On the universality of spectral limit for random matrices with martingale differences entries

F. Merlevède, C. Peligrad[1] and M. Peligrad[2]

Université Paris Est, LAMA (UMR 8050), UPEM, CNRS, UPEC.
Email: florence.merlevede@u-pem.fr

Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, Oh 45221-0025, USA.
Email:peligrc@ucmail.uc.edu, peligrm@ucmail.uc.edu

Key words and phrases. Random matrices, semicircle law, Marchenko-Pastur law, Stieltjes transform, martingale differences, Lindeberg method, random fields.
Mathematical Subject Classification (2010). 60F05, 60F15, 60G42, 60G60.

Abstract

For a class of symmetric random matrices whose entries are martingale differences adapted to an increasing filtration, we prove that under a Lindeberg-like condition, the empirical spectral distribution behaves asymptotically similarly to a corresponding matrix with independent centered Gaussian entries having the same variances. Under a slightly reinforced condition, the approximation holds in the almost sure sense. We also point out several sufficient regularity conditions imposed to the variance structure for convergence to the semicircle law or the Marchenko-Pastur law and other convergence results. In the stationary case we obtain a full extension from the i.i.d. case to the martingale case of the convergence to the semicircle law as well as to the Marchenko-Pastur one. Our results are well adapted to study several examples including non linear autoregressive conditional heteroscedastic random fields of infinite order.

1 Introduction

Some of the most celebrated theorems concerning the limiting density of empirical spectral measure for large random matrices are Wigner's (1958) semicircle law and Marchenko-Pastur (1967) law for covariance matrix. The results have been extended in various directions. In the non-identically distributed case Pastur (1973) showed that a Lindeberg-like condition is sufficient for the convergence to the semicircle law (see also Girko *et al.* (1994) and Girko (2013)). It was shown that the Lindeberg's condition is also relevant for convergence to the Marchenko-Pastur law (see Theorem 3.10 in Bai-Silverstein, 2010). Recently, Tao and Vu (2010) obtained the circular law as spectral limit for matrices with independent entries. All these results assume the independence between the entries of the matrix. An important feature of these results is that the empirical spectral measure converges in distribution for almost all points in the sample space.

For dependent entries the situation is not so well understood. Chatterjee (2006) treated exchangeable entries, whereas papers by Yin and Krishnaiah (1986), Bai and Zhou (2008), Pajor and Pastur (2009), among others, deal with the Marchenko-Pastur law under various assumptions. Several authors also considered the martingale difference type entries. Steps in this direction are papers by Götze and Tikhomirov (2004, 2006) and Götze *et al.* (2012)

[1]Supported in part by a Charles Phelps Taft Memorial Fund grant.
[2]Supported in part by a Charles Phelps Taft Memorial Fund grant, and the NSF grant DMS-1208237.

who treat the semicircle law, and papers by Adamczak (2011, 2013) and O'Rourke (2012) who deal with the Marchenko-Pastur law. These works study the universality for the empirical distribution function when the martingale difference property is defined for an entry of the matrix conditioned by the "past" which is not an ordered filtration, so the results cannot be applied to several martingale random fields useful in statistical applications. Furthermore, the conditions imposed in the stationary case lead to constant conditional variance, with respect to the "past".

There are many time series in econometric theory that can be modeled by an autoregressive process with martingale innovations which have nonconstant conditional variance (heteroscedasticity). A basic diagnostic for knowing that such a model is adequate is to look at the Wachter plot (i.e. to plot the values of the ordered eigenvalues against the quantiles of the Marchenko–Pastur law or Wigner law). Our paper provides a theoretical justification of such a procedure. Therefore, with a view towards applications, the main goal of our paper is to study the universality problem for a more general class of martingale differences which are adapted to an increasing filtration. We also impose a mild mixing condition that allows us to go beyond the constant conditional variance imposed in the previous studies, making possible to treat models that present heteroscedasticity. We provide two types of results, one concerning convergence in probability, and another concerning convergence in distribution of the empirical spectral density for almost all points in the sample space, which we believe is the first one of this type for martingale dependence. As corollaries we point out convergence to the semicircle law, the Marchenko-Pastur law as well as other limits for the limiting spectral density. For martingale differences which are selected from a stationary random field we obtain, without any additional conditions, a generalization of the empirical spectral theorems for i.i.d. In our Section 3, we point out applications of our results to non linear autoregressive conditional heteroscedastic models of infinite order, so called ARCH(∞) models and also to matrices constructed from a triangular array of one dimensional martingales.

Our method consists in comparing the Stieltjes transform of the random matrix with martingale like entries with the Stieltjes transform of a Gaussian matrix with the same covariance structure, which has interest in itself. The proofs are based on a blend of Lindeberg-like method, blocking techniques and delicate maximal inequalities. The blocking is needed to overcome the difficulties raised by selecting meaningful filtrations and mixing conditions associated to random fields. It should be noted that the Lindeberg method, without blocking, was used by Chatterjee (2006) for stydying the semicircular law for exchangeable entries, and by Tao and Wu (2011) for dealing with the universality of local eigenvalues statistics.

The paper is organized in the following way. In Section 2 we list the approximations results, spectral limit theorems, and provide a discussion of our conditions. Applications are included in Section 3. Section 4 is devoted to the main proofs. Finally, in Section 5, we carry out the proofs of some technical results which are important in themselves and also provide some background material.

All along the paper, for positive numbers a_n and b_n, the notation $a_n \ll b_n$ means that for a positive constant c, we have $a_n \leq c\, b_n$ for all n.

2 Results

Let $(X_{\ell k})_{(\ell,k)\in\mathbb{Z}^2}$ be real-valued random variables such that $\mathbb{E}(X_{\ell k}) = 0$ and $\mathbb{E}(X_{\ell k}^2) = \sigma_{\ell k}^2$, and let $(Y_{ij})_{(i,j)\in\mathbb{N}^2}$ be a sequence of independent centered real-valued Gaussian r.v.'s with $\mathbb{E}(Y_{ij}^2) = \sigma_{ij}^2$ which is in addition independent of $(X_{\ell k})_{(\ell,k)\in\mathbb{Z}^2}$. We shall assume that the variables are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We consider the symmetric $n \times n$ random matrix $\mathbf{X}_n$ such that, for any i and j in $\{1, \dots, n\}$

$$\begin{aligned} (\mathbf{X}_n)_{ij} &= X_{ij} \ \text{ for } i \geq j \ \text{ and} \\ (\mathbf{X}_n)_{ij} &= X_{ji} \ \text{ for } i < j \,. \end{aligned} \tag{1}$$

Denote by $\lambda_1^n \leq \cdots \leq \lambda_n^n$ the eigenvalues of

$$\mathbb{X}_n := \frac{1}{n^{1/2}} \mathbf{X}_n \tag{2}$$

and define its distribution function by

$$\mathbf{F}^{\mathbb{X}_n}(t) = \frac{1}{n} \sum_{1 \leq k \leq n} I(\lambda_k \leq t) \,,$$

where $I(A)$ denotes the indicator of an event A.

Similarly we define $\mathbf{Y}_n$ and $\mathbb{Y}_n$ and $\mathbf{F}^{\mathbb{Y}_n}(t)$.

The Levy distance between two distribution functions F and G is defined by

$$d(F, G) = \inf\{\varepsilon > 0 \ : \ F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon\} \,.$$

It is well-known that a sequence of distribution functions $F_n(x)$ converges to a distribution function $F(x)$ at all continuity points x of F if and only if $d(F_n, G) \to 0$. We shall refer to this convergence as weak convergence and denote $F_n \Rightarrow F$. In this paper we are interested in two types of results.

1. Convergence in probability. There is a distribution function $\mathbf{F}$ such that for all positive ϵ

$$\lim_{n \to \infty} \mathbb{P}(d(\mathbf{F}^{\mathbb{X}_n}, \mathbf{F}) > \epsilon) = 0 \,. \tag{3}$$

By abusing the language, for simplicity, we shall denote this type of convergence $\mathbf{F}^{\mathbb{X}_n} \Rightarrow \mathbf{F}$ in probability.

2. Convergence almost sure. There is a distribution function $\mathbf{F}$ such that

$$\mathbb{P}(\lim_{n \to \infty} d(\mathbf{F}^{\mathbb{X}_n}, \mathbf{F}) = 0) = 1. \tag{4}$$

In the sequel the last convergence will be denoted $\mathbf{F}^{\mathbb{X}_n} \Rightarrow \mathbf{F}$ a.s.

The Stieltjes transform of $\mathbf{F}^{\mathbb{X}_n}$ is given by

$$S^{\mathbb{X}_n}(z) = \int \frac{1}{x - z} d\mathbf{F}^{\mathbb{X}_n}(x) = \frac{1}{n} \mathrm{Tr}(n^{-1/2} \mathbf{X}_n - z \mathbf{I}_n)^{-1} \,, \tag{5}$$

where $z = u + iv \in \mathbb{C}^+$ (the set of complex numbers with positive imaginary part), and $\mathbf{I}_n$ is the identity matrix of order n.

In order to introduce the filtration we shall use lexicographic order on $\mathbb{Z}^2$: if $\mathbf{i} = (i_1, i_2)$ and $\mathbf{j} = (j_1, j_2)$ are distinct elements of $\mathbb{Z}^2$ the notation $\mathbf{j} \leq_{\mathrm{lex}} \mathbf{i}$ means that either $i_1 \leq j_1$ or $i_1 = j_1$ and $i_2 \leq j_2$ and the notation $\mathbf{j} <_{\mathrm{lex}} \mathbf{i}$ means that either $i_1 < j_1$ or $i_1 = j_1$ and $i_2 < j_2$. For any non-negative integer a, we introduce now a set of indexes

$$B_{ij}^a = \{(u, v) \in \mathbb{Z}^2; \max(|u - i|, |v - j|) \geq a, \ (u, v) \leq_{\mathrm{lex}} (i, j)\} \tag{6}$$

and for $i \geq j$ the filtration

$$\begin{aligned} \mathcal{F}_{ij}^a &= \sigma(X_{uv} : (u, v) \in B_{ij}^a \text{ and } v \leq u) \ \text{ if } B_{ij}^a \neq \emptyset \,, \\ \mathcal{F}_{ij}^a &= \{\emptyset, \Omega\} \ \text{ if } B_{ij}^a = \emptyset \text{ and} \\ \mathcal{F}_{ji}^a &= \mathcal{F}_{ij}^a \,. \end{aligned} \tag{7}$$

Note that X_{ij} is adapted to $\mathcal{F}^0_{ij}$, which is an increasing filtration in lexicographic order. Our first result compares the distribution of the spectral density of a matrix of martingale difference with the spectral density of a matrix with Gaussian independent entries, defined above. Here and everywhere in the paper we use the standard notation $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ (for X a real or complex-valued random variable).

Theorem 1. *Assume that for all* $1 \le j \le i$,

$$\mathbb{E}(X_{ij}|\mathcal{F}^1_{ij}) = 0 \ a.s. \tag{8}$$

and that

$$\sup_n \frac{1}{n^2} \sum_{1\le j\le i\le n} \sigma^2_{ij} < \infty \,. \tag{9}$$

Assume in addition that

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} \|\mathbb{E}(X^2_{ij} - \sigma^2_{ij}|\mathcal{F}^a_{ij})\|_1 = 0\,, \tag{10}$$

and for any $\varepsilon > 0$,

$$\frac{1}{n^2} \sum_{1\le j\le i\le n} \mathbb{E}(X^2_{ij} I(|X_{ij}| > \varepsilon n^{1/2})) \to 0\,. \tag{11}$$

Then, for all $z \in \mathbb{C}^+$,

$$S^{\mathbb{X}_n}(z) - S^{\mathbb{Y}_n}(z) \to 0 \text{ in probability.} \tag{12}$$

Under a slightly stronger moment condition we obtain an almost sure result.

Theorem 2. *Assume condition (8) is satisfied. Assume also that for some non-decreasing function* $h(x) \ge 1$ *such that* $x^{-1}h(x)$ *is non-increasing and* $\sum_n (nh(n))^{-1} < \infty$, *there exists a positive constant* C *such that*

$$\sup_{i,j} \mathbb{E}(X^2_{ij} h(|X_{ij}|)) \le C\,, \tag{13}$$

and the following condition holds

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} |\mathbb{E}(X^2_{ij} - \sigma^2_{ij}|\mathcal{F}^a_{ij})| = 0 \ a.s. \tag{14}$$

Then for all $z \in \mathbb{C}^+$

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

The relevance of these two theorems is that they make possible to transport the limit results from Gaussian random matrices to matrices with martingale structure. It is well known that in order to establish the convergence of empirical spectral distribution of a sequence of matrices, one needs only to show the convergence of their Stieltjes transforms and the limiting spectral distribution can be obtained from the limiting Stieltjes transform (see Theorem B.9 in Bai-Silverstein (2010), or Corollary 1 in Geronimo and Hill (2003), combined with arguments on page 38 in Bai-Silverstein (2010), based on Vitali's convergence theorem).

With the notations in definitions (3) and (4), let us give two corollaries of the above theorems:

Corollary 3. *Assume that* $(X_{ij})_{(i,j)\in\mathbb{Z}^2}$ *is as in Theorem 1. Furthermore, assume that,*

$$\mathbf{F}^{\mathbb{Y}_n} \Rightarrow \mathbf{F} \text{ in probability,}$$

where $\mathbf{F}$ *is a nonrandom distribution function. Then,*

$$\mathbf{F}^{\mathbb{X}_n} \Rightarrow \mathbf{F} \text{ in probability.}$$

The following corollary is a direct consequence of Theorem 2 and Theorem B.9 in Bai-Silverstein (2010).

Corollary 4. *Assume that $(X_{ij})_{(i,j)\in\mathbb{Z}^2}$ is as in Theorem 2. Furthermore, assume that,*

$$\mathbf{F}^{\mathbb{Y}_n} \Rightarrow \mathbf{F} \quad a.s.$$

where $\mathbf{F}$ is a nonrandom distribution function. Then,

$$\mathbf{F}^{\mathbb{X}_n} \Rightarrow \mathbf{F} \quad a.s.$$

Remark 5. *Our Theorem 1 also holds if the random variables X_{ij} are replaced by a triangular array $X_{n,ij}$ with $j \le i$. For this case the filtration is defined as $\mathcal{F}^a_{n,ij} = \sigma(X_{n,uv} : (u,v) \in B^a_{ij}$ and $v \le u$). The conditions of Theorem 1 should be modified accordingly, meaning that the additional index n should be added in all the conditions.*

Remark 6. *By the contractivity properties of the conditional expectation, the conditions in Theorem 1 could be imposed to larger sigma algebras $\mathcal{K}^a_{ij}$ such that $\mathcal{F}^a_{ij} \subseteq \mathcal{K}^a_{ij}$. For the selection $\mathcal{K}^a_{ij} = \mathcal{F}^1_{ij}$ for all a, condition (10) is implied by*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} \|\mathbb{E}(X^2_{ij} - \sigma^2_{ij}|\mathcal{F}^1_{ij})\|_1 = 0\,, \tag{16}$$

which is similar to Götze et al. (2012) martingale difference condition but with a smaller filtration. The advantage of our condition (10) is that is well adjusted to take care of martingale differences which form a stationary random field.

Remark 7. *We cannot use the same simple argument to enlarge the filtration used in Theorem 2. However the proof of this theorem is based on moment estimates and we notice that the conclusion of Theorem 2 holds if we replace condition (14) by the following condition:*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} |\mathbb{E}(X^2_{ij} - \sigma^2_{ij}|\mathcal{F}^1_{ij})| = 0 \ a.s.$$

Remark 8. *A careful analysis of the proof of Theorem 2 reveals that under a stronger stationarity assumption, condition (13) can be replaced by a weaker condition. More precisely, we infer that we can replace condition (13) by the following one: There is a random variable X such that*

$$\sup_{i,j} \mathbb{P}(|X_{ij}| > x) \le \mathbb{P}(|X| > x)\,,$$

with

$$\mathbb{E}(X^2 \ln(1+|X|)) < \infty\,.$$

Furthermore, in the strictly stationary case we can assume only the existence of moments of order two (see Theorem 11).

Convergence results. Our results can be combined with all the available results for orthogonal Gaussian ensembles to obtain various limiting laws.

1. **Convergence to the semicircle law.**

Let $g(x)$ and $G(x)$ denote the density and the distribution function of the standard semicircle law:

$$g(x) = \frac{1}{2\pi}\sqrt{4-x^2}I(|x|\le 2),\ G(x) = \int_{-\infty}^{x} g(u)du.$$

Combining Theorem 1 with Theorem 1.1 in Götze and Tikhomirov (2004) we obtain under additional regularity condition the following result:

Corollary 9. *Assume besides the conditions of Theorem 1 that*

$$\frac{1}{n^2}\sum_{1\le j\le i\le n}|\sigma_{ij}^2-1|\to 0\,. \tag{17}$$

Then,

$$\mathbf{F}^{\mathbb{X}_n}\Rightarrow G \text{ in probability }.$$

Corollary 10. *If the conditions of Theorem 2 and (17) are satisfied then,*

$$\mathbf{F}^{\mathbb{X}_n}\Rightarrow G \ \ a.s.$$

We consider next a symmetric random matrix which is constructed with variables $(X_{ij})_{1\le j\le i\le n}$ from a stationary real-valued random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$. This means that for all n and any $\mathbf{t},\mathbf{u}_1,\dots,\mathbf{u}_n$ in $\mathbb{Z}^2$ such that $\mathbf{u}_1<_{\text{lex}}\mathbf{u}_2<_{\text{lex}}\dots<_{\text{lex}}\mathbf{u}_n$, $(X_{\mathbf{u}_1},X_{\mathbf{u}_2},..,X_{\mathbf{u}_n})$ has the same distribution as $(X_{\mathbf{u}_1+\mathbf{t}},X_{\mathbf{u}_2+\mathbf{t}},..,X_{\mathbf{u}_n+\mathbf{t}})$.

In this case we have the following generalization of the semicircle law from an i.i.d. to the martingale difference sequences:

Theorem 11. *Assume that $\mathbb{X}_n$ is defined by (2) and based on a stationary real-valued random field $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$. Let $\mathcal{F}^{\infty}_{\mathbf{0}}=\cap_{a\in\mathbb{N}}\mathcal{F}^{a}_{\mathbf{0}}$ where $\mathbf{0}=(0,0)$. Assume that*

$$\mathbb{E}X_{\mathbf{0}}^2=1,\ \mathbb{E}(X_{\mathbf{0}}|\mathcal{F}^1_{\mathbf{0}})=0 \ a.s. \text{ and } \mathbb{E}(X_{\mathbf{0}}^2|\mathcal{F}^{\infty}_{\mathbf{0}})=1 \ a.s.$$

Then,

$$\mathbf{F}^{\mathbb{X}_n}\ \Rightarrow G \quad a.s.$$

2. **Convergence to the Marchenko-Pastur law.**

The sample covariance matrix is very important in multivariate statistical inference. Suppose we have real matrices $\mathbf{X}=\mathbf{X}_{np}=(X_{ij})_{1\le i\le p,1\le j\le n}$. The sample covariance matrix is simply defined as

$$\mathbf{A}=\frac{1}{n}\mathbf{X}\mathbf{X}^T\,,$$

where $\mathbf{X}^T$ is the transpose matrix of $\mathbf{X}$. We shall assume that $p/n\to y$ where $y\in(0,\infty)$. In the context of independent entries with the same mean, variance 1 and satisfying (11) (where the sum extends over $1\le i\le p$ and $1\le j\le n$), the limiting spectral distribution follows the standard Marchenko-Pastur law with the density

$$\tilde g_y(x)=\frac{1}{2\pi xy}\sqrt{(c-x)(x-b)}I(b\le x\le c)$$

and a point mass $1-1/y$ at the origin if $y>1$, where $b=(1-\sqrt{y})^2$ and $c=(1+\sqrt{y})^2$. See Theorem 3.10 in Bai-Silverstein (2010) and the references therein.

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

$$\mathbf{B}_N=\frac{1}{\sqrt{n}}\begin{pmatrix}\mathbf{0} & \mathbf{X}^T\\ \mathbf{X} & \mathbf{0}\end{pmatrix}.$$

The eigenvalues of $\mathbf{B}_N^2$ are the eigenvalues of $n^{-1}\mathbf{X}^T\mathbf{X}$ together with the eigenvalues of $n^{-1}\mathbf{X}\mathbf{X}^T$. Therefore the following relation holds: for any $z\in\mathbb{C}^+$

$$S_{\mathbf{A}}(z)=z^{-1/2}\frac{N}{2p}S_{\mathbf{B}_N}(z^{1/2})+\frac{n-p}{2pz}\,. \tag{18}$$

This relationship together with our results make it possible to formulate the convergence to Marchenko-Pastur law for martingale difference entries. For instance we can give the following result which follows easily by using Theorem 1 together with Remark 5, applied to the matrix $\mathbf{B}_N:=n^{-1/2}(b_{i,j})_{1\le j\le i\le N}$ where $b_{i,j}=X_{i-n,j}\mathbf{1}_{i\ge n+1}\mathbf{1}_{1\le j\le n}$.

Theorem 12. *Suppose we have matrices* $\mathbf{X} = (X_{ij})_{1\le i\le p, 1\le j\le n}$ *of centered, square integrable real-valued r.v.'s with the same variance equals to* 1 *and* $p/n \to y$ *where* $y \in (0,\infty)$. *Assume that for all* (i,j) *such that* $1 \le i \le p$ *and* $1 \le j \le n$,

$$\mathbb{E}(X_{ij}|\sigma(X_{\mathbf{u}}; \mathbf{u} <_{lex} (i,j)) = 0 \ a.s.$$

Assume in addition that

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{p} \sum_{j=1}^{n} \|\mathbb{E}(X_{ij}^2 - 1|\sigma(X_{\mathbf{u}}; \mathbf{u} \in B_{ij}^a))\|_1 = 0\,,$$

where B_{ij}^a *is defined by* (6), *and for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{p} \sum_{j=1}^{n} \mathbb{E}(X_{ij}^2 I(|X_{ij}| > \varepsilon n^{1/2})) = 0\,.$$

Then,

$$\mathbf{F}^{\mathbf{X}\mathbf{X}^T/n} \Rightarrow \tilde{G}_y \quad \text{in probability}\,,$$

where $\tilde{G}_y$ *is the standard Marchenko-Pastur distribution function.*

When the entries of the matrices $\mathbf{X} = (X_{ij})_{1\le i\le p, 1\le j\le n}$ come from a stationary random field, we can formulate an almost sure result. The proof of the next result is omitted since it is based on the relationship (18) and follows the lines of the proof of Theorem 11 (with obvious modifications). Namely, we prove that the Stieltjes transform of $\mathbf{B}_N$ converges almost surely to the Stieltjes transform of the same matrix but with the X_{ij}'s replaced by independent real-valued Gaussian random variables with same variance.

Theorem 13. *Suppose we have matrices* $\mathbf{X} = (X_{ij})_{1\le i\le p, 1\le j\le n}$ *with* $(X_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^2}$ *a strictly stationary real-valued random field. For any* $a \in \mathbb{N}$, *let* $B_{\mathbf{0}}^a$ *be defined in* (6), $\widetilde{\mathcal{F}}_{\mathbf{0}}^a = \sigma(X_{\mathbf{u}}; \mathbf{u} \in B_{\mathbf{0}}^a)$ *and* $\widetilde{\mathcal{F}}_{\mathbf{0}}^\infty = \cap_{a\in\mathbb{N}} \widetilde{\mathcal{F}}_{\mathbf{0}}^a$ *(here* $\mathbf{0} = (0,0)$*). Assume that* $p/n \to y$ *where* $y \in (0,\infty)$ *and*

$$\mathbb{E}X_{\mathbf{0}}^2 = 1,\ \mathbb{E}(X_{\mathbf{0}}|\widetilde{\mathcal{F}}_{\mathbf{0}}^1) = 0 \ a.s. \ \text{and}\ \mathbb{E}(X_{\mathbf{0}}^2|\widetilde{\mathcal{F}}_{\mathbf{0}}^\infty) = 1 \ a.s.$$

Then,

$$\mathbf{F}^{\mathbf{X}\mathbf{X}^T/n} \Rightarrow \tilde{G}_y \quad a.s.$$

where $\tilde{G}_y$ *is the standard Marchenko-Pastur distribution function.*

Note that the above theorem extends the Marchenko-Pastur convergence theorem from the i.i.d. case to the martingale differences case without additional moment assumption.

3. **Other convergence results.**

Our results could be also combined with other theorems for Gaussian structures. If, for instance, the covariance structure is of the form

$$cov(X_{ij}, X_{uv}) = a_i^2 a_j^2 I(i=u) I(j=v) + a_i^2 a_j^2 I(i=v) I(j=u)\,. \tag{19}$$

with

$$\max_{j\ge 1} |a_j| < \infty\,, \tag{20}$$

then

$$cov(X_{ij}, X_{uv}) = V(i,u)V(j,v) + V(i,v)V(j,u)\,,$$

where $V(i,u) = a_i^2 I(i=u)$. We note that condition (2.1) in Boutet de Monvet and Khorunzhy (1999) is satisfied and their Theorem 2.2 applies via our Theorems 1 or 2 where we reduced the study to independent Gaussian variables. This is exactly the function $V(i,u)$ treated in their Remark (iv) on page 918. The spectral limit can be specified uniquely by the relations (2.9a) and (2.9b) in Boutet de Monvet and Khorunzhy (1999) provided the following limit exists

$$\nu(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{1\le j\le n} I(a_j^2 \le t)\,. \tag{21}$$

More precisely we obtain

$$\mathbf{F}^{\mathbb{X}_n} \Rightarrow \mathbf{F} \text{ a.s.} \tag{22}$$

where the Stieltjes transform of $\mathbf{F}$ is given by the relation

$$S(z) = \int_0^\infty \frac{d\nu(\lambda)}{-z - \lambda g(z)}\,,$$

where $g(z)$ is solution of the equation

$$g(z) = \int_0^\infty \frac{\lambda d\nu(\lambda)}{-z - \lambda g(z)} \quad z \in \mathbb{C}\backslash\mathbb{R}\,.$$

This equation is uniquely solvable in the class of analytic functions f defined on $\mathbb{C}\backslash\mathbb{R}$ satisfying the conditions

$$\lim_{x\to\infty} xf(ix) < \infty, \quad \operatorname{Im} f(z) \operatorname{Im} z > 0 \text{ for } z \in \mathbb{C}\backslash\mathbb{R}\,.$$

Therefore we can formulate the following corollary:

Corollary 14. *Assume that (X_{ij}) are as in Theorem 2 and conditions (19), (20) and (21) are satisfied. Then, the convergence (22) holds.*

This result can be applied if (a_j^2) are selected from a stationary and ergodic sequence of random variables (A_k^2) with distribution function $\nu(t)$ and such that $|A_k| < Y$ a.s. for some positive random variable Y. In this case, there is a subset $\Omega' \subset \Omega$, with $\mathbb{P}(\Omega') = 1$ such that for all $\omega \in \Omega'$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1\le j\le n} I(A_j^2 \le t)(\omega) = \nu(t) \text{ and } |A_k(\omega)| < Y(\omega)\,.$$

Then, for $a_k^2 = A_k^2(\omega)$, the convergence (22) holds.

3 Applications

We mention now three applications of our results to classes of random matrices with martingale differences entries which could not be treated by the previous results in the literature. Notice that such results are relevant to statistical procedures. They give, for instance, theoretical justification to use the so-called Wachter plot introduced in [24].

Example 1. We consider a non linear ARCH(∞) random field $(X_{ij})_{(i,j)\in\mathbb{Z}^2}$ given by

$$X_{ij} = \xi_{ij}(c + \sum_{(k,\ell)>_{\text{lex}}(0,0)} g_{k\ell}(X_{i-k,j-\ell}))\,, \tag{23}$$

where $(\xi_{ij})_{(i,j)\in\mathbb{Z}^2}$ is a sequence of centered i.i.d. real-valued random variables such that $\|\xi_{\mathbf{0}}\|_2 = 1$, $c > 0$ and the $g_{k\ell}$ are functions from $\mathbb{R}$ to $\mathbb{R}$ such that for any $(x,y) \in \mathbb{R}^2$,

$$|g_{k\ell}(x) - g_{k\ell}(y)| \le \alpha_{k\ell}|x - y|\,.$$

If $\sum_{(i,j)>_{\rm lex}(0,0)} \alpha_{ij} < 1$ then, by Corollary 2 p. 121 in Doukhan and Truquet (2007), there exists a unique stationary solution of equation (23). This solution is in $\mathbb{L}^2$ and can be written as $X_{ij} = g((\xi_{i-k,j-\ell})_{(k,\ell)\geq_{\rm lex}(0,0)})$. Denote $\sigma^2 = \mathbb{E}(X_{\mathbf{0}}^2)$. Based on this stationary random field we construct the symmetric random matrix $\mathbb{X}_n$.

For any non-negative integer a, consider the sigma algebras $\mathcal{G}_{ij}^a$ and $\widetilde{\mathcal{F}}_{ij}^a$ defined by

$$\mathcal{G}_{ij}^a = \sigma(\xi_{uv} : (u,v) \in B_{ij}^a) \text{ and } \widetilde{\mathcal{F}}_{ij}^a = \sigma(X_{uv} : (u,v) \in B_{ij}^a),$$

with B_{ij}^a defined by (6). Note that $\tilde{\mathcal{F}}_{ij}^a \subseteq \mathcal{G}_{ij}^a$. Therefore

$$\mathbb{E}(X_{ij}|\widetilde{\mathcal{F}}_{ij}^1) = (c + \sum_{(k,\ell)>_{\rm lex}(0,0)} g_{k\ell}(X_{i-k,j-\ell}))\mathbb{E}(\xi_{ij}|\widetilde{\mathcal{F}}_{ij}^1) = 0 \text{ a.s.}$$

In addition, since $\mathcal{G}_{\mathbf{0}}^\infty = \cap_{a\in\mathbb{N}}\mathcal{G}_{\mathbf{0}}^a$ is trivial and $\widetilde{\mathcal{F}}_{\mathbf{0}}^\infty := \cap_{a\in\mathbb{N}}\widetilde{\mathcal{F}}_{\mathbf{0}}^a \subseteq \mathcal{G}_{\mathbf{0}}^\infty$, it follows that $\mathbb{E}(X_{\mathbf{0}}^2|\widetilde{\mathcal{F}}_{\mathbf{0}}^\infty) = \sigma^2$ a.s. Therefore all the conditions of Theorem 11 and also of Theorem 13 are satisfied and therefore their conclusions hold for $\mathbb{X}_n/\sigma$.

Example 2. Consider a real-valued martingale differences sequence $(D_i)_{i\geq 1}$ adapted to the natural filtrations $\mathcal{F}_k = \sigma(D_j, 1 \leq j \leq k)$, and with finite second moment. Let (γ_{ij}) be a matrix of real-valued random variables which are independent of $(D_i)_{i\geq 0}$ and with finite second moments. Then construct the symmetric matrix by using the lexicographic order in the following way:

$$X_{ij} = \gamma_{ij} D_{u(i,j)} \text{ where } u(i,j) = \frac{(i-1)i}{2} + j \text{ for } 1 \leq j \leq i \leq n;$$
$$X_{ij} = X_{ji} \text{ for } 1 \leq i < j \leq n.$$

For clarity we sketch below the lower half of this matrix. The rest is completed by symmetry.

$$\mathbf{D}^n = \begin{pmatrix} \gamma_{11}D_1 & & & & \cdots & \\ \gamma_{21}D_2 & \gamma_{22}D_3 & & & & \cdots \\ \gamma_{31}D_4 & \gamma_{32}D_5 & \gamma_{33}D_6 & & & \cdots \\ \gamma_{41}D_7 & \gamma_{42}D_8 & \gamma_{43}D_9 & \gamma_{44}D_{10} & \cdots & \\ \cdots & & & & & \cdots \\ \gamma_{n1}D_{1+n(n-1)/2} & \cdots & & & & \gamma_{nn}D_{n(n+1)/2} \end{pmatrix}$$

For any non-negative integer a, let us introduce the filtrations

$$\Gamma_{ij}^a = \sigma(\gamma_{uv} : (u,v) \in B_{ij}^a),$$

where B_{ij}^a is defined in (6).

The following result is valid.

Corollary 15. *Assume that for some positive δ we have $\sup_i \mathbb{E}|D_i|^{2+\delta} < \infty$ and that there is a positive constant c such that $\sup_{i,j}|\gamma_{ij}| < c$ a.s. Assume also that*

$$\lim_{a\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{a\leq i\leq n}|\mathbb{E}(D_i^2|\mathcal{F}_{i-a}) - \mathbb{E}D_i^2| = 0 \ a.s.,$$

$$\lim_{a\to\infty}\limsup_{n\to\infty}\frac{1}{n^2}\sum_{1\leq j\leq i\leq n}|\mathbb{E}(\gamma_{ij}^2|\Gamma_{ij}^a) - \mathbb{E}\gamma_{ij}^2| = 0 \ a.s.$$

Then the conclusion of Theorem 2 holds.

The proof of this corollary is a consequence of Theorem 2 via the following remark which uses the proof of Theorem 2 and Remark 23:

Remark 16. *The conclusion of Theorem 2 holds if we replace condition (14) by the following condition:*
For any non-negative integer a, there is a filtration $\mathcal{K}_{ij}^a$ satisfying for any $j \le i$: $\mathcal{F}_{ij}^0 \subseteq \mathcal{K}_{ij}^0$, $\mathcal{K}_{ij}^a \subseteq \mathcal{K}_{ij}^0$, $\mathcal{K}_{ij}^0 \subseteq \mathcal{K}_{i+1,j}^0$ and $\mathcal{K}_{i-a,j}^0 \subseteq \mathcal{K}_{ij}^a$ for $i \ge a+1$, and such that

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} |\mathbb{E}(X_{ij}^2 - \sigma_{ij}^2 | \mathcal{K}_{ij}^a)| = 0 \ \textit{a.s.} \tag{24}$$

Proof of Corollary 15. To prove the result, we first introduce the following notations: for any non-negative integer a, let

$$v(i,j,a) = \frac{(i-1)i}{2} + (j-a)\mathbf{1}_{j\ge a+1} + (j-1)\mathbf{1}_{1\le j\le a},$$

$$\mathcal{G}_{ij}^a = \mathcal{F}_{v(i,j,a)} \ \text{ and } \ \mathcal{K}_{ij}^a = \Gamma_{ij}^a \vee \mathcal{G}_{ij}^a \, .$$

It is easy to see that for any $j \le i$, the filtration $\mathcal{K}_{ij}^a$ satisfies the inclusion properties of Remark 16. Now, by the independence between the sequences (D_i) and (γ_{ij}), we have

$$\mathbb{E}(\gamma_{ij} D_{u(i,j)} | \mathcal{K}_{ij}^1) = \mathbb{E}(\gamma_{ij} | \Gamma_{ij}^1) \mathbb{E}(D_{u(i,j)} | \mathcal{F}_{u(i,j)-1}) = 0 \text{ a.s.}$$

According to Theorem 2 and Remark 16, the corollary will follow if we shall check the condition (24) for $\mathcal{K}_{ij}^a$ defined above. Simple algebra shows that

$$\begin{aligned} |\mathbb{E}[\gamma_{ij}^2 D_{u(i,j)}^2 - \mathbb{E}(\gamma_{ij}^2)\mathbb{E}(D_{u(i,j)}^2) | \mathcal{K}_{ij}^a]| &\le \mathbb{E}(\gamma_{ij}^2 | \Gamma_{ij}^a) |\mathbb{E}(D_{u(i,j)}^2 | \mathcal{G}_{ij}^a) - \mathbb{E}(D_{u(i,j)}^2)| \\ &\quad + \mathbb{E}(D_{u(i,j)}^2) |\mathbb{E}(\gamma_{ij}^2 | \Gamma_{ij}^a) - \mathbb{E}(\gamma_{ij}^2)| \, . \end{aligned}$$

Clearly, under the conditions of Corollary 15, condition (24) will hold if we prove that

$$\lim_{a\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{1\le j\le i\le n} |\mathbb{E}(D_{u(i,j)}^2 | \mathcal{G}_{ij}^a) - \mathbb{E}(D_{u(i,j)}^2)| = 0 \text{ a.s.} \tag{25}$$

With this aim, we write

$$\begin{aligned} &\frac{1}{n^2} \sum_{1\le j\le i\le n} |\mathbb{E}(D_{u(i,j)}^2 | \mathcal{G}_{ij}^a) - \mathbb{E}(D_{u(i,j)}^2)| \\ &\qquad \le \frac{1}{n^2} \sum_{i=a+1}^{n} \sum_{j=a+1}^{i} |\mathbb{E}(D_{u(i,j)}^2 | \mathcal{F}_{u(i,j)-a}) - \mathbb{E}(D_{u(i,j)}^2)| \\ &\qquad\qquad + \frac{1}{n^2} \sum_{j=1}^{a} \sum_{i=j}^{n} \mathbb{E}(D_{u(i,j)}^2) + \frac{1}{n^2} \sum_{j=1}^{a} \sum_{i=j}^{n} \mathbb{E}(D_{u(i,j)}^2 | \mathcal{F}_{u(i,j)-1}) \, . \end{aligned} \tag{26}$$

By assumption, the first term in the right-hand side is going to zero when we first let n tend to infinity and after a. Clearly the second one is going to zero as n is going to infinity since we have $\sup_i \mathbb{E}(D_i^2) < \infty$. To handle the third term, we use the following decomposition:

$$\frac{1}{n^2} \sum_{j=1}^{a} \sum_{i=j}^{n} \mathbb{E}(D_{u(i,j)}^2 | \mathcal{F}_{u(i,j)-1}) \le \frac{a}{n} + \frac{1}{n^{2+\delta/2}} \sum_{j=1}^{a} \sum_{i=j}^{n} \mathbb{E}(|D_{u(i,j)}|^{2+\delta} I(|D_{u(i,j)}| > n^{1/2}) | \mathcal{F}_{u(i,j)-1}) \, ,$$

where δ is such that $\sup_i \mathbb{E}(|D_i|^{2+\delta}) < \infty$. Since

$$\sum_{n\ge 1} \frac{1}{n^{2+\delta/2}} \sum_{j=1}^{a} \sum_{i=j}^{n} \mathbb{E}(|D_{u(i,j)}|^{2+\delta}) < \infty \, ,$$

we conclude easily that the last term in the right-hand side of (26) converges to zero as n tend to infinity. This ends the proof of condition (24) and therefore of the corollary. $\Diamond$

We list below another corollary which follows from our Theorem 1 and whose proof is straightforward.

Corollary 17. *Assume that* (γ_{ij}) *is a sequence of constants satisfying* $\sup_{(i,j)} |\gamma_{ij}| < \infty$ *and assume*

$$\frac{1}{n}\sum_{1\le i\le n}\mathbb{E}(D_i^2) < \infty\,, \tag{27}$$

$$\lim_a \limsup_n \frac{1}{n}\sum_{a\le i\le n}\|\mathbb{E}(D_i^2|\mathcal{F}_{i-a}) - \mathbb{E}D_i^2\|_1 = 0\,, \tag{28}$$

and for any $\varepsilon > 0$,

$$\frac{1}{n^2}\sum_{1\le i\le n^2}\mathbb{E}(D_i^2 I(|D_i| > \varepsilon n^{1/2})) \to 0\,. \tag{29}$$

Then the conclusion of Theorem 1 holds.

In particular, if the sequence (γ_{ij}) is constant, the only relevant conditions in these two last corollaries are imposed on the differences of martingale. Notice also that if $(D_i, i \in \mathbb{Z})$ is a strictly stationary sequence of martingale differences in $\mathbb{L}^2$, the conditions (27) and (29) are obviously satisfied and (28) becomes $\mathbb{E}(D_0^2|\mathcal{F}_{-\infty}) = \mathbb{E}(D_0^2)$ in $\mathbb{L}^1$, where $\mathcal{F}_{-\infty} = \cap_{i\in\mathbb{Z}}\sigma(D_k, k \le i)$. This last condition is equivalent to $\mathbb{E}(D_0^2|\mathcal{F}_{-\infty}) = \mathbb{E}(D_0^2)$ a.s. and it holds if the sequence is ergodic or strong mixing.

Example 3. Consider p independent copies $(D_j^{(i)})_{j\in\mathbb{Z}}$, $i = 1,\dots,p$ of a real-valued martingale differences sequence $(D_i)_{i\in\mathbb{Z}}$ with respect to the natural filtration $\mathcal{F}_j = \sigma(D_k, k \le j)$, such that $\mathbb{E}(D_i^2) = 1$ for any $i \in \mathbb{Z}$. Let $D_{ij} = D_j^{(i)}$ and $\mathbf{X} = \mathbf{X}_{np} = (D_{ij})_{1\le i\le p, 1\le j\le n}$. Applying our Theorem 12, the following corollary holds for the sample covariance matrix:

Corollary 18. *Assume that conditions* (28) *and* (29) *hold, and that* $p/n \to y \in (0,\infty)$. *Then* $\mathbf{F}^{\mathbf{X}\mathbf{X}^T/n} \Rightarrow \tilde{G}_y$ *a.s., where* $\tilde{G}_y$ *is the standard Marchenko-Pastur distribution function.*

Proof of Corollary 18. By using the fact that for any $i \in \{2,\dots,p\}$, $\sigma((D_j^{(i)})_{j\in\mathbb{Z}})$ is independent of $\sigma(D_j^{(k)})_{j\in\mathbb{Z}}, 1 \le k \le i-1)$, we can easily verify that all the conditions of Theorem 12 are satisfied under the assumptions of Corollary 18. Therefore, setting $\mathbf{A}_n = n^{-1}\mathbf{X}\mathbf{X}^T$ we obtain $\mathbf{F}^{\mathbf{A}_n} \Rightarrow \tilde{G}_y$ in probability, or equivalently, for any $z \in \mathbb{C}^+$, $S^{\mathbb{A}_n}(z) \to S_y(z)$ in probability, where $S_y(z)$ is the Stieltjes transform of $\tilde{G}_y$. Furthermore, since both Stieljes transforms are bounded, the convergence in probability implies $\mathbb{E}(S^{\mathbb{A}_n}(z)) \to S_y(z)$. Now, since the rows of $\mathbf{X}$ are independent, for any $z \in \mathbb{C}^+$ we obtain $S^{\mathbb{A}_n}(z) - \mathbb{E}(S^{\mathbb{A}_n}(z)) \to 0$ a.s. (see, for instance, Lemma 4.1 in [1]). So, overall, under the conditions of Corollary 18, we get that $S^{\mathbb{A}_n}(z)$ converges almost surely to $S_y(z)$ that is equivalent to $\mathbf{F}^{\mathbf{A}_n} \Rightarrow \tilde{G}_y$ a.s. $\Diamond$

4 Proofs

4.1 Proof of Theorem 1

We start this section with some notations. For a function f of one variable x we denote by $d^i f = d^i f/dx^i$, the derivative of order i with respect to x. For a multivariate function we use the notations $\partial_k^i f = \partial^i f/\partial^i x_k$ for the partial derivative of order i with respect to the variable x_k. Also $\partial_{jk}^2 f = \partial^2 f/\partial x_j \partial x_k$ means the derivatives with respect to x_j of the derivative with respect to x_k, and so on.

Let $k_n = n(n+1)/2$ and $\mathbf{x}_n = (x_{ij})_{1\le j\le i\le n}$ be a vector of $\mathbb{R}^{k_n}$. Let $\mathbf{A}_n(\mathbf{x}_n)$ be the symmetric matrix of order n defined by

$$(\mathbf{A}_n(\mathbf{x}_n))_{ij} = \begin{cases} \frac{1}{\sqrt{n}} x_{ij} & 1 \le j \le i \le n \\ \frac{1}{\sqrt{n}} x_{ji} & 1 \le i < j \le n \,. \end{cases} \tag{30}$$

It is convenient to introduce a function notation for the Stieltjes transform defined in (5). Let $z \in \mathbb{C}^+$ and let $s_n(\mathbf{x}_n, z)$ be the function defined from $\mathbb{R}^{k_n}$ to $\mathbb{C}$ by

$$s(\mathbf{x}_n) = s_n(\mathbf{x}_n, z) = S^{\mathbf{A}_n(\mathbf{x}_n)}(z) = \frac{1}{n}\mathrm{Tr}(\mathbf{A}_n(\mathbf{x}_n) - z\mathbf{I}_n)^{-1} \,, \tag{31}$$

where $\mathbf{I}_n$ is the identity matrix of order n and for simplicity here and in the sequel we deleted the variable z and the index n from the notation of $s_n(\mathbf{x}_n, z)$. So we write $s(\mathbf{x}_n)$ instead of $s_n(\mathbf{x}_n, z)$ when no confusion is possible. The partial derivatives of the function $s(\mathbf{x}_n)$ have been estimated in Chatterjee (2006). There are three positive constants c_1, c_2 and c_3 depending on $\mathrm{Im}\, z$ such that

$$|\partial_u s(\mathbf{x}_n)| \le \frac{c_1}{n^{3/2}}; \; |\partial^2_{uv} s(\mathbf{x}_n)| \le \frac{c_2}{n^2} \text{ and } |\partial^3_{uvw} s(\mathbf{x}_n)| \le \frac{c_3}{n^{5/2}} \text{ for all } u, v, w. \tag{32}$$

The proof is mainly based on the generalized Lindeberg method for random vectors, which is given in Proposition 19 of Section 5. We shall order the indexes of the variables $(X_{ij})_{1\le j\le i\le n}$ by using the lexicographic order. These indexes are denoted by $u_1 <_{\ell\mathrm{ex}} u_2 <_{\ell\mathrm{ex}} \cdots <_{\ell\mathrm{ex}} u_{k_n}$. Here is the enumeration for the indexes in the lower half part of the matrix

$$\begin{pmatrix} u_1 & \dots & & & & \\ u_2 & u_3 & \dots & & & \\ u_4 & u_5 & u_6 & \dots & & \\ u_7 & u_8 & u_9 & u_{10} & \dots & \\ u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & \\ \dots & & & & & \dots \\ u_{1+n(n-1)/2} & \dots & & & & u_{k_n} \end{pmatrix}.$$

With the above notations, we have that $S^{\mathbb{X}_n} - S^{\mathbb{Y}_n} = s(\mathbf{X}_n) - s(\mathbf{Y}_n)$ where $\mathbf{X}_n = (X_{u_\ell})_{1\le \ell\le k_n}$ and $\mathbf{Y}_n = (Y_{u_\ell})_{1\le \ell\le k_n}$. To prove the theorem, we shall show in what follows that

$$\lim_{n\to\infty} \mathbb{E}|s(\mathbf{X}_n) - s(\mathbf{Y}_n)| = 0 \,. \tag{33}$$

We start the proof by truncating the random variables. Let $\varepsilon > 0$. For any integer $\ell \in [1, k_n]$, we then define

$$T_{u_\ell} = T_{n,u_\ell} = X_{u_\ell} I(|X_{u_\ell}| \le \varepsilon\sqrt{n}) \,. \tag{34}$$

As in our previous notation, when no confusion is possible, to ease the notation, we shall use the notation of T_{u_ℓ} instead of T_{n,u_ℓ}, but we shall keep always in mind the dependence of n. Since $\mathbb{E}(X_{u_\ell} - T_{u_\ell})^2 \le \mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| > \varepsilon\sqrt{n}))$, by (11), it follows that

$$\frac{1}{n^2} \sum_{1\le \ell\le k_n} \mathbb{E}(X_{u_\ell} - T_{u_\ell})^2 \to 0 \text{ as } n \to \infty \,. \tag{35}$$

Denoting $\mathbf{T}_n = (T_{n,u_\ell})_{1\le \ell\le k_n}$, and using Lemma 20 from Section 5 allowing to compare two Stieltjes transforms, we have

$$\mathbb{E}|s(\mathbf{X}_n) - s(\mathbf{T}_n)|^2 \ll \frac{1}{n^2} \sum_{1\le \ell\le k_n} \mathbb{E}(X_{u_\ell} - T_{u_\ell})^2 \,.$$

Taking into account (35), it follows that

$$\mathbb{E}|s(\mathbf{X}_n) - s(\mathbf{T}_n)|^2 \to 0 \,\text{as}\, n \to \infty. \tag{36}$$

Let us consider now a vector $\mathbf{Z}_n = (Z_{n,u_\ell})_{1\le \ell \le k_n}$ of independent centered real-valued Gaussian random variables, independent of $\mathbf{X}_n$ and such that, for all ℓ, we have $\mathbb{E}Z^2_{n,u_\ell} = \mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| \le \varepsilon\sqrt{n}))$. We denote for short $Z_{n,u_\ell} = Z_{u_\ell}$. Let $\mathbb{Z}_n$ be the matrix constructed as in (1). At this point we use the Gaussian interpolation trick as stated in Lemma 21 (given for the convenience in Section 5), which allows us to compare the expectations of functions of two Gaussian vectors in term of differences of their covariances. This lemma together with relation (32) give:

$$|\mathbb{E}(s(\mathbf{Y}_n)) - \mathbb{E}(s(\mathbf{Z}_n))| \ll \frac{1}{n^2} \sum_{1\le \ell \le k_n} |\mathbb{E}X^2_{u_\ell} - \mathbb{E}Z^2_{u_\ell}| = \frac{1}{n^2} \sum_{1\le \ell \le k_n} \mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| > \varepsilon\sqrt{n}))\,.$$

Hence, using (11), it follows that

$$\lim_{n\to\infty} |\mathbb{E}(s(\mathbf{Y}_n)) - \mathbb{E}(s(\mathbf{Z}_n))| = 0\,. \tag{37}$$

Since $\mathbf{Y}_n$ and $\mathbf{Z}_n$ have independent components, it is well-known (see for instance the proof on page 34 in Bai-Silverstein, 2010) that $s(\mathbf{Y}_n) - \mathbb{E}s(\mathbf{Y}_n) \to 0$ a.s. and also $s(\mathbf{Z}_n) - \mathbb{E}s(\mathbf{Z}_n) \to 0$ a.s. By combining these last two almost sure convergence results with (37), we get that $s(\mathbf{Y}_n) - s(\mathbf{Z}_n) \to 0$ a.s. Since the Stieltjes transforms are bounded, we also derive that

$$\mathbb{E}|s(\mathbf{Y}_n) - s(\mathbf{Z}_n)| \to 0 \;\text{ as } n \to \infty\;. \tag{38}$$

Therefore by (36) and (38), we note that the convergence (33), and then the conclusion of the theorem will follow if we prove that

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \mathbb{E}|s(\mathbf{T}_n) - s(\mathbf{Z}_n)| = 0\,. \tag{39}$$

With this aim, we shall apply the approximation in Proposition 19. Let a be a fixed but arbitrary positive integer. For $u_\ell = (i,j)$, $1 \le j \le i \le n$, let $B_{u_\ell}(a)$ be the set

$$B_{u_\ell}(a) = \{(u,v) \in \mathbb{N}^2 \,:\, v \le u \;\text{ and }\; (u,v) \in B^1_{ij} \setminus B^a_{ij}\} \tag{40}$$

with B^1_{ij} and B^a_{ij} defined by (6). For example, when $a = 2$, the indexes (u,v) that belong to the set $B_{u_\ell}(2)$ are described by the points in the next matrix (below $u_\ell = (i,j)$ with $i \ge j+2$).

$$\begin{pmatrix} & & & & & \\ & & & & & \\ & & & & & \\ & \cdot & \cdot & \cdot & & \\ & \cdot & u_\ell & & & \end{pmatrix}$$

Denote

$$\mathbf{C}_{u_\ell} = \mathbf{C}_{n,u_\ell} = (T_{u_1}, \dots, T_{u_{\ell-1}}, 0, Z_{u_{\ell+1}}, \dots, Z_{u_{k_n}})$$

and

$$\mathbf{U}_{u_{\ell-1}} = \mathbf{U}_{n,u_{\ell-1}} = (U_{u_1}, \dots, U_{u_{\ell-1}})\,,$$

where $U_{u_i} = 0$ if $u_i \in B_{u_\ell}(a)$ and $U_{u_i} = T_{u_i}$ if $u_i \in B^a_{u_\ell}$. Set also

$$\widetilde{\mathbf{C}}^a_{u_\ell} = \widetilde{\mathbf{C}}^a_{n,u_\ell} = (\mathbf{U}_{u_{\ell-1}}, 0, Z_{u_{\ell+1}}, \dots, Z_{u_{k_n}})\,.$$

By Proposition 19 and (32), we get that

$$s(\mathbf{T}_n) - s(\mathbf{Z}_n) := \; R_{1,n} + \; R_{2,n}(a) + \; R_{3,n}(a)\,,$$

where

$$R_{1,n} = \sum_{1\le \ell \le k_n} (T_{u_\ell} - Z_{u_\ell})\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\,, \tag{41}$$

$$R_{2,n}(a) = \frac{1}{2}\sum_{1\le \ell \le k_n} (T^2_{u_\ell} - Z^2_{u_\ell})\partial^2_{u_\ell} s(\widetilde{\mathbf{C}}^a_{u_\ell})\,, \tag{42}$$

and

$$|R_{3,n}(a)| \le c_3 \frac{1}{n^{5/2}} \sum_{1\le \ell \le k_n} (T^2_{u_\ell} + Z^2_{u_\ell}) \sum_{u_k \in B_{u_\ell}(a)} |T_{u_k}| + c_3 \frac{1}{n^{5/2}} \sum_{1\le \ell \le k_n} (|T_{u_\ell}|^3 + |Z_{u_\ell}|^3)\,. \tag{43}$$

We first handle the term $R_{1,n}$ and we write

$$|R_{1,n}| \le \Big|\sum_{1\le \ell \le k_n} T_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big| + \Big|\sum_{1\le \ell \le k_n} Z_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big|\,. \tag{44}$$

Since the r.v.'s $Z_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})$, $1 \le \ell \le k_n$, are orthogonal, by using (32), we get

$$\mathbb{E}\Big|\sum_{1\le \ell \le k_n} Z_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big|^2 \ll \frac{1}{n^3}\sum_{1\le \ell \le k_n} \mathbb{E}(Z^2_{u_\ell}) \le \frac{1}{n^3}\sum_{1\le \ell \le k_n} \mathbb{E}(X^2_{u_\ell})\,.$$

Then, by (9), it follows that

$$\mathbb{E}\Big|\sum_{1\le \ell \le k_n} Z_{n,u_\ell}\partial_{u_\ell} s(\mathbf{C}_{n,u_\ell})\Big|^2 \to 0\,, \text{ as } n \to \infty. \tag{45}$$

To analyze the first term in the right-hand side of (44) we use the following decomposition:

$$T_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell}) = D_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell}) + \mathbb{E}(T_{u_\ell}|\mathcal{F}^1_{u_\ell})\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\,,$$

where

$$D_{u_\ell} = D_{n,u_\ell} = T_{u_\ell} - \mathbb{E}(T_{u_\ell}|\mathcal{F}^1_{u_\ell})\,.$$

By using the fact that $\mathbb{E}(X_{u_\ell}|\mathcal{F}^1_{u_\ell}) = 0$ a.s. and (32), we get

$$\Big|\sum_{1\le \ell \le k_n} \mathbb{E}(T_{u_\ell}|\mathcal{F}^1_{u_\ell})\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big| \ll \frac{1}{n^{3/2}}\sum_{1\le \ell \le k_n} \mathbb{E}(|X_{u_\ell}|I(|X_{u_\ell}| > \varepsilon\sqrt{n})|\mathcal{F}^1_{u_\ell})\,.$$

Therefore, by condition (11),

$$\mathbb{E}\Big|\sum_{1\le \ell \le k_n} \mathbb{E}(T_{u_\ell}|\mathcal{F}^1_{u_\ell})\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big| \ll \frac{1}{\varepsilon n^2}\sum_{1\le \ell \le k_n} \mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| > \varepsilon\sqrt{n})) \to 0\,, \text{ as } n \to \infty.$$

On the other hand, since the r.v.'s $D_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})$, $1 \le \ell \le k_n$, are orthogonal, by using (32), we get

$$\mathbb{E}\Big|\sum_{1\le \ell \le k_n} D_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big|^2 \ll \frac{1}{n^3}\sum_{1\le \ell \le k_n} \mathbb{E}(D^2_{u_\ell})\,.$$

But, by the properties of the conditional expectation, $\mathbb{E}(D^2_{u_\ell}) \le \mathbb{E}(T^2_{u_\ell}) \le \mathbb{E}(X^2_{u_\ell})$. Hence, by using (9), it follows that

$$\mathbb{E}\Big|\sum_{1\le \ell \le k_n} D_{n,u_\ell}\partial_{u_\ell} s(\mathbf{C}_{n,u_\ell})\Big|^2 \to 0\,, \text{ as } n \to \infty.$$

So, overall,

$$\mathbb{E}\Big|\sum_{1\le \ell\le k_n} T_{n,u_\ell}\partial_{u_\ell} s(\mathbf{C}_{n,u_\ell})\Big| \to 0 \quad \text{as } n\to\infty,$$

which combined with (45) proves that

$$\mathbb{E}|R_{1,n}| \to 0 \text{ as } \ n\to\infty.$$

We estimate now the term $\mathbb{E}|R_{3,n}(a)|$. We first note that the cardinality of $B_{u_\ell}(a)$ is smaller than $b=2a(a-1)\le 2a^2$. Therefore, by the level of truncation, we derive

$$\sum_{u_i\in B_{u_\ell}(a)} |T_{u_i}| \le 2a^2\varepsilon\sqrt{n}\,.$$

Moreover $\mathbb{E}(T_{u_\ell}^2+Z_{u_\ell}^2)\le 2\sigma_{u_\ell}^2$ and $\mathbb{E}|T_{u_\ell}|^3\le \varepsilon n^{1/2}\sigma_{u_\ell}^2$. On another hand, since Z_{u_ℓ} is a Gaussian r.v., it follows that

$$\mathbb{E}|Z_{u_\ell}|^3 \le 2(\mathbb{E}Z_{u_\ell}^2)^{3/2} = 2(\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}|\le \varepsilon n^{1/2})))^{3/2} \le 2\varepsilon n^{1/2}\sigma_{u_\ell}^2\,.$$

Therefore, the above considerations show that

$$\mathbb{E}|R_{3,n}(a)| \ll \frac{a^2\varepsilon}{n^2}\sum_{1\le \ell\le k_n}\sigma_{u_\ell}^2\,.$$

Whence by (9), for any positive integer a,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\mathbb{E}|R_{3,n}(a)| = 0\,.$$

It remains to analyze $\mathbb{E}|R_{2,n}(a)|$. We shall use the following decomposition:

$$2R_{2,n}(a) = \sum_{1\le \ell\le k_n}(T_{u_\ell}^2 - \mathbb{E}(T_{u_\ell}^2|\mathcal{F}_{u_\ell}^a))\partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) + \sum_{1\le \ell\le k_n}(\mathbb{E}(T_{u_\ell}^2|\mathcal{F}_{u_\ell}^a) - Z_{u_\ell}^2)\partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) := I_n(a) + II_n(a). \quad (46)$$

The analysis of $I_n(a)$ is tedious and is based on a blocking technique which introduces martingale structure. The estimate is done in Lemma 22 of Section 5, which we shall use with $p=2$, $K=\varepsilon n^{1/2}$, $A_{u_\ell}=\partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a)$ and $\mathcal{G}=\sigma(\mathbf{Z}_n)$. Note that by (32), $\max_{1\le\ell\le k_n}|A_{u_\ell}|\le c_2 n^{-2}$. It follows that, for any positive integers n and a,

$$\mathbb{E}|I_n(a)| \ll \varepsilon\sqrt{a}\,. \quad (47)$$

To handle the second term $II_n(a)$ in (46), we first apply the triangle inequality and use (32) to get

$$|II_n(a)| \le \frac{c_2}{n^2}\sum_{1\le\ell\le k_n}|\mathbb{E}(T_{u_\ell}^2|\mathcal{F}_{u_\ell}^a) - \mathbb{E}Z_{u_\ell}^2| + |\sum_{1\le\ell\le k_n}(Z_{u_\ell}^2 - \mathbb{E}Z_{u_\ell}^2)\partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a)|\,. \quad (48)$$

Note that $\mathbb{E}Z_{u_\ell}^2=\mathbb{E}T_{u_\ell}^2$. Therefore

$$\frac{1}{n^2}\sum_{1\le\ell\le k_n}\|\mathbb{E}(T_{u_\ell}^2|\mathcal{F}_{u_\ell}^a) - \mathbb{E}(Z_{u_\ell}^2)\|_1$$

$$\le \frac{1}{n^2}\sum_{1\le\ell\le k_n}\|\mathbb{E}(X_{u_\ell}^2|\mathcal{F}_{u_\ell}^a) - \mathbb{E}(X_{u_\ell}^2)\|_1 + \frac{2}{n^2}\sum_{1\le\ell\le k_n}\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}|>\varepsilon\sqrt{n}))\,.$$

Hence, taking into account conditions (10) and (11), it follows that

$$\lim_{a\to\infty}\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n^2}\sum_{1\le\ell\le k_n}\|\mathbb{E}(T^2_{n,u_\ell}|\mathcal{F}^a_{u_\ell})-\mathbb{E}Z^2_{n,u_\ell}\|_1=0\,. \tag{49}$$

We handle now the last term in the right-hand side of (48). Set

$$d'_{u_\ell}=d'_{n,u_\ell}=(Z^2_{u_\ell}-\mathbb{E}Z^2_{u_\ell})\partial^2_{u_\ell}s(\widetilde{\mathbf{C}}^a_{u_\ell})\,,$$

and observe that the r.v.'s $(d'_{u_\ell})_{\ell\ge 1}$ are orthogonal. Therefore, by (32),

$$\mathbb{E}|\sum_{1\le\ell\le k_n}d'_{u_\ell}|^2\le\sum_{1\le\ell\le k_n}\mathbb{E}|d'_{u_\ell}|^2\ll\frac{1}{n^4}\sum_{1\le\ell\le k_n}\mathbb{E}|Z_{u_\ell}|^4\,.$$

But by the definition of Z_{n,u_ℓ}, we have

$$\mathbb{E}|Z_{u_\ell}|^4\le 3(\mathbb{E}Z^2_{u_\ell})^2=3(\mathbb{E}(X^2_{u_\ell}I(|X_{u_\ell}|\le\varepsilon n^{1/2})))^2\le 3\varepsilon^2 n\sigma^2_{u_\ell}\,.$$

So, by (9),

$$\mathbb{E}|\sum_{1\le\ell\le k_n}d'_{u_\ell}|^2\ll\frac{\varepsilon^2 n}{n^4}\sum_{1\le\ell\le k_n}\sigma^2_{u_\ell}\to 0\ \text{ as } n\to\infty\,. \tag{50}$$

Therefore, from (49) and (50), it follows that

$$\lim_{a\to\infty}\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\mathbb{E}|\mathit{II}_n(a)|=0\,.$$

Hence, letting ε tend to zero in (47), we get that

$$\lim_{a\to\infty}\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\mathbb{E}|R_{2,n}(a)|=0\,.$$

This ends the proof of the theorem. ◊

4.2 Proof of Theorem 2

We shall use the same notations as those introduced in the proof of Theorem 1, and we also start with a truncation argument. For any integer ℓ belonging to $[1,k_n]$, let T_{n,u_ℓ} be defined as in (34) but with $\varepsilon=1$. Therefore, all along the proof, we set

$$T_{u_\ell}:=T_{n,u_\ell}=X_{u_\ell}I(|X_{u_\ell}|\le\sqrt{n})\,, \tag{51}$$

$\mathbf{T}_n=(T_{n,u_\ell})_{1\le\ell\le k_n}$ and $\mathbf{X}_n=(X_{u_\ell})_{1\le\ell\le k_n}$. In the rest of the proof, we shall write T_{u_ℓ} instead of T_{n,u_ℓ} when no confusion is possible. We start by proving that

$$|s(\mathbf{X}_n)-s(\mathbf{T}_n)|\to 0\ \text{ a.s. as } n\to\infty\,. \tag{52}$$

To compare the difference of the Stieltjes transforms above we use Lemma 20. For $z=u+iv$ with $v>0$, this gives

$$\begin{aligned}|s(\mathbf{X}_n)-s(\mathbf{T}_n)|^2&\le\frac{1}{n^2v^4}\sum_{1\le\ell\le k_n}(X_{u_\ell}-T_{u_\ell})^2\\&\le\frac{1}{n^2v^4}\sum_{1\le\ell\le k_n}X^2_{u_\ell}I(|X_{u_\ell}|>n^{1/2}):=v^{-4}U_n\,.\end{aligned} \tag{53}$$

Hence, by the Borel-Cantelli lemma, in order to prove (52), it is enough to prove that, for any $\varepsilon > 0$,

$$\sum_{r\geq 0}\mathbb{P}\big(\max_{2^r\leq j<2^{r+1}} U_j > \varepsilon\big) < \infty\,.$$

It is easy to see that by monotonicity (for instance, for $j \geq n$, we have $X_{u_\ell}^2 I(|X_{u_\ell}| > j^{1/2}) \leq X_{u_\ell}^2 I(|X_{u_\ell}| > n^{1/2})$), we have

$$\max_{2^r\leq j<2^{r+1}} U_j \leq \frac{1}{2^{2r}}\sum_{1\leq \ell\leq k_{2^{r+1}}} X_{u_\ell}^2 I(|X_{u_\ell}| > 2^{r/2})\,.$$

Therefore, by using Markov inequality, we have to establish that

$$\sum_{r\geq 0}\frac{1}{2^{2r}}\sum_{1\leq \ell\leq k_{2^{r+1}}}\mathbb{E}\big(X_{u_\ell}^2 I(|X_{u_\ell}| > 2^{r/2})\big) < \infty\,.$$

or, equivalently,

$$\sum_{n\geq 1}\frac{1}{n^3}\sum_{\ell=1}^{k_{2n}}\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > n^{1/2})) < \infty\,.$$

This holds because of the following computation. By changing the order of summation, and since $k_n \leq n^2$,

$$\sum_{n\geq 1}\frac{1}{n^3}\sum_{1\leq \ell\leq k_{2n}}\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > n^{1/2})) \leq \mathbb{E}\Big(\sum_{\ell\geq 1}X_{u_\ell}^2\sum_{n\geq\sqrt{\ell}/2}\frac{1}{n^3}I(|X_{u_\ell}| > n^{1/2})\Big)$$
$$\ll \sum_{\ell\geq 1}\frac{1}{\ell}\mathbb{E}(X_{u_\ell}^2 I(\sqrt{2}|X_{u_\ell}| > \ell^{1/4}))\,.$$

We continue the estimate in the following way:

$$\sum_{\ell\geq 1}\frac{1}{\ell}\mathbb{E}(X_{u_\ell}^2 I(\sqrt{2}|X_{u_\ell}| > \ell^{1/4})) \leq \sum_{\ell\geq 1}\frac{1}{\ell h(\ell^{1/4}/\sqrt{2})}\mathbb{E}(X_{u_\ell}^2 h(|X_{u_\ell}|)) \ll \sum_{\ell\geq 1}\frac{1}{\ell h(\ell)} < \infty\,,$$

where we used the fact that $h(\cdot)$ is a non-decreasing function, and condition (13).

Therefore, by taking into account (52), to prove the theorem, it suffices to show that

$$|s(\mathbf{T}_n) - s(\mathbf{Y}_n)| \to 0 \ \text{ a.s. as } n\to\infty\,, \tag{54}$$

where $\mathbf{Y}_n = (Y_{u_\ell})_{1\leq \ell\leq k_n}$. With this aim, we shall use Proposition 19 as in the proof of Theorem 1. This leads to the following estimate:

$$s(\mathbf{T}_n) - s(\mathbf{Y}_n) := \ R_{1,n} + \ R_{2,n}(a) + \ R_{3,n}(a)\,, \tag{55}$$

where $R_{1,n}$, $R_{2,n}(a)$ and $R_{3,n}(a)$ are respectively defined in (41), (42) and (43) with the following modifications: the T_{n,u_ℓ}'s are defined by (51) and the Z_{n,u_ℓ}'s are replaced by the Y_{u_ℓ}'s in all the terms involved in the decomposition.

We first prove that

$$|R_{1,n}| \to 0\,, \ \text{a.s. as } n\to\infty. \tag{56}$$

With this aim, as in the proof of Theorem 1, we use the following decomposition:

$$|R_{1,n}| \leq \Big|\sum_{1\leq \ell\leq k_n}\mathbb{E}(T_{u_\ell}|\mathcal{F}_{u_\ell}^1)\partial_{u_\ell}s(\mathbf{C}_{u_\ell})\Big| + \Big|\sum_{1\leq \ell\leq k_n}D_{u_\ell}\partial_{u_\ell}s(\mathbf{C}_{u_\ell})\Big| + \Big|\sum_{1\leq \ell\leq k_n}Y_{u_\ell}\partial_{u_\ell}s(\mathbf{C}_{u_\ell})\Big|\,,$$

where $D_{u_\ell} := D_{n,u_\ell} = T_{n,u_\ell} - \mathbb{E}(T_{n,u_\ell}|\mathcal{F}^1_{u_\ell})$. Hence, by taking into account (32) and the fact that $\mathbb{E}(X_{u_\ell}|\mathcal{F}^1_{u_\ell}) = 0$ a.s., we get that

$$|R_{1,n}| \leq c_1 \frac{1}{n^{3/2}} \sum_{1\leq \ell \leq k_n} |\mathbb{E}(X_{u_\ell} I(|X_{u_\ell}| > n^{1/2})|\mathcal{F}^1_{u_\ell})| \\ + \Big| \sum_{1\leq \ell \leq k_n} D_{u_\ell} \partial_{u_\ell} s(\mathbf{C}_{u_\ell}) \Big| + \Big| \sum_{1\leq \ell \leq k_n} Y_{u_\ell} \partial_{u_\ell} s(\mathbf{C}_{u_\ell}) \Big| . \quad (57)$$

We treat each term in the right hand side separately. To show that the first term in the right-hand side converges almost surely to zero, namely:

$$\frac{1}{n^{3/2}} \sum_{1\leq \ell \leq k_n} |\mathbb{E}(X_{u_\ell} I(|X_{u_\ell}| > n^{1/2})|\mathcal{F}^1_{u_\ell})| \to 0 \text{ a.s., as } n \to \infty, \quad (58)$$

it suffices to prove (by using as before dyadic arguments), that, for any $\varepsilon > 0$,

$$\sum_{n\geq 1} \frac{1}{n} \mathbb{P}\big(\max_{n\leq k<2n} \frac{1}{k^{3/2}} \sum_{1\leq \ell \leq k^2} |\mathbb{E}(X_{u_\ell} I(|X_{u_\ell}| > k^{1/2})|\mathcal{F}^1_{u_\ell})| \geq \varepsilon \big) < \infty . \quad (59)$$

But,

$$\sum_{n\geq 1} \frac{1}{n} \mathbb{E}\big(\max_{n\leq k<2n} \frac{1}{k^{3/2}} \sum_{1\leq \ell \leq k^2} |\mathbb{E}(X_{u_\ell} I(|X_{u_\ell}| > k^{1/2})|\mathcal{F}^1_{u_\ell})| \big) \\ \leq \sum_{n\geq 1} \frac{1}{n^{5/2}} \sum_{1\leq \ell \leq 4n^2} \mathbb{E}(|X_{u_\ell}| I(|X_{u_\ell}| > n^{1/2})) ,$$

and, since $h(\cdot)$ is a non-decreasing sequence, by (13),

$$\sum_{n\geq 1} \frac{1}{n^{5/2}} \sum_{1\leq \ell \leq 4n^2} \mathbb{E}(|X_{u_\ell}| I(|X_{u_\ell}| > n^{1/2})) \ll \sum_{\ell \geq 1} \frac{1}{\ell^{3/4}} \mathbb{E}(|X_{u_\ell}| I(\sqrt{2}|X_{u_\ell}| > \ell^{1/4})) \\ \ll \sum_{\ell\geq 1} \frac{1}{\ell h(\ell^{1/4}/\sqrt{2})} \mathbb{E}(X^2_{u_\ell} h(|X_{u_\ell}|)) \ll \sum_{\ell \geq 1} \frac{1}{\ell h(\ell)} < \infty . \quad (60)$$

Therefore (60) combined with Markov's inequality implies (59), which in turn implies (58). We prove now that

$$\Big| \sum_{1\leq \ell \leq k_n} D_{u_\ell} \partial_{u_\ell} s(\mathbf{C}_{u_\ell}) \Big| \to 0 \text{ a.s. as } n \to \infty. \quad (61)$$

We start by noticing that $(D_{u_\ell} \partial_{u_\ell} s(\mathbf{C}_{u_\ell}))_{1\leq \ell \leq k_n}$ is a martingale difference sequence adapted to the increasing filtration $\sigma(X_{u_1}, \dots, X_{u_\ell}, \mathbf{Y}_n)$. Hence, by Burkholder's inequality for complex-valued martingales (see, for instance, Lemma 2.12 Bai-Silverstein, 2010), and using (32), Cauchy-Schwartz's inequality and the properties of conditional expectation, we obtain

$$\mathbb{E}\Big| \sum_{1\leq \ell \leq k_n} D_{u_\ell} \partial_{u_\ell} s(\mathbf{C}_{u_\ell}) \Big|^4 \ll \frac{1}{(n^{3/2})^4} \mathbb{E}\Big(\sum_{1\leq \ell \leq k_n} D^2_{u_\ell} \Big)^2 \ll \frac{k_n}{n^6} \sum_{1\leq \ell \leq k_n} \mathbb{E}(T^4_{u_\ell}) .$$

By using the fact that $x^{-2}h(x)$ is non-increasing and condition (13), we derive that

$$\sum_{n\geq 1} \frac{k_n}{n^6} \sum_{1\leq \ell \leq k_n} \mathbb{E}(T^4_{u_\ell}) \ll \sum_{n\geq 1} \frac{1}{n^3 h(\sqrt{n})} \sum_{1\leq \ell \leq k_n} \mathbb{E}(X^2_{u_\ell} h(X_{u_\ell})) \\ \ll \sum_{n\geq 1} \frac{1}{n h(\sqrt{n})} \ll \sum_{n\geq 1} \frac{1}{n h(n)} < \infty , \quad (62)$$

which proves (61) by using Borel-Cantelli lemma. We show now that

$$\Big| \sum_{1\le \ell\le k_n} Y_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big| \to 0 \ \text{ a.s. as } n\to\infty. \tag{63}$$

To proof it we note that $(Y_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell}))_{1\le\ell\le k_n}$ is a reversed martingale differences sequence adapted to the decreasing filtration $\sigma(\mathbf{X}_n, Y_{u_{\ell+1}},\dots,Y_{u_n})$. So, using Burkholder's inequality for complex-valued reversed martingale differences, together with (32), we derive that

$$\mathbb{E}\Big| \sum_{1\le \ell\le k_n} Y_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big|^4 \ll \frac{1}{(n^{3/2})^4}\mathbb{E}\Big(\sum_{1\le \ell\le k_n} Y_{u_\ell}^2\Big)^2 \ll \frac{k_n}{n^6}\sum_{1\le \ell\le k_n}\mathbb{E}(Y_{u_\ell}^4)\,.$$

But, $\mathbb{E}(Y_{u_\ell}^4) = 3(\mathbb{E}(Y_{u_\ell}^2))^2 = 3(\mathbb{E}(X_{u_\ell}^2))^2$. Therefore

$$\sum_{n\ge 1}\mathbb{E}\Big| \sum_{1\le \ell\le k_n} Y_{u_\ell}\partial_{u_\ell} s(\mathbf{C}_{u_\ell})\Big|^4 \ll \sum_{n\ge1}\frac{1}{n^2} < \infty\,,$$

which proves (63) by using Borel-Cantelli lemma. Starting from (57), and gathering (58), (61) and (63), the almost sure convergence (56) follows.

We prove now that, for any fixed positive integer a,

$$|R_{3,n}(a)| \to 0 \text{ a.s. as } n\to\infty. \tag{64}$$

By simple algebraic computations involving the inequality $b^2c \le b^3 + c^3$ for any positive numbers b and c, and the estimate of the cardinality of $B_{u_\ell}(a)$ we obtain

$$|R_{3,n}(a)| \ll \frac{a^2}{n^{5/2}}\sum_{1\le \ell\le k_n}|T_{u_\ell}|^3 + \frac{a^2}{n^{5/2}}\sum_{1\le \ell\le k_n}|Y_{u_\ell}|^3\,. \tag{65}$$

Using the fact that $\mathbb{E}(|Y_{u_\ell}|^3) \le 2(\mathbb{E}(Y_{u_\ell}^2))^{3/2} = 2(\mathbb{E}(X_{u_\ell}^2))^{3/2}$, we derive that

$$\sum_{n\ge1}\frac{1}{n^{7/2}}\mathbb{E}\Big(\max_{n\le k<2n}\sum_{1\le \ell\le k^2}|Y_{u_\ell}|^3\Big) \ll \sum_{n\ge1}\frac{1}{n^{7/2}}\sum_{1\le \ell\le 4n^2}\mathbb{E}(|Y_{u_\ell}|^3) \ll \sum_{n\ge1}\frac{1}{n^{3/2}} < \infty\,,$$

which shows, by standard arguments, that the second term in (65) converges almost surely to zero as $n\to\infty$. To end the proof of (64), it remains to show that the first term in (65) converges almost surely to zero as $n\to\infty$. By using standard dyadic arguments and Markov's inequality, we infer that this holds provided that

$$\sum_{n\ge1}\frac{1}{n^{7/2}}\mathbb{E}\Big(\max_{n\le k<2n}\sum_{1\le \ell\le k^2}|X_{u_\ell}|^3 I(|X_{u_\ell}|\le k^{1/2})\Big) < \infty\,. \tag{66}$$

By simple computations involving the fact that $x^{-1}h(x)$ is non-increasing, and condition (13), we get

$$\begin{aligned}&\sum_{n\ge1}\frac{1}{n^{7/2}}\mathbb{E}\Big(\max_{n\le k<2n}\sum_{1\le \ell\le k^2}|X_{u_\ell}|^3 I(|X_{u_\ell}|\le k^{1/2})\Big)\\ &\quad\le \sum_{n\ge1}\frac{1}{n^{7/2}}\sum_{1\le \ell\le 4n^2}\mathbb{E}(|X_{u_\ell}|^3 I(|X_{u_\ell}|\le (2n)^{1/2}) \ll \sum_{n\ge1}\frac{1}{nh(\sqrt{n})} \ll \sum_{n\ge1}\frac{1}{nh(n)} < \infty\,,\end{aligned} \tag{67}$$

which proves (66) and ends the proof of (64).

It remains to handle the term $R_{2,n}(a)$ in (55). Let $\delta \in]0, 1/6[$ and, for any integer ℓ belonging to $[1, k_n]$, denote

$$\bar{X}_{u_\ell} = X_{u_\ell} I(|X_{u_\ell}| \le n^\delta) \ \text{ and } \ \tilde{X}_{u_\ell} = X_{u_\ell} I(n^\delta < |X_{u_\ell}| \le n^{1/2}) .$$

Using the fact that $T_{u_\ell}^2 = \bar{X}_{u_\ell}^2 + \tilde{X}_{u_\ell}^2$, we shall use the following decomposition:

$$\begin{aligned} R_{2,n}(a) = & \sum_{1 \le \ell \le k_n} (\bar{X}_{u_\ell}^2 - \mathbb{E}(\bar{X}_{u_\ell}^2 | \mathcal{F}_{u_\ell}^a))) \partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) \\ & + \sum_{1 \le \ell \le k_n} (\mathbb{E}(\bar{X}_{u_\ell}^2 | \mathcal{F}_{u_\ell}^a) - Y_{u_\ell}^2) \partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) + \sum_{1 \le \ell \le k_n} \tilde{X}_{u_\ell}^2 \partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) \\ := & I_{1,n}(a) + I_{2,n}(a) + I_{3,n}(a) . \end{aligned} \tag{68}$$

By Lemma 22 from Section 5 applied with $K = n^\delta$, $p = 4$, $A_{u_\ell} = \partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a)$ (so by (32), $b_n = c_2 n^{-2}$) and $\mathcal{G} = \sigma(\mathbf{Y}_n)$, we get that

$$\mathbb{E}|I_{1,n}(a)|^4 \ll \frac{a^3}{n^{2-6\delta}} .$$

Therefore, since $2 - 6\delta > 1$,

$$\sum_n \mathbb{E}|I_n(a)|^4 \ll a .$$

So, for any positive integer a, the Borel-Cantelli lemma implies that

$$I_{1,n}(a) \to 0 \text{ a.s. as } n \to \infty . \tag{69}$$

To handle the term $I_{2,n}(a)$ in (68), we apply first the triangle inequality. Combined with (32), this leads to

$$|I_{2,n}(a)| \le \frac{1}{n^2} \sum_{1 \le \ell \le k_n} |\mathbb{E}(\bar{X}_{u_\ell}^2 | \mathcal{F}_{u_\ell}^a) - \mathbb{E}Y_{u_\ell}^2| + \Big| \sum_{1 \le \ell \le k_n} (Y_{u_\ell}^2 - \mathbb{E}Y_{u_\ell}^2) \partial_{u_\ell}^2 s(\widetilde{\mathbf{C}}_{u_\ell}^a) \Big| . \tag{70}$$

By simple computations, we have that

$$\begin{aligned} \frac{1}{n^2} \sum_{1 \le \ell \le k_n} |\mathbb{E}(\bar{X}_{u_\ell}^2 | \mathcal{F}_{u_\ell}^a) - \mathbb{E}Y_{u_\ell}^2| \le & \frac{1}{n^2} \sum_{1 \le \ell \le k_n} |\mathbb{E}(X_{u_\ell}^2 | \mathcal{F}_{u_\ell}^a) - \mathbb{E}X_{u_\ell}^2| \\ & + \frac{1}{n^2} \sum_{1 \le \ell \le k_n} |\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > n^\delta) | \mathcal{F}_{u_\ell}^a)| . \end{aligned} \tag{71}$$

By condition (14), the first term in the right-hand side of (71) converges almost surely to 0 by letting first n tend to infinity and then a tend to infinity. To show that the second term in the right-hand side of (71) converges to zero, we use again standard dyadic arguments and Markov's inequality, and infer that it holds if

$$\sum_{n \ge 1} \frac{1}{n^3} \sum_{1 \le \ell \le 4n^2} \mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > n^\delta)) < \infty . \tag{72}$$

Since $h(\cdot)$ is non-decreasing, by using (13), we get that

$$\begin{aligned} \sum_{n \ge 1} \frac{1}{n^3} \sum_{1 \le \ell \le 4n^2} \mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > n^\delta)) \le & \sum_{n \ge 1} \frac{1}{n^3 h(n^\delta)} \sum_{1 \le \ell \le 4n^2} \mathbb{E}(X_{u_\ell}^2 h(|X_{u_\ell}|)) \\ & \le \sum_{n \ge 1} \frac{1}{n h(n^\delta)} \le \sum_{n \ge 1} \frac{1}{n h(n)} < \infty , \end{aligned}$$

proving (72). To show that the last term in (70) is convergent to 0 a.s., note that the random variables d'_{u_ℓ} defined by $d'_{u_\ell} = (Y^2_{u_\ell} - \mathbb{E}Y^2_{u_\ell})\partial^2_{u_\ell} s(\widetilde{\mathbf{C}}^a_{u_\ell})$ are orthogonal. Moreover, by (32), $\mathbb{E}|d'_{u_\ell}|^2 \ll n^{-4}\mathbb{E}(Y^4_{u_\ell}) = 3n^{-4}(\mathbb{E}(X^2_{u_\ell}))^2$. So

$$\sum_{n\geq 1}\mathbb{E}|\sum_{1\leq \ell\leq k_n} d'_{u_\ell}|^2 \leq \sum_{n\geq 1}\sum_{1\leq \ell\leq k_n}\mathbb{E}|d'_{u_\ell}|^2 \ll \sum_{n\geq 1}\frac{n^2}{n^4} < \infty\,,$$

which combined with the Borel-Cantelli lemma, implies that the last term in (70) converges to 0 a.s. This completes the proof of the fact that

$$\lim_{a\to\infty}\limsup_{n\to\infty}|I_{2,n}(a)| = 0\,. \tag{73}$$

To handle the last term in (68) we note that by (32),

$$|I_{3,n}(a)| \ll \frac{1}{n^2}\sum_{1\leq \ell\leq k_n} X^2_{u_\ell} I(|X_{u_\ell}| > n^\delta)\,. \tag{74}$$

Using once again standard dyadic arguments and Markov's inequality, we infer that $I_{3,n}(a) \to 0$ a.s. as $n\to\infty$ by (72). Therefore combining this fact with (69) and (73) proves that

$$\lim_{a\to\infty}\limsup_{n\to\infty}|R_{2,n}(a)| = 0\,. \tag{75}$$

Finally, the decomposition (55) together with (56), (64) and (75) implies (54) which completes the proof of the theorem. $\Diamond$

4.3 Proof of Theorem 11

We will follow the steps of the proof of Theorem 2 and in addition we shall use the stationarity assumption and ergodic theorems. We have to prove the counterparts of (52), (56), (64), and (75). We shall just mention the differences. To show that the almost sure convergence (52) holds, we notice that by taking into account (53), it suffices to show that

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n^2}\sum_{1\leq \ell\leq k_n} X^2_{u_\ell} I(|X_{u_\ell}| > M) = 0 \text{ a.s.} \tag{76}$$

which follows by applying the ergodic theorem for stationary random fields (see, for instance, Georgii (1988)).

Furthermore, to prove (56), we first modify the proof of (58). Let M be a fixed positive real fixed and notice that for any $n\geq M^2$,

$$\frac{1}{n^{3/2}}\sum_{1\leq \ell\leq k_n}|\mathbb{E}(X_{u_\ell} I(|X_{u_\ell}| > n^{1/2})|\mathcal{F}^1_{u_\ell})| \leq \frac{1}{n^2}\sum_{1\leq \ell\leq k_n}\mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| > M)|\mathcal{F}^1_{u_\ell})\,.$$

Applying once again the ergodic theorem for stationary random fields, we get

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n^2}\sum_{1\leq \ell\leq k_n}\mathbb{E}(X^2_{u_\ell} I(|X_{u_\ell}| > M)|\mathcal{F}^1_{u_\ell}) = 0 \text{ a.s.}$$

proving then that (58) holds. The additional change in the proof of (56) is in the proof of (61), and more specifically in the successive computations given in (62). By taking into account the stationarity and Fubini's theorem, we modify these computations as follows:

$$\sum_{n\geq 1}\frac{k_n}{n^6}\sum_{1\leq \ell\leq k_n}\mathbb{E}(T^4_{u_\ell}) \leq \mathbb{E}\Big(X^4_{\mathbf{0}}\sum_{n\geq 1}\frac{1}{n^2}I(|X_{\mathbf{0}}| \leq n^{1/2})\Big) \ll \mathbb{E}(X^2_{\mathbf{0}}) < \infty\,.$$

On another hand, to show that (64) holds, the only modification consists in the proof that the first term in the right-hand side of (65) converges almost surely to zero when n to infinity. With this aim, it suffices to write that for any positive real M,

$$\frac{1}{n^{5/2}} \sum_{1 \le \ell \le k_n} |T_{u_\ell}|^3 \le \frac{M^3}{n^{1/2}} + \frac{1}{n^2} \sum_{1 \le \ell \le k_n} X_{u_\ell}^2 I(|X_{u_\ell}| > M)$$

and to apply the ergodic theorem for stationary random fields as before (notice that by stationarity, the second term in the right-hand side of (65) could be shown to converge almost surely to zero when n to infinity by using also the ergodic theorem).

We indicate now the differences in the proof of (75). To deal with the first term in the right-hand side of (71), we notice that, by the ergodic theorem for stationary random fields,

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1 \le j \le i \le n} |\mathbb{E}(X_{ij}^2 - 1|\mathcal{F}_{ij}^a)| = \mathbb{E}\big(|\mathbb{E}(X_{\mathbf{0}}^2 - 1|\mathcal{F}_{\mathbf{0}}^a)||\mathcal{I}\big) \text{ a.s.}$$

where $\mathcal{I}$ is the invariant σ-field. Note that, by Proposition 1 in Dedecker (1998), $\mathcal{I}$ is included in the $\mathbb{P}$-completion of $\mathcal{F}_{\mathbf{0}}^a$ for all a. Whence, the sequence $\mathbb{E}\big(|\mathbb{E}(X_{\mathbf{0}}^2 - 1|\mathcal{F}_{\mathbf{0}}^a)||\mathcal{I}\big)_{a\ge1}$ is almost surely decreasing, and therefore convergent almost surely. Since by assumption, $\mathbb{E}(X_{\mathbf{0}}^2|\mathcal{F}_{\mathbf{0}}^\infty) = 1$ a.s., by the reverse martingale theorem it follows that $\lim_{a\to\infty} \mathbb{E}(X_{\mathbf{0}}^2 - 1|\mathcal{F}_{\mathbf{0}}^a) = 0$ a.s. and in $\mathbb{L}^1$. All these arguments prove that the first term in the right-hand side of (71) converges almost surely to zero by letting first n tend to infinity and after a tend to infinity. On another hand, in order to prove that the second term in the right-hand side of (71) converges almost surely to zero when n tends to infinity, it suffices to show that, for any positive integer a,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \sum_{1 \le \ell \le k_n} |\mathbb{E}(X_{u_\ell}^2 I(|X_{u_\ell}| > M)|\mathcal{F}_{u_\ell}^a)| = 0 \text{ a.s.},$$

which follows by the ergodic theorem for stationary random fields. Similarly, the ergodic theorem for stationary random fields together with the bound in (74) allows us to prove that $I_{3,n}(a)$ converges almost surely to zero when n tends to infinity.

Therefore, under the conditions of Theorem 11, the conclusion of Theorem 2 holds. Furthermore, condition (17) is satisfied, hence the result follows from Corollary 9. $\diamondsuit$

5 Technical Results

Below we give an approximation theorem needed for the proof of the main theorems. A related approximation result is in Chatterjee (2006).

Proposition 19. *Suppose that* $\mathbf{X} := (X_1, \dots, X_m)$ *and* $\mathbf{Z} := (Z_1, \dots, Z_m)$ *are random vectors in* $\mathbb{R}^m$. *Suppose that* $f : \mathbb{R}^m \to \mathbb{C}$ *is a function three times differentiable with bounded partial derivatives*

$$|\partial_{uvw}^3 f(x)| \le L_3 \quad \textit{for all } x \textit{ all } u,\ v,\ w.$$

Let B_k *be a subset of the set* $\{i \in \mathbb{N} : 1 \le i < k\}$. *Denote by* $B_k^c = \{i \in \mathbb{N} : 1 \le i < k\} \backslash B_k$. *Define a vector* $\mathbf{U}_{k-1} = (U_1, \dots, U_{k-1})$ *such that* $U_i = 0$ *if* $i \in B_k$ *and* $U_i = X_i$ *if* $i \in B_k^c$. *Then*

$$f(\mathbf{X}) - f(\mathbf{Z}) = R_1 + R_2 + R_3$$

where

$$R_1 = \sum_{1 \le k \le m} (X_k - Z_k) \partial_k f(X_1, \dots, X_{k-1}, 0, Z_{k+1}, \dots, Z_m),$$

$$R_2 = \frac{1}{2} \sum_{1 \le k \le m} (X_k^2 - Z_k^2) \partial_k^2 f(\mathbf{U}_{k-1}, 0, Z_{k+1}, \dots, Z_m),$$

and
$$|R_3| \le L_3 \sum_{1\le k\le m} (X_k^2 + Z_k^2) \sum_{u\in B_k} |X_u| + L_3 \sum_{1\le k\le m} |X_k|^3 + L_3 \sum_{1\le k\le m} |Z_k|^3.$$

Proof. For any $k \in \{0,\dots,m\}$, we define the following vectors

$$Y_k = (X_1,\dots,X_k,Z_{k+1},\dots,Z_m) \text{ and } Y_k^{(0)} = (X_1,\dots,X_{k-1},0,Z_{k+1},\dots,Z_m)\,.$$

Then, we have the telescoping decomposition:

$$f(\mathbf{X}) - f(\mathbf{Z}) = \sum_{1\le k\le m} \big(f(Y_k) - f(Y_k^0) + f(Y_k^0) - f(Y_{k-1})\big)\,.$$

By applying the Taylor expansion of order two, we get

$$f(Y_k) - f(Y_k^0) = X_k \partial_k f(Y_k^0) + \frac{1}{2} X_k^2 \partial_k^2 f(Y_k^0) + R_3'\,,$$

where $|R_3'| \le L_3 |X_k|^3$. By writing a similar expansion for $f(Y_{k-1}) - f(Y_k^0)$ leads to

$$f(\mathbf{X}) - f(\mathbf{Z}) = \sum_{1\le k\le m} (X_k - Z_k)\partial_k f(Y_k^0) + \frac{1}{2}\sum_{1\le k\le m} [X_k^2 - Z_k^2]\partial_k^2 f(Y_k^0) + R_3''\,, \qquad (77)$$

where
$$|R_3''| \le L_3 \sum_{1\le k\le m} (|X_k|^3 + |Z_k|^3)\,.$$

We continue to estimate the second term in the right-hand side of (77). Let $V_k = X_k^2 - Z_k^2$ and write

$$V_k\partial_k^2 f(Y_k^{(0)}) = V_k \partial_k^2 f(\mathbf{U}_{k-1},0,Z_{k+1},\dots,Z_m) + \big(V_k\partial_k^2 f(Y_k^{(0)}) - V_k\partial_k^2 f(\mathbf{U}_{k-1},0,Z_{k+1},\dots,Z_m)\big)\,.$$

By Taylor expansion of first order and taking into account the bounds for the derivatives, we have
$$|V_k\partial_k^2 f(Y_k^{(0)}) - V_k\partial_k^2 f(\mathbf{U}_{k-1},0,Z_{k+1},\dots,Z_m)| \le L_3|V_k| \sum_{u\in B_k} |X_u|.$$

Finally set

$$R_3 = R_3'' + \frac{1}{2}\sum_{1\le k\le m} \big(V_k\partial_k^2 f(Y_k^{(0)}) - V_k\partial_k^2 f(\mathbf{U}_{k-1},0,Z_{k+1},\dots,Z_m)\big)\,,$$

and the result follows. ◊

We state next Lemma 2.1 in Götze *et al.* (2012).

Lemma 20. *Let* $\mathbf{x} = (x_{ij})_{1\le j\le i\le n}$ *and* $\mathbf{y} = (y_{ij})_{1\le j\le i\le n}$ *two elements of* $\mathbb{R}^{k_n}$ *where* $k_n = n(n+1)/2$. *Let* $z = u + iv \in \mathbb{C}^+$ *and* $s(\cdot) := s(\cdot, z)$ *be the function from* $\mathbb{R}^{k_n}$ *to* $\mathbb{C}$ *defined by* (31). *Then*

$$|s(\mathbf{x}) - s(\mathbf{y})| \le \frac{1}{v^2}\Big(\frac{1}{n^2}\sum_{i=1}^n (x_{ii} - y_{ii})^2 + 2\sum_{i=1}^n\sum_{j=1}^{i-1} (x_{ij} - y_{ij})^2\Big)^{1/2}.$$

The following lemma is an easy consequence of the well-known Gaussian interpolation. For reference we cite Talagrand (2010) Section 1.3, Lemma 1.3.1.

Lemma 21. *Suppose that* $\mathbf{Y} = (Y_1, \ldots, Y_m)$ *and* $\mathbf{Z} = (Z_1, \ldots, Z_m)$ *are Gaussian centered random vectors in* $\mathbb{R}^m$ *with independent components. Suppose that* $f : \mathbb{R}^m \to \mathbb{C}$ *is a function twice differentiable with bounded partial derivatives*

$$|\partial_u f(\mathbf{x})| \le L_1 \text{ and } |\partial_u^2 f(\mathbf{x})| \le L_2 \text{ for all } \mathbf{x},\ u.$$

Then

$$|\mathbb{E}f(\mathbf{Y}) - \mathbb{E}f(\mathbf{Z})| \le \frac{L_2}{2} \sum_{i=1}^{n} |\mathbb{E}Y_i^2 - \mathbb{E}Z_i^2| .$$

In the next lemma we compute moments of some terms which appear in the proofs of Theorems 1 and 2. Before stating it, for reader convenience, let us recall some notations: $k_n = n(n+1)/2$ and $(u_\ell, 1 \le \ell \le k_n)$ are double indexes ordered in the strict lexicographic order. To be more precise, for any integer $\ell \in [1, k_n]$, if i is the integer in $[1, n]$ such that $\frac{i(i-1)}{2} + 1 \le \ell \le \frac{i(i+1)}{2}$, then $\ell = \frac{i(i-1)}{2} + j$ with $j \in \{1, \ldots, i\}$ and $u_\ell = (i, j)$.

Lemma 22. *Let* a *and* K *be two positive integers. For any integer* $\ell \in [1, k_n]$, *let*

$$\bar{X}_{u_\ell} = X_{u_\ell} I(|X_{u_\ell}| \le K) .$$

Let $\mathcal{G}$ *be a sigma algebra independent of* $\sigma\{(X_{ij})_{i,j \in \mathbb{Z}^2}\}$ *and* $\mathcal{F}^a_{u_\ell}$ *be defined by* (7). *Let* $(A_{u_\ell})_{1 \le \ell \le k_n}$ *be a sequence of complex-valued random variables such that* A_{u_ℓ} *is* $\mathcal{F}^a_{u_\ell} \vee \mathcal{G}$*-measurable and*

$$\max_{1 \le \ell \le k_n} |A_{u_\ell}| \le b_n \ \ a.s.$$

Assume that condition (9) *holds. Then for any* $p \ge 2$,

$$\mathbb{E}\Big| \sum_{1 \le \ell \le k_n} (\bar{X}^2_{u_\ell} - \mathbb{E}(\bar{X}^2_{u_\ell} | \mathcal{F}^a_{u_\ell})) A_{u_\ell} \Big|^p \ll K^{2(p-1)} b_n^p (a^{p/2} n^{3p/2} + a^{p-1} n^{p+1}) .$$

Proof. The proof is based Burkholder's inequality for differences of martingale with complex valued random variables. Because the filtration $\mathcal{F}^a_{u_\ell}$ is not nested we shall apply a blocking procedure. Let $v_n = [n/a]$ where $[x]$ denotes the integer part of x. Setting

$$d_{i,j} = (\bar{X}^2_{ij} - \mathbb{E}(\bar{X}^2_{ij} | \mathcal{F}^a_{ij})) A_{ij} \, , \text{ if } 1 \le j \le i \le n$$

and

$$d_{i,j} = 0 \, , \text{ if } 1 \le i < j \le n.$$

For pointing out an adapted martingale structure, we decompose the sum in the following way:

$$\sum_{1 \le \ell \le k_n} (\bar{X}^2_{u_\ell} - \mathbb{E}(\bar{X}^2_{u_\ell} | \mathcal{F}^a_{u_\ell})) A_{u_\ell} = \sum_{m=1}^{a} \sum_{k=0}^{v_n - 1} \sum_{j=1}^{ka+m} d_{ka+m,j} + \sum_{i=v_n a+1}^{n} \sum_{j=1}^{i} d_{i,j} \, ,$$

implying that

$$\Big\| \sum_{1 \le \ell \le k_n} (\bar{X}^2_{u_\ell} - \mathbb{E}(\bar{X}^2_{u_\ell} | \mathcal{F}^a_{u_\ell})) A_{u_\ell} \Big\|_p \le \sum_{m=1}^{a} \sum_{j=1}^{n} \Big\| \sum_{k=0}^{v_n - 1} d_{ka+m,j} \Big\|_p + \sum_{i=v_n a+1}^{n} \sum_{j=1}^{i} \|d_{i,j}\|_p \, . \qquad (78)$$

To handle the first term in the right-hand side of the above inequality, we note that for m and j fixed, $(d_{ka+m,j})_{k \ge 0}$ is a complex-valued sequence of martingale differences with respect to the filtration $\mathcal{F}^0_{ka+m,j} \vee \mathcal{G}$. To see this, just note that $d_{ka+m,j}$ is adapted to $\mathcal{F}^0_{ka+m,j} \vee \mathcal{G}$ and we also have, for $k \geq 1$, $\mathcal{F}^0_{(k-1)a+m,j} \subset \mathcal{F}^a_{ka+m,j}$. Then, using also that $A_{ka+m,j}$ is $\mathcal{F}^a_{ka+m,j} \vee \mathcal{G}$-measurable and that $\mathcal{G}$ is independent of $\sigma(X_{u_i}\,,\, 1 \leq i \leq k_n)$, we get for $k \geq 1$,

$$\mathbb{E}(d_{ka+m,j}|\mathcal{F}^0_{(k-1)a+m,j} \vee \mathcal{G}) = \\ \mathbb{E}(A_{ka+m,j}\mathbb{E}(X^2_{ka+m,j} - \mathbb{E}(X^2_{ka+m,j}|\mathcal{F}^a_{ka+m,j})|\mathcal{F}^a_{ka+m,j})|\mathcal{F}^0_{(k-1)a+m,j} \vee \mathcal{G}) = 0 \text{ a.s.}$$

Therefore, by applying Burkholder's inequality for differences of martingale with complex valued (see, for instance, Lemma 2.12 in Bai-Silverstein, 2010), it follows that there exists a universal positive constant C_p depending only on p such that, for any $m \in \{1, \dots, a\}$ and any $j \in \{1, \dots, n\}$,

$$\Big\| \sum_{k=0}^{v_n-1} d_{ka+m,j} \Big\|_p^p \leq C_p \Big\| \sum_{k=0}^{v_n-1} |d_{ka+m,j}|^2 \Big\|_{p/2}^{p/2}. \tag{79}$$

But, for any $1 \leq j \leq i \leq n$,

$$|d_{i,j}| \leq b_n|\bar{X}^2_{ij} - \mathbb{E}(\bar{X}^2_{ij}|\mathcal{F}^a_{ij})| \leq 2b_nK^2\,, \tag{80}$$

implying that

$$\Big| \sum_{k=0}^{v_n-1} |d_{ka+m,j}|^2 \Big|^{p/2} \leq 2^{p-1}K^{2(p-1)}b_n^{p-1} \sum_{k=0}^{v_n-1} |d_{ka+m,j}|\,.$$

Hence, starting from (79) and using the above upper bound, we get

$$\Big\| \sum_{k=0}^{v_n-1} d_{ka+m,j} \Big\|_p^p \leq C_p 2^{p-1}K^{2(p-1)}v_n^{(p-2)/2}b_n^{p-1} \sum_{k=0}^{v_n-1} \mathbb{E}(|d_{ka+m,j}|)\,.$$

which combined with the first part of (80) entails

$$\Big\| \sum_{k=0}^{v_n-1} d_{ka+m,j} \Big\|_p^p \leq C_p 2^p K^{2(p-1)} b_n^p v_n^{(p-2)/2} \sum_{k=0}^{v_n-1} \mathbb{E}(X^2_{ka+m,j})\mathbf{1}_{ka+m \geq j}\,.$$

Therefore, using Hölder's inequality and the above inequality, we derive

$$\begin{aligned}\Big(\sum_{m=1}^{a} \sum_{j=1}^{n} \Big\| \sum_{k=0}^{v_n-1} d_{ka+m,j} \Big\|_p \Big)^p &\leq (an)^{p-1} C_p 2^p K^{2(p-1)} b_n^p v_n^{(p-2)/2} \sum_{m=1}^{a} \sum_{k=0}^{v_n-1} \sum_{j=1}^{ka+m} \mathbb{E}(X^2_{ka+m,j}) \\ &\leq (an)^{p-1} C_p 2^p K^{2(p-1)} b_n^p v_n^{(p-2)/2} \sum_{i=1}^{n} \sum_{j=1}^{i} \mathbb{E}(X^2_{ij})\,.\end{aligned}$$

By taking into account condition (9) and the fact that $v_n \leq n/a$, it follows that

$$\Big(\sum_{m=1}^{a} \sum_{j=1}^{n} \Big\| \sum_{k=0}^{v_n-1} d_{ka+m,j}\mathbf{1}_{ka+m \geq j} \Big\|_p \Big)^p \ll a^{p/2}K^{2(p-1)}b_n^p n^{3p/2}\,. \tag{81}$$

We handle now the second term in the right-hand side of inequality (78). With this aim, we use Hölder's inequality and (80) to get

$$\begin{aligned}\Big(\sum_{i=v_na+1}^{n} \sum_{j=1}^{i} \|d_{i,j}\|_p \Big)^p &\leq (n - v_na)^{p-1}n^{p-1} \sum_{i=v_na+1}^{n} \sum_{j=1}^{i} \mathbb{E}(|d_{i,j}|^p) \\ &\leq a^{p-1}n^{p-1}(2b_nK^2)^{p-1}(2b_n) \sum_{i=1}^{n} \sum_{j=1}^{i} \mathbb{E}(X^2_{ij})\,.\end{aligned}$$

Hence condition (9) implies

$$\Big(\sum_{i=v_n a+1}^{n} \sum_{j=1}^{i} \|d_{i,j}\|_p \Big)^p \ll a^{p-1} K^{2(p-1)} b_n^p n^{p+1} \,. \tag{82}$$

The lemma follows by taking into account the upper bounds (81) and (82) in (78). $\Diamond$

Remark 23. *Our proof shows that the conclusion of the lemma still holds if we replace the filtration $\mathcal{F}^a_{ij}$ by a larger filtration $\mathcal{K}^a_{ij}$ (for $a \geq 0$ fixed) with the following properties: for any $j \leq i$, $\mathcal{F}^0_{ij} \subseteq \mathcal{K}^0_{ij}$, $\mathcal{K}^a_{ij} \subseteq \mathcal{K}^0_{ij}$, $\mathcal{K}^0_{ij} \subseteq \mathcal{K}^0_{i+1,j}$ and $\mathcal{K}^0_{i-a,j} \subseteq \mathcal{K}^a_{ij}$ for $i \geq a+1$. Moreover in the statement of the lemma, the filtration $\mathcal{G}$ has to be assumed to be independent of $\sigma(\bigcup_{i,j} \mathcal{K}^0_{ij})$. For instance, we can take $\mathcal{K}^a_{ij} = \sigma(X_{uv} : (u,v) \in B^a_{ij})$ where B^a_{ij} is defined in (6).*

Acknowledgements. The authors would like to thank Wlodek Bryc and Richard Bradley for useful discussions. The authors are also grateful to the referee for numerous suggestions which improved the presentation of the paper.

References

[1] Adamczak, R. (2011). On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries. *Electronic Journal of Probability* **16** 1065-1095.

[2] Adamczak, R. (2013). Some remarks on the Dozier-Silverstein theorem for random matrices with dependent entries. *Random Matrices: Theory. Appl.* **02** 1-46.

[3] Arnold, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **19** 191-198.

[4] Bai, Z. and J.W. Silverstein (2010). *Spectral analysis of large dimensional random matrices.* Springer, New York, second edition.

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

[6] Boutet de Monvel A. and A. Khorunzhy (1999). On the norm and eigenvalue distribution of large random matrices. *Ann. Probab.* **27** 913-944.

[7] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061-2076.

[8] Dedecker, J. (1998). A central limit theorem for stationary random fields. *Probab. Theory Related Fields* **110** 397-426.

[9] Doukhan, P. and L. Truquet. (2007). A fixed point approach to model random fields. *Alea* **3** 111-132.

[10] Georgii, H.O. (1988). *Gibbs Measures and Phase Transitions.* De Gruyter.

[11] Geronimo J.S. and T. Hill. (2003). Necessary and Sufficient Condition that the Limit of Stieltjes Transforms is a Stieltjes Transform. *J. of Approx. Theory* **121** 54-60.

[12] Girko, V. (2013). The Generalized Circular Law. *Random Operators and Stochastic Equations* **21** 67–109.

[13] Girko, V., Kirsch, W. and A. Kutzelnigg (1994). A necessary and sufficient condition for the semicircle law. *Random Operators and Stochastic Equations* **2** 195-202.

[14] Götze, F. and A. Tikhomirov (2004). Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem.* S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), **311** (Veroyatn. i Stat. 7) 92-23, 299.

[15] Götze, F. and A.N. Tikhomirov (2006). Limit theorems for spectra of random matrices with martingale structure. *Teor. Veroyatn. Primen.*, **51** 171-192.

[16] Götze, F., Naumov, A. and A. Tikhomirov (2012). Semicircle law for a class of random matrixes with dependent entries. arXiv:math/0702386v1

[17] Marchenko, V.A. and L.A. Pastur. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1** 457–483.

[18] O'Rourke, S. (2012). A note on the Marchenko-Pastur law for a class of random matrices with dependent entries. arXiv:1201.3554.

[19] Pajor, A. and L. Pastur. (2009). On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution, *Studia Math.* **195** 11-29.

[20] Pastur, L.A. (1973). Spectra of random selfadjoint operators. *Uspehi Mat. Nauk* **28** 3-64.

[21] Talagrand M. (2010). *Mean Field Models for Spin Glasses. Vol 1. Basic Examples.* Springer.

[22] Tao, T and V. Vu (2010). Random matrices: universality of ESDs and the circular law. With an appendix by Manjunath Krishnapur. *Ann. Probab.* **38** 2023–2065.

[23] Tao, T and V. Vu (2011). Random matrices: Universality of local eigenvalue statistics, *Acta Mathematica* **206** 127-204.

[24] Wachter, K.W. (1976). Proceedings of the Computer Science and Statistics 9th Annual Symposium on the Interface, pp. 299–308.

[25] Wigner E.P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* **67** 325-327.

[26] Yin, Y.Q. and P.R. Krishnaiah. (1986) Limit Theorem for the Eigenvalues of the Sample Covariance Matrix when the Underlying Distribution is Isotropic, *Theory Probab. Appl.* **30** 861-867.