Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2010, Vol. 46, No. 3, 677–707
DOI: 10.1214/09-AIHP329

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Limiting spectral distribution of XX' matrices

Arup Bose[a], Sreela Gangopadhyay[a] and Arnab Sen[b]

[a] *Statistics and Mathematics Unit, Indian Statistical Institute, 202 B. T. Road, Kolkata 700108, INDIA.*
E-mail: bosearu@gmail.com; abose@isical.ac.in
[b] *Department of Statistics, UC Berkeley, 367 Evans Hall Berkeley, CA 94720, USA*

Received 5 August 2008; revised 23 June 2009; accepted 16 July 2009

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 covariance matrix. In a recent article Bryc, Dembo and Jiang [*Ann. Probab.* **34** (2006) 1–38] establish the LSD for 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 equivalence classes and relating the limits of the counts to certain volume calculations. Bose and Sen [*Electron. J. Probab.* **13** (2008) 588–628] have developed this method further and have provided a general framework which deals with symmetric matrices with entries coming from an independent sequence.

In this article we enlarge the scope of the above approach to consider matrices of the form $A_p = \frac{1}{n}XX'$ where X is a $p \times n$ matrix with real entries. We establish some general results on the existence of the spectral distribution of such matrices, appropriately centered and scaled, when $p \to \infty$ and $n = n(p) \to \infty$ and $p/n \to y$ with $0 \le y < \infty$. As examples we show the existence of the spectral distribution when X is taken to be the appropriate asymmetric Hankel, Toeplitz, circulant and reverse circulant matrices. In particular, when $y = 0$, the limits for all these matrices coincide and is the same as the limit for the symmetric Toeplitz derived in Bryc, Dembo and Jiang [*Ann. Probab.* **34** (2006) 1–38]. In other cases, we obtain new limiting spectral distributions for which no closed form expressions are known. We demonstrate the nature of these limits through some simulation results.

Résumé. Une des méthodes pour obtenir la limite des distributions spectrales (LSD) des grandes matrices aléatoires est la fameuse méthode des moments, basée sur la formule des traces. Son succès a été clairement établi pour différents types de matrices telles que les matrices de Wigner et les matrices de covariance. Dans un article récent, Bryc, Dembo et Jiang [*Ann. Probab.* **34** (2006) 1–38] ont obtenu la LSD pour des matrices de Toeplitz et de Hankel en utilisant cette méthode. Ils arrivent à estimer les traces des moments de telles matrices en séparant les différents termes par classes d'équivalence et en reliant les asymptotiques des dénombrements afférents avec les calculs de certains volumes. Bose et Sen [*Electron. J. Probab.* **13** (2008) 588–628] ont développé cette idée et ont donné un cadre général pour traiter de matrices symétriques dont les entrées viennent d'une suite indépendante.

Dans cet article, nous généralisons cette approche pour considérer des matrices de la form $A_p = \frac{1}{n}XX'$ où X est une matrice $p \times n$ avec des entrées réelles. Nous démontrons un résultat général d'existence de la LSD de telles matrices, correctement recentrées et rééchelonnées, quand p et n tendent vers l'infini de telle façon que p/n tende vers $y \in (0, \infty)$. Par exemple, nous montrons l'existence de la LSD quand X est la matrice asymétrique de Hankel, de Toeplitz, circulante ou circulante inverse. En particulier, quand $y = 0$, les limites correspondent à celles obtenues par Bryc, Dembo et Jiang [*Ann. Probab.* **34** (2006) 1–38]. Sinon, nous obtenons de nouvelles lois limites pour lesquelles aucune expression explicite n'est connue. Nous étudions ces lois par quelques simulations.

MSC: 60F05; 60F15; 62E20; 60G57

Keywords: Large dimensional random matrix; Eigenvalues; Sample covariance matrix; Toeplitz matrix; Hankel matrix; Circulant matrix; Reverse circulant matrix; Spectral distribution; Bounded Lipschitz metric; Limiting spectral distribution; Moment method; Volume method; Almost sure convergence; Convergence in distribution

1. Introduction

1.1. *Random matrices and spectral distribution*

Random matrices were introduced in mathematical statistics by Wishart [20]. In 1950 Wigner successfully used random matrices to model the nuclear energy levels of a disordered particle system. The seminal work, Wigner [19] laid the foundation of this subject. This theory has grown into a separate field now, with application in many branches of sciences that includes areas as diverse as multivariate statistics, operator algebra, number theory, signal processing and wireless communication.

Spectral properties of random matrices is in general a rich and attractive area for mathematicians. In particular, when the dimension of the random matrix is large, the problem of studying the behavior of the eigenvalues in the bulk of the spectrum has arisen naturally in many areas of science and has received considerable attention. Suppose A_p is a $p \times p$ random matrix. Let $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ (or $\mathbb{C}$) denote its eigenvalues. When the eigenvalues are real, our convention will be to always write them in ascending order. The empirical spectral measure μ_p of A_p is the random measure on $\mathbb{R}$ (or $\mathbb{C}$) given by

$$\mu_p = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}, \tag{1}$$

where δ_x is the Dirac delta measure at x. The random probability distribution function on $\mathbb{R}$ (or $\mathbb{C}$) corresponding to μ_p, is known as the *Empirical Spectral Distribution* (ESD) of A_p. We will denote it by F^{A_p}.

If $\{F^{A_p}\}$ converges weakly (as p tends to infinity), either almost surely or in probability, to some (nonrandom) probability distribution, then that distribution is called the *Limiting Spectral Distribution* (LSD) of $\{A_p\}$. Proving existence of LSD and deriving its properties for general patterned matrices has drawn significant attention in the literature. We refer to Bai [1] and Bose and Sen [6] for information on several interesting situations where the LSD exists and can be explicitly specified.

Possibly the two most significant matrices whose LSD have been extensively studied, are the Wigner and the sample covariance matrices. For simplicity, let us assume for the time being that all random sequences under consideration are independent and uniformly bounded.

1. *Wigner matrix*. In its simplest form, the Wigner matrix $W_n^{(s)}$ [18,19], 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 $\{x_{ij}\colon 1 \le i \le j\}$, we can visualize the Wigner matrix as

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

It is well known that almost surely,

$$\text{LSD of } n^{-1/2} W_n^{(s)} \text{ is the } \textit{semicircle law } W, \tag{3}$$

with the density function

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

2. *Sample covariance matrix (S matrix)*. Suppose $\{x_{jk}\colon j, k = 1, 2, \dots\}$ is a double array of i.i.d. real random variables with mean zero and variance 1. The matrix

$$A_p(W) = n^{-1} W_p W_p' \quad \text{where } W_p = \big((x_{ij})\big)_{1 \le i \le p, 1 \le j \le n} \tag{5}$$

is called a sample covariance matrix (in short an S matrix). Let I_p denote the identity matrix of order p. The following results are well known. See [1] or [6] for moment method and Stieltjes transform method proofs. To have the historical idea of the successive development of these results see [3] for (a) and [9,11,13,17,21,22] for (b).

(a) If $p \to \infty$ and $p/n \longrightarrow 0$ then almost surely,

$$\text{LSD of } \sqrt{\frac{n}{p}}\big(A_p(W) - I_p\big) \text{ is the semicircle law given in (4) above.} \tag{6}$$

(b) If $p \to \infty$ and $p/n \longrightarrow y \in (0, \infty)$ then almost surely,

$$\text{LSD of } A_p(W) \text{ is the Marčenko–Pastur law } MP_y \text{ given in (8) below.} \tag{7}$$

The *Marčenko–Pastur law*, MP_y, has a positive mass $1 - \frac{1}{y}$ at the origin if $y > 1$. Elsewhere it has a density:

$$MP_y(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

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

Suppose $p = n$. Then W_n above is a square matrix with i.i.d. elements, and the Wigner matrix $W_n^{(s)}$ is obtained from W_n by retaining the elements above and on the diagonal of W_n and letting symmetry dictate the choice of the rest of the elements of $W_n^{(s)}$. We loosely say that W_n is the *asymmetric* version of $W_n^{(s)}$. It is interesting to note the following:

(i) The LSD of $n^{-1/2}W_n^{(s)}$ in (3) and that of $\sqrt{\frac{n}{p}}(\frac{1}{n}W_pW_p' - I_p)$ in (6) are identical.

(ii) If W and M are random variables obeying respectively, the semicircle law and the Marčenko–Pastur law with $y = 1$, then $M \stackrel{\mathcal{D}}{=} W^2$ in law. That is, the LSD of $n^{-1/2}W_n^{(s)}$ and that of $\frac{1}{n}W_pW_p'$ bears this squaring relationship when $p/n \to 1$.

1.2. *The problem and its motivation*

In view of the above discussion, it thus is natural to study the LSD of matrices of the form $A_p(X) = (1/n)X_pX_p'$ where X_p is a $p \times n$ suitably patterned (asymmetric) random matrix. Asymmetry is used loosely. It just means that X_p is not necessarily symmetric. In particular one may ask the following questions.

(i) Suppose $p/n \to y, 0 < y < \infty$. When does the LSD of $A_p(X) = \frac{1}{n}X_pX_p'$ exist?

(ii) Suppose $p/n \to 0$. When does the LSD of $\sqrt{\frac{n}{p}}(A_p(X) - I_p)$ exist? Note that unlike the S matrix, the mean of $A_p(X)$ is not in general equal to I_p and hence other centering/normalisations may also need to be allowed.

(iii) Suppose X_p is a $p \times n$ (asymmetric) patterned matrix and $A_p(X) = \frac{1}{n}X_pX_p'$ for which the limit in (ii) holds. Call it A. Now consider the square symmetric matrix $X_n^{(s)}$ obtained from X_p by letting $p = n$ and symmetrising it by retaining the elements of X_p above and on the diagonal and allowing symmetry to dictate the rest of the elements. Suppose as $n \to \infty$, the LSD of $n^{-1/2}X_n^{(s)}$ exists. Call it B. In general, is there any relation between B and A?

A further motivation to study the LSD of $\frac{1}{n}X_pX_p'$ comes from wireless communications theory. Many results on the information-theoretic limits of various wireless communication channels make substantial use of asymptotic random matrix theory, as the size of the matrix increases. An extended survey of results and works in this area may be found in [16]. Also see [15] for a review of some of the existing mathematical results that are relevant to the analysis of the properties of random matrices arising in wireless communications. The study of the asymptotic distribution of the *singular values* is seen as an important aspect in the analysis and design of wireless communication channels.

A typical wireless communication channel may be described by the linear vector memoryless channel:

$$y = X_p\theta + \varepsilon,$$

where θ is the n-dimensional vector of the signal input, y is the p-dimensional vector of the signal output, and the p dimensional vector ε is the additive noise. X_p, in turn, is the $p \times n$ random matrix, generally with complex entries.

Silverstein and Tulino [15] emphasize the asymptotic distribution of the squared singular-values of X_p under various assumptions on the joint distribution of the random matrix coefficients where n and p tend to infinity while the aspect ratio, $p/n \to y, 0 < y < \infty$. In their model, generally speaking, the channel matrix X_p can be viewed as $X_p = f(A_1, A_2, \dots, A_k)$ where $\{A_i\}$ are some independent random (rectangular) matrices with complex entries, each having its own meaning in terms of the channel. In most of the cases they studied, some A_i have all i.i.d. entries, while the LSD of the other matrices $A_j A_j^*$, $j \neq i$ are assumed to exist. Then the LSD of $X_p X_p^*$ is computed in terms of the LSD of $A_j A_j^*$, $j \neq i$.

Specifically, certain CDMA channels can be modeled (e.g. see [16], Chapter 3, and [15]) as $X_p = CSA$ where C is a $p \times p$ random symmetric Toeplitz matrix, S is a matrix with i.i.d. complex entries independent of C, and A is an $n \times n$ deterministic diagonal matrix. One of the main theorems in [15] establishes the LSD of $X_p = CSA$ for random $p \times p$ matrix C, not necessarily Toeplitz, under the added assumption that the LSD of the matrices CC' exists. When C is a random symmetric Topelitz matrix, the existence of the LSD of CC' is immediate from the recent work of Bryc, Dembo and Jiang [7] and, Hammond and Miller [10]. This also motivates us to study the LSD of $X_p X_p'$ matrices in some generality where X_p is not necessarily square and symmetric.

1.3. *The moment method*

We shall use the *method of moments* to study this convergence. To briefly describe this method, suppose $\{Y_p\}$ is a sequence of random variables with distribution functions $\{F_p\}$ such that $\mathbb{E}(Y_p^h) \to \beta_h$ for every positive integer h, and $\{\beta_h\}$ satisfies *Carleman's condition* (see [8], page 224):

$$\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty. \tag{9}$$

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

$$\beta_h = \beta_h(F) = \int x^h \,\mathrm{d}F(x) \tag{10}$$

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

Now suppose $\{A_p\}$ is a sequence of $p \times p$ symmetric matrices (with possibly random entries), and let, by a slight abuse of notation, $\beta_h(A_p)$ denote the hth moment of the ESD of A_p. Suppose there is a sequence of nonrandom $\{\beta_h\}_{h=1}^{\infty}$ satisfying Carleman's condition such that,

(M1) *First moment condition*: For every $h \geq 1$, $\mathbb{E}[\beta_h(A_p)] \to \beta_h$ and
(M2) *Second moment condition*: For every $h \geq 1$, $\mathrm{Var}[\beta_h(A_p)] \to 0$.

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*: For every $h \geq 1$, $\mathbb{E}[\beta_h(A_p) - \mathbb{E}(\beta_h(A_p))]^4 = \mathrm{O}(p^{-2})$.

For any symmetric $p \times p$ matrix A with eigenvalues $\lambda_1, \dots, \lambda_p$, the *trace formula*

$$\beta_h(A) = p^{-1} \sum_{i=1}^{p} \lambda_i^h = p^{-1} \operatorname{Tr}(A^h)$$

is invoked to compute moments. Moment method proofs are known for the results on Wigner matrix and the sample covariance matrix given earlier. See for example, [6].

1.4. *Brief overview of our results*

Let $\mathbb{Z}$ be the set of all integers. Let $\{x_\alpha\}_{\alpha \in \mathcal{Z}}$ be an *input sequence* where $\mathcal{Z} = \mathbb{Z}$ or $\mathbb{Z}^2$ and $\{x_\alpha\}$ are independent with $\mathbb{E}x_\alpha = 0$ and $\mathbb{E}x_\alpha^2 = 1$. Let $L_p : \{1, 2, \dots, p\} \times \{1, 2, \dots, n = n(p)\} \to \mathcal{Z}$ be a sequence of functions which we call *link functions*. For simplicity of notation, we will write L for L_p. We shall later impose suitable restriction on the link functions L_p and sufficient probabilistic assumptions on the variables. Consider the matrices

$$X_p = \big((x_{L_p(i,j)})\big)_{1\le i\le p, 1\le j\le n} \quad \text{and} \quad A_p = A_p(X) = (1/n)X_p X_p'.$$

The function L_p defines an appropriate pattern and we may call these matrices *patterned* matrices. The assumptions on the link function restrict the number and manner in which any element of the input sequence may occur in the matrix.

We shall consider two different regimes:

Regime I. $p \to \infty$, $p/n \to y \in (0,\infty)$. *In this case, we consider the spectral distribution of* $A_p = A_p(X) = n^{-1}X_p X_p'$.

Regime II. $p \to \infty$, $p/n \to 0$. *In this case we consider the spectral distribution of* $(np)^{-1/2}(X_p X_p' - nI_p) = \sqrt{\frac{n}{p}}(A_p - I_p)$ *where* I_p *is the identity matrix of order* p.

We will use the following assumptions on the input sequences in two regimes.

Assumption R1. *In Regime I*, $\{x_\alpha\}$ *are independent with mean zero and variance one, and are either uniformly bounded or identically distributed.*

Assumption R2. *In Regime II*, $\{x_\alpha\}$ *are independent with mean zero and variance* 1. *Further,* $\lambda \ge 1$ *is such that* $p = \mathrm{O}(n^{1/\lambda})$ *and* $\sup_\alpha \mathbb{E}(|x_\alpha|^{4(1+1/\lambda)+\delta}) < \infty$ *for some* $\delta > 0$.

The moment method requires all moments to be finite. However, in Regime I, our basic assumption will be that the input sequence has finite second moment. To deal with this, we first prove a truncation result which shows that in Regime I, under appropriate assumption on the link function, without loss of generality, we may assume the input sequence to be uniformly bounded. The situation in Regime II is significantly more involved and there we assume existence of higher moments.

We then establish a negligibility result which implies that for nonzero contribution to the limiting moments, only summands in the trace formula which are "pair matched" matter. See next section for details on "matching." The existence of the limit of the sum of pair matched terms and then the computation of the limit establishes the LSD.

Quite interestingly, under reasonable assumptions on the allowed pattern, *if* the limit of empirical moments exist, they automatically satisfy Carleman's condition and thus ensures the existence of the LSD. However, we are unaware of any general existing method of imposing suitable restrictions on the link function to guarantee the existence of limits of moments.

As examples of our method, we let X_p to be the nonsymmetric Toeplitz, Hankel, reverse circulant and circulant matrices. We show that the LSD exists in both Regimes I and II, thereby answering questions (i) and (ii) in the previous section affirmatively for these matrices. The LSD in Regime I are all new.

However, in Regime II, the LSD of all four matrices above are identical to the LSD obtained by Bryc, Dembo and Jiang [7] for the *symmetric* Toeplitz matrix $T_n^{(s)}$. This implies that the answer to question (iii) is not straightforward and needs further investigation.

Closed form expressions for the LSD or for its moments do not seem to be easily obtainable. We provide a few simulations to demonstrate the nature of the LSD and it is an interesting problem to derive detailed properties of the LSD in any of these cases.

1.5. *Examples*

In this section we will consider some specific nonsymmetric matrices of interest. In particular, let X_p be the asymmetric versions of the four matrices Toeplitz, Hankel, circulant and reverse circulant. Then it turns out that in Regime II, the LSD for $\sqrt{\frac{n}{p}}(A_p(X_p) - I_p)$ exists and is identical to the LSD for $n^{-1/2}T_n^{(s)}$ obtained by Bryc, Dembo and Jiang

[7] for the symmetric $n \times n$ Toeplitz matrix $T_n^{(s)}$. In Regime I, the LSD for $A_p(X_p)$ exist but they are all different. The proofs of the results given below are presented in Section 5.

1.5.1. *The asymmetric Toeplitz matrix*

The $n \times n$ symmetric Toeplitz matrix is defined as $T_n^{(s)} = ((x_{|i-j|}))$. It is known that if the input sequence $\{x_i\}$ has mean zero and variance one and is either independent and uniformly bounded or, is i.i.d., then

$$n^{-1/2}T_n^{(s)} \text{ converges weakly almost surely to } \mathcal{L}_T \quad \text{(say)}.$$

See for example [6,7] and [10]. It may be noted that the moments of $\mathcal{L}_T$ may be represented as volumes of certain subsets of the unit hypercubes but are not known in any explicit form.

Let

$$T_p = T = \big((x_{i-j})\big)_{p\times n}, \qquad A_p(T) = n^{-1}T_pT_p'.$$

Note that T_p is the nonsymmetric Toeplitz matrix.

Theorem 1. (i) [*Regime I*] *Assume R*1 *holds. If* $\frac{p}{n} \to y \in (0,\infty)$ *then empirical spectral distribution* $F^{A_p(T)}$ *converges in distribution almost surely to a nonrandom distribution which does not depend on the distribution of* $\{x_i\}$.

(ii) [*Regime II*] *Assume R*2 *holds. Then* $F^{\sqrt{\frac{n}{p}}(A_p(T)-I_p)}$ *converges in distribution to* $\mathcal{L}_T$ *almost surely.*

1.5.2. *The asymmetric Hankel matrix*

Suppose $\{x_i, i = 0, \pm1, \pm2, \ldots\}$ is an independent input sequence. Let $H_n^{(s)} = ((x_{i+j}))$ be the $n\times n$ symmetric Hankel matrix. It is known that if $\{x_i\}$ has mean zero and variance one and is either uniformly bounded or i.i.d., then

$$F^{n^{-1/2}H_n^{(s)}} \text{ converges weakly almost surely.}$$

See for example [6] and [7].

Now let $H = H_{p\times n}$ be the asymmetric Hankel matrix where the (i,j)th entry is x_{i+j} if $i \ge j$ and $x_{-(i+j)}$ if $i < j$. Let $H_p^{(s)} = ((x_{i+j}))_{p,n}$ be the rectangular Hankel matrix with symmetric link function. Let

$$A_p(H) = n^{-1}H_pH_p' \quad \text{and} \quad A_p\big(H^{(s)}\big) = n^{-1}H_p^{(s)}H_p^{(s)\prime}.$$

We then have the following theorem.

Theorem 2. (i) [*Regime I*] *Assume R*1 *holds. If* $\frac{p}{n} \to y \in (0,\infty)$ *then* $F^{A_p(H)}$ *converges almost surely to a nonrandom distribution which does not depend on the distribution of* $\{x_i\}$.

(ii) [*Regime II*] *Assume R*2 *holds. Then* $F^{\sqrt{\frac{n}{p}}(A_p(H)-I_p)}$ *converges almost surely to* $\mathcal{L}_T$. *The same limit continues to hold if* $A_p(H)$ *is replaced* $A_p(H^{(s)})$.

1.5.3. *The asymmetric reverse circulant matrix*

The *symmetric* Reverse Circulant $R_n^{(s)}$ has the link function $L(i,j) = (i+j) \bmod n$. The LSD of $n^{-1/2}R_n^{(s)}$ has been discussed in [5] and [6]. In particular it is known that if $\{x_i\}$ are independent with mean zero and variance 1 and are either (i) uniformly bounded or (ii) identically distributed, then the LSD R of $n^{-1/2}R_n^{(s)}$ exists almost surely and has the density,

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

with moments

$$\beta_{2h+1}(R) = 0 \quad \text{and} \quad \beta_{2h}(R) = h! \quad \text{for all } h \ge 0.$$

Let $R_p^{(s)}$ be the $p \times n$ matrix with link function $L(i, j) = (i + j) \bmod n$ and $R_p = R_{p\times n}$ be the *asymmetric* version of $R_p^{(s)}$ with the link function

$$L(i, j) = \begin{cases} (i + j) \bmod n & \text{for } i \geq j, \\ -\big[(i + j) \bmod n\big] & \text{for } i < j. \end{cases} \tag{11}$$

So, in effect, the rows of R_p are the first p rows of an (asymmetric) Reverse Circulant matrix. Let

$$A_p(R) = n^{-1} R_p R_p' \quad \text{and} \quad A_p\big(R^{(s)}\big) = n^{-1} R_p^{(s)} R_p^{(s)\prime}.$$

We then have the following theorem.

Theorem 3. (i) [*Regime* I] *Assume* R1 *holds. If* $\frac{p}{n} \to y \in (0, \infty)$ *then* $F^{A_p(R)}$ *converges almost surely to a nonrandom distribution which does not depend on the distribution of* $\{x_i\}$.

(ii) [*Regime* II] *Assume* R2 *holds. Then* $F^{\sqrt{\frac{n}{p}}(A_p(R)-I_p)}$ *converges almost surely to* $\mathcal{L}_T$. *The same limit continues to hold if* $A_p(R)$ *is replaced* $A_p(R^{(s)})$.

1.6. *The asymmetric circulant matrix*

The square circulant matrix is well known in the literature. Its eigenvalues can be explicitly computed and are closely related to the periodogram. Its LSD is the bivariate normal distribution. See for example [5]. Its symmetric version is treated in [6] and [14]. Let

$$C_p = C_{p\times n} = \big((x_{L(i,j)})\big) \quad \text{where } L(i, j) = (n - i + j) \bmod n, \quad \text{and} \quad A_p(C) = p^{-1} C_p C_p'.$$

Theorem 4. (i) [*Regime* I] *Assume* R1 *holds. If* $\frac{p}{n} \to y \in (0, \infty)$ *then* $F^{A_p(C)}$ *converges almost surely to a nonrandom distribution which does not depend on the distribution of* $\{x_i\}$.

(ii) [*Regime* II] *Assume* R2 *holds. Then* $F^{\sqrt{\frac{n}{p}}(A_p(C)-I_p)}$ *converges almost surely to* $\mathcal{L}_T$.

2. Basic notation, definitions and assumptions

Examples of link functions that correspond to matrices of interest are as follows:

0. Covariance matrix: $L_p(i, j) = (i, j)$.
1. Asymmetric Toeplitz matrix: $L_p(i, j) = i - j$.
2. Asymmetric Hankel matrix: $L_p(i, j) = \operatorname{sgn}(i - j)(i + j)$ where

$$\operatorname{sgn}(l) = \begin{cases} 1 & \text{if } l \geq 0, \\ -1 & \text{if } l < 0. \end{cases} \tag{12}$$

3. Asymmetric Reverse Circulant: $L_p(i, j) = \operatorname{sgn}(i - j)(i + j) \bmod n$.
4. Asymmetric Circulant: $L_p(i, j) = (n + j - i) \bmod n$.

For any set G, $\#G$ and $|G|$ will stand for the number of elements of G. We shall use the following general assumptions on the link function.

Assumptions on link function L.

A. *There exists positive integer* Φ *such that for any* $\alpha \in \mathcal{Z}$ *and* $p \geq 1$:

(i) $\#\{i\colon L_p(i, j) = \alpha\} \leq \Phi$ *for all* $j \in \{1, 2, \ldots, n\}$ *and*
(ii) $\#\{j\colon L_p(i, j) = \alpha\} \leq \Phi$ *for all* $i \in \{1, 2, \ldots, p\}$.

A′. *For any* $\alpha \in \mathcal{Z}$ *and* $p \geq 1$, $\#\{i\colon L_p(i, j) = \alpha\} \leq 1$ *for all* $j \in \{1, 2, \ldots, n\}$.

B. $k_p \alpha_p = \mathrm{O}(np)$ *where*

$$k_p = \#\{\alpha\colon\ L_p(i,j) = \alpha, 1 \le i \le p, 1 \le j \le n\}, \quad \alpha_p = \max_{\alpha \in \mathcal{Z}} \left|L_p^{-1}(\alpha)\right|.$$

Assumption A stipulates that with increasing dimensions, the number of times any fixed input variable appears in any given row or column, remains bounded. Assumption A′ stipulates that no input appears more than once in any column. Assumption B makes sure that no particular input appears too many times in the matrix. Clearly A′ implies A(i) but B is not related to either A or A′. Consider the link function $L_p(i,j) = (1,1)$ if $i = j$ and $L_p(i,j) = (i,j)$ if $i \ne j$. This link function satisfies A(ii) and A′ but not B. On the other hand if $L_p(i,j) = 1$ for all (i,j) then it satisfies B but not A. It is easy to verify that the link functions listed above satisfy these assumptions. It may also be noted that the symmetric Toeplitz link function does *not* satisfy Assumption A′.

Towards applying the moment method, we need a few notions, most of which are given in details in [6] in the context of symmetric matrices. These concepts and definitions will remain valid in Regime II also, with appropriate changes.

The trace formula. Let $A_p = n^{-1} X_p X_p'$. Then the hth moment of ESD of A_p is given by

$$p^{-1} \operatorname{Tr} A_p^h = p^{-1} n^{-h} \sum_{1 \le i_1, i_2, \ldots, i_h \le n} x_{L_p(i_1,i_2)} x_{L_p(i_3,i_2)} \cdots x_{L_p(i_{2h-1},i_{2h})} x_{L_p(i_1,i_{2h})}. \tag{13}$$

Circuits. Any function $\pi : \{0, 1, 2, \ldots, 2h\} \to \mathbb{Z}_+$ is said to be a *circuit* if

(i) $\pi(0) = \pi(2h)$,
(ii) $1 \le \pi(2i) \le p\ \forall 0 \le i \le h$ and
(iii) $1 \le \pi(2i-1) \le n\ \forall 1 \le i \le h$.

The *length* $l(\pi)$ of π is taken to be $(2h)$. A circuit depends on h and p but we will suppress this dependence.

Matched circuits. Let

$$\xi_\pi(2i-1) = L\big(\pi(2i-2), \pi(2i-1)\big), \quad 1 \le i \le h \quad \text{and}$$
$$\xi_\pi(2i) = L\big(\pi(2i), \pi(2i-1)\big), \quad 1 \le i \le h.$$

These will be called *L-values*.

A circuit π is said to have an *edge of order* $e (1 \le e \le 2h)$ if it has an L-value repeated exactly e times. Any π with *all* $e \ge 2$ will be called *L-matched* (in short *matched*). For any such π, given any i, there is at least one $j \ne i$ such that $\xi_\pi(i) = \xi_\pi(j)$.

Define for any circuit π,

$$\mathbb{X}_\pi = \prod_{i=1}^{h} x_{\xi_\pi(2i-1)} x_{\xi_\pi(2i)}. \tag{14}$$

Due to the mean zero and independence assumption, if π has at least one edge of order one then $\mathbb{E}(\mathbb{X}_\pi) = 0$.

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 $\{\xi_{\pi_1}(i) = \xi_{\pi_1}(j) \Leftrightarrow \xi_{\pi_2}(i) = \xi_{\pi_2}(j)\}$. This defines an equivalence relation between the circuits.

Words. Equivalence classes arising from the above equivalence relation may be identified with partitions of $\{1, 2, \ldots, 2h\}$ – to any partition we associate a *word* w of length $l(w) = 2h$ of letters where the first occurrence of each letter is in alphabetical order. For example, if $h = 3$, then the partition $\{\{1,3,6\},\{2,4\},\{5\}\}$ is represented by the word $ababca$. Let $|w|$ denote the number of different letters that appear in w.

The notion of order e edges, matching, nonmatching for π, carries over to words in a natural manner. For instance, the word $ababa$ 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 since $\mathbb{E}(\mathbb{X}_\pi) = 0$ for nonmatched words. The following fact is obvious.

Word count.

$$\#\{w\colon\ w \text{ is of length } 2h \text{ and has only order 2 edges with } |w| = h\} = \frac{(2h)!}{2^h h!}. \tag{15}$$

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

$$\Pi(w) = \bigl\{\pi\colon\ w[i] = w[j] \Leftrightarrow \xi_\pi(i) = \xi_\pi(j)\bigr\}.$$

If $\pi \in \Pi(w)$, then clearly the number of distinct L-values equals $|w|$. Notationally,

$$\#\bigl\{\xi_\pi(i)\colon\ 1 \le i \le 2h\bigr\} = |w|.$$

Note that now, with the above definition, we may rewrite the trace formula as

$$p^{-1}\operatorname{Tr} A_p^h = p^{-1}n^{-h} \sum_{\pi\colon\ \pi \text{ circuit}} \mathbb{X}_\pi = p^{-1}n^{-h} \sum_w \sum_{\pi \in \Pi(w)} \mathbb{X}_\pi,$$

where the first sum is taken over all words with length $(2h)$. After taking expectation, only the matched words survive and

$$\mathbb{E}\bigl[p^{-1}\operatorname{Tr} A_p^h\bigr] = p^{-1}n^{-h} \sum_{w \text{ matched}} \sum_{\pi \in \Pi(w)} \mathbb{E}\mathbb{X}_\pi.$$

To deal with (M4), we need multiple circuits. The following notions will be useful for dealing with multiple circuits:

Jointly matched and cross-matched circuits. k circuits $\pi_1, \pi_2, \ldots, \pi_k$ 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 class $\Pi^*(w)$. Define for any (matched) word w,

$$\Pi^*(w) = \bigl\{\pi\colon\ w[i] = w[j] \Rightarrow \xi_\pi(i) = \xi_\pi(j)\bigr\}. \tag{16}$$

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

Vertex and generating vertex. Each $\pi(i)$ of a circuit π will be called a *vertex*. Further, $\pi(2i), 0 \le i \le h$ will be called *even vertices* or *p-vertices* and $\pi(2i-1), 1 \le i \le h$ will be called *odd vertices* or *n-vertices*.

A vertex π 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.

We will call an odd generating vertex $\pi(2i-1)$ Type I if $\pi(2i)$ is also an even generating vertex. Otherwise we will call $\pi(2i-1)$ a Type II odd generating vertex.

Obviously, number of generating vertices in π is $|w| + 1$. By Assumption A on L function given earlier, a circuit of a fixed length is completely determined, *up to a finitely many choices* by its generating vertices. Hence, in Regime I under the assumption A, we obtain the simple but crucial estimate

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

Let $[a]$ denote the integer part of a. For Regime II, we will further assume A$'$. In that case, it shall be shown that

$$\bigl|\Pi^*(w)\bigr| = \mathrm{O}\bigl(p^{1+[(|w|+1)/2]} n^{[|w|/2]}\bigr).$$

3. Regime I

In Regime I, $p \to \infty$ such that $p/n \to y \in (0, \infty)$ and we consider the spectral distribution of $A_p = A_p(X) = n^{-1} X_p X_p'$. There are two main theorems in this section. Theorem 5 is proved under Assumption B. We show that the input sequence may be taken to be bounded without loss of generality. This allows us to deal with bounded sequences in all examples later. This result is proved sketchily in [6], with special reference to the sample covariance matrix. We provide a detailed proof for clarity and completeness. Theorem 6 is proved under Assumptions A and B. We show that the ESD of $\{A_p(X)\}$ is almost surely tight and any subsequential limit is sub Gaussian. Invoking the trace formula, we show that the LSD exists iff the moments converge. Further, in the limit, only pair matched terms potentially contribute.

To prove Theorem 5, we will need the following notion and result.

The *bounded Lipschitz metric d_{BL}* is defined on the space of probability measures as:

$$d_{BL}(\mu,\nu)=\sup\left\{\int f\,\mathrm{d}\mu-\int f\,\mathrm{d}\nu:\ \|f\|_\infty+\|f\|_L\le 1\right\} \tag{17}$$

where

$$\|f\|_\infty=\sup_x\bigl|f(x)\bigr|,\qquad \|f\|_L=\sup_{x\neq y}\bigl|f(x)-f(y)\bigr|/|x-y|.$$

Recall that convergence in d_{BL} implies weak convergence of measures.

The following inequalities provide estimate of the metric distance d_{BL} in terms of trace. Proofs may be found in [2] or [1] and uses Lidskii's theorem (see [4], page 69).

Lemma 1. (i) *Suppose A, B are $n\times n$ symmetric real matrices. Then*

$$d_{BL}^2\bigl(F^A,F^B\bigr)\le\left(\frac{1}{n}\sum_{i=1}^n\bigl|\lambda_i(A)-\lambda_i(B)\bigr|\right)^2\le\frac{1}{n}\sum_{i=1}^n\bigl(\lambda_i(A)-\lambda_i(B)\bigr)^2\le\frac{1}{n}\operatorname{Tr}(A-B)^2. \tag{18}$$

(ii) *For any square matrix S, let F^S denote its empirical spectral distribution. Suppose X and Y are $p\times n$ real matrices. Let $A=XX'$ and $B=YY'$ Then*

$$d_{BL}^2\bigl(F^A,F^B\bigr)\le\left(\frac{1}{p}\sum_{i=1}^p\bigl|\lambda_i(A)-\lambda_i(B)\bigr|\right)^2\le\frac{2}{p^2}\operatorname{Tr}(A+B)\operatorname{Tr}\bigl[(X-Y)(X-Y)'\bigr]. \tag{19}$$

Theorem 5. *Let $p\to\infty$, $\frac{p}{n}\to y\in(0,\infty)$ and Assumption B holds so that $\alpha_p k_p=\mathrm{O}(np)$. Suppose for every bounded mean zero and variance one i.i.d. input sequence, $F^{n^{-1}X_pX_p'}$ converges weakly to some fixed nonrandom distribution G almost surely. Then the same limit continues to hold if the input sequence is i.i.d. with mean zero and variance one.*

Proof. Without loss of generality we shall assume that $\mathcal{Z}=\mathbb{Z}$ and also write X for X_p. For $t>0$, denote

$$\mu(t)=\mathbb{E}\bigl[x_0\mathbb{I}\bigl(|x_0|>t\bigr)\bigr]=-\mathbb{E}\bigl[x_0\mathbb{I}\bigl(|x_0|\le t\bigr)\bigr]$$

and let

$$\sigma^2(t)=\operatorname{Var}\bigl(x_0\mathbb{I}\bigl(|x_0|\le t\bigr)\bigr)=\mathbb{E}\bigl[x_0^2\mathbb{I}\bigl(|x_0|\le t\bigr)\bigr]-\mu(t)^2.$$

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

$$x_i^*=\frac{x_i\mathbb{I}(|x_i|\le t)+\mu(t)}{\sigma(t)}=\frac{x_i-\bar{x}_i}{\sigma(t)}\quad\text{where }\bar{x}_i=x_i\mathbb{I}\bigl(|x_i|>t\bigr)-\mu(t)=x_i-\sigma(t)x_i^*. \tag{20}$$

It is easy to see that $\mathbb{E}(\bar{x}_0^2)=1-\sigma^2(t)-\mu(t)^2\to 0$ as t tends to infinity. Further, $\{x_i^*\}$ are i.i.d. bounded, mean zero and variance one random variables. Let us replace the entries $x_{L_p(i,j)}$ of the matrix X_p by the truncated version

$x^*_{L_p(i,j)}$ (respectively $\bar{x}_{L_p(i,j)}$) and denote this matrix by Y (respectively $\bar{X}_p$). By triangle inequality and (19),

$$\begin{aligned}
&d^2_{BL}\big(F^{n^{-1}X_pX'_p}, F^{n^{-1}YY'}\big)\\
&\quad\le 2d^2_{BL}\big(F^{n^{-1}X_pX'_p}, F^{n^{-1}\sigma(t)^2YY'}\big) + 2d^2_{BL}\big(F^{n^{-1}YY'}, F^{n^{-1}\sigma(t)^2YY'}\big)\\
&\quad\le \frac{2}{p^2n^2}\operatorname{Tr}\big(X_pX'_p + \sigma(t)^2YY'\big)\operatorname{Tr}\big(X_p - \sigma(t)Y\big)\big(X_p - \sigma(t)Y'\big)\\
&\qquad + \frac{2}{p^2n^2}\operatorname{Tr}\big(YY' + \sigma(t)^2YY'\big)\operatorname{Tr}\big(\sigma(t)Y - Y\big)\big(\sigma(t)Y - Y\big)'.
\end{aligned}$$

To tackle the first term on the right side above,

$$\begin{aligned}
&\operatorname{Tr}\big(X_pX'_p + \sigma(t)^2YY'\big)\\
&\quad= \sum_{i=1}^{p}\sum_{k=1}^{n} x^2_{L_p(i,k)} + \sigma(t)^2\sum_{i=1}^{p}\sum_{k=1}^{n} x^{*2}_{L_p(i,k)}\\
&\quad\le \alpha_pk_p\left(\frac{\sum_{i=1}^{k_p} x_i^2}{k_p}\right) + \sum_{i=1}^{p}\sum_{k=1}^{n}(x_{L_p(i,k)} - \bar{x}_{L_p(i,k)})^2\\
&\quad\le \alpha_pk_p\left(\frac{\sum_{i=1}^{k_p} x_i^2}{k_p}\right) + \alpha_nk_p\left(\frac{\sum_{i=1}^{k_p} x_i^2 I\{|x_i| > t\}}{k_p}\right) + |\mu(t)|\alpha_pk_p\left(\frac{\sum_{i=1}^{k_p}|x_i|}{k_p}\right) + \alpha_pk_p\left(\frac{\sum_{i=1}^{k_p}\bar{x}_i^2}{k_p}\right).
\end{aligned}$$

Therefore using $\alpha_pk_p = \mathrm{O}(np)$ and the SLLN, we can see that $(np)^{-1}(\operatorname{Tr}(X_pX'_p + \sigma(t)^2YY'))$ is bounded. Now,

$$\frac{1}{np}\big|\operatorname{Tr}\big(X_p - \sigma(t)Y\big)\big(X_p - \sigma(t)Y\big)'\big| = \frac{1}{np}\big|\operatorname{Tr}(\bar{X}_p\bar{X}'_p)\big| \le \frac{1}{np}\alpha_pk_p\left(\frac{1}{k_p}\sum_{i=1}^{k_p}\bar{x}_i^2\right),$$

which is bounded by $C\mathbb{E}(\bar{x}_i^2)$ almost surely, for some constant C. Here we use the condition $\alpha_pk_p = \mathrm{O}(np)$ and SLLN for $\{\bar{x}_i^2\}$. Since $\mathbb{E}(\bar{x}_i^2) \to 0$ as $t \to \infty$ we can make the right side tend to 0 almost surely, by first letting p tend to infinity, and then letting t tend to infinity. This takes care of the first term.

To tackle the second term,

$$\begin{aligned}
T_2 &= \frac{2}{n^2p^2}\big[\operatorname{Tr}\big(\big(\sigma(t)^2\big)YY' + YY'\big)\operatorname{Tr}\big(\big(\sigma(t)Y - Y\big)\big(\sigma(t)Y - Y\big)'\big)\big]\\
&= \frac{2}{n^2p^2}\big(\sigma(t)^2 + 1\big)\big(\sigma(t) - 1\big)^2\big(\operatorname{Tr}\big(YY'\big)\big)^2\\
&\le \frac{2}{n^2p^2}\big(\sigma(t)^2 + 1\big)\big(\sigma(t) - 1\big)^2\left(\sum_{i=1}^{p}\sum_{k=1}^{n} x^{*2}_{L_p(i,k)}\right)^2\\
&\le \frac{2}{n^2p^2}\big(\sigma(t)^2 + 1\big)\frac{(\sigma(t) - 1)^2}{\sigma(t)^2}\left[\alpha_pk_p\left(\frac{\sum_{i=1}^{k_p} x_i^2}{k_p}\right) + \alpha_pk_p\left(\frac{\sum_{i=1}^{k_p} x_i^2 I\{|x_i| > t\}}{k_p}\right)\right.\\
&\qquad \left. + |\mu(t)|\alpha_pk_p\left(\frac{\sum_{i=1}^{k_p}|x_i|}{k_p}\right) + \alpha_pk_p\left(\frac{\sum_{i=1}^{k_p}\bar{x}_i^2}{k_p}\right)\right]^2.
\end{aligned}$$

Again using the condition $\alpha_pk_p = \mathrm{O}(np)$ and $\sigma(t) \to 1$ as $t \to \infty$ we get $T_2 \to 0$ almost surely. This completes the proof of the theorem. $\square$

To investigate the existence of the LSD of $A_p = A_p(X) = (1/n)X_pX'_p$, in view of Theorem 5, we shall assume that the input sequence is i.i.d. bounded, with mean zero and variance 1. Recall the two formulae given earlier:

$$p^{-1}\operatorname{Tr} A_p^h = p^{-1}n^{-h} \sum_{\pi:\,\pi \text{ circuit}} \mathbb{X}_\pi = p^{-1}n^{-h} \sum_{w} \sum_{\pi \in \Pi(w)} \mathbb{X}_\pi,$$

where the first sum is taken over all words with length $(2h)$ and

$$\mathbb{E}\bigl[p^{-1}\operatorname{Tr} A_p^h\bigr] = p^{-1}n^{-h} \sum_{w \text{ matched}} \sum_{\pi \in \Pi(w)} \mathbb{E}\mathbb{X}_\pi.$$

Theorem 6. *Let $A_p = (1/n)X_pX'_p$ where the entries of X_p are bounded, independent with mean zero and variance 1 and L_p satisfies Properties A and B. Let $p/n \to y$, $0 < y < \infty$. Then*

(i) *If w is a matched word of length $(2h)$ with an edge of order ≥ 3, then $p^{-1}n^{-h}\sum_{\pi \in \Pi(w)} \mathbb{E}\mathbb{X}_\pi \to 0$.*
(ii) *For each $h \geq 1$,*

$$\left|\mathbb{E}\bigl[p^{-1}\operatorname{Tr} A_p^h\bigr] - \sum_{\substack{w \text{ matched} \\ |w|=h}} p^{-1}n^{-h}\bigl|\Pi^*(w)\bigr|\right| \to 0.$$

(iii) *For each $h \geq 1$, $p^{-1}\operatorname{Tr} A_p^h - \mathbb{E}[p^{-1}\operatorname{Tr} A_p^h] \xrightarrow{a.s.} 0$.*
(iv) *If we denote $\beta_h = \limsup_p \sum_{\substack{w \text{ matched} \\ |w|=h}} p^{-1}n^{-h}|\Pi^*(w)|$, then $\{\beta_h\}_{h\geq 1}$ satisfies Carleman's condition.*

The sequence of ESD, $\{F^{A_p(X)}\}$ is almost surely tight. Any subsequential limit is sub Gaussian. The LSD exists iff $\lim_p \sum_{\substack{w \text{ matched} \\ |w|=h}} p^{-1}n^{-h}|\Pi^*(w)|$ exists for each h. The same LSD continues to hold if the input sequence is i.i.d. (not necessarily bounded) with mean zero and variance one.

Remark 1. *Without assuming anything further on the link function L_p, the above limits need not exist. However in Section 1.5, we have seen several examples where the limits do indeed exist.*

Proof of Theorem 6. For the special case of covariance matrix, a proof can be found in [6]. The same arguments may be adapted to general link functions. For the sake of completeness, here are the essential steps.

Recall that circuits which have at least one edge of order 1 contribute zero. Thus, consider all circuits which have at least one edge of order at least 3 and all other edges of order at least 2. Let $N_{h,3+}$ be the number of such circuits of length $(2h)$.

Suppose first that $p = n$. Then $y = 1$. In this case π has uniform range, $1 \leq \pi(i) \leq n$, $1 \leq i \leq 2h$. Then, from the arguments of Bryc, Dembo and Jiang [7], $n^{-(1+h)}N_{h,3+} \to 0$.

Now for general $y > 0$, the range of $\pi(i)$ is not same for every i. For odd vertices, it is from 1 to n and for even vertices, it is 1 to p. However, this case is easily reduced to the previous case $(p = n)$ as follows: let $\widetilde{\Pi}(w)$ be the possibly larger class of same circuits but with range $1 \leq \pi(i) \leq \max(p, n)$, $1 \leq i \leq 2h$. Then, there is a constant C, such that

$$p^{-1}n^{-h} \sum_{\substack{w \text{ has one edge} \\ \text{of order at least 3}}} \bigl|\Pi(w)\bigr| \leq C\bigl[\max(p,n)\bigr]^{-(h+1)} \sum_{\substack{w \text{ has one edge} \\ \text{of order at least 3}}} \bigl|\widetilde{\Pi}(w)\bigr| \to 0$$

from the previous case. This proves (i). Statement (ii) is then a consequence.

For (iii), it is enough to show that

$$\mathbb{E}\bigl[p^{-1}\operatorname{Tr} A_p^h - \mathbb{E}p^{-1}\operatorname{Tr} A_p^h\bigr]^4 = \mathrm{O}\bigl(p^{-2}\bigr).$$

The proof is essentially same as the proof of Lemma 2(b) in [6]. For the sake of completeness, we give the full proof here. We write the fourth moment as

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

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 $\mathbb{E}\mathbb{X}_{\pi_j} = 0$. Hence, using independence, $\mathbb{E}[\prod_{i=1}^4 (\mathbb{X}_{\pi_i} - \mathbb{E}\mathbb{X}_{\pi_i})] = \mathbb{E}[\mathbb{X}_{\pi_j} \prod_{i=1,i\neq j}^4 (\mathbb{X}_{\pi_i} - \mathbb{E}\mathbb{X}_{\pi_i})] = 0$.

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 again by independence,

$$\mathbb{E}\Bigg[\prod_{i=1}^4 (\mathbb{X}_{\pi_i} - \mathbb{E}\mathbb{X}_{\pi_i})\Bigg] = \mathbb{E}\big[(\mathbb{X}_{\pi_j} - \mathbb{E}\mathbb{X}_{\pi_j})\big]\mathbb{E}\Bigg[\prod_{i=1,i\neq j}^4 (\mathbb{X}_{\pi_i} - \mathbb{E}\mathbb{X}_{\pi_i})\Bigg] = 0.$$

So it is clear that for non-zero contribution, the quadruple of circuits must be jointly matched and cross matched. Observe that the total number of edges in each circuit of the quadruple is $(2h)$. So total number of edges over 4 circuits is $(8h)$. Since they are at least pair matched, there can be at most $(4h)$ distinct edges i.e. distinct L-values. In [7] the authors estimated the number of quadruples of circuits which are jointly matched and cross matched. They used a method of counting which they applied only to Toeplitz and Hankel link function. But that method works as well for any general link function satisfying Assumption A. Using this method, we can say that apart from $\pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$ and $\pi_4(0)$ which are always generating vertices, there can be at most $(4h-2)$ many generating vertices to determine $(4h)$ distinct L-values. So total number of generating vertices obtained using Bryc, Dembo and Jiang [7] method is at most $(4h+2)$. Since $\frac{p}{n} \to y, 0 < y < \infty$, it is easy to see that there is a constant K which depends on y and h, such that

$$\frac{1}{p^4}\mathbb{E}\big[\operatorname{Tr} A_p^h - \mathbb{E}\big(\operatorname{Tr} A_p^h\big)\big]^4 \le K\frac{p^{4h+2}}{p^{4h+4}} = \mathrm{O}\big(p^{-2}\big). \tag{21}$$

This guarantees that if there is convergence, it is almost sure. So (iii) is proved.

By Assumption A, for any matched word w of length $(2h)$ with $|w| = h$, we have $|\Pi^*(w)| \le n^{h+1}\Phi^h$ which combined with the fact (15) yields the validity of Carleman's condition, proving (iv).

The other claims of the theorem are easy consequences of all the discussion so far. □

4. Regime II

In this case, $p \to \infty$ and $p/n \to 0$. From the experience of the existing results for the sample covariance matrix (see for example [2]), the following scaled and centered version is required. Define

$$N_p = N_p(X) = \left(\frac{n}{p}\right)^{1/2}\left(\frac{1}{n}X_p X_p' - I_p\right) = \sqrt{\frac{n}{p}}\big(A_p(X) - I_p\big). \tag{22}$$

Straightforward extension of Theorem 5 is not possible to this situation. In Theorem 7 we show how the matrix N_p defined above may be approximated by the following matrix $B_p = B_p(X)$ with bounded entries under additional moment assumption.

$$(B_p)_{ij} = \frac{1}{\sqrt{np}} \sum_{l=1}^n \hat{x}_{L(i,l)}\hat{x}_{L(j,l)} \quad \text{if } i \neq j \text{ and } (B_p)_{ii} = 0, \tag{23}$$

where $\tilde{x}_\alpha = x_\alpha \mathbb{I}(|x_\alpha| \le \varepsilon_p n^{1/4})$, $\hat{x}_\alpha = \tilde{x}_\alpha - \mathbb{E}\tilde{x}_\alpha$ and ε_p will be chosen later.

Theorem 8 establishes the relationship between the existence of the LSD and the limits of the empirical moments.

In Theorem 9 we prove an interesting result in Regime II. Suppose $L^{(1)}$ and $L^{(2)}$ are two link functions satisfying Assumptions A(ii), A′ and B and they agree on the set $\{(i,j)\colon 1 \le i \le p, 1 \le j \le n \text{ and } i < j\}$. Then we show that the corresponding matrices $N_p(X)$ have identical LSD.

In Theorem 10 we work with Assumption A and show the closeness of LSD in probability when two link functions agree on the above set.

Theorem 7. *Let $p, n \to \infty$ so that $p/n \to 0$ and Assumption R2 holds. Suppose L_p satisfies Assumptions A and B. Then there exists a nonrandom sequence $\{\varepsilon_p\}$ with the property that $\varepsilon_p \downarrow 0$ but $\varepsilon_p p^{1/4} \uparrow \infty$ as $p \uparrow \infty$ such that $D(F^{B_p}, F^{N_p}) \to 0$ a.s. where D is the metric for convergence in distribution on the space of all probability distribution functions.*

Proof. Let

$$\tilde{X}_p = \bigl((\tilde{x}_{L(i,j)})\bigr)_{1\le i\le p, 1\le j\le n} \quad \text{and} \quad \tilde{N}_p = \frac{1}{\sqrt{np}}\bigl(\tilde{X}_p\tilde{X}'_p - nI_p\bigr),$$

$$\sup_x \bigl|F^{N_p}(x) - F^{\tilde{N}_p}(x)\bigr| \le p^{-1}\operatorname{rank}(X_p - \tilde{X}_p) \le p^{-1}\sum_{i=1}^{p}\sum_{j=1}^{n}\eta_{L(i,j)} \quad \text{where } \eta_\alpha = \mathbb{I}\bigl(|x_\alpha| > \varepsilon_p n^{1/4}\bigr).$$

The first inequality above follows from [1], Lemma 2.6.

Write $q_p = \sup_\alpha \mathbb{P}(|x_\alpha| > \varepsilon_p n^{1/4})$. We claim that there exists a sequence $\varepsilon_p \downarrow 0$ going to zero arbitrarily slowly such that

$$q_p \le \varepsilon_p n^{-(1+1/\lambda)+\delta/8}. \tag{24}$$

To establish the claim, for simplicity, assume $n = n(p)$ is an increasing function of p. Fix any $\varepsilon > 0$. We have

$$n^{(1+1/\lambda)+\delta/8}\sup_\alpha \mathbb{P}\bigl(|x_\alpha| > \varepsilon n^{1/4}\bigr) \le \varepsilon^{-4(1+1/\lambda)-\delta/2}\sup_\alpha \mathbb{E}|x_\alpha|^{4(1+1/\lambda)+\delta/2}\mathbb{I}\bigl(|x_\alpha| > \varepsilon n^{1/4}\bigr) \to 0 \tag{25}$$

since the random variables $\{|x_\alpha|^{4(1+1/\lambda)+\delta/2}\}$ are uniformly integrable.

Given $m \ge 1$, by (25) find an integer p_m such that $n \ge n_m := n(p_m)$ implies

$$n^{(1+1/\lambda)+\delta/8}\sup_\alpha \mathbb{P}\bigl(|x_\alpha| > m^{-1}n^{1/4}\bigr) \le m^{-1}.$$

Define $\varepsilon_p = 1/m$ if $p_m \le p < p_{m+1}$ and $\varepsilon_p = n(p_1)^{(1+1/\lambda)+\delta/8}$ for $p < p_1$. Note that by choosing the integers in the sequence $p_1 < p_2 < \cdots$ as large as we want, we can make ε_p go to zero as slowly as we like. Clearly, ε_p satisfies the inequality (24).

For any $\beta > 0$, and with Y_i independent Bernoulli with $E(Y_i) \le q_p$,

$$\begin{aligned}
\mathbb{P}\Bigl(\sup_x \bigl|F^{N_p}(x) - F^{\tilde{N}_p}(x)\bigr| > \beta\Bigr) &\le \mathbb{P}\left(p^{-1}\sum_{i=1}^{p}\sum_{j=1}^{n}\eta_{L(i,j)} > \beta\right) \\
&\le \mathbb{P}\left(\alpha_p p^{-1}\sum_{i=1}^{k_p} Y_i > \beta\right) \\
&\le (\beta p)^{-1}\alpha_p\sum_{i=1}^{k_p}\mathbb{E}Y_i \quad \text{by Markov inequality} \\
&\le (\beta p)^{-1}\alpha_p k_p q_p \\
&\le Cnq_p = \mathrm{o}\bigl(n^{-(1/\lambda+\delta/8)}\bigr) = \mathrm{o}\bigl(p^{-(1+\lambda\delta/8)}\bigr),
\end{aligned}$$

where the constant C is such that $\alpha_p k_p \le C\beta np$. Hence by Borel-Cantelli lemma,

$$\sup_x \bigl|F^{N_p}(x) - F^{\tilde{N}_p}(x)\bigr| \to 0 \quad \text{a.s.}$$

Let

$$\hat{X}_p = \bigl((\hat{x}_{L(i,j)})\bigr)_{1\le i\le p, 1\le j\le n} \quad \text{and} \quad \hat{N}_p = \frac{1}{\sqrt{np}}\bigl(\hat{X}_p\hat{X}_p' - nI_p\bigr).$$

Using Lemma 1(i) and (ii),

$$\sup_x \bigl|F^{\hat{N}_p}(x) - F^{\tilde{N}_p}(x)\bigr| \le \left(p^{-1}\sum_{i=1}^{p}\bigl|\lambda_i(\hat{N}_p) - \lambda_i(\tilde{N}_p)\bigr|\right)^2 \le \frac{1}{np^3}\operatorname{Tr}\bigl(\tilde{X}_p\tilde{X}_p' + \hat{X}_p\hat{X}_p'\bigr)\operatorname{Tr}\bigl(\mathbb{E}(\tilde{X}_p)\mathbb{E}(\tilde{X}_p')\bigr).$$

Using the moment condition, the condition on the truncation level, the condition $\alpha_p k_p = \mathrm{O}(np)$, and an appropriate strong law of large numbers for independent random variables, it is easy to show that the above expression tends to 0 almost surely. We omit the tedious details.

On the other hand,

$$\begin{aligned}
d_{BL}^2\bigl(F^{\hat{N}_p}, F^{B_p}\bigr) &\le \frac{1}{p}\operatorname{Tr}(\hat{N}_p - B_p)^2 \\
&\le \frac{1}{2np^2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\bigl(\hat{x}_{L(i,l)}^2 - \mathbb{E}\hat{x}_{L(i,l)}^2\bigr)\right)^2 + \frac{1}{2np^2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\bigl(1 - \mathbb{E}\hat{x}_{L(i,l)}^2\bigr)\right)^2 = M + N \quad \text{say.}
\end{aligned}$$

Note that for every i, j, there exists α such that

$$0 \le 1 - \mathbb{E}\hat{x}_{L(i,j)}^2 = \mathbb{E}x_\alpha^2\mathbb{I}\bigl(|x_\alpha| > \varepsilon_p n^{1/4}\bigr) + \bigl(\mathbb{E}x_\alpha\mathbb{I}\bigl(|x_\alpha| > \varepsilon_p n^{1/4}\bigr)\bigr)^2 \le \frac{1}{\varepsilon_p^2 n^{1/2}}\bigl(\mathbb{E}x_\alpha^4 + \mathbb{E}^2x_\alpha^2\bigr). \tag{26}$$

From (26), it is immediate that

$$N \le \frac{n^2p}{2np^2}\frac{1}{n\varepsilon_p^4}\Bigl(\sup_\alpha\bigl(\mathbb{E}x_\alpha^4 + \mathbb{E}^2x_\alpha^2\bigr)\Bigr)^2 \to 0 \quad \text{since } \varepsilon_p p^{1/4} \to \infty.$$

Now we will deal with the first term, M.

$$\begin{aligned}
\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\bigl(\hat{x}_{L(i,l)}^2 - \mathbb{E}\hat{x}_{L(i,l)}^2\bigr)\right)^2 &= \sum_\alpha a_\alpha\bigl(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\bigr)^2 + \sum_{\alpha\ne\alpha'} b_{\alpha,\alpha'}\bigl(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\bigr)\bigl(\hat{x}_{\alpha'}^2 - \mathbb{E}\hat{x}_{\alpha'}^2\bigr) \\
&= T_{1p} + T_{2p} \quad \text{(say)},
\end{aligned}$$

where $a_\alpha, b_{\alpha,\alpha'} \ge 0$. Obviously,

$$\#\{\alpha \in \mathcal{Z}\colon a_\alpha \ge 1\} \le k_p \quad \text{and} \quad \#\bigl\{\bigl(\alpha,\alpha'\bigr) \in \mathcal{Z}^2\colon \alpha \ne \alpha', b_{\alpha,\alpha'} \ge 1\bigr\} \le k_p^2.$$

Also, $a_\alpha \le \alpha_p$ for all α and $b_{\alpha,\alpha'} \le \Phi^2\alpha_p$ for all $\alpha \ne \alpha'$. Hence,

$$\begin{aligned}
\sum_p \frac{1}{4n^2p^4}\mathbb{E}T_{1p}^2 &= \sum_p \frac{1}{4n^2p^4}\sum_\alpha a_\alpha^2\mathbb{E}\bigl(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\bigr)^4 \\
&\quad + \sum_p \frac{1}{4n^2p^4}\sum_{\alpha\ne\alpha'} a_\alpha a_{\alpha'}\mathbb{E}\bigl(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\bigr)^2\mathbb{E}\bigl(\hat{x}_{\alpha'}^2 - \mathbb{E}\hat{x}_{\alpha'}^2\bigr)^2
\end{aligned}$$

$$\leq \sum_p \frac{\alpha_p^2 k_p}{4n^2 p^4} \sup_\alpha \mathbb{E}\big(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\big)^4 + \sum_p \frac{\alpha_p^2 k_p^2}{4n^2 p^4} \sup_\alpha \mathbb{E}^2\big(\hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2\big)^2$$

$$\leq \sup_\alpha \sum_p \frac{\alpha_p^2 k_p}{4n^2 p^4} \big(n^{1/4}\varepsilon_p\big)^4 \mathbb{E}x_\alpha^4 + \sup_\alpha \sum_p \frac{\alpha_p^2 k_p^2}{4n^2 p^4} \mathbb{E}^2 x_\alpha^4 < \infty,$$

where we use the fact that $\alpha_p k_p = \mathrm{O}(np)$ and $\alpha_p \leq \Phi p$. Thus, $T_{1p}/(2np^2) \to 0$ a.s.

It now remains to tackle T_{2p}. Let $y_\alpha = \hat{x}_\alpha^2 - \mathbb{E}\hat{x}_\alpha^2, \alpha \in \mathcal{Z}$. Then $\{y_\alpha\}$ are mean zero independent random variables.

$$\begin{aligned}
\sum_p \frac{1}{4n^2 p^4} \mathbb{E}T_{2p}^2 &= \sum_p \frac{1}{4n^2 p^4} \mathbb{E}\Big(\sum_{\alpha \neq \alpha'} b_{\alpha,\alpha'} y_\alpha y_{\alpha'}\Big)^2 \\
&\leq \sum_p \frac{1}{4n^2 p^4} \sum_{\alpha \neq \alpha'} b_{\alpha,\alpha'}^2 \mathbb{E}y_\alpha^2 y_{\alpha'}^2 \\
&\leq \sup_\alpha \sum_p \frac{k_p^2 \alpha_p^2}{4n^2 p^4} \mathbb{E}^2 x_\alpha^2 < \infty.
\end{aligned}$$

Thus by Borel-Cantelli lemma, $T_{2p}/(np^2) \to 0$ almost surely and this completes the proof. □

Remark 2. (a) *From results of* [3], *it is known that for the special case of sample covariance matrix*, *finite fourth moment is needed for the above approximation to work*, *and inter alia*, *for the existence of the LSD almost surely. Here the link function is more general*, *necessitating slightly higher moments*.

(b) *If we carefully follow the above proof*, *finiteness of the* $4(1 + 1/\lambda) + \delta$*-th moment was only needed in proving that* $\sup_x |F^{N_p}(x) - F^{\tilde{N}_p}(x)| \xrightarrow{a.s.} 0$. *If we impose the weaker assumption* $\sup_\alpha \mathbb{E}x_\alpha^4 < \infty$, *then* $q_p \leq \varepsilon_p/n$ *for suitably chosen sequence* $\{\varepsilon_p\} \downarrow 0$ *and* $\sup_x |F^{N_p}(x) - F^{\tilde{N}_p}(x)| \xrightarrow{P} 0$ *holds and hence* $D(F^{B_p}, F^{N_p}) \to 0$ *in probability.*

Having approximated N_p by B_p, we now need to establish the behaviour of the moments of B_p. This is done through a series of Lemma, finally leading to Theorem 8. In the subsequent discussion, we will use the following notation.

$$\varPi_{\neq}(w) := \big\{\pi \in \varPi(w)\colon\ \pi(2i-2) \neq \pi(2i)\ \forall 1 \leq i \leq h\big\},$$
$$\varPi_{\neq}^*(w) := \big\{\pi \in \varPi^*(w)\colon\ \pi(2i-2) \neq \pi(2i)\ \forall 1 \leq i \leq h\big\}.$$

Analogous to $\varPi(w)$ and $\varPi^*(w)$, we define, for several words $w_1, w_2, \ldots, w_k$,

$$\varPi(w_1, w_2, \ldots, w_k) := \big\{(\pi_1, \pi_2, \ldots, \pi_k)\colon\ w_i[s] = w_j[\ell] \Leftrightarrow \xi_{\pi_i}(s) = \xi_{\pi_j}(\ell), 1 \leq i, j \leq k\big\},$$
$$\varPi^*(w_1, w_2, \ldots, w_k) := \big\{(\pi_1, \pi_2, \ldots, \pi_k)\colon\ w_i[k] = w_j[\ell] \Rightarrow \xi_{\pi_i}(k) = \xi_{\pi_j}(\ell), 1 \leq i, j \leq k\big\}.$$

(Definition of ξ_π has been given earlier.)

Moreover,

$$\begin{aligned}
\varPi_{\neq}(w_1, w_2, \ldots, w_k) := \big\{&(\pi_1, \pi_2, \ldots, \pi_k) \in \varPi(w_1, w_2, \ldots, w_k)\colon \\
&\pi_i(0) \neq \pi_i(2), \ldots, \pi_i(2h-2) \neq \pi_i(2h),\ 1 \leq i \leq k\big\}.
\end{aligned}$$

For Lemmas 2–5, we will always assume that the link function L satisfies Assumption A(ii) and A′.

Lemma 2. *Fix* $h \geq 1$ *and a matched word* w *of length* $2h$. *Then we have*

$$\big|\varPi_{\neq}(w)\big| \leq K_h p^{1+[(|w|+1)/2]} n^{[|w|/2]}, \tag{27}$$

where K_h is some constant depending on h.

Proof. Let k be the number of odd generating vertices, denoted by $\pi(2i_1-1), \pi(2i_2-1), \dots, \pi(2i_k-1)$ where $1 \le i_1 < i_2 < \cdots < i_k < h$. We wish to emphasize that $i_k < h$ since $\{\xi_\pi(2h-1), \xi_\pi(2h)\}$ cannot be a matched edge as $\xi_\pi(2h-1) \neq \xi_\pi(2h)$ by A′.

Recall the definition of Type I and Type II generating vertex given in Section 2. Let t be the number of Type I odd generating vertices. Since total number of generating vertices apart from $\pi(0)$ is $|w|$, we have $t \le [|w|/2]$.

Now, fix a Type II generating vertex $\pi(2i-1)$. Also, suppose we have already made our choices for the vertices $\pi(j)$, $j < 2i-1$ which come before $\pi(2i-1)$. Since $\pi(2i)$ is not a generating vertex,

$$\xi_\pi(2i) = \xi_\pi(j) \quad \text{for some } j < 2i-1. \tag{28}$$

(Note that since $\pi(2i-2) \neq \pi(2i)$, we cannot have $\xi_\pi(2i) = \xi_\pi(2i-1)$ by Assumption A′). Now that the value of $\xi_\pi(j)$ has been fixed, for each value of $\pi(2i)$, there can be at most Φ many choices for $\pi(2i-1)$ such that (28) is satisfied. Thus, we can only have at most $p\Phi$ many choices for the generating vertex $\pi(2i-1)$. In short, there are only $\mathrm{O}(p)$ possibilities for a Type II odd generating vertex to choose from.

With the above crucial observation before us, it is now easy to conclude

$$\left|\Pi_{\neq}(w)\right| = \mathrm{O}\bigl(p^{\#\text{even generating vertices}+\#\text{Type II vertices}} n^{\#\text{Type I vertices}}\bigr) = \mathrm{O}\bigl(p^{1+[(|w|+1)/2]} n^{[|w|/2]}\bigr). \qquad \square$$

Lemma 3. (i) *For every $h \ge 1$ even,*

$$\left| p^{-1}\mathbb{E}\operatorname{Tr} B_p^h - \sum_{w \text{ matched}, |w|=h} p^{-1-h/2} n^{-h/2} \left|\Pi^*_{\neq}(w)\right| \right| \to 0.$$

(ii) *For every $h \ge 1$ odd*, $\lim_{p\to\infty} p^{-1}\mathbb{E}\operatorname{Tr} B_p^h = 0$.

Proof. Let $\hat{\mathbb{X}}_\pi$ be as defined in (14) with x_α replaced by $\hat{x}_\alpha$. From the fact that $\mathbb{E}\hat{x}_\alpha = 0$ and $B_{ii} = 0$ for all i, we have

$$p^{-1}\mathbb{E}\operatorname{Tr} B_p^h = p^{-(1+h/2)} n^{-h/2} \sum_{w \text{ matched}} \sum_{\substack{\pi \in \Pi(w),\\ \pi(2i-2)\neq\pi(2i), \forall 1\le i\le h}} \mathbb{E}\hat{\mathbb{X}}_\pi.$$

Fix a matched word w of length $2h$. It induces a partition on $2h$ L-values $\xi_\pi(1), \xi_\pi(2), \xi_\pi(3), \dots, \xi_\pi(2h)$ resulting in $|w|$ many groups (partition blocks) where the values of ξ_π within a group are same but across the groups they are different. Let C_k be the number of groups of size k. Clearly,

$$C_2 + C_3 + \cdots + C_{2h} = |w| \quad \text{and} \quad 2C_2 + 3C_3 + \cdots + 2hC_{2h} = 2h.$$

Note that

$$\sup_\alpha \left|\mathbb{E}\hat{x}_\alpha^2 - 1\right| = o(1) \quad \text{and} \quad \sup_\alpha \mathbb{E}|\hat{x}_\alpha|^k \le \bigl(\varepsilon_p n^{1/4}\bigr)^{k-2} \quad \forall k \ge 2.$$

Thus if $\pi \in \Pi(w)$,

$$\mathbb{E}|\hat{x}_{\xi_\pi(1)}\hat{x}_{\xi_\pi(2)}\cdots\hat{x}_{\xi_\pi(2h)}| \le \bigl(\varepsilon_p n^{1/4}\bigr)^{0\cdot C_2 + 1\cdot C_3 + \cdots + (2h-2)\cdot C_{2h}} \le (\varepsilon_p n^{1/4})^{2h-2|w|}.$$

Using Lemma 2,

$$\begin{aligned}\sum_{\substack{\pi \in \Pi(w),\\ \pi(2i-2)\neq\pi(2i)\forall 1\le i\le h}} \mathbb{E}|\hat{x}_{\xi_\pi(1)}\hat{x}_{\xi_\pi(2)}\cdots\hat{x}_{\xi_\pi(2h)}| &\le \bigl(\varepsilon_p n^{1/4}\bigr)^{2h-2|w|}\bigl(K_h p^{1+[(|w|+1)/2]} n^{[|w|/2]}\bigr)\\ &= (\varepsilon_p)^{2h-2|w|} n^{h/2-(|w|/2-[|w|/2])} p^{1+[(|w|+1)/2]}.\end{aligned}$$

Case I. Either $|w| < h$ or $|w| = h$ with h odd. Then

$$\frac{1}{p^{1+h/2}n^{h/2}} \sum_{\substack{\pi \in \Pi(w), \\ \pi(2i-2) \neq \pi(2i) \forall 1 \leq i \leq h}} \mathbb{E}|\hat{x}_{\xi_\pi(1)}\hat{x}_{\xi_\pi(2)} \cdots \hat{x}_{\xi_\pi(2h)}| \to 0. \tag{29}$$

Case II. If $|w| = h$ with h even, then

$$\begin{aligned}
&\lim_p \frac{1}{p^{1+h/2}n^{h/2}} \sum_{\substack{\pi \in \Pi(w), \\ \pi(2i-2) \neq \pi(2i) \forall 1 \leq i \leq h}} \mathbb{E}\hat{x}_{\xi_\pi(1)}\hat{x}_{\xi_\pi(2)} \cdots \hat{x}_{\xi_\pi(2h)} \\
&\quad = \lim_p \frac{1}{p^{1+h/2}n^{h/2}} \big|\Pi_{\neq}(w)\big|\big(1 + o(1)\big)^h \\
&\quad = \lim_p \frac{1}{p^{1+h/2}n^{h/2}} \big|\Pi_{\neq}(w)\big| \\
&\quad = \lim_p \frac{1}{p^{1+h/2}n^{h/2}} \big|\Pi^*_{\neq}(w)\big|.
\end{aligned}$$

This completes the proof. □

Fix jointly matched and cross-matched words (w_1, w_2, w_3, w_4) of length $(2h)$ each. Let

$$\kappa = \text{total number of distinct letters in } w_1, w_2, w_3 \text{ and } w_4. \tag{30}$$

Lemma 4. *For some constants C_h depending on h,*

$$\begin{aligned}
&\#\big\{(\pi_1, \pi_2, \pi_3, \pi_4) \in \Pi(w_1, w_2, w_3, w_4)\colon \pi_i(0) \neq \pi_i(2), \ldots, \pi_i(2h-2) \neq \pi_i(2h), i = 1, 2, 3, 4\big\} \\
&\quad \leq \begin{cases} C_h p^{2+2h} n^{2h} & \text{if } \kappa = 4h \text{ or } 4h-1, \\ C_h p^{4+[(1+\kappa)/2]} n^{[\kappa/2]} & \text{if } \kappa \leq 4h-2. \end{cases}
\end{aligned} \tag{31}$$

Proof. *Case* I. $\kappa = 4h - 1$ or $4h$. Since the circuits are cross-matched, upon reordering the circuits, making circular shift and counting anti-clockwise if necessary, we may assume, without loss of generality, $\xi_{\pi_1}(2h)$ does not match with any L-value in π_1 when $\kappa = 4h - 1$. Because of cross matching, when $\kappa = 4h$, we may further assume that $\xi_{\pi_2}(2h)$ does not match with any L-value in π_1 or π_2.

We first fix the values of the p-vertices $\pi_i(0)$, $1 \leq i \leq 4$, all of which are even generating vertices. Then we scan all the vertices from left to right, one circuit after another. We will then, as argued in [7], obtain a total $(4h + 2)$ generating vertices instead of $(4h + 4)$ generating vertices which we would have obtained by usual counting.

In our dynamic counting, we scan the L-values in the following order:

$$\xi_{\pi_1}(1), \xi_{\pi_1}(2), \xi_{\pi_1}(3), \ldots, \xi_{\pi_1}(2h), \xi_{\pi_2}(1), \xi_{\pi_2}(2), \ldots, \xi_{\pi_2}(2h), \xi_{\pi_3}(1), \ldots, \xi_{\pi_4}(2h).$$

From the arguments given in Lemma 2, it is clear that for an odd vertex $\pi_i(2j-1)$ to be Type I, both $\xi_{\pi_i}(2j-1)$ and $\xi_{\pi_i}(2j)$ have to be the first appearances of two distinct L values. So, total number of Type I n-generating vertices is at most $\kappa/2$.

Case II. $\kappa \leq 4h - 2$. We again apply the crucial fact that for an odd vertex $\pi_i(2j-1)$ to be Type I, we need both $\xi_{\pi_j}(2j-1)$ and $\xi_{\pi_j}(2j)$ to be the first appearances of two distinct L-values, while the circuits being scanned from left to right and one after another. Since there are exactly κ distinct L-values, the number of Type I odd vertices is not more than $\kappa/2$. Combining this with the fact that total number of generating vertices equals $(\kappa + 4)$, we get the required bound. □

Lemma 5. *For each fixed $h \geq 1$,*

$$\sum_{p=1}^{\infty} \mathbb{E}\big[p^{-1}\big(\operatorname{Tr} B_p^h - \mathbb{E}(\operatorname{Tr} B_p^h)\big)\big]^4 < \infty.$$

Proof. As argued earlier and from the fact that the diagonal elements of B_p are all zero, we have

$$\mathbb{E}\big[p^{-1}\big(\operatorname{Tr} B_p^h - \mathbb{E}\big(\operatorname{Tr} B_p^h\big)\big)\big]^4 = p^{-4-2h} n^{-2h} \sum_{*} \sum^{*} \mathbb{E}\left(\prod_{i=1}^{4} (\hat{\mathbb{X}}_{\pi_i} - \mathbb{E}\hat{\mathbb{X}}_{\pi_i})\right),$$

where the outer sum $\sum_*$ is over all quadruples of words (w_1, w_2, w_3, w_4), each of length $2h$ and which are jointly matched and cross-matched. The inner sum $\sum^*$ is over all quadruples of circuits $\{(\pi_1, \pi_2, \pi_3, \pi_4) \in \Pi(w_1, w_2, w_3, w_4)$: $\pi_i(0) \neq \pi_i(2), \ldots, \pi_i(2h-2) \neq \pi_i(2h), i = 1, 2, 3, 4\}$.

Note that by definition of κ, $\kappa \leq 4h$ for any jointly matched quadruple of words (w_1, w_2, w_3, w_4) of total length $8h$. Fix w_1, w_2, w_3, w_4, jointly matched and cross-matched.

Case I. $\kappa = 4h$ or $4h - 1$. Then the maximum power with which any $\hat{x}_\alpha$ can occur in $\prod_{i=1}^4 \hat{\mathbb{X}}_{\pi_i}$ is bounded by 4. But since $\sup_\alpha \mathbb{E}(\hat{x}_\alpha)^4 < \infty$, we immediately have $|\mathbb{E}(\prod_{i=1}^4 (\hat{\mathbb{X}}_{\pi_i} - \mathbb{E}\hat{\mathbb{X}}_{\pi_i}))| < \infty$. Thus, by Lemma 4,

$$p^{-4-2h} n^{-2h} \sum_{*:\kappa \in \{4h-1, 4h\}} \sum^{*} \mathbb{E}\left(\prod_{i=1}^{4} (\hat{\mathbb{X}}_{\pi_i} - \mathbb{E}\hat{\mathbb{X}}_{\pi_i})\right) = p^{-4-2h} n^{-2h} \mathrm{O}\big(p^{2+2h} n^{2h}\big) = \mathrm{O}\big(p^{-2}\big).$$

Case II. Suppose now that $\kappa = 4h - k$, $k \geq 2$. Borrowing notation from Lemma 3, we have

$$C_2 + C_3 + \cdots + C_{8h} = \kappa \quad \text{and} \quad 2C_2 + 3C_3 + \cdots + 8hC_{8h} = 8h.$$

These two equations immediately give

$$C_3 + 2C_4 + \cdots + (8h-2)C_{8h} = 8h - 2\kappa = 2k.$$

It is also easy to see that

$$C_{2k+2+i} = 0 \quad \forall i \geq 1 \quad \text{and} \quad C_5 + 2C_6 + \cdots + (2k-2)C_{2k+2} \leq 2k - 2.$$

Since the input variables are truncated at $\varepsilon_p n^{1/4}$ and since $\sup_\alpha E(x_\alpha^4) < \infty$,

$$\sup_\alpha \mathbb{E}\left|\prod_{i=1}^{4} \hat{\mathbb{X}}_{\pi_i}\right| \leq \mathbb{E}(\hat{x}_\alpha)^4 n^{\frac{C_5 + 2C_6 + \cdots + (2k-2)C_{2k+2}}{4}} = \mathrm{O}\big(n^{(2k-2)/4}\big).$$

Thus, by Lemma 4, for any $k \geq 2$,

$$\begin{aligned} p^{-4-2h} n^{-2h} \sum_{*:\kappa \in \{4h-k\}} \sum^{*} \mathbb{E}\left(\prod_{i=1}^{4} (\hat{\mathbb{X}}_{\pi_i} - \mathbb{E}\hat{\mathbb{X}}_{\pi_i})\right) &\leq p^{-4-2h} n^{-2h} C_h p^{4+[(4h-k+1)/2]} n^{[(4h-k)/2]} \mathrm{O}\big(n^{(k-1)/2}\big) \\ &= \mathrm{O}\big(p^{-4-2h} p^{-2h} p^{4+[(4h-k+1)/2]} p^{[(4h-k)/2]}\big) \mathrm{O}\big(p^{(k-1)/2}\big) \\ &= \mathrm{O}\big(p^{-4-4h} p^{(4+4h-k)+(k-1)/2}\big) = \mathrm{O}\big(p^{-(k/2+1)}\big). \end{aligned}$$

For a quick explanation of the first equality above, just note that total power of n in the previous expression is negative. □

We are now ready to summarize the results of Theorem 7 and Lemma 2–5 in the following theorem in Regime II.

Theorem 8. *Let $p, n \to \infty$ so that $p/n \to 0$ and Assumption R2 holds. Suppose L_p satisfies Assumptions A(ii), A′ and B. Suppose $\{\varepsilon_p\}$ satisfying $\{\varepsilon_p\} \downarrow 0$ and $\varepsilon_p p^{1/4} \to \infty$ is appropriately chosen. Then*

(i) *For every $h \geq 1$ even,*

$$p^{-1}\mathbb{E}\operatorname{Tr} B_p^h - \sum_{w \text{ matched}, |w|=h} p^{-(1+h/2)} n^{-h/2} \left|\Pi^*_{\neq}(w)\right| \to 0.$$

(ii) *For every $h \geq 1$ odd, $\lim_{p\to\infty} p^{-1}\mathbb{E}\operatorname{Tr} B_p^h = 0$.*

(iii) *For each $h \geq 1$, $p^{-1}\operatorname{Tr} B_p^h - \mathbb{E} p^{-1}\operatorname{Tr} B_p^h \xrightarrow{a.s.} 0$.*

(iv) *$\beta_{2h} \equiv \limsup_p \sum_{\substack{w \text{ matched} \\ |w|=2h}} p^{-(1+h)} n^{-h} |\Pi^*_{\neq}(w)|$, satisfies Carleman's condition.*

As a consequence, the sequence $\{F^{N_p}\}$ is almost surely tight. Every subsequential limit is symmetric and sub Gaussian. The LSD of $\{F^{N_p}\}$ exists almost surely, iff $\lim \sum_{\substack{w \text{ matched} \\ |w|=2h}} p^{-(1+h)} n^{-h} |\Pi^*_{\neq}(w)|$ exists. These give the $(2h)$th moment of the LSD.

Below we will deal with two matrices with different link functions $L^{(1)}$ or $L^{(2)}$. The corresponding relevant quantities will now be denoted with added superscripts (1) and (2) respectively.

Theorem 9. *Let $L^{(1)}$ and $L^{(2)}$ be two link functions satisfying Assumptions A(ii), A′ and B and agreeing on the set $\{(i, j)\colon 1 \leq i \leq p, 1 \leq j \leq n, i < j\}$. Then, for each matched word w of length $(4h)$ with $|w| = 2h$,*

$$\frac{1}{p^{h+1} n^h} \left| \Pi^{*(1)}_{\neq}(w) - \Pi^{*(2)}_{\neq}(w) \right| \to 0.$$

Hence, in Regime II under Assumption R2, $F^{N_p^{(i)}}$, $i = 1, 2$ have identical asymptotic behaviour.

Proof of Theorem 9. Define for each link function $L^{(i)}$, $i = 1, 2$

$$\Gamma_j^{(i)} := \left\{\pi \in \Pi^{*(i)}_{\neq}(w)\colon 1 \leq \pi(2j+1) \leq p\right\}, \quad j = 1, 2, \ldots, 2h.$$

Now, it is enough to prove that for a fixed j,

$$\frac{1}{p^{h+1} n^h} \left|\Gamma_j^{(i)}\right| \to 0.$$

Consider the transformation $\pi \mapsto \hat{\pi}$, where $\hat{\pi}$ is also a circuit with

$$\hat{\pi}(0) = \pi(2j), \qquad \hat{\pi}(1) = \pi(2j+1), \ldots, \pi(4h-1) = \pi(2j-1), \qquad \hat{\pi}(4h) = \pi(2j).$$

Then it is easy to show that the map $\pi \mapsto \hat{\pi}$ is a bijection between $\Gamma_j^{(i)}$ and $\Gamma_1^{(i)}$ and therefore, $|\Gamma_j^{(i)}| = |\Gamma_1^{(i)}|$.

Observe that $\pi(1)$ is always a Type I odd generating vertex. By Lemma 2, we have $|\Pi^{*(i)}(w)| = \mathrm{O}(p^{h+1} n^h)$. Since, in the definition of $\Gamma_1^{(i)}$ we are restricting one of the Type I odd generating vertex to be a p generating vertex, we are going to lose a factor of n in the bound and pick up a factor of p instead. Thus, $|\Gamma_1^{(i)}| = \mathrm{O}(p^{h+2} n^{h-1})$ and hence

$$\frac{1}{p^{h+1} n^h} \left|\Gamma_j^{(i)}\right| = \mathrm{O}\left(\frac{p}{n}\right) \to 0. \qquad \square$$

The symmetric Toeplitz link function $L(i, j) = |i - j|$ does not satisfy Assumption A′ but the asymmetric Toeplitz link function $L(i, j) = i - j$ does. Hence the above result is not applicable for this pair of link functions. However, under Assumption A, we can directly claim the closeness of LSD, but only in probability.

Theorem 10. *Assume $\{x_\alpha\}$ are independent with mean zero and $\sup_\alpha \mathbb{E}x_\alpha^4 < \infty$. Suppose $L^{(1)}$ and $L^{(2)}$ are two link functions such that $L^{(1)}(i,j) = L^{(2)}(i,j)$ on the set $\{(i,j)\colon 1 \le i \le p, p \le j \le n\}$ and both satisfy Assumption A. Set $X = ((x_{L^{(1)}(i,j)}))_{p\times n}$ and $Y = ((x_{L^{(2)}(i,j)}))_{p\times n}$. Then*

$$d_{BL}\big(F^{\sqrt{n/p}(A_p(X)-I_p)}, F^{\sqrt{n/p}(A_p(Y)-I_p)}\big) \xrightarrow{P} 0.$$

Proof. Let $X = [X_0\colon Z]$ and $Y = [Y_0\colon Z]$ where X_0 and Y_0 are $p \times p$ sub-matrices of X and Y respectively. Note that

$$\begin{aligned}\mathbb{E}d_{BL}^2\big(F^{\sqrt{n/p}(A_p(X)-I_p)}, F^{\sqrt{n/p}(A_p(Y)-I_p)}\big) &\le n^{-1}p^{-2}\mathbb{E}\operatorname{Tr}\big(XX' - YY'\big)^2\\ &= n^{-1}p^{-2}\mathbb{E}\operatorname{Tr}\big(X_0X_0' - Y_0Y_0'\big)^2\\ &\le 2n^{-1}\big(\mathbb{E}\operatorname{Tr}\big(p^{-1}X_0X_0'\big)^2 + \mathbb{E}\operatorname{Tr}\big(p^{-1}Y_0Y_0'\big)^2\big).\end{aligned}$$

Calculations similar to those done in Regime I now imply that

$$\mathbb{E}p^{-1}\operatorname{Tr}\big(p^{-1}X_0X_0'\big)^2 \le K \quad \text{and} \quad \mathbb{E}p^{-1}\operatorname{Tr}\big(p^{-1}Y_0Y_0'\big)^2 \le K,$$

for some constant K. Since $p/n \to 0$, the result follows immediately. □

5. Proofs of Theorems 1–4

For convenience of counting, in Regime I, the pair matched words are classified as follows. In a pair matched circuit of length $(2h)$, there are h distinct L-values and hence $(h+1)$ generating vertices. But the number of odd (or even) generating vertices depends on the corresponding word. Note that there is always at least one even generating vertex $\pi(0)$, and the number of even generating vertices is bounded by h. Let

$$\mathcal{W}_{t,h} = \big\{w\colon w \text{ is pair-matched of length } 2h \text{ with } (t+1)\ \textit{even} \text{ generating vertices}\big\}. \tag{32}$$

Obviously the total number of odd generating vertices for a word in $\mathcal{W}_{t,h}$ is $(h-t)$.

Symmetric words. In most of the examples, due to the circuit restriction, only words where each letter appears in an odd and an even position have positive contribution in the limit. We call these *symmetric* words. Let

$$\mathcal{W}_{t,h}^0 = \{w\colon w \in \mathcal{W}_{t,h} \text{ and } w \text{ is symmetric}\}.$$

5.1. *Proof of Theorem 1*

(i) In view of Theorem 6 we will only have to show that, for each $h \ge 1$ and each word $w \in \mathcal{W}_{t,h}$, $\lim_p p^{-1}n^{-h}|\Pi^*(w)|$ exists.

We first show that the circuit condition implies that words which are *not symmetric* do not contribute in the limit.

Lemma 6. *Suppose $p/n \to y \in (0,\infty)$. Let w be any pair-matched word of length $(2h)$ which is not symmetric. Then in Regime I*,

$$p^{-1}n^{-h}\big|\Pi^*(w)\big| \to 0 \quad as\, p \to \infty. \tag{33}$$

Proof. Fix a pair-matched non-symmetric word w of length $(2h)$ and hence $|w| = h$. Let S be the set of $(h+1)$ indices corresponding to generating vertices of w. Now because of the circuit condition $\pi(0) - \pi(2h) = 0$, we must have

$$\xi_\pi(1) - \xi_\pi(2) + \xi_\pi(3) - \cdots + \xi_\pi(2h-1) - \xi_\pi(2h) = 0. \tag{34}$$

Let us enumerate S, left to right, as $\{0, i_1, i_2, \ldots, i_h\}$ and for each $i_t \in S \setminus \{0\}$, let j_t be its matching index, so that, $\xi_\pi(i_t) = \xi_\pi(j_t), i_t < j_t$.

Since w is not symmetric, there exists at least one pair of matching indices of the same parity. Because of Eq. (34) number of pairs of matching L-values with odd indices is same as the number of pairs of matching L-values with even indices. Consider the set $\mathcal{P}$ of all indices of $S \setminus \{0\}$ whose matching counterpart is of the same parity. Let $i_{\max} = \max \mathcal{P}$. Let $j_{\max}$ be the matching index for $i_{\max}$.

So for any $i \in S \setminus \{0\}$ with $i > i_{\max}$, $\xi_\pi(i)$ has a matching L-value with index of opposite parity and hence if they are substituted in Eq. (34), they have same value but opposite sign. Therefore, they cancel out each other.

Now, according to our convention, we start choosing generating vertices from the left end of the circuit. We stop when we reach $\pi(i_{\max})$. By this process we have fully determined the values of $\{\xi_\pi(t)\}$ for all $t < i_{\max}$. On the other hand, if we consider $\xi_\pi(t)$ with $t > i_{\max}$, then we immediately realize that:

1. Either its value is already determined. This is the case when its matching counterpart appears at the left of $\xi_\pi(i_{\max})$.
2. Or, its matching counterpart has index of opposite parity as we have observed before.

Thus in Eq. (34), except $\xi_\pi(i_{\max})$ and $\xi_\pi(j_{\max})$, all other $\{\xi_\pi(t)\}$ are either already determined or get cancelled with their own counterpart. So, (34) forces $\xi_\pi(i_{\max}) + \xi_\pi(j_{\max}) = 2\xi_\pi(i_{\max})$ to take some particular value. Therefore, $\pi(i_{\max})$ has no free choice though it is a generating vertex. This is a contradiction and the proof of the lemma is complete. □

Therefore, going back to the proof of part (i) of the theorem,

$$\beta_h = \lim_p \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}_{t,h}^0} \frac{1}{p^{t+1} n^{h-t}} \left|\Pi^*(w)\right|.$$

Fix a $w \in \mathcal{W}_{t,h}^0$. If $w[2i] = w[2j+1]$ then we have following restriction,

$$\pi(2i-1) - \pi(2i) = \pi(2j+1) - \pi(2j).$$

Let G be the set of indices of size $(h+1)$ corresponding to all generating vertices. It is easy to see that if we consider the above h linear restrictions on the vertices of the circuits and do not take into account the circuit condition $\pi(0) = \pi(2h)$, then each dependent vertex can be written in a *unique* manner as an integral linear combination of generating vertices which occur to the left of that particular vertex in the circuit. Mathematically,

$$\pi(i) = \sum_{j: j \le i, j \in G} a_{i,j} \pi(j) \quad \text{for some } a_{i,j} \in \mathbb{Z}.$$

Note that for $i \in G$, $a_{j,i} = \mathbb{I}\{i = j\}$ and since w is *symmetric*, we have $\pi(2h) = \pi(0)$ so that the circuit condition is automatically satisfied.

To compute the scaled limit of $|\Pi^*(w)|$ introduce the following notation. Define

$$t_{2i} = \frac{\pi(2i)}{p}, \qquad t_{2i+1} = \frac{\pi(2i+1)}{n} \quad \text{and} \quad y_n = p/n.$$

From the above discussion, if $i \notin G \cup \{2h\}$, t_i can be written in a unique manner as a linear combination of $t_G := \{t_j\colon\ j \in G\}$, namely,

$$t_{2i-1} = L_{2i-1,n}^T(t_G) := \sum_{2j-1 \in G, 2j-1 \le 2i-1} a_{2i-1,2j-1} t_{2j-1} + \sum_{2j \in G, 2j \le i} y_n a_{2i-1,2j} t_{2j},$$

$$t_{2i} = L_{2i,n}^T(t_G) := \sum_{2j-1 \in G, 2j-1 \le 2i} (1/y_n) a_{2i,2j-1} t_{2j-1} + \sum_{2j \in G, 2j \le i} a_{2i,2j} t_{2j}.$$

It is obvious that these linear combinations $L^T_{i,n}(t_G)$ depend on the word w but we suppress this dependence for notational brevity. Thus the number of elements in $\Pi^*(w)$ can be expressed alternatively as follows:

$$\begin{aligned}\left|\Pi^*(w)\right| &= \#\Big\{(t_0, t_1, \dots, t_{2h})\colon\ t_{2i} \in \{1/p, 2/p, \dots, p/p\}, t_{2i-1} \in \{1/n, 2/n, \dots, n/n\}\\ &\qquad \text{and } \frac{p}{n}t_{2i} - t_{2i+1} = \frac{p}{n}t_{2j} - t_{2j-1} \text{ if } w[2i+1] = w[2j]\Big\}\\ &= \#\big\{t_G\colon\ t_g \in \{1/p, 2/p, \dots, p/p\} \text{ if } g \text{ is even and } t_g \in \{1/n, 2/n, \dots, n/n\} \text{ if } g \text{ is odd}, g \in G,\\ &\qquad 0 < L^T_{i,n}(t_G) \le 1, \forall i \notin G \cup \{2h\}\big\}.\end{aligned}$$

It does not take us long to recognize the above complicated expression as a multi-dimensional Riemann sum. Therefore from the theory of Riemann integration, its convergence follows and we have

$$\begin{aligned}\beta_h &= \sum_{t=1}^{h-1} y^t \sum_{w \in \mathcal{W}^0_{t,h}} \lim_{p\to\infty} \frac{1}{p^{t+1}n^{h-t}}\left|\Pi^*(w)\right|\\ &= \sum_{t=1}^{h-1} y^t \sum_{w \in \mathcal{W}^0_{t,h}} \underbrace{\int_0^1\int_0^1\cdots\int_0^1}_{h+1} \mathbb{I}\big(0 < L^T_i(t_G) \le 1, \forall i \notin G \cup \{2h\}\big)\,\mathrm{d}t_G,\end{aligned}$$

where $L^T_i(t_G)$ is same as $L^T_{i,n}(t_G)$ with all y_n being replaced by y. This replacement can be justified by Polya's theorem as we can think of each t_i as a discrete uniform random variable converging to a $U(0,1)$ random variable which is continuous. The above argument shows that in case $y_n \to y \in (0,\infty)$, the LSD of $A_p(T)$ exists.

(ii) We now consider the case $\frac{p}{n} \to 0$ and prove part (ii) of the theorem. In view of Theorem 8, we only need to show that for each matched word of length $(4h)$, with $|w| = 2h$,

$$\lim_{p\to\infty} p^{-(1+h)}n^{-h}\left|\Pi^*_{\neq}(w)\right| \text{ exists.}$$

Note that if $\pi \in \Pi^*_{\neq}(w)$, then $\xi_\pi(i) \neq \xi_\pi(i+1)$ for all i odd. Hence, there can be only two types of matching between the L-values as listed below:

1. *Double bond*. A matching is said to have a double bond if there exists two consecutive odd-even L-values which match pairwise with another two consecutive odd-even L-values. There can be again two possibilities:
   (a) *Crossing*. $\xi_\pi(2i+1) = \xi_\pi(2j+2)$ and $\xi_\pi(2i+2) = \xi_\pi(2j+1)$ for some $i < j$.
   (b) *Non-crossing*. $\xi_\pi(2i+1) = \xi_\pi(2j+1)$ and $\xi_\pi(2i+2) = \xi_\pi(2j+2)$ for some $i < j$.
2. *Single bond*. The remaining types of pairing will be termed as single bond. They give rise to following type of equations:

$$\xi_\pi(2i+1) = \xi_\pi(s) \quad \text{and} \quad \xi_\pi(2i+2) = \xi_\pi(t) \quad \text{where } \{s,t\} \neq \{2j+1, 2j+2\} \text{ for all } j.$$

Claim. *Let w be a matched word of length $4h$. If w has a single bond, then*

$$\lim_{p\to\infty} p^{-(1+h)}n^{-h}\left|\Pi^*_{\neq}(w)\right| = 0.$$

Proof. Recall the definition of a Type I generating vertex. It is clear that if $\pi(2i-1)$ is Type I, then

$$\xi_\pi(s) = \xi_\pi(2i-1) \quad \text{or} \quad \xi_\pi(s) = \xi_\pi(2i) \quad \Rightarrow \quad s > 2i.$$

We show that number of Type I odd generating vertices is strictly less than h. Then the proof will follow immediately since the total number of generating vertices is $(2h+1)$.

Total number of odd vertices (generating, non-generating together) is $(2h)$. Let us form two mutually exclusive and exhaustive sets U and V where U contains all odd vertices involved in double bonds and V contains all the rest of the odd vertices. Quite clearly, if $\{\xi_\pi(2u_1-1), \xi_\pi(2u_1)\} = \{\xi_\pi(2u_2-1), \xi_\pi(2u_2)\}$ is a double bond with $u_1 < u_2$, then $\pi(2u_1-1), \pi(2u_2-1) \in U$ and $\pi(2u_1-1)$ is a Type I odd generating vertex. Thus, total number of Type I odd generating vertices in U is $(1/2)|U|$. Next we argue that the total number of Type I generating vertices in V is strictly less than $(1/2)|V|$ and hence, the total number of Type I odd generating vertices is strictly less than h.

Note that exactly half of the odd vertices, which are involved in double bond matching, are Type I odd vertices. Now, let us count the number of Type I odd vertices which are involved in single bond matchings. We list the Type I odd generating vertices in V as

$$V_1 := \{\pi(2g_1-1), \pi(2g_2-1), \ldots, \pi(2g_s-1)\}$$

and the rest of the vertices of V by

$$V_2 := V \setminus V_1 = \{\pi(2d_1-1), \pi(2d_2-1), \ldots, \pi(2d_t-1)\}.$$

For $i \neq j$, write $2i-1 \leftrightarrow 2j-1$ if $\{\xi_{\pi(2i-1)}, \xi_{\pi(2i)}\} \cap \{\xi_{\pi(2j-1)}, \xi_{\pi(2j)}\} \neq \varnothing$. From the definition of Type I odd generating vertex, it is clear that $2g_i-1 \leftrightarrow 2g_j-1$ is not possible.

We claim that $2d_i-1 \leftrightarrow 2g_l-1$ and $2d_i-1 \leftrightarrow 2g_m-1, l \neq m$ cannot occur simultaneously. Because if that happens, then we have

$$\begin{aligned} &\pi(2d_i-2) - \pi(2d_i-1) = \pi(2g_a) - \pi(2g_a-1), \\ &\pi(2d_i) - \pi(2d_i-1) = \pi(2g_b-2) - \pi(2g_b-1), \qquad \{a,b\} = \{l,m\}. \end{aligned}$$

Subtracting we get,

$$\pi(2g_b-1) - \pi(2g_a-1) = \pi(2d_i-2) - \pi(2d_i) + \pi(2g_b-2) - \pi(2g_a).$$

Vertices on the right side are all even and hence the number of choices on the right side is $O(p)$. On the other hand, in the left side we have two Type I odd vertices each of which has free choices of the order n. This is an impossibility.

So in summary, the relation $\leftrightarrow$ associates a vertex in V_1 with two vertices in V_2 (single bond), but a vertex in V_2 is not associated to two distinct vertices in V_1. Therefore, $|V_1| < |V_2|$. So the total number of Type I odd generating vertices in V is strictly less than $|V|/2$. Thus, total number of Type I odd vertices is strictly less than $(|U|+|V|)/2 = h$ which concludes the proof of the claim. $\square$

Reverting to the proof of part (ii), we may now, for the rest of our calculation, consider only those words which produce no single bond. By the circuit constraint, we have

$$\xi_\pi(1) - \xi_\pi(2) + \cdots + \xi_\pi(2i-1) - \xi_\pi(2i) + \cdots + \xi_\pi(4h-1) - \xi_\pi(4h) = 0. \tag{35}$$

Note that if i forms a non-crossing double bond with j then $\xi_\pi(2i-1) - \xi_\pi(2i) = \xi_\pi(2j-1) - \xi_\pi(2j)$. If w has at least one non-crossing double bond then (35) leads to a nontrivial restriction on the vertices of the circuit reducing the number of even generating vertices by one and thus $p^{-(1+h)} n^{-h} |\Pi^*_{\neq}(w)| \to 0$. Thus we may restrict ourselves to those words which give rise to only crossing double bonds. Let us fix one such word w of length $(4h)$, with $|w| = 2h$. Now let us consider a pair of equations forming a crossing double bond:

$$\begin{aligned} &\pi(2i) - \pi(2i+1) = \pi(2j+2) - \pi(2j+1), \\ &\pi(2i+2) - \pi(2i+1) = \pi(2j) - \pi(2j+1) \quad \text{for } i < j. \end{aligned} \tag{36}$$

In the above equations $\pi(2i+1)$ is a Type I odd generating vertex and $\pi(2j+1)$ is a non generating odd vertex which pairs up with $\pi(2i+1)$. Note that,

$$\begin{aligned} \pi(2j+1) &= \pi(2j+2) - \pi(2i) + \pi(2i+1) \\ &= \pi(2j) - \pi(2i+2) + \pi(2i+1). \end{aligned}$$

Since $-p < \pi(2j+2) - \pi(2i) < p$, if $\pi(2i+1)$ is chosen freely between p and $(n-p)$, we do not have any restriction on even vertices imposed by odd vertices, that is, even vertices can be chosen independent of the choice of odd vertices satisfying the following restrictions:

$$\pi(2i) - \pi(2i+2) = \pi(2j+2) - \pi(2j). \tag{37}$$

But if $\pi(2i+1) \in \{1, 2, \dots, p-1\} \cup \{n-p+1, n-p+2, \dots, n\}$, then the choice of even vertices is restricted by the choice of $\pi(2i+1)$ because of the constraint $1 \le \pi(2j+1) \le n$.

Define a new word $\hat{w}$ of length $2h$, so that

$$\hat{w}[i] = \hat{w}[j] \quad \text{iff} \quad w[2i-1] = w[2j] \quad \text{and} \quad w[2i] = w[2j-1].$$

It is easy to see that $|\hat{w}| = h$. Let $\hat{\pi}$ be a circuit of length $2h$ given by $\hat{\pi}(i) = \pi(2i)$.

Since $\frac{p}{n} \to 0$,

$$\begin{aligned} p^{-(1+h)} n^{-h} \bigl|\Pi^*_{\not\equiv}(w)\bigr| &= p^{-(1+h)} n^{-h} \bigl|\Pi^*_{\not\equiv}(w) \cap \bigl\{\pi\colon\ p \le \pi(2i-1) \le n-p, \forall 1 \le i \le 2h\bigr\}\bigr| + \mathrm{o}(1) \\ &= p^{-(1+h)} \bigl|\bigl\{\hat{\pi}\colon\ \hat{w}[i] = \hat{w}[j] \Rightarrow \hat{\pi}(i) - \hat{\pi}(i+1) = \hat{\pi}(j+1) - \hat{\pi}(j)\bigr\}\bigr| + \mathrm{o}(1). \end{aligned}$$

The above restriction on $\hat{\pi}$ is precisely the same restriction on pair matched circuits of length $2h$ that is obtained in the symmetric Toeplitz matrix. See for example [6] and [7]. Also, note that every word of length $2h$ with h letters can be obtained through this procedure. Therefore, in this case, the LSD of $\sqrt{\frac{n}{p}}(A_p(T) - I_p)$ is $\mathcal{L}_T$.

5.2. *Proof of Theorem* 2

(i) By Theorem 6, we know that the hth moment of Hankel LSD is given by

$$\beta_h = \lim_{p\to\infty} \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}_{t,h}} \frac{1}{p^{t+1} n^{h-t}} \bigl|\Pi^*(w)\bigr|. \tag{38}$$

We note that $\Pi^*(w) \subseteq \Pi^*_{\tilde{H}}(w)$ where $\Pi^*_{\tilde{H}}(w)$ is as defined in (16) with symmetric Hankel link function $L(i, j) = i + j$ and each vertex having same range between 1 and $\max(p, n)$ since in the latter case we have more circuits but fewer restrictions. From the arguments given in [6] for symmetric Hankel link function, it follows that for any non-symmetric word w, $n^{-(h+1)}|\Pi^*(w)| \to 0$. Thus, (38) reduces to

$$\beta_h = \lim_{p\to\infty} \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}^0_{t,h}} \frac{1}{p^{t+1} n^{h-t}} \bigl|\Pi^*(w)\bigr|.$$

Let us first consider the case when the link function is symmetric Hankel, i.e. $L(i, j) = i + j$. In that case for a word $w \in \mathcal{W}^0_{t,h}$, if $w[2i] = w[2j+1]$, we have the restriction,

$$\pi(2i-1) + \pi(2i) = \pi(2j) + \pi(2j+1).$$

Just as in the Toeplitz case, we can express each vertex in a unique manner as an integral linear combination of generating vertices occurring to its left.

$$\pi(i) = \sum_{j : j \le i, j \in G} b_{i,j} \pi(j) \quad \text{for some } b_{i,j} \in \mathbb{Z}, \tag{39}$$

with $b_{j,i} = \mathbb{I}\{i = j\}$ if $i \in G$ and $\pi(2h) = \pi(0)$ since w is symmetric.

We scale the vertices as before and call them t_i. Similar to the Toeplitz case, we define $L^H_{i,n}(t_G)$ as the linear combination which expresses t_i in terms of "free cordinates" t_G and $L^H_i(t_G)$ as its limiting version. It immediately yields

$$\lim_{p\to\infty} \frac{1}{p^{t+1}n^{h-t}} \left|\Pi^*(w)\right| = \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{h+1} \mathbb{I}\big(0 < L^H_i(t_G) \le 1, \forall i \notin G \cup \{2h\}\big)\, \mathrm{d}t_G. \tag{40}$$

Now consider the asymmetric link function. Here instead of calculating the $(h+1)$-dimensional Eucledean volume of the entire set $\{0 \le L^H_i(t_G) \le 1, \forall i \notin S \cup \{2h\}\}$, we need to take into account the restricting hyperplanes that arise from asymmetric nature of the link function since it assumes different signs around the diagonal. So, unlike the symmetric case, this imposes the following extra restrictions in addition to the usual equality between two L-values $\xi_\pi(2i)$ and $\xi_\pi(2j+1)$:

$$\text{either } \pi(2i-1) \le \pi(2i), \qquad \pi(2j-1) \le \pi(2j) \quad \text{or} \quad \pi(2i-1) > \pi(2i), \qquad \pi(2j-1) > \pi(2j).$$

Thus the size of the set $\Pi^*(w)$ is same as the cardinality of $\{t_G\colon 0 < L^H_{i,n}(t_G) \le 1, \forall i \notin G \cup \{2h\}, \operatorname{sgn}(L^H_{2i+1,n}(t_G) - y_n L^H_{2i,n}(t_G)) = \operatorname{sgn}(L^H_{2j+1,n}(t_G) - y_n L^H_{2j,n}(t_G)) \text{ if } w[2i+1] = w[2j]\}$, where each $t_i, i \in G$ takes values in $\{1/n, 2/n, \dots, n/n\}$ or $\{1/p, 2/p, \dots, p/p\}$ depending on whether i is odd or even.

This entire set may be written as a product of indicator functions in terms of $\{L^H_{i,n}(t_G)\}$, albeit in a complicated manner. When summed over t_G, in the limit this equals the corresponding Riemann integral where the indicators are replaced by their limits and y_n is replaced by y. Let us denote the giant indicator function by $f^H(t_G)$. So, we have

$$\beta_h = \sum_{t=1}^{h-1} y^t \sum_{w \in \mathcal{W}^0_{t,h}} \lim_{p\to\infty} \frac{1}{pn^h} \left|\Pi^*(w)\right| = \sum_{t=1}^{h-1} y^t \sum_{w \in \mathcal{W}^0_{t,h}} \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{h+1} f^H(t_G)\, \mathrm{d}t_G. \tag{41}$$

(ii) By Theorem 9, it is equivalent to prove the existence of the LSD for the symmetric Hankel case. Along this line, we can just imitate the argument for the asymmetric Toeplitz matrix in Regime II. Here also the essential contribution comes from the words having only crossing double bonds. But we now have a different pair of equations instead of (36) in Toeplitz case,

$$\begin{aligned} &\pi(2i) + \pi(2i+1) = \pi(2j+2) + \pi(2j+1), \\ &\pi(2i+2) + \pi(2i+1) = \pi(2j) + \pi(2j+1) \quad \text{for } i < j. \end{aligned} \tag{42}$$

But once we cancel the odd vertices as we did in the Toeplitz case, we are again reduced to the Toeplitz type restrictions on even vertices. We omit the details. Hence we conclude that the LSD exists and is $\mathcal{L}_T$.

5.3. *Proof of Theorem* 3

(i) We need to show that for each $h \ge 1$ the following limit exists.

$$\beta_h = \lim_{p\to\infty} \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}_{t,h}} \frac{1}{p^{t+1}n^{h-t}} \left|\Pi^*(w)\right|. \tag{43}$$

Similar to Hankel case, we note that $\Pi^*(w) \subseteq \Pi^*_{\tilde{R}}(w)$ where $\Pi^*_{\tilde{R}}(w)$ is as defined in (16) with symmetric reverse circulant link function $L(i,j) = (i+j) \bmod n$. In [6], it has been argued that for symmetric reverse circulant link function, the circuit condition enforces that for any non-symmetric word w, $n^{-(h+1)}|\Pi^*(w)| \to 0$. Thus, the sum in (43) is only over symmetric words,

$$\beta_h = \lim_{p\to\infty} \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}^0_{t,h}} \frac{1}{p^{t+1}n^{h-t}} \left|\Pi^*(w)\right|.$$

It now remains to show that for each symmetric word $w \in \mathcal{W}^0_{t,h}$, $p^{-(t+1)}n^{-(h-t)}|\Pi^*(w)|$ converges.

Let us first consider the symmetric link function $L(i,j) = (i+j) \bmod n$. In this case, if $w[i] = w[j]$, we obtain the restriction of the following type:

$$\big(\pi(i) + \pi(i-1)\big) \bmod n = \big(\pi(j) + \pi(j-1)\big) \bmod n,$$

which is equivalent to

$$\big(\pi(i) + \pi(i-1)\big) - \big(\pi(j) + \pi(j-1)\big) \in n\mathbb{Z}.$$

We now have exactly the same set of equations as in the Hankel case (39), with some added relaxations given as follows

$$\pi(i) - \sum_{j: j \le i, j \in G} b_{i,j}\pi(j) \in n\mathbb{Z} \quad \text{for all } i \notin G.$$

If $i \notin G \cup \{2h\}$, we can choose a *unique* integer $k_{i,n} = k_{i,n}(\pi(j)\colon\ j \in G)$ such that $1 \le \sum_{j: j \le i, j \in G} b_{i,j}\pi(j) + k_{i,n} \le n$. Thus once we fix the generating vertices, there is exactly one choice for each of the non-generating odd vertices. For the non-generating even vertices, things are a bit complicated.

For a real number a, let

$$\wp(a) := \max\{m \in \mathbb{Z}\colon\ m < a\}, \qquad \varrho(a) := a - \wp(a).$$

Using these notation, $k_{i,n}$ can be written as the following

$$\frac{k_{i,n}}{n} = \frac{1}{n}\varrho\bigg(\sum_{j: j \le i, j \in G} b_{i,j}\pi(j)\bigg).$$

Now let us fix $2i \notin G$, $i \ne h$. We have at least $[y_n]$ choices for the vertex $\pi(2i)$. Moreover, we can have an additional choice for $\pi(2i)$ if

$$\sum_{j: j \le 2i, j \in G} b_{2i,j}\pi(j) + k_{2i,n} \le p - [y_n]n.$$

Dividing by n, the above condition can be rewritten as

$$y_n L^H_{2i,n}(t_G) + \varrho\big(y_n L^H_{2i,n}(t_G)\big) \le y_n - [y_n].$$

Let $S = \{2i\colon\ 2i \notin G, i \ne h\}$. From the above discussion, we can conclude that for the symmetric link function, the size of the set $\Pi^*(w)$ when $w \in \mathcal{W}^0_{t,h}$ is given by

$$[y_n]^{h-t-1}p^{t+1}n^{h-t} + \sum_{\varnothing \ne S' \subseteq S} \#\big\{t_G\colon\ y_n L^H_{2i,n}(t_G) + \varrho\big(y_n L^H_{2i,n}(t_G)\big) \le y_n - [y_n], \forall 2i \in S'\big\}.$$

Note that ϱ has discontinuities only at integer points. Therefore, we have the following convergence

$$\begin{aligned}
&\lim_{p\to\infty} \frac{1}{p^{t+1}n^{h-t}}\big|\Pi^*(w)\big| \\
&\quad = [y]^{h-t-1} + \sum_{\varnothing \ne S' \subseteq S} \int_0^1 \cdots \int_0^1 \mathbb{I}\big(yL^H_{2i}(t_G) + \varrho\big(yL^H_{2i}(t_G)\big) \le y - [y], \forall 2i \in S'\big)\,\mathrm{d}t_G.
\end{aligned}$$

Coming back to the asymmetric case, we now have extra restrictions,

$$\operatorname{sgn}\big(\pi(2i+1) - \pi(2i)\big) = \operatorname{sgn}\big(\pi(2j+1) - \pi(2j)\big) \quad \text{if } w[2i] = w[2j+1].$$

We can now incorporate restrictions $\operatorname{sgn}(t_{2i+1} - yt_{2i}) = \operatorname{sgn}(t_{2j+1} - yt_{2j})$ in the integral as we did in the Hankel case. The even non-generating vertices, except for $\pi(2h)$ which is constrained to equal to $\pi(0)$, do not have a unique choice once we determine the generating vertices, instead there are $[y_n] + 1$ choices for each of them as observed above.

In the symmetric case, for many of the choices, the integrand was simply equal to 1. But here due to the additional constraints regarding signs, the integrand is not necessarily equal to 1 and simplification is not possible anymore and all the relevant indicators will appear in the integrand. We omit details.

(ii) In Regime II, invoking Theorem 9, it suffices to work with the symmetric link function $L(i, j) = (i + j) \bmod n$. We again imitate the proof of the Toeplitz case. Here a typical restriction in a word containing only crossing double bonds reads as,

$$\begin{aligned}
&\big(\pi(2i) + \pi(2i+1)\big) \bmod n = \big(\pi(2j+2) + \pi(2j+1)\big) \bmod n, \\
&\big(\pi(2i+2) + \pi(2i+1)\big) \bmod n = \big(\pi(2j) + \pi(2j+1)\big) \bmod n \text{ for } i < j.
\end{aligned} \tag{44}$$

As in the case of Toeplitz and Hankel matrices, we choose generating odd vertices between p and $(n - p)$, the only restriction that even vertices need to satisfy is

$$\big(\pi(2i) - \pi(2i+2)\big) \bmod n = \big(\pi(2j+2) - \pi(2j)\big) \bmod n.$$

But since p is negligible compared to n, and an even vertex can take values between 1 and p, this is equivalent to usual Toeplitz restriction

$$\big(\pi(2i) - \pi(2i+2)\big) = \big(\pi(2j+2) - \pi(2j)\big).$$

Hence, the LSD exists and the limit is exactly $\mathcal{L}_T$.

5.4. *Proof of Theorem* 4

(i) As in the previous examples, we need to prove that for each $h \geq 1$ the following limit exists.

$$\beta_h = \lim_{p\to\infty} \sum_{t=0}^{h-1} \left(\frac{p}{n}\right)^t \sum_{w \in \mathcal{W}_{t,h}} \frac{1}{p^{t+1} n^{h-t}} \big|\Pi^*(w)\big|.$$

We now obtain exactly the same set of equations as in the Toeplitz case with some added relaxations given as follows

$$\pi(i) - \sum_{j: j \leq i, j \in G} a_{i,j} \pi(j) \in n\mathbb{Z} \quad \text{for all } i \notin G.$$

In the Toeplitz proof, we already argued that for any non-symmetric word, $\sum_{j:j\in G} a_{2h,j}\pi(j) = \pi(0)$ induces a nontrivial restriction on the generating vertices. On the other hand, we can have a bounded ($\leq [y_n] + 1$) number of choices for the non-generating vertices. Thus non-symmetric words do not contribute in the limit.

The rest of the proof is exactly similar to the (symmetric) reverse circulant case. We omit details to avoid repetition.

(ii) We note that on and above the diagonal, the circulant link function exactly matches with the link function $L(i, j) = j - i$. But the link function $L(i, j) = j - i$ is nothing but the asymmetric Toeplitz link function once we index input random variables as $\{x_{-i} : i \in \mathbb{Z}\}$. We may now invoke Theorem 9 to conclude the proof.

5.5. *Some comments on the four examples*

Suppose $p/n \to y \in (0, \infty)$ and R1 holds. We make the following observations.

(i) *The LSD of $A_p(T)$ and $A_p(H^{(s)})$ are identical*. This has been observed in Remark 1.2 of [7] for $p = n$. The same argument extends to the general case of rectangular matrices. As before let T_p be the $p \times n$ asymmetric Toeplitz matrix and P_n be the symmetric permutation matrix $P_n := ((\mathbb{I}\{i + j = n + 1\}))_{i,j=1}^n$. Then note that $H_p^{(s)} := T_p P_n$ is the $p \times n$ Hankel matrix, with symmetric link, for the input sequence $\{x_{n+1+k}\colon k \geq 0\}$. This implies $H_p^{(s)} H_p^{(s)\prime} = T_p(P_n P_n')T_p' = T_p T_p'$ since $P_n P_n' = I_n$. Therefore, the assertion follows.

(ii) *The LSD of $A_p(C)$ and $A_p(R^{(s)})$ are identical.* To see this, first note the distribution of singular values for the $p \times n$ matrix C_p with the usual circulant link function $L(i,j) = (n + j - i) \bmod n$ will be unchanged if we use a new link function $L(i,j) = (n + i - j) \bmod n$. To convince ourselves, all we need to do is to take $\{x_{(n-k) \bmod n}\colon k = 0, 1, \dots, n-1\}$ as the input sequence in the second case.

Second, observe that if $\widehat{C}_p$ is the $p \times n$ "modified" circulant matrix with the link function $L(i,j) = (n + i - j) \bmod n$, then $R_p^{(s)} := \widehat{C}_p P_n$ is the $p \times n$ symmetric reverse circulant matrix for the input sequence $\{x_{1+k}\colon k \ge 0\}$. The claim follows immediately.

(iii) *The LSD of $A_p(H)$ and $A_p(H^{(s)})$ are different.* Note that if $p = n$ then the square of Wigner and covariance matrices has the same LSD. It may be tempting to believe that the same holds for other link function also when $p = n$. This is not true and the LSD for $A_p(H)$ and $A_p(H^{(s)})$ are different.

To understand why this is so, note that Bose and Sen [6] define a certain class of symmetric words known as *Catalan words*. It turns out that for the Wigner link function, the non-Catalan words do not contribute in the limit and for any Catalan word w,

$$\lim n^{-(h+1)} \left| \Pi^*(w) \right| = 1$$

for both, $A_p(W)$ and $A_p(W^{(s)})$ and hence the LSD for $A_p(W)$ and $A_p(W^{(s)})$ are equal.

For the symmetric and asymmetric Hankel link functions, we still have for any Catalan word w,

$$\lim n^{-(h+1)} \left| \Pi^*(w) \right| = 1.$$

However, there are now additional contributions from the words which are non-Catalan but symmetric (e.g. $w = abcabc$) and they do not agree for H and $H^{(s)}$. Indeed, the contributions from such words for asymmetric link function are strictly less than those for symmetric link functions due to additional sign constraints. So, hth moment of the LSD of $A_p(H)$ is strictly greater than hth moment of the LSD of $A_p(H^{(s)})$. See also Fig. 3.

(iv) Suppose $p = yn$ where y is an integer. Then the hth moment of the LSD of $A_p(C)$ has a closed form expression and is given by

$$\beta_h^C = \sum_{t=0}^{h-1} y^t \left| \mathcal{W}_{t,h}^0 \right| a^{h-t-1} = y^{h-1} \sum_{t=0}^{h-1} \left| \mathcal{W}_{t,h}^0 \right| = h! a^{h-1}.$$

Suppose Y is distributed as the reverse circulant LSD $f_R(\cdot)$. Let ξ be a Bernoulli random variable having mass $(1 - 1/y)$ at zero and independent of Y. Then $A_p(C)$ has LSD $\sim a\xi Y^2$. This is easily verified by noting that

$$\mathbb{E}\big(a\xi Y^2\big)^h = a^h \mathbb{E}(\xi) \mathbb{E}\big(Y^{2h}\big) = a^{h-1} h!.$$

(v) In Regime II the symmetric link function $L(i,j) = |i - j|$ does not obey Assumption A'. However, since it obeys Assumption A, by Theorem 10, $\sqrt{\frac{n}{p}}(A_p(T^{(s)}) - I_p)$ has the same LSD as for the asymmetric case, namely $\mathcal{L}_T$.

5.6. *Simulations*

(i) The histogram from 50 replications for the ESD of $A_p(T)$ when $p = 300$, $p/n = 1/3$ is given in Fig. 1, illustrating Theorem 1(i).

It is not too difficult to show that the support is unbounded. The more interesting evidence is that the support of the LSD excludes a neighbourhood of zero. Recall that for the S matrix with $y < 1$, the infimum of the support is $(1 - \sqrt{y})^2$. It will be interesting to prove that the infimum of the support in this case is also strictly positive and find its value. Such a result would be of interest due to numerical technique of "pre-multiplication" by patterned matrices which is used to solve large systems of sparse equations, see for example [12].

(ii) By Theorem 1(ii), in Regime II, the LSD for the asymmetric Toeplitz matrix exists. By Theorem 10, the same LSD continues to hold for symmetric Toeplitz matrices. In Fig. 2 we report the result of a simulation of these matrices. The two histograms of ESD, based on 30 replications each, are similar and the apparent difference could be only due to the finite sample effect, $p = 200$, $p/n = 0.01$.

(iii) Figure 3 shows the histograms of the ESD from 50 replications for $A_p(X)$ with $p = n = 500$ where X is the symmetric and the asymmetric Hankel matrix. This illustrates Theorem 2(i).

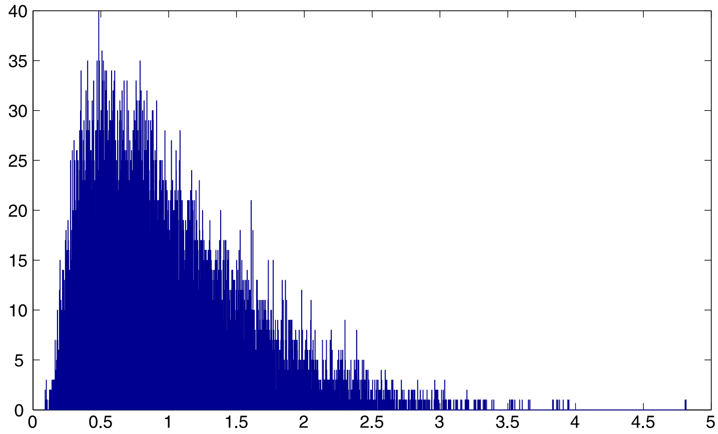

Fig. 1. Histogram for empirical spectral distribution for 50 realizations of $(1/n)TT'$ with $U(-\sqrt{3}, \sqrt{3})$ entries where T is a 300×900 asymmetric Toeplitz matrix.

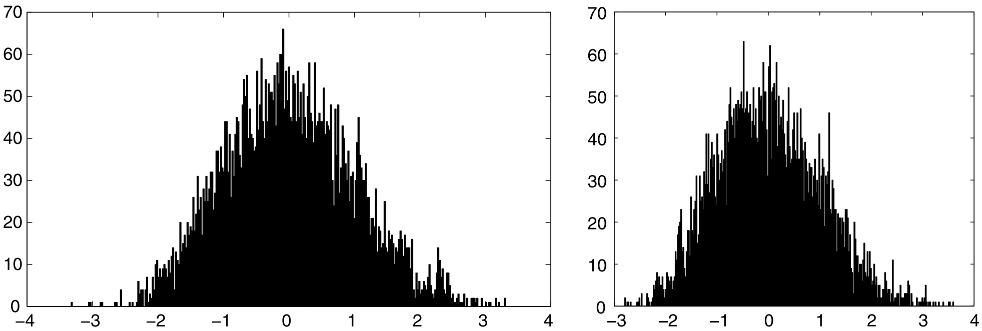

Fig. 2. Histograms for empirical spectral distribution for 30 realizations of $\frac{1}{\sqrt{np}}(TT' - nI)$ with $N(0, 1)$ entries where T is a 200×20000 Toeplitz matrix with asymmetric link function (left) and symmetric link function (right).

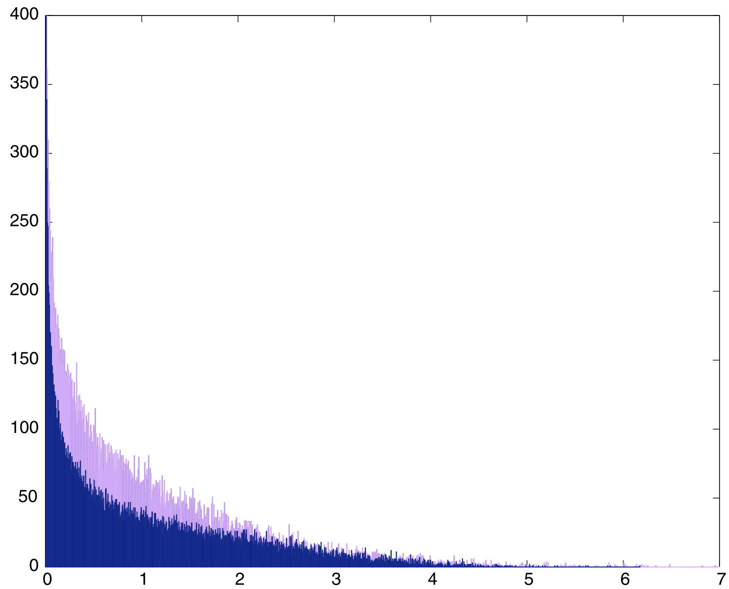

Fig. 3. Histograms for empirical spectral distribution for 50 realizations of $(1/n)HH'$ with $N(0, 1)$ entries where H is a 500×500 Hankel matrix with symmetric link function (color: gray) and asymmetric link function (color: black).

Acknowledgment

We thank the Referee for the detailed comments and suggestions. These have led to a significant improvement in the exposition of the paper.

References

[1] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statist. Sinica* **9** (1999) 611–677 (with discussions). MR1711663
[2] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Science Press, Beijing, 2006.
[3] Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.* **16** (1988) 863–875. MR0929083
[4] R. Bhatia. *Matrix Analysis*. Springer, New York, 1997. MR1477662
[5] A. Bose and J. Mitra. Limiting spectral distribution of a special circulant. *Statist. Probab. Lett.* **60** (2002) 111–120. MR1945684
[6] A. Bose and A. Sen. Another look at the moment method for large dimensional random matrices. *Electron. J. Probab.* **13** (2008) 588–628. MR2399292
[7] W. Bryc, A. Dembo and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34** (2006) 1–38. MR2206341
[8] W. Feller. *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York, 1966. MR0210154
[9] U. Grenander and J. W. Silverstein. Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **32** (1977) 499–519. MR0476178
[10] C. Hammond and S. J. Miller. Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* **18** (2005) 537–566. MR2167641
[11] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1982) 1–38. MR0650926
[12] E. Kaltofen. Asymptotically fast solution of Toeplitz-like singular linear systems. In *ISSAC* 297–304. ACM, New York, 1994. Available at http://www4.ncsu.edu/~kaltofen/bibliography/94/Ka94_issac.pdf.
[13] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb. (N.S.)* **72** (1967) 507–536. MR0208649
[14] A. Massey, S. J. Miller and J. Sinsheimer. Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices. *J. Theoret. Probab.* **20** (2007) 637–662. MR2337145
[15] J. W. Silverstein and A. M. Tulino. Theory of large dimensional random matrices for engineers. *IEEE Ninth International Symposium on Spread Spectrum Techniques and Applications* 458–464, 2006.
[16] A. M. Tulino and S. Verdu. *Random Matrix Theory and Wireless Communications*. Now Publishers Inc., 2004.
[17] K. W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** (1978) 1–18. MR0467894
[18] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)* **62** (1955) 548–564. MR0077805
[19] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958) 325–327. MR0095527
[20] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** (1928) 32–52.
[21] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** (1986) 50–68. MR0862241
[22] Y. Q. Yin and P. R. Krishnaiah. Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.* **30** (1985) 810–816. MR0816299