Asymptotic Theory of Eigenvectors for Random Matrices with Diverging Spikes *

Jianqing Fan[1], Yingying Fan[2], Xiao Han[2] and Jinchi Lv[2]

Princeton University[1] and University of Southern California[2]

July 9, 2020

Abstract

Characterizing the asymptotic distributions of eigenvectors for large random matrices poses important challenges yet can provide useful insights into a range of statistical applications. To this end, in this paper we introduce a general framework of asymptotic theory of eigenvectors (ATE) for large spiked random matrices with diverging spikes and heterogeneous variances, and establish the asymptotic properties of the spiked eigenvectors and eigenvalues for the scenario of the generalized Wigner matrix noise. Under some mild regularity conditions, we provide the asymptotic expansions for the spiked eigenvalues and show that they are asymptotically normal after some normalization. For the spiked eigenvectors, we establish asymptotic expansions for the general linear combination and further show that it is asymptotically normal after some normalization, where the weight vector can be arbitrary. We also provide a more general asymptotic theory for the spiked eigenvectors using the bilinear form. Simulation studies verify the validity of our new theoretical results. Our family of models encompasses many popularly used ones such as the stochastic block models with or without overlapping communities for network analysis and the topic models for text analysis, and our general theory can be exploited for statistical inference in these large-scale applications.

1

Running title: Asymptotic Theory for Eigenvectors

Key words: Random matrix theory; Generalized Wigner matrix; Low-rank matrix; Eigenvectors; Spiked eigenvalues; Asymptotic distributions; Asymptotic normality; High dimensionality; Networks and texts

1 Introduction

The big data era has brought us a tremendous amount of both structured and unstructured data including networks and texts in many modern applications. For network and text data, we are often interested in learning the cluster and other structural information for the underlying network communities and text topics. In these large-scale applications, we are given a network data matrix or can create such a matrix by calculating some similarity measure between text documents, where each entry of the data matrix is binary indicating the absence or presence of a link, or continuous indicating the strength of similarity between each pair of nodes or documents. Such applications naturally give rise to random matrices that can be used to reveal interesting latent structures of networks and texts for effective predictions and recommendations.

Random matrix has been widely exploited to model the interactions among the nodes of a network for applications ranging from physics and social sciences to genomics and neuroscience. Random matrix theory (RMT) has a long history and was originated by Wigner in Wigner (1955) for modeling the nucleon-nucleus interactions to understand the behavior of atomic nuclei and link the spacings of the levels of atomic nuclei to those of the eigenvalues of a random matrix. See, for example, Bai (1999) for a review of some classical technical tools such as the moment method and Stieltjes transform as well as some more recent developments on the RMT, and Mehta (2004); Tao (2004); Bai and Silverstein (2006) for detailed book-length accounts of the topic of random matrices.

There is a rich recent literature in mathematics on the asymptotic behaviors of eigenvalues and eigenvectors of random matrices (Erdős et al., 2013; Bourgade et al., 2018; Bourgade and Yau, 2017; Rudelson and Vershynin, 2016; Dekel et al., 2007). The main challenge in many RMT problems is caused by the strong dependence of eigenvalues if they are close to each other. Using the terminologies in RMT, four regimes are often of interests: bulk, subcritical edge, critical edge, and supercritical regimes. The first three regimes all have eigenvalues that are highly correlated with each other, and the last regime has weaker interactions among the eigenvalues. The last regime can be further divided into two categories according to the relative strength of spiked eigenvalues compared to noise, which can be roughly understood as the signal-to-noise ratio. There have been exciting mathematical developments in the recent mathematical literature when the smallest spiked eigenvalue has the same order as the noise (Capitaine and Donati-Martin, 2018; Knowles and Yin, 2013; Bao et al., 2018). Due to the challenge caused by constant signal-to-noise ratio, these existing results often take

complicated forms and the asymptotic distributions depend generally on the noise matrix distribution in a complex way, limiting their practical usage to statisticians. In this paper, we consider the setting of *diverging spikes* where the spiked eigenvalues are an order of magnitude larger than the noise level asymptotically. Although mathematically easier, such random matrices are of great interests to statisticians, because many statistical applications such as network analysis and text analysis often fall into this regime. Yet there lack any formal results on the asymptotic expansions and asymptotic distributions of spiked eigenvectors even in this setting. This motivates our study in this paper.

There is a larger literature on the limiting distributions of eigenvalues than eigenvectors in RMT. For instance, the limiting spectral distribution of the Wigner matrix was generalized by Arnold (1967) and Arnold (1971). Marchenko and Pastur (1967) established the well-known Marchenko–Pastur law for the limiting spectral distribution of the sample covariance matrix including the Wishart matrix which plays an important role in statistical applications. In contrast, the asymptotic distribution of the largest nonspiked eigenvalue of Wigner matrix with Gaussian ensemble was revealed to be the Tracy–Widom law in Tracy and Widom (1994) and Tracy and Widom (1996). More recent developments on the asymptotic distribution of the largest nonspiked eigenvalue include Johnstone (2001), El Karoui (2007), Johnstone (2008), Erdös et al. (2011), and Knowles and Yin (2017). See also Füredi and Komlós (1981), Baik et al. (2005), Bai and Yao (2008), Knowles and Yin (2013), Pizzo et al. (2013), Renfrew and Soshnikov (2013), Knowles and Yin (2014), and Wang and Fan (2017) for the asymptotic distributions of the spiked eigenvalues of various random matrices and sample covariance matrices. For the eigenvectors, Capitaine and Donati-Martin (2018) and Bao et al. (2018) established their asymptotic distributions, which depend on the specific distribution of the Wigner matrix in a complicated way, in the challenging setting of constant signal-to-noise ratio. There is also a growing literature on the specific scenario and applications of large network matrices. To ensure consistency, Johnstone and Lu (2009) proposed the sparse principal component analysis to reduce the noise accumulation in high-dimensional random matrices. See, for example, McSherry (2001), Spielman and Teng (2007), Bickel and Chen (2009), Decelle et al. (2011), Rohe et al. (2011), Lei (2016), Abbe (2017), Jin et al. (2017), Chen and Lei (2018), and Vu (2018).

Matrix perturbation theory has been commonly used to characterize the deviations of empirical eigenvectors from the population ones, often under the average errors (Horn and Johnson, 2012). In contrast, recently Fan et al. (2018) and Abbe et al. (2019) investigated random matrices with low expected rank and provided a tight bound for the difference between the empirical eigenvector and some linear transformation of the population eigenvector through a delicate entrywise eigenvector analysis for the first-order approximation under the maximum norm. See also Paul (2007), Koltchinskii and Lounici (2016), Koltchinskii and Xia (2016), and Wang and Fan (2017) for the asymptotics of empirical eigenstructure for

3

large random matrices. Yet despite these endeavors, the precise asymptotic distributions of the eigenvectors for large spiked random matrices still remain largely unknown even for the case of Wigner matrix noise. Indeed characterizing the exact asymptotic distributions of eigenvectors in such setting can provide useful insights into a range of statistical applications that involve the eigenspaces. In this sense, the asymptotic expansions and asymptotic distributions of eigenvectors established in this paper complement the existing work in the statistics literature.

The major contribution of this paper is introducing a general framework of asymptotic theory of eigenvectors (ATE) for large spiked random matrices with diverging spikes, where the mean matrix is low-rank and the noise matrix is the generalized Wigner matrix. The generalized Wigner matrix refers to a symmetric random matrix whose diagonal and upper diagonal entries are independent with zero mean, allowing for heterogeneous variances. Our family of models includes a variety of popularly used ones such as the stochastic block models with or without overlapping communities for network analysis and the topic models for text analysis. Under some mild regularity conditions, we establish the asymptotic expansions for the spiked eigenvalues and prove that they are asymptotically normal after some normalization. For the spiked eigenvectors, we provide asymptotic expansions for the general linear combination and further establish that it is asymptotically normal after some normalization for arbitrary weight vector. We also present a more general asymptotic theory for the spiked eigenvectors based on the bilinear form. To the best of our knowledge, these theoretical results are new to the literature. Our general theory can be exploited for statistical inference in a range of large-scale applications including network analysis and text analysis. For detailed comparisons with the literature, see Section 3.6.

The rest of the paper is organized as follows. Section 2 presents the model setting and theoretical setup for ATE. We establish the asymptotic expansions and asymptotic distributions for the spiked eigenvectors as well as the asymptotic distributions for the spiked eigenvalues in Section 3. Several specific statistical applications of our new asymptotic theory are discussed in Section 4. Section 5 presents some numerical examples to demonstrate our theoretical results. We further provide a more general asymptotic theory extending the results from Section 3 using the bilinear form in Section 6. Section 7 discusses some implications and extensions of our work. The proofs of main results are relegated to the Appendix. Additional technical details are provided in the Supplementary Material.

2 Model setting and theoretical setup

2.1 Model setting

As mentioned in the introduction, we focus on the class of large spiked symmetric random matrices with low-rank mean matrices and generalized Wigner matrices of noises. It is

worth mentioning that our definition of the generalized Wigner matrix specified in Section 1 is broader than the conventional one in the classical RMT literature; see, for example, Yau (2012) for the formal mathematical definition with additional assumptions. To simplify the technical presentation, consider an $n \times n$ symmetric random matrix with the following structure

$$\mathbf{X} = \mathbf{H} + \mathbf{W}, \tag{1}$$

where $\mathbf{H} = \mathbf{V}\mathbf{D}\mathbf{V}^T$ is a deterministic latent mean matrix of low rank structure, $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_K)$ is an $n \times K$ orthonormal matrix of population eigenvectors $\mathbf{v}_k$'s with $\mathbf{V}^T\mathbf{V} = \mathbf{I}_K$, $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_K)$ is a diagonal matrix of population eigenvalues d_k's with $|d_1| \geq \cdots \geq |d_K| > 0$, and $\mathbf{W} = (w_{ij})_{1 \leq i,j \leq n}$ is a symmetric random matrix of independent noises on and above the diagonal with zero mean $\mathbb{E}w_{ij} = 0$, variances $\sigma_{ij}^2 = \mathbb{E}w_{ij}^2$, and $\max_{1 \leq i,j \leq n} |w_{ij}| \leq 1$. The rank K of the mean part is assumed typically to be a smaller order of the random matrix size n, which is referred to as matrix dimensionality hereafter for convenience. The bounded assumption on w_{ij} is made frequently for technical simplification and satisfied in many real applications such as network analysis and text analysis. It can be relaxed to $\mathbb{E}|w_{ij}|^l \leq C^{l-2}\mathbb{E}|w_{ij}|^2$, $l \geq 2$, with C some positive constant, and all the proofs and results can carry through.

In practice, it is either matrix $\mathbf{X}$ or matrix $\mathbf{X} - \mathrm{diag}(\mathbf{X})$ that is readily available to us, where $\mathrm{diag}(\cdot)$ denotes the diagonal part of a matrix. In the context of graphs, random matrix $\mathbf{X}$ characterizes the connectivity structure of a graph with self loops, while random matrix $\mathbf{X} - \mathrm{diag}(\mathbf{X})$ corresponds to a graph without self loops. In the latter case, the observed data matrix can be decomposed as

$$\mathbf{X} - \mathrm{diag}(\mathbf{X}) = \mathbf{H} + [\mathbf{W} - \mathrm{diag}(\mathbf{X})]. \tag{2}$$

Observe that $\mathbf{W} - \mathrm{diag}(\mathbf{X})$ has the similar structure as $\mathbf{W}$ in the sense of being symmetric and having bounded independent entries on and above the diagonal, by assuming that $\mathrm{diag}(\mathbf{X})$ has bounded entries for such a case. Thus models (1) and (2) share the same decomposition of a deterministic low rank matrix plus some symmetric noise matrix of bounded entries, which is roughly all we need for the theoretical framework and technical analysis. For these reasons, to simplify the technical presentation we abuse slightly the notation by using $\mathbf{X}$ and $\mathbf{W}$ to represent the observed data matrix and the latent noise matrix, respectively, in either model (1) or model (2). Therefore, throughout the paper the data matrix $\mathbf{X}$ may have diagonal entries all equal to zero and correspondingly the noise matrix $\mathbf{W}$ may have a nonzero diagonal mean matrix, and our theory covers both cases.

In either of the two scenarios discussed above, we are interested in inferring the structural information in models (1) and (2), which often boils down to the latent eigenstructure $(\mathbf{D}, \mathbf{V})$.

Since both the eigenvector matrix $\mathbf{V}$ and eigenvalue matrix $\mathbf{D}$ are unavailable to us, we resort to the observable random data matrix $\mathbf{X}$ for extracting the structural information. To this end, we conduct a spectral decomposition of $\mathbf{X}$, and denote by $\lambda_1, \cdots, \lambda_n$ its eigenvalues and $\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_n$ the corresponding eigenvectors. Without loss of generality, assume that $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and denote by $\widehat{\mathbf{V}} = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K)$ an $n \times K$ matrix of spiked eigenvectors. As mentioned before, we aim at investigating the precise asymptotic behavior of the spiked empirical eigenvalues $\lambda_1, \cdots, \lambda_K$ and spiked empirical eigenvectors $\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K$ of data matrix $\mathbf{X}$. It is worth mentioning that our definition of spikedness differs from the conventional one in that the underlying rank order depends on the magnitude of eigenvalues instead of the nonnegative eigenvalues that are usually assumed.

One concrete example is the stochastic block model (SBM), where the latent mean matrix $\mathbf{H}$ takes the form $\mathbf{H} = \mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^T$ with $\mathbf{\Pi} = (\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_n)^T \in \mathbb{R}^{n \times K}$ a matrix of community membership vectors and $\mathbf{P} = (p_{kl}) \in \mathbb{R}^{K \times K}$ a nonsingular matrix with $p_{kl} \in [0, 1]$ for $1 \leq k, l \leq K$. Here, for each $1 \leq i \leq n$, $\boldsymbol{\pi}_i \in \{\mathbf{e}_1, \cdots, \mathbf{e}_K\}$ with $\mathbf{e}_j \in \mathbb{R}^K$, $1 \leq j \leq K$, a unit vector with the kth component being one and all other components being zero. It is well known that the community information of the SBM is encoded completely in the eigenstructure of the mean matrix $\mathbf{H}$, which serves as one of our motivations for investigating the precise asymptotic distributions of the empirical eigenvectors and eigenvalues.

2.2 Theoretical setup

We first introduce some notation that will be used throughout the paper. We use $a \ll b$ to represent $a/b \to 0$ as matrix size n increases. We say that an event $\mathcal{E}_n$ holds with significant probability if $\mathbb{P}(\mathcal{E}_n) = 1 - O(n^{-l})$ for some positive constant l and sufficiently large n. For a matrix $\mathbf{A}$, we use $\lambda_j(\mathbf{A})$ to denote the jth largest eigenvalue in magnitude, and $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$, and $\|\mathbf{A}\|_\infty$ to denote the Frobenius norm, the spectral norm, and the matrix entrywise maximum norm, respectively. Denote by $\mathbf{A}_{-k}$ the submatrix of $\mathbf{A}$ formed by removing the kth column. For any n-dimensional unit vector $\mathbf{x} = (x_1, \cdots, x_n)^T$, let $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$ represent the maximum norm of the vector.

We next introduce a definition that plays a key role in proving all asymptotic normality results in this paper.

Definition 1. *A pair of unit vectors* $(\mathbf{x}, \mathbf{y})$ *of appropriate dimensions is said to satisfy the* $\mathbf{W}^l$*-CLT condition for some positive integer* l *if* $\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}$ *is asymptotically standard normal after some normalization, where CLT refers to the central limit theorem.*

Lemmas 1 and 2 below provide some sufficient conditions under which $(\mathbf{x}, \mathbf{y})$ can satisfy the $\mathbf{W}^l$-CLT condition defined in Definition 1 for $l = 1$ and 2, which is all we need for our technical analysis of asymptotic distributions. In this paper, we apply these lemmas with either $\mathbf{x}$ or $\mathbf{y}$ equal to $\mathbf{v}_k$. Therefore, a sufficient condition for the results in our paper is that $\|\mathbf{v}_k\|_\infty$ is small enough.

6

Lemma 1. *Assume that n-dimensional unit vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *satisfy*

$$\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \ll \left[\operatorname{var}(\mathbf{x}^T \mathbf{W} \mathbf{y})\right]^{1/2} = s_n. \tag{3}$$

Then $\mathbf{x}^T \mathbf{W} \mathbf{y}$ *satisfies the Lyapunov condition for CLT and we have* $(\mathbf{x}^T \mathbf{W} \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W} \mathbf{y})/s_n \xrightarrow{\mathscr{D}} N(0,1)$ *as* $n \to \infty$, *which entails that* $(\mathbf{x}, \mathbf{y})$ *satisfies the* $\mathbf{W}^l$-*CLT condition with* $l = 1$.

To introduce $\mathbf{W}^2$-CLT, for any given unit vectors $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$, we denote respectively $s_{\mathbf{x},\mathbf{y}}^2$ and $\kappa_{\mathbf{x},\mathbf{y}}$ the mean and variance of the random variable

$$\sum_{1 \le k,i \le n,\, k \le i} \sigma_{ki}^2 \Big[\sum_{1 \le l < k \le n} w_{il}(x_k y_l + y_k x_l) + \sum_{1 \le l < i \le n} w_{kl}(x_i y_l + y_i x_l)$$
$$+ (1 - \delta_{ki})\mathbb{E}w_{ii}(x_i y_k + x_k y_i) \Big]^2 + 2 \sum_{1 \le k,i \le n,\, k \le i} \gamma_{ki}(x_k y_k + x_i y_i)$$
$$\times \Big[\sum_{1 \le l < k \le n} w_{il}(x_k y_l + y_k x_l) + \sum_{1 \le l < i \le n} w_{kl}(x_i y_l + y_i x_l)$$
$$+ (1 - \delta_{ki})\mathbb{E}w_{ii}(x_i y_k + x_k y_i) \Big] + \sum_{1 \le k,i \le n,\, k \le i} \kappa_{ki}(x_k y_k + x_i y_i)^2, \tag{4}$$

where $\gamma_{ki} = \mathbb{E}w_{ki}^3$ and $\kappa_{ki} = \mathbb{E}(w_{ki}^2 - \sigma_{ki}^2)^2$ for $k \ne i$, $\gamma_{kk} = 2(\mathbb{E}\omega_{kk}^3 - \sigma_{kk}^2 \mathbb{E}\omega_{kk})$, $\kappa_{kk} = 4\mathbb{E}(\omega_{kk}^2 - \sigma_{kk}^2)^2$ with $\omega_{kk} = 2^{-1}w_{kk}$, $\sigma_{kk}^2 = \mathbb{E}\omega_{kk}^2$, and $\delta_{ki} = 1$ when $k = i$ and 0 otherwise. It is worth mentioning that the random variable given in (4) coincides with the one defined in (A.7) in Section B.2 of Supplementary Material, which is simply the conditional variance of random variable $\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y}$ given in (A.5) when expressed as a sum of martingale differences with respect to a suitably defined σ-algebra; see Section B.2 for more technical details and the precise expressions for $s_{\mathbf{x},\mathbf{y}}^2$ and $\kappa_{\mathbf{x},\mathbf{y}}$ given in (A.8) and (A.9), respectively.

Lemma 2. *Assume that n-dimensional unit vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *satisfy* $\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \to 0$, $\kappa_{\mathbf{x},\mathbf{y}}^{1/4} \ll s_{\mathbf{x},\mathbf{y}}$, *and* $s_{\mathbf{x},\mathbf{y}} \to \infty$. *Then we have* $[\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y}]/s_{\mathbf{x},\mathbf{y}} \xrightarrow{\mathscr{D}} N(0,1)$ *as* $n \to \infty$, *which entails that* $(\mathbf{x}, \mathbf{y})$ *satisfies the* $\mathbf{W}^2$-*CLT condition.*

Remark 1. *To provide more insights into the conditions of Lemmas 1 and 2, we discuss the special case of standard Wigner matrix where* $\sigma_{ij}^2 = p(1-p)$ *with* p *the expected value of entries of* $\mathbf{X}$. *Then* $s_n^2 := \operatorname{var}(\mathbf{x}^T \mathbf{W} \mathbf{y}) \in [p(1-p), 2p(1-p)]$ *and condition (3) in Lemma 1 reduces to*

$$\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \ll \sqrt{\operatorname{var}(\mathbf{x}\mathbf{W}\mathbf{y})} \sim \sqrt{p(1-p)}.$$

Moreover, (A.13) in the Supplementary Material ensures that Lemma 2 holds under the following sufficient conditions

$$\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \to 0, \quad n^{3/2}p(1-p)\|\mathbf{x}\|_\infty^2 \|\mathbf{y}\|_\infty^2 \to 0, \quad \text{and } p(1-p)n \to \infty. \tag{5}$$

Thus if either $\|\mathbf{x}\|_\infty$ *or* $\|\mathbf{y}\|_\infty$ *is small enough, both lemmas hold. Indeed in this scenario,*

direct calculations show that $\mathbf{s}_{\mathbf{x},\mathbf{y}}^2 \sim np(1-p)$.

We see from Lemmas 1 and 2 that the $\mathbf{W}^l$-CLT condition defined in Definition 1 can indeed be satisfied under some mild regularity conditions. In particular, Definition 1 is important to our technical analysis since to establish the asymptotic normality of the spiked eigenvectors and spiked eigenvalues, we first need to expand the target to the form of $\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}$ with l some positive integer plus some small order term, and then the asymptotic normality follows naturally if $(\mathbf{x}, \mathbf{y})$ satisfies the $\mathbf{W}^l$-CLT condition. To facilitate our technical presentation, let us introduce some further notation. For any $t \neq 0$ and given matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of appropriate dimensions, we define the function

$$\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, t) = -\sum_{l=0, l\neq 1}^{L} t^{-(l+1)}\mathbf{M}_1^T \mathbb{E}\mathbf{W}^l \mathbf{M}_2, \tag{6}$$

where L is some sufficiently large positive integer that will be specified later in our technical analysis. For each $1 \leq k \leq K$, any given matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ of appropriate dimensions, and n-dimensional vector $\mathbf{u}$, we further define functions

$$\mathcal{P}(\mathbf{M}_1, \mathbf{M}_2, t) = t\mathcal{R}(\mathbf{M}_1, \mathbf{M}_2, t), \quad \widetilde{\mathcal{P}}_{k,t} = \left[t^2(A_{\mathbf{v}_k,k,t}/t)'\right]^{-1}, \tag{7}$$

$$\mathbf{b}_{\mathbf{u},k,t} = \mathbf{u} - \mathbf{V}_{-k}\left[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t)\right]^{-1}\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t)^T, \tag{8}$$

where $\mathbf{D}_{-k}$ denotes the submatrix of the diagonal matrix $\mathbf{D}$ by removing the kth row and kth column,

$$A_{\mathbf{u},k,t} = \mathcal{P}(\mathbf{u}, \mathbf{v}_k, t) - \mathcal{P}(\mathbf{u}, \mathbf{V}_{-k}, t)\left[t(\mathbf{D}_{-k})^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t)\right]^{-1}\mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t), \tag{9}$$

$(\cdot)'$ denotes the derivative with respect to scalar t or complex variable z throughout the paper, and the rest of notation is the same as introduced before.

3 Asymptotic distributions of spiked eigenvectors

3.1 Technical conditions

To facilitate our technical analysis, we need some basic regularity conditions.

Condition 1. *Assume that* $\alpha_n = \|\mathbb{E}(\mathbf{W} - \mathbb{E}\mathbf{W})^2\|^{1/2} \to \infty$ *as* $n \to \infty$.

Condition 2. *There exists a positive constant* $c_0 < 1$ *such that* $\min\{|d_i|/|d_j| : 1 \leq i < j \leq K+1, d_i \neq -d_j\} \geq 1 + c_0$. *In addition, either of the following two conditions holds:*

   i) $|d_K|/(n^\epsilon \alpha_n) \to \infty$ *with some small positive constant* ϵ,

   ii) $\max_{i,j} \mathrm{var}(w_{ij}) \leq (c_1^2 \alpha_n^2)/n$ *and* $|d_K| > c\alpha_n \log n$ *with some constants* $c_1 \geq 1$ *and* $c > 4c_1(1 + 2^{-1}c_0)$.

Condition 3. *It holds that* $|d_1| = O(|d_K|)$, $|d_K|\sigma_{\min}/\alpha_n \to \infty$, $\|\mathbf{v}_k\|_\infty^2/\sigma_{\min} \to 0$, $\alpha_n^4 \|\mathbf{v}_k\|_\infty^4/$ $(\sqrt{n}\sigma_{\min}^2) \to 0$, *and* $\sigma_{\min}^2 n \to \infty$, *where* $\sigma_{\min} = \{\min_{1 \le i,j \le n,\, i \ne j} \mathbb{E}w_{ij}^2\}^{1/2}$.

Conditions 1–2 are needed in all our Theorems 1–5 and imposed for our general model (1), including the specific case of sparse models. In contrast, condition 3 is required *only* for Theorem 3 under some specific models with dense structures such as the stochastic block models with or without overlapping communities.

Condition 1 restricts essentially the sparsity level of the random matrix (e.g., given by a network). Note that it follows easily from $\max_{1 \le i,j \le n} |w_{ij}| \le 1$ that $\alpha_n \le n^{1/2}$. It is a rather mild condition that can be satisfied by very sparse networks. For example, if $\mathbb{E}w_{12}^2 = \cdots = \mathbb{E}w_{1\lfloor \log n \rfloor}^2 = 1/2$ and the other w_{1j}'s are equal to zero, then we have $\alpha_n^2 \ge 2^{-1} \log n \to \infty$. Many network models in the literature satisfy this condition; see, for example, Jin et al. (2017), Lei (2016), and Zhang et al. (2015).

Condition 2 requires that the spiked population eigenvalues of the mean matrix $\mathbf{H}$ (in the diagonal matrix $\mathbf{D}$) are simple and there is enough gap between the eigenvalues. The constant c_0 can be replaced by some $o(1)$ term and our theoretical results can still be proved with more delicate derivations. This requirement ensures that we can obtain higher order expansions of the general linear combination for each empirical eigenvector precisely. Otherwise if there exist some eigenvalues such that $d_i = d_{i+1}$, then $\widehat{\mathbf{v}}_i$ and $\widehat{\mathbf{v}}_{i+1}$ are generally no longer identifiable so we cannot derive clear asymptotic expansions for them; see also Abbe et al. (2019) for related discussions. Condition 2 also requires a gap between α_n and $|d_K|$. Since parameter α_n reflects the strength of the noise matrix $\mathbf{W}$, it requires essentially the signal part $\mathbf{H}$ to dominate the noise part $\mathbf{W}$ with some asymptotic rate. Similar condition is used commonly in the network literature; see, for instance, Abbe et al. (2019) and Jin et al. (2017).

Condition 3 restricts our attention to some specific dense network models. In particular, $|d_1| = O(|d_K|)$ assumes that the eigenvalues in $\mathbf{D}$ share the same order. The other assumptions in Condition 3 require essentially that the minimum variance of the off-diagonal entries of $\mathbf{W}$ cannot tend to zero too fast, which is used only to establish a more simplified theory under the more restrictive model; see Theorem 3.

3.2 Asymptotic distributions of spiked eigenvalues

We first present the asymptotic expansions and CLT for the spiked empirical eigenvalues $\lambda_1, \cdots, \lambda_K$. For each $1 \le k \le K$, denote by t_k the solution to equation

$$f_k(z) = 1 + d_k \Big\{ \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \left[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1}$$
$$\times \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \Big\} = 0 \tag{10}$$

when restricted to the interval $z \in [a_k, b_k]$, where

$$a_k = \begin{cases} d_k/(1 + 2^{-1}c_0), & d_k > 0 \\ (1 + 2^{-1}c_0)d_k, & d_k < 0 \end{cases} \quad \text{and} \quad b_k = \begin{cases} (1 + 2^{-1}c_0)d_k, & d_k > 0 \\ d_k/(1 + 2^{-1}c_0), & d_k < 0 \end{cases}.$$

The following lemma characterizes the properties of the population quantity t_k's defined in (10): It is unique and the asymptotic mean of λ_k.

Lemma 3. *Equation* (10) *has a unique solution in the interval* $z \in [a_k, b_k]$ *and thus* t_k's *are well defined. Moreover, for each* $1 \leq k \leq K$ *we have* $t_k/d_k \to 1$ *as* $n \to \infty$.

It is seen from Lemma 3 that when the matrix size n is large enough, the values of t_k and d_k are very close to each other. The following theorem establishes the asymptotic expansions and CLT for λ_k and reveals that t_k is in fact its asymptotic mean.

Theorem 1. *Under Conditions* 1–2, *for each* $1 \leq k \leq K$ *we have*

$$\lambda_k - t_k = \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n d_k^{-1}). \tag{11}$$

Moreover, if $\mathrm{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) \gg \alpha_n^2 d_k^{-2}$ *and the pair of vectors* $(\mathbf{v}_k, \mathbf{v}_k)$ *satisfies the* $\mathbf{W}^1$-*CLT condition, then we have*

$$\frac{\lambda_k - t_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k}{\left[\mathrm{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)\right]^{1/2}} \xrightarrow{\mathscr{D}} N(0, 1). \tag{12}$$

Capitaine et al. (2012) and Knowles and Yin (2014) established the joint distribution of the spiked eigenvalues for the deformed Wigner matrix in different settings than ours. Capitaine et al. (2012) assumed that $\mathbb{E}w_{ii}^2 = 1/2$ and $\mathbb{E}w_{ij}^2 = 1$ for $i \neq j$, while Knowles and Yin (2014) assumed that $\mathbb{E}w_{ij}^2 = 1$ for all (i, j). Under their model settings, the smallest spiked eigenvalue $|d_K|$ and the noise level α_n are of the same order, and as a result, their asymptotic distributions depend on the distributions of the Wigner matrix. In contrast, our Theorem 1 is proved in the setting of diverging spikes. Thanks to the stronger signal-to-noise ratio, the noise matrix contributes to the distributions of the spiked eigenvalues in Theorem 1 in a global way, allowing for more heterogeneity in the variances of entries of the noise matrix $\mathbf{W}$.

Theorem 1 requires that $(\mathbf{v}_k, \mathbf{v}_k)$ satisfies the $\mathbf{W}^1$-CLT condition and $\mathrm{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) \gg \alpha_n^2 d_k^{-2}$. To gain some insights into these two conditions, we will provide some sufficient conditions for such assumptions. Let us consider the specific case of $\sigma_{\min} > 0$, that is, the generalized Wigner matrix $\mathbf{W}$ is nonsparse. We will show that as long as

$$\|\mathbf{v}_k\|_\infty^2 \sigma_{\min}^{-1} \to 0 \quad \text{and} \quad \sigma_{\min} \gg \alpha_n |d_k|^{-1}, \tag{13}$$

the aforementioned two conditions in Theorem 1 hold. We first verify the $\mathbf{W}^1$-CLT condition. By Lemma 1, a sufficient condition for $(\mathbf{v}_k, \mathbf{v}_k)$ to satisfy the $\mathbf{W}^1$-CLT condition is that

$$\|\mathbf{v}_k\|_\infty^2 \ll \left[\operatorname{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)\right]^{1/2} = \left[\mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2\right]^{1/2}. \tag{14}$$

Observe that it follows from $\sum_{1 \leq i \leq n}(\mathbf{v}_k)_i^2 = \mathbf{v}_k^T \mathbf{v}_k = 1$ and $\sum_{1 \leq i \leq n}(\mathbf{v}_k)_i^4 \leq \|\mathbf{v}_k\|_\infty^2 \leq 1$ that

$$\begin{aligned}
\left[\mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2\right]^{1/2} &\geq \left[2 \sum_{1 \leq i,j \leq n,\, i \neq j} \sigma_{ij}^2 (\mathbf{v}_k)_i^2 (\mathbf{v}_k)_j^2\right]^{1/2} \\
&\geq \sigma_{\min}\left[2 \sum_{1 \leq i,j \leq n,\, i \neq j} (\mathbf{v}_k)_i^2 (\mathbf{v}_k)_j^2\right]^{1/2} = \sigma_{\min}\left[2 - 2\sum_{1 \leq i \leq n}(\mathbf{v}_k)_i^4\right]^{1/2} \\
&\geq \sigma_{\min}\left(2 - 2\|\mathbf{v}_k\|_\infty^2\right)^{1/2}, \tag{15}
\end{aligned}$$

where $(\mathbf{v}_k)_i$ stands for the ith component of vector $\mathbf{v}_k$. The assumption $\|\mathbf{v}_k\|_\infty^2 \sigma_{\min}^{-1} \to 0$ in (13) together with (15) ensures (14), which consequently entails that $(\mathbf{v}_k, \mathbf{v}_k)$ satisfies the $\mathbf{W}^1$-CLT condition.

We next check the condition $\operatorname{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) = \mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 \gg \alpha_n^2 d_k^{-2}$. It follows directly from (15) that this condition holds under (13). In fact, since Condition 2 guarantees that $\alpha_n/|d_k|$ asymptotically vanishes, the assumption $\sigma_{\min} \gg \alpha_n |d_k|^{-1}$ can be very mild. In particular, for the Wigner matrix $\mathbf{W}$ with $\sigma_{ij} \equiv 1$ for all $1 \leq i, j \leq n$, it holds that

$$\mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 = 2. \tag{16}$$

Thus the condition of $\mathbb{E}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k)^2 \gg \alpha_n^2 d_k^{-2}$ reduces to that of $\alpha_n^2 d_k^{-2} \ll 1$, which is guaranteed to hold under Condition 2.

We also would like to point out that one potential application of the new results in Theorem 1 is determining the number of spiked eigenvalues, which in the network models reduces to determining the number of non-overlapping (or possibly overlapping) communities or clusters.

3.3 Asymptotic distributions of spiked eigenvectors

We now present the asymptotic distributions of the spiked empirical eigenvectors $\widehat{\mathbf{v}}_k$ for $1 \leq k \leq K$. To this end, we will first establish the asymptotic expansions and CLT for the bilinear form

$$\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$$

with $1 \leq k \leq K$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are two arbitrary non-random unit vectors. Then by setting $\mathbf{y} = \mathbf{v}_k$, we can establish the asymptotic expansions and CLT for the general

linear combination $\mathbf{x}^T \widehat{\mathbf{v}}_k$. Although the limiting distribution of the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ is the theoretical foundation for establishing the limiting distribution of the general linear combination $\mathbf{x}^T \widehat{\mathbf{v}}_k$, due to the technical complexities we will defer the theorems summarizing the limiting distribution of $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ to a later technical section (i.e., Section 6), and present only the results for $\mathbf{x}^T \widehat{\mathbf{v}}_k$ in this section. This should not harm the flow of the paper. For readers who are also interested in our technical proofs, they can refer to Section 6 for more technical details; otherwise it is safe to skip that technical section. For each $1 \leq k \leq K$, let us choose the direction of $\widehat{\mathbf{v}}_k$ such that $\mathbf{v}_k^T \widehat{\mathbf{v}}_k \geq 0$ for the theoretical derivations, which is always possible after a sign change when needed.

Theorem 2. *Under Conditions 1–2, for each $1 \leq k \leq K$ we have the following properties:*

1) If the unit vector $\mathbf{u}$ satisfies that $|\mathbf{u}^T \mathbf{v}_k| \in [0,1)$ and $\alpha_n^{-2} d_k^2 \mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \to \infty$, then it holds that

$$t_k \left(\mathbf{u}^T \widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right) = (\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k$$
$$+ o_p \left(\left\{ \mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \right\}^{1/2} \right), \tag{17}$$

where the asymptotic mean has the expansion $A_{\mathbf{u},k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} = -\mathbf{u}^T \mathbf{v}_k + O(\alpha_n^2 d_k^{-2})$. Furthermore, if $(\mathbf{b}_{\mathbf{u},k,t_k} - \mathbf{v}_k \mathbf{v}_k^T \mathbf{u}, \mathbf{v}_k)$ satisfies the $\mathbf{W}^1$-CLT condition, then it holds that

$$\frac{t_k \left(\mathbf{u}^T \widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right) - \mathbb{E}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k \right]}{\left\{ \mathrm{var}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k \right] \right\}^{1/2}} \xrightarrow{\mathscr{D}} N(0,1).$$

2) If $(\alpha_n^{-4} d_k^2 + 1)\mathrm{var}(\mathbf{v}_k^T \mathbf{W}^2 \mathbf{v}_k) \to \infty$, then it holds that

$$2t_k^2 \left(\mathbf{v}_k^T \widehat{\mathbf{v}}_k + A_{\mathbf{v}_k,k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right) = -\mathbf{v}_k^T \left(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2 \right) \mathbf{v}_k + o_p \left\{ \left[\mathrm{var}(\mathbf{v}_k^T \mathbf{W}^2 \mathbf{v}_k) \right]^{1/2} \right\}, \tag{18}$$

where the asymptotic mean has the expansion $A_{\mathbf{v}_k,k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} = -1 + 2^{-1} t_k^{-2} \mathbf{v}_k^T \mathbb{E}\mathbf{W}^2 \mathbf{v}_k + O(\alpha_n^3 d_k^{-3})$. Furthermore, if $(\mathbf{v}_k, \mathbf{v}_k)$ satisfies the $\mathbf{W}^2$-CLT condition, then it holds that

$$\frac{2t_k^2 \left(\mathbf{v}_k^T \widehat{\mathbf{v}}_k + A_{\mathbf{v}_k,k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right)}{\left[\mathrm{var}\left(\mathbf{v}_k^T \mathbf{W}^2 \mathbf{v}_k \right) \right]^{1/2}} \xrightarrow{\mathscr{D}} N(0,1).$$

The two parts of Theorem 2 correspond to two different cases when $\mathrm{var}(\mathbf{u}^T \widehat{\mathbf{v}}_k)$ can be of different magnitude. To understand this, note that for large enough matrix size n, we have $|t_K| \gg \alpha_n$ by Condition 2 and Lemma 3. In view of (18), the asymptotic variance of $\mathbf{v}_k^T \widehat{\mathbf{v}}_k$ is equal to $\mathrm{var}(2^{-1} t_k^{-2} \mathbf{v}_k^T \mathbf{W}^2 \mathbf{v}_k)$. In contrast, in light of (17), the asymptotic variance of $\mathbf{u}^T \widehat{\mathbf{v}}_k$ with $|\mathbf{u}^T \mathbf{v}_k| \in [0,1)$ is equal to $\mathrm{var}\left[t_k^{-1}(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k \right]$. Let us consider a specific

case when $\text{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \sim 1$. By Lemma 4 in Section 6, we have

$$\text{var}\left(2^{-1}t_k^{-2}\mathbf{v}_k^T\mathbf{W}^2\mathbf{v}_k\right) = O\left(\alpha_n^2 t_k^{-4}\right) \ll \text{var}\left[t_k^{-1}(\mathbf{b}_{u,k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right] = O\left(t_k^{-2}\right).$$

This shows that the above two cases can be very different in the magnitude for the asymptotic variance of $\mathbf{u}^T\widehat{\mathbf{v}}_k$ and thus should be analyzed separately.

To gain some insights into why $\mathbf{v}_k^T\widehat{\mathbf{v}}_k$ has smaller variance, let us consider the simple case of $K = 1$. Then in view of our technical arguments, it holds that

$$
\begin{aligned}
\mathbf{v}_1^T\widehat{\mathbf{v}}_1\widehat{\mathbf{v}}_1^T\mathbf{v}_1 &= -\frac{1}{2\pi i}\oint_{\Omega_1} \frac{\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1}{1 + d_1\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1}dz \\
&= -\frac{1}{2\pi i}\oint_{\Omega_1} \frac{1}{\left[\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1\right]^{-1} + d_1}dz,
\end{aligned}
\tag{19}
$$

where $i = \sqrt{-1}$ is associated with the complex integrals represents the imaginary unit and the line integrals are taken over the contour Ω_1 that is centered at $(a_1 + b_1)/2$ with radius $c_0|d_1|/2$. Then we can see that the population eigenvalue d_1 is enclosed by the contour Ω_1. By the Taylor expansion, we can show that with significant probability,

$$\left[\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1\right]^{-1} = -z - \mathbf{v}_1^T\mathbf{W}\mathbf{v}_1 + O\left(|z|^{-1}\alpha_n^2\log n\right).$$

Substituting the above expansion into (19) results in

$$
\begin{aligned}
-\frac{1}{2\pi i}\oint_{\Omega_1} \frac{1}{\left[\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1\right]^{-1} + d_1}dz &= -\frac{1}{2\pi i}\oint_{\Omega_1} \frac{1}{d_1 - z - \mathbf{v}_1^T\mathbf{W}\mathbf{v}_1 + O(|z|^{-1}\alpha_n^2\log n)}dz \\
&= -\frac{1}{2\pi i}\oint_{\Omega_1} \frac{1}{d_1 - z} - \frac{1}{2\pi i}\oint_{\Omega_1} \frac{\mathbf{v}_1^T\mathbf{W}\mathbf{v}_1}{(d_1 - z)^2}dz + O\left(d_1^{-2}\alpha_n^2\log n\right) \\
&= -\frac{1}{2\pi i}\oint_{\Omega_1} \frac{1}{d_1 - z} + O\left(d_1^{-2}\alpha_n^2\log n\right)
\end{aligned}
\tag{20}
$$

with significant probability. Thus the asymptotic distribution of $\mathbf{v}_1^T\widehat{\mathbf{v}}_1\widehat{\mathbf{v}}_1^T\mathbf{v}_1$ is determined by $O_p(d_1^{-2}\alpha_n^2\log n)$, which has no contribution from $\mathbf{v}_1^T\mathbf{W}\mathbf{v}_1$. On the other hand, our technical analysis for $\mathbf{u}^T\widehat{\mathbf{v}}_1\widehat{\mathbf{v}}_1\mathbf{v}_1$ (which is much more complicated and can be found in the technical proofs section) reveals that the dominating term is $\mathbf{u}^T\mathbf{W}\mathbf{v}_1$ when $\mathbf{u} \neq \mathbf{v}_1$ or $-\mathbf{v}_1$. This explains why we need to treat differently the two cases of $\mathbf{u}$ close to or far away from $\mathbf{v}_1$.

3.4 A more specific structure and an application

Theorem 2 in Section 3.3 provides some general sufficient conditions to ensure the asymptotic normality for the spiked empirical eigenvectors. Under some simplified but stronger assumptions in Condition 3, the same results on the empirical eigenvectors and eigenvalues continue to hold. Note that the stochastic block models with non-overlapping or overlapping communities can both be included as specific cases of our theoretical analysis. As mentioned

13

before, we choose the direction of $\widehat{\mathbf{v}}_k$ such that $\mathbf{v}_k^T\widehat{\mathbf{v}}_k \geq 0$ for each $1 \leq k \leq K$.

Theorem 3. *Under Conditions 1–3, for each $1 \leq k \leq K$ we have the following properties:*

1) (Eigenvalues) It holds that

$$\frac{\lambda_k - t_k - \mathbb{E}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k}{\left[\mathbb{E}(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - \mathbb{E}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k)^2\right]^{1/2}} \xrightarrow{\mathscr{D}} N(0,1).$$

2) (Eigenvectors) If the unit vector $\mathbf{u}$ satisfies that $\sigma_{\min}^{-1}\|\mathbf{v}_k(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\|_\infty \to 0$ and $|\mathbf{u}^T\mathbf{v}_k| \in [0, 1-\epsilon]$ for some positive constant ϵ, then it holds that

$$\frac{t_k\left(\mathbf{u}^T\widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2}\right) - \mathbb{E}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]}{\left\{\mathrm{var}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]\right\}^{1/2}} \xrightarrow{\mathscr{D}} N(0,1). \tag{21}$$

Moreover, it also holds that

$$\frac{2t_k^2\left(\mathbf{v}_k^T\widehat{\mathbf{v}}_k + A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2}\right)}{\left[\mathrm{var}(\mathbf{v}_k^T\mathbf{W}^2\mathbf{v}_k)\right]^{1/2}} \xrightarrow{\mathscr{D}} N(0,1). \tag{22}$$

Theorem 2 also gives the asymptotic expansions for the asymptotic mean term $A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2}$. It is seen that if $|d_K|$ diverges to infinity much faster than α_n^2, then the $O(\cdot)$ terms in the asymptotic expansions of the mean become smaller order terms and thus the following corollary follows immediately from Theorem 3.

Corollary 1. *Assume that Conditions 1–3 hold. For each $1 \leq k \leq K$, if the unit vector $\mathbf{u}$ satisfies that $|\mathbf{u}^T\mathbf{v}_k| \in [0, 1-\epsilon]$ for some positive constant ϵ and $\alpha_n^{-4}d_k^2\mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \to \infty$, then we have*

$$\frac{t_k\left(\mathbf{u}^T\widehat{\mathbf{v}}_k - \mathbf{u}^T\mathbf{v}_k\right) - \mathbb{E}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]}{\left\{\mathrm{var}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]\right\}^{1/2}} \xrightarrow{\mathscr{D}} N(0,1). \tag{23}$$

Moreover, if $\alpha_n^{-6}d_k^3\mathrm{var}(\mathbf{v}_k^T\mathbf{W}^2\mathbf{v}_k) \to \infty$ then we have

$$\frac{2t_k^2\left(\mathbf{v}_k^T\widehat{\mathbf{v}}_k - 1\right) + \mathbf{v}_k^T\mathbb{E}\mathbf{W}^2\mathbf{v}_k}{\left[\mathrm{var}(\mathbf{v}_k^T\mathbf{W}^2\mathbf{v}_k)\right]^{1/2}} \xrightarrow{\mathscr{D}} N(0,1). \tag{24}$$

Theorem 3 includes the stochastic block model as a specific case. If $\mathbf{X}$ is the affinity matrix from a stochastic block model with K non-overlapping communities and the size of each community is of the same order $O(n)$, then it holds that $\|\mathbf{v}_k\|_\infty = O(n^{-1/2})$, $d_K = O(n)$, $\alpha_n \leq n^{1/2}$, and $\alpha_n\|\mathbf{v}_k\|_\infty \leq O(1)$. Thus Condition 3 can be satisfied as long as $\sigma_{\min} \gg n^{-1/4}$, leading to the asymptotic normalities in Theorem 3.

Our Theorem 3 also covers the stochastic block models with overlapping communities.

For example, the following network model was considered in Zhang et al. (2015)

$$\mathbb{E}\mathbf{X} = \boldsymbol{\Theta}\boldsymbol{\Pi}\mathbf{P}\boldsymbol{\Pi}^T\boldsymbol{\Theta}^T, \tag{25}$$

where $\boldsymbol{\Theta}$ is an $n \times n$ diagonal degree heterogeneity matrix, $\boldsymbol{\Pi}$ is an $n \times K$ community membership matrix, and $\mathbf{P}$ is a $K \times K$ nonsingular irreducible matrix with unit diagonal entries. Observe that the above model has low-rank mean matrix and thus can be connected to our general form of eigendecomposition $\mathbb{E}\mathbf{X} = \mathbf{H} = \mathbf{V}\mathbf{D}\mathbf{V}^T$. If the spiked eigenvalues and spiked eigenvectors satisfy that $|d_k| = O(n)$ and $\|\mathbf{v}_k\|_\infty = O(n^{-1/2})$ for all $1 \le k \le K$, then Condition 3 can be satisfied when $\sigma_{\min} \gg n^{-1/4}$. Consequently, the asymptotic normalities in Theorem 3 can hold.

3.5   Proofs architecture

The key mathematical tools are from complex analysis and random matrix theory. At a high level, our technical proofs consist of four steps. First, we apply Cauchy's residue theorem to represent the desired bilinear form $\mathbf{x}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{y}$ with $1 \le k \le K$ as a complex integral over a contour for a functional of the Green function associated with the original random matrix $\mathbf{X} = \mathbf{H} + \mathbf{W}$. It is worth mentioning that such an approach was used before to study the asymptotic distributions for linear combinations of eigenvectors in the setting of covariance matrix estimation for the case of i.i.d. Gaussian random matrix coupled with linear dependency. Second, we reduce the problem to one that involves a functional of the new Green function associated with only the noise part $\mathbf{W}$ by extracting the spiked part. Such a step enables us to conduct precise high order asymptotic expansions. Third, we conduct delicate high order Taylor expansions for the noise part using new Green function corresponding to the noise part. In this step, we apply the asymptotic expansion directly to the evaluated complex integral over the contour instead of an expansion of the integrand. Such a new way of asymptotic expansion is crucial to our study. Fourth, we bound the variance of $\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}$ using delicate random matrix techniques. In contrast to just counting the number of certain paths in a graph used in classical random matrix theory literature, we need to carefully bound the individual contributions toward the quantity $\alpha_n = \|\mathbb{E}(\mathbf{W} - \mathbb{E}\mathbf{W})^2\|^{1/2}$; otherwise simple counting leads to rather loose upper bound.

3.6   Comparisons with the statistics literature

In a related work, Tang and Priebe (2018) established the CLT for the entries of eigenvectors of a random adjacency matrix. Our work differs significantly from theirs in at least four important aspects. First, Tang and Priebe (2018) assumed a prior distribution on the mean adjacency matrix, while we assume a deterministic mean matrix. As a result, the asymptotic variance in Tang and Priebe (2018) is determined by the prior distribution and is the same for

each entry of an eigenvector, while in our paper the CLT for different entries of an eigenvector can be different and the asymptotic variance depends on all entries of the eigenvector. While Tang and Priebe (2018) also provided the conditional CLT under the setting of the stochastic block model, their result conditions on just one node. Second, our model is much more general than that in Tang and Priebe (2018) in that the spiked eigenvalues can have different orders and different signs. Third, we establish the CLT for the general linear combinations of the components of normalized eigenvectors and the CLT for eigenvalues, while Tang and Priebe (2018) proved the CLT for the rows of $\mathbf{\Lambda}^{1/2}\widehat{\mathbf{V}}^T$, where $\mathbf{\Lambda} \in \mathbb{R}^{K \times K}$ is the diagonal matrix formed by K spiked eigenvalues of the adjacency matrix and $\widehat{\mathbf{V}} = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K)$ is the matrix collecting the corresponding eigenvectors of the adjacency matrix. Fourth, through a dedicated analysis of the higher order expansion for the general linear combination $\mathbf{u}^T\widehat{\mathbf{v}}_k$, we uncover an interesting phase transition phenomenon that the limiting distribution of $\mathbf{u}^T\widehat{\mathbf{v}}_k$ is different when the deterministic weight vector $\mathbf{u}$ is close to or far away from $\mathbf{v}_k$ (modulo the sign), which is new to the literature.

Wang and Fan (2017) proved the asymptotic distribution of the linear form $\mathbf{v}_i^T\widehat{\mathbf{v}}_k$ with $1 \leq i, k \leq K$, where $\mathbf{v}_i$'s and $\widehat{\mathbf{v}}_k$'s are the spiked population and empirical eigenvectors for some *covariance matrix*, respectively. Their asymptotic normality results cover the case of $\mathbf{v}_1^T\widehat{\mathbf{v}}_1$ when $K = 1$, and $\mathbf{v}_i^T\widehat{\mathbf{v}}_k$ for $1 \leq i, k \leq K$ with $i \neq k$ when $K \geq 2$. Similarly, Koltchinskii and Lounici (2016) considered the sample *covariance matrix* under the Gaussian distribution assumption, and derived the asymptotic expansion of the bilinear form $\mathbf{x}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{y}$, where $\mathbf{x}, \mathbf{y}$ are two deterministic unit vectors. They also obtained the asymptotic distribution of $\mathbf{x}^T\widehat{\mathbf{v}}_k$. Different from Wang and Fan (2017) and Koltchinskii and Lounici (2016), in this paper we establish the asymptotic distribution for the *general* linear combination $\mathbf{u}^T\widehat{\mathbf{v}}_k$ for the large structured symmetric random matrix from model (1) under fairly weak regularity conditions. Our proof techniques differ from those in Wang and Fan (2017) and Koltchinskii and Lounici (2016), and are also distinct from most of existing ones in the literature.

4 Statistical applications

The new asymptotic expansions and asymptotic distributions of spiked eigenvectors and eigenvalues established in Section 3 have many natural statistical applications. Next we discuss three specific ones. See also Fan et al. (2019) for another application on testing the node membership profiles in network models.

4.1 Detecting the existence of clustering power

One potential application of Theorem 3 is to improve the results on community detection under model setting (25). Spectral methods have been used popularly in the literature for recovering the memberships of nodes in network models. For example, applying the K-means

clustering algorithm to the K spiked eigenvectors calculated from the adjacency matrix has been a prevalent method for inferring the memberships of nodes. However, it may not be true that all these K eigenvectors are useful for clustering. For example, if eigenvector $\mathbf{v}_k = \mathbf{1}/\sqrt{n}$, then it has zero clustering power and should be dropped in the K-means clustering algorithm. This is especially important in large networks because including a useless high-dimensional eigenvector may significantly increase the noise in clustering. Theorem 3 suggests that we can test the hypothesis $H_0 : \mathbf{v}_k = \mathbf{1}/\sqrt{n}$ using the test statistic $\widehat{\mathbf{v}}_k^T \mathbf{1}$. Then with the aid of Theorem 3, the asymptotic null distribution can be established and the critical value can be calculated. This naturally suggests a method for selecting important eigenvectors in community detection.

4.2 Detecting the existence of denser subgraph

Another application of Theorem 3 is to detect the existence of a denser community in a given random graph, the same problem as studied in Arias-Castro et al. (2014) and Verzelen et al. (2015). Specifically, assume that the data matrix $\mathbf{X} = (x_{ij})$ is a symmetric adjacency matrix with independent Bernoulli entries on and above the diagonal. Let $\mathbf{H} = \mathbb{E}[\mathbf{X}]$ be the mean adjacency matrix. Consider the following null and alternative hypotheses

$$H_0 : \mathbf{H} = p\mathbf{1}\mathbf{1}^T \quad \text{vs.} \quad H_1 : \mathbf{H} = p\mathbf{1}\mathbf{1}^T + (q-p)\boldsymbol{\ell}\,\boldsymbol{\ell}^{\,T},$$

where $\boldsymbol{\ell}$ is the vector with the first n_1 entries being 1 and all remaining entries 0, and $q \in (p, 1]$. It can be seen that under the alternative hypothesis, there is a denser subgraph and q measures the connectivity of nodes within it. Arias-Castro et al. (2014) and Verzelen et al. (2015) proposed tests for the above hypothesis in the setting of $n_1 = o(n)$. We focus on the same setting and in addition assume that $n^{-1} \ll p < q$ and $q \sim p$. We next discuss how to exploit our Theorem 3 to test the same hypothesis.

Under the null hypothesis, a natural estimator of p is given by $\widehat{p} = \frac{1}{n(n-1)} \sum_{1 \le i \le j \le n} x_{ij}$. Moreover, direct calculations show that

$$\mathbf{v}_1^T \mathbb{E}\mathbf{W}^2\mathbf{v}_1 = np(1-p) \quad \text{and} \quad \mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1) = p(1-p)\left[2(n-1) + p^3 + (1-p)^3\right]. \quad (26)$$

Thus the mean and variance of $\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1$ in (26) can be estimated as

$$n\widehat{p}(1-\widehat{p}) \quad \text{and} \quad \widehat{p}(1-\widehat{p})\left[2(n-1) + \widehat{p}^3 + (1-\widehat{p})^3\right], \quad (27)$$

receptively. In view of (24) in Corollary 1, since $\mathbf{v}_1 = n^{-1/2}\mathbf{1}$ under the null hypothesis $H_0 : \mathbf{H} = p\mathbf{1}\mathbf{1}^T$, a natural test statistic for testing $H_0 : \mathbf{H} = p\mathbf{1}\mathbf{1}^T$ is given by

$$T_n = \frac{2\lambda_1^2\left(n^{-1/2}\mathbf{1}^T\widehat{\mathbf{v}}_1 - 1\right) + n\widehat{p}(1-\widehat{p})}{\left[\widehat{p}(1-\widehat{p})\left[2(n-1) + \widehat{p}^3 + (1-\widehat{p})^3\right]\right]^{1/2}}.$$

It can be seen that since $\lambda_1 \approx t_1$ (see Lemma 3), the asymptotic null distribution of T_n is expected to be $N(0,1)$ by resorting to (24) in Corollary 1. On the other hand, under the alternative hypothesis, since the leading eigenvector differs from $n^{-1/2}\mathbf{1}$, the term $n^{-1/2}\mathbf{1}^T\widehat{\mathbf{v}}_1 - 1$ in the numerator of T_n is expected to take some negative value, and thus T_n is expected to have different asymptotic behavior than $N(0,1)$. In fact, we provide the proof sketch in Section D.5 of Supplementary Material on the asymptotic null and alternative distributions. In particular, we show that the asymptotic null distribution of T_n is $N(0,1)$, and if $\frac{n_1^2(q-p)^2}{np} + \frac{n_1^2(q-p)}{n} \gg 1$, then $T_n \to -\infty$ with asymptotic probability one under the alternative hypothesis.

4.3 Rank inference

Our theory can also be applied to statistical testing on the true rank K of the mean matrix $\mathbf{H}$. Rank inference is an important problem in many high-dimensional network applications. See, for example, Lei (2016), Chen and Lei (2018), and Li et al. (2020), and the importance of the problem discussed therein. Consider the following hypotheses

$$H_0 : K = K_0 \quad \text{vs.} \quad H_1 : K > K_0,$$

where K_0 is some prespecified positive integer satisfying $K_0 \leq K$. Define

$$
\begin{aligned}
\widehat{w}_{ij} &= x_{ij} - \sum_{k=1}^{K_0} \lambda_k \mathbf{e}_i^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{e}_j \\
&= w_{ij} - \sum_{k=1}^{K} \left[\lambda_k \mathbf{e}_i^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{e}_j - d_k \mathbf{e}_i^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{e}_j \right] + \sum_{k=K_0+1}^{K} \lambda_k \mathbf{e}_i^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{e}_j. \quad (28)
\end{aligned}
$$

Under the null hypothesis $H_0 : K = K_0$, the last term in (28) disappears and we can obtain the asymptotic expansion of $\widehat{w}_{ij}$ around w_{ij} explicitly by an application of Theorems 1 and 2. Then under some additional regularity conditions, it is expected that $\widehat{w}_{ij}$ is close to w_{ij}. By the independence of w_{ii}, $i = 1, \cdots, n$, it holds that

$$\frac{\sum_{i=1}^{n} w_{ii}}{\sqrt{\sum_{i=1}^{n} w_{ii}^2}} \xrightarrow{\mathscr{D}} N(0,1) \ \text{ as } n \to \infty.$$

Since $\widehat{w}_{ii} \approx w_{ii}$ under the null hypothesis, the following asymptotic distribution is expected to hold as well

$$T_n := \frac{\sum_{i=1}^{n} \widehat{w}_{ii}}{\sqrt{\sum_{i=1}^{n} \widehat{w}_{ii}^2}} \xrightarrow{\mathscr{D}} N(0,1). \quad (29)$$

This naturally suggests a statistical test based on statistic T_n for testing $H_0 : K = K_0$. Under the alternative hypothesis, since $\widehat{w}_{ij}$ contains the smallest $K - K_0$ spiked eigenvalues and the corresponding eigenvectors, its asymptotic behavior is expected to be different, and

consequently, the test can have nontrivial power. In fact, a more sophisticated version of this test constructed based on the off-diagonal entries of $\widehat{\mathbf{W}}$ was investigated recently in Han et al. (2019).

The above asymptotic distribution can also be used to construct confidence intervals for the rank K. To understand this, note that T_n defined in (29) is a function of K_0. Thus an immediate idea for the $100(1-\alpha)\%$ confidence interval construction is to identify all K_0 such that the corresponding T_n falls into the range of $[-\Phi^{-1}(1-\alpha), \Phi^{-1}(1-\alpha)]$, where $\Phi^{-1}(\cdot)$ is the inverse distribution function of the standard normal. Similar ideas can also be exploited to construct confidence intervals for other parameters in network models.

5   Simulation studies

In this section, we use simulation studies to verify the validity of our theoretical results. We consider the stochastic block model with $K = 2$ communities. Assume that the number of nodes is n, the first $n/2$ nodes belong to the first community, and the rest belong to the second one. Then the adjacency matrix $\mathbf{X}$ has the mean structure $\mathbb{E}\mathbf{X} = \mathbf{H} = \mathbf{ARA}^T$, where $\mathbf{R}$ is a 2×2 matrix of the connectivity probabilities, and $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2) \in \mathbb{R}^{n \times 2}$ with $\mathbf{a}_1 = n^{-1/2}(\mathbf{1}^T, \mathbf{0}^T)^T$ and $\mathbf{a}_2 = n^{-1/2}(\mathbf{0}^T, \mathbf{1}^T)^T$, where $\mathbf{0}, \mathbf{1} \in \mathbb{R}^{n/2}$ are vectors of zeros and ones, respectively. It is worth mentioning that $\mathbf{ARA}^T$ is *not* the eigendecomposition of the mean matrix $\mathbf{H}$, which is why we use different notation than that in model (1).

For the connectivity probability matrix $\mathbf{R}$, we consider the structure

$$\mathbf{R} = r \left(\begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array} \right),$$

where parameter r takes 6 different values 0.02, 0.05, 0.1, 0.2, 0.3, and 0.4. A similar model was considered in Abbe et al. (2019) and Lei (2016). For the connectivity matrix $\mathbf{X}$, we simulate its entries on and above the diagonal as independent Bernoulli random variables with means given by the corresponding entries in the mean matrix $\mathbf{H}$, and set the entries below the diagonal to be the same as the corresponding ones above the diagonal. We choose the number of nodes as $n = 3000$ and repeat the simulations for $10,000$ times.

To verify our theoretical results, for each simulated connectivity matrix $\mathbf{X}$ we calculate its eigenvalues and corresponding eigenvectors. For the eigenvalues, we compare the empirical distribution of

$$\frac{\lambda_k - t_k}{\left[\mathrm{var}(\mathbf{v}_k^T \mathbf{W} \mathbf{v}_k) \right]^{1/2}} \tag{30}$$

with the standard normal distribution, where t_k is the solution to equation (10). The exact expression of $\mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)$ in (10) is compli-
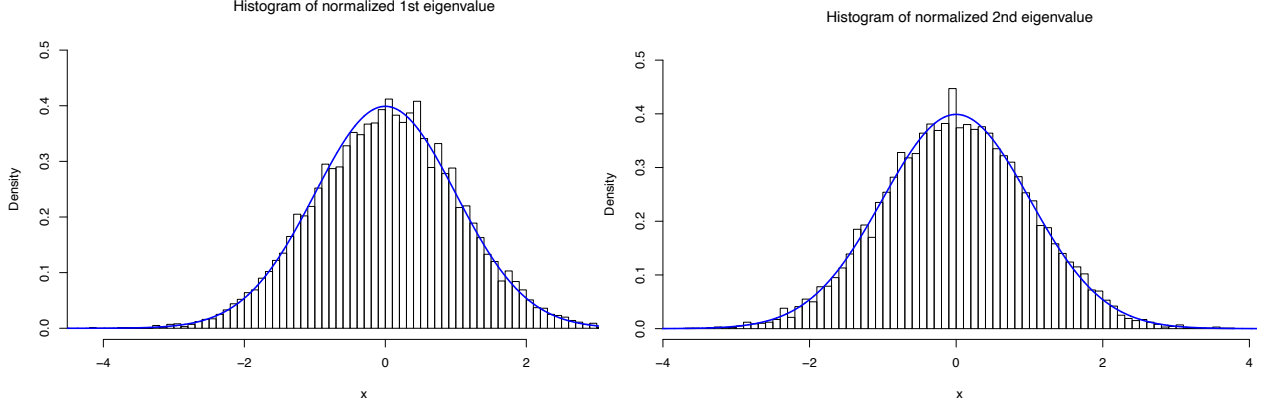
Figure 1: Histograms of the normalized eigenvalues (30) when $r = 0.4$, with the blue curves representing the standard normal density. Left panel: the first eigenvalue; right panel: the second eigenvalue.

cated. Since this term is much smaller than $\mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z)$, we can calculate an approximation of t_k by solving the equation

$$1 + d_k \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z) = 0 \tag{31}$$

using the Newton–Raphson method. Guided by the theoretical derivations, we use $L = 4$ in the asymptotic expansion of $\mathcal{R}(\mathbf{x}, \mathbf{y}, t)$ in (6) for all of our simulation examples. Tables 1–2 summarize the means and standard deviations of (30) with $k = 1$ and 2 calculated from the 10,000 repetitions as well as the p-values from the Anderson–Darling (AD) test for the normality. Figure 1 presents the histograms of the normalized first and second eigenvalues (i.e., (30) when $k = 1$ and 2) from the 10,000 repetitions.

For the eigenvectors, we evaluate the asymptotic normality of the linear combination $\mathbf{u}^T \widehat{\mathbf{v}}_k$ with $k = 1$ and 2. We experiment with three different values for $\mathbf{u}$: $\mathbf{a}_1$, $(1, 0, \cdots, 0)^T$, and $\mathbf{v}_k$. When $\mathbf{u} = \mathbf{a}_1$ or $(1, 0, \cdots, 0)^T$, we calculate the normalized statistic

$$\frac{t_k \left(\mathbf{u}^T \widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right)}{\left\{ \mathrm{var} \left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T \mathbf{v}_k \mathbf{v}_k^T) \mathbf{W} \mathbf{v}_k \right] \right\}^{1/2}}$$

using the 10,000 simulated data sets, while when $\mathbf{u} = \mathbf{v}_k$ we calculate the normalized statistic

$$\frac{2t_k^2 \left(\mathbf{v}_k^T \widehat{\mathbf{v}}_k + A_{\mathbf{v}_k,k,t_k} \widetilde{\mathcal{P}}_{k,t_k}^{1/2} \right)}{\left[\mathrm{var}(\mathbf{v}_k \mathbf{W}^2 \mathbf{v}_k) \right]^{1/2}}$$

instead. In either of the two cases above, the variance in the denominator is calculated as the sample variance from 2,000 simulated independent copies of the noise matrix $\mathbf{W}$. We compare the empirical distributions of the above two normalized statistics with the standard normal distribution. The simulation results are summarized in Tables 3–8 and Figures 2–3.

Our simulation results in Figure 1 and Tables 1–2 suggest that the normalized spiked

20

Table 1: Simulation results for $(\lambda_1 - t_1)/[\mathrm{var}(\mathbf{v}_1^T \mathbf{W} \mathbf{v}_1)]^{1/2}$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	0.0719	0.0149	-0.0068	-0.0080	-0.0024	0.0124
Standard deviation	1.0107	1.0085	0.9927	1.0115	1.0023	1.0125
AD.p-value	0.0725	0.5387	0.6263	0.2342	0.9243	0.2010

Table 2: Simulation results for $(\lambda_2 - t_2)/[\mathrm{var}(\mathbf{v}_2^T \mathbf{W} \mathbf{v}_2)]^{1/2}$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	1.0761	0.2552	0.0681	0.0272	0.0093	0.0052
Standard deviation	0.9630	0.9820	0.9872	1.0100	1.0057	1.0005
AD p-value	0.5349	0.6722	0.8406	0.1806	0.0535	0.8341

eigenvalues have distributions very close to standard Gaussian which supports our results in Theorem 1. Indeed, such a large p-value is extremely impressive given the "sample size" (the number of simulations is 10,000). In general, the simulation results for the eigenvectors support our theoretical findings in Section 3. However, the results corresponding to the first spiked eigenvector $\widehat{\mathbf{v}}_1$ (Tables 3–5) are better than those for the second spiked eigenvector $\widehat{\mathbf{v}}_2$ (Tables 4–8). This is reasonable since for the larger spiked eigenvalue, the negligible terms that we dropped in the proofs of the asymptotic normality become relatively smaller and thus have smaller finite-sample effects on the asymptotic distributions. For the linear form $\mathbf{u}^T \widehat{\mathbf{v}}_k$, when $\mathbf{u} = \mathbf{v}_k$ the convergence to standard normal is slower when compared to the case of $\mathbf{u} \neq \mathbf{v}_k$. This again supports our theoretical findings in Section 3 and explains why we need to separate the cases of $\mathbf{u} = \mathbf{v}_k$ and $\mathbf{u} \neq \mathbf{v}_k$. Such effect is especially prominent for $\mathbf{v}_2^T \widehat{\mathbf{v}}_2$, whose sample mean is -11.8020 when $r = 0.02$ as shown in Table 7. However, it is seen from the same table (and other tables) that as the spiked eigenvalue increases with r, the distribution gets closer and closer to standard Gaussian.

Table 3: Simulation results for $\mathbf{u}^T \widehat{\mathbf{v}}_1$ with $\mathbf{u} = \mathbf{a}_1$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	-0.0573	-0.0140	-0.0023	-0.0045	-0.0071	-0.0069
Standard deviation	1.0335	1.0244	1.0011	1.0001	1.0214	1.0016
AD.p-value	0.7879	0.4012	0.2417	0.5300	0.9482	0.9935

Table 4: Simulation results for $\mathbf{v}_1^T\widehat{\mathbf{v}}_1$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	-1.3288	-0.4817	-0.1900	-0.0742	-0.0409	-0.0186
Standard deviation	1.0940	1.0545	1.0338	0.9749	1.0030	1.0005
AD.p-value	0.0582	0.4251	0.0251	0.0225	0.3312	0.2912

Table 5: Simulation results for $\mathbf{u}^T\widehat{\mathbf{v}}_1$ with $\mathbf{u} = (1, 0, \cdots, 0)^T$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	0.0025	0.0021	0.0003	0.0105	0.0061	-0.0122
Standard deviation	1.0432	1.0354	0.9871	1.0016	1.0205	0.9898
AD.p-value	0.0044	0.4877	0.3752	0.1514	0.1304	0.3400

Table 6: Simulation results for $\mathbf{u}^T\widehat{\mathbf{v}}_2$ with $\mathbf{u} = \mathbf{a}_1$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	4.2611	1.0129	0.3067	0.0745	0.0219	0.0037
Standard deviation	1.2384	1.0952	1.0294	1.0098	1.0280	1.0044
AD p-value	0.3829	0.7535	0.3759	0.4105	0.9129	0.9873

Table 7: Simulation results for $\mathbf{v}_2^T\widehat{\mathbf{v}}_2$

r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	-11.8020	-4.3274	-2.0057	-0.7447	-0.3526	-0.1650
Standard deviation	1.3775	1.1192	1.0980	1.0343	1.0104	1.0089
AD p-value	0.0000	0.0011	0.0422	0.3964	0.4980	0.1186

Table 8: Simulation results for $\mathbf{u}^T\widehat{\mathbf{v}}_2$ with $\mathbf{u} = (1, 0, \cdots, 0)^T$

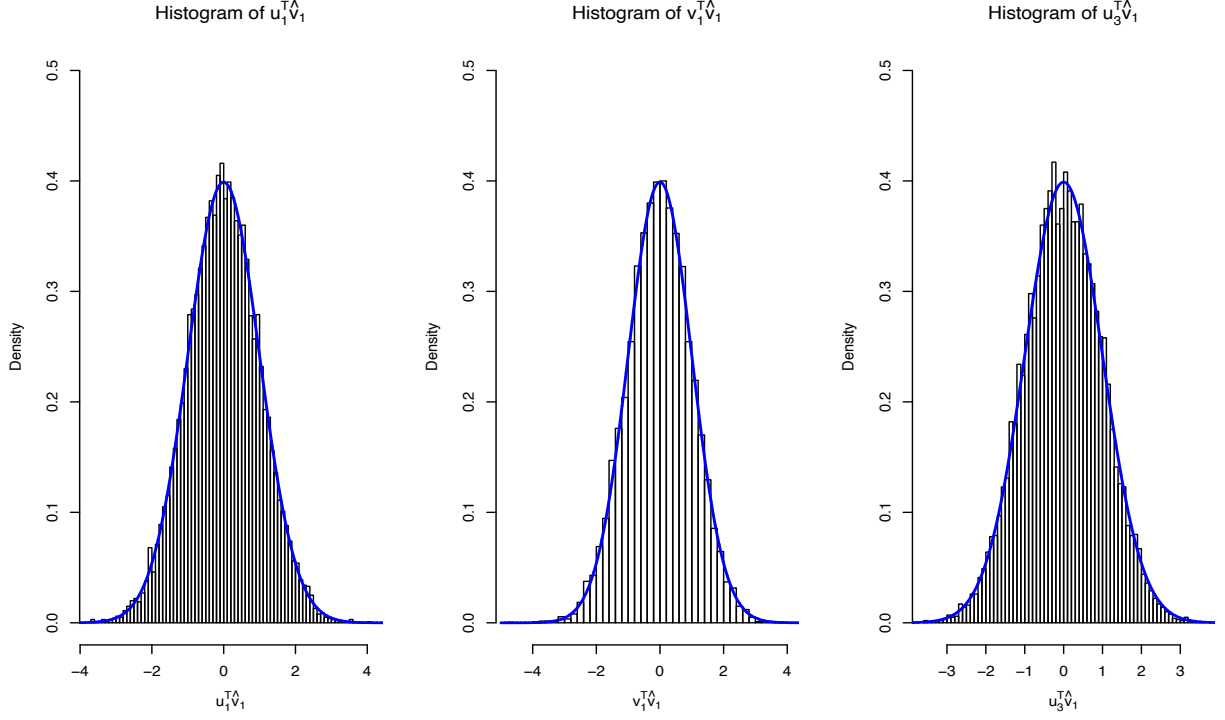| r	0.02	0.05	0.1	0.2	0.3	0.4
Mean	0.0622	0.0204	0.0018	-0.0074	-0.0119	-0.0049
Standard deviation	1.1221	1.0537	1.0272	1.0022	1.0088	0.9933
AD p-value	0.0003	0.5853	0.0930	0.6011	0.2423	0.4385

Figure 2: Histograms corresponding to the first eigenvector $\widehat{\mathbf{v}}_1$ when $r = 0.4$, with the blue curves representing the standard normal density. Left panel: $\mathbf{u}_1^T\widehat{\mathbf{v}}_1$; middle panel: $\mathbf{v}_1^T\widehat{\mathbf{v}}_1$; right panel: $\mathbf{u}_3^T\widehat{\mathbf{v}}_1$, where $\mathbf{u}_1 = \mathbf{a}_1$ and $\mathbf{u}_3 = (1, 0, \cdots, 0)^T$.
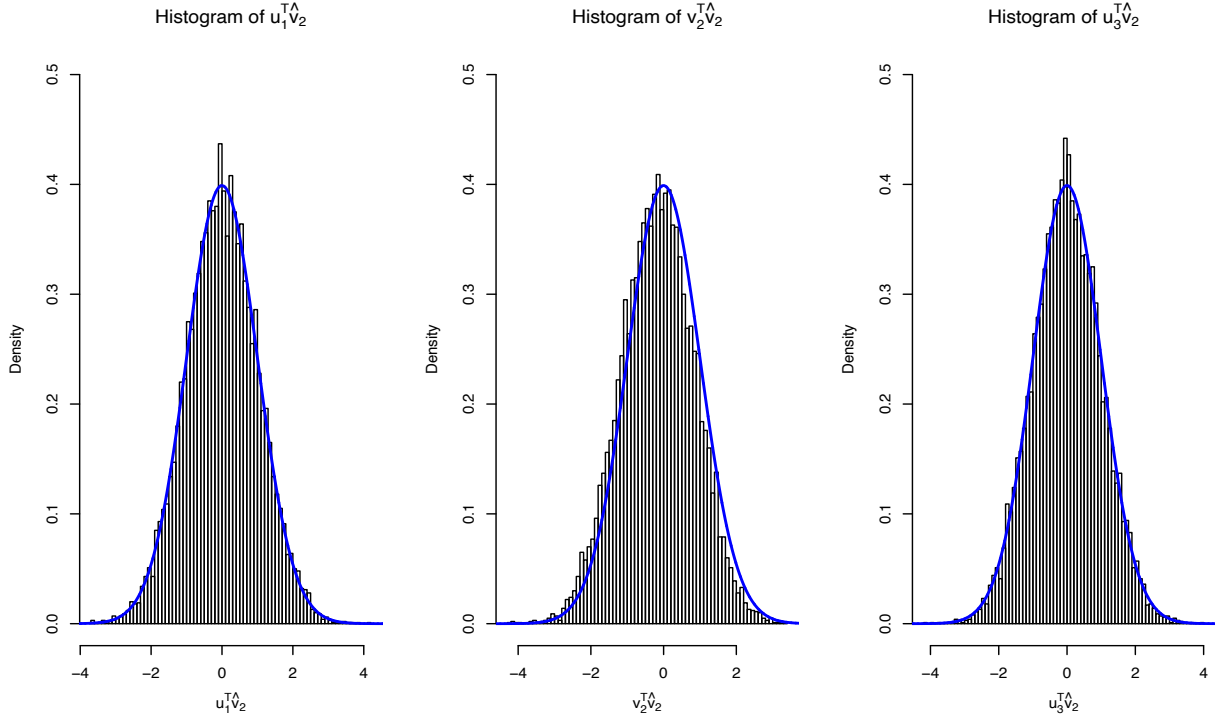

Figure 3: Histograms corresponding to the second eigenvector $\widehat{\mathbf{v}}_2$ when $r = 0.4$, with the blue curves representing the standard normal density. Left panel: $\mathbf{u}_1^T\widehat{\mathbf{v}}_2$; middle panel: $\mathbf{v}_2^T\widehat{\mathbf{v}}_2$; right panel: $\mathbf{u}_3^T\widehat{\mathbf{v}}_1$, where $\mathbf{u}_1 = \mathbf{a}_1$ and $\mathbf{u}_3 = (1, 0, \cdots, 0)^T$.

23

6 A more general asymptotic theory

As mentioned before, the asymptotic theory on the spiked eigenvectors in terms of the general linear combination and on the spiked eigenvalues presented in Section 3 is in fact a consequence of a more general asymptotic theory on the spiked eigenvectors in terms of the bilinear form. In this section, we focus our attention on such a more general asymptotic theory for the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ with $1 \leq k \leq K$, where $\mathbf{x}$ and $\mathbf{y}$ are two arbitrary n-dimensional unit vectors. See Sections 3.5 and 3.6 for detailed discussions on the technical innovations of our novel ATE theoretical framework and comparisons with the existing literature on the asymptotic distributions of eigenvectors.

For technical reasons, we will break our main results on the asymptotic distributions of the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ down to two theorems, where we consider in Theorem 4 the case when either vector $\mathbf{x}$ or vector $\mathbf{y}$ is sufficiently further away from the population eigenvector $\mathbf{v}_k$, and then we study in Theorem 5 the case when both vectors $\mathbf{x}$ and $\mathbf{y}$ are very close to $\mathbf{v}_k$. The technical treatments for these two cases are different since in the latter scenario, the first order term which determines the asymptotic distribution in Theorem 4 vanishes, and thus we need to consider higher order expansions to obtain the asymptotic distribution in Theorem 5. Let $\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k}$, $\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}$, and $\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}$ be the three rank one matrices given in (113)–(115), respectively, in the proof of Theorem 5 in Section A.6. Denote by $\sigma_k^2 = \mathrm{var}[\mathrm{tr}(\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k})]$ and

$$\widetilde{\sigma}_k^2 = \mathrm{var}\left\{ \mathrm{tr}\left[\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - \left(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2\right) \mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}\right] + \mathrm{tr}\left(\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\right) \mathrm{tr}\left(\mathbf{W}\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}\right) \right\}. \quad (32)$$

Both of the quantities above play an important role in our more general asymptotic theory.

Theorem 4. *Assume that Conditions 1–2 hold and $\mathbf{x}$ and $\mathbf{y}$ are two n-dimensional unit vectors. Then for each $1 \leq k \leq K$, if $\sigma_k^2 \gg t_k^{-4}\alpha_n^2(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|)^2 + t_k^{-4}$ we have the asymptotic expansion*

$$\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} = a_k + \mathrm{tr}(\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k}) + O_p\left\{ t_k^{-2}\alpha_n(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|) + t_k^{-2} \right\}, \quad (33)$$

where the quantity $a_k = A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k}$.

The assumption of $\sigma_k^2 \gg t_k^{-4}\alpha_n^2(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|)^2 + t_k^{-4}$ in Theorem 4 requires the variance of random variable $\mathrm{tr}(\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k})$ not too small, which at high level, requires that either vector $\mathbf{x}$ or $\mathbf{y}$ is sufficiently faraway from the population eigenvector $\mathbf{v}_k$. If $\sigma_{ij} \sim 1$ for each (i,j) pair, then such an assumption restricts essentially that $\|\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k}\|$ should not be too close to zero. This in turn ensures that the first order expansion is sufficient for deriving the asymptotic normality of $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$. Theorem 4 also entails that a simple upper bound for $\widetilde{\sigma}_k$ as defined in (32) can be shown to be $O(t_k^{-2}\alpha_n)$.

Theorem 5. *Assume that Conditions 1–2 hold and* $\mathbf{x}$ *and* $\mathbf{y}$ *are two n-dimensional unit vectors. Then for each $1 \le k \le K$, if $\sigma_k^2 = O(\widetilde{\sigma}_k^2)$ and $\widetilde{\sigma}_k^2 \gg t_k^{-6}\alpha_n^4(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|)^2 + t_k^{-6}$ we have the asymptotic expansion*

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} = a_k &+ \operatorname{tr}\left[\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - \left(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2\right)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}\right] + \operatorname{tr}\left(\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\right)\operatorname{tr}\left(\mathbf{W}\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}\right) \\
&+ O_p\left\{|t_k|^{-3}\alpha_n^2(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|) + |t_k|^{-3}\right\},
\end{aligned}
\tag{34}
$$

where the quantity a_k is given in (33).

The ATE theoretical framework for the more general asymptotic theory established in Theorems 4 and 5 is empowered by the following two technical lemmas.

Lemma 4. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, we have*

$$
\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})
\tag{35}
$$

with $l \ge 1$ some bounded positive integer and $d_{\mathbf{X}} = \|\mathbf{x}\|_\infty$.

Lemma 5. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, we have $\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(1)$ and*

$$
\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(\alpha_n^l)
\tag{36}
$$

with $l \ge 2$ some bounded positive integer.

The detailed proofs of Lemmas 4 and 5 are provided in Sections B.5 and B.6 of Supplementary Material. Our delicate technical arguments therein establish useful refinements to the classical idea of counting the number of nonzero terms from the random matrix theory. In particular, Lemma 4 is the key building block for high order Taylor expansions that involve polynomials of quantities in the lemma with different choices of $(\mathbf{x}, \mathbf{y}, l)$.

7 Discussions

In contrast to the immense literature on the asymptotic distributions for eigenvalues of large spiked random matrices, the counterpart asymptotic theory for eigenvectors has remained largely underdeveloped in statistics literature for years. Yet such a theory is much desired for understanding the precise asymptotic properties of various statistical and machine learning algorithms that build upon the spectral information of the eigenspace constructed from observed data matrix. Our work in this paper provides a first attempt with a general ATE theoretical framework for underpinning the precise asymptotic expansions and asymptotic distributions for spiked eigenvectors and spiked eigenvalues of large spiked random matrices with diverging spikes. Our results complement existing ones in the RMT literature as well as the networks literature.

The family of models in our ATE framework includes many popularly used ones for large-scale applications including network analysis and text analysis such as the stochastic block models with or without overlapping communities and the topic models. Our general asymptotic theory for eigenvectors can be exploited to develop new useful tools for precise statistical inference in these applications. It would be interesting to investigate the problem of reproducible large-scale inference as in Barber and Candès (2015); Candès et al. (2018); Lu et al. (2018); Fan et al. (2019); Fan et al. (2019) in these model settings. It would also be interesting to develop a general method to determine the rank and provide robust rank inference in such high-dimensional low-rank models. These extensions are beyond the scope of the current paper and will be interesting topics for future research.

A   Proofs of main results

Recall that Condition 2 involves two scenarios of the spike strength. We will first prove all the results under scenario i). Then in Section D of Supplementary Material, we will adapt the proofs to show that the same results also hold under scenario ii). We provide the proofs of Theorems 1–5 and Corollary 1 in this appendix. Additional technical details including the proofs of all the lemmas and further discussions on when the asymptotic normality can hold for the asymptotic expansion in Theorem 5 are contained in the Supplementary Material.

A.1   Proof of Theorem 1

The results on the asymptotic distributions of spiked eigenvalues in Theorem 1 are in fact a consequence of those on the asymptotic expansions and asymptotic distributions for the spiked eigenvectors, where a more general asymptotic theory of the eigenvectors is presented in Theorems 4–5 in Section 6. Let us define a matrix-valued function that is referred to as the Green function associated with only the noise part $\mathbf{W}$

$$\mathbf{G}(z) = (\mathbf{W} - z\mathbf{I})^{-1} \tag{37}$$

for z in the complex plane $\mathbb{C}$, where $\mathbf{I}$ stands for the identity matrix of size n. Recall that $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of matrix $\mathbf{X}$ and $\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_n$ are the corresponding eigenvectors. By Weyl's inequality, it holds that $\max|\lambda_i - d_i| \le \|\mathbf{W}\|$. Thus, in view of Condition 2 and Lemma 6 in Supplementary Material, all the spiked eigenvalues λ_k with $1 \le k \le K$ of the observed random matrix $\mathbf{X}$ have magnitudes of larger order than the eigenvalues of the noise matrix $\mathbf{W}$ with significant probability as the matrix size n increases. This entails that with significant probability, matrices $\mathbf{G}(\lambda_k)$ with $1 \le k \le K$ are well defined and nonsingular. For the rest of this proof, we restrict all the derivations on such an event that holds with asymptotic probability one.

It follows from the definition of the eigenvalue, the representation $\mathbf{X} = \mathbf{H} + \mathbf{W} = \mathbf{V}\mathbf{D}\mathbf{V}^T + \mathbf{W}$, (37), and the properties of the determinant function $\det(\cdot)$ that for each $1 \leq k \leq K$,

$$0 = \det(\mathbf{X} - \lambda_k \mathbf{I}) = \det(\mathbf{W} - \lambda_k \mathbf{I} + \mathbf{V}\mathbf{D}\mathbf{V}^T) = \det[\mathbf{G}^{-1}(\lambda_k) + \mathbf{V}\mathbf{D}\mathbf{V}^T]$$
$$= \det[\mathbf{G}^{-1}(\lambda_k)]\det[\mathbf{I} + \mathbf{G}(\lambda_k)\mathbf{V}\mathbf{D}\mathbf{V}^T],$$

which leads to $\det[\mathbf{I} + \mathbf{G}(\lambda_k)\mathbf{V}\mathbf{D}\mathbf{V}^T] = 0$ since $\det[\mathbf{G}^{-1}(\lambda_k)] = \det[\mathbf{G}(\lambda_k)]^{-1}$ is nonzero. Using the identity $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$ for matrices $\mathbf{A}$ and $\mathbf{B}$, we obtain for each $1 \leq k \leq K$,

$$0 = \det[\mathbf{I} + \mathbf{G}(\lambda_k)\mathbf{V}\mathbf{D}\mathbf{V}^T] = \det[\mathbf{I} + \mathbf{D}\mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V}], \tag{38}$$

where the second $\mathbf{I}$ represents an identity matrix of size K and we slightly abuse the notation for simplicity. Since the diagonal matrix $\mathbf{D}$ is nonsingular by assumption, it follows from (38) that

$$\det[d_k \mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V} + d_k \mathbf{D}^{-1}] = d_k \det(\mathbf{D}^{-1})\det[\mathbf{I} + \mathbf{D}\mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V}] = 0 \tag{39}$$

for each $1 \leq k \leq K$.

By the asymptotic expansions in (79), Lemmas 4 and 5, and Weyl's inequality $\max|\lambda_k - d_k| \leq \|\mathbf{W}\|$, we have for $j \neq \ell$, $d_k \mathbf{v}_j^T\mathbf{G}(\lambda_k)\mathbf{v}_\ell = -d_k O_p(\lambda_k^{-2}) = O_p(1/|d_k|)$. Thus, we can see that all off diagonal entries of matrix $d_k \mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V} + d_k \mathbf{D}^{-1}$ in (39) are of order $O_p(1/|d_k|)$. For $j \neq k$, the jth diagonal entry of $d_k \mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V} + d_k \mathbf{D}^{-1}$ equals $d_k \mathbf{v}_j^T\mathbf{G}(\lambda_k)\mathbf{v}_j + d_k/d_j$. By (78) and Lemma 4, we have $d_k \mathbf{v}_j^T\mathbf{G}(\lambda_k)\mathbf{v}_j + 1 = o_p(1)$. Moreover, by Condition 2 $|d_k/d_j - 1| \geq c$ for some positive constant c. Hence, all these diagonal entries but the kth one are of order at least $O_p(1)$. Thus the matrix $(d_k \mathbf{v}_i^T\mathbf{G}(\lambda_k)\mathbf{v}_j + \delta_{ij}d_k/d_i)_{1 \leq i,j \leq K, i,j \neq k}$ is invertible with significant probability, where $\delta_{ij} = 1$ when $i = j$ and 0 otherwise. Recall the determinant identity for block matrices from linear algebra

$$\det\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})$$

when the lower right block matrix $\mathbf{A}_{22}$ is nonsingular. Treating the kth diagonal entry of $d_k \mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V} + d_k \mathbf{D}^{-1}$ as the first block, we have with significant probability

$$\det[d_k \mathbf{V}^T\mathbf{G}(\lambda_k)\mathbf{V} + d_k \mathbf{D}^{-1}] = 0 \tag{40}$$

entailing $d_k \mathbf{v}_k^T\mathbf{G}(\lambda_k)\mathbf{v}_k + 1 = d_k \mathbf{v}_k^T\mathbf{F}_k(\lambda_k)\mathbf{v}_k$, where $\mathbf{F}_k(z) = \mathbf{G}(z)\mathbf{V}_{-k}[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}]^{-1}$ $\cdot \mathbf{V}_{-k}^T\mathbf{G}(z)$ and $\mathbf{A}_{-k}$ denotes the submatrix of matrix $\mathbf{A}$ by removing the kth column. In light of (40) and the solution $\widehat{t}_k$ to equation (94) in the proof of Theorem 4 in Section A.5,

it holds from the uniqueness of $\widehat{t}_k$ that

$$\lambda_k = \widehat{t}_k. \tag{41}$$

Therefore, combining equality (41) with asymptotic expansion $\widehat{t}_k - t_k = \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n/t_k)$ obtained in (99) completes the proof of Theorem 1.

A.2   Proof of Theorem 2

The results on the asymptotic distributions of spiked eigenvectors in Theorem 2 are also an implication of a more general asymptotic theory of the eigenvectors presented in Theorems 4–5 in Section 6 on the delicate asymptotic expansions and asymptotic distributions for the spiked eigenvectors. Recall that $\widehat{\mathbf{V}} = (\widehat{\mathbf{v}}_1, \cdots, \widehat{\mathbf{v}}_K)$ with $\widehat{\mathbf{v}}_k$ for $1 \leq k \leq K$ the empirical spiked eigenvectors of the observed random matrix $\mathbf{X}$. Without loss of generality, let us choose the direction of eigenvector $\widehat{\mathbf{v}}_k$ such that $\widehat{\mathbf{v}}_k^T \mathbf{v}_k \geq 0$. Clearly, fixing the direction of $\widehat{\mathbf{v}}_k$ does not affect the distribution of $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$; that is, its distribution stays the same when $-\widehat{\mathbf{v}}_k$ is chosen as the eigenvector. We will separately consider the two cases of $\mathbf{v}_k^T \widehat{\mathbf{v}}_k$ and $\mathbf{u}^T \widehat{\mathbf{v}}_k$ with $\mathbf{u} \neq \mathbf{v}_k$, where the former relies on the second order expansion given in (111) in the proof of Theorem 5 in Section A.6, and the latter utilizes the first order expansion given in (107) in the proof of Theorem 4 in Section A.5.

We first consider $\mathbf{v}_k^T \widehat{\mathbf{v}}_k$. Choosing $\mathbf{x} = \mathbf{y} = \mathbf{v}_k$ in Theorem 4 gives $a_k = A_{\mathbf{v}_k,k,t_k}^2 \widetilde{\mathcal{P}}_{k,t_k}$. By Lemma 5, it holds that

$$\mathcal{P}(\mathbf{v}_k, \mathbf{v}_k, t_k) = - \sum_{l=0, l \neq 2}^{L} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k}{t_k^l} = -1 + O(\alpha_n^2/t_k^2) \tag{42}$$

and

$$\|\mathcal{P}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k)\| = \| - \sum_{l=0, l \neq 2}^{L} \frac{\mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{V}_{-k}}{t_k^l} \| = O(\alpha_n^2/t_k^2). \tag{43}$$

Moreover, recalling the definition of $A_{\mathbf{u},k,t_k}$ in (9), $A_{\mathbf{u},k,t_k}$ can be rewritten as

$$A_{\mathbf{v}_k,k,t_k} = \mathcal{P}(\mathbf{v}_k, \mathbf{v}_k, t) - t_k^{-1} \mathcal{P}(\mathbf{v}_k, \mathbf{V}_{-k}, t_k) \Big(\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) \Big)^{-1} \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k).$$

Therefore, by (42)–(43), (A.16), and (91), we have

$$A_{\mathbf{v}_k,k,t_k} = -1 + O(\alpha_n^2/t_k^2) \quad \text{and} \quad \widetilde{\mathcal{P}}_{k,t_k} = 1 + O(\alpha_n^2/t_k^2). \tag{44}$$

Now recall the second order expansion of $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k \mathbf{y}$ given in (111) in the proof of Theorem 5. We next calculate the orders of each term in the expansion (111). First, we consider

28

$\mathbf{b}_{\mathbf{v}_k,k,t_k}^T$. By (43), (A.16), and the definition (7), we have

$$\|\mathbf{b}_{\mathbf{v}_k,k,t_k}^T - \mathbf{v}_k^T\| = \left\|\mathcal{R}(\mathbf{v}_k,\mathbf{V}_{-k},t_k)\left((\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right)^{-1}\mathbf{V}_{-k}^T\right\| = O(\alpha_n^2/t_k^2). \tag{45}$$

This together with (44) entails that

$$\|\mathbf{b}_{\mathbf{v}_k,k,t_k}^T + A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\| = O(\alpha_n^2/t_k^2). \tag{46}$$

It follows from Lemma 4 and (44) that

$$A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}(\mathbf{b}_{\mathbf{v}_k,k,t_k}^T + A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k/t_k = O_p(\alpha_n^2/|t_k|^3),$$

$$\begin{aligned}
\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}&\left[2\widetilde{\mathcal{P}}_{k,t_k}\left(A_{\mathbf{v}_k,k,t_k}\mathbf{b}_{\mathbf{v}_k,k,t_k}^T + A_{\mathbf{v}_k,k,t_k}\mathbf{b}_{\mathbf{v}_k,k,t_k}^T\right)\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T + \mathbf{b}_{\mathbf{v}_k,k,t_k}^T\mathbf{W}\mathbf{v}_k\mathbf{b}_{\mathbf{v}_k,k,t_k}^T\right]\mathbf{W}\mathbf{v}_k\\
&+ 2A_{\mathbf{v}_k,k,t_k}A_{\mathbf{y},k,t_k}t_k^{-2}\left(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\right)^2\\
&+ A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\left\{t_k^{-2}\mathbf{x}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{v}_k,\mathbf{V}_{-k},t)\right.\\
&\times\left.\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\right\}\\
&+ A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\left\{t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{v}_k,\mathbf{V}_{-k},t)\right.\\
&\times\left.\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\right\}\\
&+ A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{v}_k,\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k\\
&+ A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{v}_k,\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k\\
&= O_p(\frac{1}{|t_k|^2} + \frac{\alpha_n}{|t_k|^3}),
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}&(A_{\mathbf{v}_k,k,t_k}\mathbf{x}^T + A_{\mathbf{v}_k,k,t_k}\mathbf{y}^T)(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k + 3t_k^{-2}A_{\mathbf{v}_k,k,t_k}A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k\\
&= \frac{\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k}{t_k^2} + O_p(\alpha_n^3/t_k^4).
\end{aligned}$$

Substituting the above equations into (111) results in

$$\mathbf{v}_k^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{v}_k - A_{\mathbf{v}_k,k,t_k}^2\widetilde{\mathcal{P}}_{k,t_k} = -\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k/t_k^2 + O_p(|t_k|^{-2} + \alpha_n^2/|t_k|^3), \tag{47}$$

where the leading term of the asymptotic expansion now depends on the second moments of

the noise matrix $\mathbf{W}$. Recall that $\mathbf{v}_k^T\widehat{\mathbf{v}}_k \geq 0$. By (44) and (47) we have

$$
\begin{aligned}
\mathbf{v}_k^T\widehat{\mathbf{v}}_k + A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2} &= -\frac{\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k}{2A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2}t_k^2} + O_p(\alpha_n^2/t_k^3) \\
&= -\frac{\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k}{2t_k^2} + O_p(|t_k|^{-2} + \alpha_n^2/|t_k|^3). \quad (48)
\end{aligned}
$$

We now consider an arbitrary unit vector $\mathbf{u} \in \mathbb{R}^n$ with $|\mathbf{u}^T\mathbf{v}_k| \in [0,1)$ for investigating the asymptotic distributions of the general linear combinations $\mathbf{u}^T\widehat{\mathbf{v}}_k$. Recall the first order expansion given in (107) in the proof of Theorem 4 and (46) that

$$
\begin{aligned}
\mathbf{u}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{v}_k - A_{\mathbf{u},k,t_k}A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k} &= -A_{\mathbf{v}_k,k,t_k}\widetilde{\mathcal{P}}_{k,t_k}(\mathbf{b}_{\mathbf{u},k,t_k}^T + A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k/t_k \\
&\quad + O_p(\alpha_n/t_k^2). \quad (49)
\end{aligned}
$$

Then dividing (49) by $\mathbf{v}_k^T\widehat{\mathbf{v}}_k$ and using (44) and (48), we can deduce that

$$
\begin{aligned}
\mathbf{u}^T\widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2} &= \widetilde{\mathcal{P}}_{k,t_k}^{1/2}(\mathbf{b}_{\mathbf{u},k,t_k}^T + A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k/t_k + O_p(\alpha_n/t_k^2) \\
&= (\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k/t_k + O_p(\alpha_n/t_k^2). \quad (50)
\end{aligned}
$$

In view of the asymptotic expansions in (50) and (48), we can see that the desired asymptotic normalities in the two parts of Theorem 2 follow from the conditions of Lemmas 1 or 2. More specifically, for (50) if $\alpha_n^{-2}d_k^2\mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \to \infty$, then we have $\alpha_n/t_k^2 \ll \{\mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k]\}^{1/2}$ and thus the first part of Theorem 2 in (17) holds in view of (50). Furthermore, if $(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k$ is $\mathbf{W}^1$-CLT, then $(\mathbf{b}_{\mathbf{u},k,t_k} - \mathbf{v}_k\mathbf{v}_k^T\mathbf{u}, \mathbf{v}_k)$ is also $\mathbf{W}^1$-CLT and thus we have

$$
\frac{t_k\left(\mathbf{u}^T\widehat{\mathbf{v}}_k + A_{\mathbf{u},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}^{1/2}\right) - \mathbb{E}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]}{\left\{\mathrm{var}\left[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k\right]\right\}^{1/2}} \xrightarrow{\mathscr{D}} N(0,1).
$$

Similarly, the second part of Theorem 2 in (18) also holds under the condition $(\alpha_n^{-4}d_k^2 + 1)\mathrm{var}[\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k] \to \infty$ and the CLT holds if $(\mathbf{v}_k, \mathbf{v}_k)$ is $\mathbf{W}^2$-CLT. This concludes the proof of Theorem 2.

A.3   Proof of Theorem 3

The results on the asymptotic distributions of spiked eigenvalues and spiked eigenvectors in Theorem 3 are an application of those in Theorems 1 and 2 for a more specific structure of the low rank model (1), including the stochastic block model with both non-overlapping and overlapping communities as special cases.

First, note that (15) implies that the condition of Lemma 1 holds for $\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k$ under Condition 3. Consequently, $(\mathbf{v}_k, \mathbf{v}_k)$ is $\mathbf{W}^1$-CLT. In addition, (15) ensures that $\mathbb{E}(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k -$

$\mathbb{E}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k)^2 \gg \alpha_n^2/d_k^2$ under Condition 3. Therefore, it follows from Theorem 1 that the first result of Theorem 3 holds. Recall that in (A.12), $s_{\mathbf{x},\mathbf{y}}$ is defined as the expected value of the conditional variance of $\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k$. By definition, we have $\mathrm{var}[\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k] \geq s_{\mathbf{x},\mathbf{y}} \geq c\sigma_{\min}^2 n$. Thus the condition $(\alpha_n^{-4}d_k^2 + 1)\mathrm{var}[\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k] \to \infty$ in Theorem 2 is ensured by the assumptions

$$\sigma_{\min}^2 n \to \infty, \ \frac{|d_K|\sigma_{\min}}{\alpha_n} \to \infty, \ \alpha_n \leq n^{1/2}$$

in Condition 3. Moreover, by (A.13) we can see that the conditions of Lemma 2 are satisfied for $\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k$ under Condition 3. Thus $(\mathbf{v}_k, \mathbf{v}_k)$ is $\mathbf{W}^2$-CLT. Therefore, (22) holds by an application of (18) in Theorem 2.

It remains to show that the condition

$$\mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k] \gg \alpha_n^2/d_k^2 \tag{51}$$

in Theorem 2 can be guaranteed by Condition 3. Then the expansion in (17) holds. Moreover, the condition $\sigma_{\min}^{-1}\|\mathbf{v}_k[\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T]\|_\infty \to 0$ ensures that $(\mathbf{b}_{\mathbf{u},k,t_k} - \mathbf{v}_k\mathbf{v}_k^T\mathbf{u}, \mathbf{v}_k)$ is $\mathbf{W}^1$-CLT. Combining these results entails that the asymptotic normality (21) holds. Now we proceed to verify (51). Consider an arbitrary unit vector $\mathbf{u} \in \mathbb{R}^n$ satisfying $|\mathbf{u}^T\mathbf{v}_k| \in [0, 1-\epsilon]$ for some positive constant ϵ. Recalling the definition of $\mathbf{b}_{\mathbf{u},k,t}$ in (8), we have $\mathbf{b}_{\mathbf{u},k,t}^T\mathbf{v}_k = \{\mathbf{u}^T - \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t)[(\mathbf{D}_{-k})^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t)]^{-1}\mathbf{V}_{-k}^T\}\mathbf{v}_k = \mathbf{u}^T\mathbf{v}_k$. Thus it holds that $\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T = \mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{b}_{\mathbf{u},k,t_k}^T\mathbf{v}_k\mathbf{v}_k^T = \mathbf{b}_{\mathbf{u},k,t_k}^T(\mathbf{I} - \mathbf{v}_k\mathbf{v}_k^T)$. Moreover, similar to (15) we can show that

$$\left[\mathbb{E}(\mathbf{u}^T\mathbf{W}\mathbf{v}_k - \mathbb{E}\mathbf{u}^T\mathbf{W}\mathbf{v}_k)^2\right]^{1/2} \geq \sigma_{\min}(2 - 2\|\mathbf{v}_k\|_\infty^2)^{1/2}. \tag{52}$$

This ensures that there exists some positive constant c_1 such that

$$\begin{aligned}
\mathrm{var}[(\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T)\mathbf{W}\mathbf{v}_k]^2 &\geq \sigma_{\min}^2(2 - 2\|\mathbf{v}_k\|_\infty^2)\|\mathbf{b}_{\mathbf{u},k,t_k}^T - \mathbf{u}^T\mathbf{v}_k\mathbf{v}_k^T\|^2 \\
&\geq \sigma_{\min}^2(2 - 2\|\mathbf{v}_k\|_\infty^2)\|\mathbf{b}_{\mathbf{u},k,t_k}^T(\mathbf{I} - \mathbf{v}_k\mathbf{v}_k^T)\|^2 \\
&\geq c_1\sigma_{\min}^2[-(\mathbf{u}^T\mathbf{v}_k)^2 + \mathbf{b}_{\mathbf{u},k,t_k}^T\mathbf{b}_{\mathbf{u},k,t_k}],
\end{aligned} \tag{53}$$

where we have applied $\mathbf{b}_{\mathbf{u},k,t}^T\mathbf{v}_k = \mathbf{u}^T\mathbf{v}$ again in the last step.

Let $\widetilde{\mathbf{V}} = (\mathbf{v}_{K+1}, \cdots, \mathbf{v}_n)$ be an $n \times (n - K)$ matrix such that $(\mathbf{V}, \widetilde{\mathbf{V}})$ is an orthogonal matrix of size n. Then the n-dimensional unit vector $\mathbf{u}$ can be represented as $\mathbf{u} = \sum_{i=1}^n a_i\mathbf{v}_i$ for some scalars a_i's. For each $1 \leq k \leq K$, by the definition of $\mathcal{R}$ in (6) and Lemma 5 we can show that

$$\|\mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k) + t_k^{-1}\mathbf{I}\| = O\left(\frac{\alpha_n^2}{|t_k|^3}\right) \ \text{and} \ \|\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t_k) + t_k^{-1}\mathbf{u}^T\mathbf{V}_{-k}\| = O\left(\frac{\alpha_n^2}{|t_k|^3}\right). \tag{54}$$

Therefore it holds that

$$\left\|\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t_k)[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1}\mathbf{V}_{-k}^T + \sum_{1 \leq i \neq k \leq K} a_i(t_k d_i^{-1} - 1)^{-1}\mathbf{v}_i^T\right\|$$

$$= O(\alpha_n^2/t_k^2). \tag{55}$$

Then it follows from (55) and (8) that

$$\mathbf{b}_{\mathbf{u},k,t_k}^T = \sum_{i=1}^n a_i\mathbf{v}_i^T - \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t_k)[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)]^{-1}\mathbf{V}_{-k}^T$$

and

$$\left\|\mathbf{b}_{\mathbf{u},k,t_k} - a_k\mathbf{v}_k - \sum_{1 \leq i \neq k \leq K} a_i[1 + (t_k d_i^{-1} - 1)^{-1}]\mathbf{v}_i - \sum_{i=K+1}^n a_i\mathbf{v}_i\right\| = O(\alpha_n^2/t_k^2). \tag{56}$$

We denote by $\mathbf{c}_k = a_k\mathbf{v}_k + \sum_{1 \leq i \neq k \leq K} a_i[1 + (t_k d_i^{-1} - 1)^{-1}]\mathbf{v}_i + \sum_{i=K+1}^n a_i\mathbf{v}_i$. By (56), we can obtain

$$-(\mathbf{u}^T\mathbf{v}_k)^2 + \mathbf{b}_{\mathbf{u},k,t_k}^T\mathbf{b}_{\mathbf{u},k,t_k} = -a_k^2 + \|\mathbf{c}_k\|^2 + \|\mathbf{b}_{\mathbf{u},k,t_k} - \mathbf{c}_k\|^2 + 2(\mathbf{b}_{\mathbf{u},k,t_k} - \mathbf{c}_k)^T\mathbf{c}_k$$

$$= \sum_{1 \leq i \neq k \leq K} a_i^2[1 + (t_k d_i^{-1} - 1)^{-1}]^2 + \sum_{i=K+1}^n a_i^2 + O(\alpha_n^2/t_k^2)$$

$$+ \text{ some small order term,} \tag{57}$$

where the small order term takes a rather complicated form and thus we omit its expression for simplicity. Since by assumption $|\mathbf{u}^T\mathbf{v}_k| \in [0, 1 - \epsilon]$, $\mathbf{u} = \sum_{i=1}^n a_i\mathbf{v}_i$ is a unit vector, and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is an orthogonal matrix, it holds that

$$\sum_{1 \leq i \neq k \leq n}^n a_i^2 \geq 1 - (1 - \epsilon)^2. \tag{58}$$

Moreover, Condition 3 and Lemma 3 together entail that $|t_k d_i^{-1}|$ is bounded away from 0 and 1. Thus there exists some positive constant $c_2 < 1$ such that

$$[1 + (t_k d_i^{-1} - 1)^{-1}]^2 \geq c_2 \tag{59}$$

for each $1 \leq i \neq k \leq K$. Therefore, combining (53) and (57)–(59), and by the assumption $\sigma_{\min} \gg \alpha_n/t_k$, we can obtain the desired claim in (51), which completes the proof of Theorem 3.

A.4 Proof of Corollary 1

The conclusions of Corollary 1 follow directly from the results of Theorem 3.

A.5 Proof of Theorem 4

The more general asymptotic theory in Theorem 4 focuses on the first order asymptotic expansion for the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ with $\mathbf{x}$ and $\mathbf{y}$ two arbitrary n-dimensional unit vectors, while that in Theorem 5 further establishes the higher order (which is second order) asymptotic expansion for the same bilinear form. We begin with the analysis for the first order asymptotic expansion. The main ingredients of the proof are as follows. First, we represent $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ as an integral which is a functional of $\mathbf{X} = \mathbf{H} + \mathbf{W}$. By doing so we can deal with the matrix $\mathbf{H} + \mathbf{W}$ instead of the eigenvectors. Second, for the functional of $\mathbf{H} + \mathbf{W}$ obtained in the previous step we extract the $\mathbf{H}$ part from $\mathbf{H} + \mathbf{W}$ and further obtain a functional of $\mathbf{W}$. Roughly speaking, we can get an explicit function of form $f((\mathbf{W} - t\mathbf{I})^{-1})$ with $|t| \gg \|\mathbf{W}\|$. Third, by the matrix series expansion $(\mathbf{W} - t\mathbf{I})^{-1} = -\sum_{l=0}^{\infty} t^{-(l+1)} \mathbf{W}^l$, the function $f((\mathbf{W} - t\mathbf{I})^{-1})$ can be approximated by $f(-\sum_{l=0}^{L} t^{-(l+1)} \mathbf{W}^l)$ for some positive integer L. Fourth, we can then calculate the first (second or higher) order expansion of $f(-\sum_{l=0}^{L} t^{-(l+1)} \mathbf{W}^l)$ since we have an explicit expression of function f.

To facilitate our technical derivations, let us recall some basic matrix identities from the Sherman–Morrison–Woodbury formula. For any matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{F}$ of appropriate dimensions and any vectors $\mathbf{a}$ and $\mathbf{b}$ of appropriate dimensions, it holds that

$$(\mathbf{A} + \mathbf{BFC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{F}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \tag{60}$$

and

$$(\mathbf{C} + \mathbf{a}\mathbf{b}^T)^{-1}\mathbf{a} = \frac{\mathbf{C}^{-1}\mathbf{a}}{1 + \mathbf{b}^T\mathbf{C}^{-1}\mathbf{a}} \tag{61}$$

when the corresponding matrices for matrix inversion are nonsingular.

To illustrate the main ideas of our proof, we first consider the simple case of $K = 1$ and $\mathbf{x} = \mathbf{y} = \mathbf{v}_1$. The general case of $K \geq 1$ and arbitrary unit vectors will be discussed later. Let Ω_1 be a contour centered at $(a_1 + b_1)/2$ with radius $|b_1 - a_1|/2$, where the quantities a_k and b_k with $1 \leq k \leq K$ are defined in Section 3.2. Then it is seen that d_1 is enclosed by Ω_1. In view of Condition 2, Lemma 6, and Weyl's inequality, we have

$$|\lambda_1 - d_1| \leq \|\mathbf{W}\| < \min\{|d_1 - a_1|, |d_1 - b_1|\}$$

and

$$|\lambda_j - d_1| \geq |d_1| - \|\mathbf{W}\| > \max\{|d_1 - a_1|, |d_1 - b_1|\}, \; j \geq 2$$

with significant probability. We can see that the contour Ω_1 does not enclose any other eigenvalues λ_j with $j \neq 1$. Thus, by Cauchy's residue theorem from complex analysis, we have with significant probability

$$-\frac{1}{2\pi i} \oint_{\Omega_1} \frac{1}{\lambda_1 - z} dz = 1 \ \text{ and } \ -\frac{1}{2\pi i} \oint_{\Omega_1} \frac{1}{\lambda_j - z} dz = 0, \ j \geq 2,$$

where i associated with the complex integrals represents the imaginary unit $(-1)^{1/2}$ and the line integrals are taken over the contour Ω_1. Noticing that $(\mathbf{X} - z\mathbf{I})^{-1} = \sum_{i=1}^n (\lambda_j - z)^{-1} \widehat{\mathbf{v}}_j \widehat{\mathbf{v}}_j^T$, we can then obtain an integral representation of the desired bilinear form that with significant probability

$$\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1 = -\frac{\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1}{2\pi i} \oint_{\Omega_1} \frac{1}{\lambda_1 - z} dz = -\frac{1}{2\pi i} \oint_{\Omega_1} \mathbf{v}_1^T \Big(\sum_{j=1}^n \frac{\widehat{\mathbf{v}}_j \widehat{\mathbf{v}}_j^T}{\lambda_j - z} \Big) \mathbf{v}_1 dz$$

$$= -\frac{1}{2\pi i} \oint_{\Omega_1} \mathbf{v}_1^T \widetilde{\mathbf{G}}(z) \mathbf{v}_1 dz, \tag{62}$$

where the matrix-valued function $\widetilde{\mathbf{G}}(z) = (\mathbf{X} - z\mathbf{I})^{-1}$ for z in the complex plane $\mathbb{C}$ is referred to as the Green function associated with the original random matrix $\mathbf{X} = \mathbf{H} + \mathbf{W}$.

Note that by (1) and $K = 1$ for the simple case, we have $\mathbf{X} = \mathbf{H} + \mathbf{W} = d_1 \mathbf{v}_1 \mathbf{v}_1^T + \mathbf{W}$. Thus the line integral in (62) can be rewritten as

$$\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1 = -\frac{1}{2\pi i} \oint_{\Omega_1} \mathbf{v}_1^T (\mathbf{W} - z\mathbf{I} + d_1 \mathbf{v}_1 \mathbf{v}_1^T)^{-1} \mathbf{v}_1 dz. \tag{63}$$

With the aid of (60) and (61), the line integral in (63) can be further represented as

$$\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1 = -\frac{1}{2\pi i} \oint_{\Omega_1} \frac{\mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1}{1 + d_1 \mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1} dz. \tag{64}$$

To analyze the integrand of the line integral on the right hand side of (64), we first consider the term $(\mathbf{W} - z\mathbf{I})^{-1}$. Such a term admits the matrix series expansion

$$(\mathbf{W} - z\mathbf{I})^{-1} = -\sum_{l=0}^{\infty} z^{-(l+1)} \mathbf{W}^l. \tag{65}$$

Let L be the smallest positive integer such that

$$\frac{\alpha_n^{L+1} (\log n)^{(L+1)/2}}{|d_K|^{L-2}} \to 0. \tag{66}$$

Such an integer L always exists since $|d_K|/(n^\epsilon \alpha_n) \to \infty$ for small positive constant ϵ by Condition (2) and $\alpha_n \leq n^{1/2}$ by definition. Since we consider z on the contour Ω_1, it follows that $|z| \geq c|d_1|$ for some positive constant c. Thus, by (65), Condition 1, and Lemma 6 in Section B.7 of Supplementary Material, with the above choice of L in (66) we have with

probability tending to one that

$$\left\| \sum_{l=L+1}^{\infty} z^{-(l+1)} \mathbf{W}^l \right\| \leq \sum_{l=L+1}^{\infty} \frac{C^l \alpha_n^l (\log n)^{l/2}}{|z|^{l+1}} = \frac{O\{C^{L+1}\alpha_n^{L+1}(\log n)^{(L+1)/2}\}}{|z|^{L+2}} = \frac{O(1)}{|z|^4}, \quad (67)$$

where C is some positive constant. In light of (65) and (67), we can obtain the asymptotic expansion

$$\begin{aligned}
\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1 &= -\sum_{l=0}^{L} z^{-(l+1)}\mathbf{v}_1^T\mathbf{W}^l\mathbf{v}_1 - \sum_{l=L+1}^{\infty} z^{-(l+1)}\mathbf{v}_1^T\mathbf{W}^l\mathbf{v}_1 \\
&= -\sum_{l=0}^{L} z^{-(l+1)}\mathbf{v}_1^T\mathbf{W}^l\mathbf{v}_1 + \frac{O_p(1)}{d_1^4}
\end{aligned} \quad (68)$$

for z on the contour Ω_1.

Directly working with the line integral in (62) or (64) is challenging in deriving the CLT for the bilinear form $\mathbf{v}_1^T\widehat{\mathbf{v}}_1\widehat{\mathbf{v}}_1^T\mathbf{v}_1$. Next we introduce some simple facts about Cauchy's residue theorem. Assume that a complex function $f(z)$ is a holomorphic function inside Ω_1 except at one point t. Then it holds that

$$\frac{1}{2\pi i}\oint_{\Omega_1} f(z)dz = \text{Res}(f,t),$$

where $\text{Res}(f,t)$ represents the residue of function f at point t. In addition, assume that the Laurent series expansion of f around point t is given by

$$f(z) = \sum_{j=-\infty}^{\infty} a_j(z-t)^j$$

with a_j some constants. Then we have $\text{Res}(f,t) = (2\pi i)^{-1}\oint_{\Omega_1} f(z)dz = a_{-1}$. Furthermore, if $\lim_{z\to t}(z-t)f(z)$ exists then the Laurent series expansion of f entails that

$$\lim_{z\to t}(z-t)f(z) = a_{-1}. \quad (69)$$

Now let us consider the line integral in (64). Observe that the only singular point of function $\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1/[1 + d_1\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1]$ inside Ω_1 is the solution to equation

$$1 + d_1\mathbf{v}_1^T(\mathbf{W} - z\mathbf{I})^{-1}\mathbf{v}_1 = 0,$$

which we denote as $\widehat{t}_1$. Let us use $[(\mathbf{W} - \widehat{t}_1\mathbf{I})^{-1}]'$ as a shorthand notation for $h'(\widehat{t}_1)$ with

$h(t) = (\mathbf{W} - t\mathbf{I})^{-1}$. Then by Cauchy's residue theorem and in view of (64), we have

$$
\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1 = -\frac{1}{2\pi i} \oint_{\Omega_1} \frac{\mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1}{1 + d_1 \mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1} dz = -\lim_{z \to \widehat{t}_1} \frac{(z - \widehat{t}_1) \mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1}{1 + d_1 \mathbf{v}_1^T (\mathbf{W} - z\mathbf{I})^{-1} \mathbf{v}_1}
$$
$$
= -\frac{\mathbf{v}_1^T (\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1} \mathbf{v}_1}{d_1 \mathbf{v}_1^T [(\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1}]' \mathbf{v}_1}.
$$

Therefore, an application of the Taylor expansion to function $\mathbf{v}_1^T (\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1} \mathbf{v}_1 / \{d_1 \mathbf{v}_1^T [(\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1}]' \mathbf{v}_1\}$ yields

$$
-\frac{\mathbf{v}_1^T (\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1} \mathbf{v}_1}{d_1 \mathbf{v}_1^T [(\mathbf{W} - \widehat{t}_1 \mathbf{I})^{-1}]' \mathbf{v}_1} = \frac{\sum_{l=0}^L \widehat{t}_1^{-(l+1)} \mathbf{v}_1^T \mathbf{W}^l \mathbf{v}_1 + d_1^{-4} O_p(1)}{d_1 \sum_{l=0}^L (l+1) \widehat{t}_1^{-(l+2)} \mathbf{v}_1^T \mathbf{W}^l \mathbf{v}_1 + d_1^{-4} O_p(1)}. \tag{70}
$$

Note that $\widehat{t}_1$ is a random variable that depends on random matrix $\mathbf{X}$. In fact, from (99) we can see that the asymptotic expansion of $\widehat{t}_1$ is a polynomial of $\mathbf{v}_1^T \mathbf{W}^l \mathbf{v}_1$. Thus the asymptotic expansion of (70) is also a polynomial function of $\mathbf{v}_1^T \mathbf{W}^l \mathbf{v}_1$. Therefore, controlling the variance of $\mathbf{v}_1^T \mathbf{W}^l \mathbf{v}_1$ can facilitate us in identifying the leading term of the asymptotic expansion. So far we have laid out the major steps in deriving the asymptotic expansion for $\mathbf{v}_1^T \widehat{\mathbf{v}}_1 \widehat{\mathbf{v}}_1^T \mathbf{v}_1$. This can shed light on the detailed proof for the general case of $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ with $K \geq 1$.

We now move on to the general case of $K \geq 1$ and arbitrary n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$. The technical arguments for the general case are similar to those for the simple case of $K = 1$ and $\mathbf{x} = \mathbf{y} = \mathbf{v}_1$ presented above, but with more delicate technical derivations. Similarly as in (62), it follows from Cauchy's residue theorem, the definitions of the eigenvalue and eigenvector, and (1) that the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ for each $1 \leq k \leq K$ admits a natural integral representation; that is, with significant probability,

$$
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} = -\frac{1}{2\pi i} \oint_{\Omega_k} \mathbf{x}^T \widetilde{\mathbf{G}}(z) \mathbf{y} dz = -\frac{1}{2\pi i} \oint_{\Omega_k} \mathbf{x}^T \Big(\mathbf{W} - z\mathbf{I} + \sum_{j=1}^K d_j \mathbf{v}_j \mathbf{v}_j^T\Big)^{-1} \mathbf{y} dz
$$
$$
= \frac{1}{2\pi i} \oint_{\Omega_k} \frac{d_k \mathbf{x}^T \Big(\mathbf{W} - z\mathbf{I} + \sum_{1 \leq j \neq k \leq K} d_j \mathbf{v}_j \mathbf{v}_j^T\Big)^{-1} \mathbf{v}_k \mathbf{v}_k^T \Big(\mathbf{W} - z\mathbf{I} + \sum_{1 \leq j \neq k \leq K} d_i \mathbf{v}_i \mathbf{v}_i^T\Big)^{-1} \mathbf{y}}{1 + d_k \mathbf{v}_k^T \Big(\mathbf{W} - z\mathbf{I} + \sum_{1 \leq j \neq k \leq K} d_j \mathbf{v}_j \mathbf{v}_j^T\Big)^{-1} \mathbf{v}_k} dz, \tag{71}
$$

where the Green function $\widetilde{\mathbf{G}}(z)$ associated with the original random matrix $\mathbf{X}$ is defined in (62) and the line integral is taken over a contour Ω_k that is centered at $(a_k + b_k)/2$ with radius $|b_k - a_k|/2$. Then the contour Ω_k encloses the population eigenvalue d_k of the latent mean matrix $\mathbf{H}$. Note that in the representation above, we have used the results, which can

be derived from Condition 2, Lemma 6, and Weyl's inequality, that for each $j = 1, \cdots, K$,

$$|\lambda_k - d_k| \leq \|\mathbf{W}\| < \min\{|d_k - a_k|, |d_k - b_k|\},$$

$$|\lambda_j - d_k| \geq |d_j - d_k| - |\lambda_j - d_j| \geq |d_j - d_k| - \|\mathbf{W}\| > \max\{|d_k - a_k|, |d_k - b_k|\}$$

for $j \neq k$ with significant probability; that is, the contour Ω_k encloses λ_k but not any other eigenvalues with high probability.

An application of (60) leads to

$$\left(\mathbf{W} - z\mathbf{I} + \sum_{1 \leq j \neq k \leq K} d_j \mathbf{v}_j \mathbf{v}_j^T\right)^{-1} = \mathbf{G}(z) - \mathbf{G}(z)\mathbf{V}_{-k}\left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}\right]^{-1}\mathbf{V}_{-k}^T\mathbf{G}(z), \quad (72)$$

where the Green function $\mathbf{G}(z)$ associated with only the noise part $\mathbf{W}$ is defined in (37). To simplify the expression, let

$$\mathbf{F}_k(z) = \mathbf{G}(z)\mathbf{V}_{-k}[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}]^{-1}\mathbf{V}_{-k}^T\mathbf{G}(z) \qquad (73)$$

Then in view of (72), the last line integral in (71) can be further represented as

$$\mathbf{x}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{y} = \frac{1}{2\pi i}\oint_{\Omega_k} \frac{d_k\mathbf{x}^T\left[\mathbf{G}(z) - \mathbf{F}_k(z)\right]\mathbf{v}_k\mathbf{v}_k^T\left[\mathbf{G}(z) - \mathbf{F}_k(z)\right]\mathbf{y}}{1 + d_k\mathbf{v}_k^T\left[\mathbf{G}(z) - \mathbf{F}_k(z)\right]\mathbf{v}_k}dz. \qquad (74)$$

It is challenging to analyze the terms in (74) since the expression of $\mathbf{F}_k(z)$ is complicated and we need to study the asymptotic expansion of $\mathbf{F}_k(z)$ carefully. In the proof below, we will see that Lemma 4 in Section 6 is a key ingredient of the technical arguments; see Section B.5 of Supplementary Material for the proof of this lemma.

We will conduct detailed calculations for the asymptotic expansion of $\mathbf{F}_k(z)$. Let us choose L as the same positive integer as in (66). Then we have $\sum_{l=L+1}^{\infty} z^{-(l+1)}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O_p(|z|^{-4})$ for z on the contour Ω_k. It follows from Lemma 4 and Condition 2 that

$$\sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p\left\{\alpha_n|z|^{-3} + \alpha_n^2|z|^{-4} + \cdots + \alpha_n^{L-1}|z|^{-(L+1)}\right\}$$

$$= O_p(\alpha_n|z|^{-3}).$$

Therefore, similar to (68) we can show that

$$\mathbf{x}^T \mathbf{G}(z)\mathbf{y} = -z^{-1}\mathbf{x}^T\mathbf{y} - z^{-2}\mathbf{x}^T\mathbf{W}\mathbf{y} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T\mathbb{E}\mathbf{W}^l\mathbf{y} - \sum_{l=L+1}^{\infty} z^{-(l+1)}\mathbf{x}^T\mathbf{W}^l\mathbf{y}$$

$$- \sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}$$

$$= -z^{-1}\mathbf{x}^T\mathbf{y} - z^{-2}\mathbf{x}^T\mathbf{W}\mathbf{y} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T\mathbb{E}\mathbf{W}^l\mathbf{y} + O_p\left(|z|^{-4} + \alpha_n|z|^{-3}\right). \quad (75)$$

Moreover, since for $z \in \Omega_k$ we have $|z|^{-4} \le \alpha_n|z|^{-3}$ by Condition 1, we can further obtain

$$\mathbf{x}^T \mathbf{G}(z)\mathbf{y} = -z^{-1}\mathbf{x}^T\mathbf{y} - z^{-2}\mathbf{x}^T\mathbf{W}\mathbf{y} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T\mathbb{E}\mathbf{W}^l\mathbf{y} + O_p(\alpha_n|z|^{-3}). \quad (76)$$

In fact, the probabilistic event associated with the small order term $O_p(\alpha_n|z|^{-3})$ in (76) holds uniformly over z since the term $O_p(\alpha_n|z|^{-3})$ is simply $|z|^{-3}O_p(\alpha_n)$.

To simplify the technical presentation, hereafter we use the generic notation $\mathbf{u}$ to denote either $\mathbf{x}$ or $\mathbf{y}$ unless specified otherwise, which means that the corresponding derivations and results hold when $\mathbf{u}$ is replaced by $\mathbf{x}$ and $\mathbf{y}$. Since $\mathbf{x}$ and $\mathbf{y}$ can be chosen as any unit vectors, we can obtain from (76) the following asymptotic expansions by different choices of $\mathbf{x}$ and $\mathbf{y}$

$$\mathbf{u}^T \mathbf{G}(z)\mathbf{v}_k = -z^{-1}\mathbf{u}^T\mathbf{v}_k - z^{-2}\mathbf{u}^T\mathbf{W}\mathbf{v}_k - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{u}^T\mathbb{E}\mathbf{W}^l\mathbf{v}_k + O_p(\alpha_n|z|^{-3}), \quad (77)$$

$$\mathbf{v}_k^T \mathbf{G}(z)\mathbf{v}_k = -z^{-1} - z^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{v}_k^T\mathbb{E}\mathbf{W}^l\mathbf{v}_k + O_p(\alpha_n|z|^{-3}), \quad (78)$$

$$\mathbf{v}_k^T \mathbf{G}(z)\mathbf{V}_{-k} = -z^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{V}_{-k} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{v}_k^T\mathbb{E}\mathbf{W}^l\mathbf{V}_{-k} + O_p(\alpha_n|z|^{-3}), \quad (79)$$

$$\mathbf{u}^T \mathbf{G}(z)\mathbf{V}_{-k} = -z^{-1}\mathbf{u}^T\mathbf{V}_{-k} - z^{-2}\mathbf{u}^T\mathbf{W}\mathbf{V}_{-k} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{u}^T\mathbb{E}\mathbf{W}^l\mathbf{V}_{-k}$$

$$+ O_p(\alpha_n|z|^{-3}), \quad (80)$$

$$\mathbf{V}_{-k}^T \mathbf{G}(z)\mathbf{V}_{-k} = -z^{-1}\mathbf{I} - z^{-2}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{V}_{-k} - \sum_{l=2}^{L} z^{-(l+1)}\mathbf{V}_{-k}^T\mathbb{E}\mathbf{W}^l\mathbf{V}_{-k} + O_p(\alpha_n|z|^{-3}). \quad (81)$$

Thus it follows from (76)–(81) that

$$
\begin{aligned}
\mathbf{u}^T\mathbf{F}_k(z)\mathbf{v}_k &= \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \\
&\quad - z^{-2}\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k \\
&\quad - z^{-2}\mathbf{u}^T\mathbf{W}\mathbf{V}_{-k}\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \\
&\quad + z^{-2}\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{V}_{-k} \\
&\quad \times \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) + O_p(\alpha_n|z|^{-3}), \\
&= \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \\
&\quad - z^{-2}\mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k + O_p(\alpha_n|z|^{-3}) \quad (82)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{v}_k^T\mathbf{F}_k(z)\mathbf{v}_k &= \mathbf{v}_k^T\mathbf{G}(z)\mathbf{V}_{-k}\left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{V}_{-k}\right]^{-1}\mathbf{V}_{-k}^T\mathbf{G}(z)\mathbf{v}_k \\
&= \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \\
&\quad - z^{-2}\mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k + O_p(\alpha_n|z|^{-3}) \\
&= \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)\right]^{-1}\mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) + O_p(\alpha_n|z|^{-3}), \quad (83)
\end{aligned}
$$

where $\mathbf{F}_k(z)$ is defined in (73) and $\mathcal{R}$ is defined in (6).

With all the technical preparations above, we are now ready to analyze the terms in representation (74). Specifically let us consider the ratio $\{d_k\mathbf{x}^T[\mathbf{G}(z) - \mathbf{F}_k(z)]\mathbf{v}_k\mathbf{v}_k^T[\mathbf{G}(z) - \mathbf{F}_k(z)]\mathbf{y}\}/\{1 + d_k\mathbf{v}_k^T[\mathbf{G}(z) - \mathbf{F}_k(z)]\mathbf{v}_k\}$ that appears as the integrand on the left hand side of (74). Similar to (75), taking the derivative of $\mathbf{G}(z)$ we have

$$
\begin{aligned}
\mathbf{x}^T\mathbf{G}'(z)\mathbf{y} = \mathbf{x}^T(\mathbf{W} - z\mathbf{I})^{-2}\mathbf{y} &= \sum_{l=0}^{\infty}(l+1)z^{-(l+2)}\mathbf{x}^T\mathbf{W}^l\mathbf{y} \\
&= \mathcal{R}'(\mathbf{x}, \mathbf{y}, z) + 2z^{-3}\mathbf{x}^T\mathbf{W}\mathbf{y} + z^{-4}O_p(\alpha_n). \quad (84)
\end{aligned}
$$

It follows from Lemmas 4–5 that

$$
\begin{aligned}
\mathcal{R}'(\mathbf{v}_k, \mathbf{V}_{-k}, z) &= O(\alpha_n^2/z^4),\ \ \mathcal{R}'(\mathbf{v}_k, \mathbf{v}_k, z) - \frac{1}{z^2} = O(\alpha_n^2/z^4), \\
\left\|\mathcal{R}'(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) - z^{-2}\mathbf{I}\right\| &= O(\alpha_n^2/z^4). \quad (85)
\end{aligned}
$$

By (79) and Lemmas 4–5, we can conclude that

$$
\mathbf{v}_k^T\mathbf{G}(z)\mathbf{V}_{-k} = z^{-2}O_p(1) + |z|^{-3}O_p(\alpha_n^2). \quad (86)
$$

Moreover, by (80) and (A.16) we have

$$
\left\| \left\{ \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} - \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\}' \right\|
$$

$$
= \left\| \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \mathbf{V}_{-k}^T \mathbf{G}'(z) \mathbf{V}_{-k} \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \right.
$$

$$
\left. - \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathcal{R}'(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\|
$$

$$
= O \left\{ \left\| \mathbf{V}_{-k}^T \mathbf{G}'(z) \mathbf{V}_{-k} - \mathcal{R}'(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right\| \left\| \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \right\|^2 \right\}
$$

$$
+ O \left\{ \left\| \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} - \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\| \right.
$$

$$
\left. \cdot \left\| \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \mathcal{R}'(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right\| \right\}
$$

$$
= |z|^{-1} O_p(1) + z^{-2} O_p(\alpha_n) \tag{87}
$$

and

$$
\left\| \left\{ \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\}' \right\|
$$

$$
= \left\| \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathcal{R}'(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\|
$$

$$
= O(1). \tag{88}
$$

Note that in light of (84)–(87), we can obtain

$$
\mathbf{v}_k^T \mathbf{F}_k'(z) \mathbf{v}_k = 2 \mathbf{v}_k^T \mathbf{G}'(z) \mathbf{V}_{-k} \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{v}_k
$$

$$
+ \mathbf{v}_k^T \mathbf{G}(z) \mathbf{V}_{-k} \left\{ \left[\mathbf{D}_{-k}^{-1} + \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{V}_{-k} \right]^{-1} \right\}' \mathbf{V}_{-k}^T \mathbf{G}(z) \mathbf{v}_k
$$

$$
= 2 \mathcal{R}'(\mathbf{v}_k, \mathbf{V}_{-k}, z) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)
$$

$$
+ \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \left\{ \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\}' \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)
$$

$$
+ z^{-4} O_p(1) + z^{-6} O_p(\alpha_n^3). \tag{89}
$$

Combining the above result with (84) leads to

$$
d_k \mathbf{v}_k^T \left[\mathbf{G}'(z) - \mathbf{F}_k'(z) \right] \mathbf{v}_k = \frac{d_k}{z^2 \widetilde{\mathcal{P}}_{k,z}} + 2 z^{-3} d_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + z^{-4} O_p(|d_k| \alpha_n) \tag{90}
$$

for $z \in [a_k, b_k]$. Further, recalling the definition in (7) and by (88), it holds that

$$
\frac{1}{z^2 \widetilde{\mathcal{P}}_{k,z}} = \left(\frac{A_{\mathbf{v}_k, k, z}}{z} \right)' = \mathcal{R}'(\mathbf{v}_k, \mathbf{v}_k, z) - 2 \mathcal{R}'(\mathbf{v}_k, \mathbf{V}_{-k}, z) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1}
$$

$$
\times \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \left\{ \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \right]^{-1} \right\}' \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z)
$$

$$
= z^{-2} + O(\alpha_n^2 / z^4). \tag{91}
$$

40

Plugging this into (90) and by Lemmas 4–5, we have for all $z \in [a_k, b_k]$,

$$d_k \mathbf{v}_k^T \left[\mathbf{G}'(z) - \mathbf{F}_k'(z) \right] \mathbf{v}_k = d_k z^{-2} + 2z^{-3} d_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + z^{-4} O_p(|d_k| \alpha_n^2)$$
$$= d_k z^{-2} \left[1 + O_p(|z|^{-1} + |z|^{-2} \alpha_n^2) \right] = d_k z^{-2} \left[1 + o_p(1) \right]. \tag{92}$$

Thus $1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z) \right] \mathbf{v}_k$ is a monotone function with probability tending to one.

Further, in light of expressions (78) and (83) we can obtain the asymptotic expansion

$$1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z) \right] \mathbf{v}_k = f_k(z) - d_k z^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + z^{-2} O_p(\alpha_n) \tag{93}$$

for all $z \in [a_k, b_k]$, where $f_k(z)$ is defined in (10). Note that $f_k(a_k) = O(1)$, $f_k(b_k) = O(1)$, and $f_k(a_k) f_k(b_k) < 0$ as shown in the proof of Lemma 3 in Section B.4 of Supplementary Material. These results together with (92), which gives the order for the derivative of $1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z) \right] \mathbf{v}_k$, entail that there exists a unique solution $\widehat{t}_k$ to the equation

$$1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z) \right] \mathbf{v}_k = 0 \tag{94}$$

for z in the interval $[a_k, b_k]$. Using Lemma 4, we can further show that (93) becomes

$$1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z) \right] \mathbf{v}_k - f_k(z) = -\frac{d_k}{z^2} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(|z|^{-2} \alpha_n) = O_p(|z|^{-1}) \tag{95}$$

for $z \in [a_k, b_k]$. Note that $f_k(z)$ is a monotone function over $z \in [a_k, b_k]$ as shown in the proof of Lemma 3 and (A.17). Thus it follows from (94) and (95) that

$$\widehat{t}_k - t_k = O_p(1). \tag{96}$$

In fact, we can obtain a more precise order of $\widehat{t}_k - t_k$ than the initial one in (96). In view of (93) and the definition of t_k, we have

$$1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(t_k) - \mathbf{F}_k(t_k) \right] \mathbf{v}_k = -d_k t_k^{-2} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}). \tag{97}$$

By (92) and (97), an application of the mean value theorem yields

$$0 = 1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k) \right] \mathbf{v}_k = 1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(t_k) - \mathbf{F}_k(t_k) \right] \mathbf{v}_k$$
$$+ d_k \widetilde{t}_k^{-2} \left[1 + O_p(|d_k|^{-1} + |d_k|^{-2} \alpha_n^2) \right] (\widehat{t}_k - t_k), \tag{98}$$

where $\widetilde{t}_k$ is some number between t_k and $\widehat{t}_k$. The asymptotic expansions in (98) and (97) entail further that

$$\widehat{t}_k - t_k = \frac{t_k^2}{\widetilde{t}_k^2} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-1}) = \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-1}). \tag{99}$$

Now by the similar arguments as for obtaining (69), the integral (74) can be evaluated as

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} &= \frac{1}{2\pi i} \oint_{\Omega_k} \frac{d_k \mathbf{x}^T \left[\mathbf{G}(z) - \mathbf{F}_k(z)\right] \mathbf{v}_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z)\right] \mathbf{y}}{1 + d_k \mathbf{v}_k^T \left[\mathbf{G}(z) - \mathbf{F}_k(z)\right] \mathbf{v}_k} dz \\
&= \frac{\widehat{t}_k^2 \mathbf{x}^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{v}_k \mathbf{v}_k^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{y}}{\widehat{t}_k^2 \mathbf{v}_k^T \left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}_k'(\widehat{t}_k)\right] \mathbf{v}_k}.
\end{aligned} \tag{100}
$$

By (90) we have

$$
\frac{1}{\widehat{t}_k^2 \mathbf{v}_k^T \left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}_k'(\widehat{t}_k)\right] \mathbf{v}_k} = \widetilde{\mathcal{P}}_{k,\widehat{t}_k} - 2\widehat{t}_k^{-1} \widetilde{\mathcal{P}}_{k,\widehat{t}_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + \widehat{t}_k^{-2} O_p(\alpha_n) \tag{101}
$$

and (100) can be written as

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} &= \frac{\widehat{t}_k^2 \mathbf{x}^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{v}_k \mathbf{v}_k^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{y}}{\widehat{t}_k^2 \mathbf{v}_k^T \left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}_k'(\widehat{t}_k)\right] \mathbf{v}_k} \\
&= \left[\widetilde{\mathcal{P}}_{k,\widehat{t}_k} - 2\widehat{t}_k^{-1} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + \widehat{t}_k^{-2} O_p(\alpha_n)\right] \widehat{t}_k^2 \mathbf{x}^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{v}_k \mathbf{v}_k^T \\
&\quad \times \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{y}.
\end{aligned} \tag{102}
$$

Recall the definitions in (6) and (7). Then it follows from (77), (82), and (99) that

$$
\begin{aligned}
\widehat{t}_k \mathbf{u}^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right] \mathbf{v}_k &= \mathcal{P}(\mathbf{u}, \mathbf{v}_k, \widehat{t}_k) - \mathcal{P}(\mathbf{u}, \mathbf{V}_{-k}, \widehat{t}_k) \left[\widehat{t}_k \mathbf{D}_{-k}^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, \widehat{t}_k)\right]^{-1} \\
&\quad \times \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, \widehat{t}_k) - \widehat{t}_k^{-1} \mathbf{u}^T \mathbf{W} \mathbf{v}_k + \widehat{t}_k^{-1} \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, \widehat{t}_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, \widehat{t}_k)\right]^{-1} \\
&\quad \times \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n \widehat{t}_k^{-2}) \\
&= \mathcal{P}(\mathbf{u}, \mathbf{v}_k, t_k) - \mathcal{P}(\mathbf{u}, \mathbf{V}_{-k}, t_k) \left[t_k \mathbf{D}_{-k}^{-1} + \mathcal{P}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)\right]^{-1} \\
&\quad \times \mathcal{P}(\mathbf{V}_{-k}, \mathbf{v}_k, t_k) - t_k^{-1} \mathbf{u}^T \mathbf{W} \mathbf{v}_k + t_k^{-1} \mathcal{R}(\mathbf{u}, \mathbf{V}_{-k}, t_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, t_k)\right]^{-1} \\
&\quad \times \mathbf{V}_{-k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}), \\
&= A_{\mathbf{u},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{u},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}),
\end{aligned} \tag{103}
$$

where $\mathbf{u}$ stands for both $\mathbf{x}$ and $\mathbf{y}$ as mentioned before. Furthermore, by Lemma 5 and (99) we can conclude that

$$
\widetilde{\mathcal{P}}_{k,\widehat{t}_k} = \widetilde{\mathcal{P}}_{k,t_k} + O_p(\alpha_n^2 t_k^{-3}). \tag{104}
$$

Combining the representation (102) and asymptotic expansions (103)–(104), by Lemma 4

we can deduce that (100) can be further written as

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} &= \frac{\widehat{t}_k^2 \mathbf{x}^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k) \right] \mathbf{v}_k \mathbf{v}_k^T \left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k) \right] \mathbf{y}}{\widehat{t}_k^2 \mathbf{v}_k^T \left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}_k'(\widehat{t}_k) \right] \mathbf{v}_k} \\
&= \left[\widetilde{\mathcal{P}}_{k,t_k} - 2t_k^{-1} \widetilde{\mathcal{P}}_{k,t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}) \right] \left[A_{\mathbf{x},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{x},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}) \right] \\
&\quad \times \left[A_{\mathbf{y},k,t_k} - t_k^{-1} \mathbf{b}_{\mathbf{y},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}) \right] \\
&= \left[\widetilde{\mathcal{P}}_{k,t_k} - 2t_k^{-1} \widetilde{\mathcal{P}}_{k,t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}) \right] \times \left[A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} \right. \\
&\quad - t_k^{-1} \left(A_{\mathbf{x},k,t_k} \mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k} \mathbf{b}_{\mathbf{y},k,t_k}^T \right) \mathbf{W} \mathbf{v}_k \\
&\quad \left. + t_k^{-2} \mathbf{b}_{\mathbf{x},k,t_k}^T \mathbf{W} \mathbf{v}_k \mathbf{b}_{\mathbf{y},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n c_k t_k^{-2}) \right],
\end{aligned}
\tag{105}
$$

where $c_k = |A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}| + |t_k|^{-1}$.

We can expand (105), or equivalently (100), further as

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} &= \left[\widetilde{\mathcal{P}}_{k,t_k} - 2t_k^{-1} \widetilde{\mathcal{P}}_{k,t_k}^2 \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n t_k^{-2}) \right] \times \left[A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} \right. \\
&\quad \left. - t_k^{-1} \left(A_{\mathbf{x},k,t_k} \mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k} \mathbf{b}_{\mathbf{y},k,t_k}^T \right) \mathbf{W} \mathbf{v}_k + t_k^{-2} \mathbf{b}_{\mathbf{x},k,t_k}^T \mathbf{W} \mathbf{v}_k \mathbf{b}_{\mathbf{y},k,t_k}^T \mathbf{W} \mathbf{v}_k + O_p(\alpha_n c_k t_k^{-2}) \right] \\
&= \widetilde{\mathcal{P}}_{k,t_k} A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} - t_k^{-1} A_{\mathbf{x},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \left(\mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T \right) \mathbf{W} \mathbf{v}_k \\
&\quad - t_k^{-1} A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \left(\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{x},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T \right) \mathbf{W} \mathbf{v}_k \\
&\quad + t_k^{-2} \widetilde{\mathcal{P}}_{k,t_k} \left[2 \widetilde{\mathcal{P}}_{k,t_k} \left(A_{\mathbf{x},k,t_k} \mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k} \mathbf{b}_{\mathbf{y},k,t_k}^T \right) \mathbf{W} \mathbf{v}_k \mathbf{v}_k^T + \mathbf{b}_{\mathbf{x},k,t_k}^T \mathbf{W} \mathbf{v}_k \mathbf{b}_{\mathbf{y},k,t_k}^T \right] \mathbf{W} \mathbf{v}_k \\
&\quad - 2 t_k^{-3} \widetilde{\mathcal{P}}_{k,t_k}^2 \mathbf{b}_{\mathbf{x},k,t_k}^T \mathbf{W} \mathbf{v}_k \mathbf{b}_{\mathbf{y},k,t_k}^T \mathbf{W} \mathbf{v}_k \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k + O_p \left\{ \alpha_n c_k t_k^{-2} \right\}.
\end{aligned}
\tag{106}
$$

Therefore, we have characterized the terms involving t_k^{-1} for the desired first order asymptotic expansion. That is, by (106) we have

$$
\begin{aligned}
\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y} &= \widetilde{\mathcal{P}}_{k,t_k} A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} - t_k^{-1} A_{\mathbf{x},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \left(\mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T \right) \mathbf{W} \mathbf{v}_k \\
&\quad - t_k^{-1} A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \left(\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{x},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T \right) \mathbf{W} \mathbf{v}_k + O_p \left\{ (\alpha_n c_k + 1) t_k^{-2} \right\}.
\end{aligned}
\tag{107}
$$

Thus if $\sigma_k^2 = t_k^{-2} \widetilde{\mathcal{P}}_{k,t_k}^2 \mathbb{E}[(A_{\mathbf{x},k,t_k} \mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k} \mathbf{b}_{\mathbf{x},k,t_k}^T + 2 A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T) \mathbf{W} \mathbf{v}_k]^2 \gg (\alpha_n c_k + 1)^2 t_k^{-4} \sim \sigma_n^2 (|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|)^2 t_k^{-4} + t_k^{-4}$ and $(A_{\mathbf{x},k,t_k} \mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k} \mathbf{b}_{\mathbf{x},k,t_k}^T + 2 A_{\mathbf{x},k,t_k} A_{\mathbf{y},k,t_k} \widetilde{\mathcal{P}}_{k,t_k} \mathbf{v}_k^T, \mathbf{v}_k)$ is $\mathbf{W}^1$-CLT, then (33) holds, where $\sim$ means the asymptotic order. This concludes the proof of Theorem 4.

A.6 Proof of Theorem 5

We have characterized the first order asymptotic expansion for the bilinear form $\mathbf{x}^T \widehat{\mathbf{v}}_k \widehat{\mathbf{v}}_k^T \mathbf{y}$ in the proof of Theorem 4 in Section A.5, where $\mathbf{x}$ and $\mathbf{y}$ are two arbitrary n-dimensional unit vectors. We now proceed with investigating the higher order (which is second order) asymptotic expansion for the same bilinear form. More specifically, the proof of Theorem 5

involves further expansion for the $O_p\{\alpha_n c_k t_k^{-2}\}$ term given in (106).

To gain some intuition, let us recall (75) and compare with (77)–(81). By Lemma 4, we see that the order $O_p(\alpha_n|z|^{-3})$ comes from the terms of form $\mathbf{x}^T(\mathbf{W}^2-\mathbb{E}\mathbf{W}^2)\mathbf{y}/z^3$. Therefore, to obtain a higher order expansion we need to identify all terms of form $\mathbf{x}^T(\mathbf{W}^2-\mathbb{E}\mathbf{W}^2)\mathbf{y}/z^3$. It follows from (75) and Lemmas 4 and 5 that

$$
\mathbf{x}^T\mathbf{G}(z)\mathbf{y} = -z^{-1}\mathbf{x}^T\mathbf{y} - z^{-2}\mathbf{x}^T\mathbf{W}\mathbf{y} - \frac{\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y}}{z^3}
$$
$$
- \sum_{l=2}^{L} z^{-(l+1)}\mathbf{x}^T\mathbb{E}\mathbf{W}^l\mathbf{y} + O_p\left(|z|^{-4} + \alpha_n^2|z|^{-4}\right). \tag{108}
$$

Moreover, using similar arguments as for proving (101) and (103) but expanding to higher orders we can obtain

$$
\left\{\widehat{t}_k^2\mathbf{v}_k^T\left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}_k'(\widehat{t}_k)\right]\mathbf{v}_k\right\}^{-1} = \widetilde{\mathcal{P}}_{k,t_k}\Big\{1 - 2t_k^{-1}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\widetilde{\mathcal{P}}_{k,t_k}
$$
$$
\times\left[3\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k - 2(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k)^2\right]\Big\} + O_p(\alpha_n^2|t_k|^{-3}) \tag{109}
$$

and

$$
\widehat{t}_k\mathbf{u}^T\left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right]\mathbf{v}_k = A_{\mathbf{u},k,t_k} - t_k^{-1}\mathbf{u}^T\mathbf{W}\mathbf{v}_k
$$
$$
+ t_k^{-1}\mathcal{R}(\mathbf{u},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k + t_k^{-2}\mathbf{u}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k
$$
$$
- t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{u},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k
$$
$$
+ t_k^{-2}\mathcal{R}(\mathbf{u},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k
$$
$$
- t_k^{-2}\mathbf{u}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k + 2t_k^{-3}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{u},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-2}
$$
$$
\times\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k + O_p(\alpha_n^2|t_k|^{-3}), \tag{110}
$$

where $\mathbf{u}$ represents both $\mathbf{x}$ and $\mathbf{y}$ as mentioned before.

Using the representations (100) and (102), and by the asymptotic expansions (109)–(110),

44

we can obtain the $O_p(t_k^{-2})$ term for the desired second order asymptotic expansion as follows

$$\mathbf{x}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{y} = \frac{\widehat{t}_k^2\mathbf{x}^T\left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right]\mathbf{v}_k\mathbf{v}_k^T\left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right]\mathbf{y}}{\widehat{t}_k^2\mathbf{v}_k^T\left[\mathbf{G}'(\widehat{t}_k) - \mathbf{F}'_k(\widehat{t}_k)\right]\mathbf{v}_k}$$

$$= \left(\widetilde{\mathcal{P}}_{k,t_k} \times \left\{1 - 2t_k^{-1}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\widetilde{\mathcal{P}}_{k,t_k}\left[3\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k - 2(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k)^2\right]\right\}\right.$$

$$\left. + O_p(\alpha_n^2|t_k|^{-3})\right)\left[\widehat{t}_k\mathbf{x}^T\left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right]\mathbf{v}_k\right]\left[\widehat{t}_k\mathbf{v}_k^T\left[\mathbf{G}(\widehat{t}_k) - \mathbf{F}_k(\widehat{t}_k)\right]\mathbf{y}\right]$$

$$= -A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-1}\left(\mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\right)\mathbf{W}\mathbf{v}_k - A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-1}$$

$$\times \left(\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\right)\mathbf{W}\mathbf{v}_k$$

$$+ \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\left[2\widetilde{\mathcal{P}}_{k,t_k}\left(A_{\mathbf{x},k,t_k}\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k}\mathbf{b}_{\mathbf{y},k,t_k}^T\right)\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T + \mathbf{b}_{\mathbf{x},k,t_k}^T\mathbf{W}\mathbf{v}_k\mathbf{b}_{\mathbf{y},k,t_k}^T\right]\mathbf{W}\mathbf{v}_k$$

$$+ 2A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\left(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\right)^2 + A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\left\{t_k^{-2}\mathbf{x}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{x},\mathbf{V}_{-k},t)\right.$$

$$\times \left.\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\right\}$$

$$+ A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\left\{t_k^{-2}\mathbf{y}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{y},\mathbf{V}_{-k},t)\right.$$

$$\times \left.\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\right\}$$

$$+ A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{x},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k$$

$$+ A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{y},\mathbf{V}_{-k},t_k)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k$$

$$- \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}(A_{\mathbf{y},k,t_k}\mathbf{x}^T + A_{\mathbf{x},k,t_k}\mathbf{y}^T)(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k$$

$$- 3t_k^{-2}A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k + O_p\left\{(\alpha_n^2c_k + 1)|t_k|^{-3}\right\}. \tag{111}$$

In contrast to the small order term $O_p\{\alpha_n c_k t_k^{-2}\}$ in (106) from the first order asymptotic expansion, we now have the small order term $O_p\{(\alpha_n^2 c_k + 1)|t_k|^{-3}\}$ from the second order asymptotic expansion.

Let us conduct some simplifications for the expressions given in the above asymptotic expansions in (111). A combination of (106) and (111) shows that the asymptotic distribution

is determined by

$$
\begin{aligned}
& - A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-1}\left(\mathbf{b}_{\mathbf{y},k,t_k}^T + A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\right)\mathbf{W}\mathbf{v}_k \\
& \quad - A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-1}\left(\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\right)\mathbf{W}\mathbf{v}_k \\
& \quad + \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\Big[2\widetilde{\mathcal{P}}_{k,t_k}\left(A_{\mathbf{x},k,t_k}\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k}\mathbf{b}_{\mathbf{y},k,t_k}^T\right)\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T + \mathbf{b}_{\mathbf{x},k,t_k}^T\mathbf{W}\mathbf{v}_k\mathbf{b}_{\mathbf{y},k,t_k}^T\Big]\mathbf{W}\mathbf{v}_k \\
& \quad + 2A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}(\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k)^2 + A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\Big\{t_k^{-2}\mathbf{x}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{x},\mathbf{V}_{-k},t) \\
& \quad \times \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\Big\} \\
& \quad + A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\Big\{t_k^{-2}\mathbf{y}^T\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k - t_k^{-2}\mathbf{v}_k^T\mathbf{W}\mathbf{v}_k\mathcal{R}(\mathbf{y},\mathbf{V}_{-k},t) \\
& \quad \times \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T\mathbf{W}\mathbf{v}_k\Big\} \\
& \quad + A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{x},\mathbf{V}_{-k},t)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k \\
& \quad + A_{\mathbf{x},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathcal{R}(\mathbf{y},\mathbf{V}_{-k},t)\left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k \\
& \quad - \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\left(A_{\mathbf{y},k,t_k}\mathbf{x}^T + A_{\mathbf{x},k,t_k}\mathbf{y}^T\right)(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k \\
& \quad - 3t_k^{-2}A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{v}_k.
\end{aligned}
\tag{112}
$$

To further simplify the notation, we define three terms

$$
\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} = -\widetilde{\mathcal{P}}_{k,t_k}t_k^{-1}\mathbf{v}_k\left(A_{\mathbf{y},k,t_k}\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{x},k,t_k}\mathbf{b}_{\mathbf{y},k,t_k}^T + 2A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\widetilde{\mathcal{P}}_{k,t_k}\mathbf{v}_k^T\right), \tag{113}
$$

$$
\begin{aligned}
\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k} = \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}\mathbf{v}_k\Big\{ & \left[A_{\mathbf{y},k,t_k}\mathcal{R}(\mathbf{x},\mathbf{V}_{-k},t) + A_{\mathbf{x},k,t_k}\mathcal{R}(\mathbf{y},\mathbf{V}_{-k},t)\right] \\
& \times \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k},\mathbf{V}_{-k},t_k)\right]^{-1}\mathbf{V}_{-k}^T + A_{\mathbf{y},k,t_k}\mathbf{x}^T + A_{\mathbf{x},k,t_k}\mathbf{y}^T \\
& + 3A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\mathbf{v}_k^T\Big\},
\end{aligned}
\tag{114}
$$

$$
\begin{aligned}
\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k} = & \mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k} - \widetilde{\mathcal{P}}_{k,t_k}t_k^{-2}A_{\mathbf{x},k,t_k}A_{\mathbf{y},k,t_k}\mathbf{v}_k\mathbf{v}_k^T \\
& + 2\widetilde{\mathcal{P}}_{k,t_k}^2 t_k^{-2}\mathbf{v}_k\left(A_{\mathbf{x},k,t_k}\mathbf{b}_{\mathbf{x},k,t_k}^T + A_{\mathbf{y},k,t_k}\mathbf{b}_{\mathbf{y},k,t_k}^T\right).
\end{aligned}
\tag{115}
$$

Note that all the three matrices defined in (113)–(115) are of rank one and the identity $\mathbf{x}^T\mathbf{A}\mathbf{y} = \text{tr}(\mathbf{A}\mathbf{y}\mathbf{x}^T)$ holds for any matrix $\mathbf{A}$ and vectors $\mathbf{x}$ and $\mathbf{y}$. Thus in view of (113)–(115), the lengthy expression given in (112) can be rewritten in a compact form as

$$
\text{tr}\left[\mathbf{W}\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - \left(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2\right)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}\right] + \text{tr}\left(\mathbf{W}\mathbf{v}_k\mathbf{v}_k^T\right)\text{tr}\left(\mathbf{W}\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}\right). \tag{116}
$$

So far we have shown that the second order expansion of $\mathbf{x}^T\widehat{\mathbf{v}}_k\widehat{\mathbf{v}}_k^T\mathbf{y}$ is given in (111). Note that $\widetilde{\sigma}_k^2$ defined in (32) is essentially the variance of (116). Thus if $\widetilde{\sigma}_k^2 \gg (\alpha_n^2 c_k + 1)^2 t_k^{-6} \sim \sigma_n^4(|A_{\mathbf{x},k,t_k}| + |A_{\mathbf{y},k,t_k}|)^2 t_k^{-6} + t_k^{-6}$, then (116) is the leading term of (111). Furthermore, the assumption of $\sigma_k^2 = O(\widetilde{\sigma}_k^2)$ entails that the first order expansion in Theorem 4 does not dominate the second order expansion. Therefore, we see that the asymptotic distribution in Theorem 5 is determined by the joint distribution of the three random variables specified in

expression (116). This completes the proof of Theorem 5.

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research 18*(1), 6446–6531.

Abbe, E., J. Fan, K. Wang, and Y. Zhong (2019). Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics, to appear*.

Arias-Castro, E., N. Verzelen, et al. (2014). Community detection in dense random networks. *The Annals of Statistics 42*(3), 940–969.

Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl. 20*, 262–268.

Arnold, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Probability Theory and Related Fields 19*, 191–198.

Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica 9*, 611–677.

Bai, Z. D. and J. W. Silverstein (2006). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Bai, Z. D. and J. F. Yao (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44*, 447–474.

Baik, J., G. B. Arous, and S. Péché (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability 33*, 682–693.

Bao, Z., X. Ding, and K. Wang (2018). Singular vector and singular subspace distribution for the matrix denoising model. *arXiv preprint arXiv:1809.10476*.

Barber, R. F. and E. J. Candès (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics 43*, 2055–2085.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences 106*, 21068–21073.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Bourgade, P. and H.-T. Yau (2017). The eigenvector moment flow and local quantum unique ergodicity. *Communications in Mathematical Physics 350*, 231–278.

Bourgade, P., H.-T. Yau, and J. Yin (2018). Random band matrices in the delocalized phase, i: Quantum unique ergodicity and universality. *arXiv preprint arXiv:1807.01559*.

Candès, E. J., Y. Fan, L. Janson, and J. Lv (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society Series B 80*, 551–577.

Capitaine, M. and C. Donati-Martin (2018). Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of wigner matrices. *arXiv preprint arXiv:1807.07773*.

Capitaine, M., C. Donati-Martin, and D. Féral (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. Poincaré Probab. Statist. 48*, 107–133.

Chen, K. and J. Lei (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association 113*, 241–251.

Decelle, A., F. Krzakala, C. Moore, and L. Zdeborová (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E 84*, 066106.

Dekel, Y., J. R. Lee, and N. Linial (2007). Eigenvectors of random graphs: Nodal domains. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 436–448. Springer.

El Karoui, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab. 35*, 663–714.

Erdős, L., A. Knowles, H.-T. Yau, and J. Yin (2013). Delocalization and diffusion profile for random band matrices. *Communications in Mathematical Physics 323*, 367–416.

Erdös, L., H.-T. Yau, and J. Yin (2011). Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics 229*, 1435–1515.

Fan, J., Y. Fan, X. Han, and J. Lv (2019). SIMPLE: statistical inference on membership profiles in large networks. *arXiv preprint arXiv:1910.01734*.

Fan, J., W. Wang, and Y. Zhong (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Reserarch 18*, 1–42.

Fan, Y., E. Demirkaya, G. Li, and J. Lv (2019). RANK: large-scale inference with graphical nonlinear knockoffs. *Journal of the American Statistical Association, to appear*.

Fan, Y., J. Lv, M. Sharifvaghefi, and Y. Uematsu (2019). IPAD: stable interpretable forecasting with knockoffs inference. *Journal of the American Statistical Association, to appear*.

Füredi, Z. and J. Komlós (1981). The eigenvalues of random symmetric matrices. *Combinatorica 1*, 233–241.

Han, X., Q. Yang, and Y. Fan (2019). Universal rank inference via residual subsampling with application to large networks. *arXiv preprint arXiv:1912.11583*.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis (2nd edition)*. Cambridge University Press.

Jin, J., Z. T. Ke, and S. Luo (2017). Estimating network memberships by simplex vertex hunting. *https://arxiv.org/pdf/1708.07852.pdf*.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*, 295–327.

Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist. 36*, 2638–2716.

Johnstone, I. M. and A. Y. Lu (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association 104*, 682–693.

Knowles, A. and J. Yin (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math. 66*, 1663–1749.

Knowles, A. and J. Yin (2014). The outliers of a deformed Wigner matrix. *The Annals of Probability 42*, 1980–2031.

Knowles, A. and J. Yin (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields 169*, 257–352.

Koltchinskii, V. and K. Lounici (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincaré Probab. Statist. 52*, 1976–2013.

Koltchinskii, V. and D. Xia (2016). Perturbation of linear forms of singular vectors under Gaussian noise. *In: Houdré C., Mason D., Reynaud-Bouret P., Rosiński J. (eds) High Dimensional Probability VII*, 397–423.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics 44*, 401–424.

Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika 107*(2), 257–276.

Lu, Y., Y. Fan, J. Lv, and W. S. Noble (2018). DeepPINK: reproducible feature selection in deep neural networks. *Advances in Neural Information Processing Systems (NeurIPS 2018)*.

Marchenko, V. A. and L. A. Pastur (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik 1*, 457–483.

McSherry, F. (2001). Spectral partitioning of random graphs. *Proceedings of the Fourty-Second IEEE Symposium on Foundations of Computer Science, FOCS*, 529–537.

Mehta, M. L. (2004). *Random Matrices (3rd edition)*. Academic Press.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica. 17*, 1617–1642.

Pizzo, A., D. Renfrew, and A. Soshnikov (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat. 49*, 64–94.

Renfrew, D. and A. Soshnikov (2013). On finite rank deformations of Wigner matrices ii: Delocalized perturbations. *Random Matrices: Theory Appl. 2*, 1250015.

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics 39*, 1878–1915.

Rudelson, M. and R. Vershynin (2016). No-gaps delocalization for general random matrices. *Geometric and Functional Analysis 26*, 1716–1776.

Spielman, D. A. and S.-H. Teng (2007). Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and Its Applications 421*, 284–305.

Tang, M. and C. E. Priebe (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *The Annals of Statistics 46*, 2360–2415.

Tao, T. (2004). *Topics in Random Matrix Theory*. American Mathematical Society.

Tracy, C. A. and H. Widom (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys. 159*, 151–174.

Tracy, C. A. and H. Widom (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys. 177*, 727–754.

Tropp, J. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math. 12*, 389–434.

Verzelen, N., E. Arias-Castro, et al. (2015). Community detection in sparse random networks. *The Annals of Applied Probability 25*(6), 3465–3510.

Vu, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combinatorics, Probability and Computing 27*, 124–140.

Wang, W. and J. Fan (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics 45*, 1342–1374.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math. 62*, 548–564.

Yau, H.-T. (2012). Universality of generalized Wigner matrices. *Quantum Theory from Small to Large Scales: Lecture Notes of the Les Houches Summer School 95*, 675–692.

Zhang, Y., E. Levina, and J. Zhu (2015). Detecting overlapping communities in networks using spectral methods. *https://arxiv.org/pdf/1412.3432.pdf*.

Supplementary Material to "Asymptotic Theory of Eigenvectors for Random Matrices with Diverging Spikes"

Jianqing Fan, Yingying Fan, Xiao Han and Jinchi Lv

This Supplementary Material contains additional technical details. In particular, we present in Section B the proofs of all the lemmas and provide in Section C some further technical details on under what regularity conditions the asymptotic normality can hold for the asymptotic expansion in Theorem 5. Section D contains the technical details on relaxing the spike strength condition when considering scenario ii) of Condition 2 in place of scenario i), as well as the proof sketch for results in Section 4.2.

B Proofs of technical lemmas

B.1 Proof of Lemma 1

Let $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$ be two arbitrary n-dimensional unit vectors. Since $\mathbf{W}$ is a symmetric random matrix of independent entries above the diagonal, it is easy to show that

$$\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y} = \sum_{1 \leq i,j \leq n,\, i<j} w_{ij}(x_iy_j + x_jy_i) + \sum_{1 \leq i \leq n} (w_{ii} - \mathbb{E}w_{ii})(x_iy_i) \qquad (A.1)$$

and

$$s_n^2 \equiv \mathbb{E}(\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y})^2 = \sum_{1 \leq i,j \leq n,\, i<j} \mathbb{E}w_{ij}^2(x_iy_j + x_jy_i)^2 + \sum_{1 \leq i \leq n} \mathbb{E}(w_{ii} - \mathbb{E}w_{ii})^2 x_i^2 y_i^2. \quad (A.2)$$

Since w_{ij} with $1 \leq i < j \leq n$ and $w_{ii} - \mathbb{E}w_{ii}$ with $1 \leq i \leq n$ are independent random variables with zero mean, by the Lyapunov condition (see, for example, Theorem 27.3 of Billingsley (1995)) we can see that if

$$\frac{1}{s_n^3}\left[\sum_{1 \leq i,j \leq n,\, i<j} \mathbb{E}|w_{ij}|^3|x_iy_j + x_jy_i|^3 + \sum_{1 \leq i \leq n} \mathbb{E}|w_{ii} - \mathbb{E}w_{ii}|^3|x_iy_i|^3\right] \to 0,$$

then it holds that

$$\frac{\mathbf{x}^T\mathbf{W}\mathbf{y} - \mathbf{x}^T\mathbb{E}\mathbf{W}\mathbf{y}}{s_n} \xrightarrow{\mathscr{D}} N(0,1).$$

Since by assumption $\max_{1 \le i,j \le n} |w_{ij}| \le 1$ and $\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \ll s_n$, we have

$$\frac{1}{s_n^3}\left[\sum_{1 \le i,j \le n,\, i<j} \mathbb{E}|w_{ij}|^3 |x_i y_j + x_j y_i|^3 + \sum_{1 \le i \le n} \mathbb{E}|w_{ii} - \mathbb{E}w_{ii}|^3 |x_i y_i|^3 \right]$$

$$\le \frac{2}{s_n^3}\left[\sum_{1 \le i,j \le n,\, i<j} \mathbb{E}|w_{ij}|^2 |x_i y_j + x_j y_i|^3 + \sum_{1 \le i \le n} \mathbb{E}|w_{ii} - \mathbb{E}w_{ii}|^2 |x_i y_i|^3 \right]$$

$$\ll \frac{2s_n}{s_n^3}\left[\sum_{1 \le i,j \le n,\, i<j} \mathbb{E}|w_{ij}|^2 |x_i y_j + x_j y_i|^2 + \sum_{1 \le i \le n} \mathbb{E}|w_{ii} - \mathbb{E}w_{ii}|^2 |x_i y_i|^2 \right] \le 2, \quad \text{(A.3)}$$

which completes the proof of Lemma 1.

B.2   Proof of Lemma 2

The technical arguments for the proof of Lemma 2 are similar to those for the proof of Lemma 1 in Section B.1. For the case of $\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y}$, let us first consider the term $\mathbf{x}^T\mathbf{W}^2\mathbf{y}$. Such a term can be written as

$$\sum_{1 \le k,i,l \le n} w_{ki} w_{il} x_k y_l = \sum_{1 \le k,i,l \le n,\, k>l} w_{ki} w_{il}(x_k y_l + x_l y_k) + \sum_{1 \le k,i \le n} w_{ki}^2 x_k y_k$$

$$= \sum_{1 \le k,i,l \le n,\, k>l,\, k<i} w_{ki} w_{il}(x_k y_l + x_l y_k) + \sum_{1 \le k,i,l \le n,\, k>l,\, k>i} w_{ki} w_{il}(x_k y_l + x_l y_k)$$

$$+ \sum_{1 \le l < k \le n} w_{kk} w_{kl}(x_k y_l + x_l y_k) + \sum_{1 \le k,i \le n} w_{ki}^2 x_k y_k$$

$$= \sum_{1 \le k,i,l \le n,\, k>l,\, k<i} w_{ki} w_{il}(x_k y_l + x_l y_k) + \sum_{1 \le k,i,l \le n,\, i>l,\, i>k} w_{ik} w_{kl}(x_i y_l + x_l y_i)$$

$$+ \sum_{1 \le l < k \le n} w_{kk} w_{kl}(x_k y_l + x_l y_k) + \sum_{1 \le k,i \le n} w_{ki}^2 x_k y_k$$

$$= \sum_{1 \le k < i \le n} w_{ki}\left(x_k \sum_{1 \le l < k \le n} w_{il} y_l + y_k \sum_{1 \le l < k \le n} w_{il} x_l + x_i \sum_{1 \le l < i \le n} w_{kl} y_l + y_i \sum_{1 \le l < i \le n} w_{kl} x_l \right)$$

$$+ \sum_{1 \le l < k \le n} w_{kk} w_{kl}(x_k y_l + x_l y_k) + \sum_{1 \le k < i \le n} w_{ki}^2(x_k y_k + x_i y_i) + \sum_{1 \le k \le n} w_{kk}^2 x_k y_k. \quad \text{(A.4)}$$

Then it follows from (A.4) and the independence of entries w_{ki} with $1 \le k \le i \le n$ that

$$\mathbb{E}\mathbf{x}^T\mathbf{W}^2\mathbf{y} = \sum_{1 \le k,i \le n,\, k<i} \mathbb{E}w_{ki}^2(x_k y_k + x_i y_i) + \sum_{1 \le k \le n} \mathbb{E}w_{kk}^2 x_k y_k.$$

To ease the technical presentation, let us define some new notation $\omega_{kk} = 2^{-1} w_{kk}$ and

$\sigma_{kk}^2 = \mathbb{E}\omega_{kk}^2$. We can further show that

$$
\begin{aligned}
\mathbf{x}^T(\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{y} = &\sum_{1 \le k,i \le n,\, k<i} w_{ki}\Big[x_k \sum_{1 \le l<k \le n} w_{il}y_l + y_k \sum_{1 \le l<k \le n} w_{il}x_l + x_i \sum_{1 \le l<i \le n} w_{kl}y_l \\
&+ y_i \sum_{1 \le l<i \le n} w_{kl}x_l + \mathbb{E}w_{ii}(x_iy_k + x_ky_i) \Big] + \sum_{1 \le k,i \le n,\, k<i} \Big[(w_{ki}^2 - \sigma_{ki}^2)(x_ky_k + x_iy_i) \\
&+ 2(\omega_{kk}^2 - \sigma_{kk}^2)(x_ky_k + x_iy_i) \Big] + \sum_{1 \le k \le n} 2(\omega_{kk} - \mathbb{E}\omega_{kk})\Big(x_k \sum_{1 \le l<k \le n} w_{kl}y_l \\
&+ y_k \sum_{1 \le l<k \le n} w_{kl}x_l \Big),
\end{aligned}
\tag{A.5}
$$

where $\sigma_{ki}^2 = \mathbb{E}w_{ki}^2$ denotes the variance of entry w_{ki} as defined before.

We next define a σ-algebra $\mathcal{F}_t = \sigma\{\mathfrak{w}_1, \cdots, \mathfrak{w}_t\}$, where $\mathfrak{w}_t = w_{kl}$ with $t = k + 2^{-1}l(l-1)$ and $1 \le k \le l \le n$. Clearly we have $t \le 2^{-1}n(n+1)$. In fact, there is a one to one correspondence between $t \le 2^{-1}n(n+1)$ and (k,l) with $k \le l$. Suppose that such a statement is not true. Then there exist two different pairs (k_1, l_1) and (k_2, l_2) with $1 \le k_1 \le l_1 \le n$ and $1 \le k_2 \le l_2 \le n$ such that

$$
k_1 + \frac{l_1(l_1-1)}{2} = k_2 + \frac{l_2(l_2-1)}{2}.
\tag{A.6}
$$

It is easy to see that we must have $k_1 \ne k_2$ and $l_1 \ne l_2$. Without loss of generality, let us assume that $l_1 < l_2$. Then by (A.6), it holds that

$$
\frac{l_2(l_2-1)}{2} - \frac{l_1(l_1-1)}{2} = k_1 - k_2 \le k_1 - 1.
$$

On the other hand, since $l_1 < l_2$ we have

$$
\frac{l_2(l_2-1)}{2} - \frac{l_1(l_1-1)}{2} \ge \frac{l_1(l_1+1)}{2} - \frac{l_1(l_1-1)}{2} \ge l_1 \ge k_1,
$$

which contradicts the previous inequality. Thus we have shown that there is indeed a one to one correspondence between $t \le 2^{-1}n(n+1)$ and (k,l) with $k \le l$.

Assume that $t_1 \le t_2$ with $t_1 = k_1 + 2^{-1}l_1(l_1-1)$ and $t_2 = k_2 + 2^{-1}l_2(l_2-1)$. Then using the similar arguments we can show that $l_1 \le l_2$ and further $k_1 \le k_2$ when $l_1 = l_2$. This means that for $t = k + 2^{-1}l(l-1)$ with $1 \le k \le l \le n$, we have $\mathcal{F}_t = \sigma\{\mathfrak{w}_1, \cdots, \mathfrak{w}_t\} = \sigma\{w_{ij} : 1 \le i \le j < l \text{ or } 1 \le i \le k \le j = l\}$. With such a representation, we can see that the expression in (A.5) is in fact a sum of martingale differences with respect to the σ-algebra $\mathcal{F}_{k+2^{-1}i(i-1)}$. This fact entails that for $1 \le k \le i \le n$,

$$
\mathbb{E}\left[(w_{ki} - \mathbb{E}w_{ki})b_{ki} + (w_{ki}^2 - \mathbb{E}w_{ki}^2)c_{ki} | \mathcal{F}_{k+2^{-1}i(i-1)-1} \right] = 0,
$$

where $b_{ki} = x_k \sum_{1 \le l<k \le n} w_{il}y_l + y_k \sum_{1 \le l<k \le n} w_{il}x_l + x_i \sum_{1 \le l<i \le n} w_{kl}y_l + y_i \sum_{1 \le l<i \le n} w_{kl}x_l +$

$(1 - \delta_{ki})\mathbb{E}w_{ii}(x_iy_k + x_ky_i)$ with $\delta_{ki} = 1$ when $k = i$ and 0 otherwise, and $c_{ki} = x_ky_k + x_iy_i$. The conditional variance is given by

$$\sum_{1 \le k,i \le n,\, k < i} \mathbb{E}\left\{\left[w_{ki}b_{ki} + (w_{ki}^2 - \sigma_{ki}^2)c_{ki}\right]^2 |\mathcal{F}_{k+2^{-1}i(i-1)-1}\right\}$$

$$+ \sum_{1 \le k \le n} \mathbb{E}\left\{\left[(\omega_{kk} - \mathbb{E}\omega_{kk})b_{kk} + 2(\omega_{kk}^2 - \sigma_{kk}^2)c_{kk}\right]^2 |\mathcal{F}_{2^{-1}k(k+1)-1}\right\}$$

$$= \sum_{1 \le k,i \le n,\, k \le i} \sigma_{ki}^2 b_{ki}^2 + 2\sum_{1 \le k,i \le n,\, k \le i} \gamma_{ki}b_{ki}c_{ki} + \sum_{1 \le k,i \le n,\, k \le i} \kappa_{ki}c_{ki}^2, \tag{A.7}$$

where $\gamma_{ki} = \mathbb{E}w_{ki}^3$ and $\kappa_{ki} = \mathbb{E}(w_{ki}^2 - \sigma_{ki}^2)^2$ for $k \ne i$, and $\gamma_{kk} = 2(\mathbb{E}\omega_{kk}^3 - \sigma_{kk}^2\mathbb{E}\omega_{kk})$ and $\kappa_{kk} = 4\mathbb{E}(\omega_{kk}^2 - \sigma_{kk}^2)^2$.

The mean of the random variable in (A.7) can be calculated as

$$s_{\mathbf{x},\mathbf{y}}^2 = \mathbb{E}(A.7) = \sum_{1 \le k,i \le n,\, k \le i} \left[\kappa_{ki}(x_ky_k + x_iy_i)^2 + \sigma_{ki}^2 \sum_{1 \le l < k \le n} \sigma_{il}^2(x_ky_l + y_kx_l)^2 \right.$$

$$\left. + \sigma_{ki}^2 \sum_{1 \le l < i \le n} \sigma_{kl}^2(x_iy_l + y_ix_l)^2\right] + \sum_{1 \le k,i \le n,\, k \le i} \sigma_{ki}^2(1 - \delta_{ki})\left[\mathbb{E}(w_{ii} + w_{kk})\right]^2$$

$$\times (x_ky_i + x_iy_k)^2. \tag{A.8}$$

Moreover, the variance of the random variable in (A.7) is given by

$$\kappa_{\mathbf{x},\mathbf{y}} = \mathrm{var}(A.7) = \sum_{1 \le k_1,i_1,k_2,i_2 \le n,\, k_1 \le i_1,\, k_2 \le i_2} \mathbb{E}\left\{\left[\sigma_{k_1i_1}^2(z_{k_1i_1}^2 - \mathbb{E}z_{k_1i_1}^2)\right.\right.$$

$$\left. + 2\gamma_{k_1i_1}(x_{k_1}y_{k_1} + x_iy_i)z_{k_1i_1}\right]\left[\sigma_{k_2i_2}^2(z_{k_2i_2}^2 - \mathbb{E}z_{k_2i_2}^2)\right.$$

$$\left.\left. + 2\gamma_{k_2i_2}(x_{k_2}y_{k_2} + x_iy_i)z_{k_2i_2}\right]\right\}, \tag{A.9}$$

where $z_{ki} = \sum_{1 \le l < k \le n} w_{il}(x_ky_l + y_kx_l) + \sum_{1 \le l < i \le n} w_{kl}(x_iy_l + y_ix_l) + (1 - \delta_{ki})\mathbb{E}w_{ii}(x_iy_k + x_ky_i)$.

Let us recall the classical martingale CLT; see, for example, Lemma 9.12 of Bai and Silverstein (2006). If a martingale difference sequence (Y_t) with respect to a σ-algebra $\mathcal{F}_t$ satisfies the following conditions:

a) $\dfrac{\sum_{t=1}^T \mathbb{E}(Y_t^2|\mathcal{F}_{t-1})}{\sum_{t=1}^T \mathbb{E}Y_t^2} \xrightarrow{\mathrm{P}} 1$,

b) $\dfrac{\sum_{t=1}^T \mathbb{E}[Y_t^2 I(|Y_t|/\sqrt{\sum_{t=1}^T \mathbb{E}Y_t^2} \ge \epsilon)]}{\sum_{t=1}^T \mathbb{E}Y_t^2} \le \dfrac{\sum_{t=1}^T \mathbb{E}Y_t^4}{\epsilon^2(\sum_{t=1}^T \mathbb{E}Y_t^2)^2} \to 0$ for any $\epsilon > 0$,

then we have $\dfrac{\sum_{t=1}^T Y_t}{\sqrt{\sum_{t=1}^T \mathbb{E}Y_t^2}} \xrightarrow{\mathscr{D}} N(0,1)$ as $T \to \infty$, where $I(\cdot)$ denotes the indicator function. It follows from the assumption of $\kappa_{\mathbf{x},\mathbf{y}}^{1/4} \ll s_{\mathbf{x},\mathbf{y}}$ that

$$\frac{(A.7)}{\mathbb{E}(A.7)} \xrightarrow{\mathrm{P}} 1,$$

which shows that condition a) above is satisfied. Moreover, by the simple fact that for any

4

fixed i, $\mathbb{E}w_{li}^2 y_l^2 \leq 1$, and the assumptions that $s_{\mathbf{x},\mathbf{y}} \to \infty$ and $\|\mathbf{x}\|_\infty \|\mathbf{y}\|_\infty \to 0$, we have

$$
\sum_{1 \leq k,i \leq n,\, k<i} \mathbb{E}\Big\{ w_{ki} \Big[x_k \sum_{1 \leq l<k \leq n} w_{il}y_l + y_k \sum_{1 \leq l<k \leq n} w_{il}x_l + x_i \sum_{1 \leq l<i \leq n} w_{kl}y_l
$$
$$
+ y_i \sum_{1 \leq l<i \leq n} w_{kl}x_l + \mathbb{E}w_{ii}(x_iy_k + x_ky_i) \Big] \Big\}^4 + \sum_{1 \leq k \leq n} \mathbb{E}\Big[2(\omega_{kk} - \mathbb{E}\omega_{kk})
$$
$$
\times \Big(x_k \sum_{1 \leq l<k \leq n} w_{kl}y_l + y_k \sum_{1 \leq l<k \leq n} w_{kl}x_l \Big) \Big]^4
$$
$$
+ \sum_{1 \leq k,i \leq n,\, k<i} \Big\{ \mathbb{E}\big[(w_{ki}^2 - \sigma_{ki}^2)(x_ky_k + x_iy_i) \big]^4 + \mathbb{E}\big[(\omega_{kk}^2 - \sigma_{ki}^2)(x_ky_k + x_iy_i) \big]^4 \Big\} \ll s_{\mathbf{x},\mathbf{y}}^4,
$$

which entails that condition b) above is also satisfied. Therefore, an application of the martingale CLT concludes the proof of Lemma 2.

B.3 Further technical details on conditions of Lemma 2

Let us gain some further insights into the technical conditions in Lemma 2. Define $a_{kl} = x_ky_l + y_kx_l$ and note that $\kappa_{ij} = \mathbb{E}(w_{ij}^2 - \sigma_{ij}^2)^2 = \mathbb{E}w_{ij}^4 - \sigma_{ij}^4$. By the assumption of $|w_{ij}| \leq 1$, it is easy to see that $0 \leq \kappa_{ij} \leq \mathbb{E}w_{ij}^4 \leq \mathbb{E}w_{ij}^2 = \sigma_{ij}^2$. Then we can show that the random variable in (A.7) subtracted by its mean $s_{\mathbf{x},\mathbf{y}}^2$ can be represented as

$$
(A.7) - s_{\mathbf{x},\mathbf{y}}^2 = \sum_{1 \leq k,i \leq n,\, k \leq i} \sigma_{ki}^2 \Big[\sum_{1 \leq l<k \leq n} (w_{il}^2 - \sigma_{il}^2)a_{kl}^2 + \sum_{1 \leq l<i \leq n} (w_{kl}^2 - \sigma_{kl}^2)a_{il}^2
$$
$$
+ \sum_{1 \leq l_1,l_2<k \leq n,\, l_1 \neq l_2} w_{il_1} w_{il_2} a_{kl_1} a_{kl_2} + \sum_{1 \leq l_1,l_2<i \leq n,\, l_1 \neq l_2} w_{kl_1} w_{kl_2} a_{il_1} a_{il_2} \Big]
$$
$$
+ 2 \sum_{1 \leq k,i \leq n,\, k \leq i,\, 1 \leq l_1<k \leq n,\, 1 \leq l_2<i \leq n} \sigma_{ki}^2 w_{il_1} w_{kl_2} a_{kl_1} a_{il_2}
$$
$$
+ 2 \sum_{1 \leq k,i \leq n,\, k \leq i} \big[\gamma_{ki}a_{kk} + \sigma_{ki}^2 a_{ki}(1 - \delta_{ki})\mathbb{E}w_{ii} \big] \Big(\sum_{1 \leq l<k \leq n} w_{il}a_{kl} + \sum_{1 \leq l<i \leq n} w_{kl}a_{il} \Big). \quad (A.10)
$$

By (A.10) and (A.31), we have

$$
\kappa_{\mathbf{x},\mathbf{y}} = \mathbb{E}\left[(A.7) - s_{\mathbf{x},\mathbf{y}}^2\right]^2 \le C\Big\{\mathbb{E}\Big[\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2\sum_{1\le l<k\le n}(w_{il}^2-\sigma_{il}^2)a_{kl}^2\Big]^2
$$

$$
+\,\mathbb{E}\Big[\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2\sum_{1\le l<i\le n}(w_{kl}^2-\sigma_{kl}^2)a_{il}^2\Big]^2 + \mathbb{E}\Big(\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2\sum_{1\le l_1,l_2<k\le n,\, l_1\ne l_2}w_{il_1}w_{il_2}
$$

$$
\times\, a_{kl_1}a_{kl_2}\Big)^2 + \mathbb{E}\Big(\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2\sum_{1\le l_1,l_2<i\le n,\, l_1\ne l_2}w_{kl_1}w_{kl_2}a_{il_1}a_{il_2}\Big)^2
$$

$$
+\,\mathbb{E}\Big(\sum_{1\le k,i\le n,\, k\le i,\, 1\le l_1<k\le n,\, 1\le l_2<i\le n}\sigma_{ki}^2 w_{il_1}w_{kl_2}a_{kl_1}a_{il_2}\Big)^2 + \mathbb{E}\sum_{1\le k,i\le n,\, k\le i}\Big[\gamma_{ki}a_{kk}
$$

$$
+\,\sigma_{ki}^2 a_{ki}(1-\delta_{ki})\mathbb{E}w_{ii}\Big]\Big(\sum_{1\le l<k\le n}w_{il}a_{kl}+\sum_{1\le l<i\le n}w_{kl}a_{il}\Big)\Big\}
$$

$$
\le C\Big\{\Big(\sum_{1\le k,i\le n,\, k\le i}\sigma_{ki}^2\Big)^2\Big(\sum_{1\le l<k\le n}\kappa_{il}a_{kl}^4+\sum_{1\le l<i\le n}\kappa_{kl}a_{il}^4\Big)
$$

$$
+\sum_{1\le k_1,k_2,l_1,l_2\le n,\, l_1\ne l_2,\, l_1<k_1,\, l_2<k_2}\sigma_{k_1 i}^2\sigma_{k_2 i}^2\sigma_{il_1}^2\sigma_{il_2}^2 a_{k_1l_1}a_{k_1l_2}a_{k_2l_1}a_{k_2l_2}
$$

$$
+\sum_{1\le k,i_1,i_2,l_1,l_2\le n,\, l_1\ne l_2<\min\{i_1,i_2\}}\sigma_{ki_1}^2\sigma_{ki_2}^2\sigma_{kl_1}^2\sigma_{kl_2}^2 a_{i_1l_1}a_{i_1l_2}a_{i_2l_1}a_{i_2l_2}
$$

$$
+\sum_{1\le k,i,l_1,l_2\le n,\, k<i,\, l_1<k,\, l_2<i}\sigma_{ki}^2\sigma_{il_1}^2\sigma_{kl_2}^2 a_{kl_1}^2 a_{il_2}^2
$$

$$
+\sum_{1\le k,i\le n,\, k<i}\Big\{\gamma_{ki}^2 a_{kk}^2+\sigma_{ki}^4 a_{ki}^2(1-\delta_{ki})[\mathbb{E}(w_{ii}+w_{kk})]^2\Big\}\Big(\sum_{1\le l<k\le n}\sigma_{il}^2 a_{kl}^2+\sum_{1\le l<i\le n}\sigma_{kl}^2 a_{il}^2\Big)\Big\}
$$

$$
= O\left\{n\sigma_n^8\|\mathbf{x}\|_\infty^4\|\mathbf{y}\|_\infty^4\right\}, \tag{A.11}
$$

where C is some positive constant.

Given $\|\mathbf{x}\|_\infty\|\mathbf{y}\|_\infty \to 0$, it follows from (A.8) that

$$
s_{\mathbf{x},\mathbf{y}}^2 = \sum_{1\le k,i\le n,\, k<i}\Big[\kappa_{ki}(x_k y_k + x_i y_i)^2 + \sum_{1\le l<k\le n}\sigma_{il}^2(x_k y_l + y_k x_l)^2
$$

$$
+\sum_{1\le l<i\le n}\sigma_{kl}^2(x_i y_l + y_i x_l)^2\Big]
$$

$$
\ge \sigma_{\min}^2\sum_{1\le k,i\le n,\, k<i}\Big[\sum_{1\le l<k\le n}(x_k y_l + y_k x_l)^2 + \sum_{1\le l<i\le n}(x_i y_l + y_i x_l)^2\Big]
$$

$$
\ge c\sigma_{\min}^2 n, \tag{A.12}
$$

where $\sigma_{\min}^2$ is defined in Condition 3. Then we can exploit the upper bound on $\kappa_{\mathbf{x},\mathbf{y}}$ in (A.11) and the lower bound on $s_{\mathbf{x},\mathbf{y}}^2$ in (A.12) to simplify the conditions of Lemma 2, which can be reduced to

$$
\|\mathbf{x}\|_\infty\|\mathbf{y}\|_\infty \to 0, \quad \frac{\alpha_n^4\|\mathbf{x}\|_\infty^2\|\mathbf{y}\|_\infty^2}{n^{1/2}\sigma_{\min}^2} \to 0, \ \text{and}\ \sigma_{\min}^2 n \to \infty. \tag{A.13}
$$

Therefore, the conclusions of Lemma 2 hold as long as condition (A.13) is satisfied.

B.4 Proof of Lemma 3

In view of the definition of the function $f_k(z)$ defined in (10), we have

$$f_k'(z) = d_k \Big\{ \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, z) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \big[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \big]^{-1}$$
$$\times \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \Big\}'. \tag{A.14}$$

For $z \in [a_k, b_k]$, it follows from Lemma 5, Condition 2, and the definition of $\mathcal{R}$ in (6) that

$$\big\| \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) + z^{-1}\mathbf{I} \big\| = \Big\| -\sum_{l=2}^{L} z^{-(l+1)} \mathbf{V}_{-k}^T \mathbb{E}\mathbf{W}^l \mathbf{V}_{-k} \Big\|$$
$$\leq \sum_{l=2}^{L} z^{-(l+1)} \big\| \mathbf{V}_{-k}^T \mathbb{E}\mathbf{W}^l \mathbf{V}_{-k} \big\| = O(\alpha_n^2 |z|^{-3}). \tag{A.15}$$

Without loss of generality, we assume that $k \neq 1$. For l such that $|d_l| > |d_k|$, by (A.15) the diagonal entry of $\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)$ corresponding to d_l is given by

$$d_l^{-1} - z^{-1} + O(\alpha_n^2 |z|^{-3}) = (z - d_l)/(z d_l) + O(\alpha_n^2 |z|^{-3}).$$

By Condition 2, there exists some positive constant c such that $\max\{|a_k|, |b_k|\} \leq (1-c)|d_l|$. It follows that $|(z - d_l)/(z d_l)| \geq c/|z|$ and thus $|(z - d_l)/(z d_l) + O(\alpha_n^2 |z|^{-3})|^{-1} = O(|z|)$. For the remaining diagonal entry with $|d_l| < |d_k|$, there exists some positive constant c_1 such that $\min\{|a_k|, |b_k|\} \geq (1 + c_1)|d_l|$ and similarly we have $|(z - d_l)/(z d_l) + O(\alpha_n^2 |z|^{-3})|^{-1} = O(|z|)$. Thus it follows from (A.15) that the off diagonal entries of $\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z)$ are dominated by the diagonal ones, leading to

$$\Big\| \big[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \big]^{-1} \Big\| = O(|z|) \tag{A.16}$$

for all $z \in [a_k, b_k]$.

Next an application of Lemma 5 gives

$$\mathcal{R}'(\mathbf{v}_k, \mathbf{v}_k, z) = \sum_{l=0, l\neq 1}^{L} \frac{l+1}{z^{l+2}} \mathbf{v}_k^T \mathbb{E}\mathbf{W}^l \mathbf{v}_k = \frac{1}{z^2} + O(\alpha_n^2 |z|^{-4}).$$

By (A.14) and Condition 2, we have

$$\Big\{ \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, z) \big[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, z) \big]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, z) \Big\}' = O(\alpha_n^4 |z|^{-6}) = o(\alpha_n^2 |z|^{-4}).$$

Thus in view of (A.14), it holds that

$$f_k'(z) = d_k z^{-2} [1 + o(1)] \tag{A.17}$$

for $z \in [a_k, b_k]$. We can see from (A.17) that $f_k(z)$ is a monotone function over $z \in [a_k, b_k]$ when matrix size n is large enough.

Now recall that

$$f_k(d_k) = 1 + d_k \left\{ \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, d_k) - \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, d_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, d_k) \right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, d_k) \right\}.$$

By Lemma 5, we have

$$1 + d_k \mathcal{R}(\mathbf{v}_k, \mathbf{v}_k, d_k) = 1 - \sum_{l=0, l \neq 1}^{L} \frac{1}{d_k^l} \mathbf{v}_k^T \mathbb{E} \mathbf{W}^l \mathbf{v}_k = O(\alpha_n^2 d_k^{-2})$$

and

$$d_k \mathcal{R}(\mathbf{v}_k, \mathbf{V}_{-k}, d_k) \left[\mathbf{D}_{-k}^{-1} + \mathcal{R}(\mathbf{V}_{-k}, \mathbf{V}_{-k}, d_k) \right]^{-1} \mathcal{R}(\mathbf{V}_{-k}, \mathbf{v}_k, d_k) = O(\alpha_n^2 d_k^{-2}).$$

Thus it holds that $f_k(d_k) = O(\alpha_n^2 d_k^{-2}) = o(1)$. Noticing that the derivative $f_k'(z) = d_k z^{-2} [1 + o(1)] \sim d_k z^{-2} \sim |d_k|^{-1}$ and by the mean value theorem, we have $f_k(a_k) \sim o(1) + |d_k|^{-1}(a_k - d_k)$ and $f_k(b_k) \sim o(1) + |d_k|^{-1}(b_k - d_k)$, where $\sim$ represents the asymptotic order. Therefore, we see that $f_k(a_k) f_k(b_k) < 0$ and consequently the equation $f_k(z) = 0$ has a unique solution for $z \in [a_k, b_k]$, which solution satisfies that $t_k = d_k + o(d_k)$. This completes the proof of Lemma 3.

B.5 Proof of Lemma 4

The asymptotic bounds characterized in Lemma 4 play a key role in establishing the more general asymptotic theory in Theorems 4 and 5. We first assume that all the diagonal entries of $\mathbf{W} = (w_{ij})_{1 \leq i, j \leq n}$ are zero, that is, $w_{ii} = 0$. The general case of possibly $w_{ii} \neq 0$ will be dealt with later. The main idea of the proof is to calculate the moments by counting the number of nonzero terms involved in $\mathbb{E}(\mathbf{x}^T \mathbf{W}^l \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y})^2$, which is a frequently used idea in random matrix theory; see, for example, Chapter 2 of Bai and Silverstein (2006). An important difference is that bounding the order of $\mathbb{E}(\mathbf{x}^T \mathbf{W}^l \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y})^2$ by simply counting the number of nonzero terms inside is too rough for our setting since the variances of the entries of $\mathbf{W}$ can be very different from each other. Observe that the nonzero terms of the variance involve the product of w_{ij}^m with $m \geq 2$. We thus collect all such terms with the same index i but different index j, which means that we will bound $\sum_{j=1}^{n} \mathbb{E}|w_{ij}|^m \leq \alpha_n^2$ instead of using $\mathbb{E}|w_{ij}|^m \leq 1$. Then we can obtain a more accurate order since α_n^2 can be much smaller than n in general. Our technical arguments here provide useful refinements to the classical idea of counting the number of nonzero terms from the random matrix theory.

Let $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$ be two arbitrary n-dimensional unit vectors,

and $l \geq 1$ an integer. Expanding $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$ yields

$$
\begin{aligned}
&\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2 \\
&= \sum_{\substack{1 \leq i_1,\cdots,i_{l+1},j_1,\cdots,j_{l+1} \leq n, \\ i_s \neq i_{s+1}, j_s \neq j_{s+1}, 1 \leq s \leq l}} \mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big) \\
&\qquad\qquad \times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}}\big)\Big].
\end{aligned} \tag{A.18}
$$

Let $\mathbf{i} = (i_1,\cdots,i_{l+1})$ and $\mathbf{j} = (j_1,\cdots,j_{l+1})$ be two vectors taking values in $\{1,\cdots,n\}^{l+1}$. For any given vector $\mathbf{i}$, we define a graph $\mathcal{G}_{\mathbf{i}}$ whose vertices represent distinct values of the components of $\mathbf{i}$. Vertices i_s and i_{s+1} of $\mathcal{G}_{\mathbf{i}}$ are connected by undirected edges for $1 \leq s \leq l$. Similarly we can also define graph $\mathcal{G}_{\mathbf{j}}$ corresponding to $\mathbf{j}$. It can be seen that $\mathcal{G}_{\mathbf{i}}$ is a connected graph, which means that there exists some path from i_s to $i_{s'}$ for any $1 \leq s \neq s' \leq n$. Thus for each product

$$
\begin{aligned}
\mathbb{E}\Big[&\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big) \\
&\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}}\big)\Big],
\end{aligned} \tag{A.19}
$$

there exists a corresponding graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. If $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is not a connected graph, then the corresponding expectation

$$
\begin{aligned}
\mathbb{E}\Big[&\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big) \\
&\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}}\big)\Big] = 0.
\end{aligned}
$$

This shows that in order to calculate the order of $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$, it suffices to consider the scenario of connected graphs $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$.

To analyze the term in (A.59), let us calculate how many distinct vertices are contained in the connected graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since there are $2l$ edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ and $\mathbb{E}w_{ss'} = 0$ for $s \neq s'$, in order to get a nonzero value of (A.59) each edge in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ has at least one copy. Thus for each nonzero (A.59), we have l distinct edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is connected, there are at most $l+1$ distinct vertices in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Denote by $\mathcal{S}$ the set of all such pairs $(\mathbf{i},\mathbf{j})$. Combining the above arguments, we can conclude that

$$
\begin{aligned}
(A.18) = \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{S}} \mathbb{E}\Big[&\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big) \\
&\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}}\big)\Big].
\end{aligned} \tag{A.20}
$$

For notational simplicity, we denote $j_1,\cdots,j_{l+1}$ by $i_{l+2},\cdots,i_{2l+2}$ and define $\widetilde{\mathbf{i}} = (i_1,\cdots,i_{l+1}, j_1,\cdots,j_{l+1}) = (i_1,\cdots,i_{2l+2})$. We also denote $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ by $\mathcal{F}_{\widetilde{\mathbf{i}}}$ which has at most $l+1$ distinct

vertices and l distinct edges, with each edge having at least two copies. Then it holds that

$$
\begin{aligned}
|(A.60)| = \Big| &\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big[(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E} x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}) \\
&\times (x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} - \mathbb{E} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}})\big]\Big| \\
\leq &\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}\big| \\
+ &\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big| \mathbb{E}\big|x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}\big|. \quad (A.21)
\end{aligned}
$$

Observe that each expectation in (A.61) involves the product of some independent random variables, and $x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}$ and $x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}$ may share some dependency through factors $w_{ab}^{m_1}$ and $w_{ab}^{m_2}$, respectively, for some w_{ab} and nonnegative integers m_1 and m_2. Thus in light of the inequality

$$
\mathbb{E}|w_{ab}|^{m_1} \mathbb{E}|w_{ab}|^{m_2} \leq \mathbb{E}|w_{ab}|^{m_1 + m_2},
$$

we can further bound (A.61) as

$$
\begin{aligned}
(A.61) \leq 2 &\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots \\
&\times w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}\big|. \quad\quad\quad (A.22)
\end{aligned}
$$

To facilitate our technical presentation, let us introduce some additional notation. Denote by $\psi(2l+2)$ the set of partitions of the edges $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2})\}$ and $\psi_{\geq 2}(2l+2)$ the subset of $\psi(2l+2)$ whose blocks have size at least two. Let $P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)$ be the partition of $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2})\}$ that is associated with the equivalence relation $(i_{s_1}, i_{s_1+1}) \sim (i_{s_2}, i_{s_2+1})$ which is defined as if and only if $(i_{s_1}, i_{s_1+1}) = (i_{s_2}, i_{s_2+1})$ or $(i_{s_1}, i_{s_1+1}) = (i_{s_2+1}, i_{s_2})$. Denote by $|P(\widetilde{\mathbf{i}})| = m$ the number of groups in the partition $P(\widetilde{\mathbf{i}})$ such that the edges are equivalent within each group. We further denote the distinct edges in the partition $P(\widetilde{\mathbf{i}})$ as $(s_1, s_2), (s_3, s_4), \cdots, (s_{2m-1}, s_{2m})$ and the corresponding counts in each group as $r_1, \cdots, r_m$, and define $\widetilde{\mathbf{s}} = (s_1, s_2, \cdots, s_{2m})$. For the vertices, let $\phi(2m)$ be the set of partitions of $\{1, 2, \cdots, 2m\}$ and $Q(\widetilde{\mathbf{s}}) \in \phi(2m)$ the partition that is associated with the equivalence relation $a \sim b$ which is defined as if and only if $s_a = s_b$. Note that $s_{2j-1} \neq s_{2j}$

since the diagonal entries of $\mathbf{W}$ are assumed to be zero for the moment. Then we have

$$
\sum_{\mathcal{F}_{\widetilde{\mathbf{i}}},\widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big|x_{i_1}w_{i_1 i_2}w_{i_2 i_3}\cdots w_{i_l i_{l+1}}y_{i_{l+1}}x_{i_{l+2}}w_{i_{l+2}i_{l+3}}w_{i_{l+3}i_{l+4}}\cdots w_{i_{2l+1}i_{2l+2}}y_{i_{2l+2}}\big|
$$
$$
\leq \sum_{\substack{1\leq|P(\widetilde{\mathbf{i}})|=m\leq l \\ P(\widetilde{\mathbf{i}})\in\psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}}\text{ with partition }P(\widetilde{\mathbf{i}}) \\ r_1,\cdots,r_m\geq 2}} \sum_{Q(\widetilde{\mathbf{S}})\in\phi(2m)} \sum_{\substack{\widetilde{\mathbf{S}}\text{ with partition }Q(\widetilde{\mathbf{S}}) \\ 1\leq s_1,\cdots,s_{2m}\leq n}} \big|x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}\big|
$$
$$
\times \prod_{j=1}^{m}\mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}. \tag{A.23}
$$

We denote by $\mathcal{F}_{\widetilde{\mathbf{S}}}$ the graph constructed by the edges of $\widetilde{\mathbf{s}}$. Since the edges in $\widetilde{\mathbf{s}}$ are the same as those of the graph $\mathcal{F}_{\widetilde{\mathbf{i}}}$, we see that $\mathcal{F}_{\widetilde{\mathbf{S}}}$ is also a connected graph. In view of (A.63), putting term $|x_{i_1}y_{i_{l+1}}x_{i_{l+2}}y_{i_{2l+2}}|$ aside we need to analyze the summation

$$
\sum_{\substack{\widetilde{\mathbf{S}}\text{ with partition }Q(\widetilde{\mathbf{S}}) \\ 1\leq s_1,\cdots,s_{2m}\leq n}} \prod_{j=1}^{m}\mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}.
$$

If index s_{2k-1} satisfies that $s_{2k-1}\neq s$ for all $s\in\{s_1,\cdots,s_{2m}\}\setminus\{s_{2k-1}\}$, that is, index s_{2k-1} appears only in one $w_{s_{2j-1}s_{2j}}$, we call s_{2k-1} a single index (or single vertex). If there exists some single index s_{2k-1}, then we have

$$
\sum_{\substack{\widetilde{\mathbf{S}}\text{ with partition }Q(\widetilde{\mathbf{S}}) \\ 1\leq s_1,\cdots,s_{2m}\leq n}} \prod_{j=1}^{m}\mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}
$$
$$
\leq \sum_{\substack{\widetilde{\mathbf{s}}\setminus\{s_{2k-1}\}\text{ with partition }Q(\widetilde{\mathbf{s}}\setminus\{s_{2k-1}\}) \\ 1\leq s_1,\cdots,s_{2k-2},s_{2k+2},s_{2m}\leq n \\ s_{2k}=s_j\text{ for some }1\leq j\leq 2m}} \prod_{j=1}^{m}\mathbb{E}\big|w_{s_{2j-1}s_{2j}}\big|^{r_j}\sum_{s_{2k-1}=1}^{n}\mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}. \tag{A.24}
$$

Note that since graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$ is connected and index s_{2k-1} is single, there exists some j such that $s_j=s_{2k}$, which means that in the summation $\sum_{s_{2k-1}=1}^{n}\mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}$ index s_{2k} is fixed. It follows from the definition of α_n, $|w_{ij}|\leq 1$, and $r_k\geq 2$ that

$$
\sum_{s_{2k-1}=1}^{n}\mathbb{E}\big|w_{s_{2k-1}s_{2k}}\big|^{r_k}\leq\alpha_n^2.
$$

After taking the summation over index s_{2k-1}, we see that there is one less edge in $\mathcal{F}(\widetilde{\mathbf{s}})$. That is, by taking the summation above we will have one additional α_n^2 in the upper bound while removing one edge from graph $\mathcal{F}(\widetilde{\mathbf{s}})$. For the single index s_{2k}, we also have the same bound. If s_{2k_1-1} is not a single index, without loss of generality we assume that $s_{2k_1-1}=s_{2k-1}$. Then this vertex s_{2k-1} need to deal with carefully. By the assumption of

$|w_{ij}| \leq 1$, we have

$$\mathbb{E}|w_{2k-1,2k}|^{r_k}|w_{2k_1-1,2k_1}|^{r_{k_1}} \leq \mathbb{E}|w_{2k-1,2k}|^{r_k} + \mathbb{E}|w_{2k_1-1,2k_1}|^{r_{k_1}}.$$

Then it holds that

$$
\begin{aligned}
&\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j} \\
&\leq \sum_{\substack{\widetilde{\mathbf{S}} \backslash (s_{2k-1}, s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{S}} \backslash (s_{2k-1}, s_{2k_1-1})) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1, j \neq k}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j} \\
&\quad + \sum_{\substack{\widetilde{\mathbf{S}} \backslash (s_{2k-1}, s_{2k_1-1}) \text{ with partition } Q(\widetilde{\mathbf{S}} \backslash (s_{2k-1}, s_{2k_1-1})) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1, j \neq k_1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j} \qquad (\text{A.25})
\end{aligned}
$$

Note that since $\mathcal{F}_{\widetilde{\mathbf{S}}}$ is a connected graph, if we delete either edge (s_{2k-1}, s_{2k}) or edge (s_{2k_1-1}, s_{2k_1}) from graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$ the resulting graph is also connected. Then the two summations on the right hand side of (A.25) can be reduced to the case in (A.24) for the graph with edge (s_{2k-1}, s_{2k}) or (s_{2k_1-1}, s_{2k_1}) removed, since s_{2k-1} or s_{2k_1-1} is a single index in the subgraph. Similar to (A.24), after taking the summation over index s_{2k-1} or s_{2k_1-1} there are two less edges in graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$ and thus we now obtain $2\alpha_n^2$ in the upper bound.

For the general case when there are m_1 vertices belonging to the same group, without loss of generality we denote them by $w_{ab_1}, \cdots, w_{ab_{m_1}}$. If for any k graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$ is still connected after deleting edges $(a, b_1), \cdots, (a, b_{k-1}), (a, b_{k+1}), \cdots, (a, b_{m_1})$, then we repeat the process in (A.25) to obtain a new connected graph by deleting $k-1$ edges in $w_{ab_1}, \cdots, w_{ab_{m_1}}$ and thus obtain $k\alpha_n^2$ in the upper bound. Motivated by the key observations above, we carry out an iterative process in calculating the upper bound as follows.

(i) If there exists some single index in $\widetilde{\mathbf{s}}$, using (A.24) we can calculate the summation over such an index and then delete the edge associated with this vertex in $\mathcal{F}_{\widetilde{\mathbf{S}}}$. The corresponding vertices associated with this edge are also deleted. For simplicity, we also denote the new graph as $\mathcal{F}_{\widetilde{\mathbf{S}}}$. In this step, we obtain α_n^2 in the upper bound.

(ii) Repeat (i) until there is no single index in graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$.

(iii) If there exists some index associated with k edges such that graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$ is still connected after deleting any $k-1$ edges. Without loss of generality, let us consider the case of $k = 2$. Then we can apply (A.24) to obtain α_n^2 in the upper bound. Moreover, we delete k edges associated with this vertex in $\mathcal{F}_{\widetilde{\mathbf{S}}}$.

(iv) Repeat (iii) until there is no such index.

(v) If there still exists some single index, turn back to (i). Otherwise stop the iteration.

Completing the graph modification process mentioned above, we can obtain a final graph $\mathbf{Q}$ that enjoys the following properties:

i) Each edge does not contain any single index;

ii) Deleting any vertex makes the graph disconnected.

Let $\mathbf{S_Q}$ be the spanning tree of graph $\mathbf{Q}$, which is defined as the subgraph of $\mathbf{Q}$ with the minimum possible number of edges. Since $\mathbf{S_Q}$ is a subgraph of $\mathbf{Q}$, it also satisfies property ii) above. Assume that $\mathbf{S_Q}$ contains p edges. Then the number of vertices in $\mathbf{S_Q}$ is $p+1$. Denote by $q_1, \cdots, q_{p+1}$ the vertices of $\mathbf{S_Q}$ and $\deg(q_i)$ the degree of vertex q_i. Then by the degree sum formula, we have $\sum_{i=1}^{p+1} \deg(q_i) = 2p$. As a result, the spanning tree has at least two vertices with degree one and thus there exists a subgraph of $\mathbf{S_Q}$ without either of the vertices that is connected. This will result in a contradiction with property ii) above unless the number of vertices in graph $\mathbf{Q}$ is exactly one. Since l is a bounded constant, the numbers of partitions $P(\widetilde{\mathbf{i}})$ and $Q(\widetilde{\mathbf{s}})$ are also bounded. It follows that

$$(A.63) \leq Cd_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}, \tag{A.26}$$

where $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$, $d_{\mathbf{y}} = \|\mathbf{x}\|_\infty$, and C is some positive constant determined by l. Combining these arguments above and noticing that there are at most l distinct edges in graph $\mathcal{F}_{\widetilde{\mathbf{S}}}$, we can obtain

$$(A.26) \leq Cd_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \alpha_n^{2l-2} \sum_{\substack{1 \leq s_{2k_0-1}, s_{2k_0} \leq n, (s_{2k_0-1}, s_{2k_0}) = \mathbf{Q}}} \mathbb{E}|w_{s_{2k_0-1}s_{2k_0}}|^{r_{k_0}}$$

$$\leq Cd_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \alpha_n^{2l} n.$$

Therefore, we have established a simple upper bound of $O\{d_{\mathbf{x}} d_{\mathbf{y}} \alpha_n^l n^{1/2}\}$.

In fact, we can improve the aforementioned upper bound to $O(\alpha_n^{l-1})$. Note that the process mentioned above did not utilize the condition that both $\mathbf{x}$ and $\mathbf{y}$ are unit vectors, that is, $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. Since term $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|$ is involved in (A.63), we can analyze them together with random variables w_{ij}. There are four different cases to consider.

1). Two pairs of indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are equal. Without loss of generality, let us assume that $i_1 = i_{l+1} \neq i_{l+2} = i_{2l+2}$. Then it holds that $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = |x_{i_1} y_{i_1} x_{i_{l+2}} y_{i_{l+2}}| \leq 4^{-1}(x_{i_1}^2 + y_{i_1}^2)(x_{i_{l+2}}^2 + y_{i_{l+2}}^2)$. Let us consider the bound for

$$\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} x_{i_1}^2 x_{i_{l+2}}^2 \prod_{j=1}^{m} \mathbb{E}|w_{s_{2j-1}s_{2j}}|^{r_j}. \tag{A.27}$$

We assume without loss of generality that $i_1 = s_1$ and $i_{l+2} = s_2$ for this partition. Then the

13

summation in (A.27) becomes

$$\sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{S}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} x_{s_1}^2 x_{s_2}^2 \prod_{j=1}^{m} \mathbb{E} \big| w_{s_{2j-1} s_{2j}} \big|^{r_j}.$$

By repeating the iterative process (i)–(v) mentioned before, we can bound the summation for fixed s_2 and obtain an alternative upper bound

$$\sum_{s_1=1}^{n} x_{s_1}^2 \mathbb{E} \big| w_{s_1 s_2} \big|^{r_j} \leq \sum_{s_1=1}^{n} x_{s_1}^2 = 1$$

since $\mathbf{x}$ is a unit vector. Thus for this step of the iteration, we obtain 1 instead of α_n^2 in the upper bound. Since the graph is always connected during the iteration process, there exists another vertex b such that $w_{s_2 b}$ is involved in (A.27). For index s_2, we do not delete the edges containing s_2 in the graph during the iterative process (i)–(v). Then after the iteration stops, the final graph $\mathbf{Q}$ satisfies properties i) and ii) defined earlier except for vertex s_2. Since there are at least two vertices with degree one in $\mathbf{S_Q}$, we will also reach a contradiction unless the number of vertices in graph $\mathbf{Q}$ is exactly one. As a result, we can obtain the upper bound

$$(A.63) \leq C\alpha_n^{2l-4} \sum_{1 \leq s_2, b \leq n, \, (s_2, b) = \mathbf{Q}} \mathbb{E} x_{s_2}^2 \big| w_{s_2 b} \big|^r \leq C\alpha_n^{2l-2} \tag{A.28}$$

with C some positive constant. Therefore, the improved bound of $O(\alpha_n^{l-1})$ is shown for this case.

2). Indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are all distinct. Then by the triangle inequality, we have $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| \leq 4^{-1}(x_{i_1}^2 + x_{i_{l+2}}^2)(y_{i_{l+2}}^2 + y_{i_{2l+2}}^2)$. Thus this case reduces to case 1 above.

3). Indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are all equal. Then it holds that $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = x_{i_1}^2 y_{i_1}^2 \leq x_{i_1}^2$. We see that there are at most $[(2l+2-2)/2] = l$ distinct vertices in the chain $\prod_{s=1}^{2l-1} w_{i_s i_{s+1}}$ and for this case there are at most $l-1$ distinct edges in $\mathcal{F}_{\widetilde{\mathbf{i}}}$, where $[\cdot]$ denotes the integer part of a number. Compared to case 1, the maximum number of edges in the graph becomes smaller. Therefore, for this case we have

$$(A.63) \leq C\alpha_n^{2l-4} \sum_{1 \leq s_1, b \leq n, \, (s_1, b) = \mathbf{Q}} \mathbb{E} x_{s_1}^2 \big| w_{s_1 b} \big|^r \leq C\alpha_n^{2l-2}, \tag{A.29}$$

where C is some positive constant and we have assumed that $i_1 = s_1$ without loss of generality.

4). Three of the indices i_1, i_{l+1}, i_{l+2}, i_{2l+2} in $\mathcal{F}_{\widetilde{\mathbf{i}}}$ are equal. For such a case, without loss of generality let us write $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}| = |x_{i_1}^2 y_{i_1} y_{i_{2l+2}}|$. Then there are at most

14

$[(2l + 2 - 1)/2] = l$ distinct vertices in the chain $\prod_{s=1}^{2l-1} w_{i_s i_{s+1}}$ and thus for this case there are at most $l - 1$ distinct edges in $\mathcal{F}_{\tilde{\mathbf{i}}}$. Therefore, this case reduces to case 3 above.

In addition, we can also improve the upper bound to $O(\min\{d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$. The technical arguments for this refinement are similar to those for the improvement to order $O(\alpha_n^{l-1})$ above. As an example, we can bound the components of $\mathbf{y}$ by $d_{\mathbf{y}} = \|\mathbf{y}\|_\infty$, which leads to $|x_{i_{l+1}} y_{i_{l+2}} x_{i_{2l+2}}| \leq d_{\mathbf{y}}^2(x_{i_1}^2 + x_{i_{l+1}}^2)/2$. Then the analysis becomes similar to that for case 3 above. The only difference is that the length of graph $\mathcal{F}_{\tilde{\mathbf{i}}}$ is at most l instead of $l - 1$. Thus similar to (A.29), for this case we have

$$(A.63) \leq C d_{\mathbf{y}}^2 \alpha_n^{2l-2} \sum_{1 \leq s_2, b \leq n, (s_2, b) = \mathbf{Q}} \mathbb{E} x_{s_1}^2 |w_{s_1 b}|^r \leq C d_{\mathbf{y}}^2 \alpha_n^{2l}, \qquad (A.30)$$

where C is some positive constant and we have assumed that $i_1 = s_1$ or $x_{i_{l+1}} = s_1$ without loss of generality. The other one can then be used to remove a factor of α_n. Thus we can obtain the claimed upper bound $O(\min\{d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$. Therefore, combining the two aforementioned improved bounds yields the desired upper bound of $O_p(\min\{\alpha_n^{l-1}, d_{\mathbf{x}}\alpha_n^l, d_{\mathbf{y}}\alpha_n^l\})$.

We finally return to the general case of possibly $w_{ii} \neq 0$. Let us rewrite $\mathbf{W}$ as $\mathbf{W} = \mathbf{W}_0 + \mathbf{W}_1$ with $\mathbf{W}_1 = \text{diag}(w_{11}, \cdots, w_{nn})$. Then it holds that

$$\mathbf{x}^T \mathbf{W}^l \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y} = \mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} - \mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y}.$$

Recall the classical inequality

$$\mathbb{E}(X_1 + \cdots + X_m)^2 \leq m(\mathbb{E}X_1^2 + \cdots + \mathbb{E}X_m^2), \qquad (A.31)$$

where $X_1, \cdots, X_m$ are m random variables with finite second moments. Define a function

$$f(\mathbf{h}) = \prod_{i=1}^{l} \mathbf{W}_{h_i}, \qquad (A.32)$$

where the vector $\mathbf{h} = (h_1, \cdots, h_l)$ with $h_i = 0$ or 1. Then we have

$$\mathbb{E}\left[\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} - \mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y}\right]^2 = \mathbb{E}\left\{\sum_{\mathbf{h}} \mathbf{x}^T [f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\right\}^2$$

$$\leq 2^l \sum_{\mathbf{h}} \mathbb{E}\left\{\mathbf{x}^T [f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\right\}^2. \qquad (A.33)$$

This shows that we need only to consider terms of form $\mathbb{E}\{\mathbf{x}^T[f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$, each of which is a polynomial of $\mathbf{W}_0$ and $\mathbf{W}_1$.

As an example, let us analyze the term $\mathbb{E}(\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} - \mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y})^2$. Similar to

([A.18](#)), it can be shown that

$$
\mathbb{E}(\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y})^2
$$

$$
= \sum_{\substack{1 \leq i_1, \cdots, i_l, j_1, \cdots, j_l \leq n, \\ i_s \neq i_{s+1}, j_s \neq j_{s+1}, 1 \leq s \leq l}} \mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l} - \mathbb{E} x_{i_1} w_{i_1 i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l}\big)
$$

$$
\times \big(x_{j_1} w_{j_1 j_1} w_{j_1 j_2} \cdots w_{j_{l-1} j_l} y_{j_l} - \mathbb{E} x_{j_1} w_{j_1 j_1} w_{j_1 j_2} \cdots w_{j_{l-1} j_l} y_{j_l}\big) \Big]. \tag{A.34}
$$

Repeating the arguments from ([A.18](#))–([A.62](#)), we can obtain

$$
(A.34) \leq 2 \sum_{\mathcal{F}_{\tilde{\mathbf{i}}}} \mathbb{E}\big| x_{i_1} w_{i_1 i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l} x_{i_{l+1}} w_{i_{l+1} i_{l+1}} w_{i_{l+1} i_{l+2}} \cdots w_{i_{2l-1} i_{2l}} y_{i_{2l}} \big|
$$

$$
\leq 2 \sum_{\mathcal{F}_{\tilde{\mathbf{i}}}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l} x_{i_{l+1}} w_{i_{l+1} i_{l+2}} \cdots w_{i_{2l-1} i_{2l}} y_{i_{2l}} \big|.
$$

Comparing to ([A.62](#)), we can see that by replacing the diagonal entries with 1 in the expectations, the number of edges in this graph is no more than the original one in ([A.62](#)). Thus repeating all the steps before ([A.34](#)), we can deduce the bound

$$
\mathbb{E}(\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_1\mathbf{W}_0^{l-1}\mathbf{y})^2 = O(\min\{\alpha_n^{2(l-1)}, d_{\mathbf{x}}^2 \alpha_n^{2l}, d_{\mathbf{y}}^2 \alpha_n^{2l}\}).
$$

For the other expectations $\mathbb{E}\{\mathbf{x}^T[f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$, by the same reason that $\mathbf{W}_1$ is a diagonal matrix we can obtain a similar expression as ([A.34](#)) with the number of edges no larger than the original one for $\mathbb{E}(\mathbf{x}^T\mathbf{W}_0^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}_0^l\mathbf{y})^2$. Thus all the technical arguments above can be applied to $\mathbb{E}\{\mathbf{x}^T[f(\mathbf{h}) - \mathbb{E}f(\mathbf{h})]\mathbf{y}\}^2$ so we can have the same order for the upper bound as before. This shows that all the previous arguments can indeed be extended to the general case of possibly $w_{ii} \neq 0$, which concludes the proof of Lemma 4.

B.6   Proof of Lemma 5

The main idea of the proof is similar to that for the proof of Lemma 4 in Section [B.5](#). We first consider the case when all the diagonal entries of $\mathbf{W} = (w_{ij})_{1 \leq i,j \leq n}$ are zero, that is, $w_{ii} = 0$. Then we can derive a similar expression as ([A.18](#))

$$
\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = \sum_{\substack{1 \leq i_1, \cdots, i_{l+1} \leq n \\ i_s \neq i_{s+1}, 1 \leq s \leq l}} \mathbb{E}\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}\big). \tag{A.35}
$$

By the definition of graph $\mathcal{G}_{\mathbf{i}}$ in the proof of Lemma 4, we can obtain a similar expression as (A.62)

$$|(A.35)| \leq \sum_{\substack{\mathcal{G}_{\mathbf{i}} \text{ with at most } [l/2] \text{ distinct edges and } [l/2]+1 \text{ distinct vertices}}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots$$
$$\times w_{i_l i_{l+1}} y_{i_{l+1}}\big|. \tag{A.36}$$

Using similar arguments for bounding the order of the summation through the iterative process as for case 3 in the proof of Lemma 4 and noticing that $|x_{i_1} y_{i_{l+1}}| \leq 2^{-1}(x_{i_1}^2 + y_{i_{l+1}}^2)$, we can deduce the desired bound

$$\mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y} = O(\alpha_n^{l-1}), \tag{A.37}$$

where the diagonal entries of $\mathbf{W}$ have been assumed to be zero.

For the general case of $\mathbf{W}$ with possibly nonzero diagonal entries, we can apply the similar expansion as in the proof of Lemma 4 to get

$$\mathbb{E}\mathbf{x}^T (\mathbf{W}_0 + \mathbf{W}_1)^l \mathbf{y} = \sum_{\mathbf{h}} \mathbb{E}\mathbf{x}^T f(\mathbf{h})\mathbf{y}, \tag{A.38}$$

where $\mathbf{W} = \mathbf{W}_0 + \mathbf{W}_1$ with $\mathbf{W}_1 = \mathrm{diag}(w_{11}, \cdots, w_{nn})$, and vector $\mathbf{h}$ and function $f(\mathbf{h})$ are as defined in (A.32). Since by assumption $\mathbf{W}_1$ is a diagonal matrix with bounded entries, an application of similar arguments as in the proof of Lemma 4 gives

$$\mathbb{E}\mathbf{x}^T f(\mathbf{h})\mathbf{y} = O(\alpha_n^{l-1}).$$

To see this, with similar arguments as below (A.33) let us analyze the term $\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y}$ as an example. Similar to (A.35), it holds that

$$\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} = \sum_{\substack{1 \leq i_1, \cdots, i_l \leq n \\ i_s \neq i_{s+1}, 1 \leq s \leq l-1}} \mathbb{E}\left(x_{i_1} w_{i_1 i_1} w_{i_1 i_2} \cdots w_{i_{l-1} i_l} y_{i_l}\right). \tag{A.39}$$

By the assumption of $\max_{1 \leq i \leq n} |w_{ii}| \leq 1$, we can derive a similar bound as (A.36)

$$|(A.39)| \leq \sum_{\substack{\mathcal{G}_{\mathbf{i}} \text{ with at most } [(l-1)/2] \text{ distinct edges and } [(l-1)/2]+1 \text{ distinct vertices}}} \mathbb{E}\big|x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots$$
$$\times w_{i_{l-1} i_l} y_{i_l}\big|. \tag{A.40}$$

Since the number of edges is no more than that in (A.36), we can obtain the same bound

$$\mathbb{E}\mathbf{x}^T \mathbf{W}_1 \mathbf{W}_0^{l-1} \mathbf{y} = O(\alpha_n^{l-1}).$$

For the other terms in (A.38), by the same reason that $\mathbf{W}_1$ is a diagonal matrix with bounded entries we can derive similar expression as (A.40) with the number of edges no more than that in (A.36). Therefore, since l is a bounded constant we can show that $\mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O(\alpha_n^{l-1})$ for the general case of $\mathbf{W}$ with possibly nonzero diagonal entries. This completes the proof of Lemma 5.

B.7 Lemma 6 and its proof

Lemma 6. *The random matrix $\mathbf{W}$ given in (1) satisfies that for any positive constant L, there exist some positive constants C_L and σ such that*

$$\mathbb{P}\left\{\|\mathbf{W}\| \geq C_L(\log n)^{1/2}\alpha_n\right\} \leq n^{-L}, \tag{A.41}$$

where $\|\cdot\|$ denotes the matrix spectral norm and $\alpha_n = \|\mathbb{E}(\mathbf{W} - \mathbb{E}\mathbf{W})^2\|^{1/2}$.

Proof. The conclusion of Lemma 6 follows directly from Theorem 6.2 of Tropp (2012). □

C Further technical details on when asymptotic normality holds for Theorem 5

We now consider the joint distribution of the three random variables specified in expression (116) in the proof of Theorem 5 in Section A.6. To establish the joint asymptotic normality under some regularity conditions, it suffices to show that the random vector $(\text{tr}[(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - (\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}], \text{tr}((\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{v}_k\mathbf{v}_k^T), \text{tr}((\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}))$ tends to some multivariate normal distribution as matrix size n increases, where we consider the de-meaned version of this random vector for simplicity. Consequently, we need to show that for any constants c_1, c_2, and c_3 such that $c_1^2 + c_2^2 + c_3^2 = 1$, the linear combination

$$c_1\text{tr}[(\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k} - (\mathbf{W}^2 - \mathbb{E}\mathbf{W}^2)\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}] + c_2\text{tr}((\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{v}_k\mathbf{v}_k^T)$$
$$+ c_3\text{tr}((\mathbf{W} - \mathbb{E}\mathbf{W})\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}) \tag{A.42}$$

converges to a normal distribution asymptotically. Define $\mathbf{S} = \mathbf{v}_k\mathbf{v}_k^T$ and let $\mathbf{J}$, $\mathbf{L}$, and $\mathbf{Q}$ be the rescaled versions of $\mathbf{J}_{\mathbf{x},\mathbf{y},k,t_k}$, $\mathbf{L}_{\mathbf{x},\mathbf{y},k,t_k}$, and $\mathbf{Q}_{\mathbf{x},\mathbf{y},k,t_k}$, respectively, such that the asymptotic variance of each of the above three terms is equal to one. Then it remains to analyze the asymptotic behavior of the random variable

$$\sum_{1\leq k,i\leq n,\, k\leq i} w_{ki}\left\{c_1\left[\sum_{1\leq l<k\leq n} w_{il}\mathbf{L}_{kl} + \sum_{1\leq l<i\leq n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1 - \delta_{ki})(\mathbf{L}_{ki} + \mathbf{L}_{ik})\mathbb{E}w_{ii}\right]\right.$$
$$\left.+ (1 - \delta_{ki})(c_2\mathbf{S}_{ki} + c_3\mathbf{Q}_{ki})\right\} + c_1\sum_{1\leq k,i\leq n,\, k\leq i}(w_{ki}^2 - \sigma_{ki}^2)(\mathbf{L}_{kk} + \mathbf{L}_{ii}), \tag{A.43}$$

18

where $\mathbf{A}_{ij}$ indicates the (i,j)th entry of a matrix $\mathbf{A}$ and $\delta_{ki} = 1$ when $k = i$ and 0 otherwise.

Using similar arguments as in (A.7), we can show that (A.43) is in fact a sum of martingale differences with respect to the σ-algebra $\mathcal{F}_{k+2^{-1}i(i-1)-1}$. The conditional variance of the random variable given in (A.43) can be calculated as

$$
\begin{aligned}
&\sum_{1 \leq k,i \leq n,\, k \leq i} \sigma_{ki}^2 \Big\{ c_1 \Big[\sum_{1 \leq l < k \leq n} w_{il}\mathbf{L}_{kl} + \sum_{1 \leq l < i \leq n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] \\
&+ (1-\delta_{ki})(c_2\mathbf{S}_{ki} + c_3\mathbf{Q}_{ki}) \Big\}^2 + c_1^2 \sum_{1 \leq k,i \leq n,\, k \leq i} \kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})^2 \\
&+ 2c_1 \sum_{1 \leq k,i \leq n,\, k \leq i} \gamma_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) \Big\{ c_1 \Big[\sum_{1 \leq l < k \leq n} w_{il}\mathbf{L}_{kl} + \sum_{1 \leq l < i \leq n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} \\
&+ (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] + (1-\delta_{ki})(c_2\mathbf{S}_{ki}+c_3\mathbf{Q}_{ki}) \Big\}. \quad\quad\text{(A.44)}
\end{aligned}
$$

Moreover, the expectation of the random variable given in (A.44) can be shown to take the form

$$
\begin{aligned}
&c_1^2 \sum_{1 \leq k,i \leq n,\, k \leq i} \Big\{ \sigma_{ki}^2 \Big[\sum_{1 \leq l < k \leq n} \sigma_{il}^2\mathbf{L}_{kl}^2 + \sum_{1 \leq l < i \leq n} \sigma_{kl}^2\mathbf{L}_{il}^2 + \mathbf{J}_{ki}^2 + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})^2(\mathbb{E}w_{ii})^2 \Big] \\
&+ \kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})^2 \Big\} + c_2^2 \sum_{1 \leq k,i \leq n,\, k \leq i} \sigma_{ki}^2\mathbf{S}_{ki}^2 + c_3^2 \sum_{1 \leq k,i \leq n,\, k \leq i} \sigma_{ki}^2\mathbf{Q}_{ki}^2 \\
&+ 2 \sum_{1 \leq k,i \leq n,\, k \leq i} \Big[\sigma_{ki}^2(c_2\mathbf{S}_{ki}+c_3\mathbf{Q}_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big] \\
&+ 2c_1c_2 \sum_{1 \leq k,i \leq n,\, k \leq i} \gamma_{ki}\mathbf{S}_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) + 2c_1c_3 \sum_{1 \leq k,i \leq n,\, k \leq i} \gamma_{ki}\mathbf{Q}_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii}) \\
&+ 2c_2c_3 \sum_{1 \leq k,i \leq n,\, k \leq i} \sigma_{ki}^2\mathbf{S}_{ki}\mathbf{Q}_{ki} + 2c_1^2 \sum_{1 \leq k,i \leq n,\, k \leq i} \Big[\kappa_{ki}(\mathbf{L}_{kk}+\mathbf{L}_{ii})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big]. \quad\text{(A.45)}
\end{aligned}
$$

Let us consider the following three regularity conditions.

i) Assume that the six individual summation terms in (A.45) tend to some constants asymptotically. Then (A.45) tends to some constant C asymptotically. Without loss of generality, we assume that $C \neq 0$; otherwise (A.43) tends to zero in probability.

ii) Assume that $\mathrm{SD}(A.44) \ll (A.45)$, where SD stands for the standard deviation of a random variable.

iii) Assume that

$$
\begin{aligned}
&\sum_{1 \leq k,i \leq n,\, k < i} \kappa_{ki} \Big\{ \mathbb{E} \Big[\sum_{1 \leq l < k \leq n} w_{il}\mathbf{L}_{kl} + \sum_{1 \leq l < i \leq n} w_{kl}\mathbf{L}_{il} + \mathbf{J}_{ki} + (1-\delta_{ki})(\mathbf{L}_{ki}+\mathbf{L}_{ik})\mathbb{E}w_{ii} \Big]^4 \\
&+ (1-\delta_{ki})(\mathbf{S}_{ki}^4 + \mathbf{Q}_{ki}^4) \Big\} + \sum_{1 \leq k,i \leq n,\, k < i} \mathbb{E}(w_{ki}^2 - \sigma_{ki}^2)^4(\mathbf{L}_{kk}+\mathbf{L}_{ii})^4 \ll 1. \quad\text{(A.46)}
\end{aligned}
$$

We can see that conditions i) and ii) entail condition a) in the proof of Lemma 2 in Section

B.2 below (A.9), while condition iii) entails condition b). Therefore, (A.43) converges to a normal distribution asymptotically.

D   Relaxing the spike strength condition and proof sketch for results in Section 4.2

The main goal of this section is to show that all the results continue to hold when Condition 2i) is replaced with Condition 2ii), which is a weaker assumption on the spike strength. Thus from now on, we will assume Condition 2ii) instead of Condition 2i). Moreover, we provide the proof sketch for results in Section 4.2.

D.1   Replacing Condition 2i) with Condition 2ii)

Checking the proofs of our theorems, we can see that it is sufficient to show that the asymptotic expansion of $\mathbf{x}^T\mathbf{G}(z)\mathbf{y}$ remains to hold under Condition 2ii). In other words, we need to prove (76) and (108) under Condition 2ii). To accommodate the smaller magnitude of d_K in Condition 2ii), the key idea is to carefully examine the asymptotic expansions (76) and (108) as $L \to \infty$. To this end, we choose $L = \log n$ and define $c' = c/(1 + 2^{-1}c_0)$. Since $\alpha_n \le n^{1/2}$, we have the following improved version of inequality (66)

$$\frac{\alpha_n^{L+1}(C\log n)^{(L+1)/2}}{\min\{|a_K|,|b_K|\}^{L-2}} \le \frac{\alpha_n^3(C\log n)^{(L+1)/2}}{(c'\log n)^{L-2}} \le \frac{C^{(\log n+1)/2}n^{3/2}}{(\log n)^{(\log n-5)/2}c'^{\log n-2}} \to 0 \qquad (A.47)$$

for any positive constant C.

We first show that (76) holds with the choice of $L = \log n$. In view of (75), it is sufficient to establish the following two equations

$$\sum_{l=L+1}^{\infty} z^{-(2l+2)}\mathbf{x}^T\mathbf{W}^l\mathbf{y} = O_p\left(\frac{1}{|z|^4}\right) \qquad (A.48)$$

and

$$\sum_{l=2}^{L} z^{-(2l+2)}\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y} = O_p\left(\frac{\alpha_n}{|z|^3}\right) \qquad (A.49)$$

for $z \in \Omega_k$. In fact, (A.48) is a direct consequence of Lemma 6 and (A.47). In light of the definitions of a_k and b_k below (10), we can conclude that for any $z \in \Omega_k$, $|z| > 4c_1\alpha_n \log n$. Thus we see that

$$\left\{\frac{\alpha_n^{2l}(4c_1\log n)^{2l}}{|z|^{2l}}\right\} \text{ is a decreasing sequence when } l \text{ is increasing for } z \in \Omega_k. \qquad (A.50)$$

Then it follows from Lemma 7 and (A.50) that

$$\sum_{l=2}^{9} z^{-(2l+2)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} = O_p(\frac{\alpha_n}{|z|^3}), \tag{A.51}$$

$$\left| \mathbb{E} \left[\sum_{l=\sqrt{L}}^{L} z^{-(2l+2)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} \right]^2 \right| \le L \sum_{l=\sqrt{L}}^{L} |z|^{-(2l+2)} \mathbb{E} \left[\mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} \right]^2$$

$$\le CL \sum_{l=\sqrt{L}}^{L} \frac{(4c_1 l)^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}} \le CL \sum_{l=\sqrt{L}}^{L} \frac{(4c_1 \log n)^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}}$$

$$\le C(\log n)^2 \frac{\alpha_n^4 (4c_1 \log n)^{2\sqrt{\log n}}}{|z|^8 (c' \log n)^{2\sqrt{\log n}-4}} \le \frac{C(4c_1)^6 \alpha_n^4 (\log n)^6}{|z|^8 (c'/(4c_1))^{2\sqrt{\log n}-4}} \ll \frac{\alpha_n^4}{|z|^8}, \tag{A.52}$$

and

$$\left| \mathbb{E} \left[\sum_{l=10}^{\sqrt{L}} z^{-(2l+2)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} \right]^2 \right| \le \sqrt{L} \sum_{l=10}^{\sqrt{L}} |z|^{-(2l+2)} \mathbb{E} \left[\mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} \right]^2$$

$$\le C\sqrt{L} \sum_{l=10}^{\sqrt{L}} \frac{(4c_1 l)^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}} \le C\sqrt{L} \sum_{l=10}^{\sqrt{L}} \frac{(4c_1 \sqrt{\log n})^{2l} \alpha_n^{2l-2}}{|z|^{2l+2}}$$

$$\le C \log n \frac{\alpha_n^4 ((4c_1)^2 \log n)^{10}}{|z|^8 (c' \log n)^{20-4}} \ll \frac{\alpha_n^4}{|z|^8}. \tag{A.53}$$

Therefore, combining (A.51)–(A.53) yields (A.49).

To establish (108) with $L = \log n$, we need only to prove (A.48) and

$$\sum_{l=3}^{L} z^{-(2l+2)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} = O_p(\frac{\alpha_n^2}{|z|^4}), \tag{A.54}$$

where the former has been shown before. By Lemma 7, we can deduce

$$\sum_{l=3}^{9} z^{-(2l+2)} \mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} = O_p(\frac{\alpha_n^2}{|z|^4}). \tag{A.55}$$

Thus (A.54) holds by combining (A.52), (A.53), and (A.55). This concludes the proofs of the desired results.

D.2 Improvement of Lemmas 4 and 5 under Condition 2ii)

Lemma 7. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, there exists some positive constant C independent of l such that*

$$\mathbb{E} \left[\mathbf{x}^T (\mathbf{W}^l - \mathbb{E}\mathbf{W}^l) \mathbf{y} \right]^2 \le C(4c_1 l)^{2l} (\min\{\alpha_n^{l-1}, d_{\mathbf{x}} \alpha_n^l, d_{\mathbf{y}} \alpha_n^l\})^2 \tag{A.56}$$

with $l \geq 1$ some positive integer and $d_{\mathbf{x}} = \|\mathbf{x}\|_\infty$.

Lemma 8. *For any n-dimensional unit vectors $\mathbf{x}$ and $\mathbf{y}$, there exists some positive constant C independent of l such that*

$$\left| \mathbb{E}\mathbf{x}^T \mathbf{W}^l \mathbf{y} \right| \leq C(2c_1 l)^l \alpha_n^l \tag{A.57}$$

with $l \geq 2$ some bounded positive integer.

D.3 Proof of Lemma 7

The proof of Lemma 7 is a modification of that for Lemma 4. Thus we highlight only the differences of the technical arguments here. We work directly on the general case allowing for $\mathbb{E}w_{ii} \neq 0$. In view of (A.18), we have

$$\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$$
$$= \sum_{1 \leq i_1, \cdots, i_{l+1}, j_1, \cdots, j_{l+1} \leq n} \mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big)$$
$$\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} \big) \Big]. \tag{A.58}$$

Let $\mathbf{i} = (i_1, \cdots, i_{l+1})$ and $\mathbf{j} = (j_1, \cdots, j_{l+1})$ be two vectors taking values in $\{1, \cdots, n\}^{l+1}$. For any given vector $\mathbf{i}$, we define a graph $\mathcal{G}_{\mathbf{i}}$ whose vertices represent the components of $\mathbf{i}$. Vertices i_s and i_{s+1} of $\mathcal{G}_{\mathbf{i}}$ are connected by undirected edges for $1 \leq s \leq l$. Similarly we can also define graph $\mathcal{G}_{\mathbf{j}}$ corresponding to $\mathbf{j}$. It can be seen that $\mathcal{G}_{\mathbf{i}}$ is a connected graph, which means that there exists some path from i_s to $i_{s'}$ for any $1 \leq s \neq s' \leq n$. One should notice that here we allow for $i_s = i_{s+1}$ or $j_s = j_{s+1}$. Such relaxation will affect only the number of pairs $(\mathbf{i}, \mathbf{j})$, but will not affect the main arguments of the proof which are similar to the graph arguments for proving Lemma 4. Thus for each product

$$\mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big)$$
$$\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} \big) \Big], \tag{A.59}$$

there exists a corresponding graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. If $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is not a connected graph, then the corresponding expectation

$$\mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big)$$
$$\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} \big) \Big] = 0.$$

This shows that in order to calculate the order of $\mathbb{E}(\mathbf{x}^T\mathbf{W}^l\mathbf{y} - \mathbb{E}\mathbf{x}^T\mathbf{W}^l\mathbf{y})^2$, it suffices to consider the scenario of connected graphs $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$.

To analyze the term in (A.59), let us calculate how many distinct vertices are contained in the connected graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. We say that $(i_s, i_{s+1}) \in \mathcal{G}_{\mathbf{i}}$ is an *efficient edge* if $i_s \neq i_{s+1}$.

Since there are at most $2l$ efficient edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ and $\mathbb{E}w_{ss'} = 0$ for $s \neq s'$, in order to get a nonzero value of (A.59) each efficient edge in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ has at least one copy. Thus for each nonzero (A.59), we have at most l distinct efficient edges in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Since graph $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ is connected, there are at most $l+1$ distinct vertices in $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$. Denote by $\mathcal{S}$ the set of all such pairs $(\mathbf{i}, \mathbf{j})$. Combining the above arguments, we can conclude that

$$(A.18) = \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}} \mathbb{E}\Big[\big(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big)$$
$$\times \big(x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} - \mathbb{E}x_{j_1} w_{j_1 j_2} w_{j_2 j_3} \cdots w_{j_l j_{l+1}} y_{j_{l+1}} \big) \Big]. \qquad (A.60)$$

For notational simplicity, we denote $j_1, \cdots, j_{l+1}$ as $i_{l+2}, \cdots, i_{2l+2}$ and define $\widetilde{\mathbf{i}} = (i_1, \cdots, i_{l+1}, j_1, \cdots, j_{l+1}) = (i_1, \cdots, i_{2l+2})$. We also denote $\mathcal{G}_{\mathbf{i}} \cup \mathcal{G}_{\mathbf{j}}$ as $\mathcal{F}_{\widetilde{\mathbf{i}}}$ which has at most $l+1$ distinct vertices and l distinct efficient edges, with each edge having at least two copies. Then it holds that

$$|(A.60)| = \Big| \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big[(x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} - \mathbb{E}x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}})$$
$$\times (x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} - \mathbb{E}x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}})\big]\Big|$$
$$\leq \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|$$
$$+ \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} \big| \mathbb{E}\big| x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|. \qquad (A.61)$$

Observe that each expectation in (A.61) involves the product of some independent random variables, and $x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}}$ and $x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}}$ may share some dependency through factors $w_{ab}^{m_1}$ and $w_{ab}^{m_2}$, respectively, for some w_{ab} and nonnegative integers m_1 and m_2. Thus with the aid of the inequality

$$\mathbb{E}|w_{ab}|^{m_1} \mathbb{E}|w_{ab}|^{m_2} \leq \mathbb{E}|w_{ab}|^{m_1+m_2},$$

we can further bound (A.61) as

$$(A.61) \leq 2 \sum_{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}}\in\mathcal{S}} \mathbb{E}\big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots$$
$$\times w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|. \qquad (A.62)$$

To facilitate our technical presentation, let us introduce some additional notation. Denote by $\psi(2l+2)$ the set of partitions of the edges $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2}), i_s \neq i_{s+1}, s =$

$1, \cdots, 2l+1\}$ and $\psi_{\geq 2}(2l+2)$ the subset of $\psi(2l+2)$ whose blocks have size at least two. Let $P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)$ be the partition of $\{(i_1, i_2), (i_2, i_3), \cdots, (i_{2l+1}, i_{2l+2}), i_s \neq i_{s+1}, s = 1, \cdots, 2l+1\}$ that is associated with the equivalence relation $(i_{s_1}, i_{s_1+1}) \sim (i_{s_2}, i_{s_2+1})$ which is defined as if and only if $(i_{s_1}, i_{s_1+1}) = (i_{s_2}, i_{s_2+1})$ or $(i_{s_1}, i_{s_1+1}) = (i_{s_2+1}, i_{s_2})$. Denote by $|P(\widetilde{\mathbf{i}})| = m$ the number of groups in the partition $P(\widetilde{\mathbf{i}})$ such that the edges are equivalent within each group. We further denote the distinct edges in the partition $P(\widetilde{\mathbf{i}})$ as $(s_1, s_2), (s_3, s_4), \cdots, (s_{2m-1}, s_{2m})$ and the corresponding counts in each group as $r_1, \cdots, r_m$, and define $\widetilde{\mathbf{s}} = (s_1, s_2, \cdots, s_{2m})$. For the vertices, let $\phi(2m)$ be the set of partitions of $\{1, 2, \cdots, 2m\}$ and $Q(\widetilde{\mathbf{s}}) \in \phi(2m)$ the partition that is associated with the equivalence relation $a \sim b$ which is defined as if and only if $s_a = s_b$. Note that $s_{2j-1} \neq s_{2j}$ since in the partition, we consider only the off-diagonal entries (efficient edges) and for diagonal entries, we use the simple inequality $|w_{ii}| \leq 1$. Then it holds that

$$
\sum_{\substack{\mathcal{F}_{\widetilde{\mathbf{i}}}, \widetilde{\mathbf{i}} \in \mathcal{S}}} \mathbb{E} \big| x_{i_1} w_{i_1 i_2} w_{i_2 i_3} \cdots w_{i_l i_{l+1}} y_{i_{l+1}} x_{i_{l+2}} w_{i_{l+2} i_{l+3}} w_{i_{l+3} i_{l+4}} \cdots w_{i_{2l+1} i_{2l+2}} y_{i_{2l+2}} \big|
$$

$$
\leq \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|
$$

$$
\times \prod_{j=1}^{m} \mathbb{E} \big| w_{s_{2j-1} s_{2j}} \big|^{r_j}
$$

$$
\leq \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \Big(\frac{c_1^2 \alpha_n^2}{n} \Big)^m \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|.
$$

$$(\text{A.63})$$

It suffices to bound the number of graphs in the above summation. In fact, since the graph is connected there are at most $m+1$ different vertices in the graph. Moreover, there are $2l$ edges in the original graph with at most l efficient edges and the partitions corresponding to the edges have at most $(4l)^{2l}$ cases. Thus combining these arguments together we can deduce

$$
\sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \Big(\frac{c_1^2 \alpha_n^2}{n} \Big)^m \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\substack{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{s}}) \\ 1 \leq s_1, \cdots, s_{2m} \leq n}} |x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|
$$

$$
\leq d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \Big(\frac{c_1^2 \alpha_n^2}{n} \Big)^l \sum_{\substack{1 \leq |P(\widetilde{\mathbf{i}})| = m \leq l \\ P(\widetilde{\mathbf{i}}) \in \psi_{\geq 2}(2l+2)}} \sum_{\substack{\widetilde{\mathbf{i}} \text{ with partition } P(\widetilde{\mathbf{i}}) \\ r_1, \cdots, r_m \geq 2}} \sum_{Q(\widetilde{\mathbf{s}}) \in \phi(2m)} \sum_{\widetilde{\mathbf{S}} \text{ with partition } Q(\widetilde{\mathbf{s}})} 1
$$

$$
\leq d_{\mathbf{x}}^2 d_{\mathbf{y}}^2 \Big(\frac{c_1^2 \alpha_n^2}{n} \Big)^l (4l)^{2l} n^{l+1}
$$

$$
\leq (4c_1 l)^{2l} n \alpha_n^{2l} d_{\mathbf{x}}^2 d_{\mathbf{y}}^2.
$$

$$(\text{A.64})$$

Therefore, we can establish the simple upper bound that

$$\mathbb{E}\left[\mathbf{x}^T(\mathbf{W}^l - \mathbb{E}\mathbf{W}^l)\mathbf{y}\right]^2 \le C(4c_1 l)^{2l} n\alpha_n^{2l} d_{\mathbf{x}}^2 d_{\mathbf{y}}^2. \tag{A.65}$$

For the other upper bounds $C(4c_1 l)^{2l} d_{\mathbf{x}}^2 \alpha_n^{2l}$, $C(4c_1 l)^{2l} d_{\mathbf{y}}^2 \alpha_n^{2l}$, and $C(4c_1 l)^{2l} \alpha_n^{2l-2}$, the arguments are similar to those for the proof of Lemma 4. The crucial steps are considering the impact of $|x_{i_1} y_{i_{l+1}} x_{i_{l+2}} y_{i_{2l+2}}|$ from (A.27) to (A.30). For our case, we can directly prove the desired bounds $C(4c_1 l)^{2l} d_{\mathbf{x}}^2 \alpha_n^{2l}$, $C(4c_1 l)^{2l} d_{\mathbf{y}}^2 \alpha_n^{2l}$, and $C(4c_1 l)^{2l} \alpha_n^{2l-2}$ by combining the left hand side of (A.64) with the arguments from (A.27) to (A.30). This completes the proof of Lemma 7.

D.4   Proof of Lemma 8

Similar to the proof of Lemma 5, the proof of Lemma 8 is a direct modification of that of Lemma 7. Thus we omit it for brevity.

D.5   Proof sketch for results in Section 4.2

By calculating the variance of $\widehat{p}$, we have

$$\widehat{p} = p + O_p\Big(\frac{\sqrt{p(1-p)}}{n}\Big) = p\left[1 + O_p\Big(\frac{\sqrt{1-p}}{n\sqrt{p}}\Big)\right]. \tag{A.66}$$

Then the mean and variance of $\mathbf{v}_1^T \mathbf{W}^2 \mathbf{v}_1$ in (26) can be estimated as

$$\mathbf{v}_1^T \widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1 = n\widehat{p}(1-\widehat{p}) \quad \text{and} \quad \text{var}(\widehat{\mathbf{v}_1^T \mathbf{W}^2 \mathbf{v}_1}) = \widehat{p}(1-\widehat{p})\left[2(n-1) + \widehat{p}^3 + (1-\widehat{p})^3\right], \tag{A.67}$$

receptively. By Theorem 1, (A.66), and (A.67), direct calculations show that if $n^{-1} \ll p < 1$, then it holds that

$$\lambda_1 - t_1 = O_p\Big(\frac{1}{\sqrt{np}} + \sqrt{p}\Big),$$

$$\frac{\mathbf{v}_1^T \widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1}{\sqrt{\text{var}(\widehat{\mathbf{v}_1^T \mathbf{W}^2 \mathbf{v}_1})}} = \frac{\mathbf{v}_1^T \mathbb{E}\mathbf{W}^2 \mathbf{v}_1}{\sqrt{\text{var}(\mathbf{v}_1^T \mathbf{W}^2 \mathbf{v}_1)}} + o_p(1).$$

Thus if the conditions of Corollary 1 hold, by (24) we can obtain

$$\frac{2\lambda_1^2\left(\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1\right) + \mathbf{v}_1^T \widehat{\mathbb{E}\mathbf{W}^2}\mathbf{v}_1}{\left[\text{var}(\widehat{\mathbf{v}_1^T \mathbf{W}^2 \mathbf{v}_1})\right]^{1/2}} \xrightarrow{\mathscr{D}} N(0,1). \tag{A.68}$$

Since $\mathbf{v}_1 = n^{-1/2}\mathbf{1}$ under the null hypothesis, the above results together with (A.68) ensure that under the null hypothesis, statistic T_n is asymptotically standard normal.

Next we consider the case of alternative hypothesis. It can be derived that the leading

eigenvalue and eigenvector take the following forms

$$d_1 = \frac{1}{2} \left[np + n_1(q - p) + \left(n^2 p^2 + 2n_1(2n_1 - n)p(q - p) + n_1^2(q - p)^2 \right)^{1/2} \right]$$

and $\mathbf{v}_1 = (\mathbf{v}_{1,1}^T, \mathbf{v}_{1,2}^T)^T$, where $\mathbf{v}_{1,1}$ is an n_1-dimensional vector with all entries being

$$\frac{(n - n_1)p}{\sqrt{(n - n_1)(d_1 - n_1 q)^2 + n_1(n - n_1)^2 p^2}}$$

and $\mathbf{v}_{1,2}$ is an $(n - n_1)$-dimensional vector with all entries being

$$\frac{d_1 - n_1 q}{\sqrt{(n - n_1)(d_1 - n_1 q)^2 + n_1(n - n_1)^2 p^2}}.$$

With some direct calculations, we can show that under the alternative hypothesis,

$$n^{-1/2}\mathbf{1}^T\mathbf{v}_1 = \frac{(n - n_1)(d_1 - n_1(q - p))}{\sqrt{n((n - n_1)(d_1 - n_1 q)^2 + n_1(n - n_1)^2 p^2)}}. \tag{A.69}$$

Since $n_1 = o(n)$, $n^{-1} \ll p < q$, and $p \sim q$, by the Taylor expansion we can deduce

$$d_1 = np + n_1^2(q - p)\frac{4p + 5(q - p)}{4np} + O(\frac{n_1^3(q - p)^2}{n^2 p})$$

and

$$\sqrt{n((n - n_1)(d_1 - n_1 q)^2 + n_1(n - n_1)^2 p^2)}$$
$$= \sqrt{n(n - n_1)}(d_1 - n_1 q) + \frac{n_1\sqrt{n}(n - n_1)^2 p^2}{2\sqrt{(n - n_1)}(d_1 - n_1 q)} + \frac{n_1^2 p}{4} + O(\frac{n_1^3 p}{n})$$
$$= \sqrt{n(n - n_1)} \left[np - n_1(q - p) - \frac{n_1 p}{2} - \frac{n_1^2 p}{4n} + n_1^2(q - p)\frac{4p + 5(q - p)}{4np} + O(\frac{n_1^3 p}{n^2}) \right].$$

Substituting the above two equations into (A.69) yields

$$n^{-1/2}\mathbf{1}^T\mathbf{v}_1 - 1 = \sqrt{\frac{n - n_1}{n}} \left[1 + \frac{n_1}{2n} - \frac{n_1^2}{4n^2} - \frac{n_1^2(q - p)^2}{n^2 p^2} - n_1^2(q - p)\frac{4p + 5(q - p)}{4n^2 p^2} \right.$$
$$\left. + O(\frac{n_1^3}{n^3} + \frac{n_1^3(q - p)^2}{n^3 p^2}) \right] - 1$$
$$= -\frac{n_1^2(q - p)^2}{n^2 p^2} - n_1^2(q - p)\frac{4p + 5(q - p)}{4n^2 p^2} + O(\frac{n_1^3}{n^3} + \frac{n_1^3(q - p)^2}{n^3 p^2}). \tag{A.70}$$

If the conditions of Corollary 1 hold, by (24) we have

$$\frac{2\lambda_1^2 \left(\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1 \right) + \mathbf{v}_1^T\mathbb{E}\mathbf{W}^2\mathbf{v}_1}{\left[\text{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1) \right]^{1/2}} \xrightarrow{\mathscr{D}} N(0, 1). \tag{A.71}$$

This entails that

$$\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1 = O_p(\frac{\mathbf{v}_1^T\mathbb{E}\mathbf{W}^2\mathbf{v}_1 + \left[\mathrm{var}(\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1)\right]^{1/2}}{t_1^2}) = O_p(\frac{1}{np}), \qquad (\text{A.72})$$

where the last step is obtained by directly calculating the mean and variance of $\mathbf{v}_1^T\mathbf{W}^2\mathbf{v}_1$ and noting that $t_1 \sim np$. Since $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form an orthonormal basis, it follows from (A.72) that

$$\sum_{j=2}^n (\mathbf{v}_j^T\widehat{\mathbf{v}}_1)^2 = 1 - (\mathbf{v}_1^T\widehat{\mathbf{v}}_1)^2 = O_p(\frac{1}{np}). \qquad (\text{A.73})$$

Similarly, by (A.69) and the assumptions of $n_1 = o(n)$ and $q \sim p$, we can deduce

$$\sum_{j=2}^n (n^{-1/2}\mathbf{1}^T\mathbf{v}_j)^2 = O(\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2}). \qquad (\text{A.74})$$

Then it follows from (A.69), (A.72), and (A.74) that

$$n^{-1/2}\mathbf{1}^T\widehat{\mathbf{v}}_1 - 1 = n^{-1/2}\mathbf{1}^T\mathbf{v}_1\mathbf{v}_1^T\widehat{\mathbf{v}}_1 - 1 + n^{-1/2}\mathbf{1}^T\sum_{j=2}^n \mathbf{v}_j\mathbf{v}_j^T\widehat{\mathbf{v}}_1$$

$$= -\left[\frac{n_1^2(q-p)^2}{n^2p^2} + n_1^2(q-p)\frac{4p+5(q-p)}{4n^2p^2}\right] + O_p\left[\frac{n_1^3}{n^3} + \frac{n_1^3(q-p)^2}{n^3p^2} + \frac{1}{np}\right]. \qquad (\text{A.75})$$

Under the alternative hypothesis, it can be shown that the estimators in (27) are of orders $n\widehat{p}(1-\widehat{p}) = O_p(np)$ and $\widehat{p}(1-\widehat{p})\left[2(n-1) + \widehat{p}^3 + (1-\widehat{p})^3\right] = O_p(np)$, respectively, and in addition, $t_1 \sim np$. Therefore, if the conditions of Corollary 1 holds and $\frac{n_1^2(q-p)^2}{np} + \frac{n_1^2(q-p)}{n} \gg 1$, with probability tending to one we have

$$T_n \to -\infty,$$

which means that the power can tend to one asymptotically. This concludes the proof sketch for the results in Section 4.2.