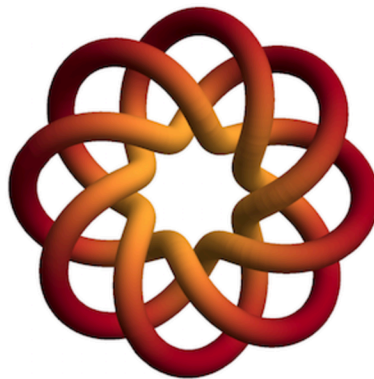


A probabilistic analysis of the Neumann series iteration

Yiting Zhang¹ and Thomas Trogdon²

¹University of California, Irvine

²Department of Applied Mathematics, University of Washington



The Minnesota Journal of Undergraduate Mathematics

Volume 7 (2021-2022 Academic Year)

The Minnesota Journal of Undergraduate Mathematics

Volume 7 (2021-2022 Academic Year)

A probabilistic analysis of the Neumann series iteration

Yiting Zhang^{* 1} and Thomas Trogon²¹University of California, Irvine²Department of Applied Mathematics, University of Washington

1. INTRODUCTION

Neumann series was introduced by Carl Neumann in 1877 in the context of potential theory [12]. Neumann series, or the more advanced Liouville-Neumann series has been applied to solve Fredholm integral equations [15]. In fact, apart from the theoretical applications of Neumann series, it plays an important role in solving computational problems. The Neumann series iteration, $\mathbf{x}^{(k)} = A\mathbf{x}^{(k-1)} + \mathbf{b}$, follows naturally from the actual Neumann series, i.e., $(I - A)^{-1} = \sum_{i=0}^{\infty} A^i$ when solving $(I - A)\mathbf{x} = \mathbf{b}$.

In this paper, we find that if the eigenvalues of an $n \times n$ symmetric matrix A fall between -1 and 1 and if the scaled extreme eigenvalues of A converge in distribution as $n \rightarrow \infty$, then after scaling, a scaled upper bound on the number of iterations needed to solve $(I - A)\mathbf{x} = \mathbf{b}$ with the Neumann series iteration will converge to the reciprocal of the limiting distribution of the largest eigenvalue.

This provides the first step in the full probabilistic analysis of the Neumann series iteration. In particular, our results show that a reasonably sharp upper bound depends only on the (rescaled) extreme eigenvalues as the matrix size tends to infinity. The limiting distributions of these eigenvalues are often universal — they are independent of distributional details of matrix entries¹. Therefore, one expects the convergence rate of the Neumann series to inherit this universality. This phenomenon has been observed in many algorithms [14, 6, 5] and rigorously established for eigenvalue algorithms [4, 3].

This paper unfolds as follows. In Section 2 we introduce the algorithm and the halting criterion. In Section 3 we state the main theorem and give two examples where it applies. The proof of the main theorem is given in Section 4.

¹This can be true, within a class of distributions.

^{*} Corresponding author

2. THE ALGORITHM

2.1. The Neumann series iteration. We first define Neumann series.

Definition 2.1. For $A \in \mathbb{R}^{n \times n}$, the Neumann series is defined formally as:

$$\sum_{i=0}^{\infty} A^i = I + A^1 + A^2 + \dots.$$

According to the above definition, we are interested in sufficient conditions for the Neumann series to converge. The following lemma and theorem provides the key for this study.

Lemma 2.2. *If the spectral norm² of A satisfies $\|A\| < 1$, then $(I - A)^{-1}$ exists, and*

$$(I - A)^{-1} = I + A + A^2 + \dots = \sum_{i=0}^{\infty} A^i.$$

Theorem 2.3. *Given $A \in \mathbb{R}^{n \times n}$ with $\|A\| < 1$ and $\mathbf{b} \in \mathbb{R}^n$, the numerical solution of $(I - A)\mathbf{x} = \mathbf{b}$ is found by applying the Neumann series iteration:*

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{0}, \\ \mathbf{x}_k &= A\mathbf{x}_{k-1} + \mathbf{b} \\ &= (I + A + A^2 + \dots + A^{k-1})\mathbf{b} \\ &= \sum_{i=0}^{k-1} A^i \mathbf{b}, \quad k = 1, 2, 3, \dots \end{aligned}$$

Here \mathbf{x}_k converges to $\mathbf{x} = (I - A)^{-1}\mathbf{b}$ as $k \rightarrow \infty$.

The proof of the above lemma and theorem can be found in [2, p. 457]. The Neumann series iteration is the numerical algorithm we will use throughout this paper.

2.2. Halting criterion. The asymptotic behavior of \mathbf{x}_k is well known by Theorem 2.3, but we are more interested in the non-asymptotic case. Given a halting criterion, we are interested in the number of iterations needed to achieve that criterion. Two natural halting times are defined as follows:

Definition 2.4. Given $\epsilon > 0$, define $k_\epsilon(A, \mathbf{b})$ and $k_\epsilon^*(A, \mathbf{b})$ by

$$\begin{aligned} k_\epsilon(A, \mathbf{b}) &= \min\{k : \|\mathbf{x} - \mathbf{x}_k\| < \epsilon\}, \\ k_\epsilon^*(A, \mathbf{b}) &= \min\{k : \|(I - A)\mathbf{x}_k - \mathbf{b}\| < \epsilon\}. \end{aligned}$$

For simplicity, our results will only concern upper bounds for the halting times:

² $\|\cdot\|$ here denotes the matrix norm induced by the ℓ^2 -norm. In fact, this lemma can be generalized to any sub-multiplicative norm.

Definition 2.5. Given $\epsilon > 0$, we define $K_\epsilon(A, \mathbf{b})$ and $K_\epsilon^*(A, \mathbf{b})$ to be

$$K_\epsilon(A, \mathbf{b}) = \min \left\{ k : \left\| \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\| < \epsilon \right\},$$

$$K_\epsilon^*(A, \mathbf{b}) = \min \left\{ k : \left\| (I - A) \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\| < \epsilon \right\}.$$

Proposition 2.6. $k_\epsilon(A, \mathbf{b}) \leq K_\epsilon(A, \mathbf{b})$ and $k_\epsilon^*(A, \mathbf{b}) \leq K_\epsilon^*(A, \mathbf{b})$.

Proof. For the first inequality, we show that

$$\|\mathbf{x} - \mathbf{x}_k\| \leq \left\| \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|.$$

The exact solution of $(I - A)\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = (I - A)^{-1}\mathbf{b}$. Based on Lemma 2.2, $\mathbf{x} = \sum_{i=0}^{\infty} A^i \mathbf{b}$. By Theorem 2.3, we also know that $\mathbf{x}_k = \sum_{i=0}^{k-1} A^i \mathbf{b}$. Therefore, we have

$$\|\mathbf{x} - \mathbf{x}_k\| = \left\| \sum_{i=0}^{\infty} A^i \mathbf{b} - \sum_{i=0}^{k-1} A^i \mathbf{b} \right\| = \left\| \sum_{i=k}^{\infty} A^i \mathbf{b} \right\| \leq \left\| \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|.$$

Similarly, for the second inequality, we show that

$$\|(I - A)\mathbf{x}_k - \mathbf{b}\| \leq \left\| (I - A) \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|.$$

If \mathbf{x} is the exact solution, we have

$$\begin{aligned} \|(I - A)\mathbf{x}_k - \mathbf{b}\| &= \|(I - A)\mathbf{x}_k - \mathbf{b} - [(I - A)\mathbf{x} - \mathbf{b}]\| \\ &= \|(I - A)(\mathbf{x}_k - \mathbf{x})\| \\ &= \left\| (I - A) \sum_{i=k}^{\infty} A^i \mathbf{b} \right\| \\ &\leq \left\| (I - A) \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|. \end{aligned}$$

Thus, $k_\epsilon(A, \mathbf{b}) \leq K_\epsilon(A, \mathbf{b})$ and $k_\epsilon^*(A, \mathbf{b}) \leq K_\epsilon^*(A, \mathbf{b})$. □

Now, to show that the upper bounds K_ϵ and K_ϵ^* are sharp we give a sufficient condition for equality to hold. Suppose λ is the largest eigenvalue of A , and $0 < \lambda < 1$. If \mathbf{b} is the eigenvector of A which corresponds to λ , we have $k_\epsilon(A, \mathbf{b}) = K_\epsilon(A, \mathbf{b})$. This can be verified by showing that $\|\mathbf{x} - \mathbf{x}_k\| = \left\| \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|$:

$$\|\mathbf{x} - \mathbf{x}_k\| = \left\| \sum_{i=k}^{\infty} A^i \mathbf{b} \right\| = \left\| \sum_{i=k}^{\infty} \lambda^i \mathbf{b} \right\| = \sum_{i=k}^{\infty} \lambda^i \|\mathbf{b}\| = \left\| \sum_{i=k}^{\infty} A^i \right\| \|\mathbf{b}\|.$$

Suppose the largest eigenvalue of $(I - A)$ is μ and $0 < \mu < 1$. If \mathbf{b} is the eigenvector of $(I - A)$ which corresponds to μ , we have $k_\epsilon^*(A, \mathbf{b}) = K_\epsilon^*(A, \mathbf{b})$. The verification is similar as before.

3. RESULTS

In this section we first state the main theorem and then provide two examples where it applies.

3.1. Main theorem.

Definition 3.1. A random variable X_n converges in distribution to X as $n \rightarrow \infty$ if

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) \rightarrow \mathbb{P}(X \leq t) = F_X(t)$$

as $n \rightarrow \infty$ at every t where $F_X(t)$ is continuous. Here $F_{X_n}(t)$ and $F_X(t)$ are the cumulative distribution functions of X_n and X , respectively.

Theorem 3.2. Suppose $(M_n)_{n \geq 1}$, $M_n \in \mathbb{R}^{n \times n}$ (or $\mathbb{C}^{n \times n}$) is a sequence of symmetric (or Hermitian) random matrices with eigenvalues:

$$-1 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n < 1, \quad \lambda_j = \lambda_j(n).$$

Suppose for some $\alpha \geq \beta > 0$, we have

$$n^\alpha(1 - \lambda_n) \rightarrow X \text{ in distribution as } n \rightarrow \infty,$$

$$n^\beta(\lambda_1 + 1) \rightarrow Y \text{ in distribution as } n \rightarrow \infty,$$

where both F_X and F_Y are continuous and supported on $[0, \infty)$. Let \mathbf{b} be a unit vector and fix $0 < \epsilon < 1/2$. Then

$$\frac{K_\epsilon(M_n, \mathbf{b})}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$.

Remark. Because ϵ is fixed, one can replace $\log(n/\epsilon^{1/\alpha})$ with $\log(n)$ in the statement of this theorem, but in our experiments this replacement slows the rate of convergence. For our demonstrations below, we will need to make further heuristic improvements on this theorem to realize a sufficiently fast rate of converge, see Appendix A.

Remark. Note that we can relax the condition on λ_1 . If we suppose that

(1) there exists a $\delta > 0$ such that $\mathbb{P}(\lambda_1 \leq 1 - \delta) \rightarrow 1$ as $n \rightarrow \infty$, and

(2) $\left\{1/n^\beta(\lambda_1 + 1)\right\}$ is a tight family of random variables,

then the conclusion of the theorem will follow. We state the theorem in the restricted case above because we are not aware of a distribution on matrices where this generality would be needed. Furthermore, the conjecture in the next remark would not hold in this more general setting.

Remark. Although Theorem 3.2 is about $K_\epsilon(M_n, \mathbf{b})$, we can state a similar theorem for $K_\epsilon^*(M_n, \mathbf{b})$. Under the setting of Theorem 3.2, for $\alpha \geq \beta > 0$, it is reasonable to conjecture

$$-\frac{K_\epsilon^*(M_n, \mathbf{b})}{\log(\epsilon)n^\alpha} \rightarrow \max\left\{\frac{1}{X}, \frac{1}{Y}\right\}$$

in distribution as $n \rightarrow \infty$.

3.2. Numerical verification.

3.2.1. Independent and identically distributed eigenvalues. Let B be an $n \times n$ matrix with independent and identically distributed standard normal entries. Construct an $n \times n$ matrix A_n by

$$A_n = Q\Lambda Q^T,$$

where Q is found by applying the QR factorization to B and

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where $(\lambda_i)_{i=1}^n$ is a collection of independent and identically distributed (iid) random variables and is uniform on $[-1, 1]$. According to [11], Q is called a Haar orthogonal matrix. With this choice of A_n , it follows that $\|A_n\| < 1$ almost surely. Therefore, the iteration in Theorem 2.3 converges with probability 1. Define

$$\begin{aligned}\lambda_{\max, n} &= \max_{1 \leq i \leq n} \lambda_i, \\ \lambda_{\min, n} &= \min_{1 \leq i \leq n} \lambda_i.\end{aligned}$$

Definition 3.3. Define $\exp(\lambda)$ to be the exponential distribution with parameter λ . The probability density function for a random variable with distribution $\exp(\lambda)$ is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Proposition 3.4. Both $n(1 - \lambda_{\max, n})$ and $n(1 + \lambda_{\min, n})$ converge in distribution to $\exp(1/2)$ as $n \rightarrow \infty$.

Proof. We only show that $n(1 - \lambda_{\max, n}) \rightarrow \exp(1/2)$ in distribution as $n \rightarrow \infty$. The proof that $n(1 + \lambda_{\min, n}) \rightarrow \exp(1/2)$ in distribution as $n \rightarrow \infty$ follows similarly. When $\lambda \geq 0$, we have

$$\begin{aligned}\mathbb{P}(\lambda_{\max, n} \leq \lambda) &= \mathbb{P}(\lambda_1 \leq \lambda, \lambda_2 \leq \lambda, \dots, \lambda_n \leq \lambda) \\ &= \prod_{i=1}^n \mathbb{P}(\lambda_i \leq \lambda) \\ &= \left(\frac{\lambda + 1}{2}\right)^n.\end{aligned}$$

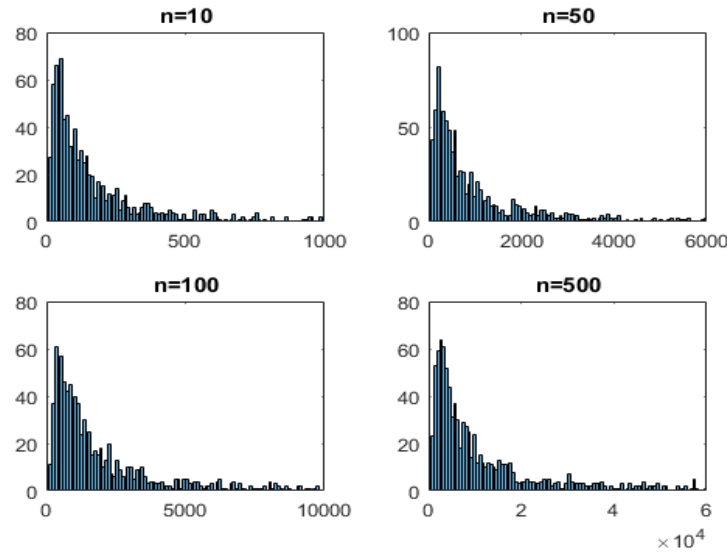


FIGURE 1. Most of the values of $K_\epsilon(A_n, \mathbf{b})$ assemble on the left of the plot. When comparing the abscissas of these four plots, we see that as n becomes larger, the range of the values $K_\epsilon(A_n, \mathbf{b})$ can achieve also becomes larger.

Define $\Lambda_{\max, n} := n(1 - \lambda_{\max, n})$, we have

$$\begin{aligned} \mathbb{P}(\Lambda_{\max, n} \leq \lambda) &= \mathbb{P}(n(1 - \lambda_{\max, n}) \leq \lambda) \\ &= 1 - \mathbb{P}\left(\lambda_{\max, n} < 1 - \frac{\lambda}{n}\right) \\ &= 1 - \left(1 + \frac{-\frac{\lambda}{2}}{n}\right)^n. \end{aligned}$$

Since $e^x = \lim_{n \rightarrow \infty} (1 + x/n)^n$, $\lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{\max, n} \leq \lambda) = 1 - e^{-\lambda/2}$. When $\lambda < 0$, by a similar argument, $\lim_{n \rightarrow \infty} \mathbb{P}(\Lambda_{\max, n} \leq \lambda) = 0$. Therefore, $n(1 - \lambda_{\max, n}) \rightarrow \exp(1/2)$ in distribution as $n \rightarrow \infty$. \square

By Theorem 3.2, with $\alpha = \beta = 1$ and $X, Y \sim \exp(1/2)$, we have

$$\frac{K_\epsilon(A_n, \mathbf{b})}{n \log(n/\epsilon)} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$. Fix $\epsilon = 10^{-3}$, Figure 1 shows the distribution of $K_\epsilon(A_n, \mathbf{b})$ for different values of n . Each plot has 10^3 samples. To verify the main theorem, it is equivalent to see if

$$\frac{n \log(n/\epsilon)}{K_\epsilon(A_n, \mathbf{b})} \rightarrow X$$

in distribution as $n \rightarrow \infty$. Figure 2 shows this convergence. Each plot has 10^3 samples. In fact, if we decompose $(n \log(n/\epsilon)) / K_\epsilon(A_n, \mathbf{b})$, we find that it involves a term that impedes the speed of convergence. Thus, the convergence in Figure 2 is quite slow. However, we can improve the speed of convergence. See Appendix A.

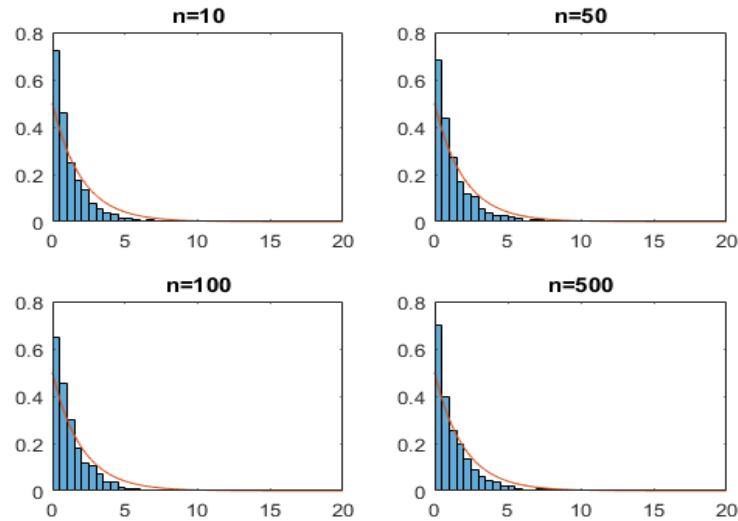


FIGURE 2. The distribution of $(n \log(n/\epsilon)) / K_\epsilon(A_n, \mathbf{b})$ becomes flatter as n becomes larger, which converges to the probability density function for $X \sim \exp(1/2)$. This is a verification of Theorem 3.2.

Remark. Note that if $K_\epsilon(A_n, \mathbf{b})$ is replaced by $k_\epsilon(A_n, \mathbf{b})$, Figure 2 seems to match better for small values of n . However, for $n = 10^3$ or larger, the distribution of the actual number of iterations has a heavier tail than the limiting distribution $X \sim \exp(1/2)$. This also appears to hold for the next example.

3.2.2. *Jacobi unitary ensemble.* The following definition can be found in [8, pg. 111].

Definition 3.5. The Jacobi ensembles are defined as the family of eigenvalue probability density functions proportional to

$$\prod_{j=1}^N (1 - \lambda_j)^{a\beta/2} (1 + \lambda_j)^{b\beta/2} \prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j|^\beta, \quad \lambda_j \in [-1, 1],$$

where λ_j 's are interpreted as eigenvalues, a, b, N are positive integers and $\beta = 1, 2$, or 4 . When $\beta = 2$ these are referred to as the Jacobi unitary ensembles.

Let $A = v^*v, B = w^*w$, where v and w are $n_1 \times n$ and $n_2 \times n$ random matrices with entries that are independent and identically distributed standard complex normal random variables. By Proposition 3.6.1 in [8, pg. 111] and Definition 3.5, we know that the eigenvalues x_1, x_2, \dots, x_n of the matrix $V = (A+B)^{-1/2}A(A+B)^{-1/2}$ have the joint density function proportional to the probability density function presented in Definition 3.5 with

$$N = n, \quad \lambda_j = 1 - 2x_j, \quad a = n_1 - n, \quad b = n_2 - n, \quad \beta = 2.$$

To verify Theorem 3.2, we will therefore focus on $W_n = I - 2V$, where $V = (A+B)^{-1/2}A(A+B)^{-1/2}$. We can express the eigenvalue correlations of W_n near 1 in terms of the Bessel

kernel [10, pg. 1576]

$$\mathbb{J}_\alpha(u, v) = \frac{J_\alpha(\sqrt{u})\sqrt{v}J'_\alpha(\sqrt{v}) - J_\alpha(\sqrt{v})\sqrt{u}J'_\alpha(\sqrt{u})}{2(u-v)},$$

where $u, v \geq 0$ and J_α is the usual Bessel function of the first kind and of order α [13]. Let $\mathbb{J}_{\alpha,s}$ be the integral operator with kernel $\mathbb{J}_\alpha(u, v)$ acting on $L^2(0, s)$. Then by Corollary 1.2 in [10, pg. 1578], for $s > 0$, we have

$$P_n\left(1 - \frac{s}{2n^2}, 1\right) \rightarrow \det(I - \mathbb{J}_{\alpha,s})$$

as $n \rightarrow \infty$, where $P_n(a, b)$ is the probability that there are no eigenvalues in the interval $(a, b) \subset (-1, 1)$ and $\det(I - \mathbb{J}_{\alpha,s})$ is the Fredholm determinant (see, for example, [1]). By definition, we have

$$P_n\left(1 - \frac{s}{2n^2}, 1\right) = \mathbb{P}\left(\lambda_n \leq 1 - \frac{s}{2n^2}\right) = \mathbb{P}\left(n^2(1 - \lambda_n) \geq \frac{s}{2}\right).$$

Therefore,

$$\mathbb{P}\left(n^2(1 - \lambda_n) \leq \frac{s}{2}\right) \rightarrow 1 - \det(I - \mathbb{J}_{\alpha,s})$$

as $n \rightarrow \infty$. Let $t = s/2$, we can rewrite it as

$$\mathbb{P}\left(n^2(1 - \lambda_n) \leq t\right) \rightarrow 1 - \det(I - \mathbb{J}_{\alpha,2t})$$

as $n \rightarrow \infty$. Similarly, we also have

$$\mathbb{P}\left(n^2(\lambda_1 + 1) \leq t\right) \rightarrow 1 - \det(I - \mathbb{J}_{\alpha,2t})$$

as $n \rightarrow \infty$. Therefore, the assumptions of Theorem 3.2 are satisfied with $\alpha = \beta = 2$. Let $n_1 = n_2 = n + 2$. Figure 3 shows the global eigenvalue distribution of W_n for different values of n . Each plot has 10^3 samples. Consider a quadrature rule [9, pg. 175]

$$\int_{-1}^1 f(x) dx \approx \sum_{j=1}^n f(x_j)w_j,$$

where w_j 's are discrete weights and x_j 's are the nodes. By a linear transformation, we find

$$\int_0^{2t} f(x) dx \approx \sum_{j=1}^n f(t(1 + x_j))(tw_j). \quad (1)$$

We use Gauss-Legendre quadrature [9]. We calculate the Bessel kernel $\mathbb{J}_\alpha(u, v)$ by applying

$$J'_\alpha(x) = J_{\alpha-1}(x) - \frac{\alpha}{x}J_\alpha(x)$$

and

$$J''_\alpha(x) = J'_{\alpha-1}(x) - \frac{\alpha}{x}J'_\alpha(x) + \frac{\alpha}{x^2}J_\alpha(x)$$

where the last formula is required when $u = v$. Then using the algorithm for calculating the Fredholm determinant from [1, pg. 874] along with (1), we evaluate the Fredholm determinant and compute the cumulative distribution function $1 - \det(I - \mathbb{J}_{\alpha,2t})$.

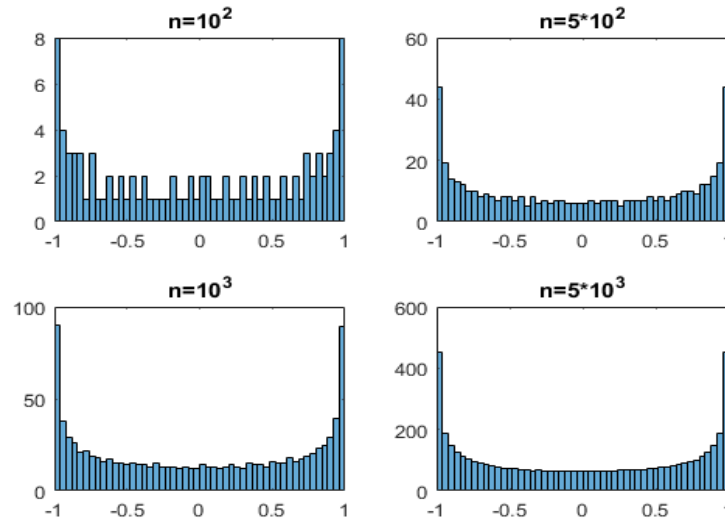


FIGURE 3. We can see that most of the eigenvalues assemble near -1 and 1 and within a small neighbourhood around 0 , the distribution of the eigenvalues is almost uniform.

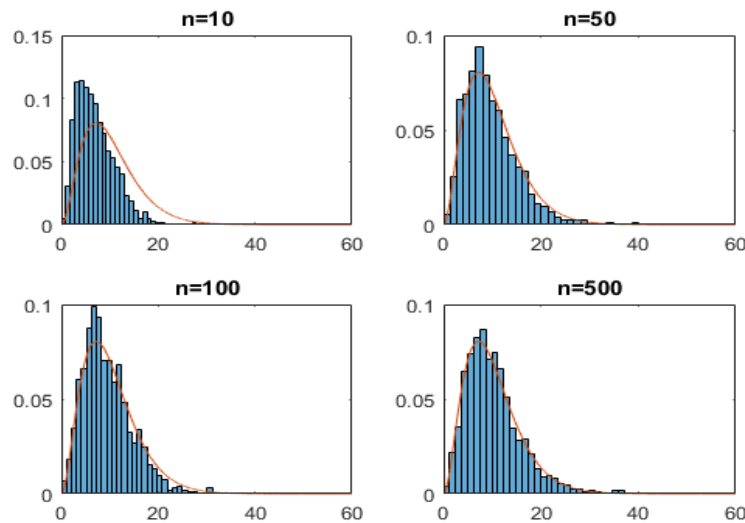


FIGURE 4. As n becomes larger, the distribution of $n^2(1 - \lambda_n)$ converges to the probability density function of $1 - \det(I - \mathbb{J}_{\alpha, 2t})$. This is a verification of Corollary 1.2 in [10, pg. 1578].

Figure 4 shows the distribution of $n^2(1 - \lambda_n)$ for different values of n and the probability density function of $1 - \det(I - \mathbb{J}_{\alpha, 2t})$, found using a central difference. Each plot has 10^3 samples. The case for $n^2(\lambda_1 + 1)$ is similar. Now, we are ready to plot the distribution of $K_\epsilon(W_n, \mathbf{b})$. Fix $\epsilon = 10^{-3}$, Figure 5 shows the distribution of $K_\epsilon(W_n, \mathbf{b})$ for different values of n . Each plot has 10^3 samples. Like the previous example, we want to see if

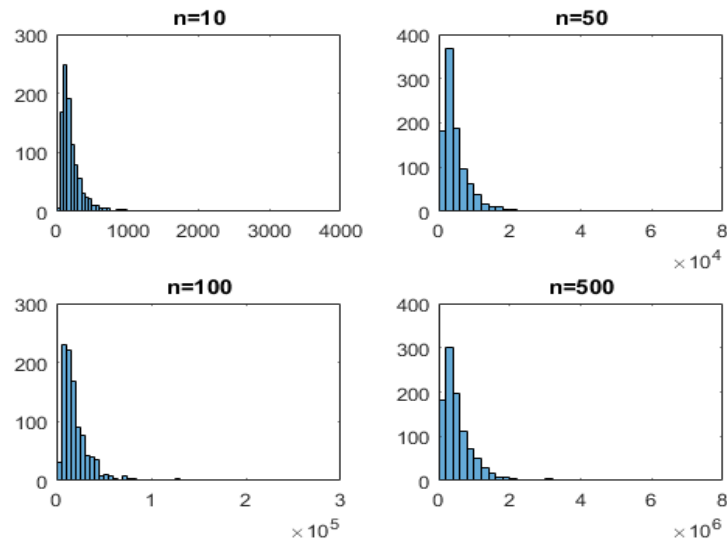


FIGURE 5. Like Figure 1, the values of $K_\epsilon(W_n, \mathbf{b})$ assemble on the left of the plot and as n becomes larger, the range of the values $K_\epsilon(W_n, \mathbf{b})$ can achieve becomes larger.

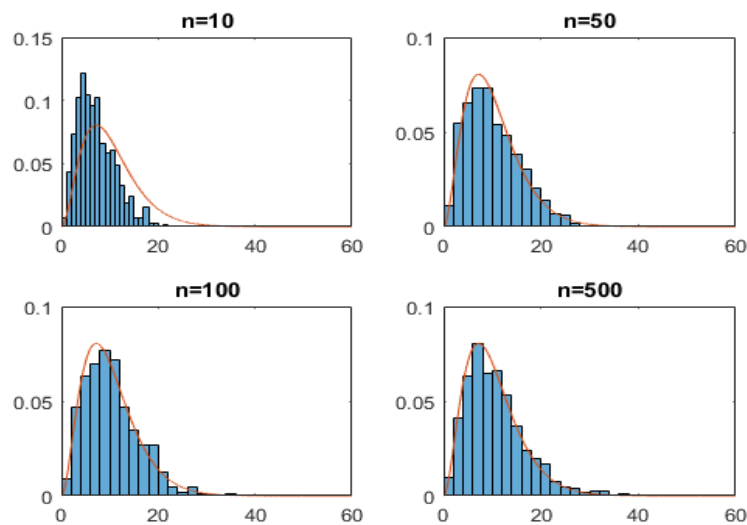


FIGURE 6. As n increases, the distribution of $\left(2n^2 \log(n/\sqrt{\epsilon})\right) / K_\epsilon(W_n, \mathbf{b})$ converges to the probability density function of $1 - \det(I - \mathbb{J}_{\alpha, 2t})$.

$\left(2n^2 \log(n/\sqrt{\epsilon})\right) / K_\epsilon(W_n, \mathbf{b})$ converges in distribution as $n \rightarrow \infty$. Figure 6 shows this convergence. Each plot has 10^3 samples.

4. LEMMAS AND THE PROOF OF THEOREM 3.2

In this section, we will first prove some lemmas and then prove the main theorem based on these lemma.

Lemma 4.1. Suppose M_n is a $n \times n$ symmetric matrix with eigenvalues

$$-1 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n < 1.$$

Then

$$\left\| (I - M_n)^{-1} - \sum_{i=0}^k M_n^i \right\| = \max \left\{ \frac{|\lambda_1|^k}{|1 - \lambda_1|}, \frac{|\lambda_n|^k}{|1 - \lambda_n|} \right\}.$$

Proof. By Lemma 2.2,

$$\left\| (I - M_n)^{-1} - \sum_{i=0}^k M_n^i \right\| = \left\| \sum_{i=k}^{\infty} M_n^i \right\| = \left\| M_n^k \sum_{i=0}^{\infty} M_n^i \right\|.$$

We decompose M_n as $M_n = U \Lambda U^T$, where Λ is a diagonal matrix formed from the eigenvalues of M_n and U is a unitary matrix. Thus,

$$\left\| M_n^k \sum_{i=0}^{\infty} M_n^i \right\| = \left\| \begin{bmatrix} \frac{\lambda_1^k}{1 - \lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\lambda_n^k}{1 - \lambda_n} \end{bmatrix} \right\| = \max \left\{ \frac{|\lambda_1|^k}{|1 - \lambda_1|}, \dots, \frac{|\lambda_n|^k}{|1 - \lambda_n|} \right\}.$$

Suppose $-1 < \lambda_1 \leq \dots \leq \lambda_{m-1} \leq 0 \leq \lambda_m \leq \dots \leq \lambda_n < 1$. Consider λ_i , where $i = m, \dots, n$. Since such λ_i 's are all positive and less than 1, we have $|\lambda_i|^k / |1 - \lambda_i| = \lambda_i^k / (1 - \lambda_i)$. Note that $\lambda_i^k / (1 - \lambda_i)$ is a strictly increasing function of $\lambda_i \in (0, 1)$ for any positive integer k , we have

$$\max \left\{ \frac{|\lambda_m|^k}{|1 - \lambda_m|}, \dots, \frac{|\lambda_n|^k}{|1 - \lambda_n|} \right\} = \frac{|\lambda_n|^k}{|1 - \lambda_n|}.$$

Now, consider λ_i , where $i = 1, \dots, m-1$. We want to show that $|\lambda_i|^k / |1 - \lambda_i|$ is a strictly decreasing function of $\lambda_i \in (-1, 0)$. Note that it is equivalent to show that $\lambda_i^k / (1 + \lambda_i)$ is a strictly increasing function of $\lambda_i \in (0, 1)$. Let $g(\lambda_i) = \lambda_i^k / (1 + \lambda_i)$ and compute

$$\begin{aligned} g'(\lambda_i) &= \frac{k\lambda_i^{k-1}(1 + \lambda_i) - \lambda_i^k}{(1 + \lambda_i)^2} \\ &= \frac{k\lambda_i^k - \lambda_i^k + k\lambda_i^{k-1}}{(1 + \lambda_i)^2} > 0. \end{aligned}$$

Therefore, $|\lambda_i|^k / |1 - \lambda_i|$ is a strictly decreasing function of $\lambda_i \in (-1, 0)$, and we have

$$\max \left\{ \frac{|\lambda_1|^k}{|1 - \lambda_1|}, \dots, \frac{|\lambda_{m-1}|^k}{|1 - \lambda_{m-1}|} \right\} = \frac{|\lambda_1|^k}{|1 - \lambda_1|}.$$

Therefore, we conclude

$$\left\| (I - M_n)^{-1} - \sum_{i=0}^k M_n^i \right\| = \max \left\{ \frac{|\lambda_1|^k}{|1 - \lambda_1|}, \frac{|\lambda_n|^k}{|1 - \lambda_n|} \right\}.$$

□

Lemma 4.2. Let³ $f, g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be strictly decreasing continuous functions such that their graphs intersect at most once. Define

$$h(x) = \max \{f(x), g(x)\}.$$

Then

$$h^{-1}(y) = \max \{f^{-1}(y), g^{-1}(y)\},$$

$$\min_{x \in \mathbb{R}^+} \max \{f(x), g(x)\} \leq y \leq \max_{x \in \mathbb{R}^+} \max \{f(x), g(x)\}.$$

Proof. Consider the case where their graphs intersect once. Without loss of generality, assume $f(x^*) = g(x^*)$, $x^* > 0$, and when $x \leq x^*$, $g(x) \geq f(x)$; when $x \geq x^*$, $f(x) \geq g(x)$. Therefore, by the definition of $h(x)$, we have

$$h(x) = \begin{cases} g(x), & x \leq x^* \\ f(x), & x \geq x^* \end{cases}$$

As a result, we get

$$h^{-1}(y) = \begin{cases} g^{-1}(y), & y \geq f(x^*) \\ f^{-1}(y), & y \leq f(x^*) \end{cases}$$

Since both f, g are strictly decreasing functions, f^{-1}, g^{-1} are also strictly decreasing functions. If $x \leq x^*$, then $f(x) \leq g(x)$. Therefore, $g^{-1}(f(x)) \geq x = f^{-1}(f(x))$. In other words, if $y = f(x)$ and $x \leq x^*$, we have $y = f(x) \geq f(x^*)$ and $f^{-1}(y) \leq g^{-1}(y)$. Similarly, if $x \geq x^*$, then $f(x) \geq g(x)$. Therefore, $g^{-1}(f(x)) \leq x = f^{-1}(f(x))$. In other words, if $y = f(x)$ and $x \geq x^*$, we have $y = f(x) \leq f(x^*)$ and $g^{-1}(y) \leq f^{-1}(y)$. Thus, we have

$$h^{-1}(y) = \max \{f^{-1}(y), g^{-1}(y)\}.$$

If their graphs do not intersect, without loss of generality, assume $f(x) > g(x)$, $x > 0$. Then

$$h(x) = f(x) = \max \{f(x), g(x)\}.$$

Therefore, $f^{-1}(g(x)) > x = g^{-1}(g(x))$. In other words, if $y = g(x)$, we have $f^{-1}(y) > g^{-1}(y)$. Thus,

$$h^{-1}(y) = f^{-1}(y) = \max \{f^{-1}(y), g^{-1}(y)\}.$$

□

Based on Lemma 4.1, to find an expression for K_ϵ , set both $|\lambda_1|^k/|1 - \lambda_1|$ and $|\lambda_n|^k/|1 - \lambda_n|$ equal to ϵ .

³ $\mathbb{R}_+ = [0, \infty)$.

Definition 4.3. Define

$$k_1 = \frac{\log \epsilon + \log |1 - \lambda_1|}{\log |\lambda_1|},$$

$$k_n = \frac{\log \epsilon + \log |1 - \lambda_n|}{\log |\lambda_n|}.$$

Based on Lemma 4.2,

$$K_\epsilon(M_n) = \max \left\{ \frac{\log \epsilon + \log |1 - \lambda_1|}{\log |\lambda_1|}, \frac{\log \epsilon + \log |1 - \lambda_n|}{\log |\lambda_n|} \right\} + \sigma$$

$$= \max \{k_1, k_n\} + \sigma,$$

where⁴ $\sigma = \lceil \max \{k_1, k_n\} \rceil - \max \{k_1, k_n\}$.

Definition 4.4. A sequence $(X_n)_{n \geq 0}$ of random variables converge to zero in probability if for every $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \epsilon) = 0.$$

Lemma 4.5. Suppose a sequence $(X_n)_{n \geq 0}$ of random variables converge in distribution to a random variable X . Suppose further that another sequence $(Y_n)_{n \geq 0}$ of random variables converge in probability to 0. Then $X_n Y_n$ converges in probability to 0.

Proof. We want to show that $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n Y_n| > \epsilon) = 0$ for every $\epsilon > 0$. The cumulative distribution function of X has an, at most, countable number of discontinuities. We can choose $(\delta_k)_{k \geq 1}$, where $\delta_k \in (0, 1/k)$ and $k \geq 1$ as the sequence converging to 0 such that for all k , ϵ/δ_k is a point of continuity for $F_{|X|}$. This is possible since X/ϵ has an, at most, countable number of discontinuities. We have

$$\mathbb{P}(|X_n Y_n| > \epsilon) = \mathbb{P}(|X_n Y_n| > \epsilon, |Y_n| \leq \delta_k) + \mathbb{P}(|X_n Y_n| > \epsilon, |Y_n| > \delta_k).$$

Given that $|X_n Y_n| > \epsilon$ and $|Y_n| \leq \delta_k$, we have $|X_n| > \epsilon/\delta_k$. Since $\mathbb{P}(|X_n Y_n| > \epsilon, |Y_n| > \delta_k) \leq \mathbb{P}(|Y_n| > \delta_k)$, we have

$$\mathbb{P}(|X_n Y_n| > \epsilon) \leq \mathbb{P}\left(|X_n| > \frac{\epsilon}{\delta_k}\right) + \mathbb{P}(|Y_n| > \delta_k).$$

Since ϵ/δ_k is a point of continuity we know that $\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n| > \delta_k) = 0$, we have for all k

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n Y_n| > \epsilon) \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(|X_n| > \frac{\epsilon}{\delta_k}\right) = \mathbb{P}\left(|X| > \frac{\epsilon}{\delta_k}\right).$$

By taking $\delta_k \rightarrow 0^+$ along $(\delta_k)_{k \geq 1}$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n Y_n| > \epsilon) = 0.$$

□

The following lemma is from [7, pg. 105].

⁴ $\lceil \cdot \rceil$ here denotes the ceiling function.

Lemma 4.6 (Converging together lemma). *Suppose a sequence $(X_n)_{n \geq 0}$ of random variables converge in distribution to a random variable X as $n \rightarrow \infty$. Suppose further that there is another sequence $(Y_n)_{n \geq 0}$ of random variables such that $Y_n - X_n$ converges to zero in probability as $n \rightarrow \infty$. Then Y_n converges to X in distribution as $n \rightarrow \infty$.*

Proof. Let F_{X_n} be the cumulative distribution function of X_n and F_X the cumulative distribution function of X . Let x be a continuity point of F_X and $\epsilon > 0$. For the upper bound on Y_n ,

$$\begin{aligned}\mathbb{P}(Y_n \leq x) &= \mathbb{P}(Y_n \leq x, |Y_n - X_n| \leq \epsilon) + \mathbb{P}(Y_n \leq x, |Y_n - X_n| > \epsilon) \\ &\leq \mathbb{P}(X_n \leq x + \epsilon) + \mathbb{P}(|Y_n - X_n| > \epsilon).\end{aligned}$$

Since $\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - X_n| > \epsilon) = 0$ and if $x + \epsilon$ is a continuity point of F_X ,

$$\mathbb{P}(X_n \leq x + \epsilon) = F_{X_n}(x + \epsilon) \rightarrow F_X(x + \epsilon)$$

in distribution as $n \rightarrow \infty$. Therefore,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n \leq x) \leq F_X(x + \epsilon)$$

for all ϵ such that $x + \epsilon$ is a continuity point of F_X . Such an ϵ exists since F_X has an, at most, countable number of discontinuities. And indeed there must exist a sequence of choices of $\epsilon > 0$ such that $\epsilon \rightarrow 0^+$ along this sequence. Take $\epsilon \rightarrow 0^+$ along this sequence,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n \leq x) \leq F_X(x).$$

For the lower bound on Y_n ,

$$\begin{aligned}\mathbb{P}(X_n \leq x - \epsilon) &= \mathbb{P}(X_n \leq x - \epsilon, |Y_n - X_n| \leq \epsilon) + \mathbb{P}(X_n \leq x - \epsilon, |Y_n - X_n| > \epsilon) \\ &\leq \mathbb{P}(Y_n \leq \epsilon) + \mathbb{P}(|Y_n - X_n| > \epsilon).\end{aligned}$$

If $x - \epsilon$ is a continuity point of F_X ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(Y_n \leq \epsilon) \geq F_X(x - \epsilon).$$

for all ϵ such that $x - \epsilon$ is a continuity point of F_X . Such ϵ exists since that F_X has an, at most, countable number of discontinuities. And indeed there must exist a sequence of choices of $\epsilon > 0$ such that $\epsilon \rightarrow 0^+$ along this sequence. Again, take $\epsilon \rightarrow 0^+$ along this sequence,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(Y_n \leq \epsilon) \geq F_X(x).$$

These two bounds imply

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \leq x) = F_X(x).$$

□

Lemma 4.7. *Suppose a sequence $(X_n)_{n \geq 0}$ of random variables converge in distribution to a random variable X as $n \rightarrow \infty$ and $X, X_n > 0$ almost surely. Set $\rho_n = 1 - X_n/n^\alpha$ and $\alpha > 0$. Then*

$$\frac{1}{n^\alpha \log |\rho_n|}$$

converges in distribution to $-1/X$ as $n \rightarrow \infty$.

Proof. By Lemma 4.6, it is equivalent to show that $1/n^\alpha \log |\rho_n|$ converges in probability to $-1/X_n$. Let

$$Y_n = \left| \frac{1}{n^\alpha \log |\rho_n|} + \frac{1}{X_n} \right|,$$

then we want to show that $\lim_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon) = 0$ for every $\epsilon > 0$. Let

$$A_n = \left\{ |1 - \rho_n| \leq \frac{1}{2} \right\} = \left\{ \left| \frac{X_n}{n^\alpha} \right| \leq \frac{1}{2} \right\},$$

then by Lemma 4.5, $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1$. Therefore,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon) &\leq \limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n) + \limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n^c) \\ &= \limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n) \end{aligned}$$

since $\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n^c) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_n^c) = 0$. Thus, we want to show that $\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n) = 0$. Let $-1/2 \leq x = X_n/n^\alpha \leq 1/2$, then $\log|1 - x| = \log(1 - x)$. By Taylor's Theorem,

$$\log(1 - x) = -x - \frac{1}{(1 - \xi)^2} x^2,$$

where $\xi \in [-\frac{1}{2}, \frac{1}{2}]$. Since $f(\xi) = 1/(1 - \xi)^2$ is a strictly increasing function of ξ , we have

$$|\log(1 - x) + x| \leq 4x^2.$$

Using $x = X_n/n^\alpha$ and multiplying both sides by n^α , we find

$$|n^\alpha \log \rho_n + X_n| \leq 4 \left(\frac{X_n^2}{n^\alpha} \right).$$

When $1/2 < x \leq 3/2$, $|\log x| \geq |x - 1|/2$. Then $|\log \rho_n| \geq |\rho_n - 1|/2$ on A_n since $1/2 \leq \rho_n \leq 3/2$, and we have

$$|n^\alpha (\log \rho_n) X_n| \geq \frac{X_n^2}{2}.$$

Therefore on A_n ,

$$Y_n = \left| \frac{1}{n^\alpha \log |\rho_n|} + \frac{1}{X_n} \right| = \left| \frac{n^\alpha \log \rho_n + X_n}{n^\alpha (\log \rho_n) X_n} \right| \leq \frac{4X_n^2/n^\alpha}{X_n^2/2} = \frac{8}{n^\alpha},$$

and

$$\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n > \epsilon, A_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(\frac{8}{n^\alpha} > \epsilon, A_n\right) = 0.$$

Thus, we conclude that

$$\frac{1}{n^\alpha \log |\rho_n|} \rightarrow -\frac{1}{X}$$

in distribution as $n \rightarrow \infty$. □

Proof of Theorem 3.2. Recall Definition 4.3:

$$k_n = \frac{\log \epsilon + \log |1 - \lambda_n|}{\log |\lambda_n|}.$$

Given that $\lambda_n = 1 - \xi_n/n^\alpha$, where $\xi_n > 0$ and $\xi_n \rightarrow X$ in distribution as $n \rightarrow \infty$, we have

$$k_n = \frac{-\alpha \log n + \log \epsilon \xi_n}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} = \frac{-\alpha \log(n/\epsilon^{1/\alpha})}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} + \frac{\log \xi_n}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|}.$$

Let

$$\widetilde{k}_n = \frac{-\alpha \log(n/\epsilon^{1/\alpha})}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|}.$$

By Lemma 4.7, we know that

$$\frac{\widetilde{k}_n}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} = -\frac{1}{n^\alpha \log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$. Moreover, by Lemma 4.5,

$$\frac{\log \xi_n}{\alpha \log(n/\epsilon^{1/\alpha})} \rightarrow 0$$

in probability as $n \rightarrow \infty$, and therefore,

$$\frac{\widetilde{k}_n - k_n}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} = \frac{\log \xi_n}{\alpha \log(n/\epsilon^{1/\alpha})} \cdot \left(-\frac{1}{n^\alpha \log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} \right) \rightarrow 0$$

in probability as $n \rightarrow \infty$ by Lemma 4.5. Finally, by Lemma 4.6, we find

$$\frac{k_n}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$. Similarly, in Definition 4.3, recall

$$k_1 = \frac{\log \epsilon + \log |1 - \lambda_1|}{\log |\lambda_1|}.$$

Given that $\lambda_1 = -1 + \xi_1/n^\beta$, where $\xi_1 > 0$ and $\xi_1 \rightarrow Y$ in distribution as $n \rightarrow \infty$, we write

$$k_1 = \frac{\log \left(2 - \frac{\xi_1}{n^\beta} \right) \epsilon}{\log \left| 1 - \frac{\xi_1}{n^\beta} \right|}.$$

Let $\zeta_1 = 2 - \xi_1/n^\beta$, then $\zeta_1 \rightarrow 2$ in probability. Thus, by Lemma 4.5, we have

$$\frac{\log \zeta_1 \epsilon}{\log(n/\epsilon^{1/\beta})} \rightarrow 0$$

in probability as $n \rightarrow \infty$. By Lemma 4.7, we have

$$\frac{1}{n^\beta \log \left| 1 - \frac{\xi_n}{n^\beta} \right|} \rightarrow -\frac{1}{X}$$

in distribution as $n \rightarrow \infty$. Therefore,

$$\begin{aligned} \frac{k_1}{\log(n/\epsilon^{1/\beta})n^\beta} &= \frac{\log \zeta_1 \epsilon}{\log(n/\epsilon^{1/\beta})n^\beta \log \left| 1 - \frac{\xi_n}{n^\beta} \right|} \\ &= \left(\frac{\log \zeta_1 \epsilon}{\log(n/\epsilon^{1/\beta})} \right) \cdot \left(\frac{1}{n^\beta \log \left| 1 - \frac{\xi_n}{n^\beta} \right|} \right) \rightarrow 0 \end{aligned}$$

in probability as $n \rightarrow \infty$ by Lemma 4.5. Given that $\alpha \geq \beta > 0$, we have

$$\frac{K_\epsilon(M_n)}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} = \max \left\{ \frac{k_1}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha}, \frac{k_n}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} \right\} + \frac{\sigma}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha}.$$

Let $1/X_n := k_n/\alpha \log(n/\epsilon^{1/\alpha})n^\alpha \rightarrow 1/X$ in distribution as $n \rightarrow \infty$ and $Y_n := k_1/\alpha \log(n/\epsilon^{1/\alpha})n^\alpha \rightarrow 0$ in probability as $n \rightarrow \infty$. Fix $0 < \epsilon < 1/2$ and define

$$A_{n,\epsilon} = \{Y_n > \epsilon\}.$$

We have

$$\mathbb{P}\left(Y_n > \frac{1}{X_n}\right) = \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}\right) + \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}^c\right).$$

Applying $\limsup_{n \rightarrow \infty}$ we find

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}\right) \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}\right) + \limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}^c\right).$$

Then

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}\right) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_{n,\epsilon}) = 0,$$

and since $1/X_n \rightarrow 1/X$ in distribution as $n \rightarrow \infty$,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}\right) &\leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}, A_{n,\epsilon}^c\right) \\ &= \limsup_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{X_n} < \epsilon\right) \\ &\leq F_{1/X}(\epsilon). \end{aligned}$$

Therefore, applying $\lim_{\epsilon \rightarrow 0^+}$ on both sides, we get

$$\lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(Y_n > \frac{1}{X_n}\right) \leq \lim_{\epsilon \rightarrow 0^+} F_{1/X}(\epsilon) = 0$$

since $1/X > 0$ and $\lim_{\epsilon \rightarrow 0^+} F_{1/X}(\epsilon) = \lim_{\epsilon \rightarrow 0^+} \mathbb{P}(1/X \leq \epsilon) = 0$. Let $Z_n = K_\epsilon(M_n) / (\alpha \log(n/\epsilon^{1/\alpha})n^\alpha)$ and

$$M_n = \left\{ Y_n > \frac{1}{X_n} \right\},$$

then we have

$$\mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon\right) = \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n\right) + \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n^c\right).$$

Applying $\lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty}$ on both sides, we get

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon\right) &\leq \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n\right) \\ &\quad + \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n^c\right). \end{aligned}$$

Since

$$\lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n\right) \leq \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}(M_n) = 0,$$

we get

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon\right) &\leq \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{X_n} - Z_n\right| \leq \epsilon, M_n^c\right) \\ &\leq \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \mathbb{P}(M_n^c) \\ &= 1. \end{aligned}$$

Thus, Z_n and $1/X_n$ have the same limiting distribution. In other words,

$$\frac{K_\epsilon(M_n)}{\alpha \log(n/\epsilon^{1/\alpha})n^\alpha} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$. □

5. ACKNOWLEDGEMENTS

This work was supported in part by NSF DMS-1753185, NSF DMS-1945652

APPENDIX A. IMPROVEMENT ON SPEED OF CONVERGENCE

Following the definitions of the proof of Theorem 3.2, we know that

$$K_\epsilon(A, \mathbf{b}) \Big/ \left(\alpha \log(n/\epsilon^{1/\alpha}) n^\alpha \right) \quad \text{and} \quad k_n \Big/ \left(\alpha \log(n/\epsilon^{1/\alpha}) n^\alpha \right)$$

have the same limiting distribution and recall

$$k_n = \frac{-\alpha \log n + \log \epsilon \xi_n}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} = \frac{-\alpha \log(n/\epsilon^{1/\alpha})}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} + \frac{\log \xi_n}{\log \left| 1 - \frac{\xi_n}{n^\alpha} \right|},$$

where $\epsilon = 10^{-3}$ and $\xi_n \rightarrow \exp(1/2)$ in distribution as $n \rightarrow \infty$ for the example in Section 3.2.1. By Lemma 4.7,

$$-\frac{1}{n^\alpha \log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} \rightarrow \frac{1}{X}$$

in distribution as $n \rightarrow \infty$. Therefore, divide k_n by $n^\alpha \log(n/\epsilon^{1/\alpha})$ and factor out the term $-1/n^\alpha \log \left| 1 - \frac{\xi_n}{n^\alpha} \right|$ giving

$$\frac{k_n}{n^\alpha \log(n/\epsilon^{1/\alpha})} = -\frac{1}{n^\alpha \log \left| 1 - \frac{\xi_n}{n^\alpha} \right|} \left(\alpha - \frac{\log \xi_n}{\log(n/\epsilon^{1/\alpha})} \right).$$

We can treat $\left(\alpha - \log \xi_n / \log(n/\epsilon^{1/\alpha}) \right)$ as a correction term and replace $\log(\xi_n)$ with its expectation. To get a faster convergence, we move $\left(\alpha - \mathbb{E}[\log \xi_n] / \log(n/\epsilon^{1/\alpha}) \right)$ to the left hand side and take the reciprocal to find

$$Z = \frac{\alpha \log(n/\epsilon^{1/\alpha}) - \mathbb{E}[\log(\xi_n)]}{\log(n/\epsilon^{1/\alpha})} \cdot \frac{n^\alpha \log(n/\epsilon^{1/\alpha})}{k_n}.$$

For the example in Section 3.2.1, we have

$$Z_1 = \frac{\log(n/\epsilon) - \mathbb{E}[\log(\xi_n)]}{\log(n/\epsilon)} \cdot \frac{n \log(n/\epsilon)}{k_n}.$$

Figure 7 shows the refinement. Each plot has 10^3 samples.

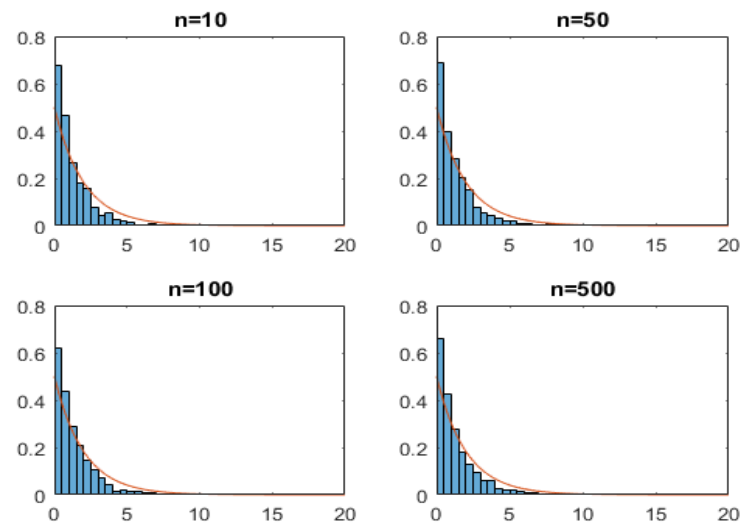


FIGURE 7. The distribution of Z_1 converges to the probability density function of $X \sim \exp(1/2)$ as n becomes larger. The figure is much like Figure 2 but with a faster speed of convergence.

REFERENCES

- [1] Folkmar Bornemann. On the numerical evaluation of fredholm determinants. *Mathematics of Computation*, 79:871–915, 04 2008.
- [2] Richard L. Burden and John Douglas Faires. *Numerical Analysis 9th Edition*. Cengage Learning, 2011.
- [3] P Deift and T Trogdon. Universality for eigenvalue algorithms on sample covariance matrices. *SIAM J. Numer. Anal.*, page (to appear), jan 2017.
- [4] P Deift and T Trogdon. Universality for the Toda Algorithm to Compute the Largest Eigenvalue of a Random Matrix. *Communications on Pure and Applied Mathematics*, 71(3):505–536, mar 2018.
- [5] P Deift and T Trogdon. Universality in numerical computation with random data: Case Studies, Analytical Results and Some Speculations. In *Abel Symposia*, volume 13, pages 221–231, mar 2018.
- [6] P A Deift, G Menon, S Olver, and T Trogdon. Universality in numerical computations with random data. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42):14973–8, oct 2014.
- [7] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4 edition, 2010.
- [8] P.J. Forrester. *Log-Gases and Random Matrices (LMS-34)*. London Mathematical Society Monographs. Princeton University Press, 2010.
- [9] Walter Gautschi. *Numerical Analysis*. Birkhäuser Basel, 2011.
- [10] Arno Kuijlaars and Maarten Vanlessen. Universality for eigenvalue correlations from the modified jacobi unitary ensemble. *Internat. Math. Research Notices*, 243:1575–1600, 04 2002.
- [11] Francesco Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society*, 54:592–604, 10 2006.
- [12] C. Neumann and I. Newton. *Untersuchungen Über Das Logarithmische und Newton'sche Potential*. B. G. Teubner, 1877.
- [13] F W J Olver, D W Lozier, R F Boisvert, and C W Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.
- [14] C W Pfrang, P Deift, and G Menon. How long does it take to compute the eigenvalues of a random symmetric matrix? *Random matrix theory, interacting particle systems, and integrable systems, MSRI Publications*, 65:411–442, 2014.
- [15] F.G. Tricomi. *Integral Equations*. (Pure and applied mathematics, v. 5). Dover Publications, 1985.

STUDENT BIOGRAPHIES

Yiting Zhang: (*Corresponding author: zjx687691@gmail.com*) Yiting Zhang graduated from the University of California, Irvine in 2019 with a B.S. in Applied Mathematics. He is currently pursuing a PhD degree in Applied Mathematics at the University of Washington. His research interests include random matrix theory and numerical linear algebra, especially how different numerical algorithms affect the results of certain linear algebra problems involving random matrices.