Electronic Journal of Probability

Vol. 13 (2008), Paper no. 21, pages 588–628.

Journal URL
http://www.math.washington.edu/~ejpecp/

Another look at the moment method for large dimensional random matrices

Arup Bose*
Statistics and Mathematics Unit
Indian Statistical Institute
202 B. T. Road, Kolkata 700108

Arnab Sen †
Department of Statistics
University of California, Berkeley
CA 94720, USA

Abstract

The methods to establish the limiting spectral distribution (LSD) of large dimensional random matrices includes the well known moment method which invokes the trace formula. Its success has been demonstrated in several types of matrices such as the Wigner matrix and the sample variance covariance matrix. In a recent article Bryc, Dembo and Jiang (2006) [7] establish the LSD for the random Toeplitz and Hankel matrices using the moment method. They perform the necessary counting of terms in the trace by splitting the relevant sets into equivalent classes and relating the limits of the counts to certain volume calculations.
We build on their work and present a unified approach. This helps provide relatively short and easy proofs for the LSD of common matrices while at the same time providing insight into the nature of different LSD and their interrelations. By extending these methods we are also able to deal with matrices with appropriate dependent entries .

Key words: Bounded Lipschitz metric, large dimensional random matrix, eigenvalues, Wigner matrix, sample variance covariance matrix, Toeplitz matrix, Hankel matrix, circulant matrix, symmetric circulant matrix, reverse circulant matrix, palindromic matrix, limiting spectral distribution, Marčenko-Pastur law, semicircular law, moment method, volume method, random probability, spectral distribution and density, Catalan numbers, m dependent sequence, linear process, almost sure convergence, convergence in distribution.

*email: abose@isical.ac.in
†email: arnab@stat.berkeley.edu

AMS 2000 Subject Classification: Primary 60F05, 60F15, 62E20, 60G57.

Submitted to EJP on November 9, 2007, final version accepted April 1, 2008.

1 Introduction

Random matrices with increasing dimension are called *large dimensional random matrices* (LDRM). They have appeared in many different areas of sciences - high dimensional data analysis, communication theory, dynamical systems, number theory, finance, combinatorics, diffusion processes, just to name a few.

In this article we deal with only real symmetric matrices so that all their eigenvalues are real. If λ is an eigenvalue of multiplicity m of an $n \times n$ matrix A_n, then the Empirical Spectral Measure puts mass m/n at λ. Note that if the entries of A_n are random, then this is a random probability measure. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all the eigenvalues, then the empirical spectral distribution function (ESD) F^{A_n} of A_n is given by

$$F^{A_n}(x) = n^{-1} \sum_{i=1}^{n} \mathrm{I}(\lambda_i \leq x).$$

Let $\{A_n\}$ be a sequence of square matrices with the corresponding ESD $\{F^{A_n}\}$. The Limiting Spectral Distribution (or measure) (LSD) of the sequence is defined as the weak limit of the sequence $\{F^{A_n}\}$, if it exists. If $\{A_n\}$ are random, the limit is in the "almost sure" or "in probability" sense.

For several matrices, the LSD is known to exist. There are several existing methods to establish such limits. One of the most fundamental methods is the method of moments: suppose $\{Y_n\}$ is a sequence of random variables with distribution functions $\{F_n\}$ such that $\mathrm{E}(Y_n^h) \to \beta_h$ for every positive integer h, and $\{\beta_h\}$ satisfies *Carleman's condition* (see Feller, 1966, page 224) [10]:

$$\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty. \tag{1}$$

Then there exists a distribution function F, such that for all h,

$$\beta_h = \beta_h(F) = \int x^h dF(x) \tag{2}$$

and $\{Y_n\}$ (or equivalently $\{F_n\}$) converges to F in distribution. We will often in short write β_h when the underlying distribution F is clear from the context.

Suppose $\{A_n\}$ is such a sequence of matrices, and let (by a slight abuse of notation) $\beta_h(A_n)$, for $h \geq 1$, denote the h-th moment of the ESD of A_n. Suppose there is a sequence of nonrandom $\{\beta_h\}_{h=1}^{\infty}$ satisfying Carleman's condition such that,

(M1) **First moment condition:** For every $h \geq 1$, $\mathrm{E}[\beta_h(A_n)] \to \beta_h$ and

(M2) **Second moment condition:** $\mathrm{Var}[\beta_h(A_n)] \to 0$.

Then the LSD is identified by $\{\beta_h\}_{h=1}^{\infty}$ and the convergence to LSD holds in probability. This convergence may be strengthened to almost sure convergence by strengthening (M2), for example, by replacing variance by the fourth moment and showing that

(M4) **Fourth moment condition:** $\mathrm{E}[\beta_h(A_n) - \mathrm{E}(\beta_h(A_n))]^4 = \mathrm{O}(n^{-2})$.

Before we discuss further details of the moment method and our unified approach, we describe some random matrices found in the literature on LDRM. The LSD of these matrices are given elsewhere in this article. The first five matrices are scaled by $n^{-1/2}$ to have a nontrivial LSD. Let $\{x_0, x_1, \ldots\}$ be a sequence of real random variables with mean zero and variance σ^2. We take $\sigma = 1$ without loss. Let $\mathbb{N}$ denote the set of all positive integers and $\mathbb{Z}_{\geq}$ the set of all nonnegative numbers.

1. Toeplitz matrix. The $n \times n$ random (symmetric) Toeplitz matrix T_n with inputs $\{x_i\}$ is the matrix whose (i,j)-th entry is $x_{|i-j|}$. So it is given by

$$T_n = \begin{bmatrix} x_0 & x_1 & x_2 & \ldots & x_{n-2} & x_{n-1} \\ x_1 & x_0 & x_1 & \ldots & x_{n-3} & x_{n-2} \\ x_2 & x_1 & x_0 & \ldots & x_{n-4} & x_{n-3} \\ & & & \vdots & & \\ x_{n-1} & x_{n-2} & x_{n-3} & \ldots & x_1 & x_0 \end{bmatrix}. \tag{3}$$

2. Hankel Matrix. Similarly, the (symmetric) Hankel matrix with inputs $\{x_i\}$ is the matrix whose (i,j)-th entry is x_{i+j}.

$$H_n = \begin{bmatrix} x_2 & x_3 & x_4 & \ldots & x_n & x_{n+1} \\ x_3 & x_4 & x_5 & \ldots & x_{n+1} & x_{n+2} \\ x_4 & x_5 & x_6 & \ldots & x_{n+2} & x_{n+3} \\ & & & \vdots & & \\ x_{n+1} & x_{n+2} & x_{n+3} & \ldots & x_{2n-1} & x_{2n} \end{bmatrix}. \tag{4}$$

Note that we have used an unconventional indexing, for our convenience. It may be noted that for the above two matrices, the (i,j)-th entry satisfies

$$a_{i,j} = x_{L(i,j)} \tag{5}$$

where $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}$ is a function independent of n. The concept of such a *link function* is going to be very crucial in our arguments.

3. Reverse circulant. This is defined as

$$RC_n = \begin{bmatrix} x_2 & x_3 & x_4 & \ldots & x_0 & x_1 \\ x_3 & x_4 & x_5 & \ldots & x_1 & x_2 \\ x_4 & x_5 & x_6 & \ldots & x_2 & x_3 \\ & & & \vdots & & \\ x_1 & x_2 & x_3 & \ldots & x_{n-1} & x_0 \end{bmatrix}. \tag{6}$$

Again, this is an unconventional indexing. The (i,j)-th element of the matrix is $x_{(i+j) \bmod n}$.

4. Symmetric Circulant. The symmetric version of the usual circulant matrix may be defined as

$$SC_n = \begin{bmatrix} x_0 & x_1 & x_2 & \ldots & x_2 & x_1 \\ x_1 & x_0 & x_1 & \ldots & x_3 & x_2 \\ x_2 & x_1 & x_0 & \ldots & x_2 & x_3 \\ & & & \vdots & & \\ x_1 & x_2 & x_3 & \ldots & x_1 & x_0 \end{bmatrix}. \tag{7}$$

So, the (i,j)-th element of the matrix is given by $x_{n/2-|n/2-|i-j||}$.

5. Palindromic matrices. These may be defined as (symmetric) matrices where the first row is a palindrome. So, for example consider the symmetric Toeplitz matrix defined earlier. Impose the restriction that the first row is a palindrome. This imposes other restrictions on the entries and we obtain the following palindromic Toeplitz matrix, see Massey, Miller and Sinsheimer [15], 2006) [MMS]:

$$PT_n = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_2 & x_1 & x_0 \\ x_1 & x_0 & x_1 & \dots & x_3 & x_2 & x_1 \\ x_2 & x_1 & x_0 & \dots & x_4 & x_3 & x_2 \\ & & & \vdots & & & \\ x_1 & x_2 & x_3 & \dots & x_1 & x_0 & x_1 \\ x_0 & x_1 & x_2 & \dots & x_2 & x_1 & x_0 \end{bmatrix}. \tag{8}$$

It may be noted that the $n \times n$ principal minor of PT_{n+1} is SC_n. Likewise we may define the palindromic version of the other matrices.

The next two matrices are the most studied matrices in the theory of random matrices. Here the input is a double sequence. Though the entries can easily be reindexed as a single sequence, for certain reasons we prefer to deal with double sequences of input for these matrices.

6. Wigner matrix. In its simplest form, a Wigner matrix W_n (Wigner, 1955, 1958) [20], [21] of order n is an $n \times n$ symmetric matrix whose entries on and above the diagonal are i.i.d. random variables with zero mean and variance one. Denoting those i.i.d. random variables by $\{x_{ij} : 1 \leq i \leq j\}$, we can visualize the Wigner matrix as

$$W_n = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1(n-1)} & x_{1n} \\ x_{12} & x_{22} & x_{23} & \dots & x_{2(n-1)} & x_{2n} \\ & & & \vdots & & \\ x_{1n} & x_{2n} & x_{3n} & \dots & x_{n(n-1)} & x_{nn} \end{bmatrix}. \tag{9}$$

7. Sample covariance matrix (S matrix). Suppose $\{x_{jk} : j,k = 1,2,\ldots\}$ is a double array of i.i.d. real random variables with mean zero and variance 1. In LDRM literature, the matrix

$$S_n = n^{-1} X_n X_n^T \quad \text{where } X_n = ((x_{ij}))_{1 \leq i \leq p, 1 \leq j \leq n} \tag{10}$$

is called a sample covariance matrix (in short an S matrix). We do not centre the matrices at the sample means as is conventional in statistics literature. This however, does not affect the LSD due to the scaling factor n^{-1} in (10).

To avoid problems of nonexistence of moments and to keep the discussion simple, assume that the random entries are **uniformly bounded**. This assumption may be relaxed to the existence of the first two moments by resorting to suitable truncation arguments (see Theorem 2). Note that the h-th moment of the ESD of an $n \times n$ matrix A_n, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ has the following nice form:

$$\beta_h(A_n) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^h = \frac{1}{n} \operatorname{Tr}(A_n^h). \tag{11}$$

Thus the major hurdle in establishing (M1) is computing at least the *leading term* in the expansion:

$$\mathrm{E}[\operatorname{Tr}(A_n^h)] = \mathrm{E}\left(\sum_{1 \leq i_1, i_2, \ldots, i_h \leq n} a_{i_1,i_2,n} a_{i_2,i_3,n} \cdots a_{i_{h-1},i_h,n} a_{i_h,i_1,n} \right) \tag{12}$$

where $a_{i,j,n}$ denotes the (i,j)-the element of A_n.

We need to do likewise for (M2) and (M4). These computations are available in the literature on a case by case basis. Every new matrix is a new (often difficult) combinatorial counting problem. Nevertheless, the moment method has been successfully applied for the Wigner, the sample covariance and the F matrices, and recently also for Toeplitz, Hankel and Markov matrices. See Bai (1999) [2] for some of the arguments in connection with Wigner, sample covariance and F matrices. For the arguments concerning Toeplitz, Hankel and Markov matrices see Bryc, Dembo and Jiang (2006) [7] [BDJ] and Hammond and Miller (2005) [12]. For the palindromic Toeplitz matrix, see [MMS] [15].

Suppose that $\{A_n\}$ are defined through an input sequence $\{x_i\}$ as in the first five matrices or through an input double sequence as in the last two matrices. Our primary goal is to present a unified way of dealing with such matrices.

Assume that the inputs are uniformly bounded, independent, with mean zero and variance σ^2. We first argue that in general, for matrices with suitable structure (that is with appropriate link function(s), L, say), the summands which have three or more powers of any inputs are indeed negligible in (M1). See Lemma 1. We also establish a general method to deal with (M4). See Lemma 2.

Thus only those summands in the trace formula potentially contribute where any input appears *exactly twice*. It also then follows from (M1) that the limit will depend solely on σ^2 and the nature of the link function (which determines the number and nature of contributing terms) but not on the distribution of the entries. Moreover, if a few entries of the matrix are different in nature from the rest, then that will not affect the LSD (see Theorem 1).

Then we show how the problem of existence of the limit of the contributing summands may be approached using the volume method of [BDJ] [7]. Using the link function, we define an equivalence relation between the terms and the total sum is split into two iterated sums. One of these is a finite sum (over the equivalence classes, called **words**) and for any fixed word, the limit of the other sum is calculated. This leads to the identification of different types of words, and their roles. Depending on the nature of the matrix, only certain words contribute positively. For some matrices, this contribution is constant across contributing words. For some others, each word contributes positively and in a nonconstant fashion (see Table 1).

It is worth noting that under suitable restriction on the L function(s), Carleman's condition is automatically satisfied (see Theorem 3).

Then we utilise and extend our general framework to deal with dependent entries. We establish some new LSD with dependent entries in Theorems 10 and 11 via suitable extensions of the earlier Lemmata (see Lemma 9 and Lemma 10).

1.1 Main results

We first prove two robustness results (Theorems 1 and 2) in Section 2. These show that in general we may assume that the entries are uniformly bounded and the same LSD persists if we change a few entries of the matrix.

In Section 3 we introduce the unified method. Lemma 1 essentially shows that only square terms contribute to the limit moments and Lemma 2 helps in verifying (M4). Theorem 3 shows that under suitable assumption on the link function, the existence of LSD and a.s. convergence of the moments are equivalent for uniformly bounded inputs. It also implies Carleman's condition and the sub Gaussian nature of the LSD.

We then apply our general approach to provide relatively short and easy proofs of the LSD of the matrices discussed above. The details for the Wigner matrix is discussed in Section 4.1. The S matrix is discussed in Section 4.2: the case when $p/n \to y \neq 0$ is presented in Section 4.2.1 and the case when $p/n \to 0$ is in Section 4.2.2. It turns out that the so called "Catalan words" are the only contributing words in these cases.

Next we deal with the Toeplitz and Hankel matrices in Section 4.3. We provide a succinct presentation of the [BDJ] [7] arguments in Section 4.3.1. All possible words contribute to the Toeplitz limit but we discover that only "symmetric words" matter for the Hankel limit. Section 4.3.2 develops interesting results on the moments of the Toeplitz and Hankel limits.

As a byproduct, a quick proof of the unboundedness of the Hankel LSD is given in Section 4.3.3.

The volume method is applied to reverse circulant in Section 4.4 and in particular again only symmetric words contribute.

Section 4.5 deals with matrices which have Gaussian limits. These include the symmetric circulant, palindromic Toeplitz and Hankel and, a doubly symmetric Hankel matrix. For the symmetric circulant, we show that all words contribute one each to the limit, and hence its LSD is Gaussian. We exploit its close connection to the other three matrices to conclude that all these matrices also have Gaussian limits.

Finally, in Section 5 we discuss an extension of our approach to suitable dependent inputs. We do not aim to provide fully general results but discuss a few specific situations. We establish that, the same LSD obtains if the input sequence is either a mean zero variance 1 i.i.d. sequence $\{x_i\}$, or is $\{x_i = \epsilon_i \dots \epsilon_{i+k-1}\}$ where $\{\epsilon_i\}$ is itself an i.i.d. sequence with mean zero and variance 1 (see Theorem 10). We also show that the LSD for Toeplitz and Hankel matrices exist when the input sequence is $X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$ where $\{\epsilon_j\}$ is i.i.d. with mean zero and variance 1 (see Theorem 11) and when $\{a_j\}$ is a suitable sequence. We provide their limit moments in terms of the limit moments when the input sequence is i.i.d.

2 Robustness of LSD

Most of the general results will be stated and proved when the $\{x_i\}$ are i.i.d. However, it must be intuitively clear that if we change a "few" of the entries then that should not alter the LSD. In our first theorem, we prove one such general robustness result. The reader will be able to establish more general non i.i.d. versions by simple modifications of our arguments.

The moment method arguments given later, crucially use the assumption of boundedness of the random variables. In the second theorem of this section, we will justify why we may always assume this condition without loss of generality. Once this is established, then in the rest of the paper we invariably provide the arguments only for the bounded case.

For the proof of the following two Theorems, we shall use the bounded Lipschitz metric. It is defined on the space of probability measures as:

$$d_{BL}(\mu,\ \nu) = \sup\{\int f d\mu - \int f d\nu : ||f||_\infty + ||f||_L \leq 1\}$$

where $||f||_\infty = \sup_x |f(x)|$, $||f||_L = \sup_{x \neq y} |f(x) - f(y)|/|x-y|$. Recall that convergence in d_{BL} implies the weak convergence of measures.

We need the following two facts. Fact 1 is an estimate of the metric distance d_{BL} in terms of trace. A proof may be found in Bai and Silverstein (2006) [3] or Bai (1999) [2] and uses Lidskii's theorem (see Bhatia, 1997, page 69) [5]. Fact 2 is the well known Cauchy's interlacing inequality (see Bhatia, 1997, page 59) [5] and its consequence.

Fact 1.
(a) Suppose A, B are $n \times n$ symmetric real matrices. Then

$$d_{BL}^2(F^A, F^B) \leq \left(\frac{1}{n} \sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)| \right)^2 \leq \frac{1}{n} \sum_{i=1}^{n} (\lambda_i(A) - \lambda_i(B))^2 \leq \frac{1}{n} \operatorname{Tr}(A-B)^2. \quad (13)$$

(b) Suppose A and B are $p \times n$ real matrices. Let $X = AA^T$ and $Y = BB^T$. Then

$$d_{BL}^2(F^X, F^Y) \leq \left(\frac{1}{p} \sum_{i=1}^{p} |\lambda_i(X) - \lambda_i(Y)| \right)^2 \leq \frac{2}{p^2} \operatorname{Tr}(X+Y) \operatorname{Tr}[(A-B)(A-B)^T]. \quad (14)$$

Fact 2.
(a) (Interlacing inequality) Suppose C is an $n \times n$ symmetric real matrix with eigenvalues $\lambda_n \geq \lambda_{n-1} \geq \ldots \geq \lambda_1$. Let D be the $(n-1) \times (n-1)$ principal submatrix of C with eigenvalues $\mu_{n-1} \geq \ldots \geq \mu_1$. Then

$$\lambda_n \geq \mu_{n-1} \geq \lambda_{n-1} \geq \mu_{n-2} \geq \ldots \geq \lambda_2 \geq \mu_1 \geq \lambda_1.$$

As a consequence, if A, B are $n \times n$ symmetric real matrices then

$$\|F^A - F^B\|_\infty \leq \frac{\operatorname{rank}(A-B)}{n}.$$

(b) Suppose A and B are $p \times n$ real matrices. Let $X = AA^T$ and $Y = BB^T$. Then

$$\|F^X - F^Y\|_\infty \leq \frac{\operatorname{rank}(A-B)}{p}.$$

Theorem 1. *Let $\{A_n\}$ be a sequence of $p \times n$ random matrices and let $\{E_n\}$ be a sequence of perturbation matrices of same order where the nonnull entries are from a triangular array of random variables $\{\epsilon_{i,n} : 1 \leq i \leq k_n\}$, $k_n \uparrow \infty$. Let $\alpha_n = \max_{1 \leq m \leq k_n} \#\{(i,j) : E_{ij,n} = \epsilon_{m,n}\}$ where $E_{ij,n}$ is the $(i,\ j)$-th entry of E_n.*

(a) Suppose $p = n$, $\{A_n,\ E_n\}$ are symmetric and the LSD of $\{n^{-1/2} A_n\}$ is F almost surely. If either of (H1) or (H2) below holds, the LSD of $\{n^{-1/2}(A_n + E_n)\}$ is also F almost surely.

(b) Suppose $\frac{p}{n} \to y \in (0, \infty)$ as $n \to \infty$ and $n^{-1} A_n A_n^T$ has LSD G almost surely. Suppose further that $(pn)^{-1} \operatorname{Tr}(A_n A_n^T)$ is almost surely bounded. If either of (H1), (H2) below holds, the LSD of $\{n^{-1} H_n H_n^T\}$ is also G almost surely where $H_n = A_n + E_n$.

(H1) $k_n \alpha_n = o(n^2)$ and $\frac{1}{k_n} \sum_{i=1}^{k_n} \epsilon_{i,n}^2$ is bounded almost surely as $n \to \infty$.

(H2) $\epsilon_{i,n}$ are row wise independent and (i) $\alpha_n = o((\log n)^{-1} n)$ and (ii) there exists $w(n)$ such that $\alpha_n k_n w(n)^2 = o(n^2)$ and $\alpha_n \sum_{i=1}^{k_n} P(|\epsilon_{i,n}| > w(n)) = o(n)$.

Proof. We first prove (a). Suppose (H1) holds. Then

$$\begin{aligned} d_{BL}^2(F^{n^{-1/2}(A_n+E_n)}, F^{n^{-1/2}A_n}) &\le n^{-2}\operatorname{Tr}(E_n^2) \\ &\le n^{-2}\alpha_n \sum_{i=1}^{k_n} \epsilon_{i,n}^2 = \frac{\alpha_n k_n}{n^2}\frac{1}{k_n}\sum_{i=1}^{k_n}\epsilon_{i,n}^2 \to 0 \ a.s. \end{aligned}$$

This proves the result when (H1) holds.

Now suppose (H2) holds. Let $\widetilde{E}_n$ be the matrix obtained from E_n replacing all occurrences of $\epsilon_{i,n}$ with $\epsilon_{i,n}\,\mathrm{I}(\epsilon_{i,n} \le w(n))$ for $1 \le i \le k_n$.

Then,

$$\begin{aligned} d_{BL}^2(F^{n^{-1/2}(A_n+\widetilde{E}_n)}, F^{n^{-1/2}A_n}) &\le n^{-2}\operatorname{Tr}(\widetilde{E}_n^2) \\ &\le n^{-2}\alpha_n \sum_{i=1}^{k_n}\epsilon_{i,n}^2\,\mathrm{I}(\epsilon_{i,n} \le w(n)) \\ &\le n^{-2}w(n)^2\alpha_n k_n \to 0. \end{aligned}$$

On the other hand, by the rank inequality,

$$\|F^{n^{-1/2}(A_n+E_n)} - F^{n^{-1/2}(A_n+\widetilde{E}_n)}\|_\infty \le \frac{\operatorname{rank}(E_n - \widetilde{E}_n)}{n}.$$

We will show $\frac{\operatorname{rank}(E_n-\widetilde{E}_n)}{n} \to 0$ almost surely. Let $p_{in} = \mathrm{P}(|\epsilon_{i,n}| > w(n))$ and $P_n = \sum_{i=1}^{k_n} p_{in}$. Note that $\operatorname{rank}(E_n - \widetilde{E}_n) \le$ number of nonzero entries in $(E_n - \widetilde{E}_n)$, which is bounded by $\alpha_n \sum_{i=1}^{k_n}\mathrm{I}(|\epsilon_{i,n}| > w(n))$.

By Bernstein's inequality (see Chow and Teicher(1997) [9]), we have, for any $\delta > 0$,

$$\begin{aligned} \mathrm{P}(\operatorname{rank}(E_n - \widetilde{E}_n) \ge \delta n) &\le \mathrm{P}\left(\alpha_n \sum_{i=1}^{k_n}\Big(\mathrm{I}(|\epsilon_{i,n}| > w(n)) - p_{in}\Big) \ge n\delta - \alpha_n P_n\right) \\ &\le 2\exp(-b_n^2/2[P_n + b_n]) \end{aligned}$$

where $b_n = \frac{n\delta - \alpha_n P_n}{\alpha_n}$. Using the given conditions, $b_n \ge K\log n$ for some large K and $\exp(-b_n^2/2[P_n+b_n]) \le Cn^{-(1+\Delta)}$ for some $C,\ \Delta > 0$. The result now follows from Borel-Cantelli Lemma.

The above arguments may be modified to prove (b). We give the main steps only. First assume (H1) holds. Then

$$\begin{aligned} d_{BL}^2(F^{n^{-1}H_nH_n^T}, F^{n^{-1}A_nA_n^T}) &\le 2(pn)^{-2}\operatorname{Tr}(H_nH_n^T + A_nA_n^T)\operatorname{Tr}(E_nE_n^T) \\ &\le 4(pn)^{-2}\left[2\operatorname{Tr}(A_nA_n^T) + \operatorname{Tr}(E_nE_n^T)\right]\operatorname{Tr}(E_nE_n^T) \to 0 \ a.s. \end{aligned}$$

since $(pn)^{-1}\operatorname{Tr}(E_nE_n^T) \to 0$ $a.s.$ (under (H1)) and, $(pn)^{-1}\operatorname{Tr}(A_nA_n^T) = O_{a.s}(1)$. This proves (b) when (H1) holds.

Now suppose (H2) holds. Let $\widetilde{E}_n$ be the matrix obtained from E_n replacing all occurrences of $\epsilon_{i,n}$ with $\epsilon_{i,n}\,\mathrm{I}(\epsilon_{i,n} \leq w(n))$ for $1 \leq i \leq k_n$. Noting that $(pn)^{-1}\operatorname{Tr}(\widetilde{E}_n\widetilde{E}_n^T) \to 0$ *a.s.*, same inequalities as above yield,

$$d_{BL}^2(F^{n^{-1}\widetilde{H}_n\widetilde{H}_n^T}, F^{n^{-1}A_nA_n^T}) \to 0 \quad \text{where } \widetilde{H}_n = A_n + \widetilde{E}_n.$$

On the other hand, by the rank inequality,

$$\|F^{n^{-1}H_nH_n^T} - F^{n^{-1}\widetilde{H}_n\widetilde{H}_n^T}\|_\infty \leq \frac{\operatorname{rank}(E_n - \widetilde{E}_n)}{p}.$$

Now we may proceed as in the proof of part (a). Note that here the matrices E_n and $\widetilde{E}_n$ are not symmetric but the above arguments remain valid. We omit the details. $\square$

Remark 1. *Suppose $\{\epsilon_{i,n}\}$ is row wise independent and $\sup_{i,n}\mathrm{E}\,|\epsilon_{i,n}|^{4+\eta} < \infty$ for some $\eta > 0$. Then the almost sure boundedness in (H1) holds by the SLLN of row-wise independent random variables.*

Now we prove a theorem showing that we may work with bounded random variables. Let $\{A_n\}$ be a sequence of $p \times n$ random matrices whose input is either a single or a double sequence of i.i.d. random variables $\{x_0, x_1, x_2, \ldots\}$ or $\{x_{ij} : i, j \geq 0\}$ so that the $(i,\ j)$-th entry of A_n is $a_{i,j,n} = x_{L_n(i,j)}$ for some function $L_n : \{1, 2, \ldots, n\}^2 \to \mathbb{Z}_\geq$ or $\mathbb{Z}^2_\geq$. Let $k_n = \#\{L_n(i,j) : 1 \leq i, j \leq n\}$ and $\alpha_n = \max_k \#\{(i,j) : L_n(i,j) = k,\ 1 \leq i,\ j \leq n\}$.

Theorem 2. *Let A_n, k_n and α_n be as above such that $k_n \to \infty$ and $k_n\alpha_n = O(n^2)$.*

(a) Let $p = n$, $\{A_n\}$ be symmetric. Suppose for every bounded, mean zero and variance one i.i.d. input sequence (double sequence), $F^{n^{-1/2}A_n}$ converges to some fixed nonrandom distribution F a.s. Then the same limit continues to hold if the input sequence is i.i.d. with variance one.

(b) Let $p/n \to y \in (0, \infty)$ as $n \to \infty$. Suppose for every bounded, mean zero and variance one i.i.d. input sequence (double sequence), $F^{n^{-1}A_nA_n^T}$ converges to some fixed nonrandom distribution G a.s. Then the same limit continues to hold if the input sequence is i.i.d. with variance one.

Proof. We will only consider the case when the input is a single sequence $\{x_0, x_1, x_2, \ldots\}$. The other case is similar.

By rank inequality, subtracting a rank one matrix $\mathrm{E}(x_0)\mathbf{1}\mathbf{1}^\mathbf{T}$ if necessary, we can assume $\{x_i\}$ have mean zero. For $t > 0$, denote

$$\mu(t) \stackrel{def}{=} \mathrm{E}[x_0\,\mathrm{I}(|x_0| > t)] = -\,\mathrm{E}[x_0\,\mathrm{I}(|x_0| \leq t)].$$

and let

$$\sigma^2(t) \stackrel{def}{=} \operatorname{Var}(x_0\,\mathrm{I}(|x_0| \leq t)) = \mathrm{E}[x_0^2\,\mathrm{I}(|x_0| \leq t)] - \mu(t)^2.$$

Since $\mathrm{E}(x_0) = 0$ and $\mathrm{E}(x_0^2) = 1$, we have $\mu(t) \to 0$ and $\sigma(t) \to 1$ as $t \to \infty$ and $\sigma^2(t) \leq 1$. Define bounded random variables

$$x_i^* = \frac{x_i\,\mathrm{I}(|x_i| \leq t) + \mu(t)}{\sigma(t)} = \frac{x_i - \bar{x}_i}{\sigma(t)} \quad \text{where } \bar{x}_i = x_i\,\mathrm{I}(|x_i| > t) - \mu(t) = x_i - \sigma(t)x_i^*. \tag{15}$$

It is easy to see that $\mathrm{E}(\bar{x}_0^2) = 1 - \sigma^2(t) - \mu(t)^2 \to 0$ as t tends to infinity. Further, $\{x_i^*\}$ are i.i.d. bounded, mean zero and variance one random variables. Let us denote the matrix A_n constructed from this sequence $\{x_i^*\}_{i\geq 0}$ by A_n^*.

We now prove (a). By triangle inequality and (13),

$$
\begin{aligned}
&d_{BL}^2\left(F^{n^{-1/2}A_n}, F^{n^{-1/2}A_n^*}\right)\\
&\quad \leq 2d_{BL}^2\left(F^{n^{-1/2}A_n}, F^{n^{-1/2}\sigma(t)A_n^*}\right) + 2d_{BL}^2\left(F^{n^{-1/2}A_n^*}, F^{n^{-1/2}\sigma(t)A_n^*}\right)\\
&\quad \leq \frac{2}{n^2}\operatorname{Tr}(A_n - \sigma(t)A_n^*)^2 + \frac{2(1-\sigma(t))^2}{n^2}\operatorname{Tr}(A_n^*)^2.
\end{aligned}
$$

Now, by the hypotheses on A_n, using strong law of large numbers, we get

$$\frac{1}{n^2}\operatorname{Tr}(A_n^*)^2 = \frac{1}{n^2}\sum_{1\leq i,j\leq n}(a_{i,j,n}^*)^2 \leq \frac{\alpha_n}{n^2}\sum_{h=0}^{k_n}{x_h^*}^2 \leq \frac{C}{k_n}\sum_{h=0}^{k_n}{x_h^*}^2 \xrightarrow{a.s.} C\,\mathrm{E}({x_0^*}^2) = C.$$

and we have $1-\sigma(t) \to 0$ as $t\to\infty$. Similarly,

$$\frac{1}{n^2}\operatorname{Tr}(A_n - \sigma(t)A_n^*)^2 \leq \frac{C}{k_n}\sum_{h=0}^{k_n}\bar{x}_h^2 \xrightarrow{a.s.} C\,\mathrm{E}(\bar{x}_0^2) \to 0 \text{ as } t\to\infty.$$

Thus

$$\limsup_n d_{BL}\left(F^{n^{-1/2}A_n}, F^{n^{-1/2}A_n^*}\right) \to 0 \quad a.s. \text{ as } t\to\infty \tag{16}$$

and the proof of (a) is complete.

To prove (b), let $\{x_i^*\}$, $\{A_n^*\}$ be as above. Now we need to use inequality (14) and the facts that there exists $C > 0$ such that almost surely, $\limsup_n (np)^{-1}\operatorname{Tr}(A_nA_n^T) \leq C$, $\limsup_n (np)^{-1}\operatorname{Tr}(A_n^*(A_n^*)^T) \leq C$ $\forall t > 0$ and $\limsup_{t\to\infty}\limsup_n (np)^{-1}\operatorname{Tr}(B_nB_n^T) = 0$ where $B_n = A_n - \sigma(t)A_n^*$. The rest of the argument is as above. We omit the details. □

3 The volume method: a unified approach

As we pointed out earlier, the entries of the matrices are taken from an input sequence. We now develop this common thread in a systematic manner. As discussed, we will assume *the input sequence to be independent and uniformly bounded with mean zero and variance one* unless otherwise mentioned.

Link function L. Let d be a positive integer. Let $L_n : \{1,2,\dots n\}^2 \to \mathbb{Z}_{\geq}^d$, $n\geq 1$ be a sequence of functions symmetric in its arguments such that $L_{n+1}(i,j) = L_n(i,j)$ whenever $1\leq i,j\leq n$.

The sequence of matrices $\{A_n\}$ under consideration will always be such that the (i,j)-th entry of A_n is of the form $a_{i,j} = x_{L_n(i,j)}$. To simplify notations we shall write $L_n = L$ and call it the **link** function. Likewise we will by abuse of notations write $\mathbb{N}^2$ as the common domain of $\{L_n\}$. The link function for the different matrices discussed so far are:

(i) Wigner matrix: $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}^2$ where $L(i,j) = (\min(i,j), \max(i,j))$.

(ii) Toeplitz matrix: $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}$ where $L(i,j) = |i-j|$.

(iii) Hankel matrix: $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}$ where $L(i,j) = i+j$.

(iv) Reverse Circulant: $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}$ where $L(i,j) = (i+j) \mod n$.

(v) Symmetric Circulant: $L : \mathbb{N}^2 \to \mathbb{Z}_{\geq}$ where $L(i,j) = n/2 - |n/2 - |i-j||$.

(vi) Sample variance covariance matrix: Here the relevant link is a pair of functions given by:

$$L1, L2 : \mathbb{N}^2 \times \mathbb{N}^2 \to \mathbb{Z}^2_{\geq},\ L1(i,j) = (i,j),\ L2(i,j) = (j,i).$$

We will primarily consider matrices where we need only one link function. The approach is easily modified to include matrices such as the S, which requires two L functions.

The trace formula. Let $A_n = ((a_{i,j}))$. Then recall that the h-th moment of ESD of $n^{-1/2}A_n$ is given by

$$\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h = \frac{1}{n^{1+h/2}} \sum_{1\leq i_1,i_2,\ldots,i_h\leq n} a_{L(i_1,i_2)}a_{L(i_2,i_3)}\cdots a_{L(i_{h-1},i_h)}a_{L(i_h,i_1)}. \tag{17}$$

Circuit: Any function $\pi : \{0,1,2,\cdots,h\} \to \{1,2,\cdots,n\}$ is said to be a *circuit* if $\pi(0) = \pi(h)$. The **length** $l(\pi)$ of π is taken to be h. A circuit depends on h and n but we will suppress this dependence.

Define

$$\mathbb{X}_\pi = x_{L(\pi(0),\pi(1))}x_{L(\pi(1),\pi(2))}\cdots x_{L(\pi(h-2),\pi(h-1))}x_{L(\pi(h-1),\pi(h))}.$$

The conditions (M1) and (M4) may be written in terms of circuits. For example,

(M1) $\mathrm{E}[\beta_h(n^{-1/2}A_n)] = \mathrm{E}[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h] = \frac{1}{n^{1+h/2}}\sum_{\pi:\ \pi \text{ circuit}} \mathrm{E}\,\mathbb{X}_\pi \to \beta_h.$

Matched Circuits: Any value $L(\pi(i-1),\pi(i))$ is said to be an *L-value* of π, and π is said to have an **edge of order** e $(1 \leq e \leq h)$ if it has an L-value repeated exactly e times. If π has at least one edge of order one then $\mathrm{E}(\mathbb{X}_\pi) = 0$. Thus only those π with all $e \geq 2$ are relevant. Such circuits will be said to be **L-matched** (in short *matched*). For any such π, given any i, there is at least one $j \neq i$ such that $L(\pi(i-1),\pi(i)) = L(\pi(j-1),\pi(j))$.

To deal with (M2) or (M4), we need multiple circuits. The following notions will be useful: k circuits $\pi_1,\pi_2,\cdots,\pi_k$ are said to be *jointly matched* if each L-value occurs at least twice across all circuits. They are said to be *cross-matched* if each circuit has at least one L-value which occurs in at least one of the other circuits.

The following extension of L-matching will be used, specially when the inputs are dependent. Suppose $f : \mathbb{Z}^d_{\geq} \to \mathbb{Z}_{\geq}$ is a function. A circuit π is said to be (L,f)-matched if for each i, there is at least one $j \neq i$ such that $f\Big(\big|L(\pi(i-1),\pi(i)) - L(\pi(j-1),\pi(j))\big|\Big) = 0$. The L matching introduced earlier is obtained as a special case with $f(x) = x$. The concepts of jointly matching and cross matching can be similarly extended.

Equivalence relation on circuits: Two circuits π_1 and π_2 (of same length) are said to be equivalent if their L values agree at exactly the same pairs (i,j). That is, iff $\big\{L(\pi_1(i-1),\pi_1(i)) = L(\pi_1(j-1),\pi_1(j)) \Leftrightarrow L(\pi_2(i-1),\pi_2(i)) = L(\pi_2(j-1),\pi(j))\big\}$. This defines an equivalence relation between the circuits.

Words: Equivalence classes may be identified with partitions of $\{1, 2, \cdots, h\}$: to any partition we associate a **word** w of length $l(w) = h$ of letters where the first occurrence of each letter is in alphabetical order. For example, if $h = 5$, then the partition $\{\{1, 3, 5\}, \{2, 4\}\}$ is represented by the word *ababa*.

The class $\Pi(w)$: Let $w[i]$ denote the i-th entry of w. The equivalence class corresponding to w will be denoted by

$$\Pi(w) = \{\pi : w[i] = w[j] \Leftrightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))\}.$$

The number of partition blocks corresponding to w will be denoted by $|w|$. If $\pi \in \Pi(w)$, then clearly, $\#\{L(\pi(i-1), \pi(i)) : 1 \leq i \leq h\} = |w|$.

The notion of order e edges, matching, nonmatching for π carries over to words in a natural manner. For instance, *ababa* is matched. The word *abcadbaa* is nonmatched, has edges of order 1, 2 and 4 and the corresponding partition is $\{\{1, 4, 7, 8\}, \{2, 6\}, \{3\}, \{5\}\}$. As pointed out, it will be enough to consider only matched words. *The total number of matched words having order* 2 *edges with* $|w| = k$ *equals* $\frac{(2k)!}{2^k k!}$.

The class $\Pi^*(w)$: Define for any (matched) word w,

$$\Pi^*(w) = \{\pi : w[i] = w[j] \Rightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))\}.$$

Note that $\Pi^*(w) \supseteqq \Pi(w)$. However, as we will see, $\Pi^*(w)$ is equivalent to $\Pi(w)$ for asymptotic considerations, but is easier to work with.

The L function for all of the above cases satisfy the following property. This property will be crucially used in the proof of the LSD of different matrices.

Property B: Let $f : \mathbb{Z}_{\geq}^d \to \mathbb{Z}_{\geq}$. Then (L, f) is said to satisfy Property B if,

$$\Delta(L, f) = \sup_n \sup_{t \in \mathbb{Z}_{\geq}^d} \sup_{1 \leq k \leq n} \#\{l : 1 \leq l \leq n,\ f(|L(k, l) - t|) = 0\} < \infty. \tag{18}$$

An example of an (L, f) which does not satisfy Property B is, $L(i, j) = \min(i, j)$ and $f(x) = x$.

The following Lemma helps to reduce the number of terms in step (M1). We need the version with general f to tackle dependent inputs later. Let $N_{h,3^+}$ be the number of (L, f) matched circuits on $\{1, 2, \ldots, n\}$ of length h with at least one edge of order ≥ 3.

Lemma 1. *(a) If (L, f) satisfies Property B, then there is a constant C such that*

$$N_{h,3^+} \leq C n^{\lfloor (h+1)/2 \rfloor}. \tag{19}$$

As a consequence, as $n \to \infty$,

$$n^{-(1+h/2)} N_{h,3^+} \to 0. \tag{20}$$

(b) Suppose $\{A_n\}$ is a sequence of $n \times n$ random matrices with input sequence $\{x_i\}$ which is uniformly bounded, independent, with mean zero variance 1 and (L, f) with $f(x) = x$ satisfies Property B. Then

(i) if h is odd,

$$\lim_n \mathrm{E}[\beta_h(n^{-1/2} A_n)] = \lim_n E[\frac{1}{n} \operatorname{Tr} \left(\frac{A_n}{\sqrt{n}}\right)^h] = 0. \tag{21}$$

(ii) if $h = 2k$, then

$$\sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi^*(w) - \Pi(w)| = 0 \tag{22}$$

and provided any of the last two limits below exists,

$$\lim_n \mathrm{E}[\beta_{2k}(n^{-1/2}A_n)] = \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi(w)| = \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi^*(w))| \tag{23}$$

Proof. Relation (19) is proved in [BDJ] [7] for L-functions corresponding to Toeplitz and Hankel matrices when $f(x) = x$. A careful scrutiny of their proof reveals that the same argument works for the present case. Relation (20) is an immediate consequence.

Relation (22) follows immediately since if $\pi \in \Pi^*(w) \setminus \Pi(w)$ then π must have an edge of order at least 4 and we can apply (20).

From mean zero and independence assumption (provided the limit below exists),

$$\lim \mathrm{E}[\beta_h(n^{-1/2}A_n)] = \lim \frac{1}{n^{1+h/2}} \sum_{\pi \text{ circuit}} \mathrm{E}\,\mathbb{X}_\pi = \sum_{w \text{ matched}} \lim \frac{1}{n^{1+h/2}} \sum_{\pi \in \Pi(w)} \mathrm{E}\,\mathbb{X}_\pi. \tag{24}$$

By Holder's inequality, $|\sum_{\pi:\pi\in\Pi(w)} \mathrm{E}\,\mathbb{X}_\pi| \le |\Pi(w)|\,\mathrm{E}(|x_{L(1,1)}|^h)$. Therefore, from part (a), matched circuits which have edges of order three or more do not contribute to the limit in (24). So, provided the limit below exists,

$$\lim E[\beta_h(n^{-1/2}A_n)] = \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+h/2}} |\Pi(w)| = \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+h/2}} |\Pi^*(w)|. \tag{25}$$

Taking $h = 2k$ establishes (23).

If $h = 2k+1$, note that then there cannot be any matched word of odd length with only order 2 edges. Hence from (25),

$$\lim_n \mathrm{E}[\beta_h(n^{-1/2}A_n)] = \lim_n E[\frac{1}{n} \operatorname{Tr} \left(\frac{A_n}{\sqrt{n}}\right)^h] = 0,$$

proving (22). The proof is now complete. □

Define, for each fixed matched word w of length $2k$ with $|w| = k$,

$$p(w) = \lim_n \frac{1}{n^{1+k}} |\Pi^*(w)| \tag{26}$$

whenever the limit exists. For any fixed word, this limit will be positive and finite only if the number of elements in the set is of exact order n^{k+1}. From Lemma 1, it follows that then the limiting $(2k)$-th moment is the finite sum

$$\beta_{2k} = \sum_{w:|w|=k,\ l(w)=2k} p(w).$$

Theorem 3 may be used to verify Carleman's condition for $\{\beta_{2k}\}$.

The next Lemma helps to verify (M4). Let $Q_{h,4}$ be the number of quadruples of circuits $(\pi_1, \pi_2, \pi_3, \pi_4)$ of length h such that they are jointly matched and cross-matched with respect to (L, f).

Lemma 2. *(a) If (L, f) satisfies Property B, then there exists a constant K such that,*

$$Q_{h,4} \leq K n^{2h+2}. \tag{27}$$

(b) Suppose $\{A_n\}$ is a sequence of $n \times n$ random matrices with input sequence $\{x_i\}$ which is uniformly bounded, independent, with mean zero variance 1 and (L, f) with $f(x) = x$ satisfies Property B. Then

$$\mathrm{E}\left[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h - \mathrm{E}\,\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h\right]^4 = \mathrm{O}(n^{-2}) \tag{28}$$

and hence (M4) holds.

Proof. Again, relation (27) is proved in [BDJ] [7] for L-functions corresponding to Toeplitz and Hankel matrices when $f(x) = x$ and the same proof works here. We omit the details.

To prove (b), we write the fourth moment as

$$\frac{1}{n^{h+2}}\,\mathrm{E}[\operatorname{Tr} A_n^h - \mathrm{E}(\operatorname{Tr} A_n^h)]^4 = \frac{1}{n^{h+2}} \sum_{\pi_1,\pi_2,\pi_3,\pi_4} \mathrm{E}[\prod_{i=1}^{4}(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})].$$

If $(\pi_1, \pi_2, \pi_3, \pi_4)$ are not jointly matched, then one of the circuits, say π_j, has an L-value which does not occur anywhere else. Also note that $\mathrm{E}\,\mathbb{X}_{\pi_j} = 0$. Hence, using independence, $\mathrm{E}[\prod_{i=1}^4(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})] = \mathrm{E}[\mathbb{X}_{\pi_j}\prod_{i=1,i\neq j}^4(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})] = 0$.

Further, if $(\pi_1, \pi_2, \pi_3, \pi_4)$ is jointly matched but is not cross matched then one of the circuits, say π_j is only self-matched, that is, none of its L-values is shared with those of the other circuits. Then by independence, $\mathrm{E}[\prod_{i=1}^4(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})] = \mathrm{E}[(\mathbb{X}_{\pi_j} - \mathrm{E}\,\mathbb{X}_{\pi_j})]\,\mathrm{E}[\prod_{i=1,i\neq j}^4(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})] = 0$.

Since the entries are bounded, so is $\mathrm{E}[\prod_{i=1}^4(\mathbb{X}_{\pi_i} - \mathrm{E}\,\mathbb{X}_{\pi_i})]$. Therefore by part (a),

$$\mathrm{E}\left[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h - \mathrm{E}\,\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h\right]^4 \leq K\frac{n^{2h+2}}{n^4(n^{h/2})^4} = \mathrm{O}(n^{-2}), \tag{29}$$

proving the Lemma completely. □

We now explore the question of when the limit in (26) exists.

Vertex and generating vertex: Any $\pi(i)$ is called a **vertex**. This vertex is said to be *generating* if either $i = 0$ or $w[i]$ is the position of the *first* occurrence of a letter. For example, if $w = abbcab$ then $\pi(0), \pi(1), \pi(2), \pi(4)$ are generating vertices. By Property B of L given earlier, a circuit of a fixed length is completely determined, *up to a finitely many choices* by its generating vertices. Obviously, number of generating vertices in π is $|w| + 1$. Hence we obtain the simple but crucial estimate

$$|\Pi^*(w)| = \mathrm{O}(n^{|w|+1}).$$

Remark 2. *Indeed, in all of the examples given so far, it turns out that we can find a set* $\Pi^{**}(w) \subseteq \Pi^*(w)$ *so that (i)* $\lim_n n^{-(k+1)}\big(|\Pi^*(w)| - |\Pi^{**}(w)|\big) = 0$ *and (ii) for any* $\pi \in \Pi^{**}(w)$ *if the generating vertices are fixed then there is* ***at most one*** *choice for the other vertices. So,* $p(w) \leq 1$. *As a further consequence,* $\beta_{2k} \leq \frac{(2k)!}{2^k k!}$. *Thus the limit moments are dominated by the Gaussian moments (see also Table 1). A general result for link functions satisfying Property B is given in Theorem 3.*

Theorem 3. *Suppose* $\{A_n = ((x_{L(i,j)}))_{i,j=1}^n\}$ *is a sequence of* $n \times n$ *random matrices with a uniformly bounded independent input sequence* $\{x_i\}$ *with mean zero variance 1. Suppose* (L, f) *satisfies Property B with* $f(x) = x$. *A nonrandom LSD* G *exists for* $\{n^{-1/2}A_n\}$, *if and only if for every* h, *almost surely,*

$$\lim \beta_h(n^{-1/2}A_n) = \beta_h(G) \text{ say.} \tag{30}$$

In that case, (i) $\beta_h(G) = 0$ *whenever* h *is odd, and (ii)* $\beta_{2k}(G) \leq \frac{(2k)!\Delta(L,f)^k}{k!2^k}$ *satisfies Carleman's condition and hence* G *is sub Gaussian. Even when the limit in (30) does not exist, the ESD of* $\{n^{-1/2}A_n\}$ *is tight almost surely and any subsequential limit* L *satisfies all the properties of* G *listed above.*

Proof. For convenience, we write F_n for $F^{n^{-1/2}A_n}$. Let $\hat{F}_n = \mathrm{E}(F_n)$ be the expected empirical spectral distribution function. Since Property B holds, we immediately have the following:

(a) (Odd moments) Lemma 1(b)(i) and Lemma 2(b) implies $\mathrm{E}[\beta_{2k-1}(F_n)] \to 0$.

(b) (Even moments) Fix a *matched* word w with $l(w) = 2k$. If $|w| = k$, then we have $(k+1)$ generating vertices. Having fixed the generating vertices arbitrarily, we have at most $\Delta(L, f)$ choices for each of the remaining k vertices. Thus, $|\Pi^*(w)| \leq \Delta(L, f)^k n^{k+1}$.

   On the other hand, it is easy to see that for any word w, if $|w| < k$ then since the number of generating vertices is less than k, $|\Pi^*(w)| = O_k(n^k)$ where the O_k term may involve k.

   Combining, we get

$$\mathrm{E}[\beta_{2k}(F_n)] \leq \sum_{w: l(w)=2k} \frac{1}{n^{k+1}} |\Pi^*(w)| \leq \frac{(2k)!}{2^k k!}\Delta(L, f)^k + O_k(n^{-1}). \tag{31}$$

(c) Lemma 2 implies that for each $h \geq 1$, $\beta_h(n^{-1/2}A_n) - \mathrm{E}[\beta_h(n^{-1/2}A_n)] \to 0$ *a.s.* Also using (a), we have $\beta_{2k-1}(F_n) \stackrel{a.s.}{\to} 0$ for all $k \geq 1$.

Only if Part: Suppose $F_n \stackrel{\mathcal{D}}{\Rightarrow} G$ almost surely. Fix x a continuity point of G. Then $F_n(x) \stackrel{a.s.}{\to} G(x)$. Using DCT, we get $\hat{F}_n(x) \to G(x)$. Hence $\hat{F}_n \stackrel{\mathcal{D}}{\Rightarrow} G$.

This implies that $\int x^{2k} d\hat{F}_n(x) = \mathrm{E}[\beta_{2k}(F_n)] \leq \frac{(2k)!\Delta(L,f)^k}{k!2^k} + O_k(n^{-1})$ for all $k \geq 1$. That is, if Y_n has distribution $\hat{F}_n$ then $\{Y_n^k\}$ is uniformly integrable for every k.

Since $\hat{F}_n \stackrel{\mathcal{D}}{\Rightarrow} G$, this implies that $\int x^{2k} d\hat{F}_n(x) \to \int x^{2k} dG(x)$ for all $k \geq 1$ and by (c) we have even moments of F_n converge to corresponding moments of G a.s.

If Part: From (c), $\mathrm{E}[\beta_{2k}(F_n)] \to \beta_{2k}(G)$. It now follows from (b) that $\beta_{2k}(G) = \int x^{2k} dG(x) \leq \frac{(2k)!\Delta(L,f)^k}{k!2^k}$ for all $k \geq 1$ and hence G satisfies Carleman's condition. The rest of the proof is now immediate by the standard moment method arguments. □

Remark 3. *It easily follows from the above theorem that in all the theorems in the next section which establish LSD for different matrices, the moments of the ESD (and also their expectations) converge to the corresponding moments of LSD almost surely if the inputs are uniformly bounded.*

Each matrix has its own contributing words: Of all the matrices we have introduced so far, the symmetric circulant, the palindromic Toeplitz, the palindromic Hankel and the doubly symmetric Hankel matrices achieve the full limit. That is, $p(w) = 1$ for all words. For all other matrices, only certain words contribute in the limit. Below we define two special type of words which arise.

Catalan words: A matched word w with $l(w) = 2k$ and $|w| = k$ will be called a Catalan word if (i) there is at least one *double* letter and (ii) sequentially deleting all double letters eventually leads to the empty word. For example, $aabbcc, abccbdda$ are Catalan words whereas $abab, abccab$ are not. The following result is well known.

Lemma 3. *The number of Catalan words of length* $2k$ *is* $\frac{(2k)!}{(k+1)!k!}$.

Proof. Let us sketch a random walk proof which will be useful while discussing the S matrix. Mark the first occurrence of any letter by $+1$ and the second occurrence by -1, obtaining a sequence $(u_1, u_2, \cdots, u_{2k})$ of $+1$ and -1 of length $2k$. For example, the Catalan words $abba$ and $abccbdda$ are represented respectively by $(1, 1, -1, -1)$ and $(1, 1, 1, -1, -1, 1, -1, -1)$. This actually provides a bijection between the Catalan words of length $2k$ and sequences $(u_1, u_2, \cdots, u_{2k})$ satisfying the following three conditions:
(I) Each u_i equals $+1$ or -1,
(II) $\sum_1^{2k} u_i = 0$ i.e. there are equal number of $+1$ and -1 in the sequence,
(III) $\sum_1^{l} u_i \geq 0 \ \forall \ l \geq 1$. We omit the details. By reflection principle, the total number of such paths is easily seen to be $\frac{(2k)!}{(k+1)!k!}$. □

Symmetric words: A matched word w of length $2k$ with $|w| = k$ is *symmetric* if each letter occurs exactly once in an odd and exactly once in an even position. Otherwise it is called asymmetric. The following result is immediate.

Lemma 4. *There are* $k!$ *symmetric words of length* $2k$ *and any Catalan word is a symmetric word.*

As we shall see later, to achieve non trivial LSD, as $n \to \infty$ (and $p \to \infty$), the scaling for all the matrices is $n^{-1/2}$ except S. For S, the scaling is 1 when $p/n \to y \neq 0, \infty$. The centering in all the above cases is zero. If $p/n \to 0$, then the S matrix has a scaling of $\sqrt{\frac{n}{p}}$ and a centering I_p. The role of different words will be clear once we deal with the different matrices in details. Table 1 summarizes the results. We provide the value of $p(w)$ for different types of matched w. The last column gives the limiting moments.

4 Examples

In the previous section we have developed a general theory based on the volume method of [BDJ] [7]. We now proceed to apply this machinery to different matrices introduced in Section 1.

Table 1: Contribution of words to moments.

MATRIX	w Cat.	w sym., not Cat.	Other w	β_{2k}
Sym. circ.	1	1	1	$\frac{(2k)!}{2^k k!}$
Palin. Toep.	1	1	1	ditto
Palin. Hank.	1	1	1	ditto
Dbly sym. Hank.	1	1	1	ditto
Rev. circ.	1	1	0	$k!$
Wigner	1	0	0	$\frac{(2k)!}{k!(k+1)!}$
S $(p/n \to 0)$	1 (Cat. in p)	0	0	$\beta_k = \frac{(2k)!}{k!(k+1)!}$
S $(p/n \to y \neq 0, \infty)$	1	0	0	$\beta_k = \sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} y^t$
Hankel	1	$0 < p_H(w) = p_T(w) < 1$	0	$\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k} \leq k!$
Toeplitz	1	$0 < p_H(w) = p_T(w) < 1$	$0 < p_T(w) < 1$	$\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k} \leq \frac{(2k)!}{k! 2^k}$

4.1 Wigner matrix: the semicircular law

The **semi-circular law** W arises as the LSD of $n^{-1/2}W_n$. It has the density function

$$p_W(s) = \begin{cases} \frac{1}{2\pi}\sqrt{4-s^2} & \text{if } |s| \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

All its odd moments are zero. The even moments are given by

$$\beta_{2k}(W) = \int_{-2}^{2} s^{2k} p_W(s) ds = \frac{(2k)!}{k!(k+1)!}. \tag{33}$$

Wigner (1955) [20] assumed the entries $\{x_i\}$ to be i.i.d. real Gaussian and established the convergence of E(ESD) of $n^{-1/2}W_n$ to the semi-circular law (32). Assuming the existence of finite moments of all orders, Grenander (1963, pages 179 and 209) [11] established the convergence of the ESD in probability. Arnold (1967) [1] obtained almost sure convergence under the finiteness of the fourth moment of the entries. Bai (1999) [2] has generalised this result by considering Wigner matrices whose entries above the diagonal are not necessarily identically distributed and have no moment restrictions except that they have finite variance.

Recall that the entries of the Wigner matrix satisfies

$$W_n(i,j) = x_{L(i,j)}, 1 \leq i,j \leq n, \quad \text{where } L(i,j) = (\min(i,j), \max(i,j)).$$

This L function satisfies Property B with $f(x) = x$.

We are now ready to state and give a short proof of the LSD of the Wigner matrix.

Theorem 4. *Let $\{w_{ij} : 1 \leq i \leq j,\ j \geq 1\}$ be a double sequence of independent random variables with $\mathrm{E}(w_{ij}) = 0$ for all $i \leq j$ and $\mathrm{E}(w_{ij}^2) = 1$ which are either (i) uniformly bounded or (ii) identically distributed. Let W_n be an $n \times n$ Wigner matrix with the entries w_{ij}. Then with probability one, $F^{n^{-1/2}W_n}$ converges weakly to the semicircular law W given in (32).*

Proof. In view of Theorems 2 and 3, it is enough to assume (i) and prove the required almost sure weak convergence.

Then in view of the discussion in Section 3, it is enough to show that (for matched word w),

$$\frac{1}{n^{1+k}}|\Pi^*(w)| \to 1 \text{ or } 0 \text{ depending on whether or not } w \text{ is a Catalan word.} \tag{34}$$

Note that if $\pi \in \Pi^*(w), w[i] = w[j] \Rightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))$. Then

EITHER $(\pi(i-1),\ \pi(i)) = (\pi(j-1),\ \pi(j))$ (constraint R1)

OR $(\pi(i-1),\ \pi(i)) = (\pi(j),\ \pi(j-1))$ (constraint R2).

For any matched word w, there are k such constraints. Since each constraint is either R1 or R2, there are at most 2^k choices in all. Let λ denote a typical choice of k constraints and $\Pi^*_\lambda(w)$ denote the subset of $\Pi^*(w)$ corresponding to λ and so we have the following finite disjoint representation:

$$\Pi^*(w) = \bigcup_\lambda \Pi^*_\lambda(w). \tag{35}$$

Fix w and λ. For $\pi \in \Pi^*_\lambda(w)$, consider the graph with $(2k+1)$ vertices $\pi(0), \pi(1), \cdots, \pi(2k)$. By abuse of notation, $\pi(i)$ thus denotes both, a vertex and its numerical value. Vertices within the following pairs are connected with a single edge:

(i) the pairs $(\pi(i-1), \pi(j-1))$ and $(\pi(i), \pi(j))$ if $w[i] = w[j]$ yields constraint R1.

(ii) the pairs $(\pi(i'-1), \pi(j'))$ and $(\pi(i'), \pi(j'-1))$ if $w[i'] = w[j']$ yields constraint R2.

(iii) the pair $(\pi(0), \pi(2k))$, ensuring that π is indeed a circuit.

So, the graph has a total of $(2k+1)$ edges. These may include loops and double edges. In any connected component, the numerical values of the vertices are of course equal.

Clearly, the number of generating vertices for π = the number of connected components and this is bounded by $(k+1)$.

Now suppose λ is such that the graph has $(k+1)$ connected components. By pigeon hole principle, there must exist a vertex say $\pi(i)$ which is connected to itself. Note that this is possible if and only if $w[i] = w[i+1]$ and R2 is satisfied. But this implies that there is a double letter $w[i]w[i+1]$ in w. Consider the reduced word w' of length $2(k-1)$ after removing this double letter. We claim that the reduced word still has a double letter.

To show this, in the original graph, coalesce the vertices $\pi(i-1)$ and $\pi(i+1)$. Delete the vertex $\pi(i)$ and remove the R2 constraint edges $(\pi(i-1), \pi(i+1))$ and $(\pi(i), \pi(i))$ but retain all the other earlier edges. For example, any *other* edge that might have existed earlier between $\pi(i-1), \pi(i+1)$ is now a loop. This gives a new graph of degree 2 with $2k+1-2 = 2(k-1)+1$ vertices and has k connected components. Proceeding as before, there must be a self-edge implying a double letter yy in w'. Proceeding inductively, after k steps, we are left with just a single vertex with a loop. In other words, w is Catalan and *all* constraints are R2.

Conversely, it is easy to verify that if w is Catalan and all constraints in λ are R2, then the number of connected components in the graph is indeed $(k+1)$. To see this, essentially retrace the steps given above. First identify a double letter. This gives an R2 constraint. Remove it and proceed inductively. For example, coalesced vertices will fall in the same connected component. We omit the details.

Denote by λ_0 the case when all constraints are R2. Note that

$$|\Pi^*_{\lambda_0}(w)| = n^{k+1}. \tag{36}$$

On the other hand suppose w is Catalan and not all constraints are R2, or, w is not Catalan and λ is arbitrary. Then the corresponding graph has at most k connected components and hence $|\bigcup_{\lambda \neq \lambda_0} \Pi^*_\lambda(w)| \leq 2^k n^k$ implying

$$\frac{1}{n^{k+1}}|\bigcup \Pi^*_{\lambda\neq\lambda_0}(w)| \to 0. \tag{37}$$

Now, (34) follows by combining (35), (36) and (37), and the proof is complete. □

Corollary 4.1. *Let $\{w_{ij} : 1 \leq i \leq j,\ j \geq 1\}$ be a double sequence of independent random variables such that $\{w_{ij} : i < j\}$ are i.i.d. with $\mathrm{Var}(w_{12}) = 1$. Suppose there exists $\epsilon_n \to 0$, such that $n^{1/2}\epsilon_n \to \infty$ and $\sum_{i=1}^n P(|w_{ii}| \geq n^{1/2}\epsilon_n) = o(n)$. Let W_n be an $n \times n$ Wigner matrix with the entries w_{ij}. Then with probability one, $F^{n^{-1/2}W_n}$ converges weakly to the semicircular law.*

Proof of Corollary 4.1. Let $\tilde{w}_{ii}$ be an i.i.d. sequence (independent of $\{w_{ij} : i < j\}$) having same distribution as that of w_{12}. Let $\widetilde{W}_n$ be the matrix obtained from W_n replacing diagonal entries w_{ii} with $\tilde{w}_{ii}$. Now $\widetilde{W}_n$ is a standard Wigner matrix and $F^{n^{-1/2}\widetilde{W}_n}$ converges to semicircular law. Take $E_n^{(1)}$ and $E_n^{(2)}$ as diagonal matrices with entries $\{-\tilde{w}_{ii}\}$ and $\{w_{ii}\}$ respectively. Then (H2) of Theorem 1 is satisfied with $k_n = n$, $\alpha_n = 1$ and $w(n) = n^{1/2}\epsilon_n$. Writing $W_n = \widetilde{W}_n + E_n^{(1)} + E_n^{(2)}$ and appealing to Theorem 1 twice, the corollary follows. □

Remark 4. *If the diagonal entries are i.i.d. μ where μ is an arbitrary probability measure and the off-diagonal entries are i.i.d. ν with variance 1 then W continues to be the LSD (see Bai and Silverstein, 2006) [3]. The weakest known general condition under which the convergence to semicircle law holds was given by Pastur (1972) [16] (see Bai(1999) [2] or, Chatterjee (2005) [8] for a completely new approach). It states that if $\{w_{ij}\}_{1\leq i,j\leq n}$ are independent mean zero, variance one random variable depending on n such that*

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} \sum_{1\leq i,j\leq n} \mathrm{E}\, w_{ij}^2 I(|w_{ij}| > \epsilon\sqrt{n}),$$

then W continues to be the LSD. The above corollary supplements all these results.

4.2 S matrix: the Marčenko-Pastur Law

The LSD of S_n was first established by Marčenko and Pastur (1967) [14]. Subsequent work on S_n may be found in Wachter (1978) [19], Jonsson (1982) [13], Yin (1986) [22], Yin and Krishnaiah (1985) [23] and Bai and Yin (1988) [4]. We state below one version. The **Marčenko-Pastur**

law, MPy, (with scale index 1) is indexed by $0 < y < \infty$. It has a positive mass $1 - \frac{1}{y}$ at the origin if $y > 1$. Elsewhere it has a density:

$$MP_y(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

where $a = a(y) = (1-\sqrt{y})^2$ and $b = b(y) = (1+\sqrt{y})^2$.

Theorem 5. *Suppose $\{x_{ij}\}$ are uniformly bounded, independent with mean zero and variance 1. Then*

(a) if $p/n \longrightarrow y \in (0,\infty)$ then F^{S_n} converges almost surely to the Marčenko-Pastur law MPy, given in (38).

(b) if $p/n \longrightarrow 0$ then the ESD of $\sqrt{\frac{n}{p}}(S_n - I_p)$ converges almost surely to the semicircular law W given in (32).

Remark 5. *When $p/n \to y > 0$, if $\{x_{ij}\}$ are i.i.d. with mean zero and variance 1, the same limit in Theorem 5(a) above continues to hold. This is due to Theorem 2(b).*

But when $p/n \to 0$, the i.i.d. unbounded case does not fall under the set up of Theorem 2 (b) and the finite level truncation used in Theorem 2 is not the optimum one. It requires different levels of truncation and more refined analysis. See Bai (1988). His arguments show that if the fourth moment of the underlying variables is finite, then the limit in Theorem 2 (b) continues to hold. For simplicity we have dealt with only the bounded case. In a separate paper we will deal with general matrices of the form XX' where we shall discuss this issue in more details.

The moments of the Marčenko-Pastur law with ratio index $y > 0$ are given by (see Bai, 1999 [2] or Bai and Silverstein, 2006 [3]),

$$\beta_k(MPy) = \sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} y^t, \ k \ge 1.$$

The following Lemma establishes the connection between these moments and the semicircle moments. This result is also available in Jonsson (1982) [13].

Lemma 5. *For every positive integer k,*

$$\sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} = \frac{(2k)!}{(k+1)!k!} = \beta_{2k}(W).$$

Proof of Lemma 5. Observe that the left side of the above expression equals the coefficient of x^k in $(1+x)^{k-1}\int(1+x)^k dx$. Which in turn is the coefficient of x^k in $\frac{1}{k+1}(1+x)^{k-1}(1+x)^{k+1}$ = coefficient of x^k in $\frac{1}{k+1}(1+x)^{2k}$. The latter is indeed the right side of the above expression.□

If $y = 1$ (for example if $p = n$) then $\beta_k(MPy) = \beta_{2k}(W) \ \ \forall \ \ k \ge 1$.

Suppose $X \sim$ semicircular law and $Y \sim$ Marčenko-Pastur law (with $y = 1$), then from the above, $Y \stackrel{\mathcal{D}}{=} X^2$. If $y \le 1$ then we still have the inequality, $\beta_k(MP_y) \le \beta_{2k}(W)$.

4.2.1 Proof of Theorem 5 when $p/n \to y > 0$.

Recall the definition of (the pair of) L functions as

$$L1, L2 : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}_{\geq} \times \mathbb{Z}_{\geq},\ L1(i,j) = (i,j),\ L2(i,j) = (j,i).$$

Then the k-th spectral moment of the S matrix can be written as follows:

$$\beta_k(S_n) = p^{-1} n^{-k} \sum_{\pi} x_{L1(\pi(0),\pi(1))} x_{L2(\pi(1),\pi(2))} x_{L1(\pi(2),\pi(3))} x_{L2(\pi(3),\pi(4))} \cdots x_{L2(\pi(2k-1),\pi(2k))}.$$

A circuit π now has the *non uniform* range $1 \leq \pi(2m) \leq p,\ 1 \leq \pi(2m+1) \leq n$. It is said to be matched if it is matched within the same $Li,\ i = 1,\ 2$ or across.

In the arguments that follow, we shall profitably use some of the earlier developments for the Wigner matrix. To do this, for any given word w, let $\widetilde{\Pi}(w)$ be the possibly larger class of circuits with the range $1 \leq \pi(i) \leq \max(p,n)$, $1 \leq i \leq 2k$. Likewise define $\widetilde{\Pi}^*(w)$.

Note that Lemma 1 was proved for a single link function but here we have two link functions. Nevertheless, the conclusions remain valid in the present case. Hence, using that and the assumption $p/n \to y > 0$, it is immediate that for some constant C,

$$p^{-1} n^{-k} \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} |\Pi^*(w) - \Pi(w)| \leq C[\max(p,n)]^{-(k+1)} \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} |\widetilde{\Pi}^*(w) - \widetilde{\Pi}(w)| \to 0 \quad (39)$$

and

$$p^{-1} n^{-k} \sum_{\substack{w \text{ has all edges} \\ \text{of order 3 or more}}} |\Pi(w)| \leq C[\max(p,n)]^{-(k+1)} \sum_{\substack{w \text{ has all edges} \\ \text{of order 3 or more}}} |\widetilde{\Pi}(w)| \to 0. \quad (40)$$

Thus in $E[\beta_k(S_n)]$ only the matched circuits with exactly order two edges potentially contribute positively and we need to calculate

$$\lim_n \sum_w \sum_{\pi \in \Pi(w)} n^{-k} p^{-1} \operatorname{E}[x_{L1(\pi(0),\pi(1))} \cdots x_{L2(\pi(2k-1),\pi(2k))}] = \lim_n \sum_{w:\ \text{matched},\ |w|=k} \frac{|\Pi^*(w)|}{n^k p}.$$

To obtain a potentially non zero contribution, we need *exactly* $(k+1)$ independent vertices (hence k dependent vertices). There is a striking similarity between the restrictions imposed on the circuits by Wigner link function and by the pair of link functions $L1, L2$ for S matrix. Note that $L1(\pi(i-1), \pi(i)) = L2(\pi(j-1), \pi(j))$ for $i \neq j$ implies an R2 constraint. On the other hand, $Lt(\pi(i-1), \pi(i)) = Lt(\pi(j-1), \pi(j))$, $t = 1$ or 2, yields an R1 constraint. However, unlike the Wigner matrix, $w[i] = w[j]$ implies exactly one of the constraints is satisfied (R1 if i and j have same parity and R2 otherwise). Hence there is a *unique* $\bar{\lambda}$ (depending on w) such that $\Pi^*(w) = \Pi^*_{\bar{\lambda}}(w)$.

As before, let λ_0 be the index when all constraints in $\Pi^*_{\lambda_0}(w)$ are R2.

Let $\widetilde{\Pi}^*_{\bar{\lambda}}(w)$ denote the class $\Pi^*_{\lambda}(w)$ but where $1 \leq \pi(i) \leq \max(p,n),\ i = 0, 1, \ldots 2k$.

If w is not Catalan then it is easy to see that $\bar{\lambda} \neq \lambda_0$. Hence from equation (37),

$$n^{-k} p^{-1} |\Pi^*_{\bar{\lambda}}(w)| \leq C[\max(p,n)]^{-(k+1)} |\bigcup_{\lambda \neq \lambda_0} \widetilde{\Pi}^*_{\lambda}(w)| \to 0. \quad (41)$$

It is also easy to see that $\bar{\lambda} = \lambda_0$ iff w is symmetric. In particular, if w is a Catalan word, then $\bar{\lambda} = \lambda_0$ and

$$\Pi^*(w) = \Pi^*_{\lambda_0}(w). \tag{42}$$

Thus from (41), we have a nontrivial contribution in the limit only when w is Catalan and hence $\bar{\lambda} = \lambda_0$ and we need to compute

$$\lim n^{-k} p^{-1} |\Pi^*_{\bar{\lambda}_0}(w)|.$$

Incidentally, we already know that $|\tilde{\Pi}^*_{\lambda_0}(w)| \sim [\max(p,n)]^{k+1}$ for every Catalan word. The difference now is that in $\Pi^*_{\lambda_0}(w)$, the circuits are of non-uniform range. The number of Catalan words is now different from the Wigner case and the limit will depend on the word. We proceed as follows.

Suppose the Catalan word w is generated by $(t+1)$ independent even vertices (with range p) and $(k-t)$ independent odd vertices (with range n), $0 \le t \le k-1$. Note that for such a word,

$$\lim_{n\to\infty} n^{-k} p^{-1} |\Pi^*_{\lambda_0}(w)| = \lim_{n\to\infty} n^{-k} p^{-1} (p^{t+1} n^{k-t}) = y^t.$$

Let $M_{t,k}$= the number of Catalan words with $(t+1)$ independent even vertices and $(k-t)$ independent odd vertices. Then

$$\lim \mathrm{E}[\beta_k(S_n)] = \sum_{t=0}^{k-1} M_{t,k} y^t.$$

To obtain $M_{t,k}$, recall the relation between Catalan words and sequences of $+1$ and -1 given in proof of Lemma 3 of Section 3. Now note that a $+1$ in the even position produces an independent even vertex. So to get $(t+1)$ even independent vertices, we should look for those sequences having exactly t many $+1$ in the even positions. Thus $M_{t,k}$ = number of sequences of length $2k$ satisfying (I)-(III) (given in proof of Lemma 3 in Section 3) and which also have t many $+1$ in the even positions.

Claim.

$$M_{t,k} = \binom{k-1}{t}^2 - \binom{k-1}{t+1}\binom{k-1}{t-1} = \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t}. \tag{43}$$

Thus, we conclude

$$\lim_{n\to\infty} \mathrm{E}[\beta_k(S_n)] = \sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} y^t = \beta_k(MPy).$$

This proves the convergence of the moments of the expected ESD.

Note that (M4) does not follow directly from the development of Section 3 since S matrix has two link functions. Nevertheless, we have seen above its close connection to Wigner matrices and (M4) can be easily verified by appropriate modifications of those arguments. We omit the details. This proves the almost sure convergence of the ESD and proof of Part (a) is now complete once we establish the above Claim.

Proof of Claim, based on reflection principle. Each valid sequence represents a simple random walk starting at origin $(0,0)$ and ending at $(2k,0)$ such that there are

(i) exactly t upward moves in k of the even steps and

(ii) no part of the walk is below the x-axis. This implies that $u_1 = +1$ and $u_{2k} = -1$.

So, consider the walk starting at $(1,1)$ and ending at $(2k-1,1)$ satisfying conditions (i) and (ii). First we count the number of paths without considering the constraint (ii). There are $\binom{k-1}{t}$ many ways of choosing t many plus ones from $(k-1)$ even steps and for each of them there are another $\binom{k-1}{t}$ many ways of choosing t many minus ones from $(k-1)$ odd steps (see Case 1 in Table 2). So, total number of such choices is $\binom{k-1}{t}^2$.

Now consider a path which satisfies (i) but violates (ii). Such a path touches $y=-1$ line at least once and hence has two consecutive upward movements. Consider the *last* time this occurs so that $u_l = +1$ and $u_{l+1} = +1$. We consider a transformation

$$(u_2, u_3, \cdots, u_l = +1, u_{l+1} = +1, \cdots, u_{2k-1}) \mapsto (u_2, u_3, \cdots, u_{l-1}, -1, -1, u_{l+2}, \cdots, u_{2k-1}).$$

The resulting sequence is a path from $(1,1)$ to $(2k-1,-3)$. This is a bijection from the set all of paths from $(1,1)$ to $(2k-1,1)$ satisfying (i) but violating (ii) to set of all paths from $(1,1)$ to $(2k-1,-3)$ having $(t-1)$ many $+1$ in the even steps (see Case 2 in Table 2). Number of such paths is $\binom{k-1}{t-1}\binom{k-1}{t+1}$.

Table 2: Path counting.

Case 1	even	odd	total
$+1$	t	$k-1-t$	$k-1$
-1	$k-1-t$	t	$k-1$
total	$k-1$	$k-1$	$2(k-1)$

Case 2	even	odd	total
$+1$	$t-1$	$k-2-t$	$k-3$
-1	$k-t$	$t+1$	$k+1$
total	$k-1$	$k-1$	$2(k-1)$

This completes the proof of the claim and finishes the proof completely. □

4.2.2 Proof of Theorem 5 when $p/n \to y = 0$

Again first assume that the variables are uniformly bounded. The k-th moment of ESD of $A_p = \sqrt{\frac{n}{p}}(S_n - I_p)$ is given by

$$\beta_k(A_p) = \frac{1}{p}\operatorname{Tr}(A_p^k) = \frac{1}{n^{k/2}p^{1+k/2}}\operatorname{Tr}(XX' - nI)^k.$$

Taking expectation, we have the following representation,

$$\mathrm{E}\big[\beta_k(A_p)\big] = \frac{1}{n^{k/2}p^{1+k/2}}\sum_{\pi}\mathrm{E}\,\mathbb{X}_\pi$$

where

$\mathbb{X}_\pi = \big(x_{\pi(0),\pi(1)}x_{\pi(2),\pi(1)} - \delta_{\pi(0),\pi(2)}\big)\ldots\big(x_{\pi(2k-2),\pi(2k-1)}x_{\pi(2k),\pi(2k-1)} - \delta_{\pi(2k-2),\pi(2k)}\big)$,
$\delta_{ij} = \mathrm{I}\{i=j\}$ and the circuit π as before has the non uniform range $1 \le \pi(2m) \le p$, $1 \le \pi(2m+1) \le n$.

Note that if a pair like $(\pi(2i), \pi(2i-1))$ or $(\pi(2i), \pi(2i+1))$ occurs only once in the product then $\mathrm{E}\,\mathbb{X}_\pi = 0$. Same holds true if for some j, the value of $\pi(2j+1)$ does not occur at least twice among $\pi(1), \pi(3), \ldots, \pi(2k-1)$.

Now we define a graph $G = (V, E)$ as follows: Let $V_1 = \{\pi(2j), 0 \le j \le k\}$ be the set of even vertices and $V_2 = \{\pi(2j-1), 1 \le j \le k\}$ be the set of odd vertices. Again by abuse of notation, $\pi(i)$ denotes both, the vertex and its numerical value. Let $V = V_1 \bigcup V_2$ and $E = \{(\pi(2l), \pi(2l+1)), (\pi(2l+2), \pi(2l+1)) : 0 \le l \le k-1\}$ (multiple edges count as one).

Note that G is always connected. So, $|V| \le |E| + 1$. On the other hand, if $\mathrm{E}\,\mathbb{X}_\pi \ne 0$, then $|E| \le k$ and $|V_2| \le [\frac{k}{2}]$. Therefore, $|V| = |V_1| + |V_2| \le k+1$ and $|V_1| \le k - |V_2| + 1$.

Thus if we fix a matching (out of finitely many choices) among the even vertices and one among the odd vertices $\pi(1), \pi(3), \ldots, \pi(2k-1)$ such that $\mathrm{E}\,\mathbb{X}_\pi \ne 0$ then we can have at most $p^{|V_1|} n^{|V_2|}$ many circuits corresponding to that type of matching. So, the contribution of that term is

$$O\left(\frac{p^{|V_1|} n^{|V_2|}}{p^{k/2+1} n^{k/2}}\right) = O\left((\frac{p}{n})^{k/2 - |V_2|}\right). \tag{44}$$

Suppose k is odd. Then $|V_2| < k/2$. Now the fact $p/n \to 0$ immediately implies from (44), that $\mathrm{E}[\beta_k(A_p)] \to 0$.

Now suppose $k = 2m$ is even. We look for all π which produce nontrivial contribution.

Again from the estimate (44), we must have $|V| = 2m+1, |E| = 2m, |V_2| = m$ and $|V_1| = m+1$. Now, observe the following:

(a) $|V_2| = m$ implies a pair partition of odd vertices. Denote it by a word w of length k. So, $\pi(2i-1) = \pi(2j-1)$ iff $w[i] = w[j]$.

(b) Each pair in E must occur exactly twice.

(c) If for some l, $(\pi(2l), \pi(2l+1)) = (\pi(2l+2), \pi(2l+1))$ or equivalently $\pi(2l) = \pi(2l+2)$, then $\mathrm{E}\,\mathbb{X}_\pi = 0$. So, consecutive even vertices cannot be equal.

(d) Note that (b) and (c) together imply that $\mathrm{E}\,\mathbb{X}_\pi = 1$.

Now suppose,

$$w[i] = w[j] \text{ i.e. } \pi(2i-1) = \pi(2j-1) \tag{45}$$

and they are different from the rest of the odd vertices. If we fixed a word w, then independent of w, there are exactly $N_1(n) = n(n-1)\ldots(n-m+1)$ choices of odd vertices satisfying the pairing imposed by w.

Consider the four pairs of vertices from E, $(\pi(2i-2), \pi(2i-1))$, $(\pi(2i), \pi(2i-1))$, $(\pi(2j-2), \pi(2j-1))$ and $(\pi(2j), \pi(2j-1))$.

By (45) and (b), they have to be matched in pairs among themselves. Also, (c) rules out the possibility that the first pair is matched with the second and, third is matched with fourth. So the other two combinations are the only possibilities. It is easy to verify that this is the same as saying that for the even vertices,

$$L(\pi(2i-2), \pi(2i)) = L(\pi(2j-2), \pi(2j)) \tag{46}$$

where L is the Wigner link function.

Let $\pi^*(i) = \pi(2i)$. This is a circuit π^* of length k. Equation (46) says that the circuit π^* is L-matched. Let $\Pi^*(w)$= all circuits π^* satisfying Wigner link function and let

$$N_2(p,w) = \sum_{\pi^* \in \Pi^*(w)} \mathrm{I}(\pi^*(i) \neq \pi^*(i+1)\ \forall i).$$

Note that depending on whether w is or is not Catalan,

$$\lim_p \frac{1}{p^{m+1}} N_2(p,w) = \lim \frac{1}{p^{m+1}} |\Pi^*(w)| = 1 \text{ or } 0.$$

Hence,

$$\begin{aligned}\lim_{n,p} \mathrm{E}[\beta_k(A_p)] &= \lim_{n,p} \frac{1}{p^{m+1} n^m} \sum_{\substack{w:\text{matched},\\ |w|=m}} N_1(n) N_2(p,w)\\ &= \lim_p \frac{1}{p^{m+1}} \sum_{\substack{w:\text{matched},\\ |w|=m}} |\Pi^*(w)| = \frac{(2m)!}{(m+1)!m!}.\end{aligned}$$

This proves the convergence of the expected ESD for the bounded case. We again omit the verification of (M4). That would establish the almost sure convergence of the ESD. □

4.3 Toeplitz and Hankel matrices with independent inputs

[BDJ] [7] established the LSD of Topelitz and Hankel matrices when the input is i.i.d. with finite second moment. See also Hammond and Miller (2005) [12]. We may prove the following version.

Theorem 6. *Suppose $\{x_i\}$ are independent with mean zero and variance 1 which are either (i) uniformly bounded or (ii) identically distributed. Then with probability one, the ESD of $n^{-1/2}T_n$ and $n^{-1/2}H_n$ converge weakly as $n \to \infty$ to nonrandom symmetric measures, say T and H respectively.*

Remark 6. *[BDJ] [7] also prove that T and H have unbounded support and H is not unimodal. The first six moments of T are $0, 1, 0, 8/3, 0$ and 11. Recently Hammond and Miller (2005) [12] obtained some useful bounds for the moments of the ESD. In particular they show that*

$$\mathrm{E}\left[\beta_{2k}\Big(\frac{T_n}{\sqrt{n}}\Big)\right] \leq \frac{(2k)!}{2^k k!} + O_k\Big(\frac{1}{n}\Big). \tag{47}$$

Note that the first term on the right above is the Gaussian $(2k)$-th moment. Also, compare this with the general bound (31).

4.3.1 The volume method proof

As always, without loss, assume variables are independent but uniformly bounded. Then we may appeal to Theorem 2 to conclude the result for the unbounded i.i.d. case. We reproduce in brief the volume method arguments of [BDJ] [7]. Though they assume the i.i.d. structure, their arguments go through for the present case of uniformly bounded independent inputs.

Recall that the L-functions for the Toeplitz and Hankel matrices are respectively given by $L(i,j) = |i-j|$ and $L(i,j) = i+j$, and they both satisfy Property B.

For the Toeplitz matrix,

$$\Pi(w) = \{\pi : w[i] = w[j] \Leftrightarrow |\pi(i-1) - \pi(i)| = |\pi(j-1) - \pi(j)|\}.$$

$$\Pi^*(w) = \{\pi : w[i] = w[j] \Rightarrow |\pi(i-1) - \pi(i)| = |\pi(j-1) - \pi(j)|\}.$$

[BDJ] [7] showed that for each k (which also follows from our general Lemma 1 (a) and Lemma 2),

$$\beta_{2k+1}(n^{-1/2}T_n) \to 0 \quad a.s.$$

Further, they showed that

$$\beta_{2k}(n^{-1/2}T_n) \to \beta_{2k}(T) \quad a.s.$$

where $\{\beta_{2k}(T)\}$ satisfies Carleman's condition (this also follows immediately from our general Theorem 3).

Similarly, for each k,

$$\beta_{2k+1}(n^{-1/2}H_n) \to 0 \text{ a.s. and } \beta_{2k}(n^{-1/2}H_n) \to \beta_{2k}(H) \text{ a.s.}$$

and again, the fact that $\{\beta_{2k}(H)\}$ satisfies Carleman's condition follows immediately from our general Theorem 3.

They proved that

$$\mathrm{E}[\beta_{2k}(T)] = \lim_{n\to\infty} \frac{1}{n^{1+k}} \sum_{\substack{w \text{ matched,} \\ l(w)=2k, |w|=k}} \sum_{\pi \in \Pi(w)} \mathrm{E}\,\mathbb{X}(\pi) = \lim_{n\to\infty} \frac{1}{n^{1+k}} \sum_{\substack{w \text{ matched,} \\ |w|=k}} |\Pi(w)|.$$

To obtain an expression for $\beta_{2k}(T)$, they proceeded as follows. For each matched word w with $|w| = k$, they showed that (the first equality below follows from Lemma 1),

$$\lim_{n\to\infty} \frac{1}{n^{1+k}} |\Pi(w)| = \lim_{n\to\infty} \frac{1}{n^{1+k}} |\Pi^*(w)| = \lim_{n\to\infty} \frac{1}{n^{1+k}} |\Pi^{**}(w)|.$$

where $\Pi^{**}(w) = \big\{\pi : w[i] = w[j] \Rightarrow \pi(i-1) - \pi(i) + \pi(j-1) - \pi(j) = 0\big\}$.

Let $v_i = \pi(i)/n$. The number of elements in $\Pi^{**}(w)$ can then be expressed as $|\Pi^{**}(w)| = \#\Big\{(v_0, v_1, \cdots, v_{2k}) : v_0 = v_{2k}, v_i \in \{0, 1/n, 2/n, \ldots, (n-1)/n\} \text{ and } v_{i-1} - v_i + v_{j-1} - v_j = 0 \text{ if } w[i] = w[j]\Big\}$.

Let $S = \{0\} \cup \{\min(i,j) : w[i] = w[j], i \neq j\}$ be the set of all indices corresponding to the generating vertices of word w. Clearly, $|S| = k+1$. If $\{v_i\}$ satisfy k equations then each v_i is a unique linear combination of $\{v_j\}$ where $j \in S$ and $j \leq i$. Denoting $\{v_i : i \in S\}$ by v_S, we write $v_i = L_i^T(v_S)\ \forall\ i = 0, 1, \cdots, 2k$. These linear functions $\{L_i^T\}$ depend on the word w but for notational simplification we suppress it. Clearly, $L_i^T(v_S) = v_i$ if $i \in S$ and also summing the k equations would imply $L_{2k}^T(v_S) = v_0$. So,

$$|\Pi^{**}(w)| = \#\big\{v_S : L_i^T(v_S) \in \{0, 1/n, 2/n, \cdots, (n-1)/n\} \text{ for all } i = 0, 1, \cdots, n\big\} \tag{48}$$

Since $\frac{1}{n^{1+k}}|\Pi^{**}(w)|$ is nothing but the $(k+1)$ dimensional Riemann sum for the function $\mathrm{I}(0 \leq L_i^T(v_S) \leq 1,\ \forall\, i \notin S \cup \{2k\})$ over $[0,1]^{k+1}$,

$$\lim_{n\to\infty} \frac{1}{n^{1+k}}|\Pi^{**}(w)| = \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^T(v_S) \leq 1,\ \forall\, i \notin S \cup \{2k\}) dv_S = p_T(w) \text{ say} \quad (49)$$

and

$$\beta_{2k}(T) = \sum_{\substack{w \text{ matched,} \\ l(w)=2k, |w|=k}} p_T(w).$$

Similar arguments will lead to the conclusion

$$\beta_{2k}(H) \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{1}{n^{1+k}} \sum_{\substack{w \text{ matched,} \\ |w|=k}} |\Pi^*(w)|$$

where $\Pi^*(w) = \{\pi : w[i] = w[j] \Rightarrow \pi(i-1) + \pi(i) = \pi(j-1) + \pi(j)\}$. We can proceed similarly to write v_i as a linear combination $L_i^H(v_S)$ of generating vertices for all $i \notin S$. As before, realizing $n^{-(k+1)}|\Pi^*(w)|$ as a Riemann sum we can conclude that it converges to the expression given below. We emphasize that unlike in the Toeplitz case, we do not automatically have $L_{2k}^H(v_S) = v_{2k}$ for every word w. Combining all these observations, we have

$$p_H(w) = \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^H(v_S) \leq 1,\ \forall\, i \notin S \cup \{2k\})\, \mathrm{I}(v_0 = L_{2k}^H(v_S)) dv_S \quad (50)$$

and

$$\beta_{2k}(H) = \sum_{\substack{w \text{ matched,} \\ l(w)=2k, |w|=k}} p_H(w).$$

In the next section, we investigate the moments of T and H and their interrelations.

4.3.2 Toeplitz and Hankel moments

Theorem 7. *(a) For any word w, $0 \leq p_T(w),\ p_H(w) \leq 1$.*

(b) For any Catalan word w, $p_T(w) = p_H(w) = 1$.

(c) If w is not a symmetric word, then $p_H(w) = 0$.

(d) $\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k}(T) \leq \frac{(2k)!}{k!2^k}$, $\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k}(H) \leq k!$.

(e) If w is a symmetric word then $p_T(w) = p_H(w)$.

(f) For each k, $\beta_{2k}(T) \geq \beta_{2k}(H)$.

Proof. By the integral representations (49) and (50) of p_T and p_H, (a) is clear.

(b) Suppose w is a Catalan word of length $2k$ and its set of generating vertices is denoted by S. Note that it is enough to show, for both Toeplitz and Hankel, that for each $j \notin S, \exists\, i \in S, i < j$ such that $v_j = v_i$ with $v_{2k} = v_0$.

To prove the above assertion, we apply induction on $|w|$. So assume that the above claim holds for all words w' such that $|w'| < |w|$. Let i and $i+1$ be the positions of the first occurrence of a double letter xx in w. It implies that $i-1, i \in S$ and $i+1 \notin S$. Recall the notation $v_i = \pi(i)/n$. Note that, for both Toeplitz and Hankel, the restriction due to the presence of this double letter xx translates into the following: $v_{i-1} = v_{i+1}$ and at the same time v_i is arbitrary (between 0 and 1).

Clearly w can be written as $w = w_1 xx w_2$ where the word $w' := w_1 w_2$ is again a Catalan word with $|w'| = k-1$. Further, note that $\mathbf{v} = (v_1, v_2, \ldots, v_{i-1}, v_{i+2}, \ldots, v_{2k})$ satisfies the $(k-1)$ equations imposed by w'. Also, if v_s in $\mathbf{v}$ corresponds to a generating vertex in w', it also corresponds to a generating vertex of w (of course w has an extra generating vertex corresponding to v_i). Since $|w'| < |w|$, by assumption the assertion holds for w'. It is now easy to see that it also holds for w.

(c) We will check that $v_0 = L_{2k}^H(v_S)$ iff w is a symmetric word. Thus if w is not a symmetric word, $p_H(w)$, will be the integral of a k dimensional object in a $(k+1)$ dimensional space, and hence equal zero, proving (c).

Let $t_i = v_{i-1} + v_i$. Since w is pair matched, let $\{(i_s, j_s),\ s \geq 1\}$ be such that $w[i_l] = w[j_l]$ and where $j_s, 1 \leq s \leq k$ is written in ascending order. Then $j_k = 2k$ and

$$t_{i_l} - t_{j_l} = 0,\ 1 \leq l \leq k. \tag{51}$$

If w is symmetric then $v_0 - v_{2k} = (t_1 + t_3 + \ldots + t_{2k-1}) - (t_2 + t_4 + \ldots + t_{2k}) = \sum_{s=1}^{k} \alpha_s (v_{i_s} - v_{j_s}) = 0$ for suitable $\alpha_s \in \{-1, 1\}$.

To show the converse, let $v_0 = L_{2k}^H(v_S)$. We have $v_{2k} = v_{2k} + \sum_{s=1}^{k} \alpha_s (t_{i_s} - t_{j_s})$ for all $\alpha_s \in \{-1, 1\}$. Fix $\alpha_k = 1$. Having fixed $\alpha_k,\ \alpha_{k-1},\ \ldots,\ \alpha_{s-1}$, we choose the value of α_s as follows: (a) if $j_s + 1 \in \{i_m, j_m\}$ for some $m > s$, set $\alpha_s = \alpha_m$ if $j_s + 1 = i_m$ and set $\alpha_s = -\alpha_m$ if $j_s + 1 = j_m$ (b) if there is no such m, choose α_s arbitrarily. By this choice of $\{\alpha_s\}$ we ensure that in $v_{2k} + \sum_{s=1}^{k} \alpha_s (t_{i_s} - t_{j_s})$ the coefficient of each $v_i, i \notin S$ cancels out. Thus, from uniqueness, $L_{2k}^H(v_S) = v_{2k} + \sum_{s=1}^{k} \alpha_s (t_{i_s} - t_{j_s})$.

Therefore, by hypothesis, $v_{2k} + \sum_{s=1}^{k} \alpha_s (t_{i_s} - t_{j_s}) - v_0 = 0$ and hence, all the coefficients of $\{v_i\}$ in LHS is zero. Since $t_{2k} = v_{2k-1} + v_{2k}$ and coefficient of t_{2k} is -1 in the left side, the coefficient of t_{2k-1} must be $+1$ on the right side, so that the coefficient of v_{2k-1} is zero. But then comparing the coefficient of v_{2k-2}, the coefficient of t_{2k-2} has to be -1. Proceeding in this manner, all t_i for odd i must have coefficient $+1$ and all t_i for even i must have coefficient -1. These imply that indices i_s and j_s occur exactly once in an odd and once in an even position. That is, the word is symmetric.

(d) The lower bounds follows from (a). As has been seen already, $0 \leq p_T(w),\ p_T(w) \leq 1$ for each word w and hence $\beta_{2k}(T) \leq \frac{(2k)!}{k! 2^k}$. There are $k!$ symmetric words and hence, $\beta_{2k}(H) \leq k!$, proving (d) completely.

(e) First consider $p_T(w)$. As pointed out earlier, $p_T(w) = \lim_{n \to \infty} \frac{1}{n^{1+k}} |\Pi^{**}(w)|$. Transforming $v_i \mapsto y_i = v_i - 1/2$, we see that $|\Pi^{**}(w)|$ is equal to $\#\big\{(y_0, y_1, \cdots, y_{2k}) : y_0 = y_{2k},\ y_i \in \{-1/2, -1/2 + 1/n, -1/2 + 2/n, \cdots, -1/2 + (n-1)/n\}$ and $y_{i-1} - y_i + y_{j-1} - y_j = 0$ if $w[i] = w[j]\big\}$ and for any fixed symmetric word w,

$$p_T(w) = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \cdots \int_{-1/2}^{1/2} \mathrm{I}(-1/2 \leq L_i^T(y_S) \leq 1/2,\ \forall\ i \notin S) dy_S. \tag{52}$$

Similarly,

$$p_H(w) = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \cdots \int_{-1/2}^{1/2} \mathrm{I}(-1/2 \leq L_i^H(y_S) \leq 1/2,\ \forall\ i \notin S) dy_S. \tag{53}$$

Note that the functions L_i^T and L_i^H are determined by the k equations as imposed by the word restriction. Denote $y_{i-1} + y_i$ and $y_{i-1} - y_i$ by p_i and q_i respectively. Suppose $(y_0, y_1, \cdots, y_{2k})$ satisfies Hankel restriction i.e. $p_i - p_j = 0$ if $w[i] = w[j]$. Now we make a transformation

$$(y_0, y_1, y_2, y_3, \cdots, y_{2k}) \mapsto (y_0, -y_1, y_2, -y_3, \cdots, y_{2k}).$$

This transformation does not change the value of the integral corresponding to $p_H(w)$. Under this volume preserving transformation, $(p_1, p_2, p_3, p_4, \cdots, p_{2k})$ maps to $(q_1, -q_2, q_3, -q_4, \cdots, -q_{2k})$. Now a typical equation in Hankel case looks like $p_r = p_s$ if $w[r] = w[s]$ before transformation. Since w is symmetric, one of r and s will be odd and other will be even. So, under the transformation the equation becomes $(-1)^{r-1}q_r = (-1)^{s-1}q_s$ or, $q_r + q_s = 0$ which is nothing but the corresponding Toeplitz equation for the word w. Hence, $p_H(w) = p_T(w)$. This proves (e).

(f) now follows trivially from (b), (c) and (e). □

Tables 3 and 4 provide a full list of partition words with their type, corresponding range of integration, and their volumes upto order 6. The second column shows the role of Catalan and symmetric words.

Table 3: $\beta_4(T) = 8/3$ and $\beta_4(H) = 2$

Word	Type	T. Range	T. Vol.	H. Range (HR)	H. Vol.
abba	Cat.	I^3	1	$HR = I^3$	1
abab	asymm.	$v_0 - v_1 + v_2 \in I$	2/3	HR $\subset$ hyperplane in $\mathbb{R}^3$	0
aabb	Cat.	I^3	1	$HR = I^3$	1

4.3.3 Unboundedness of T and H

Not many detailed properties of the limit distributions T and H are known. For instance, it is not known whether the measures T and H have densities, though simulation results suggest that they do. In [BDJ] [7], the unboundedness of LSD has been proved, separately for T and H. Now with Theorem 7 in our hand we can give a short combined proof of this fact as follows.

Proposition 1. *Let* $M_{2k} := \sum_{\substack{w:l(w)=2k,\\ w \text{ symmetric}}} p_T(w)$. *Then* $\liminf_k M_{2k}^{1/k} = \infty$. *Consequently,* $\liminf_k \beta_{2k}(H)^{1/k} = \infty$ *and both* T *and* H *have unbounded support.*

Proof. We borrow the basic approach from [BDJ] [7]. Consider relation (A.2) of [BDJ] [7] (see also relation (49)). For any pair-matched word w of length $2k$, not necessarily symmetric, we have $p_T(w) = \mathrm{P}\left(\bigcap_{i=1}^k \left\{\sum_{j=0}^k \eta_{ij} U_j \in [0,1]\right\}\right)$ for some suitable $\eta_{ij} \in \{-1, 0, 1\}$ with $\sum_j \eta_{ij} = 1$ and $\{U_j\}$ are i.i.d. $U(0,1)$.

Table 4: $\beta_6(T) = 11$ and $\beta_6(H) = 11/2$

Word	Type	T. Range	T. Vol	H. Range	H. Vol
abccba	Cat.	I^4	1	I^4	1
abccab	asymm.	$v_0 - v_1 + v_2 \in I$	2/3	contained in some hyperplane in $\mathbb{R}^4$	0
abcbca	asymm.	$v_1 - v_2 + v_3 \in I$	2/3	ditto	0
abcbac	asymm.	$v_1 - v_2 + v_3 \in I$ $v_0 - v_2 + v_3 \in I$	1/2	ditto	0
abcacb	asymm.	$v_0 - v_1 + v_3 \in I$ $v_0 - v_1 + v_2 \in I$	1/2	ditto	0
abcabc	symm., not Cat.	$v_0 - v_1 + v_3 \in I$ $v_0 - v_2 + v_3 \in I$	1/2	$v_0 + v_1 - v_3 \in I$ $-v_0 + v_2 + v_3 \in I$	1/2
abbcca	Cat.	I^4	1	I^4	1
abbcac	asymm.	$v_0 - v_1 + v_3 \in I$	2/3	contained in some hyperplane in $\mathbb{R}^4$	0
abbacc	Cat	I^4	1	I^4	1
abaccb	asymm.	$v_0 - v_1 + v_2 \in I$	2/3	contained in some hyperplane in $\mathbb{R}^4$	0
abacbc	asymm.	$v_0 - v_1 + v_2 \in I$ $v_1 - v_2 + v_3 \in I$	1/2	do	0
ababcc	asymm.	$v_0 - v_1 + v_2 \in I$	2/3	do	0
aabccb	Cat.	I^4	1	I^4	1
aabcbc	asymm.	$v_0 - v_2 + v_3 \in I$	2/3	contained in some hyperplane in $\mathbb{R}^4$	0
aabbcc	Cat.	I^4	1	I^4	1

Let $S_i = \sum_{j=0}^{k} \eta_{ij} Y_j$ where $Y_j := U_j - 1/2,\ j = 1, 2, \ldots, k$ are i.i.d. $U(-1/2, 1/2)$.

Fix $\epsilon > 0$ small. As in [BDJ] [7], define the event $A = \bigcap_j \left(|Y_j| \le \frac{1}{2\epsilon(k+1)}\right)$. Note that conditional on A, $Z_j := \epsilon(k+1)Y_j$ are again i.i.d. $U(-1/2, 1/2)$. Hence,

$$\mathrm{P}(|S_i| > 1/2|A) = \mathrm{P}(\sum_{j=0}^{k} \eta_{ij} Z_j > \epsilon(k+1)/2) \le P(\sum_{j=0}^{k} Z_j > \epsilon(k+1)/2)$$

where we use the fact that $Z_j \stackrel{\mathcal{D}}{=} -Z_j$ and adding more variables can only increase the probability of the event. By Bernstein inequality, the right side is dominated by $2\exp(-C\epsilon^2(k+1))$. This yields, for large enough k,

$$\mathrm{P}(\bigcap_{i=1}^{k} (|S_i| < 1/2)|A) \ge 1 - 2k\exp(-C\epsilon^2(k+1)) \ge 1/2.$$

Therefore, $p_T(w) = \mathrm{P}(\bigcap_i (|S_i| < 1/2)) \ge \mathrm{P}(\bigcap_i (|S_i| < 1/2)|A)P(A) \ge \frac{1}{2}\mathrm{P}(A) = \frac{1}{2}(\epsilon(k+1))^{-(k+1)}$.

Hence $M_{2k} \geq k!\frac{1}{2}(\epsilon(k+1))^{-(k+1)} \geq (3\epsilon)^{-k}$ and the rest of the argument is easy. $\square$

4.4 Application of the volume method to reverse circulant

Assuming that $E|x_i|^3 < \infty$, Bose and Mitra (2003) [6] showed that the LSD R of $\frac{1}{\sqrt{n}} RC_n$ has the density,

$$f_R(x) = |x| \exp(-x^2), \quad -\infty < x < \infty$$

with moments

$$\beta_{2k+1}(R) = 0 \text{ and } \beta_{2k}(R) = k! \text{ for all } k \geq 0.$$

This was established by them by using the normal approximation. We state the following stronger result and provide a quick proof using the volume method.

Theorem 8. *Suppose $\{x_i\}$ are independent with mean zero and variance 1 and are either (i) uniformly bounded or (ii) identically distributed. Then the LSD R, of $n^{-1/2}RC_n$ exists and is identified by $\beta_{2k+1}(R) = 0$ and $\beta_{2k}(R) = k!$ for all $k \geq 0$.*

Proof. Without loss, assume (i). Since the link function satisfies Property B with $f(x) = x$, it is enough to show that

$$\lim_n \mathrm{E}[\beta_{2k}(n^{-1/2}RC_n)] = \sum_{\substack{w:\ l(w)=2k, \text{ and} \\ \text{has only order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi^*(w)| = k! \tag{54}$$

There are $k!$ symmetric words. So, to establish (54), it is enough to show the following two claims.

Claim 1. If w (matched) is not symmetric, then $\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi^*(w)| = 0$.

Claim 2. For any symmetric word if we fix the independent vertices, then each of the remaining vertices has exactly one choice.

Proof of Claim 1. Fix any (pair matched) word w giving a partition $(i_s, j_s), 1 \leq s \leq k$.

So, $|\Pi^{**}(w)| = \sum_{(\nu_1,\nu_2,\ldots,\nu_k)\in\{-1,0,1\}^k} \#\Big\{(v_0, v_1, \cdots, v_{2k}) : v_0 = v_{2k}, v_i \in \{0, 1/n, \ldots, (n-1)/n\}$ and $v_{i_s-1} + v_{i_s} - v_{j_s-1} - v_{j_s} = \nu_s\Big\}$.

Observe that, for $\nu = (\nu_1, \nu_2, \ldots, \nu_k) \in \{-1, 0, 1\}^k$, $v_i = L_i^H(v_S) + a_i^{(\nu)}, i \notin S$ for some integer $a_i^{(\nu)}$.

Let S be the set of generating vertices. Arguing as in the Hankel case, we can easily reach the following equality (compare with (50)),
$\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi^*(w)|$

$$= \sum_{\nu\in\{-1,0,1\}^k} \underbrace{\int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^H(v_S) + a_i^{(\nu)} \leq 1,\ \forall\, i \notin S \cup \{2k\})\, \mathrm{I}(v_0 = L_{2k}^H(v_S) + a_{2k}^{(\nu)}) dv_S.$$

For the integral to be non zero, we must have $v_0 = L_{2k}^H(v_S) + a_{2k}^{(\nu)}$. Following the same recipe for the choice of $\{\alpha_s\}$ as in the proof of Theorem 7 part (c), we get $v_{2k} = v_{2k} + \sum_{s=1}^{k} \alpha_s(t_{i_s} - t_{j_s} - \nu_s) =$

$L_{2k}^H(v_S) + a_{2k}^{(\nu)}$. Hence $v_{2k} + \sum_{s=1}^k \alpha_s(t_{i_s} - t_{j_s}) + a_{2k}^{(\nu)} - v_0 = 0$ and thus coefficient of each v_i in the left side has to be zero including the constant term. This implies w has to be symmetric proving Claim 1.

Proof of Claim 2. Fix the independent vertices. We determine the dependent vertices from left to right. So, suppose a typical restriction due to (matched) word looks like $L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))$ where $i < j$ and $\pi(i-1), \pi(i)$ and $\pi(j-1)$ have already been determined. We need to find the number of choices of $\pi(j)$. The above equation can be rewritten as

$$\pi(j) \bmod n = Z \quad \text{where } Z = (L(\pi(i-1), \pi(i)) - \pi(j-1)) \bmod n \in \{0, 1, \ldots, n\}.$$

It is obvious that $\pi(j)$ can be determined uniquely from the above equation since $1 \leq \pi(j) \leq n$. We proceed inductively from left to right to get the whole circuit (i.e. $\pi(0) = \pi(2k)$) from only the independent vertices, uniquely.

This completes the proof of Claim 2 and the proof of the Theorem is complete. □

4.5 Doubly symmetric and palindromic Hankel and Toeplitz matrices

Recall that the symmetric circulant has the link function $n/2 - |n/2 - |i-j||$. This may be considered as a "doubly symmetric" version of the Toeplitz matrix whose link function is $|i-j|$.

Likewise we may consider the doubly symmetric Hankel matrix as DH_n whose link function is

$$L(i,j) = n/2 - |n/2 - (i+j) \bmod n|.$$

On the other hand, [MMS] [15] define a (symmetric) matrix to be palindromic if its first row is a palindrome. In particular, they discussed the palindromic Toeplitz matrix PT_n and also the palindromic Hankel matrix PH_n. They establish the Gaussian limit for $F^{n^{-1/2}PT_n}$ and $F^{n^{-1/2}PH_n}$. Their approach is as follows: it is known from Hammond and Miller (2005) [12] that the LSD for T_n is linked to solutions of some Diophantine equations with certain obstructions. [MMS] show that for PT_n, these restrictions are absent, yielding the Gaussian limit. They also observe a direct relation between PH_n and PT_n to obtain the same conclusion for the former. Combining their observations and ours, we have the following result with a short proof.

Theorem 9. *(a) For any input sequence, if the LSD of any one of $n^{-1/2}DH_n$ and $n^{-1/2}PH_n$ exists then the other also exists and they are equal.*

(b) For any input sequence, (i) if the LSD of any one of $n^{-1/2}PT_n$ and $n^{-1/2}SC_n$ exist, then the other also exists and they are equal and (ii) $(PT_n)^{2k} = (PH_n)^{2k}$ for every k.

(c) If the input sequence $\{x_i\}$ is independent with mean zero and variance 1 and are either (i) uniformly bounded or (ii) identically distributed, then the LSD of all four matrices $n^{-1/2}PT_n$, $n^{-1/2}SC_n$, $n^{-1/2}PH_n$ and $n^{-1/2}DH_n$ are standard Gaussian G with $\beta_{2k}(G) = \frac{(2k)!}{2^k k!}$.

Proof. Note that all the above four matrices are closely related to each other. In particular,

(i) the $n \times n$ principal minor of DH_{n+3} is PH_n.

(ii) the $n \times n$ principal minor of SC_{n+1} is PT_n.

Hence by the interlacing inequality, the claims for the equality of the LSD follow immediately.

For the rest of the proof, we may as usual, without loss, assume that the variables are uniformly bounded. Further, from Theorem 3, if the LSD exist then they have mean zero and the LSD may be identified by the convergence of the expected moments of the ESD.

Let J_n be the matrix with entries 1 in the main anti-diagonal and zero elsewhere. Then it is easy to see that

$$(PH_n)J_n = J_n(PH_n) = PT_n$$

and since $J_n^2 = I_n$,

$$(PT_n)^{2k} = (PH_n)^{2k}.$$

This shows that all four matrices will have the same LSD.

It is thus enough to show that the even moments of the symmetric circulant converge to the Gaussian moments.

There are $\frac{(2k)!}{2^k k!}$ (pair) matched word w with $l(w) = 2k$. Hence it is enough to show that for each such word,

$$\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi^*(w)| = 1 \tag{55}$$

Let $u_l = \pi(l) - \pi(l-1)$. If $w[i] = w[j]$ for $i < j$, then $|n/2 - |u_i|| = |n/2 - |u_j||$. This leads to the following six possibilities in all:

$$u_i - u_j = 0, \pm n \quad \text{OR} \quad u_i + u_j = 0, \pm n.$$

We now show that the first three possibilities are asymptotically negligible.

Lemma 6. *Fix a pair matched word w with $|w| = k$. Let N be the number of matched circuits π in $\{1, 2, \ldots, n\}$ of length $2k$ having at least one pair $i < j$ with $w[i] = w[j]$ such that $u_i - u_j = 0, \pm n$. Then, as $n \to \infty$, $N = O(n^k)$ and hence $n^{-(k+1)}N \to 0$.*

Proof. Let $(i_1, j_1), (i_2, j_2), \ldots (i_k, j_k)$ denote the pair partition corresponding to the word w, i.e. $w[i_l] = w[j_l], 1 \le l \le k$. Suppose, w.l.o.g. $u_{i_k} - u_{j_k} = 0, \pm n$. Clearly a circuit π becomes completely specified if we know $\pi(0)$ and all the 'slopes' $\{u_i\}$.

As already observed, if we fix some value for u_{i_l}, there are at most six options for u_{j_l}. We may choose the values of $\pi(0), u_{i_1}, u_{i_2}, \ldots, u_{i_{k-1}}$ in $O(n^k)$ ways and then we may choose values of $u_{j_1}, u_{j_2}, \ldots, u_{j_{k-1}}$ in $O(6^k)$ ways. For any such choice, from the sum restriction $\sum_{i=1}^{2k} u_i = \pi(2k) - \pi(0) = 0$ we know $t_{i_k} + t_{j_k}$ and on the other hand by hypothesis, $u_{i_k} - u_{j_k} = 0, +n, -n$. Thus the pair (u_{i_k}, u_{j_k}) has 6 possibilities. Thus there are most $O(n^k)$ circuits with the given restrictions and the proof of the Lemma is complete. □

It now remains to show that with $\Pi'(w) = \{\pi : \pi \text{ is a circuit }, w[i] = w[j] \Rightarrow u_i + u_j = 0, \pm n\}$, $\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi'(w)| = 1$. Suppose for some $i < j, u_i + u_j = 0, \pm n$. If we know the circuit up to position $(j-1)$ then $\pi(j)$ has to take one of the values $A - n, A, A + n$ where $A = \pi(j-1) - \pi(i) + \pi(i-1)$. Noting $-(n-2) \le A \le (2n-1)$, exactly one of the three values will fall within 1 and n and be a valid choice for $\pi(j)$. Thus we first choose the independent vertices arbitrarily, then the dependent vertices are determined, from left to right uniquely, so that $u_i + u_j = 0, \pm n$. This automatically yields $\pi(0) = \pi(2k)$ as follows: $\pi(2k) - \pi(0) = \sum_{i=1}^{2k} u_i = dn$ for some $d \in \mathbb{Z}$. But since $|\pi(2k) - \pi(0)| \le n - 1$, we must have $d = 0$. Thus $|\Pi'(w)| = n^{k+1}$ and hence $\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi'(w)| = 1$, proving the theorem. □

5 Matrices with dependent entries

We now illustrate the volume method with a few new results for matrices with dependent entries.

Theorem 10. *Suppose $x_t = \epsilon_t \epsilon_{t+1} \cdots \epsilon_{t+d-1}$. Suppose $\{\epsilon_i\}$ are independent with mean zero and variance 1 and are either (i) uniformly bounded or (ii) identically distributed. Let $A_{n,d} = ((x_{L(i,j)}))$ be a sequence of random matrices with a one-dimensional link function L having property B with $f(x) = x$. Suppose ESD of $n^{-1/2}A_{n,d}$ converges to a nonrandom G almost surely when $d = 1$. Then the same LSD continues to hold almost surely for any $d \geq 2$.*

Since there is dependence between the $\{x_i\}$, we will have to deal with more complicated matchings.

Matched circuits: Let L be a link function. Consider the $d \times h$ matrix, $M^\pi = ((m_{i,j}))$ where $m_{i,j} = L(\pi(j-1), \pi(j)) + i - 1$, $1 \leq i \leq d$, $1 \leq j \leq h$. We now say that π is d-matched if every element of M^π appears at least twice. This notion is extended to d-joint matching and d-cross matching in the obvious way. Note the following facts:

1. No two entries belonging to the same column of M^π can be equal.

2. If some entry in the j_1-th column of M^π is equal to some entry in its j_2-th column then $|L(\pi(j_1 - 1), \pi(j_1)) - L(\pi(j_2 - 1), \pi(j_2))| \leq d - 1$.

Let $N^M_{h,3+}$= Number of d-matched circuits of length h with at least one entry of M^π repeated at least thrice.

Similarly, let $Q^M_{h,4}$= Number of circuits $(\pi_1, \pi_2, \pi_3, \pi_4)$ of length h which are jointly d- matched and jointly d-cross-matched with respect to (L, f) with $f(x) = x$.

Lemma 7. *Suppose (L, f) with $f(x) = x$ satisfies Property B. Then*

(a) there is a constant $C_{h,d}$ such that

$$N^M_{h,3+} \leq C_{h,d} n^{\lfloor (h+1)/2 \rfloor}. \tag{56}$$

(b) there exists a constant $K_{h,d}$ such that,

$$Q^M_{h,4} \leq K_{h,d} n^{2h+2}. \tag{57}$$

(c) Suppose $x_t = \epsilon_t \epsilon_{t+1} \cdots \epsilon_{t+d-1}$. Suppose $\{\epsilon_i\}$ are uniformly bounded, independent with mean zero and variance 1. Let $A_{n,d} = ((x_{L(i,j)}))$ be a sequence of random matrices with a link function L having Property B with $f(x) = x$. Then for every h,

$$\mathrm{E}\left[\frac{1}{n}\mathrm{Tr}\left(\frac{A_{n,d}}{\sqrt{n}}\right)^h - \mathrm{E}\,\frac{1}{n}\mathrm{Tr}\left(\frac{A_{n,d}}{\sqrt{n}}\right)^h\right]^4 = \mathrm{O}(n^{-2}) \tag{58}$$

and hence (M4) holds.

Proof. Let $g : \mathbb{N} \to \mathbb{Z}_{\geq}$ be defined as $g(x) = \lfloor x/d \rfloor$. Obviously 0 has only d many preimages under g. So (L, g) satisfies Property B. If π is as given, then from the observations made earlier, π has at least one (L, g) match of order three or more. So an application of Lemma 1 completes the proof of (a).

(b) follows immediately from Lemma 2 by noting that the set under consideration is contained in $\big\{(\pi_1, \pi_2, \pi_3, \pi_4):$ they are jointly and cross-matched with respect to $g\big\}$.

The proof of (c) uses (b) and we skip the details. □

Lemma 8. *Each d-matched circuit π with only pair matchings is also pair-matched w.r.t. L and vice versa. Hence if $l(\pi) = h$ is odd, then no d-matched circuit π can be pair-matched.*

Proof. Let $\alpha_1 = \min_t L(\pi(t-1), \pi(t))$. Since π is d-matched, there exist at least two entries from the matrix M^π whose values equal α_1. These entries have to come from the first row since the elements in each column are strictly increasing. Let $1 \le j_1 \ne j_2 \le h$ be such that $L(\pi(j_1), \pi(j_1)) = L(\pi(j_2), \pi(j_2)) = \alpha_1$. Then the j_1-th column of the matrix M^π will be entirely matched with the j_2-th column. Now, since M^π contains only pair-matchings, no entry in the columns j_1 and j_2 can share the same values with entries from the rest of the matrix. Now drop these two columns and repeat the above procedure. Clearly, if h is odd, we cannot have a d-pair-matching. On the other hand, if h is even, we have concluded that the columns of M^π are pairwise matched with one another. This automatically induces an L-pair-matching for π and we are done. Other direction is trivial. □

Sketch of proof of Theorem 10. First assume that $\{\epsilon_i\}$ is uniformly bounded. Let $F_{n,d}$ denote the ESD of $n^{-1/2}A_{n,d}$. Lemma 7 and 8 imply that for every h, d,

$$\beta_h(F_{n,d}) - \mathrm{E}[\beta_h(F_{n,d})] \to 0 \text{ almost surely.}$$

On the other hand, the h-th expected moment of the ESD is given by

$$\mathrm{E}[\beta_h(F_{n,d})] = \frac{1}{n}\,\mathrm{E}\big[\mathrm{Tr}(n^{-1/2}A_{n,d})^h\big] = \frac{1}{n^{1+h/2}}\sum_{\pi}\mathrm{E}(\mathbb{X}_\pi) = \frac{1}{n^{1+h/2}}\sum_{\pi\ d\text{-matched}}\mathrm{E}(\mathbb{X}_\pi)$$

where $\mathbb{X}_\pi = \prod_{i=1}^h \epsilon_{L(\pi(i-1),\pi(i))}\epsilon_{L(\pi(i-1),\pi(i))+1}\cdots\epsilon_{L(\pi(i-1),\pi(i))+d-1}$.

Lemma 7(a) and Lemma 8, imply that if h is odd, then $\lim \mathrm{E}[\beta_h(F_{n,d})] = 0$ and hence for every d, $\lim \beta_h(F_{n,d}) = 0$ almost surely.

Now suppose $h = 2k$ is even. Let $\Pi(w)$ be as defined in Section 3 for ordinary L-matching. From Theorem 3, almost surely,

$$\lim n^{-(k+1)}\sum_{w:|w|=k}\Pi(w) = \lim \beta_{2k}(n^{-1/2}A_{n,1}) = \lim \mathrm{E}[\beta_{2k}(n^{-1/2}A_{n,1}) = \beta_{2k}(G).$$

On the other hand, Lemma 7 and Lemma 8 imply that for all d,

$$\lim n^{-(k+1)}\sum_{w:|w|=k}\Pi(w) = \lim \mathrm{E}[\beta_{2k}(F_{n,d})] = \lim \beta_{2k}(F_{n,d}), \quad \text{almost surely.}$$

This proves the theorem for the uniformly bounded case.

Now assume that $\{x_i\}$ is i.i.d. Then by following the arguments given in the proof of Theorem 2 it may be shown that the same LSD persists. □

We now explore another dependent situation. Though the following results can possibly be extended to other matrices, we will restrict our attention to Toeplitz and Hankel matrices.

Theorem 11. *Suppose* $X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$ *where* $\{a_j\}$ *satisfies* $\sum_j |a_j| < \infty$ *and* $\{\epsilon_i\}$ *are independent with mean zero and variance 1 such that either (i)* $\{\epsilon_i\}$ *are uniformly bounded or (ii) identically distributed and* $\sum_j j a_j^2 < \infty$. *Then with probability one, ESD of* $T_n/\sqrt{n}$ *and* $H_n/\sqrt{n}$, *with entries coming from* $\{X_t\}$, *converge weakly to non-random symmetric probability measures* T_a *and* H_a *respectively. These LSD do not depend on the distribution of* ϵ_1. *Further, the* $(2k)$*-th moment of* T_a *and* H_a *are given by*

$$\beta_{2k}(T_a) = A_{2k}\beta_{2k}(T) \text{ and } \beta_{2k}(H_a) = A_{2k}\beta_{2k}(H)$$

where

$$A_{2k} = \sum_{d=0}^{\infty}\Bigg(\sum_{\substack{m_1,m_2,\ldots,m_k \geq 0:\\ \sum_1^k m_i = d}} \prod_{j=1}^{k} a_{m_j}\Bigg)^2$$

and T *and* H *are as defined in Theorem 6.*

Remark 7. *Observe that* $A_{2k} \leq A^{2k}$ *where* $A = \sum_j |a_j|$. *Thus it follows from Theorem 11 that the limiting moments are bounded above by the Gaussian moments, that is,* $\beta_{2k}(T_a), \beta_{2k}(H_a) \leq E(Z^{2k})$ *where* $Z \sim N(0, A^2)$. *In some easy cases* $\{A_{2k}\}$ *can be explicitly calculated. For example, if* $X_t = \epsilon_t + \theta\epsilon_{t-1}$ *then* $A_{2k} = \sum_{i=0}^{k} \binom{k}{i}^2 \theta^{2i}$.

Proof of Theorem 11. We provide an outline of the proof. First assume that $\{\epsilon_j\}$ is bounded by B. This implies that $\{X_t\}$ is also uniformly bounded by AB. In the rest of the proof, L will stand for either Toeplitz or Hankel link function and B_n will stand for T_n or H_n.

Matched Circuits. Fix $h \geq 1$ and $\mathbf{m} \in \mathbb{Z}_{\geq}^{h}$. In view of the overlap of the ϵ_j in the different X_i, now it is most natural to define the following. Denote

$$\lambda_{\mathbf{m}(\pi)} = \big(L(\pi(0), \pi(1)) - m_1, L(\pi(1), \pi(2)) - m_2, \cdots, L(\pi(h-1), \pi(h)) - m_h\big).$$

A circuit π is now called $\mathbf{m}$-matched if for $1 \leq i \leq L(\pi), \exists\, j \neq i$ such that $\lambda_{\mathbf{m}(\pi)}(i) = \lambda_{\mathbf{m}(\pi)}(j)$. Fix $h \geq 1$. Define $\Pi^*_{\mathbf{m}}(w) = \{\pi : \pi \text{ is a circuit}, l(\pi) = h, w[i] = w[j] \Rightarrow \lambda_{\mathbf{m}(\pi)}(i) = \lambda_{\mathbf{m}(\pi)}(j)\}$.

We now have the following two Lemmata similar to Lemma 1(a) and Lemma 2(a). We omit their proofs.

Lemma 9. *Fix* $h \geq 1$ *and* $\mathbf{m} \in \mathbb{Z}_{\geq}^{h}$.

(a) Suppose w *is a matched word one letter occurring at least thrice. Then for some constant* C_h *independent of* $\mathbf{m}$,

$$|\Pi^*_{\mathbf{m}}(w)| \leq C_h n^{\lfloor (h+1)/2 \rfloor}. \tag{59}$$

(b) For some A_h *independent of* $\mathbf{m}$,

$$\sum_{w:\text{ matched}} |\Pi^*_{\mathbf{m}}(w)| \leq A_h n^{1+h/2}. \tag{60}$$

Lemma 10. *Let* $h \geq 1$ *and* $\mathbf{m_i} \in \mathbb{Z}_{\geq}^{h}$ *for* $i = 1, 2, \cdots, 4$. *Then there exists a constant* K_h *independent of* $\mathbf{m_1}, \ldots, \mathbf{m_4}$, *such that*

$$\#\{(\pi_1, \pi_2, \pi_3, \pi_4) : \pi_i \text{ are jointly } \mathbf{m}\text{-matched and cross-}\mathbf{m}\text{-matched}\} \leq K_h n^{2h+2}. \tag{61}$$

From the trace formula,
$$\frac{1}{n^{1+h/2}}\,\mathrm{E}\,\mathrm{Tr}(B_n^h) = \frac{1}{n^{1+h/2}}\sum_{\pi}\mathrm{E}(\mathbb{X}_\pi)$$
where
$$\mathbb{X}_\pi = \prod_{i=1}^{h} X_{L(\pi(i-1),\pi(i))} = \sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}a_{m_2}\cdots a_{m_h}\epsilon(\lambda_{\mathbf{m}}(\pi)),\ \text{ and } \epsilon(\lambda_{\mathbf{m}}(\pi)) = \prod_{i=1}^{h}\epsilon_{\lambda_{\mathbf{m}(\pi)}(i)}.$$
Hence, from the mean zero and independence assumption on $\{\epsilon_j\}$, we have
$$\frac{1}{n^{1+h/2}}\,\mathrm{E}\,\mathrm{Tr}(B_n^h) = \frac{1}{n^{1+h/2}}\sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}a_{m_2}\cdots a_{m_h}\sum_{\substack{\pi:\pi \text{ is}\\ \mathbf{m}\text{-matched}}}\mathrm{E}(\epsilon(\lambda_{\mathbf{m}}(\pi))).$$
Consider the following string of equalities:
$$\lim_{n\to\infty}\frac{1}{n^{1+h/2}}\,\mathrm{E}\,\mathrm{Tr}(B_n^h) = \lim_{n\to\infty}\frac{1}{n^{1+h/2}}\,\mathrm{E}\left[\sum_{\pi}\sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}\cdots a_{m_h}\epsilon(\lambda_{\mathbf{m}}(\pi))\right] \tag{62}$$
$$= \lim_{n\to\infty}\frac{1}{n^{1+h/2}}\,\mathrm{E}\left[\sum_{\mathbf{m}\geq\mathbf{0}}\sum_{\pi} a_{m_1}\cdots a_{m_h}\epsilon(\lambda_{\mathbf{m}}(\pi))\right] \tag{63}$$
$$= \lim_{n\to\infty}\frac{1}{n^{1+h/2}}\sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}\cdots a_{m_h}\sum_{\substack{\pi:\pi \text{ is}\\ \mathbf{m}\text{-matched}}}\mathrm{E}[\epsilon(\lambda_{\mathbf{m}}(\pi))] \tag{64}$$
$$= \sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}\cdots a_{m_h}\lim_{n\to\infty}\frac{1}{n^{1+h/2}}\sum_{\substack{\pi:\pi \text{ is}\\ \mathbf{m}\text{-matched}}}\mathrm{E}[\epsilon(\lambda_{\mathbf{m}}(\pi))] \tag{65}$$
Above equalities follow from repetitive application of DCT: we use the uniform boundedness of $\{\epsilon_j\}$ and absolute summability of $\{a_i\}$ in (63) and(64). In (65) we use Lemma 9(b) which implies that $|\frac{1}{n^{1+h/2}}\sum_{\substack{\pi:\pi \text{ is}\\ \mathbf{m}\text{-matched}}}\mathrm{E}[\epsilon_{\lambda_{\mathbf{m}}(\pi)}]| \leq A_h B^h \quad \forall n$ where B is the uniform bound on $\{\epsilon_j\}$.
Now, by applying Lemma 9(a), if $h = 2k-1$, then $\lim_{n\to\infty}\frac{1}{n^{1+h/2}}\,\mathrm{E}\,\mathrm{Tr}(B_n^h) = 0$ and if $h = 2k$ then
$$\lim_{n\to\infty}\frac{1}{n^{1+k}}\,\mathrm{E}\,\mathrm{Tr}(B_n^{2k}) = \sum_{\mathbf{m}\geq\mathbf{0}} a_{m_1}\cdots a_{m_{2k}}\lim_{n\to\infty}\frac{1}{n^{1+k}}\sum_{\substack{w:\ w \text{ is}\\ \text{matched, } |w|=k}}|\Pi^*_{\mathbf{m}}(w)|. \tag{66}$$
We will now consider only the Toeplitz case i.e. when $B_n = T_n$. Similar arguments are applicable to the Hankel case. Let $u_i = \pi(i) - \pi(i-1)$. Fix some (matched) word w with $|w| = k$ and $\pi \in \Pi^*_{\mathbf{m}}(w)$. This implies $|u_i| - m_i = |u_j| - m_j$ whenever $w[i] = w[j], i < j$. Arguments similar to those given in Proposition 4.4 in [BDJ] [7] or Lemma 6 imply that the number of circuits in which u_i and u_j have same signs for at least one pair (i,j) with $w[i] = w[j]$ is asymptotically negligible. Now let $0 < i_1 < i_2 < \ldots < i_k$ be independent vertices for w and $(i_1,j_1),(i_2,j_2),\ldots,(i_k,j_k)$ be the pair partition of $\{1,2,\ldots,2k\}$ induced by w. Define
$$\Pi'_{\mathbf{f}}(w) = \{\pi : \pi \text{ is a circuit }, u_{i_l} + u_{j_l} = f_l, 1 \leq l \leq k\} \text{ where } \mathbf{f} = (f_1,f_2,\ldots,f_k) \in M \subset \mathbb{Z}^k$$

and $M = \{(\pm(m_{i_1} - m_{j_1}), \pm(m_{i_2} - m_{j_2}), \ldots, \pm(m_{i_k} - m_{j_k})\}$. Sum restriction imposed on $\{u_i\}$ by the circuit condition implies $\Pi'_{\mathbf{f}}(w) = \emptyset$ if $\sum_{i=1}^{k} f_i \neq 0$. We now make a crucial observation that if $\sum_{i=1}^{k} f_i = 0$ then $\lim_n n^{-(k+1)}|\Pi'_{\mathbf{f}}(w)| = p_T(w)$ as $\mathbf{f}/n \to \mathbf{0}$. Clearly,

$$\lim_n \frac{1}{n^{k+1}}|\Pi^*_{\mathbf{m}}(w)| = \lim_n \sum_{\mathbf{f}:\mathbf{f}\in M} \frac{1}{n^{k+1}}|\Pi'_{\mathbf{f}}(w)| = \sum_{\mathbf{f}:\mathbf{f}\in M \text{ and } \sum_{i=1}^{k} f_i=0} p_T(w).$$

Thus after interchanging the order of summation, we get RHS of (66) is equal to

$$\sum_{\substack{w:w \text{ matched} \\ |w|=k}} p_T(w)\Bigg(\sum_{\substack{\mathbf{f}:\mathbf{f}\in\mathbb{Z}^k \\ \text{and } \sum_{i=1}^{k} f_i=0}} \prod_{j=1}^{k} b(|f_j|)\Bigg) \text{ where } b(j) = \sum_{m\geq 0} a_m a_{m+j}, j \geq 0 \tag{67}$$

The term in the big parenthesis does not involve w. It is also equal to the factor given in the statement in Theorem 11. It is easy to show that Carleman's condition is satisfied.

Also by using Lemma 10, it can be shown that

$$\mathrm{E}\left[\frac{1}{n}\mathrm{Tr}(B_n/\sqrt{n})^h - \frac{1}{n}\mathrm{E}[\mathrm{Tr}(B_n/\sqrt{n})^h]\right]^4 = O(n^{-2}).$$

This shows that the convergence of moments happens almost surely. So, we have established the result for the bounded case.

Let us now turn to the case when $\{\epsilon_j\}$ are i.i.d. mean zero, variance one but not necessarily uniformly bounded. From Phillips and Solo (1992) if $\sum_j j a_j^2 < \infty$ then we have the following SLLN:

$$\frac{1}{n}\sum_{1}^{n} X_t^2 \xrightarrow{a.s.} \mathrm{E}(X_1^2) = \sum_{j}^{\infty} a_j^2. \tag{68}$$

Now we may give the following truncation argument to reduce to the earlier case.

Let $\{\epsilon^*_i\}$ be the truncated version of $\{\epsilon_i\}$ as in (15). Let $X^*_t = \sum_{j=0}^{\infty} a_j \epsilon^*_{t-j}$. We have already established the Theorem for T^*_n, the Toeplitz matrix constructed from $\{X^*_t\}$. Following the steps of Theorem 2 and using (68) it can be easily shown that $\limsup_n d_{BL}(F^{n^{-1/2}T_n}, F^{n^{-1/2}T^*_n}) \to 0$ almost surely as the truncation level goes to infinity. Thus the Theorem remains true for the Toeplitz matrix constructed using $\{X_t\}$. This proves the Theorem completely. □

Acknowledgement. The authors are grateful to G.J. Babu of Pennsylvania State University for his help. They also appreciate the constructive comments of the anonymous Referee.

References

[1] Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20, 262–268. MR0217833

[2] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9, 611–677 (with discussions). MR1711663

[3] Bai, Z. D. and Silverstein, J. (2006). *Spectral Analysis of Large Dimensional Random Matrices.* Science Press, Beijing.

[4] Bai, Z. D. and Yin, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.*, 16, no. 2, 863–875. MR0929083

[5] Bhatia, R. (1997). *Matrix Analysis.* Springer, New York. MR1477662

[6] Bose, A. and Mitra, J. (2002). Limiting spectral distribution of a special circulant. *Stat. Probab. Letters*, 60, 1, 111–120. MR1945684

[7] [BDJ] Bryc, W., Dembo, A. and Jiang, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34, no. 1, 1–38. MR2206341

[8] Chatterjee, S. (2005). A simple invariance theorem. Available at *http://arxiv.org/abs/math.PR/0508213.*

[9] Chow, Y. S. and Teicher, H. (1997). *Probability theory: Independence, interchangeability, martingales.* Third edition, Springer-Verlag, New York. MR1476912

[10] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications.* Vol. 2, Wiley, New York. MR0210154

[11] Grenander, U. (1963). *Probabilities on Algebraic Structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Göteborg-Uppsala. MR0206994

[12] Hammond, C. and Miller, S. J. (2005). Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* 18, no. 3, 537–566. MR2167641

[13] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* 12, no. 1, 1–38. MR0650926

[14] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices, (Russian) *Mat. Sb. (N.S.)* 72 (114), 507–536. MR0208649

[15] Massey, A., Miller, S. J. and and Sinsheimer, J. (2007). Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices. *J. Theoret. Probab.* 20, 3, 637–662. MR2337145

[16] Pastur, L. (1972). The spectrum of random matrices. (Russian) *Teoret. Mat. Fiz.* 10 no. 1, 102–112. MR0475502

[17] Phillips P., Solo V. (1992). Asymptotics for linear processes *Ann. Stat.* (2), 20, 971–1001. MR1165602

[18] Sen, A. (2006). Large dimensional random matrices. *M. Stat. Project Report.*, May 2006. Indian Statistical Institute, Kolkata.

[19] Wachter, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* 6, 1–18. MR0467894

[20] Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, (2), 62, 548–564. MR0077805

[21] Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, (2), 67, 325–327. MR0095527

[22] Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20, no. 1, 50–68. MR0862241

[23] Yin, Y. Q. and Krishnaiah, P. R. (1985). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.*, 30, no. 4, 810–816. MR0816299