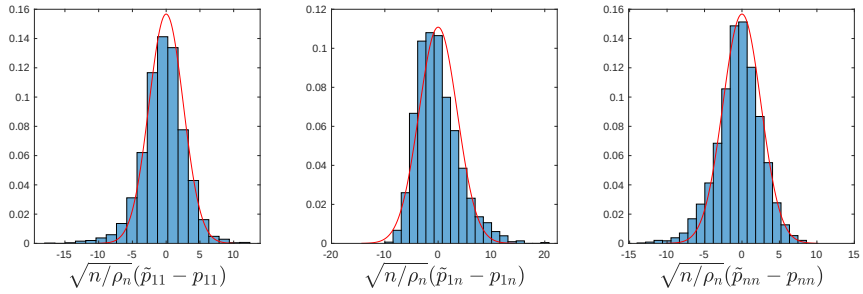

Figure 3. Numerical results of Section 5.1. The three panels are the histograms of $\sqrt{n/\rho_n}(\widetilde{p}_{11} - p_{11})$, $\sqrt{n/\rho_n}(\widetilde{p}_{1n} - p_{1n})$, and $\sqrt{n/\rho_n}(\widetilde{p}_{nn} - p_{nn})$ over the 3000 Monte Carlo replicates with the asymptotic normal densities highlighted in the red curves, respectively.

5. Simulation study

5.1. Symmetric noisy matrix completion

We first consider a synthetic example for the symmetric noisy matrix completion problem and illustrate Theorem 4.2. The setup here is similar to the two-block stochastic block model in Section 1.2. We set $n = 5000$, $a = 0.9$, $b = 1.8$, $n\rho_n = 6\log n$, $\lambda_1 = n\rho_n(a+b)/2$, $\lambda_2 = n\rho_n(a-b)/2$, $\mathbf{u}_1 = n^{-1/2}[1,\ldots,1]^{\mathrm{T}}$, $\mathbf{u}_2 = n^{-1/2}[1,\ldots,1,-1,\ldots,-1]^{\mathrm{T}}$ (the first $n/2$ entries of $\mathbf{u}_2$ are $n^{-1/2}$ and the remaining entries are $-n^{-1/2}$), and $\mathbf{X} = \sqrt{n}[\mathbf{u}_1,\mathbf{u}_2]$. Let $\epsilon_{ij} \sim \mathrm{N}(0, 0.1^2\rho_n^4)$ and $I_{ij} \sim \mathrm{Bernoulli}(\rho_n)$ independently for $i,j \in [n]$, $i \le j$, and let $\epsilon_{ij} = \epsilon_{ji}$, $I_{ij} = I_{ji}$ if $i > j$. The noisy observed matrix $\mathbf{A} = [A_{ij}]_{n\times n}$ is generated by taking $A_{ij} = (\rho_n \mathbf{x}_i^{\mathrm{T}}\mathbf{I}_{1,1}\mathbf{x}_j + \epsilon_{ij})I_{ij}/\rho_n$, where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$, $i \in [n]$. We follow the notations in Section 4.2 by letting $\widetilde{\mathbf{x}}_i$ and $\mathbf{x}_i$ be the ith row of $\widetilde{\mathbf{X}}$ and $\mathbf{X}$, respectively, where $\widetilde{\mathbf{X}}$ is defined as in Section 2.1, $\widetilde{p}_{11} = \widetilde{\mathbf{x}}_1^{\mathrm{T}}\mathbf{I}_{1,1}\widetilde{\mathbf{x}}_1$, $\widetilde{p}_{1n} = \widetilde{\mathbf{x}}_1^{\mathrm{T}}\mathbf{I}_{1,1}\widetilde{\mathbf{x}}_n$, $\widetilde{p}_{nn} = \widetilde{\mathbf{x}}_n^{\mathrm{T}}\mathbf{I}_{1,1}\widetilde{\mathbf{x}}_n$, $p_{11} = \rho_n\mathbf{x}_1^{\mathrm{T}}\mathbf{I}_{1,1}\mathbf{x}_1$, $p_{1n} = \rho_n\mathbf{x}_1^{\mathrm{T}}\mathbf{I}_{1,1}\mathbf{x}_n$, and $p_{nn} = \rho_n\mathbf{x}_n^{\mathrm{T}}\mathbf{I}_{1,1}\mathbf{x}_n$. Then by Theorem 3.1,

$$\sqrt{\frac{n}{\rho_n}}(\widetilde{p}_{11} - p_{11}) \xrightarrow{\mathcal{L}} \mathrm{N}(0, 8a^2), \quad \sqrt{\frac{n}{\rho_n}}(\widetilde{p}_{1n} - p_{1n}) \xrightarrow{\mathcal{L}} \mathrm{N}(0, 4b^2), \quad \sqrt{\frac{n}{\rho_n}}(\widetilde{p}_{nn} - p_{nn}) \xrightarrow{\mathcal{L}} \mathrm{N}(0, 8a^2).$$

We next generate 3000 independent Monte Carlo replicates of $\mathbf{A}$ from $\mathrm{SNMC}(\rho_n\mathbf{X}\mathbf{I}_{1,1}\mathbf{X}^{\mathrm{T}}, \rho_n, 0.1\rho_n^2)$. For each realization of $\mathbf{A}$, we compute the estimates $\widetilde{p}_{11}$, $\widetilde{p}_{1n}$, $\widetilde{p}_{nn}$ for p_{11}, p_{1n}, p_{nn}, and plot the histograms of $\sqrt{n/\rho_n}(\widetilde{p}_{11} - p_{11})$, $\sqrt{n/\rho_n}(\widetilde{p}_{1n} - p_{1n})$, $\sqrt{n/\rho_n}(\widetilde{p}_{nn} - p_{nn})$ in Figure 3, together with their respective asymptotic normal densities. It is clear that the empirical distributions of the Monte Carlo samples of $\sqrt{n/\rho_n}(\widetilde{p}_{11} - p_{11})$, $\sqrt{n/\rho_n}(\widetilde{p}_{1n} - p_{1n})$, and $\sqrt{n/\rho_n}(\widetilde{p}_{nn} - p_{nn})$ can be well approximated by their respective asymptotic normal distributions. This numerical observation is in agreement with the asymptotic normality established in Theorem 4.2.

5.2. Eigenvectors and their one-step refinements for random graphs

In this section, we present a simulated example of random dot product graphs. Consider a stochastic block model on n vertices with a cluster assignment rule $\tau : [n] \to \{1,2\}$ and a block probability matrix

$$\mathbf{B} = \rho_n \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix},$$

where $\rho_n \in (0,1)$ is a sparsity factor and $p, q \in (0,1)$. The adjacency matrix $\mathbf{A} = [A_{ij}]_{n\times n}$ is generated as follows: For all $i \le j$, $i, j \in [n]$, let $A_{ij} \sim \text{Bernoulli}([\mathbf{B}]_{\tau(i)\tau(j)})$ independently for $i \le j$ and let $A_{ij} = A_{ji}$ for all $i > j$. We take $\tau(i) = 1$ if $i = 1, \ldots, n/2$, and $\tau(i) = 2$ if $i = n/2+1, \ldots, n$ for simplicity. The number of vertices n is set to 5000 and we take $n\rho_n = 5(\log n)^{3/2}$ such that the conditions of Theorem 3.1 and Theorem 4.6 are both satisfied. The values of p and q are selected to be $p = 0.95$ and $q = 0.3$.

We generate 3000 independent copies of the adjacency matrix $\mathbf{A}$ from the aforementioned stochastic block model. For each realization of $\mathbf{A}$, we compute the adjacency spectral embedding $\widetilde{\mathbf{x}}$ of $\mathbf{A}$ into $\mathbb{R}$, the unscaled top eigenvector $\mathbf{u}_\mathbf{A}$ of $\mathbf{A}$, and the one-step refinement $\widehat{\mathbf{x}}$ of $\widetilde{\mathbf{x}}$. The population scaled eigenvector and the unscaled eigenvector are denoted by $\rho_n^{1/2}\mathbf{x}$ and $\mathbf{u}_\mathbf{P}$, respectively. For this specific model, it is straightforward to obtain $\mathbf{x} = \begin{bmatrix} p \ldots p\ q \ldots q \end{bmatrix}$ and $\mathbf{u}_\mathbf{P} = (np^2/2 + nq^2/2)^{-1/2} \begin{bmatrix} p \ldots p\ q \ldots q \end{bmatrix}$. The only non-zero eigenvalue of $\mathbf{P}$ is $\lambda = n\rho_n(p^2/2 + q^2/2)$. For each $i \in [n]$, we denote $\widetilde{x}_i$, $[\mathbf{u}_\mathbf{A}]_i$, $\widehat{x}_i$, x_i, and $[\mathbf{u}_\mathbf{P}]_i$ the ith coordinates of $\widetilde{\mathbf{x}}$, $\mathbf{u}_\mathbf{A}$, $\widehat{\mathbf{x}}$, $\mathbf{x}$, and $\mathbf{u}_\mathbf{P}$, respectively. Then by Theorem 3.1 and Theorem 4.6, for each $i \in [n]$, the random variables $\sqrt{n}(\widetilde{x}_i - \rho_n^{1/2}x_i)$, $\sqrt{n}(\widehat{x}_i - \rho_n^{1/2}x_i)$, and $n\rho_n^{1/2}([\mathbf{u}_\mathbf{A}]_i - [\mathbf{u}_\mathbf{P}]_i)$ converge to mean-zero Gaussians in distribution with the variances depending on p,q, and the community membership $\tau(i)$.

We take $i = 1$ as an illustrative vertex and visualize the numerical performance of $\widetilde{\mathbf{x}}$ and $\mathbf{u}_\mathbf{A}$ in Figures 4 and 5. The left panels of Figures 4, 5 are the histograms of $\sqrt{n}(\widetilde{x}_1 - \rho_n^{1/2}x_1)$ and $n\rho_n^{1/2}([\mathbf{u}_\mathbf{A}]_1 - [\mathbf{u}_\mathbf{P}]_1)$ with the corresponding asymptotic normal densities highlighted in the red curves. We see that the shapes of the two histograms are closely aligned with the limit densities, verifying the conclusion of Theorem 3.1 empirically. The right panels of Figures 4 and 5 present the boxplots of $\sqrt{n}(\widehat{x}_1 - \rho_n^{1/2}x_1)$

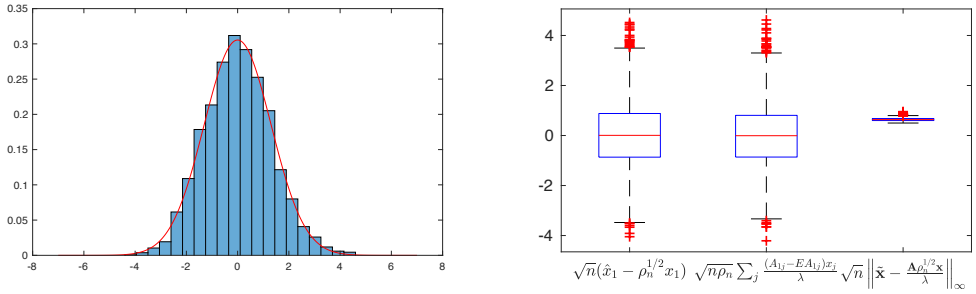

Figure 4. Numerical results for Section 5.2. Left panel: The histogram of $\sqrt{n}(\widetilde{x}_1 - \rho_n^{1/2}x_1)$ over the 3000 Monte Carlo replicates with the asymptotic normal density highlighted in the red curve. Right panel: The boxplots of $\sqrt{n}(\widehat{x}_1 - \rho_n^{1/2}x_1)$, its linear approximation $\sqrt{n\rho_n}\sum_j(A_{1j} - \mathbb{E}A_{1j})x_j/\lambda_2$, and the infinity norm of the higher-order remainder $\sqrt{n}\|\widetilde{\mathbf{x}} - \mathbf{A}\rho_n^{1/2}\mathbf{x}/\lambda\|_\infty$ across the 3000 Monte Carlo replicates.

and $n\rho_n^{1/2}([\mathbf{u}_\mathbf{A}]_1 - [\mathbf{u}_\mathbf{P}]_1)$, their linear approximations, and the infinity norms of the corresponding higher-order remainders. From the right panel of Figure 4, we can see that the dominating term for $\sqrt{n}(\widehat{x}_1 - \rho_n^{1/2}x_1)$ is $\sqrt{n\rho_n}\sum_j(A_{1j} - \mathbb{E}A_{1j})x_j/\lambda$. However, the infinity norm of the higher-order remainder $\sqrt{n}\|\widetilde{\mathbf{x}} - \mathbf{A}\rho_n^{1/2}\mathbf{x}/\lambda\|_\infty$ is not necessarily negligible. This agrees with the observation in Section 1.2. A similar observation regarding the unscaled eigenvector can be found in the right panel of Figure 5 as well.

We also compare the performance between the adjacency spectral embedding $\widetilde{\mathbf{x}}$ and its one-step refinement $\widehat{\mathbf{x}}$ in Figure 6 below. Taking $i = 1$ as an illustrative vertex, we visualize the histogram of $\sqrt{n}(\widehat{x}_1 - \rho_n^{1/2}x_1)$ in the left panel of Figure 6, overlaid with the corresponding asymptotic normal density

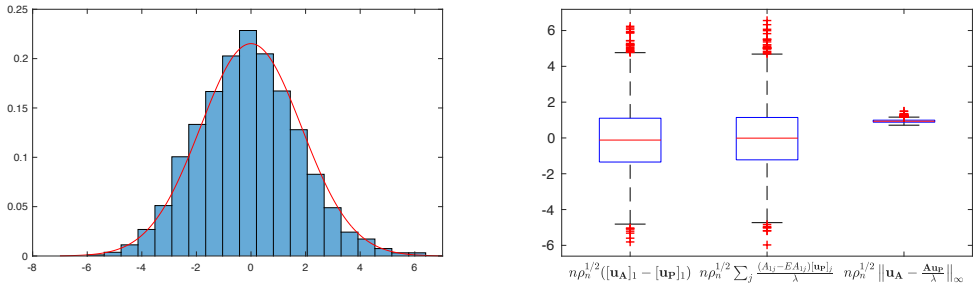

Figure 5. Numerical results for Section 5.2. Left panel: The histogram of $n\rho_n^{1/2}([\mathbf{u_A}]_1 - [\mathbf{u_P}]_1)$ over the 3000 Monte Carlo replicates with the asymptotic normal density highlighted in the red curve. Right panel: The boxplots of $n\rho_n^{1/2}([\mathbf{u_A}]_1 - [\mathbf{u_P}]_1)$, its linear approximation $n\rho_n^{1/2}\sum_j(A_{1j} - \mathbb{E}A_{1j})[\mathbf{u_P}]_j/\lambda$, and the infinity norm of the higher-order remainder $n\rho_n^{1/2}\|\mathbf{u_A} - \mathbf{Au_P}/\lambda\|_\infty$ across the 3000 Monte Carlo replicates.

in the red curve. The limit normal density is almost perfectly aligned with the histogram, verifying Theorem 4.6 numerically. The right panel compares the boxplots of the estimation errors $\|\widetilde{\mathbf{x}}w - \rho_n^{1/2}x\|_2^2$ and $\|\widehat{\mathbf{x}}w - \rho_n^{1/2}x\|_2^2$ across the 3000 Monte Carlo replicates, where w is the sign of $\mathbf{u}_\mathbf{A}^\mathrm{T}\mathbf{u_P}$. It is clear that the errors of the one-step refinement $\widehat{\mathbf{x}}$ are smaller than those of the adjacency spectral embedding, which also agrees with the observation in [82] but under a much sparser regime that $n\rho_n \propto (\log n)^{3/2}$.

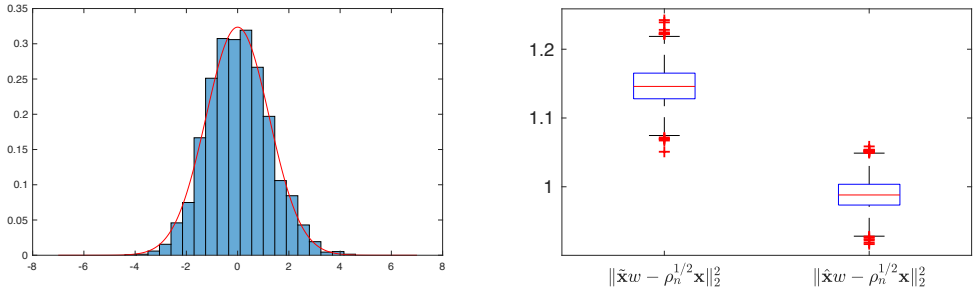

Figure 6. Numerical results for Section 5.2. Left panel: The histogram of $\sqrt{n}(\widehat{x}_1 - \rho_n^{1/2}x_1)$ over the 3000 Monte Carlo replicates with the asymptotic normal density highlighted in the red curve. Right panel: The boxplots of $\|\widetilde{\mathbf{x}}w - \rho_n^{1/2}\mathbf{x}\|_2^2$ for the adjacency spectral embedding and $\|\widehat{\mathbf{x}}w - \rho_n^{1/2}\mathbf{x}\|_2^2$ for its one-step refinement across the 3000 Monte Carlo replicates, where w is the sign of $\mathbf{u}_\mathbf{A}^\mathrm{T}\mathbf{u_P}$.

5.3. Eigenvector-based inference in random graphs

We now consider the numerical experiments for the two subsequent random graph inference tasks in Section 4.4. Consider a mixed membership stochastic block model specified as follows. The block probability matrix $\mathbf{B}$ is a 2×2 symmetric matrix with the diagonals being 0.9 and the off-diagonals being 0.1. The corresponding two pure nodes are $\mathbf{x}_1^* = [0.7071, 0.6325]^\mathrm{T}$ and $\mathbf{x}_2^* = [0.7071, -0.6325]^\mathrm{T}$.

Table 1. Numerical results for Section 5.3: Summary statistic for estimating the pure nodes.

Pure node	MSE for $\widetilde{\mathbf{x}}_{\iota_k}$	MSE for $\widehat{\mathbf{x}}_{\iota_k}$	Sample covariance for $\widetilde{\mathbf{x}}_{\iota_k}$	Sample covariance for $\widehat{\mathbf{x}}_{\iota_k}$
$\mathbf{x}_1^*$	7.4×10^{-4}	5.4×10^{-4}	$\frac{1}{n}\begin{bmatrix}1.00 & 0.90\\ 0.90 & 2.25\end{bmatrix}$	$\frac{1}{n}\begin{bmatrix}0.97 & 0.75\\ 0.75 & 1.34\end{bmatrix}$
$\mathbf{x}_2^*$	7.8×10^{-4}	5.4×10^{-4}	$\frac{1}{n}\begin{bmatrix}0.99 & -0.96\\ -0.96 & 2.36\end{bmatrix}$	$\frac{1}{n}\begin{bmatrix}0.95 & -0.75\\ -0.75 & 1.36\end{bmatrix}$

We set the number of vertices to be $n = 4500$ with $n_0 = 900$ pure nodes in each community and set the sparsity factor to be $\rho_n = 5(\log n)^{3/2}/n$. The membership profile matrix $\mathbf{\Theta}$ has the form

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{1}_{n_0} & \mathbf{0}_{n_0} \\ \mathbf{0}_{n_0} & \mathbf{1}_{n_0} \\ \mathbf{t} & \mathbf{1}_{n-2n_0} - \mathbf{t} \end{bmatrix},$$

where $\mathbf{t} \in \mathbb{R}^{n-2n_0}$ is the vector whose entries are equidistant points over $[0.2, 0.8]$ and $\mathbf{1}_{n_0} \in \mathbb{R}^{n_0}$ is the vector of all ones. Equivalently, the mixed membership stochastic block model can be written as $\mathrm{RPDG}(\rho_n^{1/2}\mathbf{X})$ with $\mathbf{X} = \mathbf{\Theta}\mathbf{X}^*$, where $\mathbf{X}^* = [\mathbf{x}_1^*, \mathbf{x}_2^*]^{\mathrm{T}}$. Let $\mathbf{x}_i$ be the ith row of $\mathbf{X}$ for each $i \in [n]$.

We draw 2000 independent Monte Carlo replicates of $\mathbf{A}$ from $\mathrm{MMSBM}(\mathbf{\Theta}, \mathbf{B}, \rho_n)$ specified above and investigate the performance of the two inference tasks in Section 4.4: The estimation of the pure nodes and the hypothesis testing of the equality of latent positions. For the first task, given a realization $\mathbf{A} \sim \mathrm{MMSBM}(\mathbf{\Theta}, \mathbf{B}, \rho_n)$, we first compute the adjacency spectral embedding $\widetilde{\mathbf{X}}$ of $\mathbf{A}$ into $\mathbb{R}^2$ and then apply Algorithm 1 to obtain the estimated pure node indices J. Next, we compute ι_k using formula (27) for $k = 1, 2$, with the tuning parameter η being 0.1. For each $k = 1, 2$, we then compute the two estimators given by $\widetilde{\mathbf{x}}_{\iota_k}$ (the estimator based on the adjacency spectral embedding) and $\widehat{\mathbf{x}}_{\iota_k}$ (the estimator based on the one-step refinement). For the second task, we let $i = 1$ and consider testing the null hypothesis $H_0 : \mathbf{x}_i = \mathbf{x}_j$ against alternative hypothesis $H_A : \mathbf{x}_i \neq \mathbf{x}_j$ for $j \in \{1901, 2101, \ldots, 3701\}$ using the two test statistics $T_{ij}^{(\mathrm{ASE})}$ and $T_{ij}^{(\mathrm{OSE})}$ defined in Section 4.4. We set the significance level to be 0.05.

For the estimation of the pure nodes, for each $k = 1, 2$, we compute the empirical mean-squared errors (MSE) for estimating $\mathbf{x}_k^*$ using $\widetilde{\mathbf{x}}_{\iota_k}$ and $\widehat{\mathbf{x}}_{\iota_k}$ across the 2000 repeated experiments. We also compute the corresponding sample covariance matrices. These numerical results are summarized in Table 1. Clearly, Table 1 suggests that the estimators $\widehat{\mathbf{x}}_{\iota_1}, \widehat{\mathbf{x}}_{\iota_2}$ based on the one-step refinement have smaller mean-squared errors and smaller variances compared to the estimators $\widetilde{\mathbf{x}}_{\iota_1}, \widetilde{\mathbf{x}}_{\iota_2}$ based on the adjacency spectral embedding. This phenomenon validates Theorem 4.9 empirically.

For the hypothesis testing of the equality of the latent positions, we compare the empirical powers of the two testing procedures across the 2000 repeated experiments. Below, Table 2 tabulates the empirical powers of tests (i.e., the numbers of successful rejections divided by 2000) based on $T_{ij}^{(\mathrm{ASE})}$ and $T_{ij}^{(\mathrm{OSE})}$ as functions of the distance $\|\mathbf{x}_i - \mathbf{x}_j\|_2$ when j varies in $\{1901, 2101, \ldots, 3701\}$. It is clear from Table 2 that the test based on $T_{ij}^{(\mathrm{OSE})}$ is more powerful than the test based on $T_{ij}^{(\mathrm{ASE})}$, which validates Theorem 4.10 empirically.

6. Discussion

In this paper, we establish the Berry-Esseen theorems for the entrywise limits of the eigenvectors for a broad class of random matrix models with low expected rank, referred to as the signal-plus-noise

Table 2. Numerical results for Section 5.3: Power comparison for testing the equality of latent positions.

$\|\mathbf{x}_i - \mathbf{x}_j\|_2$	0.31	0.37	0.42	0.48	0.53	0.59	0.65	0.70	0.76	0.82
Power of $T_{ij}^{(\mathrm{ASE})}$	0.33	0.42	0.52	0.61	0.70	0.76	0.85	0.89	0.94	0.95
Power of $T_{ij}^{(\mathrm{OSE})}$	0.35	0.45	0.56	0.67	0.75	0.80	0.88	0.92	0.95	0.97

matrix model. Our generic entrywise eigenvector limit theorem leads to new and sharp results for several concrete statistical applications: the symmetric noisy matrix completion model, the eigenvectors and their one-step refinement of random dot product graphs, the estimation of pure nodes in mixed membership stochastic block models, and the hypothesis testing of the equality of latent positions in random graphs.

Several potential future research directions are worth exploring. In terms of the general signal-plus-noise matrix model framework, we restrict ourselves within the class of symmetric random matrices whose upper diagonal entries are independent random variables. Extensions to singular vectors of rectangular random matrices may be interesting for rectangular noisy matrix completion problems, bipartite network analysis, and high-dimensional principal component analysis [4].

For the symmetric matrix completion problem, we require that the variance of the mean-zero normal errors scales at the rate ρ_n^4. It is possible to relax this requirement and assume that $\mathrm{var}(\epsilon_{ij})$ scales at the rate ρ_n^2 by modifying the proof technique in [3]. This relaxation may require additional work because Assumption 3 no longer holds when $\mathrm{var}(\epsilon_{ij}) \propto \rho_n^2$.

For random dot product graphs, we have focused on the eigenvector analysis of the graph adjacency matrix. It has also been observed that the eigenvectors of the normalized Laplacian matrix have decent performance when the graph becomes sparse [63,72]. The entrywise limit theorems for the eigenvectors of the normalized Laplacian have been established in [72] under the sparsity assumption that $n\rho_n = \omega((\log n)^4)$. An interesting future research direction is to explore the entrywise limit theorems for the eigenvectors of the normalized Laplacian when $n\rho_n = \Omega(\log n)$. In addition, there has also been a growing interest in developing limit theorems for spectral analysis of multiple graphs [8,51]. We believe that the results and the approach developed in the present work may shed some light on the entrywise estimation of the eigenvectors for multiple random graph models.

Appendix: The successive projection algorithm

This appendix provides the detailed successive projection algorithm proposed in [34], which finds the row indices corresponding to the pure nodes based on the noisy observed adjacency matrix $\mathbf{A}$. It is used to construct the estimator $\widehat{\mathbf{\Theta}}$ for the membership profile matrix $\mathbf{\Theta}$ in a mixed membership stochastic block model in Section 4.4.

Supplementary Material

Supplement to "Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals"
The supplementary material contains the proofs for Section 3 and Section 4.

Algorithm 1 Successive projection algorithm (SPA)

Input: Data matrix $\mathbf{A} = [A_{ij}]_{n\times n}$, rank d

Compute the leading eigenvectors of $\mathbf{A}$: $\mathbf{A}\mathbf{U}_\mathbf{A} = \mathbf{U}_\mathbf{A}\mathbf{S}_\mathbf{A}$, where $\mathbf{U}_\mathbf{A} \in \mathbb{O}(n,d)$, $\mathbf{S}_\mathbf{A} = \mathrm{diag}(\widehat{\lambda}_1,\ldots,\widehat{\lambda}_d)$, and $|\widehat{\lambda}_1| \geq |\widehat{\lambda}_2| \geq \ldots \geq |\widehat{\lambda}_n|$.

Let $\mathbf{R} = \mathbf{U}_\mathbf{A}\mathbf{U}_\mathbf{A}^\mathrm{T}$, $J = \varnothing$, and $k = 1$.

While $\mathbf{R} \neq \mathbf{0}_{n\times n}$ and $j \leq d$

  Set $j^* \longleftarrow \arg\max_{j\in[n]} \|\mathbf{R}\mathbf{e}_j\|_2^2$. If there are ties, pick j^* to be the smallest index.

  Set $\mathbf{u}_j \longleftarrow \mathbf{R}\mathbf{e}_{j^*}$.

  Set $\mathbf{R} \longleftarrow (\mathbf{I}_n - \|\mathbf{u}_j\|_2^{-2}\mathbf{u}_j\mathbf{u}_j^\mathrm{T})\mathbf{R}$.

  Let $J \longleftarrow J \cup \{j^*\}$.

  Set $j \longleftarrow j + 1$.

End While

Output: Set of indices J.

References

[1] ABBE, E. (2017). Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.* **18** 6446–6531.

[2] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact Recovery in the Stochastic Block Model. *IEEE Trans. Inform. Theory* **62** 471–487. https://doi.org/10.1109/TIT.2015.2490670

[3] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. https://doi.org/10.1214/19-AOS1854

[4] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise Estimation of Singular Vectors of Low-Rank Matrices With Heteroskedasticity and Dependence. *IEEE Trans. Inform. Theory* **68** 4618–4650. https://doi.org/10.1109/TIT.2022.3159085

[5] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed Membership Stochastic Blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[6] ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832.

[7] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ.

[8] ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** 1–49.

[9] ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. and SUSSMAN, D. L. (2018). Statistical Inference on Random Dot Product Graphs: a Survey. *J. Mach. Learn. Res.* **18** 1–92.

[10] ATHREYA, A., PRIEBE, C. E., TANG, M., LYZINSKI, V., MARCHETTE, D. J. and SUSSMAN, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. https://doi.org/10.1007/s13171-015-0071-x

[11] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral analysis of large dimensional random matrices* **20**. Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-0661-8

[12] BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. U.S.A.* **110** 14563–14568. https://doi.org/10.1073/pnas.1307845110

[13] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. https://doi.org/10.1016/j.aim.2011.02.007

[14] BENNETT, J., LANNING, S. et al. (2007). The netflix prize. In *Proceedings of KDD cup and workshop* **2007** 35–35. https://doi.org/10.1145/1345448.1345459

[15] BHATIA, R. (1997). *Matrix analysis* **169**. Springer Science & Business Media. https://doi.org/10.1007/978-1-4612-0653-8

[16] CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. https://doi.org/10.1214/17-AOS1541

[17] CANDES, E. J. and PLAN, Y. (2011). Tight Oracle Inequalities for Low-Rank Matrix Recovery From a Minimal Number of Noisy Random Measurements. *IEEE Trans. Inform. Theory* **57** 2342–2359. https://doi.org/10.1109/TIT.2011.2111771

[18] CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. https://doi.org/10.1007/s10208-009-9045-5

[19] CANDES, E. J. and TAO, T. (2010). The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Trans. Inform. Theory* **56** 2053–2080. https://doi.org/10.1109/TIT.2010.2044061

[20] CAPE, J. (2020). Orthogonal Procrustes and norm-dependent optimality. *Electron. J. Linear Algebra* **36** 158–168. https://doi.org/10.13001/ela.2020.5009

[21] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. https://doi.org/10.1093/biomet/asy070

[22] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. https://doi.org/10.1214/18-AOS1752

[23] CHATTERJEE, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *Ann. Statist.* **43** 177–214. https://doi.org/10.1214/14-AOS1272

[24] CHUNG, K. L. (2001). *A course in probability theory*. Academic press.

[25] DAVIS, C. and KAHAN, W. M. (1970). The Rotation of Eigenvectors by a Perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. https://doi.org/10.1137/0707001

[26] DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* **52** 1289–1306. https://doi.org/10.1109/TIT.2006.871582

[27] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. https://doi.org/10.1214/14-AOS1257

[28] ELDAR, Y. C. and KUTYNIOK, G. (2012). *Compressed sensing: theory and applications*. Cambridge university press. https://doi.org/10.1017/CBO9780511794308

[29] ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: spectral analysis beyond Davis-Kahan. In *Proceedings of Algorithmic Learning Theory* (F. JANOOS, M. MOHRI and K. SRIDHARAN, eds.). *Proc. Mach. Learn. Res.* **83** 321–358. PMLR.

[30] FAN, J., FAN, Y., HAN, X. and LV, J. (2020). Asymptotic Theory of Eigenvectors for Random Matrices With Diverging Spikes. *J. Amer. Statist. Assoc.* **0** 1–14. https://doi.org/10.1080/01621459.2020.1840990

[31] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Simple: Statistical Inference on Membership Profiles in Large Networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84** 630–653. https://doi.org/10.1111/rssb.12505

[32] FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** 1–42.

[33] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** 1980–2024.

[34] GILLIS, N. and VAVASIS, S. A. (2014). Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 698–714. https://doi.org/10.1109/TPAMI.2013.226

[35] GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99** 7821–7826. https://doi.org/10.1073/pnas.122653799

[36] GOLDBERG, D., NICHOLS, D., OKI, B. M. and TERRY, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM* **35** 61–70. https://doi.org/10.1145/138859.138867

[37] GOPALAN, P. K. and BLEI, D. M. (2013). Efficient discovery of overlapping communities in massive networks. *Proc. Natl. Acad. Sci. U.S.A.* **110** 14534–14539. https://doi.org/10.1073/pnas.1221839110

[38] GROSS, D. (2011). Recovering Low-Rank Matrices From Few Coefficients in Any Basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. https://doi.org/10.1109/TIT.2011.2104999

[39] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. https://doi.org/10.1016/0378-8733(83)90021-7

[40] HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian Estimation from Few Samples: Community Detection and Related Problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* 379–390. https://doi.org/10.1109/FOCS.2017.42

[41] JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-Rank Matrix Completion Using Alternating Minimization. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*. *STOC '13* 665–674. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2488608.2488693

[42] JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. https://doi.org/10.1214/17-AOS1630

[43] JIN, J., KE, Z. T. and LUO, S. (2023). Mixed membership estimation for social networks. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2022.12.003

[44] JOHNSTONE, I. M. and LU, A. Y. (2009). On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. PMID: 20617121. https://doi.org/10.1198/jasa.2009.0121

[45] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107. https://doi.org/10.1103/PhysRevE.83.016107

[46] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078.

[47] KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. https://doi.org/10.1214/11-AOS894

[48] KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-74978-5

[49] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. https://doi.org/10.1214/14-AOS1274

[50] LEI, L. (2019). Unified $\ell_{2\to\infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv preprint:1909.04798*. https://doi.org/10.48550/arXiv.1909.04798

[51] LEVIN, K., ATHREYA, A., TANG, M., LYZINSKI, V. and PRIEBE, C. E. (2017). A central limit theorem for an omnibus embedding of random dot product graphs. *arxiv preprint:1705.09355*. https://doi.org/10.48550/arXiv.1705.09355

[52] LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. https://doi.org/10.1093/biomet/asaa006

[53] LYZINSKI, V., SUSSMAN, D. L., TANG, M., ATHREYA, A. and PRIEBE, C. E. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electron. J. Statist.* **8** 2905–2922. https://doi.org/10.1214/14-EJS978

[54] LYZINSKI, V., TANG, M., ATHREYA, A., PARK, Y. and PRIEBE, C. E. (2017). Community Detection and Classification in Hierarchical Stochastic Blockmodels. *IEEE Trans. Netw. Sci. Eng.* **4** 13–26. https://doi.org/10.1109/TNSE.2016.2634322

[55] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2017). On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations. In *Proceedings of the 34th International Conference on Machine Learning* (D. PRECUP and Y. W. TEH, eds.). *Proc. Mach. Learn. Res.* **70** 2324–2333. PMLR.

[56] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2020). Estimating Mixed Memberships With Sharp Eigenvector Deviations. *J. Amer. Statist. Assoc.* **0** 1–13. https://doi.org/10.1080/01621459.2020.1751645

[57] NEIL, J., UPHOFF, B., HASH, C. and STORLIE, C. (2013). Towards improved detection of attackers in computer networks: New edges, fast updating, and host agents. In *2013 6th International Symposium on Resilient Control Systems (ISRCS)* 218–224. https://doi.org/10.1109/ISRCS.2013.6623779

[58] NICKEL, C. L. M. (2008). Random dot product graphs a model for social networks, PhD thesis, Johns Hopkins University.

[59] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. https://doi.org/10.1016/j.jspi.2013.09.005

[60] PRIEBE, C. E., PARK, Y., TANG, M., ATHREYA, A., LYZINSKI, V., VOGELSTEIN, J. T., QIN, Y., COCANOUGHER, B., EICHLER, K., ZLATIC, M. et al. (2017). Semiparametric spectral modeling of the Drosophila connectome. *arxiv preprint:1705.03297*. https://doi.org/10.48550/arXiv.1705.03297

[61] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. https://doi.org/10.1214/11-AOS887

[62] RUBIN-DELANCHY, P., ADAMS, N. M. and HEARD, N. A. (2016). Disassortativity of computer networks. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)* 243–247. https://doi.org/10.1109/ISI.2016.7745482

[63] SARKAR, P. and BICKEL, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43** 962–990. https://doi.org/10.1214/14-AOS1285

[64] SCHÖNEMANN, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika* **31** 1–10. https://doi.org/10.1007/BF02289451

[65] SHABALIN, A. A. and NOBEL, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. https://doi.org/10.1016/j.jmva.2013.03.005

[66] JIANBO SHI and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905. https://doi.org/10.1109/34.868688

[67] STEWART, G. W. and SUN, J.-G. (1990). *Matrix Perturbation Theory*. Academic Press.

[68] SUN, Y., BARICZ, A. and ZHOU, S. (2010). On the Monotonicity, Log-Concavity, and Tight Bounds of the Generalized Marcum and Nuttall Q-Functions. *IEEE Trans. Inform. Theory* **56** 1166–1186. https://doi.org/10.1109/TIT.2009.2039048

[69] SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. https://doi.org/10.1080/01621459.2012.699795

[70] SUSSMAN, D. L., TANG, M. and PRIEBE, C. E. (2014). Consistent Latent Position Estimation and Vertex Classification for Random Dot Product Graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 48–57. https://doi.org/10.1109/TPAMI.2013.135

[71] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. https://doi.org/10.3150/15-BEJ789

[72] TANG, M. and PRIEBE, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.* **46** 2360–2415. https://doi.org/10.1214/17-AOS1623

[73] TANG, M., SUSSMAN, D. L. and PRIEBE, C. E. (2013). Universally consistent vertex classification for latent positions graphs. *Ann. Statist.* **41** 1406–1430. https://doi.org/10.1214/13-AOS1112

[74] TANG, R., KETCHA, M., BADEA, A., CALABRESE, E. D., MARGULIES, D. S., VOGELSTEIN, J. T., PRIEBE, C. E. and SUSSMAN, D. L. (2019). Connectome smoothing via low-rank approximations. *IEEE Transactions on Medical Imaging* **38** 1446–1456. https://doi.org/10.1109/TMI.2018.2885968

[75] VAN DER VAART, A. W. (2000). *Asymptotic statistics* **3**. Cambridge university press. https://doi.org/10.1017/CBO9780511802256

[76] VERSHYNIN, R. (2012). *Introduction to the non-asymptotic analysis of random matrices* In *Compressed Sensing: Theory and Applications* 210–268. Cambridge University Press. https://doi.org/10.1017/CBO9780511794308.006

[77] WASSERMAN, S. and FAUST, K. (1994). *Social network analysis: Methods and applications* **8**. Cambridge university press. https://doi.org/10.1017/CBO9780511815478

[78] WEDIN, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. https://doi.org/10.1007/BF01932678

[79] XIA, D. and ZHOU, F. (2019). The Sup-norm Perturbation of HOSVD and Low Rank Tensor Denoising. *J. Mach. Learn. Res.* **20** 1–42.

[80] XIE, F. (2023). Supplement to "Entrywise limit theorems for eigenvectors of signal-plus-noise matrix models with weak signals". *Bernoulli*.

[81] XIE, F. and XU, Y. (2020). Optimal Bayesian estimation for random dot product graphs. *Biometrika* **107** 875–889. https://doi.org/10.1093/biomet/asaa031

[82] XIE, F. and XU, Y. (2021). Efficient Estimation for Random Dot Product Graphs via a One-Step Procedure. *Journal of the American Statistical Association* **0** 1–14. https://doi.org/10.1080/01621459.2021.1948419

[83] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press Cambridge. https://doi.org/10.1017/CBO9781107588080

[84] YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random Dot Product Graph Models for Social Networks. In *Algorithms and Models for the Web-Graph* (A. BONATO and F. R. K. CHUNG, eds.) 138–149. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77004-6_11

[85] ZHONG, Y. and BOUMAL, N. (2018). Near-Optimal Bounds for Phase Synchronization. *SIAM J. Optim.* **28** 989–1016. https://doi.org/10.1137/17M1122025