Journal of Statistical Physics (2020) 181:1746–1780
https://doi.org/10.1007/s10955-020-02644-7

Eigenvalues Outside the Bulk of Inhomogeneous Erdős–Rényi Random Graphs

Arijit Chakrabarty[1] · Sukrit Chakraborty[1] · Rajat Subhra Hazra[1]

Received: 27 February 2020 / Accepted: 19 September 2020 / Published online: 26 September 2020

Abstract

In this article, an inhomogeneous Erdős–Rényi random graph on $\{1, \ldots, N\}$ is considered, where an edge is placed between vertices i and j with probability $\varepsilon_N f(i/N, j/N)$, for $i \le j$, the choice being made independently for each pair. The integral operator I_f associated with the bounded function f is assumed to be symmetric, non-negative definite, and of finite rank k. We study the edge of the spectrum of the adjacency matrix of such an inhomogeneous Erdős–Rényi random graph under the assumption that $N\varepsilon_N \to \infty$ sufficiently fast. Although the bulk of the spectrum of the adjacency matrix, scaled by $\sqrt{N\varepsilon_N}$, is compactly supported, the kth largest eigenvalue goes to infinity. It turns out that the largest eigenvalue after appropriate scaling and centering converges to a Gaussian law, if the largest eigenvalue of I_f has multiplicity 1. If I_f has k distinct non-zero eigenvalues, then the joint distribution of the k largest eigenvalues converge jointly to a multivariate Gaussian law. The first order behaviour of the eigenvectors is derived as a byproduct of the above results. The results complement the homogeneous case derived by [18].

Keywords Adjacency matrices · Inhomogeneous Erdős–Rényi random graph · Largest eigenvalue · Scaling limit · Stochastic block model

Mathematics Subject Classification 60B20 · 05C80 · 46L54

Communicated by Abhishek Dhar.

✉ Rajat Subhra Hazra
rajatmaths@gmail.com

Arijit Chakrabarty
arijit.isi@gmail.com

Sukrit Chakraborty
sukrit049@gmail.com

1 Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

1 Introduction

Given a graph on N vertices, say, $\{1, \ldots, N\}$, let A_N denote the adjacency matrix of the graph, whose (i, j)th entry is 1 if there is an edge between vertices i and j and 0 otherwise. Important statistics of the graph are the eigenvalues and eigenvectors of A_N which encode crucial information about the graph. The present article considers the generalization of the most studied random graph, namely the Erdős–Rényi random graph (ERRG). It is a graph on N vertices where an edge is present independently with probability ε_N. The adjacency matrix of the ERRG is a symmetric matrix with diagonal entries zero, and the entries above the diagonal are independent and identically distributed Bernoulli random variables with parameter ε_N. We consider an inhomogeneous extension of the ERRG where the presence of an edge between vertices i and j is given by a Bernoulli random variable with parameter $p_{i,j}$ and these $\{p_{i,j} : 1 \leq i < j \leq N\}$ need not be same. When $p_{i,j}$ are same for all vertices i and j it shall be referred as (homogeneous) ERRG.

The mathematical foundations of inhomogeneous ERRG where the connection probabilities $p_{i,j}$ come from a discretization of a symmetric, non-negative function f on $[0, 1]^2$ was initiated in [9]. The said article considered edge probabilities given by

$$p_{i,j} = \frac{1}{N} f\left(\frac{i}{N}, \frac{j}{N}\right).$$

In that case the average degree is bounded and the phase transition picture on the largest cluster size was studied in the same article (see also [8,32] for results on inhomogeneous ERRG). The present article considers a similar set-up where the average degree is unbounded and studies the properties of eigenvalues of the adjacency matrix. The connection probabilities are given by

$$p_{i,j} = \varepsilon_N f\left(\frac{i}{N}, \frac{j}{N}\right)$$

with the assumption that

$$N\varepsilon_N \to \infty. \tag{1.1}$$

Let $\lambda_1(A_N) \geq \cdots \geq \lambda_N(A_N)$ be the eigenvalues of A_N. It was shown in [13] (see also [35] for a graphon approach) that the empirical distribution of the centered adjacency matrix converges, after scaling with $\sqrt{N\varepsilon_N}$, to a compactly supported measure μ_f. When $f \equiv 1$, the limiting law μ_f turns out to be the semicircle law. Note that $f \equiv 1$ corresponds to the (homogeneous) ERRG (see [16,31] also for the homogeneous case). Quantitative estimates on the largest eigenvalue of the homogeneous case (when $N\varepsilon_N \gg (\log N)^4$) were studied in [20,34] and it follows from their work that the smallest and second largest eigenvalue converge to the edge of the support of semicircular law. The results were improved recently in [7] and the condition on sparsity can be extended to the case $N\varepsilon_N \gg \log N$ (which is also the connectivity threshold). It was shown that inhomogeneous ERRG also has a similar behaviour. The largest eigenvalue of inhomogeneous ERRG when $N\varepsilon_N \ll \log N$ was treated in [6]. Under the assumption that $N^{\xi} \ll N\varepsilon_N$ for some $\xi \in (2/3, 1]$, it was proved in [17, Theorem 2.7] that the second largest eigenvalue of the (homogeneous) ERRG after appropriate centering and scaling converges in distribution to the Tracy–Widom law. The results were recently improved in [27]. The properties of the largest eigenvector in the homogeneous case was studied in [1,17,22,27,31].

The scaling limit of the maximum eigenvalue of inhomogenous ERRG also turns out to be interesting. The fluctuations of the maximum eigenvalue in the homogeneous case were studied in [18]. It was proved that

$$(\varepsilon_N(1-\varepsilon_N))^{-1/2}\left(\lambda_1(A_N)-\mathrm{E}[\lambda_1(A_N)]\right)\Rightarrow N(0,2).$$

The above result was shown under the assumption that

$$(\log N)^{\xi} \ll N\varepsilon_N \tag{1.2}$$

for some $\xi > 8$, which is a stronger assumption than (1.1).

It is well known that in the classical case of a (standard) Wigner matrix, the largest eigenvalue converges to the Tracy–Widom law. We note that there is a different scaling between the edge and bulk of the spectrum in ERRG. As pointed out before, the bulk is of the order $(N\varepsilon_N)^{1/2}$ and the order of the largest eigenvalue is $N\varepsilon_N$. Letting

$$W_N = A_N - \mathrm{E}(A_N), \tag{1.3}$$

where $\mathrm{E}(A_N)$ is the entrywise expectation of A_N, it is easy to see that

$$A_N = \varepsilon_N \mathbf{1}\mathbf{1}' + W_N,$$

where $\mathbf{1}$ is the $N\times 1$ vector with each entry 1. Since the empirical spectral distribution of $(N\varepsilon_N)^{-1/2}W_N$ converges to semi-circle law, the largest eigenvalue of the same converges to 2 almost surely. As $\mathrm{E}[A_N]$ is a rank-one matrix, it turns out that the largest eigenvalue of A_N scales like $N\varepsilon_N$, which is different from the bulk scaling.

The above behaviour can be treated as a special case of the perturbation of a Wigner matrix. When W_N is a symmetric random matrix with independent and identically distributed entries with mean zero and finite variance σ^2 and the deformation is of the form

$$M_N = \frac{W_N}{\sqrt{N}} + P_N,$$

the largest eigenvalue is well-studied. Motivated by the study of adjacency matrix of homogeneous ERRG, [20] studied the above deformation with $P_N = mN^{-1/2}\mathbf{1}\mathbf{1}'$, $m\neq 0$. They showed that

$$N^{-1/2}\left(\lambda_1(M_N) - Nm - \frac{\sigma^2}{m}\right)\Rightarrow N(0,2\sigma^2).$$

Since the bulk of $N^{-1/2}M_N$ lies within $[-2\sigma,\ 2\sigma]$, the largest eigenvalue is detached from the bulk. In general, when the largest eigenvalue of the perturbation has the same order as that of the maximum eigenvalue of W_N, the problem is more challenging. One of the seminal results in this direction was obtained in [3]. They exhibited a phase transition in the behaviour of the largest eigenvalue for complex Wishart matrix, which is now referred to as the BBP (Baik–Ben Arous–Péché) phase transition. It is roughly as follows. Suppose P_N is a deterministic matrix of rank k with non-trivial eigenvalues $\theta_1\ge\theta_2\cdots\ge\theta_k>0$. If $\theta_i\le\sigma$, then $\lambda_i(M_N)\to 2\sigma$ almost surely, and if $\theta_i>\sigma$ then

$$\lambda_i \to \theta_i + \frac{\sigma^2}{\theta_i}, \quad \text{almost surely.}$$

See [2,19] for further extensions. It is clear that when $\theta_i>\sigma$ the corresponding eigenvalue lies outside the bulk of the spectrum. The phase transition is also present at the level of fluctuations around 2σ or $\theta_i+\sigma^2/\theta_i$. It is known that under some moment conditions on the entries of W_N (see [11,24,25]), when $\theta_i\le\sigma$, the fluctuations are governed by the Tracy–Widom law, and when $\theta_i>\sigma$, the limiting distribution is given by the eigenvalues of a random matrix of order k. This limiting random matrix depends on the eigenvectors of P_N and also on the entries of W_N. The non-universal nature was pointed out in [11]. For example, when W_N is a Gaussian orthogonal ensemble and $P_N = \theta_1 \mathbf{1}\mathbf{1}'$ then the limit is Gaussian and if the entries of W_N are not from a Gaussian distribution, then the limit is a convolution of Gaussian and the distribution from which the entries of W_N are sampled. One can find further details in [3–5,11,12,24,25] and the survey by [29]. The case when the rank k depends on N was considered in [10,23,26]. Various applications of these results on outliers can be found in the literature, for example, [14,15,30].

The adjacency matrix of the inhomogeneous ERRG does not fall directly into purview of the above results, since W_N, as in (1.3), is a symmetric matrix, with independent entries above the diagonal, but the entries have a variance profile, which also depends on the size of the graph. The inhomogeneity does not allow the use of local laws suitable for semicircle law in an obvious way. The present article aims at extending the results obtained in [18] for the case that f is a constant to the case that f is a non-negative, symmetric, bounded, Riemann integrable function on $[0, 1]^2$ which induces an integral operator of finite rank k, under the assumption that (1.2) holds. The case $k \geq 2$ turns out to be substantially difficult than the case $k = 1$ for the following reason. If $k = 1$, that is,

$$\mathrm{E}(A_N) = u_N u_N',$$

for some $N \times 1$ deterministic column vector u_N, then with high probability it holds that

$$u_N' \left(\lambda I - W_N\right)^{-1} u_N = 1,$$

where λ is the largest eigenvalue of A_N. The above equation facilitates the asymptotic study of λ. However, when $k \geq 2$, the above equation takes a complicated form. The observation which provides a way out of this is that λ is also an eigenvalue of a $k \times k$ matrix with high probability; the same is recorded in Lemma 5.2 of Sect. 5. Besides, working with the eigenvalues of a $k \times k$ matrix needs more linear algebraic work when $k \geq 2$. For example, the proof of Lemma 5.8, which is one of the major steps in the proof of a main result, becomes a tautology when $k = 1$.

The following results are obtained in the current paper. If the largest eigenvalue of the integral operator has multiplicity 1, then the largest eigenvalue of the adjacency matrix has a Gaussian fluctuation. More generally, it is shown that the eigenvalues which correspond to isolated eigenvalues, which will be defined later, of the induced integral operator jointly converge to a multivariate Gaussian law. Under the assumption that the function f is Lipschitz continuous, the leading order term in the expansion of the expected value of the isolated eigenvalues is obtained. Furthermore, under an additional assumption, the inner product of the eigenvector with the discretized eigenfunction of the integral operator corresponding to the other eigenvalues is shown to have a Gaussian fluctuation. Some important examples of such f include the rank-one case, and the stochastic block models. It remains an open question to see if the $(k+1)$th eigenvalue follows a Tracy–Widom type scaling.

The mathematical set-up and the main results of the paper are stated in Sect. 2. Theorem 2.3 shows that of the k largest eigenvalues, the isolated ones, centred by their mean and appropriately scaled, converge to a multivariate normal distribution. Theorem 2.4 studies the first and second order of the expectation of the top k isolated eigenvalues. Theorems 2.5 and 2.6 study the behaviour of the eigenvectors corresponding to the top k isolated eigenvalues. Section 3 contains the special case when f is rank one and the example of stochastic block models. A few preparatory estimates are noted in Sect. 4, which are used later in the proofs of the main results, given in Sect. 5. The estimates in Sect. 4 are proved in Appendix.

2 The Set-Up And The Results

Let $f : [0, 1] \times [0, 1] \to [0, \infty)$ be a function which is symmetric, bounded, and Riemann integrable, that is,

$$f(x, y) = f(y, x), \quad 0 \le x, y \le 1, \tag{2.1}$$

and the set of discontinuities of f in $[0, 1] \times [0, 1]$ has Lebesgue measure zero.

The integral operator I_f with kernel f is defined from $L^2[0, 1]$ to itself by

$$\big(I_f(g)\big)(x) = \int_0^1 f(x, y)g(y)\, dy, \quad 0 \le x \le 1.$$

Besides the above, we assume that I_f is a non-negative definite operator and the range of I_f has a finite dimension.

Under the above assumptions I_f turns out to be a compact self-adjoint operator, and from the spectral theory one obtains $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_k > 0$ as the non-zero eigenvalues of I_f (where k is the dimension of the range of I_f), and eigenfunctions r_i corresponding to θ_i. Therefore, $\{r_1, \ldots, r_k\}$ is an orthonormal set in $L^2[0, 1]$, and by assumption, each r_i is Riemann integrable (see Lemma 6.1 in Appendix). Also, for any $g \in L^2[0, 1]$ one has

$$I_f(g) = \sum_{i=1}^{k} \theta_i \langle r_i, g \rangle_{L^2[0,1]} r_i.$$

Note that this gives

$$\int_0^1 \left(\sum_{i=1}^{k} \theta_i r_i(x) r_i(y) g(y) \right) dy = \int_0^1 f(x, y)g(y)\, dy \ \text{ for almost all} x \in [0, 1].$$

Since g is an arbitrary function in $L^2[0, 1]$ this immediately gives

$$f(x, y) = \sum_{i=1}^{k} \theta_i r_i(x) r_i(y), \quad \text{for almost all } (x, y) \in [0, 1] \times [0, 1].$$

Since the functions on both sides of the above equation are Riemann integrable, the corresponding Riemann sums are approximately equal, and hence there is no loss of generality in assuming that the above equality holds for every x and y.

That is, we **now assume that**

$$f(x, y) = \sum_{i=1}^{k} \theta_i r_i(x) r_i(y) \ge 0, \quad \text{for all } (x, y) \in [0, 1] \times [0, 1], \tag{2.2}$$

where $\theta_1 \ge \cdots \ge \theta_k > 0$ and $\{r_1, \ldots, r_k\}$ is an orthonormal set in $L^2[0, 1]$. The assumptions on $r_1, \ldots, r_k$ are listed below for easy reference.

Assumption F1. The functions $r_1, \ldots, r_k$ from $[0, 1]$ to $\mathbb{R}$ are bounded and Riemann integrable.

Assumption F2. For each $i = 1, \ldots, k$, r_i is Lipschitz, that is,

$$|r_i(x) - r_i(y)| \le K_i |x - y|,$$

for some fixed $K_i < \infty$. This is clearly stronger than Assumption F1, and will be needed in a few results. A consequence of this assumption is that there exists K such that

$$\left| f(x, y) - f(x', y') \right| \le K \left(|x - x'| + |y - y'| \right), \quad 0 \le x, x', y, y' \le 1. \tag{2.3}$$

Let $(\varepsilon_N : N \geq 1)$ be a real sequence satisfying

$$0 < \varepsilon_N \leq \left[\sup_{0 \leq x, y \leq 1} f(x, y) \right]^{-1}, \quad N \geq 1.$$

The following will be the bare minimum assumption for all the results.
Assumption E1. For some $\xi > 8$, fixed once and for all,

$$\lim_{N \to \infty} \frac{1}{N \varepsilon_N} (\log N)^{\xi} = 0,$$

that is, (1.2) holds. Furthermore,

$$\lim_{N \to \infty} \varepsilon_N = \varepsilon_\infty, \tag{2.4}$$

for some $\varepsilon_\infty \geq 0$. It's worth emphasizing that we do not assume that ε_N necessarily goes to zero, although that may be the case.

For one result, we shall have to make a stronger assumption on (ε_N) which is the following.
Assumption E2. As $N \to \infty$,

$$N^{-2/3} \ll \varepsilon_N \ll 1. \tag{2.5}$$

For $N \geq 1$, let $\mathbb{G}_N$ be an inhomogeneous Erdős–Rényi graph where an edge is placed between vertices i and j with probability $\varepsilon_N f(i/N, j/N)$, for $i \leq j$, the choice being made independently for each pair in $\{(i, j) : 1 \leq i \leq j \leq N\}$. Note that we allow self-loops. Let A_N be the adjacency matrix of $\mathbb{G}_N$. In other words, A_N is an $N \times N$ symmetric matrix, where $\{A_N(i, j) : 1 \leq i \leq j \leq N\}$ is a collection of independent random variable, and

$$A_N(i, j) \sim \text{Bernoulli}\left(\varepsilon_N f\left(\frac{i}{N}, \frac{j}{N} \right) \right), \quad 1 \leq i \leq j \leq N.$$

A few more notations are needed for stating the main results. For a moment, set $\theta_0 = \infty$ and $\theta_{k+1} = -\infty$, and define the set of indices i for which θ_i is isolated as follows:

$$\mathcal{I} = \{1 \leq i \leq k : \theta_{i-1} > \theta_i > \theta_{i+1}\}.$$

For an $N \times N$ real symmetric matrix M, let $\lambda_1(M) \geq \cdots \geq \lambda_N(M)$ denote its eigenvalues, as mentioned in Sect. 1. Finally, after the following definition, the main results will be stated.

Definition A sequence of events E_N occurs *with high probability*, abbreviated as w.h.p., if

$$P(E_N^c) = O\left(e^{-(\log N)^{\eta}} \right),$$

for some $\eta > 1$. For random variables Y_N, Z_N,

$$Y_N = O_{hp}(Z_N),$$

means there exists a deterministic finite constant C such that

$$|Y_N| \leq C |Z_N| \text{ w.h.p.},$$

and

$$Y_N = o_{hp}(Z_N),$$

means that for all $\delta > 0$,

$$|Y_N| \leq \delta |Z_N| \text{ w.h.p.}$$

We shall say

$$Y_N = O_p(Z_N),$$

to mean that

$$\lim_{x\to\infty} \sup_{N\geq 1} P(|Y_N| > x|Z_N|) = 0,$$

and

$$Y_N = o_p(Z_N),$$

to mean that for all $\delta > 0$,

$$\lim_{N\to\infty} P(|Y_N| > \delta|Z_N|) = 0.$$

The reader may note that if $Z_N \neq 0$ a.s., then "$Y_N = O_p(Z_N)$" and "$Y_N = o_p(Z_N)$" are equivalent to "$(Z_N^{-1} Y_N : N \geq 1)$ is stochastically tight" and "$Z_N^{-1} Y_N \xrightarrow{P} 0$", respectively. Besides, "$Y_N = O_{hp}(Z_N)$" is a much stronger statement than "$Y_N = O_p(Z_N)$", and so is "$Y_N = o_{hp}(Z_N)$" than "$Y_N = o_p(Z_N)$".

In the rest of the paper, the subscript 'N' is dropped from notations like A_N, W_N, ε_N etc. and the ones that will be introduced. The first result is about the first order behaviour of $\lambda_i(A)$.

Theorem 2.1 *Under Assumptions* E1 *and* F1, *for every* $1 \leq i \leq k$,

$$\lambda_i(A) = N\varepsilon\theta_i \left(1 + o_{hp}(1)\right).$$

An immediate consequence of the above is that for all $1 \leq i \leq k$, $\lambda_i(A)$ is non-zero w.h.p. and hence dividing by the same is allowed, as done in the next result. Define

$$e_i = \begin{bmatrix} N^{-1/2} r_i(1/N) \\ N^{-1/2} r_i(2/N) \\ \vdots \\ N^{-1/2} r_i(1) \end{bmatrix}, \quad 1 \leq i \leq k. \tag{2.6}$$

The second main result studies the asymptotic behaviour of $\lambda_i(A)$, for $i \in \mathcal{I}$, after appropriate centering and scaling.

Theorem 2.2 *Under Assumptions* E1 *and* F1, *for every* $i \in \mathcal{I}$, *as* $N \to \infty$,

$$\lambda_i(A) = \mathrm{E}\left(\lambda_i(A)\right) + \frac{N\theta_i\varepsilon}{\lambda_i(A)} e_i' W e_i + o_p(\sqrt{\varepsilon}),$$

where W *is as defined in* (1.3).

The next result is the corollary of the previous two.

Theorem 2.3 *Under Assumptions* E1 *and* F1, *if* $\mathcal{I}$ *is a non-empty set, then as* $N \to \infty$,

$$\left(\varepsilon^{-1/2}\left(\lambda_i(A) - \mathrm{E}[\lambda_i(A)]\right) : i \in \mathcal{I}\right) \Rightarrow \left(G_i : i \in \mathcal{I}\right), \tag{2.7}$$

where the right hand side is a multivariate normal random vector in $\mathbb{R}^{|\mathcal{I}|}$, *with mean zero and*

$$\mathrm{Cov}(G_i, G_j) = 2\int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x, y) \left[1 - \varepsilon_\infty f(x, y)\right] dx\, dy, \tag{2.8}$$

for all $i, j \in \mathcal{I}$.

For $i, j \in \mathcal{I}$, that W is a symmetric matrix whose upper triangular entries are independent and zero mean implies that as $N \to \infty$,

$$\begin{aligned}
&Cov\left(e_i' W e_i, e_j' W e_j\right) \\
&\sim 4 \sum_{1 \le k \le l \le N} \mathrm{Cov}\left(e_i(k) W(k,l) e_i(l), e_j(k) W(k,l) e_j(l)\right) \\
&= 4N^{-2} \sum_{1 \le k \le l \le N} r_i\left(\frac{k}{N}\right) r_i\left(\frac{l}{N}\right) r_j\left(\frac{k}{N}\right) r_j\left(\frac{l}{N}\right) \varepsilon f\left(\frac{k}{N}, \frac{l}{N}\right)\left[1 - \varepsilon f\left(\frac{k}{N}, \frac{l}{N}\right)\right] \\
&\sim 4\varepsilon \int_0^1 dx \int_x^1 dy\, r_i(x) r_i(y) r_j(x) r_j(y) f(x,y)\left[1 - \varepsilon_\infty f(x,y)\right] \\
&= 2\varepsilon \int_0^1 \int_0^1 r_i(x) r_i(y) r_j(x) r_j(y) f(x,y)\left[1 - \varepsilon_\infty f(x,y)\right] dx\, dy. \qquad (2.9)
\end{aligned}$$

With the help of the above, it may be checked that the Lindeberg–Lévy central limit theorem implies that as $N \to \infty$,

$$\left(\varepsilon^{-1/2} e_i' W e_i : i \in \mathcal{I}\right) \Rightarrow (G_i : i \in \mathcal{I}), \qquad (2.10)$$

where the right hand side is a zero mean Gaussian vector with covariance given by (2.8). Therefore, Theorem 2.3 would follow from Theorems 2.1 and 2.2 .

Remark 2.1 If $f > 0$ a.e. on $[0,1] \times [0,1]$, then the Krein–Rutman theorem (see Lemma 6.2) implies that $1 \in \mathcal{I}$, and that $r_1 > 0$ a.e. Thus, in this case, if $\varepsilon_\infty = 0$, then

$$\mathrm{Var}(G_1) = 2 \int_0^1 \int_0^1 r_1(x)^2 r_1(y)^2 f(x,y)\, dx\, dy > 0.$$

Remark 2.2 For a fixed $\theta > 1$, define

$$f(x,y) = \theta \mathbf{1}\left(x \vee y < \frac{1}{2}\right) + \mathbf{1}\left(x \wedge y > \frac{1}{2}\right), \quad 0 \le x, y \le 1.$$

In this case, the integral operator associated with f has exactly two non-zero eigenvalues, which are $\theta/2$ and $1/2$, with corresponding normalized eigenfunctions $r_1(x) = \sqrt{2}\mathbf{1}(x < 1/2)$ and $r_2(x) = \sqrt{2}\mathbf{1}(x \ge 1/2)$, respectively. Let (ε_N) satisfy Assumption E1 and suppose that $\varepsilon_\infty = 0$. Theorem 2.3 implies that as $N \to \infty$,

$$\left(\varepsilon^{-1/2}\left(\lambda_1(A) - \mathrm{E}[\lambda_1(A)]\right), \varepsilon^{-1/2}\left(\lambda_2(A) - \mathrm{E}[\lambda_2(A)]\right)\right) \Rightarrow (G_1, G_2),$$

where the right hand side has a bivariate normal distribution with mean zero. Furthermore, since $r_1 r_2$ is identically zero, it follows that G_1 and G_2 are uncorrelated and hence independent.

Remark 2.3 That the claim of Theorem 2.3 may not hold if $i \notin \mathcal{I}$ is evident from the following example. As in Remark 2.2, suppose that Assumption E1 holds and that $\varepsilon_\infty = 0$. Let

$$f(x,y) = \mathbf{1}\left(x \vee y < \frac{1}{2}\right) + \mathbf{1}\left(x \wedge y > \frac{1}{2}\right), \quad 0 \le x, y \le 1.$$

In this case, the integral operator associated with f has exactly one non-zero eigenvalue, which is $1/2$, and that has multiplicity 2, with eigenfunctions r_1 and r_2 as in Remark 2.2. In other words, f doesn't have any simple eigenvalue.

Theorem 2.3 itself implies that there exists $\beta_N \in \mathbb{R}$ such that

$$\varepsilon^{-1/2}(\lambda_1(A) - \beta) \Rightarrow G_1 \vee G_2,$$

where G_1 and G_2 are independent from normal with mean 0. Furthermore,

$$\text{Var}(G_1) = 2\int_0^1\int_0^1 r_1(x)^2 r_1(y)^2 f(x,y)\,dx\,dy = 8\int_0^1\int_0^1 \mathbf{1}\left(x \vee y \le \frac{1}{2}\right) dx\,dy = 2.$$

That is, G_1 and G_2 are i.i.d. from $N(0, 2)$. Hence, there doesn't exist a centering and a scaling by which $\lambda_1(A)$ converges weakly to a non-degenerate normal distribution.

The next main result of the paper studies asymptotics of $\text{E}(\lambda_i(A))$ for $i \in \mathcal{I}$.

Theorem 2.4 *Under Assumptions* E1 *and* F2*, it holds for all* $i \in \mathcal{I}$*,*

$$\text{E}\left[\lambda_i(A)\right] = \lambda_i(B) + O\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right),$$

where B *is a* $k \times k$ *symmetric deterministic matrix, depending on* N*, defined by*

$$B(j,l) = \sqrt{\theta_j\theta_l}N\varepsilon e_j' e_l + \theta_i^{-2}\sqrt{\theta_j\theta_l}(N\varepsilon)^{-1}\text{E}\left(e_j' W^2 e_l\right), \quad 1 \le j, l \le k,$$

and e_j *and* W *are as defined in* (2.6) *and* (1.3)*, respectively.*

The next result studies the asymptotic behaviour of the normalized eigenvector corresponding to $\lambda_i(A)$, again for isolated vertices i. It is shown that the same is asymptotically aligned with e_i, and hence it is asymptotically orthogonal to e_j. Upper bounds on rates of convergence are obtained.

Theorem 2.5 *As in Theorem* 2.4*, let Assumptions* E1 *and* F2 *hold. Then, for a fixed* $i \in \mathcal{I}$*,*

$$\lim_{N\to\infty} P\left(\lambda_i(A) \quad \textit{is an eigenvalue of multiplicity } 1\right) = 1. \tag{2.11}$$

If v *is the eigenvector, with* L^2*-norm* 1*, of* A *corresponding to* $\lambda_i(A)$*, then*

$$e_i' v = 1 + O_p\left((N\varepsilon)^{-1}\right), \tag{2.12}$$

that is, $N\varepsilon(1 - e_i' v)$ *is stochastically tight. When* $k \ge 2$*, it holds that*

$$e_j' v = O_p\left((N\varepsilon)^{-1}\right), \quad j \in \{1, \ldots, k\} \setminus \{i\}. \tag{2.13}$$

The last main result of this paper studies finer fluctuations of (2.13) under an additional condition.

Theorem 2.6 *Let* $k \ge 2$*,* $i \in \mathcal{I}$ *and Assumptions* E2 *and* F2 *hold. If* v *is as in Theorem* 2.5*, then, for all* $j \in \{1, \ldots, k\} \setminus \{i\}$*,*

$$e_j' v = \frac{1}{\theta_i - \theta_j}\left[\theta_i \frac{1}{\lambda_i(A)} e_i' W e_j + (N\varepsilon)^{-2}\frac{1}{\theta_i}\text{E}\left(e_i' W^2 e_j\right)\right] + o_p\left(\frac{1}{N\sqrt{\varepsilon}}\right).$$

Remark 2.4 An immediate consequence of Theorem 2.6 is that under Assumption E2., there exists a deterministic sequence $(z_N : N \ge 1)$ given by

$$z = \frac{1}{(N\varepsilon)^2\theta_i(\theta_i - \theta_j)}\text{E}\left(e_i' W^2 e_j\right),$$

such that as $N \to \infty$,

$$Z_{ij} = N\sqrt{\varepsilon}\left(e_j' v - z\right) \tag{2.14}$$

converges weakly to a normal distribution with mean zero, for all $i \in \mathcal{I}$ and $j \in \{1, \ldots, k\} \setminus \{i\}$. Furthermore, the convergence holds jointly for all i and j satisfying the above. This, along with (2.7), implies that the collection

$$\left(Z_{ij} : i \in \mathcal{I}, j \in \{1, \ldots, k\} \setminus \{i\}\right) \cup \left(\varepsilon^{-1/2} \left(\lambda_i(A) - \mathrm{E}\left[\lambda_i(A)\right]\right) : i \in \mathcal{I}\right)$$

converges weakly, as $N \to \infty$, to

$$\left(G_{ij} : i \in \mathcal{I}, j \in \{1, \ldots, k\} \setminus \{i\}\right) \cup (G_i : i \in \mathcal{I})$$

which is a zero mean Gaussian vector in $\mathbb{R}^{k|\mathcal{I}|}$. The covariance matrix of (G_i) is as in (2.8), and $\mathrm{Cov}(G_{ij}, G_{i'j'})$ and $\mathrm{Cov}(G_{ij}, G_{i'})$ are not hard to calculate by proceeding as in (2.9).

3 Examples and Special Cases

3.1 The Rank One Case

Let us consider the special case of $k = 1$, that is,

$$f(x, y) = \theta r(x) r(y),$$

for some $\theta > 0$, and a bounded Riemann integrable $r : [0, 1] \to [0, \infty)$ satisfying

$$\int_0^1 r(x)^2\, dx = 1.$$

In this case, Theorem 2.3 implies that

$$\varepsilon^{-1/2} \left(\lambda_1(A) - \mathrm{E}\left(\lambda_1(A)\right)\right) \Rightarrow G_1,$$

as $N \to \infty$, where

$$G_1 \sim N\left(0, \sigma^2\right),$$

with

$$\sigma^2 = 2\theta \left(\int_0^1 r(x)^3\, dx\right)^2 - 2\theta^2 \varepsilon_\infty \left(\int_0^1 r(x)^4\, dx\right)^2.$$

If r is Lipschitz and $\varepsilon_\infty = 0$, then the claim of Theorem 2.4 boils down to

$$\mathrm{E}\left[\lambda_1(A)\right] = \theta N \varepsilon e_1' e_1 + (N\varepsilon\theta)^{-1} \mathrm{E}\left(e_1' W^2 e_1\right) + O\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right), \tag{3.1}$$

where

$$e_1 = N^{-1/2} \left[r(1/N)\, r(2/N)\, \ldots r(1)\right]'.$$

Lipschitz continuity of r implies that

$$e_1' e_1 = 1 + O\left(N^{-1}\right),$$

and hence (3.1) becomes

$$\mathrm{E}\left[\lambda_1(A)\right] = \theta N \varepsilon + (N\varepsilon\theta)^{-1} \mathrm{E}\left(e_1' W^2 e_1\right) + O\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right). \tag{3.2}$$

Clearly,

$$
\begin{aligned}
&\mathrm{E}\left(e_1' W^2 e_1\right)\\
&= \frac{1}{N}\sum_{i=1}^{N} r\left(\frac{i}{N}\right)^2 \mathrm{E}\left[W^2(i,i)\right]\\
&= \frac{1}{N}\sum_{i=1}^{N} r\left(\frac{i}{N}\right)^2 \sum_{1\le j\le N,\, j\ne i} \varepsilon f\left(\frac{i}{N},\frac{j}{N}\right)\left(1-\varepsilon f\left(\frac{i}{N},\frac{j}{N}\right)\right)\\
&= \theta\varepsilon N^{-1} \sum_{1\le i\ne j\le N} r\left(\frac{i}{N}\right)^3 r\left(\frac{j}{N}\right) + O\left(N^{-1}\varepsilon^2\right)\\
&= N\theta\varepsilon \int_0^1 r(x)^3\,dx \int_0^1 r(y)\,dy + O(\varepsilon).
\end{aligned}
$$

In conjunction with (3.2) this yields

$$
\mathrm{E}\left[\lambda_1(A)\right] = \theta N\varepsilon + \int_0^1 r(x)^3\,dx \int_0^1 r(y)\,dy + O\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right).
$$

3.2 Stochastic Block Model

Another important example is the stochastic block model, defined as follows. Suppose that

$$
f(x,y) = \sum_{i,j=1}^{k} p(i,j)\mathbf{1}_{B_i}(x)\mathbf{1}_{B_j}(y), \quad 0\le x,y\le 1,
$$

where p is a $k\times k$ symmetric positive definite matrix, and $B_1,\ldots,B_k$ are disjoint Borel subsets of $[0,1]$ whose boundaries are sets of measure zero, that is, their indicators are Riemann integrable. We show below how to compute the eigenvalues and eigenfunctions of I_f, the integral operator associated with f.

Let β_i denote the Lebesgue measure of B_i, which we assume without loss of generality to be strictly positive. Rewrite

$$
f(x,y) = \sum_{i,j=1}^{k} \tilde{p}(i,j)s_i(x)s_j(y),
$$

where

$$
\tilde{p}(i,j) = p(i,j)\sqrt{\beta_i\beta_j}, \quad 1\le i,j\le k,
$$

and

$$
s_i = \beta_i^{-1/2}\mathbf{1}_{B_i}, \quad 1\le i\le k.
$$

Thus, $\{s_1,\ldots,s_k\}$ is an orthonormal set in $L^2[0,1]$. Let

$$
\tilde{p} = U'DU,
$$

be a spectral decomposition of $\tilde{p}$, where U is a $k\times k$ orthogonal matrix, and

$$
D = \mathrm{Diag}(\theta_1,\ldots,\theta_k),
$$

for some $\theta_1 \geq \cdots \geq \theta_k > 0$.

Define functions $r_1, \ldots, r_k$ by

$$\begin{bmatrix} r_1(x) \\ \vdots \\ r_k(x) \end{bmatrix} = U \begin{bmatrix} s_1(x) \\ \vdots \\ s_k(x) \end{bmatrix}, \; x \in [0, 1].$$

It is easy to see that $r_1, \ldots, r_k$ are orthonormal in $L^2[0, 1]$, and for $0 \leq x, y \leq 1$,

$$\begin{aligned} f(x, y) &= [s_1(x) \ \ldots \ s_k(x)] \, \tilde{p} \, [s_1(x) \ \ldots \ s_k(x)]' \\ &= [r_1(x) \ \ldots \ r_k(x)] \, U \tilde{p} U' \, [r_1(x) \ \ldots \ r_k(x)]' \\ &= \sum_{i=1}^{k} \theta_i r_i(x) r_i(y). \end{aligned}$$

Thus, $\theta_1, \ldots, \theta_k$ are the eigenvalues of I_f, and $r_1, \ldots, r_k$ are the corresponding eigenfunctions.

4 Estimates

In this section, we'll record a few estimates that will subsequently be used in the proof. Since their proofs are routine, they are being postponed to Appendix. Let W be as defined in (1.3). Throughout this section, Assumptions E1 and F1 will be in force.

Lemma 4.1 *There exist constants C_1, $C_2 > 0$ such that*

$$P\left(\|W\| \geq 2\sqrt{MN\varepsilon} + C_1(N\varepsilon)^{1/4}(\log N)^{\xi/4}\right) \leq e^{-C_2(\log N)^{\xi/4}}, \tag{4.1}$$

where $M = \sup_{0 \leq x, y \leq 1} f(x, y)$. Consequently,

$$\|W\| = O_{hp}\left(\sqrt{N\varepsilon}\right).$$

The notations e_1 and e_2 introduced in the next lemma and used in the subsequent lemmas should not be confused with e_j defined in (2.6). Continuing to suppress 'N' in the subscript, let

$$L = [\log N],$$

where $[x]$ is the largest integer less than or equal to x.

Lemma 4.2 *There exists $0 < C_1 < \infty$ such that if e_1 and e_2 are $N \times 1$ vectors with each entry in $[-1/\sqrt{N}, 1/\sqrt{N}]$, then*

$$\left|\mathrm{E}\left(e_1' W^n e_2\right)\right| \leq (C_1 N\varepsilon)^{n/2}, \quad 2 \leq n \leq L.$$

Lemma 4.3 *There exists $\eta_1 > 1$ such that for e_1, e_2 as in Lemma 4.2, it holds that*

$$\begin{aligned} &\max_{2 \leq n \leq L} P\left(\left|e_1' W^n e_2 - \mathrm{E}\left(e_1' W^n e_2\right)\right| > N^{(n-1)/2} \varepsilon^{n/2} (\log N)^{n\xi/4}\right) \\ &= O\left(e^{-(\log N)^{\eta_1}}\right), \end{aligned} \tag{4.2}$$

where ξ is as in (1.2). In addition,

$$e_1' W e_2 = o_{hp}\left(N\varepsilon\right). \tag{4.3}$$

Lemma 4.4 *If e_1, e_2 are as in Lemma 4.2, then*

$$\mathrm{Var}\left(e_1' W e_2\right) = O(\varepsilon), \tag{4.4}$$

and

$$\mathrm{E}\left(e_1' W^3 e_2\right) = O(N\varepsilon). \tag{4.5}$$

5 Proof of the Main Results

This section is devoted to the proof of the main results. The section is split into several subsections, each containing the proof of one main result, for the ease of reading. Unless mentioned otherwise, Assumptions E1 and F1. are made.

5.1 Proof of Theorem 2.1

We start with showing that Theorem 2.1 is a corollary of Lemma 4.1. At this point, it should be clarified that throughout this section, e_j will always be as defined in (2.6).

Proof of Theorem 2.1 For a fixed $i \in \{1, \ldots, k\}$, it follows that

$$|\lambda_i(A) - \lambda_i(\mathrm{E}(A))| \le \|W\| = O_{hp}\left((N\varepsilon)^{1/2}\right),$$

by Lemma 4.1. In order to complete the proof, it suffices to show that

$$\lim_{N\to\infty} (N\varepsilon)^{-1} \lambda_i(\mathrm{E}(A)) = \theta_i,$$

which however follows from the observation that (2.2) implies that

$$\mathrm{E}(A) = N\varepsilon \sum_{j=1}^{k} \theta_j e_j e_j'. \tag{5.1}$$

This completes the proof. □

5.2 Proof of Theorem 2.2

Proceeding towards the proof of Theorem 2.2, let us fix $i \in \mathcal{I}$, once and for all, denote

$$\mu = \lambda_i(A),$$

and let V be a $k \times k$ real symmetric matrix, depending on N which is suppressed in the notation, defined by

$$V(j,l) = \begin{cases} N\varepsilon\sqrt{\theta_j\theta_l}\, e_j' \left(I - \frac{1}{\mu}W\right)^{-1} e_l, & \text{if } \|W\| < \mu, \\ 0, & \text{else,} \end{cases}$$

for all $1 \le j, l \le k$. It should be noted that if $\|W\| < \mu$, then $I - W/\mu$ is invertible.

The proof of Theorem 2.2, which is the main content of this paper, involves several steps, and hence it is imperative to sketch the outline of the proof for the reader's convenience, which is done below.

(1) The first major step of the proof is to show that w.h.p., μ exactly equals $\lambda_i(V)$. This is done in Lemma 5.2. When $i = k = 1$, the matrix V is a scalar and hence in that case the equation boils down to $\mu = V$ which is a consequence of the resolvent equation. For higher values of k, the Gershgorin circle theorem is employed for the desired claim. Therefore, this step is a novelty of the given proof.

(2) The next step in the proof is to write μ as the solution of an equation of the form

$$\mu = \lambda_i\left(\sum_{n=0}^{L} \mu^{-n} Y_n\right) + Error,$$

for suitable matrices $Y_1, Y_2, \ldots$. This is done in Lemma 5.4.

(3) The third step is to replace Y_n by $\mathrm{E}(Y_n)$ in the equation obtained in the above step, for $n \geq 2$. This is done in Lemma 5.5.

(4) Arguably the most important step in the proof is to obtain an equation of the form

$$\mu = \bar{\mu} + \mu^{-1}\zeta + Error,$$

for some deterministic $\bar{\mu}$ depending on N and random ζ. Once again, this is achieved from the previous step with the help of the Gershgorin circle theorem and other linear algebraic tools. This is done in Lemma 5.8.

(5) The final step of the proof is to show that $\bar{\mu}$ of the above step can be replaced by $\mathrm{E}(\mu)$.

Now let us proceed towards executing the above steps for proving Theorem 2.2. As the zeroth step, we show that $V/N\varepsilon$ converges to $\mathrm{Diag}(\theta_1, \ldots, \theta_k)$, that is, the $k \times k$ diagonal matrix with diagonal entries $\theta_1, \ldots, \theta_k$, w.h.p.

Lemma 5.1 *As $N \to \infty$,*

$$V(j,l) = N\varepsilon\theta_j\left(\mathbf{1}(j=l) + o_{hp}(1)\right), 1 \leq j, l \leq k.$$

Proof For fixed $1 \leq j, l \leq k$, writing

$$\left(I - \frac{1}{\mu}W\right)^{-1} = I + O_{hp}\left(\mu^{-1}\|W\|\right),$$

we get that

$$V(j,l) = N\varepsilon\sqrt{\theta_j\theta_l}\left(e_j' e_l + \frac{1}{\mu} O_{hp}(\|W\|)\right).$$

Since

$$\lim_{N\to\infty} e_j' e_l = \mathbf{1}(j=l), \tag{5.2}$$

and

$$\|W\| = o_{hp}(\mu)$$

by Lemma 4.1 and Theorem 2.1, the proof follows. □

The next step, which is one of the main steps in the proof of Theorem 2.2, shows that the ith eigenvalues of A and V are exactly equal w.h.p.

Lemma 5.2 *With high probability,*

$$\mu = \lambda_i(V).$$

The proof of the above lemma is based on the following fact which is a direct consequence of the Gershgorin circle theorem; see Theorem 1.6, pg 8 of [33].

Fact 5.1 *Suppose that U is an $n \times n$ real symmetric matrix. Define*

$$R_l = \sum_{1 \le j \le n,\, j \ne l} |U(j,l)|,\ 1 \le l \le n.$$

If for some $1 \le m \le n$ it holds that

$$U(m,m) + R_m < U(l,l) - R_l \quad \textit{for all } 1 \le l \le m-1, \tag{5.3}$$

and

$$U(m,m) - R_m > U(l,l) + R_l, \quad \textit{for all } m+1 \le l \le n, \tag{5.4}$$

then

$$\left\{\lambda_1(U), \ldots, \lambda_n(U)\right\} \setminus \left(\bigcup_{1 \le l \le k,\, l \ne m} [U(l,l) - R_l, U(l,l) + R_l] \right) = \left\{\lambda_m(U)\right\}.$$

Remark 5.1 The assumptions (5.3) and (5.4) of Fact 5.1 mean that the Gershgorin disk containing the mth largest eigenvalue is disjoint from any other Gershgorin disk.

Proof of Lemma 5.2 The first step is to show that

$$\mu \in \left\{\lambda_1(V), \ldots, \lambda_k(V)\right\} \text{ w.h.p.} \tag{5.5}$$

To that end, fix $N \ge 1$ and a sample point for which $\|W\| < \mu$. The following calculations are done for that fixed sample point.

Let v be an eigenvector of A, with norm 1, corresponding to $\lambda_i(A)$. That is,

$$\mu v = Av = Wv + N\varepsilon \sum_{l=1}^{k} \theta_l (e_l' v) e_l, \tag{5.6}$$

by (5.1). Since $\mu > \|W\|$, $\mu I - W$ is invertible, and hence

$$v = N\varepsilon \sum_{l=1}^{k} \theta_l (e_l' v)\, (\mu I - W)^{-1}\, e_l. \tag{5.7}$$

Fixing $j \in \{1, \ldots, k\}$ and premultiplying the above by $\sqrt{\theta_j}\mu e_j'$ yields

$$\mu\sqrt{\theta_j}(e_j' v) = N\varepsilon \sum_{l=1}^{k} \sqrt{\theta_j}\theta_l (e_l' v) e_j' \left(I - \frac{1}{\mu} W\right)^{-1} e_l = \sum_{l=1}^{k} V(j,l)\sqrt{\theta_l}(e_l' v).$$

As the above holds for all $1 \le j \le k$, this means that if

$$u = \left[\sqrt{\theta_1}(e_1' v) \ldots \sqrt{\theta_k}(e_k' v)\right]', \tag{5.8}$$

then

$$Vu = \mu u. \tag{5.9}$$

Recalling that in the above calculations a sample point is fixed such that $\|W\| < \mu$, what we have shown, in other words, is that a vector u satisfying the above exists w.h.p.

In order to complete the proof of (5.5), it suffices to show that u is a non-null vector w.h.p. To that end, premultiply (5.6) by v' to obtain that

$$\mu = v'Wv + N\varepsilon\|u\|^2.$$

Dividing both sides by $N\varepsilon$ and using Lemma 4.1 implies that

$$\|u\|^2 = \theta_i + o_{hp}(1).$$

Thus, u is a non-null vector w.h.p. From this and (5.9), (5.5) follows.

Lemma 5.1 shows that for all $l \in \{1, \ldots, i-1\}$,

$$\left[V(i,i) + \sum_{1\le j\le k,\, j\ne i} |V(i,j)|\right] - \left[V(l,l) - \sum_{1\le j\le k,\, j\ne l} |V(l,j)|\right]$$
$$= N\varepsilon\,(\theta_i - \theta_l)\,(1 + o_{hp}(1)),$$

as $N \to \infty$. Since $i \in \mathcal{I}$, $\theta_i - \theta_l < 0$, and hence

$$V(i,i) + \sum_{1\le j\le k,\, j\ne i} |V(i,j)| < V(l,l) - \sum_{1\le j\le k,\, j\ne l} |V(l,j)| \text{ w.h.p.}$$

A similar calculation shows that for $l \in \{i+1, \ldots, k\}$,

$$V(i,i) - \sum_{1\le j\le k,\, j\ne i} |V(i,j)| > V(l,l) + \sum_{1\le j\le k,\, j\ne l} |V(l,j)| \text{ w.h.p.}$$

In view of (5.5) and Fact 5.1, the proof would follow once it can be shown that for all $l \in \{1, \ldots, k\} \setminus \{i\}$,

$$|\mu - V(l,l)| > \sum_{1\le j\le k,\, j\ne l} |V(l,j)| \text{ w.h.p.}$$

This follows, once again, by dividing both sides by $N\varepsilon$ and using Theorem 2.1 and Lemma 5.1. This completes the proof. □

The next step is to write

$$\left(I - \frac{1}{\mu}W\right)^{-1} = \sum_{n=0}^{\infty} \mu^{-n} W^n, \tag{5.10}$$

which is possible because $\|W\| < \mu$. Denote

$$Z_{j,l,n} = e_j' W^n e_l, \quad 1 \le j,\ l \le k,\ n \ge 0,$$

which should not be confused with Z_{ij} defined in (2.14), and for $n \ge 0$, let Y_n be a $k \times k$ matrix with

$$Y_n(j,l) = \sqrt{\theta_j\theta_l}\, N\varepsilon Z_{j,l,n}, \quad 1 \le j,\ l \le k.$$

The following bounds will be used several times.

Lemma 5.3 *It holds that*

$$\mathrm{E}\left(\|Y_1\|\right) = O\left(N\varepsilon^{3/2}\right),$$

and

$$\|Y_1\| = o_{hp}\left((N\varepsilon)^2\right).$$

Proof Lemma 4.4 implies that

$$\mathrm{Var}\left(Z_{j,l,1}\right) = O(\varepsilon), \quad 1 \le j, l \le k.$$

Hence,

$$\mathrm{E}\|Y_1\| = O\left(N\varepsilon \sum_{j,l=1}^{k} \mathrm{E}|Z_{j,l,1}|\right) = O\left(N\varepsilon \sum_{j,l=1}^{k} \sqrt{\mathrm{Var}(Z_{j,l,1})}\right) = O\left(N\varepsilon^{3/2}\right),$$

the equality in the second line using the fact that $Z_{j,l,1}$ has mean 0. This proves the first claim. The second claim follows from (4.3) of Lemma 4.3. ◻

The next step is to truncate the infinite sum in (5.10) to level L, where $L = [\log N]$ as defined before.

Lemma 5.4 *It holds that*

$$\mu = \lambda_i\left(\sum_{n=0}^{L} \mu^{-n} Y_n\right) + o_{hp}\left(\sqrt{\varepsilon}\right).$$

Proof From the definition of V, it is immediate that for $1 \le j, l \le k$,

$$V(j,l) = N\varepsilon\sqrt{\theta_j\theta_l} \sum_{n=0}^{\infty} \mu^{-n} e_j' W^n e_l \, \mathbf{1}(\|W\| < \mu),$$

and hence

$$V = \mathbf{1}(\|W\| < \mu) \sum_{n=0}^{\infty} \mu^{-n} Y_n.$$

For the sake of notational simplicity, let us suppress $\mathbf{1}(\|W\| < \mu)$. Therefore, with the implicit understanding that the sum is set as zero if $\|W\| \ge \mu$, for the proof it suffices to check that

$$\left\| \sum_{n=L+1}^{\infty} \mu^{-n} Y_n \right\| = o_{hp}(\sqrt{\varepsilon}). \tag{5.11}$$

To that end, Theorem 2.1 and Lemma 4.1 imply that

$$\left\| \sum_{n=L+1}^{\infty} \mu^{-n} Y_n \right\| \le \sum_{n=L+1}^{\infty} |\mu|^{-n} \|Y_n\| = O_{hp}\left((N\varepsilon)^{-(L-1)/2}\right).$$

In order to prove (5.11), it suffices to show that as $N \to \infty$,

$$-\log\varepsilon = o\left((L-1)\log(N\varepsilon)\right). \tag{5.12}$$

To that end, recall (1.2) to argue that

$$N^{-1} = o(\varepsilon) \tag{5.13}$$

and

$$\log\log N = O(\log(N\varepsilon)). \tag{5.14}$$

By (5.13), it follows that

$$-\log\varepsilon = O\left(\log N\right)$$

$$= o\left(\log N \log\log N\right)$$
$$= o\left((L-1)\log(N\varepsilon)\right),$$

the last line using (5.14). Therefore, (5.12) follows, which ensures (5.11), which in turn completes the proof. □

In the next step, Y_n is replaced by its expectation for $n \geq 2$.

Lemma 5.5 *It holds that*

$$\mu = \lambda_i \left(Y_0 + \mu^{-1} Y_1 + \sum_{n=2}^{L} \mu^{-n} \mathrm{E}(Y_n) \right) + o_{hp}\left(\sqrt{\varepsilon}\right).$$

Proof In view of Theorem 2.1 and Lemma 5.4, all that has to be checked is

$$\sum_{n=2}^{L} (N\varepsilon)^{-n} \|Y_n - \mathrm{E}(Y_n)\| = o_{hp}(\sqrt{\varepsilon}). \tag{5.15}$$

For that, invoke Lemma 4.3 to claim that

$$\max_{2\leq n\leq L,\, 1\leq j,l\leq k} P\left(\left|Z_{j,l,n} - \mathrm{E}(Z_{j,l,n})\right| > N^{(n-1)/2}\varepsilon^{n/2}(\log N)^{n\xi/4}\right)$$
$$= O\left(e^{-(\log N)^{\eta_1}}\right), \tag{5.16}$$

where ξ is as in (1.2).

Our next claim is that there exists $C_2 > 0$ such that for N large,

$$\bigcap_{2\leq n\leq L, 1\leq j,l\leq k} \left[\left|Z_{j,l,n} - \mathrm{E}(Z_{j,l,n})\right| \leq N^{(n-1)/2}\varepsilon^{n/2}(\log N)^{n\xi/4}\right] \tag{5.17}$$

$$\subset \left[\sum_{n=2}^{L} (N\varepsilon)^{-n} \|Y_n - \mathrm{E}(Y_n)\| \leq C_2\sqrt{\varepsilon}\left((N\varepsilon)^{-1}(\log N)^{\xi}\right)^{1/2}\right].$$

To see this, suppose that the event on the left hand side holds. Then, for fixed $1 \leq j, l \leq k$, and large N,

$$\sum_{n=2}^{L} (N\varepsilon)^{-n} \left\|Y_n(j,l) - \mathrm{E}\left[Y_n(j,l)\right]\right\|$$
$$\leq \theta_1 N\varepsilon \sum_{n=2}^{L} (N\varepsilon)^{-n} \left|Z_{j,l,n} - \mathrm{E}\left(Z_{j,l,n}\right)\right|$$
$$\leq \theta_1 \sum_{n=2}^{\infty} (N\varepsilon)^{-(n-1)} N^{(n-1)/2}\varepsilon^{n/2}(\log N)^{n\xi/4}$$
$$= \left[1 - (N\varepsilon)^{-1/2}(\log N)^{\xi/4}\right]^{-1} \theta_1\sqrt{\varepsilon}(N\varepsilon)^{-1/2}(\log N)^{\xi/2}.$$

Thus, (5.17) holds for some $C_2 > 0$.

Combining (5.16) and (5.17), it follows that

$$P\left(\sum_{n=2}^{L} (N\varepsilon)^{-n} \|Y_n - \mathrm{E}(Y_n)\| > C_2\sqrt{\varepsilon}\left((N\varepsilon)^{-1}(\log N)^{\xi}\right)^{1/2}\right)$$

$$= O\left(\log N e^{-(\log N)^{\eta_1}}\right)$$
$$= o\left(e^{-(\log N)^{(1+\eta_1)/2}}\right).$$

This, with the help of (1.2), establishes (5.15) from which the proof follows. □

The goal of the next two lemmas is replacing μ by a deterministic quantity in

$$\sum_{n=2}^{L} \mu^{-n} \mathrm{E}(Y_n).$$

Lemma 5.6 *For N large, the deterministic equation*

$$x = \lambda_i \left(\sum_{n=0}^{L} x^{-n} \mathrm{E}(Y_n) \right), \; x > 0, \tag{5.18}$$

has a solution $\tilde{\mu}$ such that

$$0 < \liminf_{N\to\infty} (N\varepsilon)^{-1} \tilde{\mu} \le \limsup_{N\to\infty} (N\varepsilon)^{-1} \tilde{\mu} < \infty. \tag{5.19}$$

Proof Define a function

$$h : (0, \infty) \to \mathbb{R},$$

by

$$h(x) = \lambda_i \left(\sum_{n=0}^{L} x^{-n} \mathrm{E}(Y_n) \right).$$

Our first claim is that for any fixed $x > 0$,

$$\lim_{N\to\infty} (N\varepsilon)^{-1} h\left(x N\varepsilon\right) = \theta_i. \tag{5.20}$$

To that end, observe that since $\mathrm{E}(Y_1) = 0$,

$$h\left(x N\varepsilon\right) = \lambda_i \left(\mathrm{E}(Y_0) + \sum_{n=2}^{L} (x N\varepsilon)^{-n} \mathrm{E}(Y_n) \right).$$

Recalling that

$$Y_0(j,l) = N\varepsilon \sqrt{\theta_j \theta_l}\, e_j' e_l, \quad 1 \le j,\; l \le k,$$

it follows by (5.2) that

$$\lim_{N\to\infty} (N\varepsilon)^{-1} \mathrm{E}(Y_0) = \mathrm{Diag}(\theta_1, \ldots, \theta_k). \tag{5.21}$$

Lemma 4.2 implies that

$$\mathrm{E}(Z_{j,l,n}) \le \left(O(N\varepsilon)\right)^{n/2},$$

uniformly for $2 \le n \le L$, and hence there exists $0 < C_3 < \infty$ with

$$\|\mathrm{E}(Y_n)\| \le (C_3 N\varepsilon)^{n/2+1}, \quad 2 \le n \le L. \tag{5.22}$$

Therefore,

$$\left\| \sum_{n=2}^{L} (xN\varepsilon)^{-n} \mathrm{E}(Y_n) \right\| \le \sum_{n=2}^{\infty} (xN\varepsilon)^{-n} (C_3 N\varepsilon)^{n/2+1} \to C_3^2 x^{-2},$$

as $N \to \infty$. With the help of (5.21), this implies that

$$\lim_{N\to\infty} (N\varepsilon)^{-1} \left(\sum_{n=0}^{L} (xN\varepsilon)^{-n} \mathrm{E}(Y_n) \right) = \mathrm{Diag}(\theta_1, \ldots, \theta_k),$$

and hence (5.20) follows. It follows that for a fixed $0 < \delta < \theta_i$.

$$\lim_{N\to\infty} (N\varepsilon)^{-1} \left[N\varepsilon(\theta_i + \delta) - h\left((\theta_i + \delta)N\varepsilon\right) \right] = \delta,$$

and thus, for large N,

$$N\varepsilon(\theta_i + \delta) > h\left((\theta_i + \delta)N\varepsilon\right).$$

Similarly, again for large N,

$$N\varepsilon(\theta_i - \delta) < h\left((\theta_i - \delta)N\varepsilon\right).$$

Hence, for N large, (5.18) has a solution $\tilde{\mu}$ in $[(N\varepsilon)(\theta_i - \delta), (N\varepsilon)(\theta_i + \delta)]$, which trivially satisfies (5.19). Hence the proof. □

Lemma 5.7 *If $\tilde{\mu}$ is as in Lemma 5.6, then*

$$\mu - \tilde{\mu} = O_{hp}\left((N\varepsilon)^{-1} \|Y_1\| + \sqrt{\varepsilon} \right).$$

Proof Lemmas 5.5 and 5.6 imply that

$$\begin{aligned}
&|\mu - \tilde{\mu}| \\
&= \left| \lambda_i \left(Y_0 + \mu^{-1} Y_1 + \sum_{n=2}^{L} \mu^{-n} \mathrm{E}(Y_n) \right) - \lambda_i \left(\sum_{n=0}^{L} \tilde{\mu}^{-n} \mathrm{E}(Y_n) \right) \right| + o_{hp}(\sqrt{\varepsilon}) \\
&\le \|\mu^{-1} Y_1\| + |\mu - \tilde{\mu}| \sum_{n=2}^{L} \mu^{-n} \tilde{\mu}^{-n} \|E(Y_n)\| \sum_{j=0}^{n-1} \mu^j \tilde{\mu}^{n-1-j} + o_{hp}(\sqrt{\varepsilon}) \\
&= |\mu - \tilde{\mu}| \sum_{n=2}^{L} \mu^{-n} \tilde{\mu}^{-n} \|E(Y_n)\| \sum_{j=0}^{n-1} \mu^j \tilde{\mu}^{n-1-j} + O_{hp}\left((N\varepsilon)^{-1} \|Y_1\| + \sqrt{\varepsilon} \right).
\end{aligned}$$

Thus,

$$|\mu - \tilde{\mu}| \left[1 - \sum_{n=2}^{L} \mu^{-n} \tilde{\mu}^{-n} \|E(Y_n)\| \sum_{j=0}^{n-1} \mu^j \tilde{\mu}^{n-1-j} \right] \le O_{hp}\left((N\varepsilon)^{-1} \|Y_1\| + \sqrt{\varepsilon} \right) \tag{5.23}$$

Equations (5.19) and (5.22) imply that

$$\left|\sum_{n=2}^{L}\mu^{-n}\tilde{\mu}^{-n}\|E(Y_n)\|\sum_{j=0}^{n-1}\mu^{j}\tilde{\mu}^{n-1-j}\right| = O_{hp}\left(\sum_{n=2}^{\infty}n(N\varepsilon)^{-(n+1)}(C_3N\varepsilon)^{n/2+1}\right)$$
$$= O_{hp}\left((N\varepsilon)^{-1}\right)$$
$$= o_{hp}(1),\ N\to\infty. \tag{5.24}$$

This completes the proof with the help of (5.23). □

The next lemma is arguably the most important step in the proof of Theorem 2.2, the other major step being Lemma 5.2.

Lemma 5.8 *There exists a deterministic $\bar{\mu}$, which depends on N, such that*

$$\mu = \bar{\mu} + \mu^{-1}Y_1(i,i) + o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right).$$

Proof Define a $k\times k$ deterministic matrix

$$X = \sum_{n=0}^{L}\tilde{\mu}^{-n}\mathrm{E}(Y_n),$$

which, as usual, depends on N. Lemma 5.7 and (5.24) imply that

$$\left\|X - \sum_{n=0}^{L}\mu^{-n}\mathrm{E}(Y_n)\right\| \le |\mu-\tilde{\mu}|\sum_{n=2}^{L}\mu^{-n}\tilde{\mu}^{-n}\|\mathrm{E}(Y_n)\|\sum_{j=0}^{n-1}\mu^{j}\tilde{\mu}^{n-1-j}$$
$$= o_{hp}\left(|\mu-\tilde{\mu}|\right)$$
$$= o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right).$$

By Lemma 5.5 it follows that

$$\mu = \lambda_i\left(\mu^{-1}Y_1 + X\right) + o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right). \tag{5.25}$$

Let

$$H = X + \mu^{-1}Y_1 - \left(X(i,i) + \mu^{-1}Y_1(i,i)\right)I,$$
$$M = X - X(i,i)I,$$

and

$$\bar{\mu} = \lambda_i(X).$$

Clearly,

$$\lambda_i\left(\mu^{-1}Y_1 + X\right) = X(i,i) + \mu^{-1}Y_1(i,i) + \lambda_i(H) = \bar{\mu} - \lambda_i(M) + \mu^{-1}Y_1(i,i) + \lambda_i(H).$$

Thus, the proof would follow with the aid of (5.25) if it can be shown that

$$\lambda_i(H) - \lambda_i(M) = o_{hp}\left((N\varepsilon)^{-1}\|Y_1\|\right). \tag{5.26}$$

If $k=1$, then $i=1$ and hence $H=M=0$. Thus, the above is a tautology in that case. Therefore, assume without loss of generality that $k\ge 2$.

Proceeding towards proving (5.26) when $k\ge 2$, set

$$U_1 = (N\varepsilon)^{-1}M, \tag{5.27}$$

and

$$U_2 = (N\varepsilon)^{-1} H. \tag{5.28}$$

The main idea in the proof of (5.26) is to observe that the eigenvector of U_1 corresponding to $\lambda_i(U_1)$ is same as that of M corresponding to $\lambda_i(M)$, and likewise for U_2 and X. Hence, the first step is to use this to get a bound on the differences between the eigenvectors in terms of $\|U_1 - U_2\|$.

An important observation that will be used later is that

$$\|U_1 - U_2\| = O_{hp}\left((N\varepsilon)^{-2}\|Y_1\|\right). \tag{5.29}$$

The second claim of Lemma 5.3 implies that the right hand side above is $o_{hp}(1)$. The same implies that for $m = 1, 2$ and $1 \leq j, l \leq k$,

$$U_m(j,l) = (\theta_j - \theta_i)\mathbf{1}(j = l) + o_{hp}(1), \quad N \to \infty. \tag{5.30}$$

In other words, as $N \to \infty$, U_1 and U_2 converge to $\text{Diag}(\theta_1 - \theta_i, \ldots, \theta_k - \theta_i)$ w.h.p. Therefore,

$$\lambda_i(U_m) = o_{hp}(1), \quad m = 1, 2. \tag{5.31}$$

Let $\tilde{U}_m$, for $m = 1, 2$, be the $(k-1) \times (k-1)$ matrix (recall that $k \geq 2$) obtained by deleting the ith row and the ith column of U_m, and let $\tilde{u}_m$ be the $(k-1) \times 1$ vector obtained from the ith column of U_m by deleting its ith entry. It is worth recording, for possible future use, that

$$\|\tilde{u}_m\| = o_{hp}(1), \quad m = 1, 2, \tag{5.32}$$

which follows from (5.30), and that

$$\|\tilde{u}_1 - \tilde{u}_2\| = O_{hp}\left((N\varepsilon)^{-2}\|Y_1\|\right), \tag{5.33}$$

follows from (5.29).

Equations (5.30) and (5.31) imply that $\tilde{U}_m - \lambda_i(U_m)I_{k-1}$ converges w.h.p. to

$$\text{Diag}(\theta_1 - \theta_i, \ldots, \theta_{i-1} - \theta_i, \theta_{i+1} - \theta_i, \theta_k - \theta_i).$$

Since $i \in \mathcal{I}$, the above matrix is invertible. Fix $\delta > 0$ such that every matrix in the closed δ-neighborhood B_δ, in the sense of operator norm, of the above matrix is invertible. Let

$$C_4 = \sup_{E \in B_\delta} \|E^{-1}\|. \tag{5.34}$$

Then, $C_4 < \infty$. Besides, there exists $C_5 < \infty$ satisfying

$$\left\|E_1^{-1} - E_2^{-1}\right\| \leq C_5\|E_1 - E_2\|, \quad E_1, E_2 \in B_\delta. \tag{5.35}$$

Fix $N \geq 1$ and a sample point such that $\tilde{U}_m - \lambda_i(U_m)I_{k-1}$ belongs to B_δ. Then, it is invertible. Define a $(k-1) \times 1$ vector

$$\tilde{v}_m = -\left[\tilde{U}_m - \lambda_i(U_m)I_{k-1}\right]^{-1} \tilde{u}_m, \quad m = 1, 2,$$

and a $k \times 1$ vector

$$v_m = \left[\tilde{v}_m(1), \ldots, \tilde{v}_m(i-1),\ 1,\ \tilde{v}_m(i), \ldots, \tilde{v}_m(k-1)\right]', \quad m = 1, 2.$$

It is immediate that

$$\|\tilde{v}_m\| \le C_4 \|\tilde{u}_m\|, \quad m = 1, 2. \tag{5.36}$$

Our next claim is that

$$U_m v_m = \lambda_i(U_m) v_m, \quad m = 1, 2. \tag{5.37}$$

This claim is equivalent to

$$[U_m - \lambda_i(U_m) I_k]\, v_m = 0. \tag{5.38}$$

Let $\bar{U}_m$ be the $(k-1) \times k$ matrix obtained by deleting the ith row of $U_m - \lambda_i(U_m) I_k$. Since the latter matrix is singular, and $\tilde{U}_m - \lambda_i(U_m) I_{k-1}$ is invertible, it follows that the ith row of $U_m - \lambda_i(U_m) I_k$ lies in the row space of $\bar{U}_m$. In other words, the row spaces of $U_m - \lambda_i(U_m) I_k$ and $\bar{U}_m$ are the same, and so do their null spaces. Thus, (5.38) is equivalent to

$$\bar{U}_m v_m = 0.$$

To see the above, observe that the ith column of $\bar{U}_m$ is $\tilde{u}_m$, and hence we can partition

$$\bar{U}_m = \left[\bar{U}_{m1}\, \tilde{u}_m\, \bar{U}_{m2}\right],$$

where $\bar{U}_{m1}$ and $\bar{U}_{m2}$ are of order $(k-1) \times (i-1)$ and $(k-1) \times (k-i)$, respectively. Furthermore,

$$\left[\bar{U}_{m1}\, \bar{U}_{m2}\right] = \tilde{U}_m - \lambda_i(U_m) I_{k-1}.$$

Therefore,

$$\bar{U}_m v_m = \tilde{u}_m + \left[\bar{U}_{m1}\, \bar{U}_{m2}\right] \tilde{v}_m = \tilde{u}_m + \left(\tilde{U}_m - \lambda_i(U_m) I_{k-1}\right) \tilde{v}_m = 0.$$

Hence, (5.38) follows, which proves (5.37).

Next, we note

$$\begin{aligned}
\|v_1 - v_2\| = \|\tilde{v}_1 - \tilde{v}_2\| &\le \|(\tilde{U}_1 - \lambda_i(U_1) I_{k-1})^{-1}\| \|\tilde{u}_1 - \tilde{u}_2\| \\
&\quad + \|(\tilde{U}_1 - \lambda_i(U_1) I_{k-1})^{-1} - (\tilde{U}_2 - \lambda_i(U_2) I_{k-1})^{-1}\| \|\tilde{u}_2\| \\
&\le C_4 \|\tilde{u}_1 - \tilde{u}_2\| + C_5 \|(\tilde{U}_1 - \lambda_i(U_1) I_{k-1}) - (\tilde{U}_2 - \lambda_i(U_2) I_{k-1})\| \|\tilde{u}_2\|,
\end{aligned}$$

C_4 and C_5 being as in (5.34) and (5.35), respectively. Recalling that the above calculation was done on an event of high probability, what we have proven, with the help of (5.29) and (5.33), is that

$$\|v_1 - v_2\| = O_{hp}\left((N\varepsilon)^{-2} \|Y_1\|\right).$$

Furthermore, (5.32) and (5.36) imply that

$$\|\tilde{v}_m\| = o_{hp}(1).$$

Finally, noting that

$$U_m(i, i) = 0, \quad m = 1, 2,$$

and that

$$v_m(i) = 1, \quad m = 1, 2,$$

it follows that

$$
\begin{aligned}
|\lambda_i(U_1) - \lambda_i(U_2)| &= \left| \sum_{1\le j\le k,\, j\ne i} U_1(i,j)v_1(j) - \sum_{1\le j\le k,\, j\ne i} U_2(i,j)v_2(j) \right| \\
&\le \sum_{1\le j\le k,\, j\ne i} |U_1(i,j)||v_1(j) - v_2(j)| \\
&\quad + \sum_{1\le j\le k,\, j\ne i} |U_1(i,j) - U_2(i,j)|v_2(j)| \\
&= O_{hp}\left(\|\tilde u_1\|\|v_1 - v_2\| + \|U_1 - U_2\|\|\tilde v_2\|\right) \\
&= o_{hp}\left((N\varepsilon)^{-2}\|Y_1\|\right).
\end{aligned}
$$

Recalling (5.27) and (5.28), (5.26) follows, which completes the proof in conjunction with (5.25). □

Now, we are in a position to prove Theorem 2.2.

Proof of Theorem 2.2 Recalling that

$$Y_1(i,i) = \theta_i N\varepsilon\, e_i' W e_i,$$

it suffices to show that

$$\mu - \mathrm{E}(\mu) = \mu^{-1}Y_1(i,i) + o_p(\sqrt{\varepsilon}). \tag{5.39}$$

Lemma 5.8 implies that

$$\mu - \bar\mu = \mu^{-1}Y_1(i,i) + o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right) = O_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right), \tag{5.40}$$

a consequence of which, combined with Lemma 5.3, is that

$$\lim_{N\to\infty} (N\varepsilon)^{-1}\bar\mu = \theta_i. \tag{5.41}$$

Thus,

$$
\begin{aligned}
\left|\frac{1}{\bar\mu}Y_1(i,i) - \frac{1}{\mu}Y_1(i,i)\right| &= O_{hp}\left((N\varepsilon)^{-2}|\mu - \bar\mu|\|Y_1\|\right) \\
&= o_{hp}\left(|\mu - \bar\mu|\right) \\
&= o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right) \\
&= o_p(\sqrt{\varepsilon}),
\end{aligned} \tag{5.42}
$$

Lemma 5.3 implying the second line, the third line following from (5.40) and the fact that

$$\|Y_1\| = O_p(N\varepsilon^{3/2}), \tag{5.43}$$

which is also a consequence of the former lemma, being used for the last line. Using Lemma 5.8 once again, we get that

$$\mu = \bar\mu + \frac{1}{\bar\mu}Y_1(i,i) + o_{hp}\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right). \tag{5.44}$$

Let

$$R = \mu - \bar\mu - \frac{1}{\bar\mu}Y_1(i,i).$$

Clearly,

$$\mathrm{E}(R) = \mathrm{E}(\mu) - \bar{\mu},$$

and (5.44) implies that for $\delta > 0$ there exists $\eta > 1$ with

$$\mathrm{E}|R| \le \delta(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\mathrm{E}\|Y_1\|) + \mathrm{E}^{1/2}\left(\mu - \bar{\mu} - \frac{1}{\bar{\mu}}Y_1(i,i)\right)^2 O(e^{-(\log N)^\eta}).$$

Lemma 5.3 implies that

$$\mathrm{E}|R| \le o(\sqrt{\varepsilon}) + \mathrm{E}^{1/2}\left(\mu - \bar{\mu} - \frac{1}{\bar{\mu}}Y_1(i,i)\right)^2 O(e^{-(\log N)^\eta}).$$

Next, (5.41) and that $|\mu| \le N^2$ a.s. imply that

$$\mathrm{E}^{1/2}\left(\mu - \bar{\mu} - \frac{1}{\bar{\mu}}Y_1(i,i)\right)^2 = O(N^2) = o(\varepsilon^{1/2}N^3) = o(\varepsilon^{1/2}e^{(\log N)^\eta}).$$

Thus,

$$\mathrm{E}|R| = o(\sqrt{\varepsilon}),$$

and hence

$$\mathrm{E}(\mu) = \bar{\mu} + o(\sqrt{\varepsilon}).$$

This, in view of (5.44), implies that

$$\mu = \mathrm{E}(\mu) + \frac{1}{\bar{\mu}}Y_1(i,i) + o_p\left((N\varepsilon)^{-1}\|Y_1\| + \sqrt{\varepsilon}\right) = \mathrm{E}(\mu) + \frac{1}{\bar{\mu}}Y_1(i,i) + o_p\left(\sqrt{\varepsilon}\right),$$

the second line following from (5.43). This establishes (5.39) with the help of (5.42), and hence the proof. □

5.3 Proof of Theorem 2.3

Theorems 2.1 and 2.2 establish Theorem 2.3 with the help of (2.10).

5.4 Proof of Theorem 2.4

Now we shall proceed toward proving Theorem 2.4. For the rest of this section, that is, this subsection and the subsequent two, Assumption F2. holds. In other words, $r_1, \ldots, r_k$ are assumed to be Lipschitz continuous and hence so is f.

The following lemma essentially proves Theorem 2.4.

Lemma 5.9 *Under Assumptions* E1 *and F2.,*

$$\mu = \lambda_i\left(Y_0 + (N\varepsilon\theta_i)^{-2}\mathrm{E}(Y_2)\right) + O_p\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right).$$

Proof Lemma 5.5 implies that

$$\mu = \lambda_i\left(\sum_{n=0}^{3}\mu^{-n}\mathrm{E}(Y_n)\right) + O_p\left(\mu^{-1}\|Y_1\| + \sum_{n=4}^{L}\mu^{-n}\|\mathrm{E}(Y_n)\|\right) + o_p(\sqrt{\varepsilon}).$$

Equation (5.43) implies that

$$\mu = \lambda_i\left(\sum_{n=0}^{3}\mu^{-n}\mathrm{E}(Y_n)\right) + O_p\left(\sqrt{\varepsilon} + \sum_{n=4}^{L}\mu^{-n}\|\mathrm{E}(Y_n)\|\right).$$

From (5.22), it follows that

$$\sum_{n=4}^{L}\mu^{-n}\|\mathrm{E}(Y_n)\| = O_p\left((N\varepsilon)^{-1}\right),$$

and hence

$$\mu = \lambda_i\left(\sum_{n=0}^{3}\mu^{-n}\mathrm{E}(Y_n)\right) + O_p\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right). \tag{5.45}$$

Lemma 4.4, in particular (4.5) therein, implies that

$$\|\mathrm{E}(Y_3)\| = O\left((N\varepsilon)^2\right),$$

and hence

$$\mu^{-3}\|\mathrm{E}(Y_3)\| = O_p\left((N\varepsilon)^{-1}\right).$$

This, in conjunction with (5.45), implies that

$$\mu = \lambda_i\left(Y_0 + \mu^{-2}\mathrm{E}(Y_2)\right) + O_p\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right). \tag{5.46}$$

An immediate consequence of the above and (5.22) is that

$$\mu = \lambda_i(Y_0) + O_p(1). \tag{5.47}$$

Applying Fact 5.1 as in the proof of Lemma 5.2, it can be shown that

$$|\lambda_i(Y_0) - Y_0(i,i)| \le \sum_{1\le j\le k,\, j\ne i}|Y_0(i,j)|. \tag{5.48}$$

Since r_i and r_j are Lipschitz functions, it holds that

$$e_i'e_j = \mathbf{1}(i=j) + O\left(N^{-1}\right).$$

Hence, it follows that

$$Y_0(i,i) = N\varepsilon\left(\theta_i + O(N^{-1})\right) = N\varepsilon\theta_i + O(\varepsilon),$$

and similarly,

$$Y_0(i,j) = O(\varepsilon), \quad j \ne i.$$

Combining these findings with (5.48) yields that

$$\lambda_i(Y_0) = N\varepsilon\theta_i + O(\varepsilon). \tag{5.49}$$

Equations (5.47) and (5.49) together imply that

$$\mu = N\varepsilon\theta_i + O_p(1). \tag{5.50}$$

Therefore,

$$\left\|\mu^{-2}\mathrm{E}(Y_2) - (N\varepsilon\theta_i)^{-2}\mathrm{E}(Y_2)\right\|$$

$$= O_p\left((N\varepsilon)^{-3}\|\mathrm{E}(Y_2)\|\right)$$
$$= O_p\left((N\varepsilon)^{-1}\right).$$

This in conjunction with (5.46) completes the proof. □

Theorem 2.4 is a simple corollary of the above lemma, as shown below.

Proof of Theorem 2.4 A consequence of Theorem 2.2 is that

$$\mu - \mathrm{E}(\mu) = O_p(\sqrt{\varepsilon}).$$

The claim of Lemma 5.9 is equivalent to

$$\lambda_i(B) - \mu = O_p\left(\sqrt{\varepsilon} + (N\varepsilon)^{-1}\right).$$

The proof follows by adding the two equations, and noting that B is a deterministic matrix. □

5.5 Proof of Theorem 2.5

Next we proceed towards the proof of Theorem 2.5, for which the following lemma will be useful.

Lemma 5.10 *Under Assumptions* E1 *and* F2*, as* $N \to \infty$,

$$e_j' \left(I - \mu^{-1}W\right)^{-n} e_l = \mathbf{1}(j = l) + O_p\left((N\varepsilon)^{-1}\right), \quad 1 \le j, l \le k, \; n = 1, 2.$$

Proof For a fixed $n = 1, 2$, expand

$$\left(I - \mu^{-1}W\right)^{-n} = I + n\mu^{-1}W + O_p\left(\mu^{-2}\|W\|^2\right).$$

The proof can be completed by proceeding along similar lines as in the proof of Lemma 5.9. □

Now we are in a position to prove Theorem 2.5.

Proof of Theorem 2.5 Theorem 2.1 implies that (2.11) holds for any $i \in \mathcal{I}$. Fix such an i, denote

$$\mu = \lambda_i(A),$$

and let v be the eigenvector of A, having norm 1, corresponding to μ, which is uniquely defined with probability close to 1.

Fix $k \ge 2$, and $j \in \{1, \ldots, k\} \setminus \{i\}$. Premultiplying (5.7) by e_j' yields that

$$e_j' v = N\varepsilon \sum_{l=1}^{k} \theta_l (e_l' v) e_j' \left(\mu I - W\right)^{-1} e_l, \quad \text{w.h.p.} \tag{5.51}$$

Therefore,

$$e_j' v \left(1 - \theta_j \frac{N\varepsilon}{\mu} e_j' \left(I - \mu^{-1}W\right)^{-1} e_j\right)$$
$$= \frac{N\varepsilon}{\mu} \sum_{1 \le l \le k,\, l \ne j} \theta_l (e_l' v) e_j' \left(I - \mu^{-1}W\right)^{-1} e_l, \quad \text{w.h.p.}$$

Lemma 5.10 implies that as $N \to \infty$,

$$1-\theta_j \frac{N\varepsilon}{\mu} e_j' \left(I-\mu^{-1}W\right)^{-1} e_j \xrightarrow{P} 1-\frac{\theta_j}{\theta_i} \neq 0.$$

Therefore,

$$\begin{aligned} e_j' v &= O_p\left(\frac{N\varepsilon}{\mu} \sum_{1\le l\le k,\, l\neq j} \theta_l (e_l' v) e_j' \left(I-\mu^{-1}W\right)^{-1} e_l\right) \\ &= O_p\left(\sum_{1\le l\le k,\, l\neq j} \left|e_j' \left(I-\mu^{-1}W\right)^{-1} e_l\right|\right) \\ &= O_p\left((N\varepsilon)^{-1}\right), \end{aligned}$$

the last line being another consequence of Lemma 5.10. Thus, (2.13) holds.

Using (5.7) once again, we get that

$$1 = (N\varepsilon)^2 \sum_{l,m=1}^{k} \theta_l \theta_m (e_l' v)(e_m' v) e_l' \left(\mu I - W\right)^{-2} e_m,$$

that is,

$$\begin{aligned} &\theta_i^2 (e_i' v)^2 e_i' \left(I-\mu^{-1}W\right)^{-2} e_i \\ &\quad = (N\varepsilon)^{-2}\mu^2 - \sum_{(l,m)\in\{1,\dots,k\}^2\setminus\{(i,i)\}} \theta_l\theta_m (e_l' v)(e_m' v) e_l' \left(I-\mu^{-1}W\right)^{-2} e_m. \end{aligned} \tag{5.52}$$

Using Lemma 5.10 once again, it follows that

$$e_i' \left(I-\mu^{-1}W\right)^{-2} e_i = 1 + O_p\left((N\varepsilon)^{-1}\right).$$

Thus, (2.12) would follow once it's shown that

$$(N\varepsilon)^{-2}\mu^2 = \theta_i^2 + O_p\left((N\varepsilon)^{-1}\right), \tag{5.53}$$

and that for all $(l,m) \in \{1,\dots,k\}^2 \setminus \{(i,i)\}$,

$$(e_l' v)(e_m' v) e_l' \left(I-\mu^{-1}W\right)^{-2} e_m = O_p\left((N\varepsilon)^{-1}\right). \tag{5.54}$$

Equation (5.53) is a trivial consequence of (5.50). For (5.54), assuming without loss of generality that $l \neq i$, (2.13) implies that

$$\begin{aligned} \left|(e_l' v)(e_m' v) e_l' \left(I-\mu^{-1}W\right)^{-2} e_m\right| &= \left|(e_m' v) e_l' \left(I-\mu^{-1}W\right)^{-2} e_m\right| O_p\left((N\varepsilon)^{-1}\right) \\ &\le \left|e_l' \left(I-\mu^{-1}W\right)^{-2} e_m\right| O_p\left((N\varepsilon)^{-1}\right) \\ &= O_p\left((N\varepsilon)^{-1}\right), \end{aligned}$$

the last line following from Lemma 5.10. Thus, (5.54) follows, which in conjunction with (5.53) establishes (2.12). This completes the proof. ◻

5.6 Proof of Theorem 2.6

Finally, Theorem 2.6 is proved below, based on Assumptions E2. and F2.

Proof of Theorem 2.6 Fix $i \in \mathcal{I}$. Recall (5.8) and (5.9), and let u be as defined in the former. Let $\tilde{u}$ be the column vector obtained by deleting the ith entry of u, $\tilde{V}_i$ be the column vector obtained by deleting the ith entry of the ith column of V, and $\tilde{V}$ be the $(k-1) \times (k-1)$ matrix obtained by deleting the ith row and ith column of V. Then, (5.9) implies that

$$\mu\tilde{u} = \tilde{V}\tilde{u} + u(i)\tilde{V}_i, \quad \text{w.h.p.} \tag{5.55}$$

Lemma 5.1 implies that

$$\left\| I_k - \mu^{-1}V - \text{Diag}\left(1 - \frac{\theta_1}{\theta_i}, \ldots, 1 - \frac{\theta_k}{\theta_i}\right) \right\| = o_{hp}(1),$$

and hence $I_{k-1} - \mu^{-1}\tilde{V}$ is non-singular w.h.p. Thus, (5.55) implies that

$$\tilde{u} = u(i)\mu^{-1}\left(I_{k-1} - \mu^{-1}\tilde{V}\right)^{-1}\tilde{V}_i, \quad \text{w.h.p.} \tag{5.56}$$

The next step is to show that

$$\left\| \mu^{-1}V - \text{Diag}\left(\frac{\theta_1}{\theta_i}, \ldots, \frac{\theta_k}{\theta_i}\right) \right\| = o_p\left(\sqrt{\varepsilon}\right). \tag{5.57}$$

To see this, use the fact that f is Lipschitz to write for a fixed $1 \le j, l \le k$,

$$\begin{aligned} V(j,l) &= N\varepsilon\sqrt{\theta_j\theta_l}\left(e_j'e_l + \mu^{-1}e_j'We_l + O_p\left(\mu^{-2}\|W\|^2\right)\right) \\ &= N\varepsilon\sqrt{\theta_j\theta_l}\left(e_j'e_l + O_p\left((N\varepsilon)^{-1}\right)\right) \\ &= N\varepsilon\theta_j\left(\mathbf{1}(j=l) + O_p\left((N\varepsilon)^{-1}\right)\right) \\ &= N\varepsilon\theta_j\left(\mathbf{1}(j=l) + o_p\left(\sqrt{\varepsilon}\right)\right), \end{aligned} \tag{5.58}$$

the last line following from the fact that

$$(N\varepsilon)^{-1} = o\left(\sqrt{\varepsilon}\right), \tag{5.59}$$

which is a consequence of (2.5). This along with (5.50) implies that

$$(N\varepsilon\theta_i)^{-1}\mu = 1 + o_p\left(\sqrt{\varepsilon}\right). \tag{5.60}$$

Combining this with (5.58) yields that

$$\mu^{-1}V(j,l) = \theta_i^{-1}\theta_j\mathbf{1}(j=l) + o_p\left(\sqrt{\varepsilon}\right).$$

Thus, (5.57) follows, an immediate consequence of which is that

$$\left\| \left(I_{k-1} - \mu^{-1}\tilde{V}\right)^{-1} - \tilde{D} \right\| = o_p\left(\sqrt{\varepsilon}\right), \tag{5.61}$$

where

$$\tilde{D} = \left[\text{Diag}\left(1 - \frac{\theta_1}{\theta_i}, \ldots, 1 - \frac{\theta_{i-1}}{\theta_i}, 1 - \frac{\theta_{i+1}}{\theta_i}, \ldots, 1 - \frac{\theta_k}{\theta_i}\right)\right]^{-1}.$$

Next, fix $j \in \{1, \ldots, k\} \setminus \{i\}$. By similar arguments as above, it follows that

$$V(i, j) = N\varepsilon\sqrt{\theta_i\theta_j}\left(\sum_{n=0}^{3}\mu^{-n}e_i'W^ne_j + O_p\left(\mu^{-4}\|W\|^4\right)\right)$$
$$= N\varepsilon\sqrt{\theta_i\theta_j}\sum_{n=0}^{3}\mu^{-n}e_i'W^ne_j + O_p\left((N\varepsilon)^{-1}\right)$$
$$= N\varepsilon\sqrt{\theta_i\theta_j}\sum_{n=1}^{2}\mu^{-n}e_i'W^ne_j + o_p\left(\sqrt{\varepsilon}\right),$$

using (5.59) once again, because

$$N\varepsilon e_i'e_j = O(\varepsilon) = o\left(\sqrt{\varepsilon}\right),$$

and

$$N\varepsilon\mu^{-3}e_i'W^3e_j = O_p\left((N\varepsilon)^{-2}\mathrm{E}(e_i'W^3e_j)\right) = o_p\left(\sqrt{\varepsilon}\right),$$

by (4.5). Thus,

$$V(i, j) - N\varepsilon\sqrt{\theta_i\theta_j}\mu^{-1}e_i'We_j = N\varepsilon\sqrt{\theta_i\theta_j}\mu^{-2}e_i'W^2e_j + o_p\left(\sqrt{\varepsilon}\right)$$
$$= N\varepsilon\sqrt{\theta_i\theta_j}\mu^{-2}\mathrm{E}\left(e_i'W^2e_j\right) + o_p\left(\sqrt{\varepsilon}\right)$$
$$= (N\varepsilon)^{-1}\theta_j^{1/2}\theta_i^{-3/2}\mathrm{E}\left(e_i'W^2e_j\right) + o_p\left(\sqrt{\varepsilon}\right),$$

the second line following from Lemma 4.3, and the last line from (5.59), (5.60) and Lemma 4.2. In particular,

$$V(i, j) = O_p(1).$$

The above in conjunction with (5.61) implies that

$$\left[\left(I_{k-1} - \mu^{-1}\tilde{V}\right)^{-1}\tilde{V}_i\right](j)$$
$$= \left(1 - \frac{\theta_j}{\theta_i}\right)^{-1}\sqrt{\theta_i\theta_j}\left[(N\varepsilon)^{-1}\theta_i^{-2}\mathrm{E}\left(e_i'W^2e_j\right) + N\varepsilon\mu^{-1}e_i'We_j\right] + o_p(\sqrt{\varepsilon}).$$

In light of (5.56), the above means that

$$e_j'v$$
$$= (e_i'v)\mu^{-1}\left(1 - \frac{\theta_j}{\theta_i}\right)^{-1}\left[(N\varepsilon)^{-1}\theta_i^{-1}\mathrm{E}\left(e_i'W^2e_j\right) + N\varepsilon\theta_i\mu^{-1}e_i'We_j + o_p(\sqrt{\varepsilon})\right]$$
$$= \mu^{-1}\left(1 - \frac{\theta_j}{\theta_i}\right)^{-1}\left[(N\varepsilon)^{-1}\theta_i^{-1}\mathrm{E}\left(e_i'W^2e_j\right) + N\varepsilon\theta_i\mu^{-1}e_i'We_j + o_p(\sqrt{\varepsilon})\right],$$

the last line following from (2.12) and (5.59). Using (5.60) once again yields that

$$N\varepsilon(e_j'v) = \frac{1}{\theta_i - \theta_j}\left[(N\varepsilon)^{-1}\theta_i^{-1}\mathrm{E}\left(e_i'W^2e_j\right) + N\varepsilon\theta_i\mu^{-1}e_i'We_j\right] + o_p(\sqrt{\varepsilon}).$$

This completes the proof. □

Acknowledgements The authors thank an anonymous referee for insightful comments that helped improve the paper significantly. RSH thanks Kavita Ramanan for a fruitful discussion. The research of AC and RSH was supported by the MATRICS grant of SERB. SC thanks Matteo Sfragara for helpful discussions.

Appendix

Lemma 6.1 *The eigenfunctions* $\{r_i : 1 \le i \le k\}$ *of the operator* I_f *are Riemann integrable.*

Proof Let $D_f \subset [0,1]\times[0,1]$ be the set of discontinuity points f. Since f is Riemann integrable, the Lebesgue measure of D_f is 0. Let

$$D_f^x = \{y \in [0,\ 1] :\ (x, y) \in D_f\},\ x \in [0,\ 1].$$

If λ is the one dimensional Lebesgue measure, then Fubini's theorem implies that

$$E = \{x \in [0,1] : \lambda(D_f^x) = 0\}$$

has full measure. Fix an $x \in E$ and consider $x_n \to x$ and observe that

$$f(x_n, y) \to f(x, y) \quad \text{for all} y \notin D_f^x.$$

Fix $1 \le i \le k$ and let θ_i be the eigenvalue with corresponding eigenfunction r_i, that is,

$$r_i(x) = \frac{1}{\theta_i}\int_0^1 f(x, y) r_i(y)\,dy. \tag{6.1}$$

Using f is bounded and $r \in L^2[0,1]$, dominated convergence theorem implies

$$r_i(x_n) = \frac{1}{\theta_i}\int_{(D_f^x)^c} f(x_n, y) r_i(y)\,dy \to \frac{1}{\theta_i}\int_0^1 f(x, y) r_i(y)\,dy = r_i(x)$$

and hence r is continuous at x. So the discontinuity points of r_i form a subset of E^c which has Lebesgue measure 0. Further, (6.1) shows that r_i is bounded and hence Riemann integrability follows. □

The following result is a version of the Perron–Frobenius theorem in the infinite dimensional setting (also known as the Krein–Rutman theorem). Since our integral operator is positive, self-adjoint and finite dimensional so the proof in this setting is much simpler and can be derived following the work of [28]. In what follows, we use for $f, g \in L^2[0,1]$, the inner product

$$\langle f,\ g\rangle = \int_0^1 f(x)g(x)dx.$$

Lemma 6.2 *Suppose* $f > 0$ *a.e. on* $[0,1]\times[0,1]$. *Then largest eigenvalue* θ_1 *of* T_f *is positive and the corresponding eigenfunction* r_1 *can be chosen such that* $r_1(x) > 0$ *for almost every* $x \in [0,1]$. *Further,* $\theta_1 > \theta_2$.

Proof First observe that

$$\begin{aligned} 0 < \theta_1 = \langle r_1,\ \theta_1 r_1\rangle = \langle r_1,\ I_f(r_1)\rangle &= |\langle r_1,\ I_f(r_1)\rangle| \\ &\le \langle u_1, I_f(u_1)\rangle \le \theta_1 \end{aligned}$$

where $u_1(x) = |r_1|(x)$ and the last inequality follows from the Rayleigh-Ritz formulation of the largest eigenvalue. Hence note that the string of inequalities is actually an equality, that is,

$$\langle r_1, I_f(r_1)\rangle = \langle u_1, I_f(u_1)\rangle.$$

Breaking $r_1 = r_1^+ - r_1^-$ implies either $r_1^+ = 0$ or $r_1^- = 0$ almost everywhere. Without loss of generality assume that $r_1 \geq 0$ almost everywhere. Using

$$\theta_1 r_1(x) = \int_0^1 f(x, y) r_1(y)\, dy$$

Note that if $r_1(x)$ is zero for some x then due to the positivity assumption on f, $r_1(y) = 0$ for almost every $y \in [0, 1]$ which is a contradiction. Hence we have that $r_1(x) > 0$ almost every $x \in [0, 1]$.

For the final claim, without loss of generality assume that $\int_0^1 r_1(x)\, dx \geq 0$. If $\theta_1 = \theta_2$, then the previous argument would give us $r_2(x) > 0$ and this will contradict the orthogonality of r_1 and r_2. □

Lemmas 4.1–4.4 are proved in the rest of this section. Therefore, the notations used here should refer to those in Sect. 4 and should not be confused with those in Sect. 5. For example, e_1 and e_2 are as in Lemma 4.2.

Proof of Lemma 4.1 Note that for any even integer k

$$\mathrm{E}(\|W\|^k) \leq \mathrm{E}(\mathrm{Tr}(W^k)). \tag{6.2}$$

Using $\mathrm{E}(W(i, j)^2) \leq \varepsilon M$ and condition (1.2) it is immediate that conditions of Theorem 1.4 of [34] are satisfied. We shall use the following estimate from the proof of that result. It follows from [34, Sect. 4]

$$\mathrm{E}(\mathrm{Tr}(W^k)) \leq K_1 N (2\sqrt{\varepsilon M N})^k \tag{6.3}$$

where K_1 is some positive constant and there exists a constant $a > 0$ such that k can be chosen as

$$k = \sqrt{2} a (\varepsilon M)^{1/4} N^{1/4}.$$

Using (6.2), (6.3) and $(1 - x)^k \leq e^{-kx}$ for $k, x > 0$,

$$\begin{aligned}
& P\left(\|W\| \geq 2\sqrt{MN\varepsilon} + C_1 (N\varepsilon)^{1/4} (\log N)^{\xi/4}\right) \\
&= K_1 N \left(1 - \frac{C_1 (N\varepsilon)^{1/4} (\log N)^{\xi/4}}{2\sqrt{MN\varepsilon} + C_1 (N\varepsilon)^{1/4} (\log N)^{\xi/4}}\right)^k \\
&\leq K_1 N \exp\left(-\frac{k C_1 (N\varepsilon)^{1/4} (\log N)^{\xi/4}}{2\sqrt{MN\varepsilon} + C_1 (N\varepsilon)^{1/4} (\log N)^{\xi/4}}\right).
\end{aligned} \tag{6.4}$$

Now plugging in the value of k in the bound (6.4) and using

$$2\sqrt{M} + C_1 (N\varepsilon)^{-1/4} (\log N)^{\xi/4} \leq 2\sqrt{M} + C_1$$

we have

$$(6.4) \leq K_1 N \exp\left(-\frac{C_1 a M^{1/4} \sqrt{2} (\log N)^{\xi/4}}{2\sqrt{M} + C_1}\right) \leq e^{-C_2 (\log N)^{\xi/4}}$$

for some constant $C_2 > 0$ and N large enough. This proves (4.1) and hence the lemma. □

Proof of Lemma 4.2 Let A be the event where Lemma 4.1 holds, that is, $\|W\| \le C\sqrt{N\varepsilon}$ for some constant C. Since the entries of e_1 and e_2 are in $[-1/\sqrt{N}, 1/\sqrt{N}]$ so $\|e_i\| \le 1$ for $i = 1, 2$. Hence on the high probability event it holds that

$$\left|\mathrm{E}\left(e_1' W^n e_2 \mathbf{1}_A\right)\right| \le (CN\varepsilon)^{n/2}.$$

We show that the above expectation on the low probability event A^c is negligible. For that first observe

$$|\mathrm{E}[(e_1' W^n e_2)^2]| \le N^{nC'}$$

for some constant $0 < C' < \infty$. Thus using Lemma 4.1 one has

$$\begin{aligned}\left|\mathrm{E}\left(e_1' W^n e_2 \mathbf{1}_{A^c}\right)\right| &\le \left|\mathrm{E}\left[(e_1' W^n e_2)^2\right]^{1/2}\right| P(A_N^c)^{1/2}\\ &\le \exp\left(nC' \log N - 2^{-1} C_2 (\log N)^{\xi/4}\right)\end{aligned}$$

Since $n \le \log N$ and $\xi > 8$ the result follows. □

Proof of Lemma 4.3 The proof is similar to the proof of Lemma 6.5 of [18]. The exponent in the exponential decay is crucial, so the proof is briefly sketched. Observe that

$$\begin{aligned}&e_1' W^n e_2 - \mathrm{E}\left(e_1' W^n e_2\right)\\ &= \sum_{i \in \{1,\dots,N\}^{n+1}} e_1(i_1) e_2(i_{n+1}) \left(\prod_{l=1}^{n} W(i_l, i_{l+1}) - \mathrm{E}\left[\prod_{l=1}^{n} W(i_l, i_{l+1})\right]\right)\end{aligned} \tag{6.5}$$

To use the independence, one can split the matrix W as $W' + W''$ where the upper triangular matrix W' has entries $W'(i, j) = W(i, j)\mathbf{1}(i \le j)$ and the lower triangular matrix W'' with entries $W''(i, j) = W(i, j)\mathbf{1}(i > j)$. Therefore the above quantity under the sum breaks into 2^n terms each having similar properties. Denote one such term as

$$L_n = \sum_{i \in \{1,\dots,N\}^{n+1}} e_1(i_1) e_2(i_{n+1}) \left(\prod_{l=1}^{n} W'(i_l, i_{l+1}) - \mathrm{E}\left[\prod_{l=1}^{n} W'(i_l, i_{l+1})\right]\right).$$

Using the fact that each entry of e_1 and e_2 are bounded by $1/\sqrt{N}$, it follows by imitating the proof of Lemma 6.5 of [18] that

$$\mathrm{E}[|L_n|^p] \le \frac{(Cnp)^{np}\,(N\varepsilon)^{np/2}}{N^{p/2}},$$

where p is an even integer and C is a positive constant, independent of n and p. Rest of the $2^n - 1$ terms arising in (6.5) have the same bound and hence

$$\begin{aligned}&P\left(\left|e_1' W^n e_2 - \mathrm{E}\left(e_1' W^n e_2\right)\right| > N^{(n-1)/2} \varepsilon^{n/2} (\log N)^{n\xi/4}\right)\\ &\le \frac{(2Cnp)^{np}\,(N\varepsilon)^{np/2}}{N^{p/2} N^{p(n-1)/2} \varepsilon^{pn/2} (\log N)^{pn\xi/4}} = \frac{(2Cnp)^{np}}{(\log N)^{pn\xi/4}}.\end{aligned}$$

Choose $\eta \in (1,\ \xi/4)$ and consider

$$p = \frac{(\log N)^\eta}{2Cn},$$

(with N large enough to make p an even integer) to get

$$P\left(\left|e_1' W^n e_2 - \mathrm{E}\left(e_1' W^n e_2\right)\right| > N^{(n-1)/2}\varepsilon^{n/2}(\log N)^{n\xi/4}\right)$$
$$\leq \exp\left(-\frac{1}{2C}(\log N)^{\eta}(\frac{\xi}{4} - \eta)\log\log N\right).$$

Note that $n \leq L$, ensures that $p > 1$. Since the bound is uniform over all $2 \leq n \leq L$, the first bound (4.2) follows.

For (4.3) one can use Hoeffding's inequality [21, Theorem 2] as follows.

Define

$$\widetilde{A}(k,l) = A(k,l)e_1(k)e_2(l), \quad 1 \leq k \leq l \leq N.$$

Since $A(k,l)$ are Bernoulli random variables, so one has $\{\widetilde{A}(k,l) : 1 \leq k \leq l \leq N\}$ are independent random variables taking values in $[-1/N, 1/N]$ and hence by Hoeffding's inequality we have, for any $\delta > 0$,

$$P\left(\left|\sum_{1\leq k\leq l\leq N}\widetilde{A}(k,l) - E\left(\sum_{1\leq k\leq l\leq N}\widetilde{A}(k,l)\right)\right| > \delta N\varepsilon\right)$$
$$\leq 2\exp\left(-\delta^2(N\varepsilon)^2\right) \leq 2\exp\left(-\delta^2(\log N)^{2\xi}\right).$$

Dealing with the case $k > l$ similarly, the desired bound on $e_1' W e_2$ follows. □

Proof of Lemma 4.4 Follows by a simple moment calculation. □

References

1. Alt, J., Ducatez, R., Knowles, A.: Delocalization transition for critical Erdős–Rényi graphs. ArXiv:2005.14180 (2020)
2. Baik, J., Silverstein, J.W.: Eigenvalues of large sample covariance matrices of spiked population models. J. Multivar. Anal. **97**(6), 1382–1408 (2006). https://doi.org/10.1016/j.jmva.2005.08.003
3. Baik, J., Ben Arous, G., Péché, S.: Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. Ann. Probab. **33**(5), 1643–1697 (2005). https://doi.org/10.1214/009117905000000233
4. Benaych-Georges, F., Nadakuditi, R.R.: The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Adv. Math. **227**(1), 494–521 (2011). https://doi.org/10.1016/j.aim.2011.02.007
5. Benaych-Georges, F., Guionnet, A., Maida, M.: Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. Electron. J. Probab. **16**, 1621–1662 (2011)
6. Benaych-Georges, F., Bordenave, C., Knowles, A.: Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. Ann. Probab. **47**(3), 1653–1676 (2019)
7. Benaych-Georges, F., Bordenave, C., Knowles, A.: Spectral radii of sparse random matrices. Ann. Inst. H. Poincaré Probab. Statist. **56**(3), 2141–2161 (2020). https://doi.org/10.1214/19-AIHP1033
8. Bhamidi, S., Van Der Hofstad, R., van Leeuwaarden, J., et al.: Scaling limits for critical inhomogeneous random graphs with finite third moments. Electron. J. Probab. **15**, 1682–1702 (2010)
9. Bollobás, B., Janson, S., Riordan, O.: The phase transition in inhomogeneous random graphs. Random Struct. Algorithms **31**(1), 3–122 (2007)
10. Capitaine, M., Péché, S.: Fluctuations at the edges of the spectrum of the full rank deformed GUE. Probab. Theory Relat. Fields **165**(1–2), 117–161 (2016). https://doi.org/10.1007/s00440-015-0628-6
11. Capitaine, M., Donati-Martin, C., Féral, D., et al.: The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. Ann. Probab. **37**(1), 1–47 (2009)
12. Capitaine, M., Donati-Martin, C., Féral, D.: Central limit theorems for eigenvalues of deformations of Wigner matrices. Ann. Probab. Stat. **48**(1), 107–133 (2012)

13. Chakrabarty, A., Hazra, R.S., den Hollander, F., Sfragara, M.: Spectra of adjacency and Laplacian matrices of inhomogeneous Erdős–Rényi random graphs. To appear in Random Matrices: Theory and Applications (2019). https://doi.org/10.1142/S201032632150009X
14. Chapon, F., Couillet, R., Hachem, W., Mestre, X.: The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation. Markov Process. Relat. Fields **20**(2), 183–228 (2014)
15. Couillet, R., Hachem, W.: Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. IEEE Trans. Inform. Theory **59**(1), 509–525 (2013). https://doi.org/10.1109/TIT.2012.2218572
16. Ding, X., Jiang, T., et al.: Spectral distributions of adjacency and Laplacian matrices of random graphs. Ann. Appl. Probab. **20**(6), 2086–2117 (2010)
17. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős–Rényi graphs II: eigenvalue spacing and the extreme eigenvalues. Commun. Math. Phys. **314**(3), 587–640 (2012). https://doi.org/10.1007/s00220-012-1527-7
18. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős–Rényi graphs I: local semicircle law. Ann. Probab. **41**(3B), 2279–2375 (2013). https://doi.org/10.1214/11-AOP734
19. Féral, D., Péché, S.: The largest eigenvalue of rank one deformation of large Wigner matrices. Commun. Math. Phys. **272**(1), 185–228 (2007). https://doi.org/10.1007/s00220-007-0209-3
20. Füredi, Z., Komlós, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**(3), 233–241 (1981). https://doi.org/10.1007/BF02579329
21. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**(301), 13–30 (1963)
22. Huang, J., Landon, B., Yau, H.-T.: Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős—Rényi graphs. Ann. Probab. **48**(2), 916–962 (2020). https://doi.org/10.1214/19-AOP1378
23. Johansson, K.: From Gumbel to Tracy–Widom. Probab. Theory Relat. Fields **138**(1–2), 75–112 (2007). https://doi.org/10.1007/s00440-006-0012-7
24. Knowles, A., Yin, J.: The isotropic semicircle law and deformation of Wigner matrices. Commun. Pure Appl. Math. **66**(11), 1663–1750 (2013). https://doi.org/10.1002/cpa.21450
25. Knowles, A., Yin, J.: The outliers of a deformed Wigner matrix. Ann. Probab. **42**(5), 1980–2031 (2014). https://doi.org/10.1214/13-AOP855
26. Lee, J.O., Schnelli, K.: Extremal eigenvalues and eigenvectors of deformed Wigner matrices. Probab. Theory Relat. Fields **164**(1–2), 165–241 (2016). https://doi.org/10.1007/s00440-014-0610-8
27. Lee, J.O., Schnelli, K.: Local law and Tracy–Widom limit for sparse random matrices. Probab. Theory Relat. Fields **171**(1–2), 543–616 (2018)
28. Ninio, F.: A simple proof of the Perron–Frobenius theorem for positive symmetric matrices. J. Phys. A Math. Gen. **9**(8), 1281 (1976)
29. Péché, S.: Deformed ensembles of random matrices. In: Proceedings of the International Congress of Mathematicians—Seoul 2014, vol. III. pp. 1159–1174. Kyung Moon SA, Seoul (2014).
30. Tiomoko Ali, H., Couillet, R.: Improved spectral community detection in large heterogeneous networks. J. Mach. Learn. Res. **18**, Paper No. 225, 49 (2017)
31. Tran, L.V., Vu, V.H., Wang, K.: Sparse random graphs: eigenvalues and eigenvectors. Random Struct. Algorithms **42**(1), 110–134 (2013). https://doi.org/10.1002/rsa.20406
32. van der Hofstad, R.: Critical behavior in inhomogeneous random graphs. Random Struct. Algorithms **42**(4), 480–508 (2013)
33. Varga, R.S.: Geršgorin and His Circles. Springer Series in Computational Mathematics, vol. 36. Springer, Berlin (2004). https://doi.org/10.1007/978-3-642-17798-9
34. Vu, V.H.: Spectral norm of random matrices. Combinatorica **27**(6), 721–736 (2007). https://doi.org/10.1007/s00493-007-2190-z
35. Zhu, Y.: Graphon approach to limiting spectral distributions of Wigner-type matrices. arXiv.1806.11246 (2018)