SPA: 2916 Model 1 pp. 1–27 (col. fig: NIL)

ARTICLE IN PRESS

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications xx (xxxx) xxx–xxx

stochastic processes and their applications

www.elsevier.com/locate/spa

Q1 On the empirical spectral distribution for matrices with long memory and independent rows

Q2 F. Merlevède[a,*], M. Peligrad[b]

[a] *Université Paris Est, LAMA (UMR 8050), UPEM, CNRS, UPEC, France*

[b] *Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, OH 45221-0025, USA*

Received 29 June 2015; received in revised form 25 February 2016; accepted 26 February 2016

Abstract

In this paper we show that the empirical eigenvalue distribution of any sample covariance matrix generated by independent samples of a stationary regular sequence has a limiting distribution depending only on the spectral density of the sequence. We characterize this limit in terms of Stieltjes transform via a certain simple equation. No rate of convergence to zero of the covariances is imposed, so, the underlying process can exhibit long memory. If the stationary sequence has trivial left sigma field the result holds without any other additional assumptions. This is always true if the entries are functions of i.i.d.

As a method of proof, we study the empirical eigenvalue distribution for a symmetric matrix with independent rows below the diagonal; the entries satisfy a Lindeberg-type condition along with mixingale-type conditions without rates. In this nonstationary setting we point out a property of universality, meaning that, for large matrix size, the empirical eigenvalue distribution depends only on the covariance structure of the sequence and is independent on the distribution leading to it. These results have interest in themselves, allowing to study symmetric random matrices generated by random processes with both short and long memory.

MSC: 60F05; 60F15; 60G42; 60G60

Keywords: Random matrices; Stieltjes transform; Martingale approximation; Lindeberg method; Empirical eigenvalue distribution; Spectral density; Sample covariance matrix

* Corresponding author.

E-mail addresses: florence.merlevede@u-pem.fr (F. Merlevède), peligrm@ucmail.uc.edu (M. Peligrad).

http://dx.doi.org/10.1016/j.spa.2016.02.016

0304-4149/

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

1. Introduction and the main result

Due to the fact that random matrices appear in many applied fields, their empirical spectral distribution is a subject of intense research. Earlier works, pioneered by the celebrated paper by Wigner [46], deal with symmetric matrices having independent entries below the diagonal. Only in the last three decades there has been an intense effort to weaken the hypotheses of independence and various forms of weak dependence have been considered. The progress was in general achieved first for Gaussian random matrices. For this case the joint distribution of eigenvalues is tractable. Among the papers for symmetric Gaussian matrices with correlated entries we mention the works of Khorunzhy and Pastur [29], Boutet de Monvel et al. [13], Boutet de Monvel and Khorunzhy [12], Chakrabarty et al. [17].

Our paper is essentially motivated by the study of large sample covariance matrices, which is a very important topic in multivariate analysis and signal processing.

The spectral analysis of large-dimensional sample covariance matrices has been actively studied starting with the seminal work of Marčenko and Pastur [32] who considered independent random samples from an independent multidimensional vector. Later, also for the independent case, Wachter [45] established the almost sure results and recently Jin et al. [28] generalized it to auto-cross covariance matrix. A big step forward was the study of the dependent case represented in numerous papers. Basically, the entries of the matrix were allowed to be linear combinations of an independent sequence. The first paper where such a model was considered is by Yin and Krishnaiah [50] followed by important contributions by Yin [49], Silverstein [40], Silverstein and Bai [41], Hachem et al. [26], Pfaffel and Schlemm [35], Yao [48], Davis et al. [20], Pan et al. [34], Liu et al. [30], Bhattacharjee and Bose [8] among others.

A departure from linear models was considered by Bai and Zhou [3] who derived the limiting spectral distribution of large sample covariance matrices provided that the true covariance matrix has bounded spectral norm and the entries satisfy a dependence type condition. This dependence condition, sometimes called "good vector condition" is satisfied for Gaussian vectors or for isotropic vectors with log-concave distribution as shown in [33]. Note that the circular law for random matrices with independent isotropic unconditional log-concave rows has been proved by Adamczak [1]. As applications of their main result, Bai and Zhou [3] exhibited the limiting spectral distributions of Spearman's rank correlation matrices, sample correlation matrices and sample covariance matrices from finite populations. When applied to linear models the conditions imposed in the paper by Bai and Zhou can be verified when the innovations are square integrable and the coefficients are absolutely summable as shown in [48,34]. It should be mentioned that the bounded spectral norm condition imposed to the true covariance matrix does not allow to derive the limiting spectral distribution of large sample covariance matrices associated with linear processes exhibiting long range dependence.

Recently, Banna and Merlevède [4] considered samples from a stationary process whose variables are functions of i.i.d. and proved, under a dependence condition implying the absolute summability of the covariances, that the asymptotic behavior of the empirical eigenvalue distribution can be obtained by analyzing a Gaussian matrix with the same covariance structure. In [5], this result has been improved and extended to large covariance matrices associated with square integrable variables that are functions of an i.i.d. random field. In this latter paper, it is proved that no extra assumptions are needed to reduce the study of empirical spectral distribution to the one of a Gaussian matrix with the same covariance structure.

Even if many models encountered in time series analysis can be rewritten as functions of an i.i.d. sequence, this assumption is not completely satisfactory since many stationary processes

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

even with trivial left sigma field cannot be in general represented as a function of an i.i.d. sequence, as shown for instance in [38]. The main goal of our paper is then to study the asymptotic behavior of the empirical eigenvalue distribution of large sample covariance matrices associated with stationary processes with variables which are not necessarily functions of an i.i.d. sequence, and to exhibit the spectral limiting distribution even for the case when they have long memory.

In Theorem 1, we find the limiting empirical eigenvalue distribution for the sample covariance matrix of a stationary process which is regular. The regularity is an ergodic-type property and includes many classes of stochastic processes which are not functions of i.i.d. (see Section 3). Our result also shows that the limit can be obtained much beyond the situation of short range dependent case which corresponds to continuous and bounded spectral densities or absolutely summable covariances. It also applies to long range dependent stationary stochastic processes. We show that the limit of the empirical spectral distribution exists and we also characterize the limit in terms of its Stieltjes transform. This limit depends only on the spectral density of the process, even for the case when it is not continuous or even square integrable. The limit can also be characterized via the free multiplicative convolution which potentially gives further insight in the process spectral density.

Furthermore, the technical theorems leading to Theorem 1 are also important. They reduce the study of the empirical spectral distribution of symmetric matrices with independent regular rows, below diagonal, to the study of the sequence of the expected value of Stieltjes transforms associated to a Gaussian matrix with the same covariance structure. These results are set in the non-stationary case for variables satisfying a certain Lindeberg condition. Their proofs are complicated by the fact that our intention was to avoid the use of rates of decay of the covariances.

In order to stress the importance of our results we include several applications to regular processes, functions of i.i.d. random variables and linear processes with martingale differences innovations. As we shall see, Theorem 1 applies to large sample covariance matrices constructed from independent copies of any stationary process whose entries are functions of i.i.d. random variables which are centered and has finite second moments. In particular the theorem applies to any causal linear process with i.i.d. innovations as soon as the process exists in $\mathbb{L}^2$, so it could have long memory.

Our proofs are a blend of probabilistic techniques for dependent structures such as the big and small block argument, martingale approximations and properties of Gaussian processes. Because our variables are correlated the method of proof is based on the Stieltjes transform, which is well adapted to handle dependent entries. The Stieltjes transform is also useful to characterize the limit.

Here are some notations used all along the paper. The notation $[x]$ is used to denote the integer part of a real x. The notation $\mathbf{0}_p$ means a row vector of size p with components equal to zero. When no confusion is possible concerning the size of a null vector $\mathbf{0}$ we will omit the index of its size. For a matrix A, we denote by A^T its transpose matrix, by $\mathrm{Tr}(A)$ its trace. We shall also use the notation $\|X\|_r$ for the $\mathbb{L}^r$-norm ($r \geq 1$) of a real-valued random variable X.

For any sequence of square matrices A_n of order n with only real eigenvalues $\lambda_{1,n} \leq \cdots \leq \lambda_{n,n}$, the spectral distribution function is defined by

$$F^{A_n}(x) = \frac{1}{n} \sum_{k=1}^{n} I(\lambda_{k,n} \leq x),$$

where $I(B)$ denotes the indicator of an event B. The general problem is to find a distribution function F such that $F^{A_n} \to F$ at all points of continuity of F, or equivalently $d(F^{A_n}, F) \to 0$, where the Lévy distance between two distribution functions F and G is defined by

$$d(F, G) = \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon \ \forall x \in \mathbb{R}\}.$$

The Stieltjes transform of F^{A_n} is given by

$$S^{A_n}(z) = \int \frac{1}{x - z} dF^{A_n}(x) = \frac{1}{n}\mathrm{Tr}(A_n - z\mathbf{I}_n)^{-1},$$

where $z = u + \mathrm{i}v \in \mathbb{C}^+$ (the set of complex numbers with positive imaginary part), and $\mathbf{I}_n$ is the identity matrix of order n. It is well-known that $\lim_{n\to\infty} d(F^{A_n}, F) = 0$ if and only if for all $z \in \mathbb{C}^+$, $S_{A_n}(z) \to S_F(z)$. We can also see, for instance, in Proposition 2.1 in [11], that the estimate of the Lévy distance between empirical spectral distribution functions associated with two matrices can be also given in terms of their Stieltjes transforms.

Let N and p be two positive integers and consider the $N \times p$ matrix

$$\mathcal{X}_{N,p} = \big(X_{ij}\big)_{1\le i\le N, 1\le j\le p}, \tag{1}$$

where X_{ij}'s are real-valued random variables. Define now the symmetric matrix $\mathbb{B}_N$ of order p by

$$\mathbb{B}_N = \frac{1}{N}\mathcal{X}_{N,p}^T \mathcal{X}_{N,p}. \tag{2}$$

The matrix $\mathbb{B}_N$ is usually referred to as the sample covariance matrix associated with the process $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$. It is also known under the name of Gram random matrix.

In Theorem 1, we consider N independent copies $(X_{ij})_{j\in\mathbb{Z}}$, $i = 1, \ldots, N$ of a stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of real-valued random variables in $\mathbb{L}^2$ and give sufficient conditions to characterize the limiting distribution of $F^{\mathbb{B}_N}$ (also known under the name of spectral limiting distribution of $\mathbb{B}_N$) when $p/N \to c \in (0, \infty)$.

Relevant to this characterization is the notion of spectral distribution function induced by the covariances of $(X_i)_{i\in\mathbb{Z}}$. By Herglotz's Theorem (see e.g. [16]), there exists a non-decreasing function G (the spectral distribution function) on $[-\pi, \pi]$ such that, for all $j \in \mathbb{Z}$, $\mathrm{Cov}(X_0, X_j) = \int_{-\pi}^{\pi} \exp(\mathrm{i}j\theta) dG(\theta)$. If G is absolutely continuous with respect to the normalized Lebesgue measure λ on $[-\pi, \pi]$, then the Radon–Nikodym derivative f of G with respect to the Lebesgue measure is called the spectral density, it is a nonnegative, even and integrable function on $[-\pi, \pi]$ which satisfies

$$c_j = \mathrm{Cov}(X_0, X_j) = \int_{-\pi}^{\pi} \exp(\mathrm{i}j\theta) f(\theta) d\theta, \qquad j \in \mathbb{Z}.$$

This setting is very natural for statistical interpretation. For instance let $\mathbf{X}$ be a $p \times 1$ random vector. In order to estimate its covariance structure $(c_j)_{1\le j\le p}$ we consider N independent observations from $\mathbf{X}$ which are the i.i.d. vectors $\mathbf{X}_i = (X_{ij})_{1\le j\le p}$. If we form a matrix having the vectors $\mathbf{X}_i$ as columns, then the covariance matrix is estimated as $\mathbb{B}_N$ defined in (2). When p is fixed and $N \to \infty$ this estimator is consistent. However when $p/N \to c \in (0, \infty)$ this is no longer true. We intend to understand the limiting spectral distribution of $F^{\mathbb{B}_N}$ when the observations are generated by a stationary sequence $(X_i)_{i\in\mathbb{Z}}$ and its covariance structure which is encapsulated in the process' spectral density.

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

We shall introduce the following regularity conditions. Define the left tail sigma field of $(X_i)_{i\in\mathbb{Z}}$ by $\mathcal{G}_{-\infty} = \bigcap_{k\in\mathbb{Z}} \mathcal{G}_k$ where $\mathcal{G}_k = \sigma(X_j, j \le k)$

$$\mathbb{E}(X_0|\mathcal{G}_{-\infty}) = 0 \quad \text{a.s.} \tag{3}$$

and for every integer k

$$\mathbb{E}(X_0 X_k|\mathcal{G}_{-\infty}) = \mathbb{E}(X_0 X_k) \quad \text{a.s.} \tag{4}$$

We point out that if (3) holds, then the process $(X_k)_{k\in\mathbb{Z}}$ is purely non deterministic. Hence, by a result of Szegö (see for instance [10, Theorem 3]) if (3) holds, the spectral density f of $(X_k)_{k\in\mathbb{Z}}$ exists and if X_0 is non degenerate,

$$\int_{-\pi}^{\pi} \log f(t)\, dt > -\infty;$$

in particular, f cannot vanish on a set of positive measure.

Theorem 1. *Consider N independent copies $(X_{ij})_{j\in\mathbb{Z}}$, $i = 1,\dots,N$ of a stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of real-valued random variables centered and in $\mathbb{L}^2$ and that satisfies the conditions (3) and (4). Assume $p/N \to c \in (0,\infty)$. Then there is a nonrandom probability distribution F such that $d(F^{\mathbb{B}_N}, F) \to 0$ a.s. Furthermore, the Stieltjes transform $S = S(z)$, $z \in \mathbb{C}^+$, of F is uniquely determined by the equation*

$$z = -\frac{1}{\underline{S}} + \frac{c}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\underline{S} + (2\pi f(\lambda))^{-1}} d\lambda, \tag{5}$$

where $\underline{S} := -(1-c)/z + cS$ and $f(\cdot)$ is the spectral density of $(X_k)_{k\in\mathbb{Z}}$.

When the variables (X_{ij}) are i.i.d. the spectral density is $f(\lambda) = \sigma^2/2\pi$ where $\sigma^2 = \mathbb{E}(X_{00}^2)$ and the solution to Eq. (5) is the well-known Marčenko–Pastur distribution, MP, whose density is given by

$$g_y(x) = \frac{1}{2\pi c\sigma^2 x}\sqrt{(a-x)(x-b)}\, I(a \le x \le b)$$

and a point mass $1 - 1/c$ at the origin if $c > 1$, where $a = \sigma^2(1-\sqrt{c})^2$ and $b = \sigma^2(1+\sqrt{c})^2$.

As a consequence or our proof, we can make the following remark which will be justified at the end of the paper:

Remark 2. The probability measure F which appears as the limit in Theorem 1 can also be described as the free multiplicative convolution $\mu_f \otimes MP$ of the probability distribution μ_f of the variable $2\pi f(U)$ where U is uniformly distributed on $[-\pi,\pi]$ with Marčenko–Pastur distribution. F has compact support if and only if f has compact support.

Potentially, via some newly developed numerical free deconvolution methods (see for instance [39]) from the limiting spectral distribution F we can find the distribution of $f(U)$ where U is uniformly distributed on $[-\pi,\pi]$. If the spectral density is monotonous on $[0,\pi]$ the distribution of $f(U)$ will uniquely determine the spectral density f. We shall not pursue the numerical methods in this paper.

Note that if $\mathcal{G}_{-\infty}$ is trivial then the conditions (3) and (4) hold. Therefore we can immediately formulate the following corollary to Theorem 1:

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

Corollary 3. *Consider N independent copies $(X_{ij})_{j\in\mathbb{Z}}$, $i = 1, \ldots, N$ of a stationary sequence $(X_i)_{i\in\mathbb{Z}}$ of real-valued random variables centered and in $\mathbb{L}^2$ with trivial left tail sigma field, $\mathcal{G}_{-\infty}$. Then the conclusion of Theorem 1 holds.*

2. Some technical results for symmetric matrices

A key step in the proof of Theorem 1 is to show that the study of the limiting spectral distribution function of $\mathbb{B}_N$ can be reduced to studying the same problem as for a Gaussian matrix with the same covariance structure. This step will be achieved with the help of some preliminary technical results concerning symmetric matrices with independent rows below the diagonal. These technical results have interest in themselves since they show that, for symmetric matrices with independent rows below the diagonal, very simple regularity conditions on the entries of each row allow to reduce the study of their limiting spectral distribution function to the one of a symmetric Gaussian matrix with the same covariance structure. In particular, this applies when the rows, below the diagonal, are independent and generated by the same stationary sequence provided it is regular, i.e. has a trivial left tail sigma-field.

To state the results of this section, let us introduce some notations. Let $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{N}^2}$ be real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In what follows, we consider the symmetric $n \times n$ random matrix $\mathbf{X}_n$ defined as follows: for any i and j in $\{1, \ldots, n\}$,

$$(\mathbf{X}_n)_{ij} = X_{ij} \quad \text{for } i \geq j \text{ and} \qquad (6)$$
$$(\mathbf{X}_n)_{ij} = X_{ji} \quad \text{for } i < j.$$

Define

$$\mathbb{X}_n := \frac{1}{n^{1/2}}\mathbf{X}_n, \qquad (7)$$

and set

$$L(A) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{i}\mathbb{E}(X_{ij}^2 I(|X_{ij}| > A)),$$

where A is a positive number.

We shall introduce now a Lindeberg's type condition:

Condition A. (1) $E(X_{\mathbf{u}}) = 0$ *for all* $u \in N^2$.
(2) *There is* $\sigma > 0$ *such that* $\sup_{\mathbf{u}\in\mathbb{N}^2} \|X_{\mathbf{u}}\|_2 \leq \sigma$.
(3) *For every* $\varepsilon > 0$ *we have* $L(\varepsilon n^{1/2}) \to 0$ *as* $n \to \infty$.

Clearly the items (2) and (3) of this condition are satisfied as soon as the family $(X_{\mathbf{u}}^2)$ is uniformly integrable or the random field is stationary.

Next result, in the nonstationary setting, shows that two mild regularity-like conditions without rates, are sufficient for reducing the study of the limiting spectral distribution of a symmetric matrix with independent rows below the diagonal to the corresponding problem for a Gaussian matrix having the same covariance structure. This result indicates that for large matrix size, the empirical distribution of the eigenvalues is universal, in the sense that it is determined only by the covariance structure of the process.

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

Theorem 4. *Assume that Condition A is satisfied and in addition that the random vectors $(R_i)_{i\geq 1}$, where $R_i = (X_{ij})_{j\in\mathbb{N}}$, are mutually independent. For any $i \geq 1$ fixed, let $\mathcal{G}_{ik} = \sigma(X_{ij}, 1 \leq j \leq k)$ and, by convention, for $k \leq 0$, $\mathcal{G}_{ik} = \{\emptyset, \Omega\}$. Then, under the following two additional assumptions:*

$$\eta_m = \sup_{i\geq j\geq m} \|\mathbb{E}(X_{ij}|\mathcal{G}_{i,j-m})\|_2 \to 0 \tag{8}$$

and

$$\gamma_m = \sup_{i\geq \ell\geq k\geq m} \|\mathbb{E}(X_{ik}X_{i\ell}|\mathcal{G}_{i,k-m}) - \mathbb{E}(X_{ik}X_{i\ell})\|_1 \to 0, \tag{9}$$

the following convergence holds: for all $z \in \mathbb{C}^+$,

$$S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z) \to 0 \quad \textit{almost surely, as } n \to \infty, \tag{10}$$

where $\mathbb{X}_n$ is defined by (7) and $\mathbb{Y}_n = \mathbf{Y}_n/\sqrt{n}$, $\mathbf{Y}_n$ being the symmetric matrix defined as in (6) and constructed from a centered real-valued Gaussian random field $(Y_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$ having the same covariance structure as $(X_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$.

Remark 5. Since $\mathbf{Y}_n$ is constructed from a centered real-valued Gaussian random field $(Y_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$ having the same covariance structure as $(X_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$, we have in particular that the random vectors $(G_i)_{i\geq 1}$, where $G_i = (Y_{ij})_{j\in\mathbb{N}}$, are mutually independent. Therefore relation (15) in the proof of Theorem 4 also holds for $\mathbb{Y}_n$. Hence, in addition to the conclusion of Theorem 4, we also have

$$S^{\mathbb{X}_n}(z) - S^{\mathbb{Y}_n}(z) \to 0 \quad \text{almost surely, as } n \to \infty,$$

provided that $(X_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$ and $(Y_\mathbf{u})_{\mathbf{u}\in\mathbb{N}^2}$ are defined on the same probability space.

Remark 6. Theorem 4 also holds if we allow the random variables X_{ij} to depend on the matrix size n. In this context we write $X_{ij}^{(n)}$ instead of X_{ij}, we adapt in a natural way Condition A and we modify conditions (8) and (9) as follows:

$$\sup_{n\geq 1}\sup_{i\geq j\geq m} \|\mathbb{E}(X_{ij}^{(n)}|\mathcal{G}_{i,j-m}^{(n)})\|_2 \to_{m\to\infty} 0$$

and

$$\sup_{n\geq 1}\sup_{i\geq \ell\geq k\geq m} \|\mathbb{E}(X_{ik}^{(n)}X_{i\ell}^{(n)}|\mathcal{G}_{i,k-m}^{(n)}) - \mathbb{E}(X_{ik}^{(n)}X_{i\ell}^{(n)})\|_1 \to_{m\to\infty} 0.$$

Comment 7. Compare to Theorem 5 in [5], we do not assume in our Theorem 4 that the entries are functions of an i.i.d. random field but rather that they satisfy the regularity assumptions (8) and (9). As a counterpart we assume that the rows of $\mathbb{X}_n$ are mutually independent. The importance of this independence assumption appears at least twice in the proof of Theorem 4. Indeed our proof consists first of reducing the almost sure convergence to zero of $S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)$ to the one of $\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)$. The fact that the rows are independent allows to use either the Burkholder–Rosenthal inequality for martingales (see for instance the arguments in the proof on page 34 in [2]) or concentration inequalities based on the Hoeffding–Azuma inequality for martingales (see for instance Theorem 1(ii) of Guntuboyina and Leeb [25]). In case where the entries are functions of an i.i.d. random field, the problem can be reduced first to study symmetric matrices associated with m-dependent random fields and then suitable concentration

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

inequalities can also be obtained. However for general random fields, the situation is not easy to handle. The second step where the independence of the rows plays a crucial role is in the comparison of $\mathbb{E}S^{\mathbb{X}_n}(z)$ with $\mathbb{E}S^{\mathbb{Y}_n}(z)$. To handle this step we use a blockwise Lindeberg-type method applied to martingale approximations of the matrices $\mathbb{X}_n$ and $\mathbb{Y}_n$ (see Step 3 of the proof). In our Lindeberg-type method, because of the martingale structure, the terms of the first order in the Taylor expansion vanish and the terms of the third order are easy to handle. However to deal with the terms of the second order more work is needed and, in order to weaken the dependence, another Taylor expansion is applied. The independence of the rows then plays a crucial role to get the upper bound (46). Without this independence assumption of the rows, we do not think that the conclusion of Theorem 4 holds without imposing rates of convergence to zero on the covariances.

Next corollary applies to stationary sequences and shows that the conclusion of Theorem 4 holds under simple regularity conditions.

Corollary 8. *Let $(X_{ij})_{j\in\mathbb{Z}}$, $i = 1, \dots, n$ be n independent copies of a stationary sequence $(X_k)_{k\in\mathbb{Z}}$ of real-valued random variables which are centered and in $\mathbb{L}^2$. Then the conclusion of Theorem 4 holds under the regularity conditions (3) and (4).*

Theorem 4 and its Remark 6 allow us to formulate the following result for Gram matrices. It will be the key step in the proof of Theorem 1.

Theorem 9. *Under the conditions of Theorem 4 and if $p/N \to c \in (0,\infty)$, the following convergence holds: for all $z \in \mathbb{C}^+$,*

$$S^{\mathbb{B}_N}(z) - \mathbb{E}S^{\mathbb{H}_N}(z) \to 0 \quad \text{almost surely, as } N \to \infty,$$

where $\mathbb{B}_N$ is defined by (2) and $\mathbb{H}_N$ is a Gram random matrix associated with a centered real-valued Gaussian process $(Y_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ having the same covariance structure as $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$.

3. Examples

As we mentioned before, conditions (3) and (4) are satisfied for a stationary sequence if the left tail sigma field $\mathcal{G}_{-\infty}$ is trivial. Processes with trivial tail sigma field are called regular (see Chapter 2, Volume 1 in [15]). It should be noted that Wiener conjectured that a necessary and sufficient condition for a stationary process to be representable as a one-sided function of a sequence of independent, identically distributed random variables is that left tail sigma field be trivial. However this conjecture was proven to be false. See for instance [38].

We give next examples of regular processes.

1. Mixing sequences. The strong mixing coefficient is defined in the following way:

$$\alpha(\mathcal{A},\mathcal{B}) = \sup\{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A},\ B \in \mathcal{B}\},$$

where $\mathcal{A}$ and $\mathcal{B}$ are two sigma algebras.

The ρ-mixing coefficient, also known as maximal coefficient of correlation, is defined as

$$\rho(\mathcal{A},\mathcal{B}) = \sup\{\mathrm{Cov}(X,Y)/\|X\|_2\|Y\|_2 : X \in \mathbb{L}^2(\mathcal{A}),\ Y \in \mathbb{L}^2(\mathcal{B})\}.$$

For the stationary sequence of random variables $(X_k)_{k\in\mathbb{Z}}$, $\mathcal{G}^n$ denotes the σ-field generated by X_i with indices $i \geq n$, and $\mathcal{G}_m$ denotes the σ-field generated by X_i with indices $i \leq m$. Then we define the sequences of mixing coefficients

$$\alpha_n = \alpha(\mathcal{G}_0,\mathcal{G}^n) \quad \text{and} \quad \rho_n = \rho(\mathcal{G}_0,\mathcal{G}^n).$$

A sequence is called strongly mixing if $\alpha_n \to 0$. It is well-known that for strongly mixing sequences the left tail sigma field is trivial; see Claim 2.17a in [15]. Examples of this type include Harris recurrent Markov chains.

If $\lim_{n\to\infty} \rho_n < 1$, then the tail sigma field is also trivial according to Section 2.5 in [14].

Note that our conditions (8) and (9) also hold without the assumptions of stationarity and of regularity. For instance, if

$$\alpha_{2,n} := \sup_{i\geq 1}\sup_{j\geq k} \alpha\big(\sigma(X_{i1}, \ldots, X_{ik}), \sigma(X_{i,k+n}, X_{i,j+n})\big) \to 0,$$

and if the variables are centered and $(X^2_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is uniformly integrable, then (8) and (9) are satisfied. Note that the condition $\alpha_{2,n} \to 0$ is not enough for regularity.

Furthermore, even a weaker degree of dependence for stationary sequences lead to the conditions (8) and (9), namely $\bar{\alpha}_{2,n} \to 0$ where

$$\bar{\alpha}_{2,n} = \sup_{v\geq u\geq n}\ \sup_{x_1,x_2\in\mathbb{R}} \left\| \mathbb{P}(X_u \leq x_1, X_v \leq x_2 | \mathcal{G}_0) - \mathbb{P}(X_u \leq x_1, X_v \leq x_2)\right\|_1 .$$

In case of a homogeneous Markov chain $(X_i)_{i\in\mathbb{Z}}$ with transition operator K and invariant measure ν, $\bar{\alpha}_{2,n} \to 0$ as soon as

$$\sup_{k\geq n}\ \sup_{g\in BV_1} \mathbb{E}_\nu |K^k(g) - \nu(g)| \to 0, \quad \text{as } n \to \infty$$

and there exists a positive constant C such that for any $n > 0$ and any bounded variation function g,

$$\|dK^n(g)\| \leq C_2 \|dg\|.$$

Above BV_1 is the set of bounded variation functions g with $\|dg\| \leq 1$ where $\|dg\|$ is the total variation norm of the measure dg. For instance, if $(X_i)_{i\in\mathbb{Z}}$ is the extended Markov chain whose transition operator K is the Perron–Frobenius operator of the so-called Liverani–Saussol–Vaienti map T_γ of parameter $\gamma \in (0, 1)$ (see [31]), then K satisfies the two above mentioned conditions and we can say that exist two positive constants C and D such that, for any $n > 0$,

$$\frac{D}{n^{(1-\gamma)/\gamma}} \leq \bar{\alpha}_{2,n} \leq \frac{C}{n^{(1-\gamma)/\gamma}}.$$

(See [21] for more details.) Note that this example of Markov chain is known not to be strong mixing.

2. Functions of i.i.d. random variables. Let $(\varepsilon_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ be i.i.d. and $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function such that, for any i, j in $\mathbb{Z}$, $X_{ij} = g(\varepsilon_{i,k}, k \leq j)$ is well defined in $\mathbb{L}^2$ and $\mathbb{E}(X_{ij}) = 0$. These are regular random fields since each row has a trivial left sigma field. Therefore for these processes, conditions (3) and (4) are satisfied. Examples include linear processes, functions of linear processes and iterated random functions (see for instance [47], among others).

For example let $X_{ij} = \sum_{k=0}^{\infty} a_k \varepsilon_{i,j-k}$, where $(\varepsilon_{i,j})$ are i.i.d. with mean 0 and finite variance, and a_k are real coefficients with $\sum_{k=1}^{\infty} a_k^2 < \infty$. In this case X_{ij} is well-defined, the process is regular, and therefore the conclusion of Theorem 1 holds. The limiting empirical eigenvalue distribution of Gram matrices associated with linear processes was investigated in several papers (see for instance [35,48,34,4]) but, all the previous known results treat only the short memory case meaning that the a_k's are absolutely summable.

For a nonstationary example we shall look at a more general linear process, based on martingale difference innovations satisfying Lindeberg's condition.

3. Linear processes with martingale entries. Assume that for any $1 \le j \le i \le n$, the (i,j)th entry of $\mathbf{X}_n$ is given by a linear process of the form

$$X_{ij} = \sum_{\ell=0}^{\infty} a_{i\ell} d_{i,j-\ell}, \tag{11}$$

where $(a_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is a sequence of real numbers and $(d_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ is a sequence of real-valued random variables satisfying the conditions below:

$\mathbf{A}_1$. $A_{n,i} = \sum_{j=0}^{n} a_{ij}^2 < \infty$ is convergent as $n \to \infty$ uniformly in $i \ge 1$.
$\mathbf{A}_2$. There is $\sigma > 0$ such that $\sup_{\mathbf{u}\in\mathbb{Z}^2} \|d_{\mathbf{u}}\|_2 < \sigma$ and for every $\varepsilon > 0$,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{i} \mathbb{E}(d_{ij}^2 I(|d_{ij}| > \varepsilon\sqrt{n})) \to 0 \quad \text{as } n \to \infty.$$

$\mathbf{A}_3$. Setting $\mathcal{F}_{ij} = \sigma(d_{ik},\ k \le j)$, $\mathbb{E}(d_{ij}|\mathcal{F}_{i,j-1}) = 0$ a.s. for any (i,j) in $\mathbb{Z}^2$ and

$$\sup_{i\ge1}\sup_{j\ge n} \|\mathbb{E}(d_{ij}^2|\mathcal{F}_{i,j-n}) - \mathbb{E}(d_{ij}^2)\|_1 \to 0 \quad \text{as } n \to \infty.$$

Corollary 10. *Assume that (X_{ij}) is a linear process as defined in (11) such that the conditions $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ hold. Assume in addition that the random vectors $(d_{i.})_{i\ge1}$, where $d_{i.} = (d_{ij})_{j\in\mathbb{Z}}$, are mutually independent. Then the conclusion of Theorem 4 holds.*

The proof of this corollary is based on standards arguments for martingales and is left to the reader.

4. Proofs

All along the proofs, we shall use the fact that the Stieltjes transform of the spectral measure is a smooth function of the matrix entries. To formalize things in a way that is suitable for our purpose, we shall adopt the same notations as in [18] and introduce the following map A which "constructs" Wigner-type matrices. Let $N = n(n+1)/2$ and write elements of $\mathbb{R}^N$ as $\mathbf{x} = (x_{ij})_{1\le j\le i\le n}$. For any $\mathbf{x}$ in $\mathbb{R}^N$, let $A(\mathbf{x})$ be the matrix defined by

$$(A(\mathbf{x}))_{ij} = \begin{cases} \frac{1}{\sqrt{n}} x_{ij} & i \ge j \\ \frac{1}{\sqrt{n}} x_{ji} & i < j. \end{cases} \tag{12}$$

Let $z \in \mathbb{C}^+$ and $s_n := s_{n,z}$ be the function defined from $\mathbb{R}^N$ to $\mathbb{C}$ by

$$s_n(\mathbf{x}) = \frac{1}{n}\mathrm{Tr}(A(\mathbf{x}) - z\mathbf{I}_n)^{-1}, \tag{13}$$

where $\mathbf{I}_n$ is the identity matrix of order n.

The function s_n, as defined above, admits partial derivatives of all orders that are uniformly bounded. In particular, denoting for any $\mathbf{u} \in \{(i,j)\}_{1\le j\le i\le n}$, $\partial_{\mathbf{u}} s_n$ for $\partial s_n/\partial x_{\mathbf{u}}$, it follows

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

from [18] that the following upper bounds hold: for any $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\{(i,j)\}_{1\le j\le i\le n}$, there exist universal positive constants c_1, c_2 and c_3 depending only on the imaginary part of z such that

$$|\partial_{\mathbf{u}} s_n| \le \frac{c_1}{n^{3/2}}, \qquad |\partial_{\mathbf{u}}\partial_{\mathbf{v}} s_n| \le \frac{c_2}{n^2} \quad \text{and} \quad |\partial_{\mathbf{u}}\partial_{\mathbf{v}}\partial_{\mathbf{w}} s_n| \le \frac{c_3}{n^{5/2}}. \tag{14}$$

4.1. Proof of Theorem 4

The proof of this theorem requires several steps. First we reduce the problem to studying the difference of expected values. Next, in order to weaken the dependence, we partition the variables in each row in big and small blocks. The big blocks are approximated by vector-valued martingale differences. Then, we replace one by one these martingale differences by Gaussian random vectors having the same covariance structure with the help of a blockwise Lindeberg-type method.

All along the proof $z = x + \mathrm{i}y$ will be a complex number in $\mathbb{C}^+$. Also, the notation $a \ll b$ means that there is a constant C depending only on $\operatorname{Im} z = y$ such that $a \le Cb$.

Step 1: Reduction of the problem to a difference of expected values.

Since the random vectors $(R_i)_{1\le i\le n}$, where $R_i = (X_{ij})_{1\le j\le i}$, are mutually independent, it is well-known (see for instance the arguments in the proof on page 34 in [2]) that

$$S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{X}_n}(z) \to 0 \quad \text{a.s.} \tag{15}$$

Hence, in order to prove Theorem 4, it suffices to show that

$$\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z) \to 0. \tag{16}$$

To prove the above convergence, notice that there is no loss of generality in assuming that the entries $(Y_{\mathbf{u}})$ of $\mathbf{Y}_n$ are independent of the entries $(X_{\mathbf{u}})$ of $\mathbf{X}_n$. Therefore, from now on, we assume that $\mathbf{Y}_n$ is a symmetric matrix constructed from a real-valued centered Gaussian random field $(Y_{\mathbf{u}})$ having the same covariance structure as $(X_{\mathbf{u}})$ and independent of $(X_{\mathbf{u}})$.

We write $S^{\mathbb{X}_n}(z)$ and $S^{\mathbb{Y}_n}(z)$ as a function of the entries on and below the diagonal, arranged row after row. More exactly, using the notation (13), we write

$$S^{\mathbb{X}_n}(z) = s_n\big(L^X\big) \quad \text{and} \quad S^{\mathbb{Y}_n}(z) = s_n\big(L^Y\big),$$

where $L^X = (L_i^X)_{1\le i\le n}$ and $L^Y = (L_i^Y)_{1\le i\le n}$ with $L_i^X = (X_{i1},\dots,X_{ii})$ and $L_i^Y = (Y_{i1},\dots,Y_{ii})$. Also, in the sequel, to further simplify the notation we shall skip the index n from s_n and we put $s = s_n := s_{n,z}$.

Step 2: Martingale approximation.

We shall introduce a martingale structure on each row. We start from the celebrated Bernstein big and small blocks argument which weakens the dependence. We partition the variables in each row in big and small blocks and show that the variables in large blocks have a dominant contribution. The large blocks are then decomposed in martingale differences and a rest which also has a smaller contribution.

Let p and q be two integers fixed for the moment. Fix i in $\{1,\dots,n\}$ and let $k_i = [i/(p+q)]$. We partition the set $\{1,\dots,i\}$ in big and small blocks with the following restriction: a big block of size p is followed by a small block of size q. We shall have the set of indexes $I_{i1}, J_{i1}, I_{i2}, J_{i2}, \dots, I_{ik_i}, J_{ik_i} J_{i,k_i+1}$ where each index set I_{ij} is of size p, each index set J_{ij}

is of size q and the last block has a size at most $p+q$. More precisely, for any i in $\{1,\dots,n\}$ and for any $j \in \{1,\dots,k_i\}$,

$$I'_j = \{(j-1)(p+q)+1,\dots,(j-1)(p+q)+p\} \quad \text{and}$$
$$J'_j = \{(j-1)(p+q)+p+1,\dots,j(p+q)\}$$

and

$$I_{ij} = \{(i,k); k \in I'_j\} \quad \text{and} \quad J_{ij} = \{(i,k); k \in J'_j\}.$$

Corresponding to this index decomposition, the vectors L_j^X and L_j^Y are partitioned in k_i+1 consecutive vectors. Setting

$$B_{ij} = (X_{\mathbf{u}})_{\mathbf{u}\in I_{ij}}, \qquad b_{ij} = (X_{\mathbf{u}})_{\mathbf{u}\in J_{ij}}, \qquad B^*_{ij} = (Y_{\mathbf{u}})_{\mathbf{u}\in I_{ij}} \quad \text{and} \quad b^*_{ij} = (Y_{\mathbf{u}})_{\mathbf{u}\in J_{ij}}$$

we write

$$L_i^X = (B_{i1}, b_{i1}, B_{i2}, b_{i2}, \dots, B_{ik_i}, b_{ik_i}, b_{i,k_i+1}) \quad \text{and}$$
$$L_i^Y = (B^*_{i1}, b^*_{i1}, B^*_{i2}, b^*_{i2}, \dots, B^*_{ik_i}, b^*_{ik_i}, b^*_{i,k_i+1}).$$

We introduce now the following vectors

$$B_i^X = (B_{i1}, \mathbf{0}_q, B_{i2}, \mathbf{0}_q, \dots, B_{ik_i}, \mathbf{0}_q, \mathbf{0}_r) \quad \text{and}$$
$$B_i^Y = (B^*_{i1}, \mathbf{0}_q, B^*_{i2}, \mathbf{0}_q, \dots, B^*_{ik_i}, \mathbf{0}_q, \mathbf{0}_r),$$

where $r = i - k_i(p+q)$. Note that B_i^X (resp. B_i^Y) is derived from L_i^X (resp. L_i^Y) where we replace the variables in b_{ij} (resp. b^*_{ij}) by 0's. In addition, for A a positive real, fixed for the moment, we set for any $\mathbf{u} \in \mathbb{Z}^2$

$$\widetilde{X}_{\mathbf{u}} := X_{\mathbf{u}} I(|X_{\mathbf{u}}| \le A),$$

and, for any $i \in \{1,\dots,n\}$,

$$\widetilde{B}_i^X = (\widetilde{B}_{i1}, \mathbf{0}_q, \widetilde{B}_{i2}, \mathbf{0}_q, \dots, \widetilde{B}_{ik_i}, \mathbf{0}_q, \mathbf{0}_r) \quad \text{where } \widetilde{B}_{ij} = (\widetilde{X}_{\mathbf{u}})_{\mathbf{u}\in I_{ij}} \text{ for } j \in \{1,\dots,k_i\}.$$

Next, for any $i \in \{1,\dots,n\}$, we consider the sigma algebras $\mathcal{F}_{i0}^X = \mathcal{F}_{i0}^Y = \{\emptyset, \Omega\}$ and for $1 \le \ell \le k_i$, $\mathcal{F}_{i\ell}^X = \sigma(B_{ij}; 1 \le j \le \ell)$ and $\mathcal{F}_{i\ell}^Y = \sigma(B^*_{ij}; 1 \le j \le \ell)$. Then, for any $\ell \in \{1,\dots,k_i\}$, we define

$$\widetilde{D}_{i\ell} = \widetilde{B}_{i\ell} - \mathbb{E}(\widetilde{B}_{i\ell} | \mathcal{F}_{i,\ell-1}^X), \tag{17}$$

and

$$D^*_{i\ell} = B^*_{i\ell} - \mathbb{E}(B^*_{i\ell} | \mathcal{F}_{i,\ell-1}^Y). \tag{18}$$

By $\mathbb{E}(\widetilde{B}_{i\ell}|\mathcal{F}_{i,\ell-1}^X)$ (resp. $\mathbb{E}(B^*_{i\ell}|\mathcal{F}_{i,\ell-1}^Y)$) we understand a vector of dimension p where each component is a component of the vector $\widetilde{B}_{i\ell}$ (resp. $B^*_{i\ell}$) conditioned with respect to $\mathcal{F}_{i,\ell-1}^X$ (resp. $\mathcal{F}_{i,\ell-1}^Y$). Note that $(\widetilde{D}_{i\ell})_{1\le\ell\le k_i}$ and $(D^*_{i\ell})_{1\le\ell\le k_i}$ are vector valued martingale differences adapted respectively to $(\mathcal{F}_{i\ell}^X)_{1\le\ell\le k_i}$ and $(\mathcal{F}_{i\ell}^Y)_{1\le\ell\le k_i}$. We then define the vectors $\widetilde{D}_i^X$ and D_i^Y

with dimension i and with a similar structure as B_i^X as follows:

$$\widetilde{D}_i^X = (\widetilde{D}_{i1}, \mathbf{0}_q, \widetilde{D}_{i2}, \mathbf{0}_q, \dots, \widetilde{D}_{ik_i}, \mathbf{0}_q, \mathbf{0}_r) \quad \text{and}$$
$$D_i^Y = (D_{i1}^*, \mathbf{0}_q, D_{i2}^*, \mathbf{0}_q, \dots, D_{ik_i}^*, \mathbf{0}_q, \mathbf{0}_r). \tag{19}$$

Setting $\widetilde{D}^X = (\widetilde{D}_i^X)_{1\le i\le n}$, we first compare $\mathbb{E}s(L^X)$ to $\mathbb{E}s(\widetilde{D}^X)$. We write

$$\mathbb{E}s(L^X) - \mathbb{E}s(\widetilde{D}^X) = \mathbb{E}\Delta_1(s) + \mathbb{E}\Delta_2(s) + \mathbb{E}\Delta_3(s),$$

where

$$\Delta_1(s) = s(L^X) - s(B^X), \qquad \Delta_2(s) = s(B^X) - s(\widetilde{B}^X)$$

and

$$\Delta_3(s) = s(\widetilde{B}^X) - s(\widetilde{D}^X),$$

with the notations $B^X = (B_i^X)_{1\le i\le n}$ and $\widetilde{B}^X = (\widetilde{B}_i^X)_{1\le i\le n}$. To control each of the $\mathbb{E}\Delta_i(s)$ for $i = 1, 2, 3$, we apply Lemma 11. Therefore, we get

$$|\mathbb{E}\Delta_1(s)|^2 \ll \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^{k_i+1} \sum_{\mathbf{u}\in J_{ij}} \mathbb{E}(X_{\mathbf{u}}^2) \ll \Big(\frac{q}{p} + \frac{q+p}{n}\Big)\sigma^2,$$

$$|\mathbb{E}\Delta_2(s)|^2 \ll \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^{k_i} \sum_{\mathbf{u}\in I_{ij}} \mathbb{E}(X_{\mathbf{u}}^2 I(|X_{\mathbf{u}}| > A)) \ll L(A),$$

and

$$|\mathbb{E}\Delta_3(s)|^2 \ll \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^{k_i} \sum_{\mathbf{u}\in I_{ij}} \|\mathbb{E}(\widetilde{X}_{\mathbf{u}}|\mathcal{F}_{i,j-1}^X)\|_2^2 \le 2\Big(L(A) + \max_{1\le j\le i\le n} \|\mathbb{E}(X_{ij}|\mathcal{G}_{i,j-q})\|_2^2\Big)$$
$$\ll \big(L(A) + \eta_q^2\big).$$

We proceed in a similar way with the matrix $\mathbb{Y}_n$. Therefore, setting $D^Y = (D_i^Y)_{1\le i\le n}$, we write

$$\mathbb{E}s(L^Y) - \mathbb{E}s(D^Y) = \mathbb{E}\Delta_1'(s) + \mathbb{E}\Delta_2'(s),$$

with the notations

$$\Delta_1'(s) = s(L^Y) - s(B^Y) \quad \text{and} \quad \Delta_2'(s) = s(B^Y) - s(D^Y),$$

where $B^Y = (B_i^Y)_{1\le i\le n}$. Applying Lemma 11 and using the fact that $(Y_{\mathbf{u}})$ has the same covariance structure as $(X_{\mathbf{u}})$, we derive

$$|\mathbb{E}\Delta_1'(s_n)|^2 \ll \Big(\frac{q}{p} + \frac{q+p}{n}\Big)\sup \mathbb{E}(Y_{\mathbf{u}}^2) \ll \Big(\frac{q}{p} + \frac{q+p}{n}\Big)\sigma^2.$$

On another hand, Lemmas 11 and 13 imply that

$$|\mathbb{E}\Delta_2'(s)|^2 \ll \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^{k_i} \sum_{\mathbf{u}\in I_{ij}} \|\mathbb{E}(Y_{\mathbf{u}}|\mathcal{F}_{i,j-1}^Y)\|_2^2 \ll \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^{k_i} \sum_{\mathbf{u}\in I_{ij}} \|\mathbb{E}(X_{\mathbf{u}}|\mathcal{F}_{i,j-1}^X)\|_2^2$$
$$\ll \max_{1\le j\le i\le n} \|\mathbb{E}(X_{ij}|\mathcal{G}_{i,j-q})\|_2^2 \ll \eta_q^2.$$

Overall we have the decomposition

$$\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z) = \mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y) + E_n(p,q,A), \tag{20}$$

with

$$|E_n(p,q,A)| \ll \Big(\Big(\frac{q}{p} + \frac{q+p}{n}\Big)^{1/2}\sigma + L^{1/2}(A) + \eta_q\Big).$$

Step 3: The study of $\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y)$.

The aim of this step is to prove the following upper bound:

$$\big|\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y)\big| \ll \frac{1}{n^{1/2}} p^2\sigma^2(A + aA + \sigma) + pL(A) + \eta_q^2 + p\gamma_{aq}. \tag{21}$$

With this aim, we shall proceed as in the Lindeberg method, that is first write $\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y)$ as telescoping sums and then treat each term in the sums with the help of the Taylor expansion.

Step 3.1. Writing the difference as telescoping sums. To study $\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y)$ we first decompose the difference according to the rows and after that we study the rows separately. With this aim we introduce a telescoping sum where each term is a difference of two functions whose arguments differ only by one row. Namely we write

$$\begin{aligned}&\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y)\\ &\quad = \sum_{i=1}^{n}\Big(\mathbb{E}s\big(\widetilde{D}^X_{[1,i-1]}, \widetilde{D}^X_i, D^Y_{[i+1,n]}\big) - \mathbb{E}s\big(\widetilde{D}^X_{[1,i-1]}, D^Y_i, D^Y_{[i+1,n]}\big)\Big),\end{aligned} \tag{22}$$

where $\widetilde{D}^X_{[a,b]} = (\widetilde{D}^X_a, \dots \widetilde{D}^X_b)$ and $D^Y_{[a,b]} = (D^Y_a, \dots D^Y_b)$ with $\widetilde{D}^X_i$ and D^Y_i defined in (19). Now for every i fixed denote by

$$s_i(\mathbf{x}) := s\big(\widetilde{D}^X_{[1,i-1]}, \mathbf{x}, D^Y_{[i+1,n]}\big).$$

Note that s_i is a random function from $\mathbb{R}^i$ to $\mathbb{C}$. With this notation

$$\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y) = \sum_{i=1}^{n}\mathbb{E}(s_i(\widetilde{D}^X_i) - s_i(D^Y_i)).$$

From now on, for easier notation, it will be convenient to extend the vectors $(\widetilde{D}_{i\ell})_{1\le\ell\le k_i}$ and $(D^*_{i\ell})_{1\le\ell\le k_i}$ defined in (17) and (18) as follows:

$$\widetilde{D}'_{i\ell} = (\widetilde{D}_{i\ell}, \mathbf{0}_q) \quad \text{and} \quad D'^*_{i\ell} = (\widetilde{D}^*_{i\ell}, \mathbf{0}_q) \quad \text{for } 1\le\ell\le k_i - 1 \tag{23}$$

and

$$\widetilde{D}'_{ik_i} = (\widetilde{D}_{ik_i}, \mathbf{0}_{q+r}) \quad \text{and} \quad D'^*_{ik_i} = (D^*_{ik_i}, \mathbf{0}_{q+r}). \tag{24}$$

With these notations, as in the Lindeberg's method, we write now another telescoping sum where we change one by one the vectors $\widetilde{D}'_{i\ell}$ by $D'^*_{i\ell}$ in the argument of s_i. With this aim we write Q4

$$\begin{aligned}s_i(\widetilde{D}_i^X) - s_i(D_i^Y) &= s_i(\widetilde{D}'_{i1}, \dots, \widetilde{D}'_{ik_i}) - s_i(D'^*_{i1}, \dots, D'^*_{ik_i})\\&= \sum_{u=1}^{k_i}\Big(s_i(\widetilde{D}'_{i,[1,u-1]}, \widetilde{D}'_{iu}, D'^*_{i,[u+1,k_i]}) - s_i(\widetilde{D}'_{i,[1,u-1]}, D'^*_{iu}, D'^*_{i,[u+1,k_i]})\Big)\\&:= \sum_{u=1}^{k_i}\Big(s_{i,u}(\widetilde{D}'_{iu}) - s_{i,u}(D'^*_{iu})\Big),\end{aligned} \tag{25}$$

where $\widetilde{D}'_{i,[k,\ell]} := (\widetilde{D}'_{ik}, \dots, \widetilde{D}'_{i\ell})$ and $D'^*_{i,[k,\ell]} := (D'^*_{ik}, \dots, D'^*_{i\ell})$. Note that the $s_{i,u}$'s defined above are random functions from $\mathbb{R}^{p+q}$ to $\mathbb{C}$ if $1 \le u \le k_i - 1$ and from $\mathbb{R}^{p+q+r}$ to $\mathbb{C}$ if $u = k_i$ (where $r = i - k_i(p+q)$).

Hence starting from (22) and taking into account (25), we derive

$$\mathbb{E}s(\widetilde{D}^X) - \mathbb{E}s(D^Y) = \sum_{i=1}^{n}\sum_{u=1}^{k_i}\Big(\mathbb{E}s_{i,u}(\widetilde{D}'_{iu}) - \mathbb{E}s_{i,u}(D'^*_{iu})\Big). \tag{26}$$

Step 3.2. Taylor expansion of each term in the double sum (26). In this step, we shall treat separately each term in the double sum (26) corresponding to the ith row by using a Taylor expansion. So, in the following, i is fixed. Let us first write

$$s_{i,u}(\widetilde{D}'_{iu}) - s_{i,u}(D'^*_{iu}) = s_{i,u}(\widetilde{D}'_{iu}) - s_{i,u}(\mathbf{0}) + s_{i,u}(\mathbf{0}) - s_{i,u}(D'^*_{iu}),$$

and make a Taylor expansion of order three of both $s_{i,u}(\widetilde{D}'_{iu}) - s_{i,u}(\mathbf{0})$ and $s_{i,u}(D'^*_{iu}) - s_{i,u}(\mathbf{0})$. As we shall see the expectations of the terms of the first order will be zero whereas the expectations of the terms of the third order will be easy to handle. This will lead to the following upper bound: for any $i \in \{1, \dots, n\}$ and any $u \in \{1, \dots, k_i\}$,

$$\begin{aligned}&\Big|\mathbb{E}s_{i,u}(\widetilde{D}'_{iu}) - \mathbb{E}s_{i,u}(D'^*_{iu}) - \sum_{j,\ell=1}^{p}\mathbb{E}\Big(\big(\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}^X_{i,u-1}) - \mathbb{E}(g_{iu}^{(j)}g_{iu}^{(\ell)})\big)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\Big)\Big|\\&\quad\ll \frac{1}{n^{5/2}}p^3\sigma^2(A+\sigma),\end{aligned} \tag{27}$$

where $d_{iu}^{(j)}$ is the jth component of the vector $\widetilde{D}'_{iu}$ and $g_{iu}^{(j)}$ the jth component of the vector D'^*_{iu}. The next step will consist of proving that for any $i \in \{1, \dots, n\}$,

$$\begin{aligned}&\Big|\sum_{u=1}^{k_i}\sum_{j,\ell=1}^{p}\mathbb{E}\Big(\big(\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}^X_{i,u-1}) - \mathbb{E}(g_{iu}^{(j)}g_{iu}^{(\ell)})\big)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\Big)\Big| \ll \frac{1}{n}\eta_q^2 + \frac{1}{n^{3/2}}(Aap^2)\sigma^2\\&\quad+ \frac{p}{n^2}\sum_{j=1}^{i}\|X_{ij}^2 I(|X_{ij}| > A)\|_1 + \frac{1}{n^2}k_i p^2\gamma_{aq}.\end{aligned} \tag{28}$$

Starting from (26) and taking into account (27) and (28), the upper bound (21) will follow.

The rest of this step consists of proving (27) and (28). Using Taylor's expansion of order three, we get

$$s_{i,u}(\widetilde{D}'_{iu}) - s_{i,u}(\mathbf{0}) = \widetilde{R}_1 + \widetilde{R}_2 + \widetilde{R}_3, \tag{29}$$

where

$$\widetilde{R}_1 = \sum_{j=1}^{p} d_{iu}^{(j)} \partial_j s_{i,u}(\mathbf{0}), \qquad \widetilde{R}_2 = \frac{1}{2}\Big(\sum_{j=1}^{p} d_{iu}^{(j)} \partial_j\Big)^2 s_{i,u}(\mathbf{0})$$

and

$$\widetilde{R}_3 = \frac{1}{6}\Big(\sum_{j=1}^{p} d_{iu}^{(j)} \partial_j\Big)^3 s_{i,u}(\theta \widetilde{D}'_{iu}) \quad \text{with } \theta \in]0,1[.$$

Similarly, we get

$$s_{i,u}(D'^*_{iu}) - s_{i,u}(\mathbf{0}) = R_1^* + R_2^* + R_3^*, \tag{30}$$

where

$$R_1^* = \sum_{j=1}^{p} g_{iu}^{(j)} \partial_j s_{i,u}(\mathbf{0}) \quad \text{and} \quad R_2^* = \frac{1}{2}\Big(\sum_{j=1}^{p} g_{iu}^{(j)} \partial_j\Big)^2 s_{i,u}(\mathbf{0})$$

and

$$R_3^* = \frac{1}{6}\Big(\sum_{j=1}^{p} g_{iu}^{(j)} \partial_j\Big)^3 s_{i,u}(\theta D'^*_{iu}) \quad \text{with } \theta \in]0,1[.$$

Now notice that, for any $u \in \{1, \dots, k_i\}$ and any $j \in \{1, \dots, p\}$,

$$d_{iu}^{(j)} = \widetilde{X}_{i,(u-1)(p+q)+j} - \mathbb{E}(\widetilde{X}_{i,(u-1)(p+q)+j} | \mathcal{F}^X_{i,u-1}) := \widetilde{X}_{iu}^{(j)} - \mathbb{E}(\widetilde{X}_{iu}^{(j)} | \mathcal{F}^X_{i,u-1}), \tag{31}$$

and

$$g_{iu}^{(j)} = Y_{i,(u-1)(p+q)+j} - \mathbb{E}(Y_{i,(u-1)(p+q)+j} | \mathcal{F}^Y_{i,u-1}) := Y_{iu}^{(j)} - \mathbb{E}(Y_{iu}^{(j)} | \mathcal{F}^Y_{i,u-1}). \tag{32}$$

Therefore

$$\|d_{iu}^{(j)}\|_3^3 \le 2^3 \|\widetilde{X}_{iu}^{(j)}\|_3^3 \ll A\sigma^2,$$

and since $\mathbf{Y}_n$ has the same covariance structure as $\mathbf{X}_n$ and is a Gaussian vector,

$$\|g_{iu}^{(j)}\|_3^3 \le 2^3 \|Y_{iu}^{(j)}\|_3^3 \le 2^4 \|Y_{iu}^{(j)}\|_2^3 \ll \sigma^3.$$

Taking into account the two previous inequalities and the upper bound on the partial derivatives of order three of s given in (14), we infer that

$$|\mathbb{E}(\widetilde{R}_3) + \mathbb{E}(R_3^*)| \ll \frac{1}{n^{5/2}} p^3 \sigma^2 (A + \sigma). \tag{33}$$

On another hand, we notice that for any j, ℓ in $\{1, \dots, p\}$, $\partial_j s_{i,u}(\mathbf{0})$ and $\partial_j \partial_\ell s_{i,u}(\mathbf{0})$ are complex-valued random variables measurable with respect to the sigma algebra $\mathcal{H}_{i,u}$ defined by

$$\mathcal{H}_{i,u} = \mathcal{F}^X_{i,u-1} \vee \sigma\big((L_j^X)_{1\le j\le i-1}, (L_k^Y)_{i+1\le k\le n}\big) \vee \sigma\big(D^*_{i,u+1}, \dots, D^*_{ik_i}\big). \tag{34}$$

Hence

$$\mathbb{E}(\widetilde{R}_1) = \sum_{j=1}^{p} \mathbb{E}\big(\partial_j s_{i,u}(\mathbf{0}) \mathbb{E}(d_{iu}^{(j)} | \mathcal{H}_{i,u})\big),$$

and

$$\mathbb{E}(\widetilde{R}_2) = \frac{1}{2}\sum_{j,\ell=1}^{p} \mathbb{E}\big(\partial_j\partial_\ell s_{i,u}(\mathbf{0})\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{H}_{i,u})\big).$$

Since the rows of $\mathbf{X}_n$ are assumed to be independent and $\mathbf{Y}_n$ is assumed to be independent of $\mathbf{X}_n$, then $\sigma(d_{iu}^{(1)},\dots,d_{iu}^{(p)}) \vee \mathcal{F}_{i,u-1}^X$ is independent of

$$\sigma\big((L_j^X)_{1\le j\le i-1},(L_k^Y)_{i+1\le k\le n}\big)\vee\sigma\big(D_{i,u+1}^*,\dots,D_{ik_i}^*\big).$$

Therefore, by the properties of the conditional expectation, $\mathbb{E}(d_{iu}^{(j)}|\mathcal{H}_{i,u}) = \mathbb{E}(d_{iu}^{(j)}|\mathcal{F}_{i,u-1}^X) = 0$ and $\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{H}_{i,u}) = \mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X)$. Hence,

$$\mathbb{E}(\widetilde{R}_1) = 0 \quad\text{and}\quad \mathbb{E}(\widetilde{R}_2) = \frac{1}{2}\sum_{j,\ell=1}^{p}\mathbb{E}\big(\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\big). \tag{35}$$

We handle now the terms $\mathbb{E}(R_1^*)$ and $\mathbb{E}(R_2^*)$. With this aim we notice that by definition $(D_{iu}^* : 1 \le u \le k_i)_{1\le i\le n}$ is a centered Gaussian vector such that $\mathrm{Cov}(D_{iu}^*, D_{i'u'}^*) = \mathbf{0}_{p,p}$ if $(i,u) \neq (i',u')$. Therefore $D_{i,u}^*, i = 1,\dots,n, u = 1,\dots,k_i$ are centered Gaussian random variables in $\mathbb{R}^p$ which are mutually independent. In addition they are independent of $(X_{\mathbf{u}})$. Therefore,

$$\mathbb{E}(R_1^*) = \sum_{j=1}^{p}\mathbb{E}(g_{iu}^{(j)})\mathbb{E}\big(\partial_j s_{i,u}(\mathbf{0})\big) = 0, \tag{36}$$

and

$$\mathbb{E}(R_2^*) = \frac{1}{2}\sum_{j,\ell=1}^{p}\mathbb{E}(g_{iu}^{(j)}g_{iu}^{(\ell)})\mathbb{E}\big(\partial_j\partial_\ell s_{i,u}(\mathbf{0})\big). \tag{37}$$

So, taking into account (29), (30), (33) and (35)–(37), the upper bound (27) follows.

We turn now to the proof of (28). Recalling the notations (31) and (32), we first write

$$\begin{aligned}\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X) - \mathbb{E}(g_{iu}^{(j)}g_{iu}^{(\ell)}) &= \mathbb{E}(\widetilde{X}_{iu}^{(j)}\widetilde{X}_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X) - \mathbb{E}(Y_{iu}^{(j)}Y_{iu}^{(\ell)})\\ &\quad - \mathbb{E}(\widetilde{X}_{iu}^{(j)}|\mathcal{F}_{i,u-1}^X)\mathbb{E}(\widetilde{X}_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X) + \mathbb{E}\big(\mathbb{E}(Y_{iu}^{(j)}|\mathcal{F}_{i,u-1}^Y)\mathbb{E}(Y_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^Y)\big).\end{aligned}$$

Therefore, by the triangle inequality and Jensen's inequality,

$$\begin{aligned}&\Big|\sum_{u=1}^{k_i}\sum_{j,\ell=1}^{p}\mathbb{E}\Big(\big(\mathbb{E}(d_{iu}^{(j)}d_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X) - \mathbb{E}(g_{iu}^{(j)}g_{iu}^{(\ell)})\big)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\Big)\Big|\\ &\quad\le \sum_{u=1}^{k_i}\sum_{j,\ell=1}^{p}\Big|\mathbb{E}\Big(\big(\mathbb{E}(\widetilde{X}_{iu}^{(j)}\widetilde{X}_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X) - \mathbb{E}(Y_{iu}^{(j)}Y_{iu}^{(\ell)})\big)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\Big)\Big|\\ &\qquad + \sum_{u=1}^{k_i}\mathbb{E}\Big|\sum_{j,\ell=1}^{p}\mathbb{E}(\widetilde{X}_{iu}^{(j)}|\mathcal{F}_{i,u-1}^X)\mathbb{E}(\widetilde{X}_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^X)\partial_j\partial_\ell s_{i,u}(\mathbf{0})\Big|\\ &\qquad + \sum_{u=1}^{k_i}\Big|\sum_{j,\ell=1}^{p}\mathbb{E}\big(\mathbb{E}(Y_{iu}^{(j)}|\mathcal{F}_{i,u-1}^Y)\mathbb{E}(Y_{iu}^{(\ell)}|\mathcal{F}_{i,u-1}^Y)\big)\mathbb{E}\big(\partial_j\partial_\ell s_{i,u}(\mathbf{0})\big)\Big|\\ &\quad := T_1 + T_2 + T_3.\end{aligned} \tag{38}$$

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

As we shall see the terms T_2 and T_3 will be handled by using Lemma 12 whereas to give a suitable upper bound for the term T_1 we will use an additional Taylor expansion.

Let us first handle T_3. Recalling the notation (23) and (24) and setting

$$C_{i,u} = \big(\widetilde{D}^X_{[1,i-1]}, \widetilde{D}'_{i1}, \ldots, \widetilde{D}'_{i,u-1}, \mathbf{0}, D'^*_{i,u+1}, \ldots, D'^*_{i,k_i}, D^Y_{[i+1,n]}\big), \tag{39}$$

we note that $\mathbb{E}(Y^{(j)}_{iu}|\mathcal{F}^Y_{i,u-1})\mathbb{E}(Y^{(\ell)}_{iu}|\mathcal{F}^Y_{i,u-1})$ is independent of $\partial_j\partial_\ell s_{i,u}(C_{i,u})$. This is because of the independence between $\mathbf{Y}_n$ and $\mathbf{X}_n$ together with the independence between the vectors $(\mathbb{E}(Y^{(j)}_{iu}|\mathcal{F}^Y_{i,u-1}), \mathbb{E}(Y^{(\ell)}_{iu}|\mathcal{F}^Y_{i,u-1}))$ and $(D'^*_{i,u+1}, \ldots, D'^*_{i,u_{k_i}}, D^Y_{[i+1,n]})$. To prove the latter independence, it suffices to notice that $(\mathbb{E}(Y^{(j)}_{iu}|\mathcal{F}^Y_{i,u-1}), \mathbb{E}(Y^{(\ell)}_{iu}|\mathcal{F}^Y_{i,u-1}), D'^*_{i,u+1}, \ldots, D'^*_{i,u_{k_i}}, D^Y_{[i+1,n]})$ is a Gaussian vector and that $(\mathbb{E}(Y^{(j)}_{iu}|\mathcal{F}^Y_{i,u-1}), \mathbb{E}(Y^{(\ell)}_{iu}|\mathcal{F}^Y_{i,u-1}))$ and $(D'^*_{i,u+1}, \ldots, D'^*_{i,u_{k_i}}, D^Y_{[i+1,n]})$ are uncorrelated. So, we can bound T_3 as follows:

$$T_3 \le \sum_{u=1}^{k_i} \mathbb{E}\Big| \sum_{j,k\in I'_u} \mathbb{E}(Y_{ij}|\mathcal{F}^Y_{i,u-1})\mathbb{E}(Y_{ik}|\mathcal{F}^Y_{i,u-1})\partial_{ij}\partial_{ik} s(C_{i,u})\Big|,$$

where we recall that $I'_u = \{(u-1)(p+q)+1, \ldots, (u-1)(p+q)+p\}$. An application of Lemma 12 gives

$$\Big|\sum_{j,k\in I'_u} \mathbb{E}(Y_{ij}|\mathcal{F}^Y_{i,u-1})\mathbb{E}(Y_{ik}|\mathcal{F}^Y_{i,u-1})\partial_{ij}\partial_{ik} s(C_{i,u})\Big| \ll \frac{1}{n^2}\sum_{j\in I'_u}\big(\mathbb{E}(Y_{ij}|\mathcal{F}^Y_{i,u-1})\big)^2.$$

Whence, using in addition Lemma 13, we derive

$$T_3 \ll \frac{1}{n^2}\sum_{u=1}^{k_i}\sum_{j\in I'_u} \|\mathbb{E}(Y_{ij}|\mathcal{F}^Y_{i,u-1})\|_2^2 \ll \frac{1}{n^2}\sum_{u=1}^{k_i}\sum_{j\in I'_u} \|\mathbb{E}(X_{ij}|\mathcal{F}^X_{i,u-1})\|_2^2.$$

Since $\mathcal{F}^X_{i,u-1} \subset \mathcal{G}_{i,\ell-q}$ for any $\ell \in \{(u-1)(p+q)+1, \ldots, (u-1)(p+q)+p\}$, it follows that

$$T_3 \ll \frac{1}{n^2}\sum_{j=1}^{i} \|\mathbb{E}(X_{ij}|\mathcal{G}_{i,j-q})\|_2^2 \ll \frac{1}{n}\eta_q^2. \tag{40}$$

To treat T_2 we proceed as in the proof of relation (40), and infer that

$$T_2 \ll \frac{1}{n^2}\sum_{j=1}^{i} \|\mathbb{E}(\widetilde{X}_{ij}|\mathcal{G}_{i,j-q})\|_2^2 \ll \frac{1}{n}\eta_q^2 + \frac{1}{n^2}\sum_{j=1}^{i}\|X_{ij}^2 I(|X_{ij}|>A)\|_1. \tag{41}$$

We handle now the term T_1 in (38). Using the notation (39) and the fact that $\mathbf{Y}_n$ has the same covariance structure as $\mathbf{X}_n$, we start by rewriting T_1 as follows:

$$\begin{aligned} T_1 &= \sum_{u=1}^{k_i}\sum_{j,\ell\in I'_u} \big|\mathbb{E}\big(\big(\mathbb{E}(\widetilde{X}_{ij}\widetilde{X}_{i\ell}|\mathcal{F}^X_{i,u-1}) - \mathbb{E}(X_{ij}X_{i\ell})\big)\partial_{ij}\partial_{i\ell}s(C_{i,u})\big)\big| \\ &= \sum_{u=1}^{k_i}\sum_{j,\ell\in I'_u} \big|\mathbb{E}\big(\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell} - \mathbb{E}(X_{ij}X_{i\ell})\big)\partial_{ij}\partial_{i\ell}s(C_{i,u})\big)\big|, \end{aligned} \tag{42}$$

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

where for the second equality we used the fact that $\partial_{ij}\partial_{i\ell}s(C_{i,u})$ is measurable with respect to $\mathcal{H}_{i,u}$ defined by (34) and that $\sigma\big((X_{i,(u-1)(p+q)+j})_{1\le j\le p}\big)\vee\mathcal{F}^X_{i,u-1}$ is independent of

$$\sigma\big((L^X_j)_{1\le j\le i-1},(L^Y_k)_{i+1\le k\le n}\big)\vee\sigma\big(D^*_{i,u+1},\dots,D^*_{ik_i}\big).$$

To treat the summands in (42), we further weaken the dependence by suppressing some variables in $C_{i,u}$ which are "close" to $\widetilde{X}_{ij}\widetilde{X}_{i\ell}$. Let a be a positive integer fixed for the moment. Then, setting,

$$C^{(a)}_{i,u}=\big(\widetilde{D}^X_{[1,i-1]},\widetilde{D}'_{i1},\dots,\widetilde{D}'_{i,u-a},\mathbf{0},D'^*_{i,u+1},\dots,D'^*_{i,k_i},D^Y_{[i+1,n]}\big)\quad\text{if } u\ge a+1,$$

and

$$C^{(a)}_{i,u}=\big(\widetilde{D}^X_{[1,i-1]},\mathbf{0},D'^*_{i,u+1},\dots,D'^*_{i,k_i},D^Y_{[i+1,n]}\big)\quad\text{if } 1\le u\le a,$$

we write

$$\big|\mathbb{E}\big(\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}-\mathbb{E}(X_{ij}X_{i\ell})\big)\partial_{ij}\partial_{i\ell}s(C_{i,u})\big)\big|\le I_1+I_2 \tag{43}$$

where

$$I_1=\big|\mathbb{E}\big(\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}-\mathbb{E}(X_{ij}X_{i\ell})\big)\partial_{ij}\partial_{i\ell}\big(s(C_{i,u})-s(C^{(a)}_{i,u})\big)\big)\big|$$

and

$$I_2=\big|\mathbb{E}\big(\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}-\mathbb{E}(X_{ij}X_{i\ell})\big)\partial_{ij}\partial_{i\ell}s(C^{(a)}_{i,u})\big)\big|.$$

By using the multivariate Taylor expansion of first order for $\partial_{ij}\partial_{i\ell}s$, taking into account the definitions of $C_{i,u}$ and $C^{(a)}_{i,u}$ and then by using (14), we derive, after simple computations, that for any j and ℓ in I'_u,

$$\begin{aligned}I_1\ll\frac{1}{n^{5/2}}\Big\{&\mathbf{1}_{u\ge a+1}\sum_{v=u-a+1}^{u-1}\sum_{r\in I'_v}\|\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}-\mathbb{E}(X_{ij}X_{i\ell})\big)\big(\widetilde{X}_{ir}-\mathbb{E}(\widetilde{X}_{ir}|\mathcal{F}^X_{i,v-1})\big)\|_1\\&+\mathbf{1}_{u\le a}\sum_{v=1}^{u-1}\sum_{r\in I'_v}\|\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}-\mathbb{E}(X_{ij}X_{i\ell})\big)\big(\widetilde{X}_{ir}-\mathbb{E}(\widetilde{X}_{ir}|\mathcal{F}^X_{i,v-1})\big)\|_1\Big\},\end{aligned}$$

leading to

$$I_1\ll\frac{1}{n^{5/2}}(Aap)\sigma^2. \tag{44}$$

Next, using (14) again and the definition of the conditional expectation, we infer that

$$I_2\ll\frac{1}{n^2}\|\mathbb{E}\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}|\sigma(C^{(a)}_{i,u})\big)-\mathbb{E}(X_{ij}X_{i\ell})\|_1.$$

Notice now that, since $\mathbf{X}_n$ and $\mathbf{Y}_n$ are assumed to be independent and since the rows of $\mathbf{X}_n$ are independent, $\mathbb{E}\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}|\sigma(C^{(a)}_{i,u})\big)=\mathbb{E}\big(\widetilde{X}_{ij}\widetilde{X}_{i\ell}|\mathcal{F}^X_{i,u-a}\big)$. Therefore, after simple computations based on the definition of $\widetilde{X}_{ij}$ and on the fact that $A\|X_{ij}I(|X_{ij}|>A)\|_1\le\|X^2_{ij}I(|X_{ij}|>A)\|_1$, we obtain

$$I_2 \ll \frac{1}{n^2}\|\mathbb{E}(X_{ij}X_{i\ell}|\mathcal{F}^X_{i,u-a}) - \mathbb{E}(X_{ij}X_{i\ell})\|_1 + \frac{1}{n^2}\|X_{ij}I(|X_{ij}|>A)\|_2\|X_{i\ell}I(|X_{i\ell}|>A)\|_2. \tag{45}$$

Starting from (42) and taking into account (43)–(45), we get

$$T_1 \ll \frac{1}{n^{3/2}}(Aap^2)\sigma^2 + \frac{p}{n^2}\sum_{j=1}^{i}\|X^2_{ij}I(|X_{ij}|>A)\|_1 + \frac{1}{n^2}k_i p^2\gamma_{aq}. \tag{46}$$

So, overall, starting from the decomposition (38) and taking into account the upper bounds (40), (41) and (46), the upper bound (28) follows. This ends the proof of Step 3.

Step 4: End of the proof.

Starting from (20), taking $A=\varepsilon\sqrt{n}$ and considering the upper bound (21), we get

$$\left|\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)\right| \ll p^2\sigma^2\left(\varepsilon + a\varepsilon + \frac{1}{n^{1/2}}\sigma\right) + pL(\varepsilon\sqrt{n}) + \eta_q^2 + p\gamma_{aq} + \left(\frac{q}{p}+\frac{q+p}{n}\right)^{1/2}\sigma + L^{1/2}(\varepsilon\sqrt{n}) + \eta_q.$$

Therefore, when $n\to\infty$, we obtain for all p,q,a, and ε,

$$\limsup_{n\to\infty}\left|\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)\right| \ll p^2\sigma^2(\varepsilon+a\varepsilon) + \eta_q^2 + \eta_q + p\gamma_{aq} + (q/p)^{1/2}\sigma.$$

Now we let $\varepsilon\to 0$ and obtain

$$\limsup_{n\to\infty}\left|\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)\right| \ll \eta_q^2 + \eta_q + p\gamma_{aq} + (q/p)^{1/2}\sigma.$$

Then we let $a\to\infty$, and, by our hypotheses, for any p and q we obtain

$$\limsup_{n\to\infty}\left|\mathbb{E}S^{\mathbb{X}_n}(z) - \mathbb{E}S^{\mathbb{Y}_n}(z)\right| \ll \eta_q^2 + \eta_q + (q/p)^{1/2}\sigma.$$

Now we can let p and q tend to ∞ in such a way $q/p\to 0$ to obtain the desired result. ♢

4.2. Proof of Corollary 8

By the reverse martingale convergence theorem and condition (3), we get that $\lim_{n\to\infty}\mathbb{E}(X_0|\mathcal{G}_{-n}) = \mathbb{E}(X_0|\mathcal{G}_{-\infty}) = 0$ a.s. So, since X_0 belongs to $\mathbb{L}^2$, this last convergence implies that condition (8) holds. We prove now that under the conditions of the corollary, condition (9) is satisfied. Note first that, by stationarity, this latter condition reads as

$$\sup_u\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n}) - \mathbb{E}(X_0X_u)\|_1 \to 0 \quad \text{as } n\to\infty. \tag{47}$$

To prove that (47) holds we shall prove that

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sup_{u\geq p+1}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n}) - \mathbb{E}(X_0X_u)\|_1 = 0, \tag{48}$$

and that

$$\lim_{p\to\infty}\limsup_{n\to\infty}\max_{1\leq u\leq p}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n}) - \mathbb{E}(X_0X_u)\|_1 = 0. \tag{49}$$

To prove (48), we note that

$$\begin{aligned}\sup_{u\geq p+1}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n})-\mathbb{E}(X_0X_u)\|_1 &\leq \sup_{u\geq p+1}\|\mathbb{E}(X_0X_u|\mathcal{G}_0)-\mathbb{E}(X_0X_u)\|_1\\ &= \sup_{u\geq p+1}\|X_0\mathbb{E}(X_u|\mathcal{G}_0)-\mathbb{E}(X_0X_u)\|_1\\ &\leq 2\|X_0\|_2\cdot\sup_{u\geq p+1}\|\mathbb{E}(X_u|\mathcal{G}_0)\|_2\leq 2\|X_0\|_2\cdot\|\mathbb{E}(X_0|\mathcal{G}_{-p})\|_2.\end{aligned}$$

This shows that (48) holds since (8) does under (3). We turn now to the proof of (49). By the reverse martingale convergence theorem

$$\begin{aligned}\lim_{n\to\infty}\max_{1\leq u\leq p}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n})-\mathbb{E}(X_0X_u)\|_1 &= \max_{1\leq u\leq p}\lim_{n\to\infty}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-n})-\mathbb{E}(X_0X_u)\|_1\\ &= \sup_{1\leq u\leq p}\|\mathbb{E}(X_0X_u|\mathcal{G}_{-\infty})-\mathbb{E}(X_0X_u)\|_1,\end{aligned}$$

which is equal to zero by condition (4). This ends the proof of (49) and then of the corollary. $\diamondsuit$

4.3. *Proof of Theorem 9*

It is well-known that for deriving the limiting spectral distribution of $\mathbb{B}_N$ it is enough to study the Stieltjes transform of the following symmetric matrix of order $n = N + p$:

$$\mathbb{X}_n = \frac{1}{\sqrt{N}}\begin{pmatrix}\mathbf{0}_{p,p} & \mathcal{X}_{N,p}^T\\ \mathcal{X}_{N,p} & \mathbf{0}_{N,N}\end{pmatrix}.$$

Indeed the eigenvalues of $\mathbb{X}_n^2$ are the eigenvalues of $N^{-1}\mathcal{X}_{N,p}^T\mathcal{X}_{N,p}$ together with the eigenvalues of $N^{-1}\mathcal{X}_{N,p}\mathcal{X}_{N,p}^T$. Since these two latter matrices have the same nonzero eigenvalues, the following relation holds: for any $z\in\mathbb{C}^+$, $S_{\mathbb{B}_N}(z) = z^{-1/2}\frac{n}{2p}S_{\mathbb{X}_n}(z^{1/2}) + \frac{N-p}{2pz}$ (see, for instance, page 549 in [37] for additional arguments leading to the relation above). Obviously a similar equation holds for the Gram random matrix $\mathbb{H}_N$ associated with $(Y_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$, namely: $S_{\mathbb{H}_N}(z) = z^{-1/2}\frac{n}{2p}S_{\mathbb{Y}_n}(z^{1/2}) + \frac{N-p}{2pz}$, where $\mathbb{Y}_n$ is defined as $\mathbb{X}_n$ but with $X_{\mathbf{u}}$ replaced by $Y_{\mathbf{u}}$. Therefore, in order to prove the theorem, it suffices to show that, for any $z\in\mathbb{C}^+$,

$$\lim_{N\to\infty}\left|S_{\mathbb{X}_n}(z)-\mathbb{E}(S_{\mathbb{Y}_n}(z))\right| = 0 \quad \text{a.s.} \tag{50}$$

Note now that $\mathbb{X}_n := n^{-1/2}[x_{ij}^{(n)}]_{i,j=1}^n$ where $x_{ij}^{(n)} = \sqrt{\frac{n}{N}}X_{i-p,j}\mathbf{1}_{i\geq p+1}\mathbf{1}_{1\leq j\leq p}$ if $1\leq j\leq i\leq n$, and $x_{ij}^{(n)} = x_{ji}^{(n)}$ if $1\leq i<j\leq n$. Similarly we can write $\mathbb{Y}_n := n^{-1/2}[y_{ij}^{(n)}]_{i,j=1}^n$ where the $y_{ij}^{(n)}$'s are defined as the $x_{ij}^{(n)}$'s but with $X_{i-p,j}$ replaced by $Y_{i-p,j}$. The theorem then follows by applying Remark 6 of Theorem 4 to the matrices $\mathbb{X}_n$ and $\mathbb{Y}_n$ defined above. $\diamondsuit$

4.4. *Proof of Theorem 1*

According to Theorem B.9. in Bai and Silverstein [2], Theorem 1 will follow if one can prove that Q5

$$\mathbb{P}\big(S^{\mathbb{B}_N}(z)\big)\to S(z), \quad \forall z\in\mathbb{C}^+ = 1, \tag{51}$$

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

where $S(z)$ is the Stieltjes transform of a non-random measure F which is uniquely determined by Eq. (5). On the other hand, by well-known arguments involving Vitali's convergence theorem (see also Lemma 2.14 in [2]), to prove (51), it is enough to prove that for any $z \in \mathbb{C}^+$,

$$\lim_{N\to\infty} S^{\mathbb{B}_N}(z) = S(z) \quad \text{a.s.}$$

According to Theorem 9, this last convergence is reduced to show that, for any $z \in \mathbb{C}^+$

$$\lim_{N\to\infty} \mathbb{E}(S^{\mathbb{H}_N}(z)) = S(z), \tag{52}$$

where $\mathbb{H}_N$ is the Gram matrix associated with a Gaussian random field $(Y_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ having the same covariance structure as $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$.

With this aim, we shall apply Theorem 1.1 in [40]. Consider N independent copies $(g_{ij})_{j\in\mathbb{Z}}$, $i = 1, \dots, N$ of a sequence $(g_k)_{k\in\mathbb{Z}}$ of i.i.d. standard normal random variables. Set

$$\Gamma_p := \begin{pmatrix} c_0 & c_1 & \cdots & c_{p-1} \\ c_1 & c_0 & & c_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p-1} & c_{p-2} & \cdots & c_0 \end{pmatrix} \quad \text{where } c_k = \mathrm{Cov}(X_0, X_k).$$

Using the stationarity of the Gaussian process $(Y_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$, we can easily verify that the random vector $((Y_{1j})_{1\le j\le p}, \dots, (Y_{Nj})_{1\le j\le p})$ has the same distribution as $(\mathbf{g}_1\Gamma_p^{1/2}, \dots, \mathbf{g}_N\Gamma_p^{1/2})$ where for any $i \in \{1, \dots, N\}$, $\mathbf{g}_i = (g_{ij})_{1\le j\le p}$ and $\Gamma_p^{1/2}$ is the symmetric non-negative square root of Γ_p. Therefore, for any $z \in \mathbb{C}^+$,

$$\mathbb{E}(S^{\mathbb{H}_N}(z)) = \mathbb{E}(S^{\Gamma_p^{1/2}\mathbb{G}_N\Gamma_p^{1/2}}(z)), \tag{53}$$

where $\mathbb{G}_N = \frac{1}{N}\mathcal{G}_{N,p}^T\mathcal{G}_{N,p}$ with $\mathcal{G}_{N,p} = (g_{ij})_{1\le i\le N, 1\le j\le p}$.

By a version of the Szegö's theorem for Toeplitz forms generated by real-valued integrable functions (see Theorem 3 in [43] or Theorem 1 in [42]) we obtain that

$$F^{\Gamma_p} \text{ converges to a probability distribution } H \text{ as } p \to \infty, \tag{54}$$

where H is the distribution of $2\pi f(U)$ with U a uniformly distributed random variable on $[-\pi, \pi)$. Therefore, for any φ which is continuous and bounded,

$$\int \varphi(x)dH(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi(2\pi f(\lambda))d\lambda. \tag{55}$$

Now, according to Theorem 1.1 in [40], if $p/N \to c \in (0, \infty)$ there is a nonrandom probability distribution F such that

$$d(F^{\Gamma_p^{1/2}\mathbb{G}_N\Gamma_p^{1/2}}, F) \to 0 \quad \text{a.s.} \tag{56}$$

Furthermore, the Stieltjes transform $S = S(z)$, $z \in \mathbb{C}^+$, of F satisfies the equation

$$S = \int \frac{1}{x(1 - c - czS) - z} dH(x).$$

Setting $\underline{S} := -(1-c)/z + cS$, this last equation becomes

$$z = -\frac{1}{\underline{S}} + c \int \frac{x}{1 + x\underline{S}} dH(x). \tag{57}$$

Let us mention that $\underline{S}$ is also a Stieltjes transform (see relation (1.3) in [40] or [23]), so $\operatorname{Im} \underline{S} > 0$ for $z \in \mathbb{C}^+$.

Combining (56) with (53), it follows that for any $z \in \mathbb{C}^+$, the convergence (52) holds and that S is the Stieltjes transform of a probability distribution that is determined by Eq. (57). Since the function $\varphi(x) := x/(1 + x\underline{S})$ is continuous and bounded by $1/\operatorname{Im} \underline{S}$, the relation (57) can be then rewritten via identity (55) as in Eq. (5). $\diamondsuit$

4.5. Comments on Remark 2

In the proof of Theorem 1 we have shown that the limiting spectral distribution F coincides to that one of the matrices of the form $\Gamma_p^{1/2} \mathbb{G}_N \Gamma_p^{1/2}$ where Γ_p is a symmetric Toeplitz matrix whose entries are the Fourier coefficients of $f(x)$. Also $\mathbb{G}_N$ is the Wishart matrix formed by i.i.d. real-valued standard normal variables. Since the Wishart matrix has eigenvectors that are uniformly distributed with Haar measure, the matrices Γ_p and $\mathbb{G}_N$ are asymptotically free. The limiting distribution F can be expressed by using the notion of free multiplicative convolution $\mu_f \otimes MP$ between μ_f the limiting spectral distribution of Γ_p, which is the distribution of $2\pi f(U)$ where U is a random variable uniformly distributed on $[-\pi, \pi)$, with MP, the Marčenko–Pastur distribution. For further details on the computation of free multiplicative convolution and deconvolution see [44,7,22,36,39,6]. Q6

Also note that F has compact support if and only if μ_f has compact support (see Proposition 3.4 in [19]). By the definition of μ_f this happens if and only if f has compact support.

We also mention that by the Proposition 1 in [48] one can describe precisely the support of the LSD F in case where the spectral density f is associated with a linear process with short memory.

5. Conclusion

The sample covariance matrix based on repeated independent samples from a vector is a consistent estimator of the real covariances when the number of samples is increasing. When the number of dimensions is increasing this is no longer the case. However, if the number of variables in the vector grows proportionally with the number of samples, the limiting spectral distribution (LSD for short) for the sample covariance matrix exists under mild conditions. This always holds if the stationary sequence is regular in the sense described in this paper. For instance the LSD exists if the entries of the process are functions of i.i.d. and even less, if the stationary process has trivial sigma field. This is important because many statistics in multivariate analysis can be expressed as a function of the eigenvalues. Furthermore, the LSD of the sample covariance matrix indicates a strong and fascinating relationship between the LSD and the process' spectral distribution. The LSD has the form of the free convolution of two distributions, one generated by the process' spectral density with Marčenko–Pastur distribution. It is expected that, via free deconvolution methods the process' spectral distribution could be recovered in some cases. For instance, if the spectral density is monotone on $[0, \pi)$ then it can be obtained from the LSD and in this class of spectral densities, the process' spectral density is in one to one correspondence to the LSD.

It will be interesting to investigate whether the LSD exists for any covariance matrix constructed based on a stationary and ergodic sequence and to express it in function of the process' spectral measure.

Uncited references

Q7

[9] and [27].

Acknowledgments

Q8 Q9 This research was supported in part by a Charles Phelps Taft Memorial Fund grant and the NSF grants DMS-1208237 and DMS-1512936. We are indebted to two anonymous referees for many interesting and constructive suggestions. We would also want to thank to Wlodek Bryc for useful discussions.

Appendix

In this section, we collect several results useful for our proofs.

The first result we mention is Lemma 2.1 in [24] that allows to compare the difference between two Stieltjes transforms.

Lemma 11. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric* $n \times n$ *matrices with real entries. Then, for any* $z = x + \mathrm{i}y \in \mathbb{C} \setminus \mathbb{R}$,

$$|S^{\mathbf{A}}(z) - S^{\mathbf{B}}(z)| \le \frac{1}{y^2\sqrt{n}} |\mathrm{Tr}(\mathbf{A} - \mathbf{B})^2|^{1/2}.$$

In addition, concerning the partial derivatives of second order, the following lemma will be also useful. We could not find this lemma in the literature, so we shall provide a short proof.

Lemma 12. *Let* $z \in \mathbb{C}^+$ *and* $s_n := s_{n,z}$ *be defined by* (13). *Let* $(a_{ij})_{1\le j\le i\le n}$ *and* $(b_{ij})_{1\le j\le i\le n}$ *be real numbers. Then, there exists an universal positive constant* c_4 *depending only on the imaginary part of* z *such that for any subset* $\mathcal{I}_n$ *of* $\{(i,j)\}_{1\le j\le i\le n}$ *and any element* $\mathbf{x}$ *of* $\mathbb{R}^N$,

$$\Big| \sum_{\mathbf{u}\in\mathcal{I}_n} \sum_{\mathbf{v}\in\mathcal{I}_n} a_{\mathbf{u}} b_{\mathbf{v}} \partial_{\mathbf{u}} \partial_{\mathbf{v}} s_n(\mathbf{x}) \Big| \le \frac{c_4}{n^2} \Big(\sum_{\mathbf{u}\in\mathcal{I}_n} a_{\mathbf{u}}^2 \sum_{\mathbf{v}\in\mathcal{I}_n} b_{\mathbf{v}}^2 \Big)^{1/2}.$$

Proof. Setting $G = (A(\mathbf{x}) - z\mathbf{I}_n)^{-1}$, we have

$$\partial_{\mathbf{u}} \partial_{\mathbf{v}} s_n = \frac{1}{n} \mathrm{Tr}(G \partial_{\mathbf{u}} A G \partial_{\mathbf{v}} A G) + \frac{1}{n} \mathrm{Tr}(G \partial_{\mathbf{v}} A G \partial_{\mathbf{u}} A G).$$

(See the equality (20) in [18].) Whence, with the notations

$$\widetilde{A} := \sum_{\mathbf{u}\in\mathcal{I}_n} a_{\mathbf{u}} \partial_{\mathbf{u}} A \quad \text{and} \quad \widetilde{B} := \sum_{\mathbf{u}\in\mathcal{I}_n} b_{\mathbf{u}} \partial_{\mathbf{u}} A,$$

it follows that

$$\sum_{\mathbf{u}\in\mathcal{I}_n} \sum_{\mathbf{v}\in\mathcal{I}_n} a_{\mathbf{u}} b_{\mathbf{v}} \partial_{\mathbf{u}} \partial_{\mathbf{v}} s_n = \frac{1}{n} \mathrm{Tr}(G^2 \widetilde{A} G \widetilde{B}) + \frac{1}{n} \mathrm{Tr}(G \widetilde{A} G^2 \widetilde{B}).$$

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

Recall now the following facts: Let B and C be two complex valued matrices of order n. Then, $|\mathrm{Tr}(BC)| \leq \|B\|_2\|C\|_2$ where $\|B\|_2^2 = \sum_{i=1}^n \sum_{j=1}^n b_{ij}^2$ (the b_{ij}'s being the entries of B) and $\max\{\|BC\|_2, \|CB\|_2\} \leq \max_{1\leq i\leq n} |\eta_i|. \|C\|_2$ if B admits a spectral decomposition with eigenvalues $\eta_1, \dots, \eta_n$. Therefore using the above facts, together with the facts that $(\partial_{\mathbf{u}} A)_{ij} = n^{-1/2}$ if $(i, j) = \mathbf{u}$ or $(j, i) = \mathbf{u}$ and 0 otherwise, and that G admits a spectral decomposition with eigenvalues bounded by $1/y$ with $y = \mathrm{Im}\,(z)$, we get

$$\frac{1}{n}|\mathrm{Tr}(G^2\widetilde{A}G\widetilde{B})| \leq \|G^2\widetilde{A}\|_2\|G\widetilde{B}\|_2 \leq \frac{1}{y^3}\frac{2}{n^2}\Big(\sum_{\mathbf{u}\in\mathcal{I}_n} a_{\mathbf{u}}^2 \sum_{\mathbf{v}\in\mathcal{I}_n} b_{\mathbf{v}}^2\Big)^{1/2}.$$

A similar bound being valid for $n^{-1}\mathrm{Tr}(G\widetilde{A}G^2\widetilde{B})$, the lemma follows. $\diamondsuit$

Another key result we use for dealing with Gaussian vectors is:

Lemma 13. *Let $X = (X_k)_{1\leq k\leq n}$ and $Y = (Y_k)_{1\leq k\leq n}$ be two vectors in $\mathbb{L}^2$ which have the same covariance structure. Assume in addition that Y is Gaussian. Then, for all $u \leq k$ we have*

$$\|\mathbb{E}(Y_k|\mathcal{F}_u^Y)\|_2 \leq \|\mathbb{E}(X_k|\mathcal{F}_u^X)\|_2,$$

where $\mathcal{F}_u^Y = \sigma(Y_i, i \leq u)$ and $\mathcal{F}_u^X = \sigma(X_i, i \leq u)$.

Proof. To prove the inequality above, it suffices to notice the following facts. Let

$$\mathcal{V}_u^Y = \overline{\mathrm{span}}(1, (Y_j,\ 1 \leq j \leq u)) \quad \text{and} \quad \mathcal{V}_u^X = \overline{\mathrm{span}}(1, (X_j,\ 1 \leq j \leq u)),$$

where the closure is taken in $\mathbb{L}^2$. Denote by $\Pi_{\mathcal{V}_u^Y}(\cdot)$ the orthogonal projection on $\mathcal{V}_u^Y$ and by $\Pi_{\mathcal{V}_u^X}(\cdot)$ the orthogonal projection on $\mathcal{V}_u^X$. Since $(Y_j)_{1\leq j\leq n}$ is a Gaussian vector $\mathbb{E}(Y_k|\mathcal{F}_u^Y) = \Pi_{\mathcal{V}_u^Y}(Y_k)$ a.s. and in $\mathbb{L}^2$. On another hand, since $(Y_k)_{1\leq k\leq n}$ has the same covariance structure as $(X_k)_{1\leq k\leq n}$, we observe that

$$\|\Pi_{\mathcal{V}_u^Y}(Y_k)\|_2 = \|\Pi_{\mathcal{V}_u^X}(X_k)\|_2.$$

But, by the definition of the conditional expectation, $\|X_k - \mathbb{E}(X_k|\mathcal{F}_u^X)\|_2 \leq \|X_k - \Pi_{\mathcal{V}_u^X}(X_k)\|_2$. Hence, by Pythagoras theorem,

$$\|\Pi_{\mathcal{V}_u^X}(X_k)\|_2 \leq \|\mathbb{E}(X_k|\mathcal{F}_u^X)\|_2.$$

Combining all the observations above, the lemma follows. $\diamondsuit$

References

[1] R. Adamczak, On the Marčenko–Pastur and circular laws for some classes of random matrices with dependent entries, Electron. J. Probab. 16 (2011) 1065–1095.

[2] Z. Bai, J. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., Springer, New York, 2010.

[3] Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns, Statist. Sinica 18 (2008) 425–442.

[4] M. Banna, F. Merlevède, Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes, J. Theoret. Probab. 28 (2015) 745–783.

[5] M. Banna, F. Merlevède, M. Peligrad, On the limiting spectral distribution for a large class of random matrices with correlated entries, Stoch. Proc. Appl. 125 (2015) 2700–2726.

[6] T. Benaych-Georges, M. Debbah, Free deconvolution: from theory to practice, 2008. http://www.cmapx.polytechnique.fr/~benaych/paper-florent.pdf.

[7] H. Bercovici, D. Voiculescu, Free convolutions of measures with unbounded support, Indiana Univ. Math. J. 42 (1993) 733–773.

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

[8] M. Bhattacharjee, A. Bose, Large sample behaviour of high dimensional autocovariance matrices, 2014. Preprint available at: http://www.isical.ac.in/~statmath/html/publication/large_sample_behaviour.pdf.
[9] P. Billingsley, Convergence of Probability Measures, John Wiley & Sons, 1968.
[10] N.H. Bingham, Szegö's theorem and its probabilistic descendants, Probab. Surv. 9 (2012) 287–324.
[11] S.G. Bobkov, F. Götze, A.N. Tikhomirov, On concentration of empirical measures and convergence to the semi-circle law, J. Theoret. Probab. 23 (2010) 792–823.
[12] A. Boutet de Monvel, A. Khorunzhy, On the norm and eigenvalue distribution of large random matrices, Ann. Probab. 27 (1999) 913–944.
[13] A. Boutet de Monvel, A. Khorunzhy, V. Vasilchuk, Limiting eigenvalue distribution of random matrices with correlated entries, Markov Process. Related Fields 2 (1996) 607–636.
[14] R.C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions, Probab. Surv. 2 (2005) 107–144.
[15] R.C. Bradley, Introduction to Strong Mixing Conditions. Vol. 1–3, Kendrick Press, Heber City, UT, 2007.
[16] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer, New York, 1991.
[17] A. Chakrabarty, R.S. Hazra, D. Sarkar, From random matrices to long range dependence, 2014. arXiv:math/1401.0780.
[18] S. Chatterjee, A generalization of the Lindeberg principle, Ann. Probab. 34 (2006) 2061–2076.
[19] R. Couillet, W. Hachem, Analysis of the limiting spectral measure of large random matrices of the separable covariance type, Random Matrices Theory Appl. 3 (4) (2014).
[20] R.A. Davis, O. Pfaffel, R. Stelzer, Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails, Stochastic Process. Appl. 124 (2014) 18–50.
[21] J. Dedecker, S. Gouëzel, F. Merlevède, Some almost sure results for unbounded functions of intermittent maps and their associated Markov chains, Ann. Inst. H. Poincaré Probab. Statist. 46 (2010) 796–821.
[22] A. Edelman, N.R. Rao, Free probability, sample covariance matrices and tochastic Eigen-Inference, 1998. http://hdl.handle.net/1721.1/30241.
[23] J. Geronimo, T. Hill, Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform, J. Approx. Theory 121 (2003) 54–60.
[24] F. Götze, A. Naumov, A. Tikhomirov, Semicircle law for a class of random matrixes with dependent entries, 2012. arXiv:math/0702386v1.
[25] A. Guntuboyina, H. Leeb, Concentration of the spectral measure of large Wishart matrices with dependent entries, Electron. Comm. Probab. 14 (2009) 334–342.
[26] W. Hachem, P. Loubaton, J. Najim, The empirical eigenvalue distribution of a Gram matrix: From independence to stationarity, Markov Process. Related Fields 11 (2005) 629–648.
[27] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, in: Probability and Mathematical Statistics, Academic Press, New York, London, 1980.
[28] B. Jin, C. Wang, Z.D. Bai, K.K. Nair, M. Harding, Limiting spectral distribution of a symmetrized auto-cross covariance matrix, Ann. Appl. Probab. 24 (2014) 1199–1225.
[29] A. Khorunzhy, L. Pastur, On the eigenvalue distribution of the deformed Wigner ensemble of random matrices, in: V.A. Marchenko (Ed.), Spectral Operator Theory and Related Topics, in: Adv. Soviet Math., vol. 19, Amer. Math. Soc., Providence, RI, 1994, pp. 97–127.
[30] H. Liu, A. Aue, D. Paul, The Marčenko–Pastur law for linear time series, Ann. Statist. 43 (2015) 675–712.
[31] C. Liverani, B. Saussol, S. Vaienti, A probabilistic approach to intermittency, Ergodic Theory Dynam. Systems 19 (1999) 671–685.
[32] V. Marčenko, L. Pastur, Distribution of eigenvalues for some sets of random matrices, Mat. Sb. 72 (1967) 507–536.
[33] A. Pajor, L. Pastur, On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution, Studia Math. 195 (2009) 11–29.
[34] G. Pan, J. Gao, Y. Yang, Testing independence among a large number of high-dimensional random vectors, J. Amer. Statist. Assoc. 109 (2014) 600–612.
[35] O. Pfaffel, E. Schlemm, Eigenvalue distribution of large sample covariance matrices of linear processes, Probab. Math. Statist. 31 (2011) 313–329.
[36] N.J. Rao, A. Edelman, Free probability, sample covariance matrices, and Signal processing, in: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, Vol. 5, 2006.
[37] R. Rashidi Far, T. Oraby, W. Bryc, R. Speicher, On slow-fading MIMO systems with nonseparable correlation, IEEE Trans. Inform. Theory 54 (2008) 544–553.
[38] M. Rosenblatt, A comment on a conjecture of N. Wiener, Statist. Probab. Lett. 79 (2009) 347–348.
[39] O. Ryan, M. Debbah, Free deconvolution for signal processing applications, in: ISIT 2007, 2007, pp. 1846–1850.
[40] J. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, J. Multivariate Anal. 55 (1995) 331–339.

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016

[41] J. Silverstein, Z. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 175–192. P.. 1
[42] P. Tilli, A note on the spectral distribution of Toeplitz matrices, Linear Multilinear Algebra 45 (1998) 147–159. 2
[43] E.E. Tyrtyshnikov, N.L. Zamarashkin, Spectra of multilevel Toeplitz matrices: advanced theory via simple matrix relations, Linear Algebra Appl. 270 (1998) 15–27. Q10 3
[44] D.V. Voiculescu, K.J. Dykema, A. Nica, Free Random Variables, Amer. Math. Soc., Providence, RI, 1992. 4
[45] K.W. Wachter, The strong limits of random matrix spectra for sample matrices of independent elements, Ann. Probab. 6 (1978) 1–18. 5
[46] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. of Math. 67 (1958) 325–327. 6
[47] W.B. Wu, M. Woodroofe, A central limit theorem for iterated random functions, J. Appl. Probab. 37 (2000) 748–755. 7
[48] J. Yao, A note on a Marčenko–Pastur type theorem for time series, Statist. Probab. Lett. 82 (2012) 22–28. 8
[49] Y.Q. Yin, Limiting spectral distribution for a class of random matrices, J. Multivariate Anal. 20 (1986) 50–68. 9
[50] Y.Q. Yin, P.R. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices, J. Multivariate Anal. 13 (1983) 489–507. 10

Please cite this article in press as: F. Merlevède, M. Peligrad, On the empirical spectral distribution for matrices with long memory and independent rows, Stochastic Processes and their Applications (2016), http://dx.doi.org/10.1016/j.spa.2016.02.016