Fluctuations of Eigenvalues and Concentration Inequalities for Patterned Random Matrices

Fred Rajasekaran

Advisor: Professor Todd Kemp

Department of Mathematics

University of California San Diego

In partial fulfillment of the requirements

for the Mathematics Honors Program

June 2023

Acknowledgements

I would like to thank Professor Todd Kemp for teaching me random matrix theory and for providing helpful advice throughout this project. I would also like to thank Professor Andrej Zlatoš for teaching me analysis, and I would further like to thank him and Thomas Grubb for getting me started on my mathematics research journey. In addition, I want to express my gratitude towards Professor Joel Tropp for providing helpful advice related to the proof methods in Chapter 3.

Contents

1 Introduction and Background **1**

1.1 What is a Random Matrix? 1

1.2 Wigner Matrices and the Semicircle Law 2

1.3 Some Other Types of Patterned Random Matrices 5

1.4 Non-Asymptotic Random Matrix Theory 6

1.5 Outline of Main Results 9

2 Fluctuations of Eigenvalues for Random Matrices with Correlated Entries **10**

2.1 Introduction 10

2.1.1 Definitions of Other Correlated Matrix Models 12

2.2 Proof of Theorem 1.2 for Toeplitz Models: Outline and First Steps 15

2.3 Bounding the Spectral Norm of a Gaussian Toeplitz Matrix with Correlated Entries 18

2.4 Fluctuations for Other Matrix Models with Correlated Entries 23

2.4.1 Reverse Circulant Matrices with Correlated Entries 24

2.4.2 Circulant Matrices with Correlated Entries 27

2.4.3 Symmetric Circulant Matrices with Correlated Entries 28

2.4.4 Hankel Matrices with Correlated Entries 28

2.5 Fluctuations of Eigenvalues with Correlation Decay 29

2.6 Towards Universality 31

2.6.1 Removing the Conditions on the Covariances 31

2.6.2 Moving Beyond Gaussian Entries 32

3 Matrix Concentration Inequalities with Sub-Gaussian Coefficients **33**

3.1 Introduction 33

3.2 Proof of Main Theorem and Corollary 34

3.3 Application to Patterned Random Matrices 37

Abstract

In this thesis, we first review some basic concepts in random matrix theory and then we show how these concepts motivate and help solve two independent problems. The first result is a Gaussian central limit theorem for eigenvalue statistics of inhomogeneous matrix models which interpolate between Wigner matrices and certain patterned random matrices. The second result is a non-asymptotic concentration inequality for the operator norm of random matrices with sub-Gaussian entries.

Chapter 1

Introduction and Background

1.1 What is a Random Matrix?

The field of random matrix theory began in the 1920s when Wishart introduced random matrices in mathematical statistics in [Wis28]. However, the field gained popularity after Wigner used ensembles of random matrices to study the nuclei of heavy atoms in [Wig51] and [Wig55]. The study of random matrices has since evolved from just being a mathematical tool into a field of its own, with applications in mathematics, physics, finance, computer science, and more. For more applications of the theory of random matrices, see [EW13] and the references therein.

So what is a random matrix? It is simply a matrix-valued random variable, or in other words, a matrix whose entries are random variables. In the background of this entire thesis, there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined. We will not directly refer to this probability space, but implicitly all probabilities are with respect to $(\Omega, \mathcal{F}, \mathbb{P})$. We will also assume all random variables are real-valued unless otherwise noted. Hence given n^2 random variables $\{X_{ij}\}_{1\leq i,j\leq n}$, an $n \times n$ random matrix X is

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nn} \end{bmatrix}. \tag{1.1}$$

When studying random matrices, we are often interested in understanding the eigenvalues of the given matrix. Similar questions can be asked about the eigenvectors (see [OVW16] for a review of this area), though they are much less understood. Eigenvectors will not be considered in this thesis and instead we will focus on the eigenvalues of random matrices, and in order to study them, we encapsulate them in the *empirical spectral distribution*.

Definition 1.1.1. *Let X be an $n\times n$ Hermitian random matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Then the empirical spectral distribution (ESD) of X is the random probability measure on $\mathbb{R}$ defined as $\mu_X = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}$.*

In other words, given a random matrix, we define a random probability measure on $\mathbb{R}$ that puts a point mass of $\frac{1}{n}$ at each eigenvalue (counted with multiplicity) of the associated random matrix. Note that X is assumed to be Hermitian, so all of its eigenvalues are real and thus μ_X is a probability measure on $\mathbb{R}$. A similar construction works for non-symmetric matrices, and in this case the ESDs will be probability measures on $\mathbb{C}$. However, for the majority of this thesis we will be considering symmetric random matrices unless otherwise noted, and hence all ESDs are measures on the real line.

The distribution of μ_X is a probability measure on the space of probability measures on $\mathbb{R}$. An important consequence of this definition is that for any $f \in C(\mathbb{R})$, we have $\int_{\mathbb{R}} f\, d\mu_X = \frac{1}{n}\sum_{i=1}^n f(\lambda_i)$ as random variables, and one can talk about convergence of ESDs. In order to understand the limiting behavior of the eigenvalues of a random matrix, we study the empirical spectral distribution of a sequence of matrices as the dimension tends to infinity. In other words, given a sequence of (square) random matrices $X_n \in M_{n\times n}(\mathbb{R})$, the limiting empirical spectral distribution is defined as $\mu := \lim_{n\to\infty} \mu_{X_n}$. This convergence of measures is defined weakly, and the type of weak convergence (e.g. convergence in probability or almost sure convergence) is dependent on the context. Hence we say that μ_{X_n} converges to μ weakly in probability (resp. almost surely) if for all $f \in C_c(\mathbb{R})$, $\int_{\mathbb{R}} f\, d\mu_{X_n}$ converges to $\int_{\mathbb{R}} f\, d\mu$ in probability (resp. almost surely).

One of the beautiful phenomena in random matrix theory is that often the limiting empirical spectral distribution turns out to be deterministic. We will see examples of this in Sections 1.2 and 1.3. The goal of the remainder of this introduction is to briefly introduce the reader to some of the important areas in random matrix theory in order to place the body of this thesis within a broader context. The main theorems in Chapter 2 pertain to models that interpolate between the ensembles discussed in Sections 1.2 (Wigner matrices) and 1.3 (certain types of patterned random matrices). Section 1.4 describes a sub-field of random matrix theory known as non-asymptotic random matrix theory, and techniques from this area are used to prove the main theorems in Chapter 2. Finally, the main result in Chapter 3 is purely non-asymptotic.

1.2 Wigner Matrices and the Semicircle Law

In this section we review a couple of benchmark results in random matrix theory, then describe how they relate to the main results in this thesis. We will not provide any proofs, though they can be found in Chapters 1 and 3 of [AGZ10].

The first main result is Wigner's semicircle law, and in order to state it we need to define Wigner matrices. In words, Wigner matrices are symmetric matrices with the maximum amount of independence among the entries, but to be more precise, let $\{Y_{ij}\}_{1\le i\le j\le n}$ be independent identically distributed random variables. Then define Y_n entry-wise as

$$(Y_n)_{ij} = \begin{cases} Y_{ij} & i \le j \\ Y_{ji} & i > j \end{cases}. \tag{1.2}$$

Written out in a matrix,

$$Y_n = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1n} \\ Y_{12} & Y_{22} & Y_{23} & \dots & Y_{2n} \\ Y_{13} & Y_{23} & Y_{33} & \dots & Y_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{1n} & Y_{2n} & Y_{3n} & \dots & Y_{nn} \end{bmatrix}. \tag{1.3}$$

For simplicity, we will further assume that $E[Y_{ij}] = 0$ and $E[(Y_{ij})^2] = 1$. Wigner's semicircle law is a statement about the limiting empirical spectral distribution of matrices of the above type, so we need to take the dimension to infinity. However, an observant reader may have already seen the following issue: if we take $n \to \infty$, the largest eigenvalue of Y_n will tend to ∞ as well. This is easily seen by the fact that if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of Y_n (with λ_n the largest in magnitude) $n\lambda_n^2 \geq \lambda_1^2 + \cdots + \lambda_n^2 = \mathrm{Tr}(Y^2) = \sum_{i,j=1}^n Y_{ij}^2$, and $\sum_{i,j=1}^n Y_{ij}^2$ grows like n^2 as $n \to \infty$ by the strong law of large numbers. Hence we introduce the following scaling to define Wigner matrices.

Definition 1.2.1. *Let Y_n be the sequence of matrices defined above. Then a Wigner matrix X_n is defined as $n^{-1/2}Y_n$.*

The $n^{-1/2}$ scaling may seem a bit peculiar, though it turns out that this is appropriate. For a simple argument explaining why, see section 1 of [Kem13]. This $n^{-1/2}$ scaling will appear in multiple places throughout this thesis, and in the Wigner case it turns out that the $n^{-1/2}$ scaling forces the limiting ESD to have compact support.

We are now ready to state the main theorem of this section, which is Wigner's semicircle law.

Theorem 1.2.2. *Let X_n be a sequence of Wigner matrices as defined above, and let μ_{X_n} be the associated empirical spectral distribution. Then as $n \to \infty$, μ_{X_n} converges weakly almost surely to the semicircle law σ, where σ is the following (deterministic) probability measure on $\mathbb{R}$:*

$$\sigma(dx) = \frac{1}{2\pi}\sqrt{(4 - x^2)_+}dx. \tag{1.4}$$

The measure σ defined above has a density whose graph is a (scaled) semicircle of radius two centered at zero, hence the name "the semicircle law". Figure 1.1 contains a plot of the histogram of the eigenvalues of a 4000×4000 Wigner matrix, and the structure of the spectrum is apparent. The above theorem is a statement about eigenvalues in the "bulk", i.e. about an order n number of eigenvalues. One can also study "local" statistics, in which one considers the behavior of small fractions of eigenvalues or even individual eigenvalues.

An example of a local result is the *Tracy-Widom Law*, which describes the behavior of the largest eigenvalue of certain types of Wigner matrices. The semicircle law hints that the top eigenvalue converges to $+2$, but we can study this convergence directly and further understand how the largest eigenvalue fluctuates around $+2$. The model we will consider here is the *Gaussian Unitary Ensemble* (GUE).

Definition 1.2.3. *For $k \geq j \geq 1$, let $\{Z_{jk}\}$ and $\{Z'_{jk}\}$ be two independent families of $N(0,1)$ random variables. Then define the $n \times n$ Hermitian complex matrix X_n whose entries are*

$$[X_n]_{jk} = \overline{[X_n]_{kj}} = n^{-1/2}\frac{1}{\sqrt{2}}(Z_{jk} + iZ'_{jk}), 1 \leq j < k \leq n \tag{1.5}$$

$$[X_n]_{jj} = n^{-1/2}Z_{jj}, 1 \leq j \leq n. \tag{1.6}$$

The above sequence X_n is called a Gaussian Unitary Ensemble GUE_n.

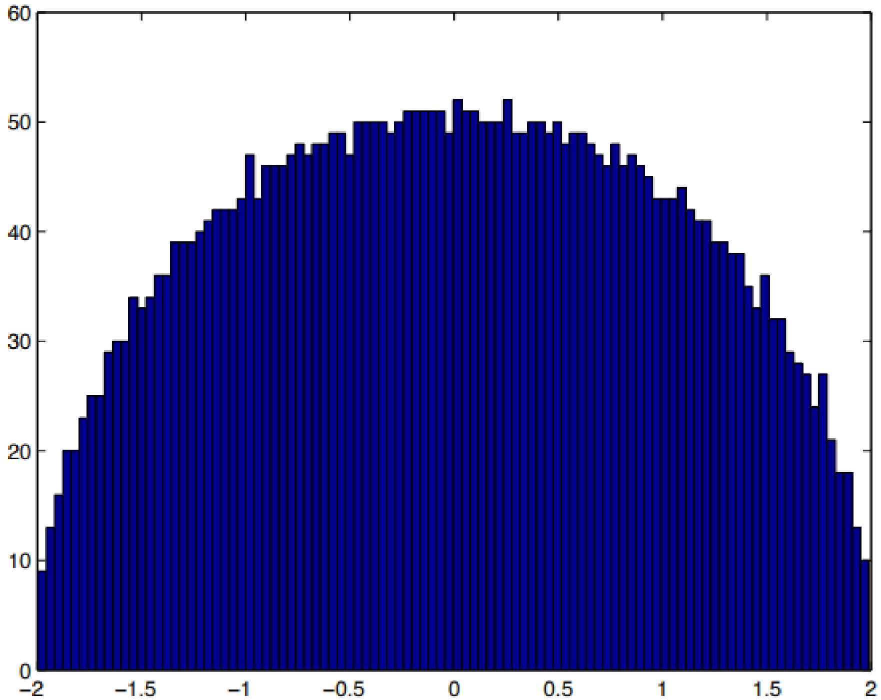

Figure 1.1: A histogram of the eigenvalues of a 4000×4000 Wigner matrix. Image taken from [Kem13].

Remark 1.2.4. *The name Gaussian Unitary Ensemble comes from the fact that the joint law of the entries is invariant under conjugation by unitary matrices.*

Theorem 1.2.5. *Let λ_n be the largest eigenvalues of a GUE_n. Then the limiting CDF of $n^{2/3}(\lambda_n - 2)$ is given by*

$$\lim_{n\to\infty} \mathbb{P}(n^{2/3}(\lambda_n - 2) \le t) = \exp\left(-\int_t^\infty (x-t)q(x)^2\,dx\right). \tag{1.7}$$

Here, q is a solution of the Painlevé II equation $q''(x) - xq(x) + 2q(x)^3 = 0$. Asymptotically, as $x \to \infty$, $q(x)$ is approximately the Airy function, which is the solution of the ODE $u''(x) = xu(x)$.

Remark 1.2.6. *Here, we considered complex matrices (the GUE_n) but analogous Tracy-Widom-type results hold for real matrices (the Gaussian Orthogonal Ensemble) and matrices with quaternion entries (the Gaussian Symplectic Ensemble).*

The above statement is fully local, as it describes the behavior of a single eigenvalue as the dimension tends to infinity. Further, instead of describing the deterministic limiting value of λ_n, the Tracy-Widom law describes the fluctuations of λ_n about its limit (with the appropriate scaling). In Chapter 2, the two main theorems are in some way a mixture of the semicircle law and the Tracy-Widom law. In Chapter 2, we study the eigenvalues in the bulk (i.e. we study all eigenvalues a once), and we prove two theorems about how

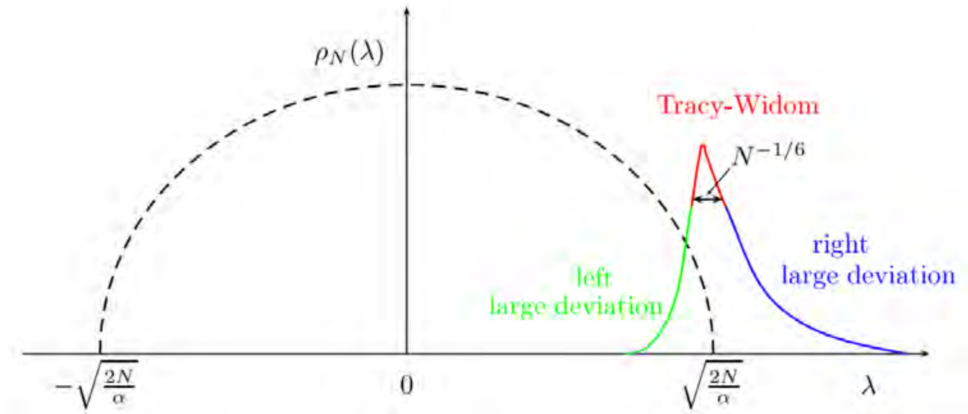

Figure 1.2: The eigenvalues of an unnormalized GUE_N (i.e. without the $N^{-1/2}$ scaling). Here, N denotes the size of the matrix, and α denotes the variance of the entries (in the context of this section $\alpha = 1$) The density of eigenvalues in the bulk is given by the semicircle $\rho_N(\lambda)$, and the fluctuations of the top eigenvalue are shown in color. Over a scale of $O(N^{-1/6})$ (which in the further $N^{-1/2}$ normalized case is $N^{-2/3}$), the maximum eigenvalue has the Tracy-Widom distribution, which is depicted in red. Image taken from [NM11].

certain statistics of theses eigenvalue fluctuate around their means. In particular, we study the fluctuations of *linear eigenvalue statistics*.

Definition 1.2.7. *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of an $n \times n$ matrix X_n. Then for a fixed test function f, the linear eigenvalue statistics are defined as*

$$\sum_{i=1}^{n} f(\lambda_i). \tag{1.8}$$

Note that the linear eigenvalue statistics for a random matrix X_n and a test function f are precisely the same as integrating f against the ESD (up to a factor of n). Hence, when $f = x^p$ for some $p \in \mathbb{N}$, these linear eigenvalue statistics return the moments of the ESD (multiplied by n).

When we describe their fluctuations, we get a result of the form

$$\varphi(n) \left(\int_{\mathbb{R}} f(x)\, d\mu_{X_n} - \int_{\mathbb{R}} f(x)\, d\bar{\mu}_n \right) \xrightarrow{n\to\infty} Y, \tag{1.9}$$

where $\varphi(n)$ describes some appropriate scaling, $\bar{\mu}_n$ is an "averaged" empirical spectral distribution, and Y is a limiting random variable that describes the fluctuations. Note that $\bar{\mu}_n$ is an "average" measure in the sense that $\mathbb{E}[\int_{\mathbb{R}} f(x)\, d\mu_{X_n}] = \int_{\mathbb{R}} f(x)\, d\bar{\mu}_n$, and its existence follows from the Riesz Representation Theorem.

1.3 Some Other Types of Patterned Random Matrices

While Wigner matrices play an important role in the study of random matrices, the field is not limited solely to models with independent entries. In [Bai99], Bai proposed the study of the limiting ESD of certain types of patterned random matrices with additional linear structure among the entries, and in [BDJ06], Bryc, Dembo, and Jiang provided solutions to these problems. In particular, they studied the limiting empirical spectral distributions of random Toeplitz, Hankel, and Markov matrices. The precise definitions of Toeplitz and Hankel models are given below and will be provided again later on when necessary (Markov matrices are not relevant to this work).

Definition 1.3.1. *An $n \times n$ symmetric random Toeplitz matrix T_n is defined as follows. Let $X_0, X_1, X_2, \ldots$ be independent random variables. Then*

$$T_n = \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{n-2} & X_{n-1} \\ X_1 & X_0 & X_1 & \cdots & X_{n-3} & X_{n-2} \\ X_2 & X_1 & X_0 & \cdots & X_{n-4} & X_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ X_{n-2} & X_{n-3} & X_{n-4} & \cdots & X_0 & X_1 \\ X_{n-1} & X_{n-2} & X_{n-3} & \cdots & X_1 & X_0 \end{bmatrix}. \tag{1.10}$$

In other words, $(T_n)_{ij} = X_{|i-j|}$.

Definition 1.3.2. *An $n \times n$ random Hankel matrix is defined as follows. Let $X_1, X_2, \ldots, X_{2n-1}$ be independent random variables. Then*

$$H_n = \begin{bmatrix} X_1 & X_2 & X_3 & \cdots & X_{n-1} & X_n \\ X_2 & X_3 & X_4 & \cdots & X_n & X_{n+1} \\ X_3 & X_4 & X_5 & \cdots & X_{n+1} & X_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_n & X_{n+1} & X_{n+2} & \cdots & X_{2n-2} & X_{2n-1} \end{bmatrix} \tag{1.11}$$

In other words, $(H_n)_{ij} = X_{i+j-1}$.

Bryc, Dembo, and Jiang showed that when the entries have mean zero and variance $\frac{1}{n}$, the limiting spectral distribution for Toeplitz and Hankel matrices is a nonrandom, symmetric distribution with unbounded support. Since then, the spectra of patterned random matrices have been extensively studied (see the book of Bose [Bos18] and the references therein). Chapter 2 concerns generalized versions of random Toeplitz and Hankel matrices, as well as generalized versions of other patterned models, namely circulant, reverse circulant, and symmetric circulant matrices (which are defined in Chapter 2). These generalized models interpolate between Wigner matrices and the corresponding patterned random matrices. These are the types of matrices that we will prove a fluctuations result for their linear statistics.

1.4 Non-Asymptotic Random Matrix Theory

In the above sections, we have discussed "classical" random matrix theory. This theory involves specialized models (usually assuming i.i.d. entries), and the ensembles in question are usually highly structured and invariant as the dimension tends to infinity. For example any Toeplitz matrix has independent diagonals and identically distributed entries along each diagonal, and this condition is invariant as $n \to \infty$. In classical random matrix theory, we can usually find exact asymptotics of the limiting spectral distribution (and a variety of other quantities of interest).

There is, however, another sub-field of random matrix theory, in which researchers study random matrices of a fixed dimension. This area is called non-asymptotic random matrix theory. Researchers in this area study highly inhomogeneous random matrices of a fixed dimension, and instead of getting exact results in the limit, they prove high probability estimates (that depend on the dimension).

In this section we will give a quick overview of some of the results in non-asymptotic random matrix theory (relevant to this thesis) and how they can be used to further understand the spectrum of random matrices. For a detailed introduction to the field, see the surveys of Tropp [Tro15] and Vershynin [Ver12], and the book by Vershynin [Ver18].

A large part of non-asymptotic random matrix theory is bounding the largest singular value (either in expectation or with high probability) of a random matrix, and one of the approaches to doing this is through *matrix concentration inequalities*. This theory exploits the principle that in practice, most random matrices X of interest can be written as a sum of independent random matrices in the following way

$$X = \sum_{i=1}^{k} Z_i \tag{1.12}$$

where Z_i are independent random matrices of common dimension. Trivially, we can write any random matrix in this form with $k = 1$, but in many cases k can be large and depend on the dimension. In these scenarios, we can get useful bounds on the largest singular value. We will see many examples of this phenomenon applied to patterned random matrices later on in this thesis, but the following is an example of how this idea is applied to Wigner matrices:

Example 1.4.1. *Let $E_{ij} \in M_n(\mathbb{R})$ for $1 \leq i \leq j \leq n$ be defined by*

$$(E_{ij})_{kl} = \begin{cases} n^{-1/2} & (i,j) = (k,l) \text{ or } (i,j) = (l,k) \\ 0 & \text{otherwise.} \end{cases} \tag{1.13}$$

Now for $1 \leq i \leq j \leq n$, let X_{ij} be independent, identically distributed random variables with mean 0 and variance 1. Then any Wigner matrix can be written as $\sum_{1 \leq i \leq j \leq n} X_{ij} E_{ij}$ where. Here, $\{X_{ij}E_{ij}\}_{1 \leq i \leq j \leq n}$ is a set of independent random matrices.

Then matrix concentration inequalities can bound how much the operator norms of matrices that can be written in the form of equation (1.12) differ from their means. In other words, they provide tail estimates of the form

$$\mathbb{P}(||X - \mathbb{E}[X]|| \geq t) \leq \dots \tag{1.14}$$

where the expectation of a matrix is taken entry-wise. In the situations considered in this thesis, we assume that the entries of the random matrices are centered, so $\mathbb{E}[X]$ is just the zero matrix.

The ideas and main results in matrix concentration inequalities have their roots in the analogous theory for real-valued random variables. Concentration inequalities for sums of independent random variables have been extensively studied, and important results in this area include the simple Markov and Chebyshev inequalities, as well as Bernstein inequalities and Chernoff bounds. Many of these have matrix analogues which are used to bound the operator norm of random matrices, as is evidenced in the following comparison between Bernstein inequalities for real valued random variables and matrices taken from [Tro15].

Theorem 1.4.2. (Bernstein Inequality for real-valued random variables)

Let $S_1, \dots, S_n$ be independent, centered, real random variables, and assume that each one is uniformly bounded:

$$\mathbb{E}S_k = 0 \textit{ and } |S_k| \leq L \textit{ for each } k = 1, \dots, n \tag{1.15}$$

for some $L > 0$. Then if $Z = \sum_{k=1}^{n} S_k$, let $\nu(Z)$ denote the variance of Z:

$$\nu(Z) = \mathbb{E}Z^2 = \sum_{k=1}^{n} \mathbb{E}S_k^2. \tag{1.16}$$

Then

$$\mathbb{P}(|Z| \geq t) \leq 2\exp\left(\frac{-t^2/2}{\nu(Z) + Lt/3}\right) \textit{ for all } t \geq 0. \tag{1.17}$$

With this in mind, there is also the following Matrix Bernstein Inequality, which is strikingly similar.

Theorem 1.4.3. (Bernstein Inequality for matrices)

Let $\mathbf{S}_1, \dots, \mathbf{S}_n$ be independent, centered random matrices with common dimension $d_1 \times d_2$, and assume that each one is uniformly bounded

$$\mathbb{E}\mathbf{S}_k = \mathbf{0} \textit{ and } ||\mathbf{S}_k|| \leq L \textit{ for each } k - 1, \dots, n. \tag{1.18}$$

Then let

$$\mathbf{Z} = \sum_{k=1}^{n} \mathbf{S}_k, \tag{1.19}$$

and let $\nu(\mathbf{Z})$ denote the matrix variance statistic of $\mathbf{Z}$:

$$\nu(\mathbf{Z}) = \max\{||\mathbb{E}(\mathbf{Z}\mathbf{Z}^*)||, ||\mathbb{E}(\mathbf{Z}^*\mathbf{Z})||\} \tag{1.20}$$

$$= \max\left\{\left\|\sum_{k=1}^{n} \mathbb{E}(\mathbf{S}_k\mathbf{S}_k^*)\right\|, \left\|\sum_{k=1}^{n} \mathbb{E}(\mathbf{S}_k^*\mathbf{S}_k)\right\|\right\}. \tag{1.21}$$

Then

$$\mathbb{P}(||Z|| \geq t) \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{\nu(Z) + Lt/3}\right) \textit{ for all } t \geq 0. \tag{1.22}$$

This correspondence between matrix and ordinary concentration inequalities appears in non-asymptotic random matrix theory quite frequently, and we prove such an example in Chapter 3. The result in Chapter 3 resembles the Matrix Bernstein Inequality, though the models considered in the chapter are constructed as series of independent sub-Gaussian random variables multiplied by deterministic matrices. Further, the main theorem in Chapter 3 has an analogue for real-valued sub-Gaussian random variables, much like the relation between the Matrix Bernstein Inequality and Bernstein Inequality for real-valued random variables.

1.5 Outline of Main Results

This thesis contains two papers which appear in Chapters 2 and 3. The first result pertains to how the eigenvalues for certain types of random matrices fluctuate around their averaged empirical spectral distribution. This result is a central limit theorem for statistics related to the eigenvalues of generalized patterned random matrices. The matrices considered in Chapter 2 (defined in 2.1) are highly inhomogeneous in their correlation structure, and in order to get a handle on their eigenvalues we employ combinatorial arguments and non-asymptotic techniques. To the best of our knowledge, this is the first type of result where a Gaussian central limit theorem for these eigenvalue statistics has been proved for inhomogeneous random matrices.

The second paper is a purely non-asymptotic result. It is a concentration inequality for random matrices that can be constructed as a series of sub-Gaussian random variables multiplied by deterministic Hermitian matrices. Similar types of results have been proven for random matrices with Gaussian and Rademacher entries (both of which are sub-Gaussian), but our result applies in full generality to all sub-Gaussian random variables. The key tool in this paper is a bound on the matrix moment generating function using the moment bounds of sub-Gaussian random variables.

Chapter 2

Fluctuations of Eigenvalues for Random Matrices with Correlated Entries

2.1 Introduction

In this paper, we study the fluctuations of linear statistics of eigenvalues for patterned random matrix models with correlated entries. If X_n is an $n \times n$ random matrix, its linear eigenvalue statistics are defined as

$$\sum_{i=1}^{n} f(\lambda_i) \tag{2.1}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of X_n and f is a fixed test function.

The linear eigenvalue statistics of various random matrix models have been extensively studied in the literature. One is often interested in the fluctuations of linear statistics, and Gaussian central limit theorem results for various random matrix models have recently been proved. For fluctuations results for Wigner matrix models, see [Cha09], [AZ06], [Joh98], [Sos02], [SS98], [Shc11], [LP09], and [SW13] and the references therein. For random Toeplitz matrix models, see [Cha09] and [LSW12], and for circulant matrices see [AS17] and [AS18]. For a non-Gaussian fluctuations result, see [SMS22] for an example with odd monomial test functions and random Hankel matrices. The fluctuations of eigenvalues for symmetric circulant and reverse circulant matrices were studied in [MS21] and [AS17].

To the best of our knowledge, the current linear statistics fluctuations results in the literature do not consider any models in which the matrix entries are allowed to have some general correlation structure. In this paper, we study the linear statistics of models that resemble the symmetric random Toeplitz, circulant, reverse circulant, symmetric circulant, and Hankel models, except we allow correlations among certain entries. This work is a generalization of Section 4.3 of [Cha09] and the related results in [AS17].

The random Toeplitz model with correlated entries is defined as follows, and is a more general version of the model studied in [FL13]. The corresponding circulant, reverse circulant, symmetric circulant, and Hankel models with correlated entries are defined in

Section 2.1.1.

Definition 2.1.1. *For $n \in \mathbb{N}$, a random matrix $X = X_n$ is called a random Toeplitz matrix with correlated entries if it is symmetric and has identically distributed, independent diagonals up to symmetry. The entries of X are assumed to satisfy $\mathbb{E}(X_{ij}) = 0$.*

An $n \times n$ Toeplitz matrix with correlated entries is determined by n independent random vectors of dimensions varying from 1 through n, $\{\mathbf{X}_k\}_{k=0}^{n-1}$ with $\mathbf{X}_k \in \mathbb{R}^{n-k}$. Then the entries of the matrix are given by $X_{ij} = [\mathbf{X}_{|i-j|}]_{\min(i,j)}$. That is,

$$X = \begin{bmatrix} [\mathbf{X}_0]_1 & [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & \cdots & [\mathbf{X}_{n-1}]_1 \\ [\mathbf{X}_1]_1 & [\mathbf{X}_0]_2 & [\mathbf{X}_1]_2 & \cdots & [\mathbf{X}_{n-2}]_2 \\ [\mathbf{X}_2]_1 & [\mathbf{X}_1]_2 & [\mathbf{X}_0]_3 & \cdots & [\mathbf{X}_{n-3}]_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ [\mathbf{X}_{n-1}]_1 & [\mathbf{X}_{n-2}]_2 & [\mathbf{X}_{n-3}]_3 & \cdots & [\mathbf{X}_0]_n \end{bmatrix} \tag{2.2}$$

We refer to the vector $\mathbf{X}_k$ as the kth diagonal of the correlated Toeplitz matrix.

The most important example is the Gaussian case where the diagonal vectors are all jointly Gaussian vectors. In this case, the distribution of the entries of the matrix is completely determined by their covariances. For notation, we use

$$c_k(i,j) = \mathrm{Cov}([\mathbf{X}_k]_i, [\mathbf{X}_k]_j) = \mathbb{E}([\mathbf{X}_k]_i[\mathbf{X}_k]_j). \tag{2.3}$$

Note that the values $c_k(i,j)$ also depend on n and they cannot be chosen with complete freedom, since the $(n-k) \times (n-k)$ covariance matrix of each diagonal must be positive semidefinite. Also, due to the independent diagonals condition the covariance between any two entries on different diagonals is 0. The covariances between entries in the upper triangle of an Toeplitz matrix can thus be written as

$$\mathrm{Cov}(X_{ij}, X_{kl}) = \mathbb{E}([\mathbf{X}_{|i-j|}]_i[\mathbf{X}_{|k-l|}]_k) = \delta_{|i-j|=|k-l|}c_{|i-j|}(i,k), \qquad i \le j,\ k \le l. \tag{2.4}$$

As stated above, the covariances $c_k(i,j)$ completely determine the distribution of the entries in the Gaussian case. When $c_k(i,j) = \delta_{ij}$ for all i,j,k, we have a Gaussian Wigner matrix and when $c_k(i,j) = 1$ for all i,j,k, we have a Gaussian symmetric random Toeplitz matrix. Hence, one can think of correlated Toeplitz matrices as a model that interpolates between the well-studied Wigner and random Toeplitz matrix models.

In this paper, we prove Gaussian central limit theorems for the linear eigenvalue statistics of random Toeplitz matrices with correlated entries. We also prove a similar result for correlated versions of circulant, reverse circulant, symmetric circulant, and Hankel matrices (we give precise definitions of these models in Section 2.1.1). We have the following theorem.

Theorem 2.1.2. *Let X_n be an $n \times n$ random Toeplitz, circulant or symmetric circulant matrix with correlated centered Gaussian entries with covariances labeled by $c_k(i,j)$. Assume that there exists constants m, $M \in (0,\infty)$ such that $m \le \mathbb{E}(X_{ij}^2) \le M$ for all i,j, and n. Further, assume that there exists $\gamma > 0$ such that $c_k(i,j) \ge \gamma$ for all i,j,k, and n. Let p be a positive integer and let $W_n = \mathrm{Tr}(X_n^p)$. Then as $n \to \infty$,*

$$\frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} \text{ converges in total variation to } N(0,1). \tag{2.5}$$

In the case when X_n is a reverse circulant or Hankel matrix with correlated entries, the theorem still holds under the further assumption that p is restricted to be an even positive integer.

The condition on the parity of p is most likely required in the case of the Hankel matrices with correlated entries. In the uncorrelated model, it was shown in [SMS22] that the fluctuations of the linear eigenvalue statistics converge to a non-Gaussian limit for odd monomial test functions. However, it is not clear whether this condition is required in the reverse circulant case.

Furthermore, the reverse circulant, circulant, and Hankel models with correlated entries (defined in Section 2.4) are not symmetric and hence do not necessarily have real eigenvalues.

The above theorem allows for very general correlation structures among the diagonals. The covariances along each diagonal can fluctuate wildly with n, and as long as they stay bounded away from 0, the Gaussian central limit theorem still holds. The next theorem examines the fluctuations of the linear eigenvalue statistics in the regime where the covariances among the entries uniformly converge to 0.

Theorem 2.1.3. *Let X_n be an $n \times n$ random Toeplitz, circulant, reverse circulant, symmetric circulant, or Hankel matrix with correlated centered Gaussian entries with covariances labeled by $c_k(i,j)$. Assume that there exists constants m, $M \in (0,\infty)$ such that $m \leq \mathbb{E}(X_{ij}^2) \leq M$ for all i,j and for all n. Further, assume $c_k(i,j) = o(n^{-1/3})$ for all i,j,k, and n with $i \neq j$. Let p be a positive integer and let $W_n = \mathrm{Tr}(X_n^p)$. Then as $n \to \infty$,*

$$\frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} \text{ converges in total variation to } N(0,1). \tag{2.6}$$

The set up of the paper is as follows. We first prove, in detail, Theorem 3.1.3 in the case when X_n is a random Toeplitz matrix with correlated entries. Thus in Section 2.2 we give an outline of the correlated Toeplitz matrix proof and compute some necessary bounds on the variance. In Section 2.3 we compute a bound on the operator norm of an arbitrary Toeplitz matrix with correlated entries and then use this bound to complete the proof of Theorem 3.1.3. In Section 2.4, we comment on how the Toeplitz proof of Theorem 3.1.3 can be adapted for each of the other matrix models. In Section 2.5, we prove Theorem 2.1.3. Finally, in Section 2.6 we conjecture about how Theorem 3.1.3 may be extended, both by removing the conditions on the covariances and by generalizing to sub-Gaussian entries.

2.1.1 Definitions of Other Correlated Matrix Models

In this section we define other versions of patterned random matrix models that allow for correlations among the entries. In [AS17], Adhikari and Saha proved Gaussian fluctuations results for the uncorrelated versions of the following matrices. In all of the following models, for $k = 1, \ldots, n$, let

$$c_k(i,j) = \mathrm{Cov}([\mathbf{X}_k]_i, [\mathbf{X}_k]_j) = \mathbb{E}([\mathbf{X}_k]_i [\mathbf{X}_k]_j). \tag{2.7}$$

Reverse Circulant Matrices with Correlated Entries

Definition 2.1.4. *An $n \times n$ reverse circulant matrix is defined as*

$$RC_n = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_n \\ x_2 & x_3 & x_4 & \cdots & x_n & x_1 \\ x_3 & x_4 & x_5 & \cdots & x_1 & x_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n & x_1 & x_2 & \cdots & x_{n-2} & x_{n-1} \end{bmatrix} \tag{2.8}$$

In other words, $(RC_n)_{ij} = x_{(i+j-1) \mod n}$. The $(j+1)$-th row is obtained by shifting the j-th row by one position to the left.

In order to introduce the correlations, we have n independent centered random vectors $\mathbf{X}_k \in \mathbb{R}^n$ for $k = 1, \ldots, n$. Then the random reverse circulant matrix with correlated entries can be written as

$$X_n = n^{-1/2} \begin{bmatrix} [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & [\mathbf{X}_3]_1 & \cdots & [\mathbf{X}_{n-1}]_1 & [\mathbf{X}_n]_1 \\ [\mathbf{X}_2]_2 & [\mathbf{X}_3]_2 & [\mathbf{X}_4]_2 & \cdots & [\mathbf{X}_n]_2 & [\mathbf{X}_1]_2 \\ [\mathbf{X}_3]_3 & [\mathbf{X}_4]_3 & [\mathbf{X}_5]_3 & \cdots & [\mathbf{X}_1]_3 & [\mathbf{X}_2]_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ [\mathbf{X}_n]_n & [\mathbf{X}_1]_n & [\mathbf{X}_2]_n & \cdots & [\mathbf{X}_{n-2}]_n & [\mathbf{X}_{n-1}]_n \end{bmatrix} \tag{2.9}$$

Circulant Matrices with Correlated Entries

Definition 2.1.5. *An $n \times n$ circulant matrix is defined as*

$$C_n = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_n \\ x_n & x_1 & x_2 & \cdots & x_{n-2} & x_{n-1} \\ x_{n-1} & x_n & x_1 & \cdots & x_{n-3} & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_2 & x_3 & x_4 & \cdots & x_n & x_1 \end{bmatrix} \tag{2.10}$$

In other words, $(C_n)_{ij} = x_{(j-i+1) \mod n}$. The $(j+1)$-th row is obtained by shifting the j-th row by one position to the right.

In order to introduce the correlations, we have n independent centered random vectors $\mathbf{X}_k \in \mathbb{R}^n$ for $k = 1, \ldots, n$. Then the random circulant matrix with correlated entries can be written as

$$X_n = n^{-1/2} \begin{bmatrix} [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & [\mathbf{X}_3]_1 & \cdots & [\mathbf{X}_{n-1}]_1 & [\mathbf{X}_n]_1 \\ [\mathbf{X}_n]_2 & [\mathbf{X}_1]_2 & [\mathbf{X}_2]_2 & \cdots & [\mathbf{X}_{n-2}]_2 & [\mathbf{X}_{n-1}]_2 \\ [\mathbf{X}_{n-1}]_3 & [\mathbf{X}_n]_3 & [\mathbf{X}_1]_3 & \cdots & [\mathbf{X}_{n-3}]_3 & [\mathbf{X}_{n-2}]_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ [\mathbf{X}_2]_n & [\mathbf{X}_3]_n & [\mathbf{X}_4]_n & \cdots & [\mathbf{X}_{n-1}]_n & [\mathbf{X}_1]_n \end{bmatrix} \tag{2.11}$$

Symmetric Circulant Matrices with Correlated Entries

Definition 2.1.6. *An $n \times n$ symmetric circulant matrix is defined as*

$$SC_n = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_2 & x_1 \\ x_1 & x_0 & x_1 & \cdots & x_3 & x_2 \\ x_2 & x_1 & x_0 & \cdots & x_4 & x_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1 & x_2 & x_3 & \cdots & x_1 & x_0 \end{bmatrix} \tag{2.12}$$

In other words, $(SC_n)_{ij} = x_{\frac{n}{2}-|\frac{n}{2}-|i-j||}$. The $(j+1)$-th row is obtained by shifting the j-th row by one position to the right.

In the correlated model (similar to the uncorrelated model), the structure of the random matrix slightly depends on the parity of n. When n is odd we have n independent centered Gaussian random vectors $\mathbf{X}_k \in \mathbb{R}^n$ for $k = 1, \dots, n$. When n is even, we have $n-1$ independent Gaussian random vectors $X_k \in \mathbb{R}^n$ for $k = 1, \dots \frac{n}{2}-1, \frac{n}{2}+1, \dots n$ and 1 independent Gaussian random vector $X_{n/2} \in R^{n/2}$. Then the random symmetric circulant matrix with correlated entries can be written as

$$X_n = n^{-1/2} \begin{bmatrix} [\mathbf{X}_0]_1 & [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & \cdots & [\mathbf{X}_2]_{n-1} & [\mathbf{X}_1]_n \\ [\mathbf{X}_1]_1 & [\mathbf{X}_0]_2 & [\mathbf{X}_1]_2 & \cdots & [\mathbf{X}_3]_{n-1} & [\mathbf{X}_2]_n \\ [\mathbf{X}_2]_1 & [\mathbf{X}_1]_2 & [\mathbf{X}_0]_3 & \cdots & [\mathbf{X}_4]_{n-1} & [\mathbf{X}_3]_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ [\mathbf{X}_1]_n & [\mathbf{X}_2]_n & [\mathbf{X}_3]_n & \cdots & [\mathbf{X}_1]_{n-1} & [\mathbf{X}_0]_n \end{bmatrix} \tag{2.13}$$

The following example exhibits the difference in structure due to the parity of the size of the matrix.

Example 2.1.7. *We have*

$$X_4 = \frac{1}{2} \begin{bmatrix} [\mathbf{X}_0]_1 & [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & [\mathbf{X}_1]_4 \\ [\mathbf{X}_1]_1 & [\mathbf{X}_0]_2 & [\mathbf{X}_1]_2 & [\mathbf{X}_2]_2 \\ [\mathbf{X}_2]_1 & [\mathbf{X}_1]_2 & [\mathbf{X}_0]_3 & [\mathbf{X}_1]_3 \\ [\mathbf{X}_1]_4 & [\mathbf{X}_2]_2 & [\mathbf{X}_1]_3 & [\mathbf{X}_0]_4 \end{bmatrix} \quad \textit{and} \quad X_5 = \frac{1}{\sqrt{5}} \begin{bmatrix} [\mathbf{X}_0]_1 & [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & [\mathbf{X}_2]_4 & [\mathbf{X}_1]_5 \\ [\mathbf{X}_1]_1 & [\mathbf{X}_0]_2 & [\mathbf{X}_1]_2 & [\mathbf{X}_2]_2 & [\mathbf{X}_2]_5 \\ [\mathbf{X}_2]_1 & [\mathbf{X}_1]_2 & [\mathbf{X}_0]_3 & [\mathbf{X}_1]_3 & [\mathbf{X}_2]_3 \\ [\mathbf{X}_2]_4 & [\mathbf{X}_2]_2 & [\mathbf{X}_1]_3 & [\mathbf{X}_0]_4 & [\mathbf{X}_1]_4 \\ [\mathbf{X}_1]_5 & [\mathbf{X}_2]_5 & [\mathbf{X}_2]_3 & [\mathbf{X}_1]_4 & [\mathbf{X}_0]_5 \end{bmatrix} \tag{2.14}$$

Hankel Matrices with Correlated Entries

Definition 2.1.8. *An $n \times n$ Hankel matrix is defined as*

$$H_n = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_n \\ x_2 & x_3 & x_4 & \cdots & x_n & x_{n+1} \\ x_3 & x_4 & x_5 & \cdots & x_{n+1} & x_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n & x_{n+1} & x_{n+2} & \cdots & x_{2n-2} & x_{2n-1} \end{bmatrix} \tag{2.15}$$

In other words, $(H_n)_{ij} = x_{i+j-1}$.

In order to introduce the correlations we have n vectors $\mathbf{X}_k \in \mathbb{R}^k$ for $k = 1, \ldots, n$ and $n-1$ vectors $\mathbf{X}_k \in \mathbb{R}^{2n-k}$ for $k = n+1, \ldots 2n-1$. Then the Hankel matrix with correlated entries is

$$X_n = n^{-1/2} \begin{bmatrix} [\mathbf{X}_1]_1 & [\mathbf{X}_2]_1 & [\mathbf{X}_3]_1 & \cdots & [\mathbf{X}_{n-1}]_1 & [\mathbf{X}_n]_1 \\ [\mathbf{X}_2]_2 & [\mathbf{X}_3]_2 & [\mathbf{X}_4]_2 & \cdots & [\mathbf{X}_n]_2 & [\mathbf{X}_{n+1}]_1 \\ [\mathbf{X}_3]_3 & [\mathbf{X}_4]_3 & [\mathbf{X}_5]_3 & \cdots & [\mathbf{X}_{n+1}]_2 & [\mathbf{X}_{n+2}]_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ [\mathbf{X}_n]_n & [\mathbf{X}_{n+1}]_{n-1} & [\mathbf{X}_{n+2}]_{n-2} & \cdots & [\mathbf{X}_{2n-2}]_2 & [\mathbf{X}_{2n-1}]_1 \end{bmatrix} \tag{2.16}$$

2.2 Proof of Theorem 1.2 for Toeplitz Models: Outline and First Steps

This section and the following section are devoted to proving Theorem 3.1.3 for random Toeplitz matrices with correlated entries.

In order to prove the theorem we use the technique of [Cha09], in which Chatterjee gave a method for proving central limit theorems for linear statistics of eigenvalues of random matrices via second order Poincaré inequalities. Let $(X_{ij})_{1\le i,j\le n}$ be a collection of jointly Gaussian random variables with $n^2 \times n^2$ covariance matrix Σ. Let $X = n^{-1/2}(X_{ij})_{1\le i,j\le n}$. Then we have the following proposition.

Proposition 2.2.1. *(Proposition 4.4. of [Cha09]) Take an entire function f and define f_1, f_2 as*

$$f_1(z) = \sum_{m=1}^{\infty} m|b_m|x^{m-1} \qquad f_2(z) = \sum_{m=2}^{\infty} m(m-1)|b_m|z^{m-2} \tag{2.17}$$

where $f(z) = \sum_{m=0}^{\infty} b_m z^m$. Let λ denote the operator norm of X. Let $a = (\mathbb{E}f_1(\lambda)^4)^{1/4}$ and $b = (\mathbb{E}f_2(\lambda)^4)^{1/4}$. Suppose $W = \mathrm{Re}\mathrm{Tr}f(X)$ has finite fourth moment and let $\sigma^2 = \mathrm{Var}(W)$. Let Z be a normal random variable with the same mean and variance as W. Then

$$d_{TV}(W, Z) \le \frac{2\sqrt{5}||\Sigma||^{3/2}ab}{\sigma^2 n} \tag{2.18}$$

We use this proposition to prove Theorem 3.1.3. Here we consider monomials, so $f = x^p$, $f_1 = px^{p-1}$, and $f_2 = p(p-1)x^{p-2}$, and X will be an $n \times n$ Toeplitz matrix with correlated entries (which we denote as X_n) with covariances in $[\gamma, 1]$. We need to bound three terms, $||\Sigma||$ (where $||\cdot||$ is the operator norm), ab, and σ^2. The bound for $||\Sigma||$ is almost the same as in [Cha09] for random Toeplitz matrices, and the argument for bounding σ^2 from below is also modeled off of Chatterjee's argument with some modifications to deal with the covariances. The bound for ab is new, and in order to bound the term we need to employ the matrix concentration inequalites in [Tro18].

For completeness, we give the proof of the bound on the operator norm of the covariance matrix Σ via the Gershgorin circle theorem.

Lemma 2.2.2. *For any $n \times n$ random Toeplitz matrix with correlated entries with Σ as above, $||\Sigma|| \le 2Mn$.*

Proof. Let $\sigma_{ij,kl} = \mathrm{Cov}(X_{ij}, X_{kl})$, so $\sigma_{ij,kl} = 0$ if $|i-j| \neq |k-l|$ and if $|i-j| = |k-l|$, $\sigma_{ij,kl} \leq 1$. By the Gershgorin circle theorem, if $\lambda_1 \dots \lambda_{n^2}$ are the eigenvalues of Σ, we have

$$||\Sigma|| = \max(|\lambda_m|) \leq \sup_{1\leq i,j\leq n} \sum_{k,l=1}^{n} |\sigma_{ij,kl}|. \tag{2.19}$$

For any fixed i, j, there are at most $2n$ values of k, l such that $|i-j| = |k-l|$. Hence each term in the maximum is bounded by $2Mn$ (the M factor comes from applying Hölder's inequality to each $c_k(i,j)$) and thus so is $||\Sigma||$. □

As mentioned above, the proof for bounding the variance from below is similar to Chatterjee's argument for random Toeplitz matrices. The main difference is employing Wick's theorem to show that the covariances of products of the entries are positive, and then using the bound γ to show the variance grows at least linearly.

Remark 2.2.3. *This is the only part of the proof that uses the lower bound γ of the covariances. This bound is sufficient to get linear growth of the variance, but may not be necessary. However, one cannot drop all conditions on the covariances and still maintain linear growth of the variance (see Corollary 1 or Theorem 2 of [SS98], in which σ^2 converges to a finite limit for the Wigner matrix model).*

Lemma 2.2.4. *With W_n as in Theorem 1.2 for the Toeplitz model with correlated entries,* $\mathrm{Var}(W_n) \geq Kn$ *for some constant K that only depends on p and γ.*

Proof. It suffices to prove the lemma under the assumption that all of the entries are standard Gaussian random variables. In the case when they are not, normalize each entry and factor out the normalization coefficients and apply the same argument (see the proof of Theorem 4.2 in [Cha09]).

We first show that products of the entries all have nonnegative covariances. In this proof, $X_{ij} \coloneqq (X_n)_{ij}$ where X_n is the correlated Toeplitz matrix. For any collections of non-negative integers $(\alpha_{ij})_{1\leq i\leq j\leq n}$ and $(\beta_{ij})_{1\leq i\leq j\leq n}$ we have

$$\mathrm{Cov}\left(\prod X_{ij}^{\alpha_{ij}}, \prod X_{ij}^{\beta_{ij}}\right) = \mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}+\beta_{ij}}\right) - \mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}}\right)\mathbb{E}\left(\prod X_{ij}^{\beta_{ij}}\right) \tag{2.20}$$

and by the independent diagonals condition, these products factor as (with $k = j - i$ and using the fact that $X_{ji} = X_{ij}$)

$$\mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}+\beta_{ij}}\right) - \mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}}\right)\mathbb{E}\left(\prod X_{ij}^{\beta_{ij}}\right) = \prod_{k=0}^{n-1}\mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}+\beta_{ij}}\right) - \prod_{k=0}^{n-1}\mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}}\right)\mathbb{E}\left(\prod X_{ij}^{\beta_{ij}}\right) \tag{2.21}$$

where the products on the right hand side inside the expectation are now taken over all $1 \leq i \leq j \leq n$ such that $j - i = k$. We show that for each k, the term in the product on the left is at least its corresponding term on the right. Fix k (so all i, j below are such that $j - i = k$) and consider

$$\mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}+\beta_{ij}}\right) - \mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}}\right)\mathbb{E}\left(\prod X_{ij}^{\beta_{ij}}\right). \tag{2.22}$$

Then since the X_{ij} are centered multivariate Gaussian random variables, by Wick's theorem if $\sum \alpha_{ij}$ or $\sum \beta_{ij}$ is odd, then the term on the right is 0. Thus it remains to consider the case when $\sum \alpha_{ij} = 2n$ and $\sum \beta_{ij} = 2m$ for some positive integers m, n. For notation purposes, enumerate the $2n + 2m$ X_{ij}'s by $X_1, \ldots X_{2n+2m}$ such that the first $X_1, \ldots, X_{2n}$ correspond to an X_{ij} which is raised to some α_{ij}, and $X_{2n+1}, \ldots X_{2n+2m}$ correspond to an X_{ij} which is raised to some β_{ij}. Here, $P_2(k)$ denotes the set of pair partitions of $[k]$. In this case by Wick's Theorem

$$\mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}+\beta_{ij}}\right) - \mathbb{E}\left(\prod X_{ij}^{\alpha_{ij}}\right)\mathbb{E}\left(\prod X_{ij}^{\beta_{ij}}\right) = \tag{2.23}$$

$$\sum_{\pi \in P_2(2(n+m))} \prod_{\{i,j\}\in\pi} \mathbb{E}(X_iX_j) - \left(\sum_{\pi\in P_2(2n)} \prod_{\{i,j\}\in\pi} \mathbb{E}(X_iX_j)\right)\left(\sum_{\pi\in P_2(2m)} \prod_{\{i,j\}\in\pi} \mathbb{E}(X_{i+2n}X_{j+2n})\right) \tag{2.24}$$

The map $\phi : P_2(2n) \times P_2(2m) \mapsto P_2(2(n+m))$ where $\phi(\pi, \sigma) = \pi \cup (2n + \sigma)$ (where addition is done element wise) is an injection. Thus every term in the double sum on the right has a corresponding term on the left. Since each of the $c_k(i, j) \geq 0$, $\mathbb{E}(X_iX_j) \geq 0$ for all i, j so equation (2.20) is nonnegative.

Now

$$W_n = \mathrm{Tr}(X_n^p) = n^{-p/2} \sum_{1\leq i_1,\ldots,i_p\leq n} X_{i_1i_2}X_{i_2i_3}\ldots X_{i_pi_1}. \tag{2.25}$$

Then since each of the terms above will have positive covariance, for any partition π of any subset of $\{1, \ldots, n\}^p$,

$$\begin{aligned}
\mathrm{Var}(W_n) &= n^{-p}\mathrm{Var}\left(\sum_{1\leq i_1,\ldots,i_p\leq n} X_{i_1i_2}X_{i_2i_3}\ldots X_{i_pi_1}\right) && (2.26)\\
&= n^{-p}\mathrm{Var}\left(\sum_{S\in\pi}\left(\sum_{(i_1,\ldots,i_p)\in S} X_{i_1i_2}X_{i_2i_3}\ldots X_{i_pi_1}\right)\right) && (2.27)\\
&\geq n^{-p}\sum_{S\in\pi}\mathrm{Var}\left(\sum_{(i_1,\ldots,i_p)\in S} X_{i_1i_2}X_{i_2i_3}\ldots X_{i_pi_1}\right) && (2.28)
\end{aligned}$$

Now construct a partition by taking distinct positive integers $1 \leq a_1, a_2, \ldots a_{p-1} \leq \lceil \frac{n}{3p} \rceil$ and let $D_{a_1,\ldots,a_{p-1}}$ be the set of all $1 \leq i_1, \ldots i_p \leq n$ such that $i_{k+1} - i_k = a_k$ for $k = 1, \ldots, p-1$ and $1 \leq i_1 \leq \lceil \frac{n}{3} \rceil$. Then $|D_{a_1,\ldots,a_{p-1}}| = \lceil \frac{n}{3} \rceil$ (since the choice of i_1 fixes all other i_k). Since the a_i's are distinct, $X_{i_ki_{k+1}}$ is independent from $X_{i_ji_{j+1}}$ for all $j \neq k$. Thus

$$\mathrm{Var}\left(\sum_{(i_1,\ldots i_p),\in D} X_{i_1i_2}\ldots X_{i_pi_1}\right) = \sum_{(i_1,\ldots i_p),\in D}\ \sum_{(i_1',\ldots,i_p')\in D} \mathrm{Cov}(X_{i_1i_2}\ldots X_{i_pi_1}, X_{i_1'i_2'}\ldots X_{i_p'i_1'}). \tag{2.29}$$

Bounding the terms in the sum, using independence and the fact that the X_{ij} are centered,

$$\mathrm{Cov}(X_{i_1 i_2}\dots X_{i_p i_1}, X_{i'_1 i'_2}\dots X_{i'_p i'_1}) = \mathbb{E}(X_{i_1 i_2}\dots X_{i_p i_1} X_{i'_1 i'_2}\dots X_{i'_p i'_1}) - \mathbb{E}(X_{i_1 i_2}\dots X_{i_p i_1})\mathbb{E}(X_{i'_1 i'_2}\dots X_{i'_p i'_1}) \tag{2.30}$$

$$= \mathbb{E}(X_{i_1 i_2} X_{i'_1 i'_2})\dots \mathbb{E}(X_{i_p i_1} X_{i'_p i'_1}) \tag{2.31}$$

$$\geq \gamma^p \tag{2.32}$$

Thus (2.29) $\geq \gamma^p |D|^2 \geq \frac{n^2\gamma^p}{9}$.

The number of ways to choose $a_1, \dots, a_{p-1}$ satisfying the restrictions is

$$\lceil \frac{n}{3p} \rceil (\lceil \frac{n}{3p} \rceil - 1)\dots(\lceil \frac{n}{3p} \rceil - p + 2). \tag{2.33}$$

Since we can assume without loss of generality that $n \geq 4p^2$, the above quantity can be lower bounded by $(n/12p)^{p-1}$ (since $\lceil \frac{n}{3p} \rceil - p + 2 \geq \frac{n}{3p} - p \geq \frac{n}{3p} - \frac{n}{4p} = \frac{n}{12p}$). Finally note that if $(a_1, \dots, a_{p-1}) \neq (a'_1, \dots a'_{p-1})$ then $D_{a_1,\dots,a_{p-1}}$ and $D_{a'_1,\dots a'_{p-1}}$ are disjoint. Then applying (2.28),

$$\mathrm{Var}(W_n) \geq n^{-p} \frac{n^{p-1}}{(12p)^{p-1}} \frac{n^2\gamma^p}{9} = Kn. \tag{2.34}$$

□

The last part of the proof is to get bounds on ab, which is done in the next section.

2.3 Bounding the Spectral Norm of a Gaussian Toeplitz Matrix with Correlated Entries

In order to bound the term ab in (2.18), one must first bound the spectral norm of an arbitrary random Toeplitz matrix with correlated entries. In order to bound the spectral norm, the following matrix Khintchine inequality is used from [Tro18]. If $H_1, \dots, H_k$ are fixed Hermitian matrices of common dimension n, and $\gamma_1, \dots, \gamma_k$ are standard normal random variables, then the random matrix

$$X = \sum_{i=1}^{k} \gamma_i H_i \tag{2.35}$$

is called a *Hermitian matrix Gaussian series.* The following result gives bounds on the operator norm of such matrices.

Theorem 2.3.1. *(Corollary 2.4 of [Tro18]) Consider a Hermitian matrix Gaussian series $X = \sum_{i=1}^{k} \gamma_i H_i$ with dimension n. Introduce the matrix standard deviation parameter*

$$\sigma(X) = ||\mathrm{Var}(X)||^{1/2} = \left\| \sum_{i=1}^{k} H_i^2 \right\|^{1/2}. \tag{2.36}$$

Then

$$\frac{1}{\sqrt{2}} \cdot \sigma(X) \leq \mathbb{E}||X|| \leq \sqrt{e(1 + 2\log n)} \cdot \sigma(X) \tag{2.37}$$

were $||\cdot||$ denotes the spectral norm.

With this in mind, we can then get a bound on the spectral norm.

Theorem 2.3.2. *If X_n is an $n \times n$ Toeplitz matrix with correlated entries and λ_n is the spectral norm of X_n, then $\mathbb{E}(\lambda_n) \leq C\sqrt{\log n}$, where C is a constant independent of n.*

Remark 2.3.3. *Note that this theorem holds for any correlated Toeplitz matrix, i.e. there are no restrictions on the covariances $c_k(i,j)$ as long as the corresponding covariance matrices are positive semidefinite.*

The first step in proving the theorem is to write any correlated Toeplitz matrix as a Hermitian matrix Gaussian series. Suppose $\mathbf{X}_k$ is a length $n-k$ vector of jointly Gaussian random variables with mean 0, variance 1, and covariance matrix Σ. If A is a matrix such that $AA^T = \Sigma$ and $\mathbf{Z}_k$ is length $n-k$ vector of i.i.d. standard normal random variables, then $A\mathbf{Z}_k$ has the same distribution as $\mathbf{X}_k$. Thus we can write

$$\mathbf{X}_k = \sum_{j=1}^{n-k} [\mathbf{Z}_k]_j \mathbf{a}_j. \tag{2.38}$$

Here, $[\mathbf{Z}_k]_j$ are independent standard normal Gaussian random variables and $\mathbf{a}_j$ is the jth column of A. We will also denote $\mathbf{a}^j$ as the jth row of A.

Remark 2.3.4. *Since each of the entries of the random matrix have variance at most M, this forces the Euclidean length of each row of A is at most M. This fact will be useful when bounding the operator norm of the matrix standard deviation parameter in the Gaussian series.*

Remark 2.3.5. *Such a decomposition of Σ always exists since Σ is positive semidefinite.*

Thus we can construct any diagonal of a correlated Toeplitz matrix in this manner and constructing the full matrix is just a matter of piecing together the diagonals. For any diagonal vector $\mathbf{X}_k$ of a Toeplitz matrix with correlated standard Gaussian entries, let Σ be its $(n-k) \times (n-k)$ covariance matrix and A be such that $AA^T = \Sigma$. Then for $l = 1, \ldots n-k$, let $B_{n,k,l}$ be an $n \times n$ matrix such that

$$(B_{n,k,l})_{ij} = \begin{cases} 0 & |i-j| \neq k \\ (A)_{min(i,j),l} & |i-j| = k. \end{cases} \tag{2.39}$$

$B_{n,k,l}$ should be thought of as the lth column of A pasted along the kth diagonal of an $n \times n$ matrix, with 0's elsewhere. The A matrices depend on n and k but for the sake of notation we omit the indices. Then the random Toeplitz matrix with correlated entries can be written as

$$X_n = n^{-1/2} \sum_{k=0}^{n-1} \sum_{l=1}^{n-k} Z_{k,l} B_{n,k,l}, \tag{2.40}$$

where each of the $Z_{k,l}$ are i.i.d. standard normal random variables. We are now ready to prove Theorem 2.2.

Proof. (Proof of Theorem 2.3.2)

Again it suffices to prove the theorem under the assumption that the entries are standard Gaussians. The bounds for the general case will just contain extra factors of M, which are independent of n.

In order to prove the theorem, we need to show that $|| \sum_{k=0}^{n-1} \sum_{l=1}^{n-k} B_{n,k,l}^2||^{1/2} \lesssim n^{1/2}$. Since

$$\left\| \sum_{k=0}^{n-1} \sum_{l=1}^{n-k} B_{n,k,l}^2 \right\| \leq \sum_{k=0}^{n-1} \left\| \sum_{l=1}^{n-k} B_{n,k,l}^2 \right\|, \tag{2.41}$$

it suffices to show that for any k, $|| \sum_{l=1}^{n-k} B_{n,k,l}^2|| \leq C$ for some constant C which is independent of n and k (in this case we will see that $C = 4$ works). The reason why such a bound holds is due to the fact stated in Remark 2.3.4. If the $B_{n,k,l}$ were all diagonal matrices (as is the case when $k = 0$), then $\sum_{l=1}^{k} B_{n,k,l}^2 = I_n$ which has operator norm 1. When the $B_{n,k,l}$ are not diagonal the entries "mix" when the matrix is squared, and we need to bound the impact of this mixing on the operator norm of the matrix.

To prove the above heuristics rigorously, we explicitly compute

$$(B_{n,k,l}^2)_{ij} = \sum_{p=1}^{n} (B_{n,k,l})_{ip} (B_{n,k,l})_{pj}. \tag{2.42}$$

First consider the case when $0 < k \leq \frac{n}{2}$. Since $(B_{n,k,l})_{ij}$ is nonzero only when $|i-j| = k$, the term in the sum is nonzero when $|i-p| = |j-p| = k$. This condition can only hold for some p if $|i-j| = 0$ or $|i-j| = 2k$. Thus $(B_{n,k,l}^2)_{ij} = 0$ if $|i-j| \neq 0$ and $|i-j| \neq 2k$. Now consider the case when $i = j$. Again, here A (which depends on n and k) is the matrix associated to $B_{n,k,l}$, i.e. the kth diagonal of $B_{n,k,l}$ is the lth column of A.

$$\begin{align} (B_{n,k,l}^2)_{ii} &= \sum_{p=1}^{n} (B_{n,k,l})_{ip} (B_{n,k,l})_{pi} \tag{2.43} \\ &= (B_{n,k,l})_{i,i+k} (B_{n,k,l})_{i+k,i} \mathbb{1}_{\{i \in [0,n-k]\}} + (B_{n,k,l})_{i,i-k} (B_{n,k,l})_{i-k,i} \mathbb{1}_{\{i \in [k,n]\}} \tag{2.44} \\ &= \begin{cases} ((A)_{il})^2 + ((A)_{i-k,l})^2 & i \in [k, n-k] \\ ((A)_{il})^2 & i \in [0, k] \\ ((A)_{i-k,l})^2 & i \in [n-k, n]. \end{cases} \tag{2.45} \end{align}$$

The last equality comes from equation (2.39).

Now consider the case when $|i-j| = 2k$. Due to symmetry it suffices to consider the case in the upper triangle when $j = i+2k$. So $(B_{n,k,l}^2)_{i,i+2k} = \sum_{p=1}^{n} (B_{n,k,l})_{ip} (B_{n,k,l})_{p,i+2k}$, and for this to be nonzero requires $p = i+k$. Then we get

$$\begin{align} (B_{n,k,l}^2)_{i,i+2k} &= (B_{n,k,l})_{i,i+k} (B_{n,k,l})_{i+k,i+2k} \tag{2.46} \\ &= (A)_{il} (A)_{i+k,l}. \tag{2.47} \end{align}$$

Putting this and the previous result together gives

$$(B^2_{n,k,l})_{ij} = \begin{cases} ((A)_{il})^2 + ((A)_{i-k,l})^2 & i \in [k, n-k], i = j \\ ((A)_{il})^2 & i \in [0, k], i = j \\ ((A)_{i-k,l})^2 & i \in [n-k, n], i = j \\ (A)_{\min(i,j),l}(A)_{\min(i,j)+k,l} & |i-j| = 2k \\ 0 & \text{otherwise.} \end{cases} \tag{2.48}$$

Then summing over l and using Remark 2.3.4,

$$\left(\sum_{l=1}^{n-k} B^2_{n,k,l}\right)_{ij} = \begin{cases} 2 & i \in [k, n-k], i = j \\ 1 & i \in [0, k] \cup [n-k, n], i = j \\ \mathbf{a}^{\min(i,j)} \cdot \mathbf{a}^{\min(i,j)+k} & |i-j| = 2k \\ 0 & \text{otherwise.} \end{cases} \tag{2.49}$$

By the Cauchy-Schwarz inequality and Remark 2.3.4, $|\mathbf{a}^{\min(i,j)} \cdot \mathbf{a}^{\min(i,j)+k}| \leq 1$. Hence by the Gershgorin Circle Theorem,

$$\left\| \sum_{l=1}^{n-k} B^2_{n,k,l} \right\| \leq 4 \tag{2.50}$$

since each Gershgorin disk is centered at 1 or 2 and has radius at most 2.

In the general case with non-unit variances, the above equation becomes

$$\left(\sum_{l=1}^{n-k} B^2_{n,k,l}\right)_{ij} = \begin{cases} 2M & i \in [k, n-k], i = j \\ M & i \in [0, k] \cup [n-k, n], i = j \\ \mathbf{a}^{\min(i,j)} \cdot \mathbf{a}^{\min(i,j)+k} & |i-j| = 2k \\ 0 & \text{otherwise.} \end{cases} \tag{2.51}$$

and $|\mathbf{a}^{\min(i,j)} \cdot \mathbf{a}^{\min(i,j)+k}| \leq M^2$ so

$$\left\| \sum_{l=1}^{n-k} B^2_{n,k,l} \right\| \leq 2M + 2M^2 \tag{2.52}$$

.

Back in the scenario with normalized entries, for the case when $k = 0$, each $B_{n,0,l}$ is a diagonal matrix, so it follows from Remark 2.3 that $\sum_{l=1}^{n} B^2_{n,0,l} = I_n$ and thus $\left\|\sum_{l=1}^{n} B^2_{n,0,l}\right\| = 1$. When $k > \frac{n}{2}$, a similar computation to the case when $k \leq \frac{n}{2}$ can be done, and we see that there is no "mixing" of matrix entries upon squaring. In this case, from equation (2.42), $(B^2_{n,k,l})_{ij}$ is only nonzero when $|i-j| = 0$ or $|i-j| = 2k$. Since $k > \frac{n}{2}$, $|i-j| \neq 2k$ for any i, j, so $B^2_{n,k,l}$ is diagonal. When $i = j$ we have $(B^2_{n,k,l})_{ii} = \sum_{p=1}^{n} (B_{n,k,l})_{ip}(B_{n,k,l})_{pi}$, and for the terms to be nonzero we need $p = i+k$ or $p = i-k$. $p = i+k$ can only happen when $i \in [0, n-k]$ and $p = i-k$ can only happen if $i \in [k, n]$. Thus, from equation (2.39),

$$(B^2_{n,k,l})_{ij} = \begin{cases} ((A)_{il})^2 & i \in [0, n-k], i = j \\ ((A)_{i-k,l})2 & \in i \in [k, n], i = j \\ 0 & \text{otherwise.} \end{cases} \tag{2.53}$$

Summing over l and applying Remark 2.3.4, we see that $\sum_{l=1}^{n-k} B_{n,k,l}^2$ is the identity matrix with some diagonal elements replaced by 0's, and thus $\left\|\sum_{l=1}^{n-k} B_{n,k,l}^2\right\| \leq 1$.

Hence we have shown for any k, $|| \sum_{l=1}^{k} B_{n,k,l}^2|| \leq 4$, and thus $|| \sum_{k=0}^{n-1} \sum_{l=1}^{n-k} B_{n,k,l}^2||^{1/2} \leq 2\sqrt{n}$ as desired.

Then, applying the Matrix Khintchine inequality,

$$\mathbb{E}||X_n|| \leq n^{-1/2}\sqrt{e(1+2\log n)}(2\sqrt{n}) \leq C\sqrt{\log n}$$

for n large enough and some constant C which does not depend on n. □

Lemma 2.3.6. $\mathbb{E}(\lambda_n^k) \leq (Ck\log n)^{k/2}$ *for any n and k, where C is universal constant.*

Proof. This is proved via concentration of measure techniques (see [Led01]). From equation (11.2) in [Kem13],

$$\mu(\{|\lambda_n - \mathbb{E}(\lambda_n)| \geq t\}) \leq 2e^{-t^2/2||F||_{Lip}^2} \tag{2.54}$$

where $||F||_{Lip}$ is the value of the Lipschitz function that gives the spectral norm of A_n. Now we use the layer cake representation. Let $\kappa_k(dt) = kt^{p-1}dt$ on $[0, \infty)$. The corresponding cumulative function is $\phi_k(x) = \int_0^x d\kappa_k = x^k$. Then by Proposition 12.5 of [Kem13] applied to the random variable $X_n = |\lambda_n - \mathbb{E}(\lambda_n)|$,

$$\mathbb{E}(X_n^k) = \int_0^\infty \mathbb{P}(X_n \geq t)kt^{k-1}dt \leq 2k\int_0^\infty t^{k-1}e^{-t^2/||F||_{Lip}^2} \tag{2.55}$$

and then substituting $s = \frac{t}{\sqrt{2}||F||_{Lip}}$ the above becomes

$$\begin{aligned}
2k\left(\sqrt{2}||F||_{Lip}^{k-1}\int_0^\infty s^{k-1}e^{-s^2}ds(\sqrt{2}||F||_{Lip})\right) &= 2k(2||F||_{Lip}^2)^{k/2}\int_0^\infty s^{k-1}e^{-s^2}ds && (2.56)\\
&= C^{k/2}k\int_0^\infty s^{k-1}e^{-s^2}ds && (2.57)\\
&= C^{k/2}k\Gamma\left(\frac{k}{2}\right) && (2.58)\\
&\leq C_1^{k/2}k\left(\frac{k}{2}\right)^{k/2-1} && (2.59)\\
&= C_2^{k/2}k^{k/2}. && (2.60)
\end{aligned}$$

Then (with C possibly changing between inequalities)

$$\begin{aligned}
||\lambda_n||_k &\leq ||\lambda_n - \mathbb{E}(\lambda_n)||_k + ||\mathbb{E}(\lambda_n)||_k && (2.61)\\
&\leq C(k^{1/2} + \sqrt{\log n}) && (2.62)\\
&\leq C(k^{1/2}\sqrt{\log n}) && (2.63)
\end{aligned}$$

where $||\cdot||_k$ denotes the L^k-norm, and the last inequality is for n larger than e^2 and $k \geq 2$ (the case when $k = 1$ is Theorem 2.3.2). □

Now we can finally prove Theorem 3.1.3.

Proof. *(Proof of Theorem 3.1.3)*
Following Proposition 2.2.1, $f_1(x) = px^{p-1}$ and $f_2(x) = p(p-1)x^{p-2}$ so $ab \le p^3(\mathbb{E}(\lambda_n^{4p}))^{1/2}$. Let Z_n be a gaussian random variable with the same mean and variance as W_n. Then by Proposition 2.2.1 and Lemma 2.2.2,

$$d_{TV}(W_n, Z_n) \le \frac{Cp^3(\mathbb{E}(\lambda_n^{4p}))^{1/2}\sqrt{n}}{\mathrm{Var}(W_n)} \tag{2.64}$$

where C is a universal constant. From Lemma 2.3.6, the $p^3(\mathbb{E}(\lambda_n^{4p}))^{1/2}$ term is bounded by $p^3(Cp\log n)^p$. Incorporating Lemma 2.2.4,

$$d_{TV}(W_n, Z_n) \le \frac{C^p p^{p+3}(\log n)^p}{\sqrt{n}} \tag{2.65}$$

and this goes to 0 as $n \to \infty$. Note that in this final step, the constant C now depends on p (it contains a factor of γ^{-p}). Centering and normalizing proves the theorem. $\square$

2.4 Fluctuations for Other Matrix Models with Correlated Entries

The arguments and methods used above to prove the central limit theorem for generalized Toeplitz matrices extend to other matrix models. In [AS17], Adhikari and Saha used Chatterjee's total variation bound to prove fluctuations results for circulant, symmetric circulant, reverse circulant, and Hankel matrices. In this section, we extend these results to the four corresponding matrix models that allow for correlated entries. The proofs for each of these follow the same structure by computing upper bounds on $||\Sigma||$ and $||X_n||$ and lower bounds on $\mathrm{Var}(W_n)$. The bound for Σ is essentially the same for all of the models (including the Toeplitz matrix with correlated entries). Furthermore, in this section we will assume that all entries of the random matrices have unit variance. The proofs can be adapted to the general case in the same way as for the Toeplitz matrices with correlated entries.

Lemma 2.4.1. *Let X_n be an $n \times n$ circulant, reverse circulant, symmetric circulant, or Hankel matrix with standard Gaussian correlated entries. Let Σ denote its covariance matrix. Then $||\Sigma|| \le 2n$.*

Proof. For any i, j, $(X_n)_{ij}$ is correlated with at most $2n - 1$ other entries of X_n. Then by the Gershgorin circle theorem,

$$||\Sigma|| \le 1 + (2n-1) \sup_{k, i \neq j} c_k(i,j) \le 2n. \tag{2.66}$$

$\square$

2.4.1 Reverse Circulant Matrices with Correlated Entries

The proof structure is similar to that of the proof for the random Toeplitz model with correlated entries using Proposition 2.2.1.

Lemma 2.4.2. *With W_n as in Theorem 3.1.3 and X_n a reverse circulant matrix with correlated entries,* $\mathrm{Var}(W_n) \geq Kn$ *for some constant K that only depends on p and γ.*

To prove the lemma we adapt the proof of the bound for the variance in Lemma 10 of [AS17]. We can use the same partition argument and incorporate the bound γ on the covariances. The following proof is thus a version of the proof of Lemma 10 in [AS17], with the appropriate adjustments to allow for correlated entries.

Proof. Again letting $X_{ij} \coloneqq (X_n)_{ij}$, we have

$$W_n = \mathrm{Tr}(X_n^p) = n^{-p/2} \sum_{1 \leq i_1, \ldots, i_p \leq n} X_{i_1 i_2} X_{i_2 i_3} \ldots X_{i_p i_1}. \tag{2.67}$$

Showing that all of the terms in the above sum are positively correlated is done in a similar way to the proof of Lemma 2.2.4, except the products on the right hand side of equation (2.20) factor with the condition of $(j + i - 1) \mod n$ instead of $j - i$. Then the same argument using Wick's theorem shows that the terms in the above sum are positively correlated.

Then for any partition π of any subset of $\{1, \ldots, \frac{n}{3}\}^p$,

$$\mathrm{Var}(W_n) \geq n^{-p} \sum_{S \in \pi} \mathrm{Var}\left(\sum_{(i_1, \ldots i_p) \in S} X_{i_1 i_2} X_{i_2 i_3} \ldots X_{i_p i_1} \right). \tag{2.68}$$

Now we construct an appropriate partition following the proof of Lemma 10 in [AS17]. Let

$$\mathcal{A} = \left\{ (a_1, \ldots, a_p) \in \mathbb{N}^p : \frac{kn}{3p} + 1 \leq a_k \leq \frac{(k+1)n}{3p}, k = 1, 2, \ldots, p-1 \right\} \tag{2.69}$$

and define

$$D_{a_1, \ldots, a_p} = \{(i_1, \ldots, i_p) : 1 \leq i_1 \leq \frac{n}{3p}, i_k + i_{k+1} - 1 = a_k, k = 1, \ldots, p\}, \tag{2.70}$$

where $i_{p+1} = i_1$ and $(a_1, \ldots, a_p) \in \mathcal{A}$. Note that the a_i's are distinct and because of this $X_{i_k i_{k+1}}$ is independent from $X_{i_j i_{j+1}}$ for $j \neq k$. Thus given some $D_{a_1, \ldots, a_p}$ (which we denote as D in the following computation),

$$\mathrm{Var}\left(\sum_{(i_1, \ldots i_p) \in D} X_{i_1 i_2} \ldots X_{i_p i_1} \right) = \sum_{(i_1, \ldots i_p) \in D} \sum_{(i'_1, \ldots i'_p) \in D} \mathrm{Cov}(X_{i_1 i_2} \ldots X_{i_p i_1}, X_{i'_1 i'_2} \ldots X_{i'_p i'_1}). \tag{2.71}$$

Then following equation (2.30),

$$\mathrm{Cov}(X_{i_1 i_2} \ldots X_{i_p i_1}, X_{i'_1 i'_2} \ldots X_{i'_p i'_1}) \geq \gamma^p, \tag{2.72}$$

so

$$\operatorname{Var}\left(\sum_{(i_1,\ldots i_p)\in D} X_{i_1 i_2}\ldots X_{i_p i_1}\right) \geq \gamma^p |D|^2. \tag{2.73}$$

Note that if $(a_1,\ldots,a_{p-1}) \neq (a'_1,\ldots a'_{p-1})$ then $D_{a_1,\ldots,a_{p-1}}$ and $D_{a'_1,\ldots a'_{p-1}}$ are disjoint. Then from equation (2.68),

$$\begin{aligned}
\operatorname{Var}(W_n) &\geq n^{-p} \sum_{(a_1,\ldots,a_p)\in\mathcal{A}} \operatorname{Var}\left(\sum_{(i_1,\ldots,i_p)\in D} X_{i_1 i_2}\ldots X_{i_p i_1}\right) && (2.74)\\
&\geq n^p |\mathcal{A}||D|^2 \gamma^p && (2.75)\\
& && (2.76)
\end{aligned}$$

For a fixed value of a_p, $|\mathcal{A}| = (\frac{n}{3p})^{p-1}$. From the definition of $a_1,\ldots,a_p$, we have

$$i_p = \begin{cases} a_{p-1} - a_{p-2} + \cdots - a_2 + a_1 - i_1 + 1 & \text{if } p \text{ is even,} \\ a_{p-1} - a_{p-2} + \cdots + a_2 - a_1 + i_1 & \text{if } p \text{ is odd.} \end{cases} \tag{2.77}$$

This implies that $a_p = i_p + i_1 = a_{p-1} - a_{p-2} + \cdots - a_2 + a_1$, when p is even. Thus if p is even then a_p is determined by $a_1,\ldots a_{p-1}$ and it does not depend on i_1, but if p is odd then a_p depends on $a_1,\ldots,a_2,\ldots a_{p-1}$ and i_1.

Thus when p is even, for a fixed choice of $a_1,\ldots a_{p-1}$ the number of elements in $D_{a_1,\ldots a_p}$ is the same as the number of ways of choosing i_1, so

$$|D_{a_1,\ldots a_p}| = \frac{n}{3p} \tag{2.78}$$

when p is even. Thus, when p is even

$$\operatorname{Var}(W_n) \geq n^{-p} \left(\frac{n}{3p}\right)^{p-1} \left(\frac{n}{3p}\right)^2 = Kn \tag{2.79}$$

where K only depends on γ and p. □

The last part of the proof of Theorem 3.1.3 for reverse circulant matrices is a bound on the operator norm.

Lemma 2.4.3. *If X_n is an $n \times n$ reverse circulant matrix with correlated entries and λ_n is the spectral norm of X_n, then $\mathbb{E}(\lambda_n) \leq C\sqrt{\log n}$, where C is a constant independent of n.*

In order to prove the above Lemma, we need to be slightly more careful than in the proof of Theorem 2.3.2. Note that Theorem 2.3.1 requires the matrices in the Gaussian series to be self-adjoint. However, generalized reverse circulant matrices with correlated entries are not symmetric, and thus neither will be the coefficient matrices from the construction in the beginning of Section 2.3.

We can get around this via the following dilation argument. If A is an $n \times n$ matrix (not necessarily self-adjoint), then its *dilation* is

$$\tilde{A} := \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}. \tag{2.80}$$

Here, we are only considering real matrices, so $A^* = A^T$. Then since $\tilde{A}$ is a $2n \times 2n$ self-adjoint matrix we have

$$||\tilde{A}||^2 = ||\tilde{A}^2|| = \max\{||A^T A||, ||AA^T||\} = ||A||^2 \tag{2.81}$$

where $||\cdot||$ denotes the operator norm. Hence we will use the dilations of the coefficient matrices in the matrix series to compute the bound on the operator norm via matrix concentration inequalities.

Proof. We construct the matrices $B_{n,k,l}$ in a similar manner as done in section 3. Recall that n is the size of the matrix, k labels the Gaussian vector $\mathbf{X}_k$, and l corresponds to the lth column of A, which is such that $AA^T = \Sigma$ where Σ is the covariance matrix of $\mathbf{X}_k$. Again, A depends on k and n, but we forgo the labeling for notation purposes. We then have

$$(B_{n,k,l})_{ij} = \begin{cases} 0 & i+j-1 \not\equiv k \mod n \\ A_{il} & i+j-1 \equiv k \mod n \end{cases} \tag{2.82}$$

Then we can write

$$X_n = n^{-1/2} \sum_{k=1}^{n} \sum_{l=1}^{n} Z_{k,l} B_{n,k,l} \tag{2.83}$$

where the $Z_{k,l}$ are standard normal random variables. Then applying the dilations,

$$\tilde{X_n} = n^{-1/2} \sum_{k=1}^{n} \sum_{l=1}^{n} Z_{k,l} \widetilde{B_{n,k,l}}. \tag{2.84}$$

Hence we must now show $\left\| \sum_{k=1}^n \sum_{l=1}^n \widetilde{B_{n,k,l}}^2 \right\|^{1/2} \lesssim n^{1/2}$. Now

$$\widetilde{B^2_{n,k,l}} = \begin{bmatrix} B_{n,k,l}B^T_{n,k,l} & 0 \\ 0 & B^T_{n,k,l}B_{n,k,l} \end{bmatrix} \tag{2.85}$$

so we compute

$$(B_{n,k,l}B^T_{n,k,l})_{ij} = \sum_{i=1}^{n} (B_{n,k,l})_{ip}(B_{n,k,l})_{jp} \tag{2.86}$$

and for this to be nonzero, from equation (2.82) we need $i+p-1 \equiv k \mod n$ and $j+p-1 \equiv k \mod n$ which forces $i=j$, so $B_{n,k,l}B^T_{n,k,l}$ and $B^T_{n,k,l}B_{n,k,l}$ are diagonal. Then from equation (2.82),

$$(B_{n,k,l}B^T_{n,k,l})_{ii} = A^2_{il} \tag{2.87}$$

and similarly for $B^T_{n,k,l}B_{n,k,l}$. Then by the fact that the rows of A have length 1, $\sum_{l=1}^n \widetilde{B_{n,k,l}}^2 = I_{2n\times 2n}$, and it follows that $\left\| \sum_{k=1}^n \sum_{l=1}^n \widetilde{B_{n,k,l}}^2 \right\|^{1/2} \leq n^{1/2}$ (in fact we have equality here). $\square$

2.4.2 Circulant Matrices with Correlated Entries

The proof of the theorem again involves two parts along with Lemma 2.4.1.

Lemma 2.4.4. *With W_n as in Theorem 3.1.3 and X_n a circulant matrix with correlated entries,* $\mathrm{Var}(W_n) \geq Kn$ *for some constant K that only depends on p and γ.*

Proof. This can be proved in similar manner to Lemma 2.2.4. Notice that the upper triangle of a circulant matrix with correlated entries is identical to that of a Toeplitz matrix with correlated entries. The argument form Chatterjee used in Lemma 2.2.4 only involves a partition involving the upper triangle (since $a_1, \ldots, a_{p-1}$ are all positive and $i_{k+1} - i_k = a_k$), and the only point where the lower triangle is involved is the term $X_{i_p i_1}$. However, the bound in equation (2.30) still holds (up to a constant factor) even if $X_{i_p i_1}$ is correlated with one other $X_{i_k i_{k+1}}$. To see this, suppose $X_{i_p i_1}$ is not independent of $X_{i_k i_{k+1}}$ for some k. Then

$$\mathrm{Cov}(X_{i_1 i_2} \ldots X_{i_p i_1}, X_{i'_1 i'_2} \ldots X_{i'_p i'_1}) = \mathbb{E}(X_{i_1 i_2} \ldots X_{i_p i_1} X_{i'_1 i'_2} \ldots X_{i'_p i'_1}) - \mathbb{E}(X_{i_1 i_2} \ldots X_{i_p i_1})\mathbb{E}(X_{i'_1 i'_2} \ldots X_{i'_p i'_1}) \tag{2.88}$$

$$= \mathbb{E}(X_{i_1 i_2} X_{i'_1 i'_2}) \ldots \mathbb{E}(X_{i_p i_1} X_{i'_p i'_1} X_{i_k i_{k+1}} X_{i'_k i'_{k+1}}) \tag{2.89}$$

$$\geq \gamma^{p-2}(3\gamma^2) \tag{2.90}$$

$$= 4\gamma^p. \tag{2.91}$$

The second equality comes from independence. If $p \geq 3$, then $X_{i_1 i_2}$ is independent of $X_{i_2 i_3}$, and in the case when $p = 2$, $X_{i_1 i_2}$ is independent of $X_{i_2 i_1}$ due to the lack of symmetry of circulant matrices. The first inequality comes from applying Wick's theorem to the expectation of the four Gaussians in the line above.

Thus following the rest of the proof of the correlated Toeplitz case completes the proof of the Lemma.

□

Lemma 2.4.5. *If X_n is an $n \times n$ circulant matrix with correlated entries and λ_n is the spectral norm of X_n, then $\mathbb{E}(\lambda_n) \leq C\sqrt{\log n}$, where C is a constant independent of n.*

Proof. This is proved via matrix concentration inequalities and a dilation argument. The corresponding $B_{n,k,l}$ matrices are defined by

$$(B_{n,k,l})_{ij} = \begin{cases} 0 & j - i + 1 \not\equiv k \mod n \\ A_{il} & j - i + 1 \equiv k \mod n \end{cases}. \tag{2.92}$$

A similar computation to the reverse circulant case shows that

$$(B_{n,k,l} B_{n,k,l}^T)_{ii} = A_{il}^2. \tag{2.93}$$

Since each of the rows of A have length 1, $\sum_{l=1}^n \widetilde{B_{n,k,l}}^2 = I_{2n \times 2n}$ where $\widetilde{B_{n,k,l}}$ is the dilation

$$\widetilde{B_{n,k,l}} = \begin{bmatrix} 0 & B_{n,k,l}^T \\ B_{n,k,l} & 0 \end{bmatrix}. \tag{2.94}$$

It then follows that $\left\| \sum_{k=1}^n \sum_{l=1}^n \widetilde{B_{n,k,l}}^2 \right\|^{1/2} \leq n^{1/2}$ and thus the lemma is proved. □

2.4.3 Symmetric Circulant Matrices with Correlated Entries

The following two Lemmas complete the proof.

Lemma 2.4.6. *With W_n as in Theorem 3.1.3 and X_n a symmetric circulant matrix with correlated entries,* $\mathrm{Var}(W_n) \geq Kn$ *for some constant K that only depends on p and γ.*

Proof. Note that when $|i-j| \leq \frac{n}{2}$, $X_{ij} = [\mathbf{X}_{|i-j|}]_{\min(i,j)}$ just as in the Toeplitz case. The argument adapted from [Cha09] used in the Toeplitz case only considers partitions where $|i_k - i_{k-1}| \leq \lceil \frac{n}{3p} \rceil$, so thus the argument applies in the case for symmetric circulant matrices with correlated entries, and we get $\mathrm{Var}(W_n) \geq Kn$ for the same constant K as the Toeplitz case. □

Lemma 2.4.7. *If X_n is an $n \times n$ symmetric circulant matrix with correlated entries and λ_n is the spectral norm of X_n, then $\mathbb{E}(\lambda_n) \leq C\sqrt{\log n}$, where C is a constant independent of n.*

Proof. Note that for any symmetric circulant matrix X_n, we can write $X_n = Y_n + Z_n$ where Y_n and Z_n are Toeplitz matrices with correlated entries with diagonals 0 to $\lceil \frac{n}{2} \rceil$ replaced by 0's in Z_n and diagonals $\lceil \frac{n}{2} \rceil + 1$ to $n-1$ replaced by 0's in Y_n. Then applying the same argument via matrix concentration inequalities as in the Toeplitz case to Y_n and Z_n yields

$$||Y_n|| \leq C\sqrt{\log n} \qquad \text{and} \qquad ||Z_n|| \leq C\sqrt{\log n} \tag{2.95}$$

where C is a constant independent of n. It then follows that $||X_n|| \leq 2C\sqrt{\log n}$ as desired. □

2.4.4 Hankel Matrices with Correlated Entries

We then have the following two Lemmas needed to complete the proof.

Lemma 2.4.8. *With W_n as in Theorem 3.1.3 and X_n a Hankel matrix with correlated entries,* $\mathrm{Var}(W_n) \geq Kn$ *for some constant K that only depends on p and γ.*

Proof. We can apply the same combinatorial argument as in the reverse circulant case here. □

Lemma 2.4.9. *If X_n is an $n \times n$ Hankel matrix with correlated entries and λ_n is the spectral norm of X_n, then $\mathbb{E}(\lambda_n) \leq C\sqrt{\log n}$, where C is a constant independent of n.*

Proof. Any Hankel matrix can be viewed as the the first $n \times n$ block of a $2n \times 2n$ reverse circulant matrix with correlated entries. Thus let H_n be any Hankel matrix with correlated entries and RC_{2n} be a reverse circulant matrix with its first $n \times n$ block being H_n. Then

$$||H_n|| \leq ||RC_{2n}||. \tag{2.96}$$

Then by Lemma 2.4.3,

$$\mathbb{E}||H_n|| \leq \mathbb{E}||RC_{2n}|| \leq C\sqrt{\log(2n)} \leq \tilde{C}\sqrt{\log n}. \tag{2.97}$$

□

2.5 Fluctuations of Eigenvalues with Correlation Decay

In this section we prove Theorem 2.1.3. The proof follows the same general structure as that of Theorem 3.1.3 by using Proposition 2.2.1. In order to prove the theorem as stated, we need tighter bounds on the operator norm. These bounds result from the sharper matrix concentration inequalities in [BBH23].

The set up is as follows. As in Theorem 2.3.1, let

$$X = \sum_{i=1}^{k} \gamma_i H_i \tag{2.98}$$

where $\gamma_1, \dots, \gamma_k$ are independent standard normal random variables, and $H_1, \dots, H_k$ are nonrandom Hermitian matrices of common dimension n, and again let

$$\sigma(X) = ||\mathbb{E}(X^2)|| = \left\| \sum_{i=1}^{k} H_i^2 \right\|. \tag{2.99}$$

Further, define

$$\nu(X) = ||\mathrm{Cov}(X)|| \tag{2.100}$$

where $\mathrm{Cov}(X)$ is viewed as a matrix in $M_{n^2 \times n^2}(\mathbb{R})$ and $\mathrm{Cov}(X)_{ij,kl} = \mathrm{Cov}(X_{ij}, X_{kl})$. We then have the following result, which is equation (1.11) in [BBH23] (a corollary of their Theorem 1.2).

Theorem 2.5.1. *With X, $\sigma(X)$, and $\nu(X)$ defined as above, $\mathbb{E}(||X||) \lesssim \sigma(X) + \nu(X)(\log n)^{3/2}$.*

Remark 2.5.2. *Here $\nu(X)$ involves the correlation matrix for the random matrix X, i.e. it includes the $n^{-1/2}$ normalization for each of the entries. This is in contrast to Chatterjee's statement in Proposition 2.2.1 in which we compute the operator norm of the covariance matrix of the Gaussian input sequence before we rescale by $n^{-1/2}$.*

Using the above theorem, we have following Lemma, which states that if we have any polynomial decay in the correlations, then the operator norm of any of the correlated models converges to a finite limit in expectation.

Lemma 2.5.3. *Let X_n be a random Toeplitz, circulant, reverse circulant, symmetric circulant, or Hankel matrix with correlated entries with standard Gaussian entries with covariances along each diagonal labeled by $c_k(i,j)$. Further, assume $c_k(i,j) = n^{-\alpha}$ for some $\alpha > 0$ for all i, j, k with $i \neq j$. Then $\mathbb{E}(||X_n||)$ converges to a finite number as $n \to \infty$.*

Proof. In Theorem 2.3.2 and Lemmas 2.4.3, 2.4.5, 2.4.7, and 2.4.9, we showed that $\sigma(X)$ is bounded by a constant. In all of the models, each entry of X_n is correlated with at most $2n - 1$ other entries. Hence by the Gershgorin circle theorem,

$$\nu(X) \leq \frac{1}{n} \left(1 + (2n-1) \sup_{k, i \neq j} c_k(i,j) \right). \tag{2.101}$$

The n^{-1} factor comes from including the normalization factor $n^{-1/2}$ in $\mathrm{Cov}(X)$. Since $c_k(i,j) = o(n^{-\alpha})$, $\nu(x) = o(n^{-\alpha})$, and thus from Theorem 2.5.1, $\mathbb{E}(||X_n||) \leq C + o(n^{-\alpha})$. $\square$

Remark 2.5.4. *Theorem 2.5.1 and Lemma 2.5.3 is only needed to get rid of the* $\sqrt{\log n}$ *from the operator norm of the correlated matrix models. Theorem 2.1.3 holds without the result of Lemma 2.5.3 if we restrict the decay in covariances to* $c_k(i,j) = o(n^-\alpha)$ *for any* $\alpha > \frac{1}{3}$.

Remark 2.5.5. *Combining the above Lemma with the concentration arguments in Lemma 2.3.6 shows that the operator norm of these correlated models has a sub-Gaussian distribution when there is any polynomial decay in the correlations among the entries. For more about sub-Gaussian random variables, see section 2.5 of [Ver18].*

Proof. (Proof of Theorem 2.1.3)

The following proof works for any of the patterned matrix models with correlations considered in this paper. Let X_n be a random Toeplitz, circulant, reverse circulant, symmetric circulant, or Hankel matrix with correlated entries with standard Gaussian entries with covariances along each diagonal labeled by $c_k(i,j)$. Let p be a positive integer. The structure of the proof is similar to that of the case where the correlations are bounded away from 0 in that we show the fraction on the right hand side of equation (2.18) converges to 0.

From the computations in Sections 2.3 and 2.4, we know that the expectation of the operator norm of X_n is at most $\sqrt{\log n}$. Thus from Lemma 2.3.6 and Lemma 2.5.3, the term ab in equation (2.18) is bounded by $p^{p+3}C^p$ where C is independent of n.

Next we bound σ^2 from below by a constant. Recall from the proof of Lemma 2.2.4 that the covariance of arbitrary products of the entries of X_n is nonnegative. Thus

$$\mathrm{Var}(W_n) = n^{-p}\mathrm{Var}\left(\sum_{1\le i_1,\dots,i_p\le n} X_{i_1i_2}X_{i_2i_3}\dots X_{i_pi_1}\right) \tag{2.102}$$

$$\ge n^{-p}\sum_{1\le i_1,\dots,i_p\le n} \mathrm{Var}(X_{i_1i_2}X_{i_2i_3}\dots X_{i_pi_1}). \tag{2.103}$$

In order to apply a partition argument via Wick's Theorem to the variances in the sum, it will be again easier to relabel the $2p$ random variables $X_{i_1i_2}, X_{i_2i_3}, \dots, X_{i_pi_1}, X_{i_1i_2}, X_{i_2i_3}, \dots, X_{i_pi_1}$ by enumerating them from 1 to $2p$. Further, define the partition $\tau := \{\{1,p+1\},\{2,p+2\},\dots,\{p,2p\}\}$. Then we have

$$\mathrm{Var}(X_{i_1i_2}X_{i_2i_3}\dots X_{i_pi_1}) = \sum_{\pi\in P_2(2p)}\prod_{\{i,j\}\in\pi}\mathbb{E}(X_iX_j) \tag{2.104}$$

$$-\left(\sum_{\pi\in P_2(p)}\prod_{\{i,j\}\in\pi}\mathbb{E}(X_iX_j)\right)\left(\sum_{\pi\in P_2(p)}\prod_{\{i,j\}\in\pi}\mathbb{E}(X_{i+p}X_{j+p})\right) \tag{2.105}$$

$$\tag{2.106}$$

Now note that every term in the double sum on the right appears in the double sum on the left (when p is odd the terms on the right are 0). However, τ does not appear in the double sum due its blocks crossing between the sets $\{1,\dots,p\}$ and $\{p+1,\dots,2p\}$. Hence

$$\mathrm{Var}(X_{i_1i_2}X_{i_2i_3}\dots X_{i_pi_1}) \ge \prod_{\{i,j\}\in\tau}\mathbb{E}(X_iX_j) = 1 \tag{2.107}$$

and it follows that

$$\sigma^2 = \mathrm{Var}(W_n) \geq 1. \tag{2.108}$$

The last part of the proof is to bound the operator norm of the covariance matrix Σ. Following the proof of Lemma 2.4.1,

$$||\Sigma|| \leq 1 + (2n-1) \sup_{k, i \neq j} c_k(i,j). \tag{2.109}$$

Under the assumption that $c_k(i,j) = o(n^{-1/3})$, $||\Sigma|| = o(n^{2/3})$ and $||\Sigma||^{3/2} = o(n)$. Then combining this with the lower bound for the variance and upper bound on ab via Lemmas 2.5.3 and 2.3.6, and plugging into equation (2.18) we get

$$d_{TV}(W_n, Z_n) = o(1). \tag{2.110}$$

□

2.6 Towards Universality

2.6.1 Removing the Conditions on the Covariances

One obvious extension of Theorem 3.1.3 and Theorem 2.1.3 is to allow for completely arbitrary correlation structures. Hence we have the following conjecture.

Conjecture 2.6.1. *Let X_n be an $n \times n$ random Toeplitz, circulant or symmetric circulant matrix with correlated standard Gaussian entries. Let $p \geq 2$ be a positive integer and let $W_n = \mathrm{Tr}(X_n^p)$. Then as $n \to \infty$,*

$$\frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} \text{ converges in total variation to } N(0,1). \tag{2.111}$$

In the case when X_n is a reverse circulant or Hankel matrix with correlated entries, the theorem still holds under the further assumption that p is restricted to be an even positive integer.

In the proofs of Theorems 3.1.3 and 2.1.3 the fraction in equation (2.18) vanishes as $n \to \infty$ for different reasons. Hence in order to prove Conjecture 2.6.1, one would need to adopt new methods to more precisely simultaneously bound all three terms that we bounded in our proofs, or an entirely new approach in general. Furthermore, there is the fact that Theorem 2.1.3 holds for Hankel matrices with correlated entries when p is odd, but the linear eigenvalue statistics of a Hankel matrix for odd p converge to a non-Gaussian limit (see Theorem 2 of [SMS22]). Hence there is most likely a phase transition for the limiting statistics of odd monomial test functions for linear eigenvalue statistics of Hankel matrices, and a similar phenomenon may occur in the case of reverse circulant matrices. We also did not consider correlated models with negative correlations.

2.6.2 Moving Beyond Gaussian Entries

In [AS18], Adhikari and Saha were able to prove a Gaussian central limit theorem for circulant matrices with sub-Gaussian entries. They further assumed that the entries belonged to a certain class of random variables whose laws can be written as C^2 functions with bounded derivatives of a standard Gaussian, denoted $\mathcal{L}(c_1, c_2)$ where c_1 and c_2 are the respective bounds for the first and second derivatives (see Definition 2.1 of [Cha09]). There are multiple equivalent definitions of sub-Gaussian random variables (see Proposition 2.5.2 of [Ver18]), and one such definition is the following.

Definition 2.6.2. *A random variable X is said to be sub-Gaussian if there exists a constant K such that the tails of x satisfy*

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/K^2} \quad \textit{for all } t \geq 0. \tag{2.112}$$

We conjecture that similar extensions to correlated models studied with sub-Gaussian entries hold due to the sub-Gaussian tails inducing similar $\sqrt{\log n}$ bounds on the operator norm. Meckes proved in [Mec07] that the operator norm of a random Toeplitz matrix (and other models considered in this paper) with sub-Gaussian entries is asymptotically bounded above by $\sqrt{\log n}$. Due to concentration of measure, the uncorrelated models have the most dependence among their entries and thus are the "worst-case" scenario for bounding the operator norm. Thus it is natural to conjecture that for any of the correlated models in this paper with sub-Gaussian entries, the operator norm is asymptotically bounded above by $\sqrt{\log n}$. We then have the following conjecture.

Conjecture 2.6.3. *Let X_n be an $n \times n$ random Toeplitz, circulant or symmetric circulant matrix with correlated symmetric standard (mean 0 variance 1) sub-Gaussian entries. Let $p \geq 2$ be a positive integer and let $W_n = \mathrm{Tr}(X_n^p)$. Then as $n \to \infty$,*

$$\frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} \textit{ converges in total variation to } N(0, 1). \tag{2.113}$$

In the case when X_n is a reverse circulant or Hankel matrix with correlated entries, the theorem still holds under the further assumption that p is restricted to be an even positive integer.

In order to prove this, new methods will most likely need to be used. Proposition 2.2.1 from [Cha09] only holds for Gaussian entries, though Theorems 2.2 and 3.1 of the same paper can be applied to more general distributions of the entries (and this was how Adhikari and Saha proved their sub-Gaussian universality result in [AS18]). However, these results only hold under the assumption that each entry of the matrix can be written as a function of independent random variables. It is not straightforward to construct a random matrix with a general correlation structure and sub-Gaussian entries in this manner.

Chapter 3

Matrix Concentration Inequalities with Sub-Gaussian Coefficients

3.1 Introduction

In this paper we prove a matrix concentration inequality for matrix series with sub-Gaussian coefficients.

Definition 3.1.1. *A random variable X is said to be sub-Gaussian if there exists a constant κ such that $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/\kappa^2}$.*

In other words, the tails of the CDF decay faster than a Gaussian. However, for the purposes of this paper, it will be easier to use the following equivalent definition, which states the sub-Gaussian condition in terms of moment bounds of the random variable.

Definition 3.1.2. *A random variable X is said to be sub-Gaussian if there exists a constant K such that $\mathbb{E}|X^p| \leq K^p p^{p/2}$. We will refer to such a random variable as sub-Gaussian with moment bound K.*

In the above definition, it is not hard to prove (via the layercake representation) that $K \leq 3\kappa$ where κ is from Definition 3.1.1. For a full proof of this fact and the equivalence of the two definitions, see Proposition 2.5.2 in [Ver18].

The class of sub-Gaussian random variables is quite large. It contains Gaussians, Bernoulli, Rademacher, and any uniform random variable on a compact set. More generally, any bounded random variable is sub-Gaussian.

Hence in this paper, the object of study is random matrices with sub-Gaussian entries, and we will prove a concentration inequality for the operator norm. For a general overview of the field of matrix concentration inequalities, see [Tro15] and the references therein. Much of the work in the study of matrix concentration for matrix series has been dealing with the important examples where the coefficients are Gaussian or Rademacher random variables. However, to the best of our knowledge, there are no results for the general case of sub-Gaussian entries.

Hence we have the following theorem, which bounds the operator norm of self-adjoint matrix series with sub-Gaussian coefficients.

Theorem 3.1.3. *Let $A_1, \ldots, A_n \in M_{d\times d}(\mathbb{R})$ be fixed (nonrandom) Hermitian matrices, and let $\gamma_1, \ldots, \gamma_n$ be independent symmetric sub-Gaussian random variables with uniform moment bound K. Then if $X = \sum_{i=1}^n \gamma_i A_i$,*

$$\mathbb{P}(\lambda_{\max}(X) \geq t) \leq de^{-t^2/8K^2\sigma^2} \tag{3.1}$$

where $\lambda_{\max}$ is the largest eigenvalue of X. In particular,

$$\mathbb{P}(||X|| \geq t) \leq 2de^{-t^2/8K^2\sigma^2} \tag{3.2}$$

This theorem can also be viewed as a matrix version of Proposition 5 in [Mec07]. Much like how Proposition 5 of [Mec07] can be proved via the Laplace transform, Theorem 3.1.3 will be proved with the matrix Laplace transform. We also have the following corollary, which follows easily from Theorem 3.1.3 and the layercake representation.

Corollary 3.1.4. *With X and σ defined as above, we have*

$$\mathbb{E}[||X||] \leq (2^{7/2}K\sigma)\sqrt{\log(2ed)}. \tag{3.3}$$

In this case, the entries of X can also by asymmetric. In the case where they are symmetric, the coefficient $2^{7/2}$ can be reduced to $2^{3/2}$.

In Section 3.2, we prove Theorem 3.1.3 via the matrix Laplace transform, and we also prove Corollary 3.1.4. In Section 3.3, we apply Corollary 3.1.4 to give simple proofs for bounds on the operator norm of various patterned random matrix models with sub-Gaussian entries.

3.2 Proof of Main Theorem and Corollary

The proof is an adaptation of the argument to prove Theorem 4.1 in [Tro12]. The moment bounds from the sub-Gaussian condition allow us to adapt the matrix Laplace transform argument used for Gaussian matrix series. In order to prove the Theorem, we need the following result, which is Corollary 3.7 of [Tro12].

Proposition 3.2.1. *(Corollary 3.7 of [Tro12]) Consider a finite sequence $\{\mathbf{X}_k\}$ of independent, random, self-adjoint matrices with dimension d. Assume there is a function $g : (0, \infty) \to [0, \infty]$ and a sequence $\{\mathbf{A}_k\}$ of self-adjoint matrices that satisfy the relations*

$$\mathbb{E}\left(e^{\theta \mathbf{X}_k}\right) \preccurlyeq e^{g(\theta)\mathbf{A}_k} \quad \text{for } \theta > 0. \tag{3.4}$$

Define the scale parameter

$$\rho := \lambda_{\max}\left(\sum_k \mathbf{A}_k\right). \tag{3.5}$$

Then, for all $t \in \mathbb{R}$,

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \mathbf{X}_k\right)\right) \leq d \inf_{\theta>0} e^{-\theta t + g(\theta)\rho}. \tag{3.6}$$

The following Lemma will help with commuting the semidefinite partial ordering with limits of sequences of matrices.

Lemma 3.2.2. *Let $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$ be sequences of $d \times d$ self-adjoint matrices such that $A_n \preccurlyeq B_n$ for all n. If $A_n \to A$ and $B_n \to B$, then $A \preccurlyeq B$.*

Proof. We need to show $B - A \succcurlyeq 0$, i.e. for any $x \in R^d$, $x^T(B-A)x \geq 0$. Since all matrix norms induce the same topology, $A_n \to A$ and $B_n \to B$ with respect to all matrix norms. Then

$$\begin{aligned} x^T(B-A)x &= x^TBx - x^TAx && (3.7)\\ &= \lim_{n\to\infty} x^TB_nx - x^TA_nx && (3.8)\\ &= \lim_{n\to\infty} (x^T(B_n - A_n)x) && (3.9)\\ &\geq 0 \text{ for any } x \in \mathbb{R}^d. && (3.10) \end{aligned}$$

The second equality is due to convergence in the Hilbert-Schmidt norm, since $(A_n)_{ij} \to (A)_{ij}$ and $(B_n)_{ij} \to (B)_{ij}$ for all $1 \leq i, j \leq d$. □

With the above Lemma we can prove the following, which states that the relation from Proposition 3.2.1 in equation (3.4) is satisfied in the sub-Gaussian case.

Lemma 3.2.3. *Suppose that $\mathbf{A}$ is a self-adjoint matrix. Let γ be symmetric sub-Gaussian with moment bound K. Then*

$$\mathbb{E}(e^{\gamma\theta\mathbf{A}}) = e^{2K^2\theta^2\mathbf{A}^2} \tag{3.11}$$

Proof. We can absorb θ into $\mathbf{A}$, so it suffices to assume $\theta = 1$. The moments for a symmetric sub-Gaussian random variable γ satisfy $\mathbb{E}(\gamma^p) \leq K^p p^{p/2}$ for p even and $\mathbb{E}(\gamma^p) = 0$ for p odd. Then

$$\mathbb{E}(e^{\gamma\mathbf{A}}) = \mathbf{I} + \sum_{p=1}^{\infty} \frac{\mathbb{E}[\gamma^{2p}]\mathbf{A}^{2p}}{(2p)!}. \tag{3.12}$$

Now since $\mathbf{A}$ is self-adjoint, for any p $\mathbf{A}^{2p}$ has all non-negative eigenvalues and is thus positive semidefinite. Then since $\mathbb{E}(\gamma^{2p}) \leq K^{2p}(2p)^p$, $K^{2p}(2p)^p\mathbf{A}^{2p} - \mathbb{E}(\gamma^{2p})\mathbf{A}^{2p} = C_p\mathbf{A}^{2p}$ where $C_p \geq 0$ and thus $C_p\mathbf{A}^{2p}$ is positive semidefinite since it has only nonnegative eigenvalues. Hence $K^{2p}(2p)^p\mathbf{A}^{2p} \succcurlyeq \mathbb{E}(\gamma^{2p})\mathbf{A}^{2p}$ for any p. Then combining Lemma 3.2.2 with the fact that the sum of two positive semidefinite matrices is again positive semidefinite, we get

$$\begin{aligned} \mathbf{I} + \sum_{p=1}^{\infty} \frac{\mathbb{E}[\gamma^{2p}]\mathbf{A}^{\mathbf{2p}}}{(2p)!} &\preccurlyeq \mathbf{I} + \sum_{p=1}^{\infty} \frac{K^{2p}(2p)^p\mathbf{A}^{\mathbf{2p}}}{(2p)!} && (3.13)\\ &= \mathbf{I} + \sum_{p=1}^{\infty} \frac{(2\mathbf{A}^2K^2)^p p^p}{(2p)!} && (3.14)\\ &\preccurlyeq \mathbf{I} + \sum_{p=1}^{\infty} \frac{(2\mathbf{A}^2K^2)^p}{p!} && (3.15)\\ &= e^{2K^2\mathbf{A}^2}. && (3.16) \end{aligned}$$

The second semidefinite ordering inequality comes from $\frac{p^p}{(2p)!} \leq \frac{1}{p!}$ and Lemma 3.2.2. □

Now we are ready to prove Theorem 3.1.3.

Proof. (Proof of Theorem 3.1.3)
From the above lemma we get

$$\mathbb{E}(e^{\gamma_k \theta \mathbf{A}_k}) \preccurlyeq e^{g(\theta)\mathbf{A}_k^2} \tag{3.17}$$

where $g(\theta) = 2K^2\theta^2$ for $\theta > 0$. Recall that

$$\sigma^2 = \left\| \sum_k A_k^2 \right\| = \lambda_{\max}\left(\sum_k A_k^2 \right). \tag{3.18}$$

Then Proposition 3.2.1 gives

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \gamma_k \mathbf{A}_k \right) \geq t \right) \leq d \inf_{\theta>0} e^{-\theta t + g(\theta)\sigma^2}. \tag{3.19}$$

Picking $\theta = \frac{t}{4K^2\sigma^2}$ minimizes this infinum, and plugging in this value gives

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \gamma_k \mathbf{A}_k \right) \geq t \right) \leq de^{t^2/8K^2\sigma^2}. \tag{3.20}$$

Furthermore, applying the same argument to the series $X = \sum_{k=1}^d \gamma_k(-\mathbf{A}_k)$

$$\mathbb{P}\left(-\lambda_{\max}\left(\sum_k \gamma_k \mathbf{A}_k \right) \geq t \right) \leq de^{t^2/8K^2\sigma^2}, \tag{3.21}$$

so

$$\mathbb{P}\left(\left\| \sum_k \gamma_k \mathbf{A}_k \right\| \geq t \right) \leq 2de^{t^2/8K^2\sigma^2}. \tag{3.22}$$

□

Then, Corollary 3.1.4 follows from a symmetrization argument and the layercake represenation.

Proof. (Proof of Corollary 3.1.4) When the entries of X are not symmetric, we can use the following symmetrization argument from Section 2.3.2 of [Tao12], which was also used by Meckes in [Mec07]. Let $\tilde{X}$ be an independent copy of X. Then

$$\mathbb{E}(X - \tilde{X}|X) = X \tag{3.23}$$

and since the function $X \to ||X||$ is convex, by Jensen's inequality,

$$\mathbb{E}(||X - \tilde{X}|||X) \geq ||X|| \tag{3.24}$$

and then $\mathbb{E}(||X||) \leq \mathbb{E}(||X - \tilde{X}||)$, and $X - \tilde{X}$ is symmetric. If $X = \sum_{i=1}^n \gamma_i A_i$, then $X - \tilde{X} = \sum_{i=1}^n \gamma_i A_i - \sum_{i=1}^n \tilde{\gamma}_i A_i$ where $\tilde{\gamma}_i$ is an independent copy of γ_i. Then $\sigma(X - \tilde{X}) = \left\| \sum_{i=1}^n (2A_i)^2 \right\|^{1/2} = 2\sigma(X)$. Further, the moment bound in the sub-Gaussian condition of $\gamma - \tilde{\gamma}$ is at most twice the moment bound of γ. To see this, compute $\mathbb{E}[(\gamma - \tilde{\gamma})^{2p}]$ for some positive integer p by expanding via the binomial theorem, and then use independence and the fact that γ and $\tilde{\gamma}$ share the same sub-Gaussian moment bound. Thus the symmetrization argument introduces a factor of 4 into the bound of $\mathbb{E}||X||$.

In the case when the entries of X are assumed to be symmetric, we have from the layercake representation,

$$\mathbb{E}(||X||^2) = \int_0^\infty \mathbb{P}(||X|| > \sqrt{t})dt \tag{3.25}$$

$$\leq \int_0^{8K^2\sigma^2 \log(2d)} \mathbb{P}(||X|| > \sqrt{t})dt + 2d \int_{8K^2\sigma^2 \log(2d)}^\infty e^{-t/2\sigma^2} dt \tag{3.26}$$

$$\leq 8K^2\sigma^2 \log(2ed). \tag{3.27}$$

Then applying Jensen's inequality,

$$\mathbb{E}(||X||) \leq \sqrt{\mathbb{E}(||X||^2)} \leq (2^{3/2} K\sigma)\sqrt{\log(2ed)}. \tag{3.28}$$

□

3.3 Application to Patterned Random Matrices

In this section we use Theorem 3.1.3 to obtain simpler proofs of the results in [Mec07] and the appendix of [AS17]. A further benefit of using Corollary 3.1.4 to prove the spectral norm bound is that it explicitly gives the dependence on the constant from the sub-Gaussian condition.

An $n \times n$ symmetric random Toeplitz matrix T_n is defined as follows. Let $X_0, X_1, X_2, \ldots$ be independent random variables. Then $(T_n)_{ij} = X_{|i-j|}$,

$$T_n = \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{n-2} & X_{n-1} \\ X_1 & X_0 & X_1 & \cdots & X_{n-3} & X_{n-2} \\ X_2 & X_1 & X_0 & \cdots & X_{n-4} & X_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ X_{n-2} & X_{n-3} & X_{n-4} & \cdots & X_0 & X_1 \\ X_{n-1} & X_{n-2} & X_{n-3} & \cdots & X_1 & X_0 \end{bmatrix}. \tag{3.29}$$

Theorem 3.3.1. *Let T_n be a $n \times n$ symmetric random Toeplitz matrix with sub-Gaussian entries with uniform moment bound K. Then*

$$\mathbb{E}(||T_n||) \leq 2^{9/2} K \sqrt{n \log(2en)}. \tag{3.30}$$

The operator norm of symmetric random Toeplitz matrices was further analyzed in [SV13], though they assumed all entries had second moment equal to one and all higher order moments were uniformly bounded.

Proof. Any Toeplitz matrix can be constructed as a matrix series with the Hermitian matrices

$$(B_k)_{ij} = \begin{cases} 1 & |i-j| = k \\ 0 & |i-j| \neq k \end{cases} \tag{3.31}$$

for $k = 0, \dots, n-1$. Bounding $||B_k^2||$ via the Gershgorin circle theorem, an explicit computation gives $(B_k)_{ii} \leq 2$ for any $i = 1, \dots, n$ and summing over row i, $\sum_{j=1, j\neq i}^{n} (B_k^2)_{ij} \leq 2$ for any $i = 1, \dots, n$ (these inequalities come from the conditions on the entries of B_k and the formula for the ij-th entry of B_k^2). Then the Gershgorin circle theorem implies $||B_k^2|| \leq 4$ for any k, and thus $\sigma(X) = ||\sum_{k=0}^{n-1} B_k^2||^{1/2} \leq (\sum_{k=0}^{n-1} ||B_k^2||)^{1/2} \leq 2\sqrt{n}$ □

One can also use Corollary 3.1.4 to prove spectral norm bounds for non-symmetric matrices. If X is a (not necessarily Hermitian) square matrix, define it's *dilation* as

$$\tilde{X} \coloneqq \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}. \tag{3.32}$$

Then $\tilde{X}$ is Hermitian and

$$||\tilde{X}||^2 = ||\tilde{X}^2|| = \left\| \begin{bmatrix} XX^* & 0 \\ 0 & X^*X \end{bmatrix} \right\| = \max\{||X^*X||, ||XX^*||\} = ||X||^2. \tag{3.33}$$

In this paper, all matrices are real so $X^* = X^T$. We apply this to give a simpler proof of Lemma 16 in [AS17], which bounded the operator norm of a random circulant matrix. An $n \times n$ random circulant matrix C_n is defined as $(C_n)_{ij} = X_{(j-i) \mod n}$.

$$C_n = \begin{bmatrix} X_0 & X_1 & X_2 & \cdots & X_{n-2} & X_{n-1} \\ X_{n-1} & X_0 & X_1 & \cdots & X_{n-3} & X_{n-2} \\ X_{n-2} & X_{n-3} & X_0 & \cdots & X_{n-4} & X_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_2 & X_3 & X_4 & \cdots & X_0 & X_1 \\ X_1 & X_2 & X_3 & \cdots & X_{n-1} & X_0 \end{bmatrix}. \tag{3.34}$$

We have the following theorem.

Theorem 3.3.2. *Let C_n be a $n \times n$ symmetric random circulant matrix with sub-Gaussian entries with uniform moment bound K. Then*

$$\mathbb{E}(||C_n||) \leq 16K\sqrt{n \log(2en)}. \tag{3.35}$$

Proof. Any circulant matrix can be constructed as a matrix series with the matrices

$$(B_k)_{ij} = \begin{cases} 1 & j - i \equiv k \mod n \\ 0 & j - i \not\equiv k \mod n \end{cases} \tag{3.36}$$

for $k = 0, 1, \dots, n-1$. Then the dilation becomes $\tilde{C}_n = \sum_{k=0}^{n-1} X_k \tilde{B}_k$, where $\tilde{B}_k$ is the dilation of B_k. Then a straightforward computation yields $\sum_{k=0}^{n} B_k B_k^T = nI_n$ and similarly for $B_k^T B_k$, where I_n is the $n \times n$ identity matrix. Thus $||\sum_{k=0}^{n-1} \tilde{B}_k|| = 2n$, and it follows that $||C_n|| = ||\tilde{C}_n|| \leq 16K\sqrt{n \log(2en)}$. □

References

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Vol. 118. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010, pp. xiv+492 (cit. on p. 2).

[AS17] Kartick Adhikari and Koushik Saha. "Fluctuations of eigenvalues of patterned random matrices". In: *J. Math. Phys.* 58.6 (2017), pp. 063301, 20 (cit. on pp. 10, 12, 23, 24, 37, 38).

[AS18] Kartick Adhikari and Koushik Saha. "Universality in the fluctuation of eigenvalues of random circulant matrices". In: *Statist. Probab. Lett.* 138 (2018), pp. 1–8 (cit. on pp. 10, 32).

[AZ06] Greg W. Anderson and Ofer Zeitouni. "A CLT for a band matrix model". In: *Probab. Theory Related Fields* 134.2 (2006), pp. 283–338 (cit. on p. 10).

[Bai99] Z. D. Bai. "Methodologies in spectral analysis of large-dimensional random matrices, a review". In: *Statist. Sinica* 9.3 (1999). With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author, pp. 611–677 (cit. on p. 5).

[BBH23] Afonso S. Bandeira, March T. Boedihardjo, and Ramon van Handel. *Matrix Concentration Inequalities and Free Probability*. 2023. arXiv: `2108.06312 [math.PR]` (cit. on p. 29).

[BDJ06] Wł odzimierz Bryc, Amir Dembo, and Tiefeng Jiang. "Spectral measure of large random Hankel, Markov and Toeplitz matrices". In: *Ann. Probab.* 34.1 (2006), pp. 1–38 (cit. on p. 6).

[Bos18] Arup Bose. *Patterned random matrices*. CRC Press, Boca Raton, FL, 2018, pp. xxi+267 (cit. on p. 6).

[Cha09] Sourav Chatterjee. "Fluctuations of eigenvalues and second order Poincaré inequalities". In: *Probab. Theory Related Fields* 143.1-2 (2009), pp. 1–40 (cit. on pp. 10, 15, 16, 28, 32).

[EW13] Alan Edelman and Yuyang Wang. "Random matrix theory and its innovative applications". In: *Advances in applied mathematics, modeling, and computational science*. Vol. 66. Fields Inst. Commun. Springer, New York, 2013, pp. 91–116 (cit. on p. 1).

[FL13] Olga Friesen and Matthias Löwe. "A phase transition for the limiting spectral density of random matrices". In: *Electron. J. Probab.* 18 (2013), no. 17, 17 (cit. on p. 10).

[Joh98] Kurt Johansson. “On fluctuations of eigenvalues of random Hermitian matrices”. In: *Duke Math. J.* 91.1 (1998), pp. 151–204 (cit. on p. 10).

[Kem13] Todd Kemp. “Math 247a: Introduction to random matrix theory”. In: *Lecture notes* (2013) (cit. on pp. 3, 4, 22).

[Led01] Michel Ledoux. *The concentration of measure phenomenon.* Vol. 89. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001, pp. x+181 (cit. on p. 22).

[LP09] A. Lytova and L. Pastur. “Central limit theorem for linear eigenvalue statistics of random matrices with independent entries”. In: *Ann. Probab.* 37.5 (2009), pp. 1778–1840 (cit. on p. 10).

[LSW12] Dang-Zheng Liu, Xin Sun, and Zheng-Dong Wang. “Fluctuations of eigenvalues for random Toeplitz and related matrices”. In: *Electron. J. Probab.* 17 (2012), no. 95, 22 (cit. on p. 10).

[Mec07] Mark W. Meckes. “On the spectral norm of a random Toeplitz matrix”. In: *Electron. Comm. Probab.* 12 (2007), pp. 315–325 (cit. on pp. 32, 34, 36, 37).

[MS21] Shambhu Nath Maurya and Koushik Saha. “Fluctuations of linear eigenvalue statistics of reverse circulant and symmetric circulant matrices with independent entries”. In: *J. Math. Phys.* 62.4 (2021), Paper No. 043506, 31 (cit. on p. 10).

[NM11] Celine Nadal and Satya N. Majumdar. “A simple derivation of the Tracy-Widom distribution of the maximal eigenvalue of a Gaussian unitary random matrix”. In: *J. Stat. Mech. Theory Exp.* 4 (2011), P04001, 29 (cit. on p. 5).

[OVW16] Sean O’Rourke, Van Vu, and Ke Wang. “Eigenvectors of random matrices: a survey”. In: *J. Combin. Theory Ser. A* 144 (2016), pp. 361–442 (cit. on p. 1).

[Shc11] M. Shcherbina. “Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices”. In: *Zh. Mat. Fiz. Anal. Geom.* 7.2 (2011), pp. 176–192, 197, 199 (cit. on p. 10).

[SMS22] Kiran Kumar A. S., Shambhu Nath Maurya, and Koushik Saha. *A non-Gaussian limit for linear eigenvalue statistics of Hankel matrices.* 2022. arXiv: `2209.08252 [math.PR]` (cit. on pp. 10, 12, 31).

[Sos02] Alexander Soshnikov. “Gaussian limit for determinantal random point fields”. In: *Ann. Probab.* 30.1 (2002), pp. 171–187 (cit. on p. 10).

[SS98] Ya. Sinai and A. Soshnikov. “Central limit theorem for traces of large random symmetric matrices with independent matrix elements”. In: *Bol. Soc. Brasil. Mat. (N.S.)* 29.1 (1998), pp. 1–24 (cit. on pp. 10, 16).

[SV13] Arnab Sen and Bálint Virág. “The top eigenvalue of the random Toeplitz matrix and the sine kernel”. In: *Ann. Probab.* 41.6 (2013), pp. 4050–4079 (cit. on p. 37).

[SW13] Philippe Sosoe and Percy Wong. “Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices”. In: *Adv. Math.* 249 (2013), pp. 37–87 (cit. on p. 10).

[Tao12] Terence Tao. *Topics in random matrix theory*. Vol. 132. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2012, pp. x+282 (cit. on p. 36).

[Tro12] Joel A. Tropp. "User-friendly tail bounds for sums of random matrices". In: *Found. Comput. Math.* 12.4 (2012), pp. 389–434 (cit. on p. 34).

[Tro15] Joel A. Tropp. "An Introduction to Matrix Concentration Inequalities". In: *Foundations and Trends® in Machine Learning* 8.1-2 (2015), pp. 1–230 (cit. on pp. 7, 8, 33).

[Tro18] Joel A. Tropp. "Second-order matrix concentration inequalities". In: *Appl. Comput. Harmon. Anal.* 44.3 (2018), pp. 700–736 (cit. on pp. 15, 18).

[Ver12] Roman Vershynin. "Introduction to the non-asymptotic analysis of random matrices". In: *Compressed sensing*. Cambridge Univ. Press, Cambridge, 2012, pp. 210–268 (cit. on p. 7).

[Ver18] Roman Vershynin. *High-dimensional probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. An introduction with applications in data science, With a foreword by Sara van de Geer. Cambridge University Press, Cambridge, 2018, pp. xiv+284 (cit. on pp. 7, 30, 32, 33).

[Wig51] Eugene P Wigner. "On the statistical distribution of the widths and spacings of nuclear resonance levels". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 47. 4. Cambridge University Press. 1951, pp. 790–798 (cit. on p. 1).

[Wig55] Eugene P. Wigner. "Characteristic vectors of bordered matrices with infinite dimensions". In: *Ann. of Math. (2)* 62 (1955), pp. 548–564 (cit. on p. 1).

[Wis28] John Wishart. "The generalised product moment distribution in samples from a normal multivariate population". In: *Biometrika* (1928), pp. 32–52 (cit. on p. 1).