Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes

Marwa Banna and Florence Merlevède

Université Paris Est, LAMA (UMR 8050), UPEMLV, CNRS, UPEC, 5 Boulevard Descartes, 77 454 Marne La Vallée, France.
E-mail: marwa.banna@univ-mlv.fr; florence.merlevede@univ-mlv.fr

Abstract

In this paper we derive an extension of the Marčenko-Pastur theorem to a large class of weak dependent sequences of real-valued random variables having only moment of order 2. Under a mild dependence condition that is easily verifiable in many situations, we derive that the limiting spectral distribution of the associated sample covariance matrix is characterised by an explicit equation for its Stieltjes transform, depending on the spectral density of the underlying process. Applications to linear processes, functions of linear processes and ARCH models are given.

Key words: Sample covariance matrices, weak dependence, Lindeberg method, Marčenko-Pastur distributions, limiting spectral distribution.
Mathematical Subject Classification (2010): 60F99, 60G10, 62E20.

1 Introduction

A typical object of interest in many fields is the sample covariance matrix $\mathbf{B}_n = n^{-1}\sum_{j=1}^n \mathbf{X}_j^T\mathbf{X}_j$ where $(\mathbf{X}_j)$, $j = 1, \dots, n$, is a sequence of $N = N(n)$-dimensional real-valued row random vectors. The interest in studying the spectral properties of such matrices has emerged from multivariate statistical inference since many test statistics can be expressed in terms of functionals of their eigenvalues. The study of the empirical distribution function (e.d.f.) $F^{\mathbf{B}_n}$ of the eigenvalues of $\mathbf{B}_n$ goes back to Wishart 1920's, and the spectral analysis of large-dimensional sample covariance matrices has been actively developed since the remarkable work of Marčenko and Pastur (1967) stating that if $\lim_{n\to\infty} N/n = c \in (0, \infty)$, and all the coordinates of all the vectors $\mathbf{X}_j$'s are i.i.d. (independent identically distributed), centered and in $\mathbb{L}^2$, then, with probability one, $F^{\mathbf{B}_n}$ converges in distribution to a non-random distribution (the original Marčenko-Pastur's theorem is stated for random variables having moment of order four, for the proof under moment of order two only, we refer to Yin (1986)).

Since the Marčenko-Pastur's pioneering paper, there has been a large amount of work aiming at relaxing the independence structure between the coordinates of the $\mathbf{X}_j$'s. Yin (1986) and Silverstein (1995) considered a linear transformation of independent random variables which leads to the study of the empirical spectral distribution of random matrices of the form $\mathbf{B}_n = n^{-1}\sum_{j=1}^n \Gamma_N^{1/2}\mathbf{Y}_j^T\mathbf{Y}_j\Gamma_N^{1/2}$ where Γ_N is an $N\times N$ non-negative definite Hermitian random matrix, independent of the $\mathbf{Y}_j$'s which are i.i.d and such that all their coordinates are i.i.d. In the latter paper, it is shown that if $\lim_{n\to\infty} N/n = c \in (0,\infty)$ and F^{Γ_N} converges almost surely in distribution to a non-random probability distribution function (p.d.f.) H on $[0,\infty)$, then, almost surely, $F^{\mathbf{B}_n}$ converges in distribution to a (non-random) p.d.f. F that is characterized in terms of its Stieltjes transform which satisfies a certain equation. Some further investigations on the model above mentioned can be found Silverstein and Bai (1995) and Pan (2010).

A natural question is then to wonder if other possible correlation patterns of coordinates can be considered, in such a way that, almost surely (or in probability), $F^{\mathbf{B}_n}$ still converges in distribution to a non-random p.d.f. The recent work by Bai and Zhou (2008) is in this direction. Assuming that the $\mathbf{X}_j$'s are i.i.d. and a very general dependence structure of their coordinates, they derive the limiting spectral distribution (LSD) of $\mathbf{B}_n$. Their result has various applications. In particular, in case when the $\mathbf{X}_j$'s are independent copies of $\mathbf{X} = (X_1, \dots, X_N)$ where $(X_k)_{k\in\mathbb{Z}}$ is a stationary linear process with centered i.i.d. innovations, applying their Theorem 1.1, they prove that, almost surely, $F^{\mathbf{B}_n}$ converges in distribution to a non-random p.d.f. F, provided that $\lim_{n\to\infty} N/n = c \in (0,\infty)$, the coefficients of the linear process are absolutely summable and the innovations have a moment of order four (see their Theorem 2.5). For this linear model, let us mention that in a recent paper, Yao (2012) shows that the Stieltjes transform of the limiting p.d.f. F satisfies an explicit equation that depends on c and on the spectral density of the underlying linear process. Still in the context of the linear model described above but, relaxing the equidistribution assumption on the innovations, and using a different approach than the one considered in the papers by Bai and Zhou (2008) and by Yao (2012), Pfaffel and Schlemm (2011) also derive the LSD of $\mathbf{B}_n$ still assuming moments of order four for the innovations plus a polynomial decay of the coefficients of the underlying linear process.

In this work, we extend such Marčenko-Pastur type theorems along another direction. We shall assume that the $\mathbf{X}_j$'s are independent copies of $\mathbf{X} = (X_1, \dots, X_N)$ where $(X_k)_{k\in\mathbb{Z}}$ is a stationary process of the form $X_k = g(\cdots, \varepsilon_{k-1}, \varepsilon_k)$ where the ε_k's are i.i.d. real valued random variables and $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function such that X_k is a proper centered random variable. Assuming that X_0 has a moment of order two only, and imposing a dependence condition expressed in terms of conditional expectation, we prove that if $\lim_{n\to\infty} N/n = c \in (0,\infty)$, then almost surely, $F^{\mathbf{B}_n}$ converges in distribution to a non-random p.d.f. F whose Stieltjes transform satisfies an explicit equation that depends on c and on the spectral density of the underlying stationary process $(X_k)_{k\in\mathbb{Z}}$ (see our Theorem 2.1). The imposed dependence condition is directly related to the physical mechanisms of the underlying process, and is easy verifiable in many situations. For instance, when $(X_k)_{k\in\mathbb{Z}}$ is a linear process with i.i.d. innovations, our dependence condition is satisfied, and then our Theorem 2.1 applies, as soon as the coefficients of the linear process are absolutely summable and the innovations have a moment of order two only, which improves Theorem 2.5 in Bai and Zhou (2008) and Theorem 1.1 in Yao (2012). Other models, such as functions of linear processes and ARCH models, for which our Theorem 2.1 applies, are given in Section 3.

Let us now give an outline of the method used to prove our Theorem 2.1. Since the $\mathbf{X}_j$'s are independent, the result will follow if we can prove that the expectation of the Stieltjes transform of $F^{\mathbf{B}_n}$, say $S_{F^{\mathbf{B}_n}}(z)$, converges to the Stieltjes transform of F, say $S(z)$, for any complex number z with positive imaginary part. With this aim, we shall consider a sample covariance matrix $\mathbf{G}_n = n^{-1}\sum_{j=1}^n \mathbf{Z}_j^T\mathbf{Z}_j$ where the $\mathbf{Z}_j$'s are independent copies of $\mathbf{Z} = (Z_1, \dots Z_N)$ where $(Z_k)_{k\in\mathbb{Z}}$ is a sequence of Gaussian random variables having the same covariance structure as the underlying process $(X_k)_{k\in\mathbb{Z}}$. The $\mathbf{Z}_j$'s will be assumed to be independent of the $\mathbf{X}_j$'s. Using the Gaussian structure of $\mathbf{G}_n$, the convergence of $\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)$ to $S(z)$ will follow by Theorem 1.1 in Silverstein (1995). The main step of the proof is then to show that the difference between the expectations of the Stieltjes transform of $F^{\mathbf{B}_n}$ and that of $F^{\mathbf{G}_n}$ converges to zero. This will be achieved by approximating first $(X_k)_{k\in\mathbb{Z}}$ by an m-dependent sequence of random variables that are bounded. This leads to a new sample covariance matrix $\bar{\mathbf{B}}_n$. We then handle the difference between $\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)$ and $\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)$ with the help of the so-called Lindeberg method used in the multidimensional case. Lindeberg method is known to be an efficient tool to derive limit theorems and, from our knowledge, it has been used for the first time in the context of random matrices by Chatterjee (2006). With the help of this method, he proved the LSD of Wigner matrices associated with exchangeable random variables.

The paper is organized as follows: in Section 2, we specify the model and state the LSD result for the sample covariance matrix associated with the underlying process. Applications to linear processes, functions of linear processes and ARCH models are given in Section 3. Section 4 is devoted to the proof of the main result, whereas some technical tools are stated and proved in Appendix.

Here are some notations used all along the paper. For any non-negative integer q, the notation $\mathbf{0}_q$ means a row vector of size q. For a matrix A, we denote by A^T its transpose matrix, by $\mathrm{Tr}(A)$ its trace, by $\|A\|$ its spectral norm, and by $\|A\|_2$ its Hilbert-Schmidt norm (also called the Frobenius norm). 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 square matrix A of order N with only real eigenvalues, the empirical spectral distribution of A is defined as

$$F^A(x) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\{\lambda_k \leq x\}} ,$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of A. The Stieltjes transform of F^A is given by

$$S_{F^A}(z) = \int \frac{1}{x - z} dF^A(x) = \frac{1}{N} \mathrm{Tr}(A - z\mathbf{I})^{-1} ,$$

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

Finally, the notation $[x]$ is used to denote the integer part of any real x and, for two reals a and b, the notation $a \wedge b$ means $\min(a, b)$, whereas the notation $a \vee b$ means $\max(a, b)$.

2 Main result

We consider a stationary causal process $(X_k)_{k\in\mathbb{Z}}$ defined as follows: let $(\varepsilon_k)_{k\in\mathbb{Z}}$ be a sequence of i.i.d. real-valued random variables and let $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function such that, for any $k \in \mathbb{Z}$,

$$X_k = g(\xi_k) \quad \text{with} \quad \xi_k := (\ldots, \varepsilon_{k-1}, \varepsilon_k) \tag{2.1}$$

is a proper random variable, $\mathbb{E}(g(\xi_k)) = 0$ and $\|g(\xi_k)\|_2 < \infty$.

The framework (2.1) is very general and it includes many widely used linear and nonlinear processes. We refer to the papers by Wu (2005, 2011) for many examples of stationary processes that are of form (2.1). Following Priestley (1988) and Wu (2005), $(X_k)_{k\in\mathbb{Z}}$ can be viewed as a physical system with ξ_k (respectively X_k) being the input (respectively the output) and g being the transform or data-generating mechanism.

For n a positive integer, we consider n independent copies of the sequence $(\varepsilon_k)_{k\in\mathbb{Z}}$ that we denote by $(\varepsilon_k^{(i)})_{k\in\mathbb{Z}}$ for $i = 1, \ldots, n$. Setting $\xi_k^{(i)} = \big(\ldots, \varepsilon_{k-1}^{(i)}, \varepsilon_k^{(i)}\big)$ and $X_k^{(i)} = g(\xi_k^{(i)})$, it follows that $(X_k^{(1)})_{k\in\mathbb{Z}}, \ldots, (X_k^{(n)})_{k\in\mathbb{Z}}$ are n independent copies of $(X_k)_{k\in\mathbb{Z}}$. Let now $N = N(n)$ be a sequence of positive integers, and define for any $i \in \{1, \ldots, n\}$, $\mathbf{X}_i = \big(X_1^{(i)}, \ldots, X_N^{(i)}\big)$. Let

$$\mathcal{X}_n = (\mathbf{X}_1^T | \ldots | \mathbf{X}_n^T) \quad \text{and} \quad \mathbf{B}_n = \frac{1}{n} \mathcal{X}_n \mathcal{X}_n^T . \tag{2.2}$$

In what follows, $\mathbf{B}_n$ will be referred to as the sample covariance matrix associated with $(X_k)_{k\in\mathbb{Z}}$. To derive the limiting spectral distribution of $\mathbf{B}_n$, we need to impose some dependence structure on $(X_k)_{k\in\mathbb{Z}}$. With this aim, we introduce the projection operator: for any k and j belonging to $\mathbb{Z}$, let

$$P_j(X_k) = \mathbb{E}(X_k|\xi_j) - \mathbb{E}(X_k|\xi_{j-1}) .$$

We state now our main result.

Theorem 2.1 *Let $(X_k)_{k\in\mathbb{Z}}$ be defined in (2.1) and $\mathbf{B}_n$ by (2.2). Assume that*

$$\sum_{k\geq 0} \|P_0(X_k)\|_2 < \infty \,, \tag{2.3}$$

and that $c(n) = N/n \to c \in (0,\infty)$. Then, with probability one, $F^{\mathbf{B}_n}$ tends to a non-random probability distribution F, whose Stieltjes transform $S = S(z)$ $(z \in \mathbb{C}^+)$ satisfies the equation

$$z = -\frac{1}{\underline{S}} + \frac{c}{2\pi}\int_0^{2\pi} \frac{1}{\underline{S} + \big(2\pi f(\lambda)\big)^{-1}} d\lambda \,, \tag{2.4}$$

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

Let us mention that, in the literature, the condition (2.3) is referred to as the Hannan-Heyde condition and is known to be essentially optimal for the validity of the central limit theorem for the partial sums (normalized by $\sqrt{n}$) associated with an adapted regular stationary process in $\mathbb{L}^2$. As we shall see in the next section, the quantity $\|P_0(X_k)\|_2$ can be computed in many situations including non linear models. We would like to mention that the condition (2.3) is weaker than the 2-strong stability condition introduced by Wu (2005, Definition 3) that involves a coupling coefficient.

Remark 2.2 Under the condition (2.3), the series $\sum_{k\geq 0} |\mathrm{Cov}(X_0, X_k)|$ is finite (see for instance the inequality (4.61)). Therefore (2.3) implies that the spectral density $f(\cdot)$ of $(X_k)_{k\in\mathbb{Z}}$ exists, is continuous and bounded on $[0, 2\pi)$. It follows that Proposition 1 in Yao (2012) concerning the support of the limiting spectral distribution F still applies if (2.3) holds. In particular, F is compactly supported. Notice also that condition (2.3) is essentially optimal for the covariances to be absolutely summable. Indeed, for a causal linear process with non-negative coefficients and generated by a sequence of i.i.d. real-valued random variables centered and in $\mathbb{L}^2$, both conditions are equivalent to the summability of the coefficients.

Remark 2.3 Let us mention that each of the following conditions is sufficient for the validity of (2.3):

$$\sum_{n\geq 1} \frac{1}{\sqrt{n}} \|\mathbb{E}(X_n|\xi_0)\|_2 < \infty \quad \text{or} \quad \sum_{n\geq 1} \frac{1}{\sqrt{n}} \|X_n - \mathbb{E}(X_n|\mathcal{F}_1^n)\|_2 < \infty \,, \tag{2.5}$$

where $\mathcal{F}_1^n = \sigma(\varepsilon_k \,, 1 \leq k \leq n)$. A condition as the second part of (2.5) is usually referred to as a near epoch dependence type condition. The fact that the first part of (2.5) implies (2.3) follows from Corollary 2 in Peligrad and Utev (2006). Corollary 5 of the same paper asserts that the second part of (2.5) implies its first part.

Remark 2.4 Since many processes encountered in practice are causal, Theorem 2.1 is stated for the one-sided process $(X_k)_{k\in\mathbb{Z}}$ having the representation (2.1). With non-essential modifications in the proof, the same result holds when $(X_k)_{k\in\mathbb{Z}}$ is a two-sided process having the representation

$$X_k = g(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots) \,, \tag{2.6}$$

where $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. real-valued random variables. Assuming that X_0 is centered and in $\mathbb{L}^2$, condition (2.3) has then to be replaced by the following condition: $\sum_{k\in\mathbb{Z}} \|P_0(X_k)\|_2 < \infty$.

Remark 2.5 One can wonder if Theorem 2.1 extends to the case of functionals of another strictly stationary sequence which can be strong mixing or absolutely regular, even if this framework and ours have different range of applicability. Actually, many models encountered in econometric theory have the representation (2.1) whereas, for instance, functionals of absolutely regular (β-mixing) sequences occur naturally as orbits of chaotic dynamical systems. In this situation, we do not think that Theorem 2.1 extends in its full generality without requiring an additional near epoch dependence type condition. It is outside the scope of this paper to study such models which will be the object of further investigations.

3 Applications

In this section, we give two different classes of models for which the condition (2.3) is satisfied and then for which our Theorem 2.1 applies. Other classes of models, including non linear time series such as iterative Lipschitz models or chains with infinite memory, which are of the form (2.1) and for which the quantities $\|P_0(X_k)\|_2$ or $\|\mathbb{E}(X_k|\xi_0)\|_2$ can be computed may be found in Wu (2011).

3.1 Functions of linear processes

In this section, we shall focus on functions of real-valued linear processes. Define

$$X_k = h\Big(\sum_{i\geq 0} a_i \varepsilon_{k-i}\Big) - \mathbb{E}\Big(h\Big(\sum_{i\geq 0} a_i \varepsilon_{k-i}\Big)\Big)\,, \tag{3.1}$$

where $(a_i)_{i\in\mathbb{Z}}$ is a sequence of real numbers in ℓ^1 and $(\varepsilon_i)_{i\in\mathbb{Z}}$ is a sequence of i.i.d. real-valued random variables in $\mathbb{L}^1$. We shall give sufficient conditions in terms of the regularity of the function h, for the condition (2.3) to be satisfied.

Denote by $w_h(\cdot)$ the modulus of continuity of the function h on $\mathbb{R}$, that is:

$$w_h(t) = \sup_{|x-y|\leq t} |h(x) - h(y)|\,.$$

Corollary 3.1 *Assume that*

$$\sum_{k\geq 0} \|w_h(|a_k \varepsilon_0|)\|_2 < \infty\,, \tag{3.2}$$

or

$$\sum_{k\geq 1} \frac{\big\|w_h\big(\sum_{\ell\geq 0} |a_{k+\ell}||\varepsilon_{-\ell}|\big)\big\|_2}{k^{1/2}} < \infty\,. \tag{3.3}$$

Then, provided that $c(n) = N/n \to c \in (0,\infty)$, the conclusion of Theorem 2.1 holds for $F^{\mathbf{B}_n}$ where $\mathbf{B}_n$ is the sample covariance matrix of dimension N defined by (2.2) *and associated with $(X_k)_{k\in\mathbf{Z}}$ defined by (3.1).*

Example 1. Assume that h is γ-Hölder with $\gamma \in]0,1]$, that is: there is a positive constant C such that $w_h(t) \leq C|t|^\gamma$. Assume that

$$\sum_{k\geq 0} |a_k|^\gamma < \infty \quad \text{and} \quad \mathbb{E}(|\varepsilon_0|^{(2\gamma)\vee 1}) < \infty\,,$$

then the condition (3.2) is satisfied and the conclusion of Corollary 3.1 holds. In particular, when h is the identity, which corresponds to the fact that X_k is a causal linear process, the conclusion of Corollary 3.1 holds as soon as $\sum_{k\geq 0} |a_k| < \infty$ and ε_0 belongs to $\mathbb{L}^2$. This improves Theorem 2.5 in Bai and Zhou (2008) and Theorem 1 in Yao (2012) that require ε_0 to be in $\mathbb{L}^4$.

Example 2. Assume $\|\varepsilon_0\|_\infty \leq M$ where M is a finite positive constant, and that $|a_k| \leq C\rho^k$ where $\rho \in (0,1)$ and C is a finite positive constant, then the condition (3.3) is satisfied and the conclusion of Corollary 3.1 holds as soon as $\sum_{k\geq 1} k^{-1/2} w_h\big(\rho^k MC(1-\rho)^{-1}\big) < \infty$. Using the usual comparison between series and integrals, it follows that the latter condition is equivalent to

$$\int_0^1 \frac{w_h(t)}{t\sqrt{|\log t|}} dt < \infty \,. \tag{3.4}$$

For instance if $w_h(t) \leq C|\log t|^{-\alpha}$ with $\alpha > 1/2$ near zero, then the above condition is satisfied.

Let us now consider the special case of functionals of Bernoulli shifts (also called Raikov or Riesz-Raikov sums). Let $(\varepsilon_k)_{k\in\mathbb{Z}}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(\varepsilon_0 = 1) = \mathbb{P}(\varepsilon_0 = 0) = 1/2$ and let, for any $k \in \mathbb{Z}$,

$$Y_k = \sum_{i\geq 0} 2^{-i-1}\varepsilon_{k-i} \ \text{ and } \ X_k = h(Y_k) - \int_0^1 h(x)dx \,, \tag{3.5}$$

where $h \in \mathbb{L}^2([0,1])$, $[0,1]$ being equipped with the Lebesgue measure. Recall that Y_n, $n \geq 0$, is an ergodic stationary Markov chain taking values in $[0,1]$, whose stationary initial distribution is the restriction of Lebesgue measure to $[0,1]$. As we have seen previously, if h has a modulus of continuity satisfying (3.4), then the conclusion of Theorem 2.1 holds for the sample covariance matrix associated with such a functional of Bernoulli shifts. Since for Bernoulli shifts, the computations can be done explicitly, we can even derive an alternative condition to (3.4), still in terms of regularity of h, in such a way that (2.3) holds.

Corollary 3.2 . *Assume that*

$$\int_0^1\int_0^1 (h(x)-h(y))^2 \frac{1}{|x-y|}\Big(\log\Big(\log\frac{1}{|x-y|}\Big)\Big)^t dxdy < \infty \,, \tag{3.6}$$

for some $t > 1$. Then, provided that $c(n) = N/n \to c \in (0,\infty)$, the conclusion of Theorem 2.1 holds for $F^{\mathbf{B}_n}$ where $\mathbf{B}_n$ is the sample covariance matrix of dimension N defined by (2.2) *and associated with $(X_k)_{k\in\mathbf{Z}}$ defined by (3.5).*

As a concrete example of a map satisfying (3.6), we can consider the function

$$g(x) = \frac{1}{\sqrt{x}}\frac{1}{(1+\log(2/x))^4}\sin\Big(\frac{1}{x}\Big) \ , \ 0 < x < 1$$

(see the computations pages 23-24 in Merlevède *et al* (2006) showing that the above function satisfies (3.6)).

Proof of Corollary 3.1. To prove the corollary, it suffices to show that the condition (2.3) is satisfied as soon as (3.2) or (3.3) holds. Let $(\varepsilon_k^*)_{k\in\mathbb{Z}}$ be an independent copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$. Denoting by $\mathbb{E}_\varepsilon(\cdot)$ the conditional expectation with respect to $\varepsilon = (\varepsilon_k)_{k\in\mathbb{Z}}$, we have that, for any $k \geq 0$,

$$\begin{aligned}\|P_0(X_k)\|_2 &= \Big\|\mathbb{E}_\varepsilon\Big(h\Big(\sum_{i=0}^{k-1} a_i\varepsilon_{k-i}^* + \sum_{i\geq k} a_i\varepsilon_{k-i}\Big) - h\Big(\sum_{i=0}^{k} a_i\varepsilon_{k-i}^* + \sum_{i\geq k+1} a_i\varepsilon_{k-i}\Big)\Big)\Big\|_2 \\ &\leq \|w_h\big(\big|a_k(\varepsilon_0 - \varepsilon_0^*)\big|\big)\|_2 \,.\end{aligned}$$

Next, by the subadditivity of $w_h(\cdot)$, $w_h(|a_k(\varepsilon_0 - \varepsilon_0^*)|) \leq w_h(|a_k\varepsilon_0|) + w_h(|a_k\varepsilon_0^*|)$. Whence, $\|P_0(X_k)\|_2 \leq 2\|w_h(|a_k\varepsilon_0|)\|_2$. This proves that the condition (2.3) is satisfied under (3.2).

We prove now that if (3.3) holds then so does the condition (2.3). According to Remark 2.3, it suffices to prove that the first part of (2.5) is satisfied. With the same notations as before, we have that, for any $\ell \geq 0$,

$$\mathbb{E}(X_\ell|\xi_0) = \mathbb{E}_\varepsilon\Big(h\Big(\sum_{i=0}^{\ell-1} a_i\varepsilon_{\ell-i}^* + \sum_{i\geq \ell} a_i\varepsilon_{\ell-i}\Big) - h\Big(\sum_{i\geq 0} a_i\varepsilon_{\ell-i}^*\Big)\Big) \,.$$

Hence, for any non-negative integer ℓ,

$$\|\mathbb{E}(X_\ell|\xi_0)\|_2 \le \Big\| w_h\Big(\sum_{i\ge\ell}|a_i(\varepsilon_{\ell-i}-\varepsilon^*_{\ell-i})|\Big)\Big\|_2 \le 2\Big\| w_h\Big(\sum_{i\ge\ell}|a_i||\varepsilon_{\ell-i}|\Big)\Big\|_2 ,$$

where we have used the subadditivity of $w_h(\cdot)$ for the last inequality. This latter inequality entails that the first part of (2.5) holds as soon as (3.3) does. □

Proof of Corollary 3.2. By Remark 2.3, it suffices to prove that the second part of (2.5) is satisfied as soon as (3.6) is. Actually we shall prove that (3.6) implies that

$$\sum_{n\ge1}(\log n)^t\|X_n-\mathbb{E}(X_n|\mathcal{F}_1^n)\|_2^2<\infty , \tag{3.7}$$

which clearly entails the second part of (2.5) since $t>1$. An upper bound for the quantity $\|X_n-\mathbb{E}(X_n|\mathcal{F}_1^n)\|_2^2$ has been obtained in Ibragimov and Linnik (1971, Chapter 19.3). Setting $A_{jn}=[j2^{-n},(j+1)2^{-n})$ for $j=0,1,\ldots,2^n-1$, they obtained (see the pages 372-373 of their monograph) that

$$\|X_n-\mathbb{E}(X_n|\mathcal{F}_1^n)\|_2^2 \le 2^n\sum_{j=0}^{2^n-1}\int_{A_{j,n}}\int_{A_{j,n}}(h(x)-h(y))^2dxdy .$$

Since

$$\sum_{j=0}^{2^n-1}\int_{A_{j,n}}\int_{A_{j,n}}(h(x)-h(y))^2dxdy \le \int_0^1\int_0^1(h(x)-h(y))^2\mathbf{1}_{|x-y|\le 2^{-n}}dxdy ,$$

it follows that

$$\sum_{n\ge1}(\log n)^t\|X_n-\mathbb{E}(X_n|\mathcal{F}_1^n)\|_2^2 \le \int_0^1\int_0^1\sum_{n:2^{-n}\ge|x-y|}2^n(\log n)^t(h(x)-h(y))^2\mathbf{1}_{|x-y|\le 2^{-n}}dxdy .$$

This latter inequality together with the fact that for any $u\in(0,1)$, $\sum_{n:2^{-n}\ge u}(\log n)^t \le Cu^{-1}(\log(\log u^{-1}))^t$ for some positive constant C, prove that (3.7) holds under (3.6). □

3.2 ARCH models

Let $(\varepsilon_k)_{k\in\mathbb{Z}}$ be an i.i.d. sequence of zero mean real-valued random variables such that $\|\varepsilon_0\|_2=1$. We consider the following ARCH(∞) model described by Giraitis *et al.* (2000):

$$Y_k=\sigma_k\varepsilon_k \quad \text{where} \quad \sigma_k^2=a+\sum_{j\ge1}a_jY_{k-j}^2 , \tag{3.8}$$

where $a\ge0$, $a_j\ge0$ and $\sum_{j\ge1}a_j<1$. Such models are encountered when the volatility $(\sigma_k^2)_{k\in\mathbb{Z}}$ is unobserved. In that case, the process of interest is $(Y_k^2)_{k\in\mathbb{Z}}$ and, in what follows, we consider the process $(X_k)_{k\in\mathbb{Z}}$ defined, for any $k\in\mathbb{Z}$, by:

$$X_k=Y_k^2-\mathbb{E}(Y_k^2) \quad \text{where } Y_k \text{ is defined in (3.8).} \tag{3.9}$$

Notice that, under the above conditions, there exists a unique stationary solution of equation (3.8) satisfying (see Giraitis *et al.* (2000)):

$$\sigma_k^2=a+a\sum_{\ell=1}^{\infty}\sum_{j_1,\ldots,j_\ell=1}^{\infty}a_{j_1}\ldots a_{j_\ell}\varepsilon_{k-j_1}^2\ldots\varepsilon_{k-(j_1+\cdots+j_\ell)}^2 . \tag{3.10}$$

Corollary 3.3 *Assume that ε_0 belongs to $\mathbb{L}^4$ and that*

$$\|\varepsilon_0\|_4^2 \sum_{j\geq 1} a_j < 1 \text{ and } \sum_{j\geq n} a_j = O(n^{-b}) \text{ for some } b > 1/2 . \tag{3.11}$$

Then, provided that $c(n) = N/n \to c \in (0, \infty)$, the conclusion of Theorem 2.1 holds for $F^{\mathbf{B}_n}$ where $\mathbf{B}_n$ is the sample covariance matrix of dimension N defined by (2.2) *and associated with $(X_k)_{k\in\mathbb{Z}}$ defined by (3.9).*

Proof of Corollary 3.3. By Remark 2.3, it suffices to prove that the first part of (2.5) is satisfied as soon as (3.11) is. With this aim, let us notice that, for any integer $n \geq 1$,

$$\begin{aligned}
\|\mathbb{E}(X_n|\xi_0)\|_2 &= \|\varepsilon_0\|_4^2 \|\mathbb{E}(\sigma_n^2|\xi_0) - \mathbb{E}(\sigma_n^2)\|_2 \\
&\leq 2a\|\varepsilon_0\|_4^2 \Big\| \sum_{\ell=1}^{\infty} \sum_{j_1,\dots,j_\ell=1}^{\infty} a_{j_1} \dots a_{j_\ell} \varepsilon_{n-j_1}^2 \dots \varepsilon_{n-(j_1+\dots+j_\ell)}^2 \mathbf{1}_{j_1+\dots+j_\ell \geq n} \Big\|_2 \\
&\leq 2a\|\varepsilon_0\|_4^2 \sum_{\ell=1}^{\infty} \sum_{j_1,\dots,j_\ell=1}^{\infty} \sum_{k=1}^{\ell} a_{j_1} \dots a_{j_\ell} \mathbf{1}_{j_k \geq [n/\ell]} \|\varepsilon_0\|_4^{2\ell} \leq 2a\|\varepsilon_0\|_4^2 \sum_{\ell=1}^{\infty} \ell \kappa^{\ell-1} \sum_{k=[n/\ell]}^{\infty} a_k \,,
\end{aligned}$$

where $\kappa = \|\varepsilon_0\|_4^2 \sum_{j\geq 1} a_j$. So, under (3.11), there exists a positive constant C not depending on n such that $\|\mathbb{E}(X_n|\xi_0)\|_2 \leq C n^{-b}$. This upper bound implies that the first part of (2.5) is satisfied as soon as $b > 1/2$. □

Remark 3.4 Notice that if we consider the sample covariance matrix associated with $(Y_k)_{k\in\mathbb{Z}}$ defined in (3.8), then its LSD follows directly by Theorem 2.1 since $P_0(Y_k) = 0$, for any positive integer k.

4 Proof of Theorem 2.1

To prove the theorem it suffices to show that for any $z \in \mathbb{C}^+$,

$$S_{F^{\mathbf{B}_n}}(z) \to S(z) \text{ almost surely.} \tag{4.1}$$

Since the columns of $\mathcal{X}_n$ are independent, by Step 1 of the proof of Theorem 1.1 in Bai and Zhou (2008), to prove (4.1), it suffices to show that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big) = S(z) \,, \tag{4.2}$$

where $S(z)$ satisfies the equation (2.4).

The proof of (4.2) being very technical, for reader convenience, let us describe the different steps leading to it. We shall consider a sample covariance matrix $\mathbf{G}_n := \frac{1}{n}\mathcal{Z}_n\mathcal{Z}_n^T$ (see (4.32)) such that the columns of $\mathcal{Z}_n$ are independent and the random variables in each column of $\mathcal{Z}_n$ form a sequence of Gaussian random variables whose covariance structure is the same as that of the sequence $(X_k)_{k\in\mathbb{Z}}$ (see Section 4.2). The aim will be then to prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \big|\mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)\big| = 0 \,, \tag{4.3}$$

and

$$\lim_{n\to\infty} \mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big) = S(z) \,. \tag{4.4}$$

The proof of (4.4) will be achieved in Section 4.4 with the help of Theorem 1.1 in Silverstein (1995) combined with arguments developed in the proof of Theorem 1 in Yao (2012). The proof of (4.3) will be divided in several steps. First, to "break" the dependence structure, we introduce a parameter m, and approximate $\mathbf{B}_n$ by a sample covariance matrix $\bar{\mathbf{B}}_n := \frac{1}{n}\bar{\mathcal{X}}_n\bar{\mathcal{X}}_n^T$ (see (4.16)) such that the columns of $\bar{\mathcal{X}}_n$ are independent and the random variables in each column of $\bar{\mathcal{X}}_n$ form of an m-dependent sequence of random variables bounded by $2M$, with M a positive real (see Section 4.1). This approximation will be done in such a way that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{M\to\infty}\limsup_{n\to\infty}\Big|\mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)\Big| = 0\,. \tag{4.5}$$

Next, the sample Gaussian covariance matrix $\mathbf{G}_n$ is approximated by another sample Gaussian covariance matrix $\widetilde{\mathbf{G}}_n$ (see (4.34)), depending on the parameter m and constructed from $\mathbf{G}_n$ by replacing some of the variables in each column of $\mathcal{Z}_n$ by zeros (see Section 4.2). This approximation will be done in such a way that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\Big|\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big)\Big| = 0\,. \tag{4.6}$$

In view of (4.5) and (4.6), the convergence (4.3) will then follow if we can prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{M\to\infty}\limsup_{n\to\infty}\Big|\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big)\Big| = 0\,. \tag{4.7}$$

This will be achieved in Section 4.3 with the help of the Lindeberg method. The rest of this section is devoted to the proofs of the convergences (4.3)-(4.7).

4.1 Approximation by a sample covariance matrix associated with an m-dependent sequence.

Let $N \geq 2$ and m be a positive integer fixed for the moment and assumed to be less than $\sqrt{N/2}$. Set

$$k_{N,m} = \left[\frac{N}{m^2+m}\right]\,, \tag{4.8}$$

where we recall that $[\,\cdot\,]$ denotes the integer part. Let M be a fixed positive number that depends neither on N, nor on n, nor on m. Let φ_M be the function defined by $\varphi_M(x) = (x \wedge M) \vee (-M)$. Now for any $k \in \mathbb{Z}$ and $i \in \{1,\dots,n\}$ let

$$\widetilde{X}^{(i)}_{k,M,m} = \mathbb{E}\Big(\varphi_M(X^{(i)}_k)|\varepsilon^{(i)}_k,\dots,\varepsilon^{(i)}_{k-m}\Big) \text{ and } \bar{X}^{(i)}_{k,M,m} = \widetilde{X}^{(i)}_{k,M,m} - \mathbb{E}\big(\widetilde{X}^{(i)}_{k,M,m}\big)\,. \tag{4.9}$$

In what follows, to soothe the notations, we shall write $\widetilde{X}^{(i)}_{k,m}$ and $\bar{X}^{(i)}_{k,m}$ instead of respectively $\widetilde{X}^{(i)}_{k,M,m}$ and $\bar{X}^{(i)}_{k,M,m}$, when no confusion is allowed. Notice that $\big(\bar{X}^{(1)}_{k,m}\big)_{k\in\mathbb{Z}},\dots,\big(\bar{X}^{(n)}_{k,m}\big)_{k\in\mathbb{Z}}$ are n independent copies of the centered and stationary sequence $\big(\bar{X}_{k,m}\big)_{k\in\mathbb{Z}}$ defined by

$$\bar{X}_{k,m} = \widetilde{X}_{k,m} - \mathbb{E}\big(\widetilde{X}_{k,m}\big) \text{ where } \widetilde{X}_{k,m} = \mathbb{E}\Big(\varphi_M(X_k)|\varepsilon_k,\dots,\varepsilon_{k-m}\Big)\,,\ k \in \mathbb{Z}\,. \tag{4.10}$$

This implies in particular that: for any $i \in \{1,\dots,n\}$ and any $k \in \mathbb{Z}$,

$$\|\bar{X}^{(i)}_{k,m}\|_\infty = \|\bar{X}_{k,m}\|_\infty \leq 2M\,. \tag{4.11}$$

For any $i \in \{1,\dots,n\}$, note that $\big(\bar{X}^{(i)}_{k,m}\big)_{k\in\mathbb{Z}}$ forms an m-dependent sequence, in the sense that $\bar{X}^{(i)}_{k,m}$ and $\bar{X}^{(i)}_{k',m}$ are independent if $|k-k'| > m$. We write now the interval $[1,N]\cap\mathbb{N}$ as a union of disjoint sets as follows:

$$[1,N]\cap\mathbb{N} = \bigcup_{\ell=1}^{k_{N,m}+1} I_\ell \cup J_\ell\,,$$

where, for $\ell \in \{1, \dots, k_{N,m}\}$,

$$I_\ell := \big[(\ell-1)(m^2+m)+1\,,\,(\ell-1)(m^2+m)+m^2\big] \cap \mathbb{N}, \tag{4.12}$$
$$J_\ell := \Big[(\ell-1)(m^2+m)+m^2+1\,,\,\ell(m^2+m)\Big] \cap \mathbb{N}\,,$$

and, for $\ell = k_{N,m}+1$,

$$I_{k_{N,m}+1} = \big[k_{N,m}(m^2+m)+1\,,\,N\big] \cap \mathbb{N}\,,$$

and $J_{k_{N,m}+1} = \emptyset$. Note that $I_{k_{N,m}+1} = \emptyset$ if $k_{N,m}(m^2+m) = N$.

Let now $\big(\mathbf{u}_\ell^{(i)}\big)_{\ell \in \{1,\dots,k_{N,m}\}}$ be the random vectors defined as follows. For any ℓ belonging to $\{1, \dots, k_{N,m}-1\}$,

$$\mathbf{u}_\ell^{(i)} = \Big(\big(\bar{X}_{k,m}^{(i)}\big)_{k \in I_\ell}, \mathbf{0}_m \Big)\,. \tag{4.13}$$

Hence, the dimension of the random vectors defined above is equal to m^2+m. Now, for $\ell = k_{N,m}$, we set

$$\mathbf{u}_{k_{N,m}}^{(i)} = \Big(\big(\bar{X}_{k,m}^{(i)}\big)_{k \in I_{k_{N,m}}}, \mathbf{0}_r \Big)\,, \tag{4.14}$$

where $r = m + N - k_{N,m}(m^2+m)$. This last vector is then of dimension $N - (k_{N,m}-1)(m^2+m)$. Notice that the random vectors $\big(\mathbf{u}_\ell^{(i)}\big)_{1 \le i \le n, 1 \le \ell \le k_{N,m}}$ are mutually independent.

For any $i \in \{1, \dots, n\}$, we define now row random vectors $\bar{\mathbf{X}}^{(i)}$ of dimension N by setting

$$\bar{\mathbf{X}}^{(i)} = \big(\mathbf{u}_\ell^{(i)}\,,\, \ell = 1, \dots, k_{N,m}\big)\,, \tag{4.15}$$

where the $\mathbf{u}_\ell^{(i)}$'s are defined in (4.13) and (4.14). Let

$$\bar{\mathcal{X}}_n = \big(\bar{\mathbf{X}}^{(1)T}| \dots |\bar{\mathbf{X}}^{(n)T}\big) \quad \text{and} \quad \bar{\mathbf{B}}_n = \frac{1}{n}\bar{\mathcal{X}}_n \bar{\mathcal{X}}_n^T\,. \tag{4.16}$$

In what follows, we shall prove the following proposition.

Proposition 4.1 *For any $z \in \mathbb{C}^+$, the convergence* (4.5) *holds true with $\mathbf{B}_n$ and $\bar{\mathbf{B}}_n$ as defined in* (2.2) *and* (4.16) *respectively.*

To prove the proposition above, we start by noticing that, by integration by parts, for any $z = u + iv \in \mathbb{C}^+$,

$$\Big|\mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)\Big| \le \mathbb{E}\Big| \int \frac{1}{x-z} dF^{\mathbf{B}_n}(x) - \int \frac{1}{x-z} dF^{\bar{\mathbf{B}}_n}(x)\Big|$$
$$= \mathbb{E}\Big| \int \frac{F^{\mathbf{B}_n}(x) - F^{\bar{\mathbf{B}}_n}(x)}{(x-z)^2} dx\Big| \le \frac{1}{v^2}\,\mathbb{E}\int \big|F^{\mathbf{B}_n}(x) - F^{\bar{\mathbf{B}}_n}(x)\big| dx\,. \tag{4.17}$$

Now, $\int \big|F^{\mathbf{B}_n}(x) - F^{\bar{\mathbf{B}}_n}(x)\big| dx$ is nothing else but the Wasserstein distance of order 1 between the empirical measure of $\mathbf{B}_n$ and that of $\bar{\mathbf{B}}_n$. To be more precise, if $\lambda_1, \dots, \lambda_N$ denote the eigenvalues of $\mathbf{B}_n$ in the non-increasing order, and $\bar{\lambda}_1, \dots, \bar{\lambda}_N$ the ones of $\bar{\mathbf{B}}_n$, also in the non-increasing order, then, setting $\eta_n = \frac{1}{N}\sum_{k=1}^N \delta_{\lambda_k}$ and $\bar{\eta}_n = \frac{1}{N}\sum_{k=1}^N \delta_{\bar{\lambda}_k}$, we have that

$$\int \big|F^{\mathbf{B}_n}(x) - F^{\bar{\mathbf{B}}_n}(x)\big| dx = W_1(\eta_n, \bar{\eta}_n) = \inf \mathbb{E}|X - Y|\,,$$

where the infimum runs over the set of couples of random variables (X, Y) on $\mathbb{R} \times \mathbb{R}$ such that $X \sim \eta_n$ and $Y \sim \bar{\eta}_n$. Arguing as in Remark 4.2.6 in Chafaï *et al* (2012), we have

$$W_1(\eta_n, \bar{\eta}_n) = \frac{1}{N} \min_{\pi \in \mathcal{S}_N} \sum_{k=1}^{N \wedge n} |\lambda_k - \bar{\lambda}_{\pi(k)}|\,,$$

where π is a permutation belonging to the symmetric group $\mathcal{S}_N$ of $\{1,\dots,N\}$. By standard arguments, involving the fact that if x,y,u,v are real numbers such that $x\le y$ and $u>v$, then $|x-u|+|y-v|\ge|x-v|+|y-u|$, we get that $\min_{\pi\in\mathcal{S}_N}\sum_{k=1}^{N\wedge n}|\lambda_k-\bar\lambda_{\pi(k)}|=\sum_{k=1}^{N\wedge n}|\lambda_k-\bar\lambda_k|$. Therefore,

$$W_1(\eta_n,\bar\eta_n)=\int\big|F^{\mathbf{B}_n}(x)-F^{\bar{\mathbf{B}}_n}(x)\big|dx=\frac{1}{N}\sum_{k=1}^{N\wedge n}|\lambda_k-\bar\lambda_k|\,. \tag{4.18}$$

Notice that $\lambda_k=s_k^2$ and $\bar\lambda_k=\bar s_k^2$ where the s_k's (respectively the $\bar s_k$'s) are the singular values of the matrix $n^{-1/2}\mathcal{X}_n$ (respectively of $n^{-1/2}\bar{\mathcal{X}}_n$). Hence, by Cauchy-Schwarz's inequality,

$$\sum_{k=1}^{N\wedge n}|\lambda_k-\bar\lambda_k|\le\Big(\sum_{k=1}^{N\wedge n}\big|s_k+\bar s_k\big|^2\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}\big|s_k-\bar s_k\big|^2\Big)^{1/2}$$
$$\le 2^{1/2}\Big(\sum_{k=1}^{N\wedge n}\big(s_k^2+\bar s_k^2\big)\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}\big|s_k-\bar s_k\big|^2\Big)^{1/2}\le 2^{1/2}\Big(\mathrm{Tr}(\mathbf{B}_n)+\mathrm{Tr}(\bar{\mathbf{B}}_n)\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}\big|s_k-\bar s_k\big|^2\Big)^{1/2}.$$

Next, by Hoffman-Wielandt's inequality (see e.g. Corollary 7.3.8 in Horn and Johnson (1985)),

$$\sum_{k=1}^{N\wedge n}\big|s_k-\bar s_k\big|^2\le n^{-1}\mathrm{Tr}\big((\mathcal{X}_n-\bar{\mathcal{X}}_n)(\mathcal{X}_n-\bar{\mathcal{X}}_n)^T\big)\,.$$

Therefore,

$$\sum_{k=1}^{N\wedge n}|\lambda_k-\bar\lambda_k|\le 2^{1/2}n^{-1/2}\Big(\mathrm{Tr}(\mathbf{B}_n)+\mathrm{Tr}(\bar{\mathbf{B}}_n)\Big)^{1/2}\Big(\mathrm{Tr}\big((\mathcal{X}_n-\bar{\mathcal{X}}_n)(\mathcal{X}_n-\bar{\mathcal{X}}_n)^T\big)\Big)^{1/2}. \tag{4.19}$$

Starting from (4.17), considering (4.18) and (4.19), and using Cauchy-Schwarz's inequality, it follows that

$$\Big|\mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big)-\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)\Big|$$
$$\le\frac{2^{1/2}}{v^2}\frac{1}{Nn^{1/2}}\|\mathrm{Tr}(\mathbf{B}_n)+\mathrm{Tr}(\bar{\mathbf{B}}_n)\|_1^{1/2}\|\mathrm{Tr}\big((\mathcal{X}_n-\bar{\mathcal{X}}_n)(\mathcal{X}_n-\bar{\mathcal{X}}_n)^T\big)\|_1^{1/2}. \tag{4.20}$$

By the definition of $\mathbf{B}_n$,

$$\frac{1}{N}\mathbb{E}\big(|\mathrm{Tr}(\mathbf{B}_n)|\big)=\frac{1}{nN}\sum_{i=1}^{n}\sum_{k=1}^{N}\|X_k^{(i)}\|_2^2=\|X_0\|_2^2\,, \tag{4.21}$$

where we have used that for each i, $\big(X_k^{(i)}\big)_{k\in\mathbb{Z}}$ is a copy of the stationary sequence $(X_k)_{k\in\mathbb{Z}}$. Now, setting

$$\mathcal{I}_{N,m}=\bigcup_{\ell=1}^{k_{N,m}}I_\ell\ \text{ and }\ \mathcal{R}_{N,m}=\{1,\dots,N\}\backslash\mathcal{I}_{N,m}\,, \tag{4.22}$$

recalling the definition (4.16) of $\bar{\mathbf{B}}_n$, using the stationarity of the sequence $(\bar X_{k,m}^{(i)})_{k\in\mathbb{Z}}$, and the fact that $\mathrm{card}(\mathcal{I}_{N,m})=m^2k_{N,m}\le N$, we get

$$\frac{1}{N}\mathbb{E}\big(|\mathrm{Tr}(\bar{\mathbf{B}}_n)|\big)=\frac{1}{nN}\sum_{i=1}^{n}\sum_{k\in\mathcal{I}_{N,m}}\|\bar X_{k,m}^{(i)}\|_2^2\le\|\bar X_{0,m}\|_2^2\,.$$

Next,

$$\|\bar{X}_{0,m}\|_2 \leq 2\|\widetilde{X}_{0,m}\|_2 \leq 2\|\varphi_M(X_0)\|_2 \leq 2\|X_0\|_2\,. \tag{4.23}$$

Therefore,

$$\frac{1}{N}\mathbb{E}\big(|\mathrm{Tr}(\bar{\mathbf{B}}_n)|\big) \leq 4\|X_0\|_2^2\,. \tag{4.24}$$

Now, by definition of $\mathcal{X}_n$ and $\bar{\mathcal{X}}_n$,

$$\frac{1}{Nn}\mathbb{E}\big(|\mathrm{Tr}\big(\mathcal{X}_n - \bar{\mathcal{X}}_n\big)\big(\mathcal{X}_n - \bar{\mathcal{X}}_n\big)^T|\big)$$
$$= \frac{1}{nN}\sum_{i=1}^{n}\sum_{k\in\mathcal{I}_{N,m}}\|X_k^{(i)} - \bar{X}_{k,m}^{(i)}\|_2^2 + \frac{1}{nN}\sum_{i=1}^{n}\sum_{k\in\mathcal{R}_{N,m}}\|X_k^{(i)}\|_2^2\,.$$

Using stationarity, the fact that $\mathrm{card}(\mathcal{I}_{N,m}) \leq N$ and

$$\mathrm{card}(\mathcal{R}_{N,m}) = N - m^2 k_{N,m} \leq \frac{N}{m+1} + m^2\,, \tag{4.25}$$

we get that

$$\frac{1}{Nn}\mathbb{E}\big(|\mathrm{Tr}\big(\mathcal{X}_n - \bar{\mathcal{X}}_n\big)\big(\mathcal{X}_n - \bar{\mathcal{X}}_n\big)^T|\big) \leq \|X_0 - \bar{X}_{0,m}\|_2^2 + (m^{-1} + m^2N^{-1})\|X_0\|_2^2\,. \tag{4.26}$$

Starting from (4.20), considering the upper bounds (4.21), (4.24) and (4.26), we derive that there exists a positive constant C not depending on (m, M) and such that

$$\limsup_{n\to\infty}\Big|\mathbb{E}\big(S_{F^{\mathbf{B}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)\Big| \leq \frac{C}{v^2}\big(\|X_0 - \bar{X}_{0,m}\|_2 + m^{-1/2}\big)\,.$$

Therefore, Proposition 4.1 will follow if we can prove that

$$\lim_{m\to\infty}\limsup_{M\to\infty}\|X_0 - \bar{X}_{0,m}\|_2 = 0\,. \tag{4.27}$$

Let us introduce now the sequence $(X_{k,m})_{k\in\mathbb{Z}}$ defined as follows: for any $k \in \mathbb{Z}$,

$$X_{k,m} = \mathbb{E}\big(X_k|\varepsilon_k, \ldots, \varepsilon_{k-m}\big)\,. \tag{4.28}$$

With the above notation, we write that

$$\|X_0 - \bar{X}_{0,m}\|_2 \leq \|X_0 - X_{0,m}\|_2 + \|X_{0,m} - \bar{X}_{0,m}\|_2\,.$$

Since X_0 is centered, so is $X_{0,m}$. Then $\|X_{0,m} - \bar{X}_{0,m}\|_2 = \|X_{0,m} - \mathbb{E}(X_{0,m}) - \bar{X}_{0,m}\|_2$. Therefore, recalling the definition (4.10) of $\bar{X}_{0,m}$, it follows that

$$\|X_{0,m} - \bar{X}_{0,m}\|_2 \leq 2\|X_{0,m} - \widetilde{X}_{0,m}\|_2 \leq 2\|X_0 - \varphi_M(X_0)\|_2 \leq 2\|\big(|X_0| - M\big)_+\|_2\,. \tag{4.29}$$

Since X_0 belongs to $\mathbb{L}^2$, $\lim_{M\to\infty}\|\big(|X_0| - M\big)_+\|_2 = 0$. Therefore, to prove (4.27) (and then Proposition 4.1), it suffices to prove that

$$\lim_{m\to\infty}\|X_0 - X_{0,m}\|_2 = 0\,. \tag{4.30}$$

Since $(X_{0,m})_{m\geq 0}$ is a martingale with respect to the increasing filtration $(\mathcal{G}_m)_{m\geq 0}$ defined by $\mathcal{G}_m = \sigma(\varepsilon_{-m}, \ldots, \varepsilon_0)$, and is such that $\sup_{m\geq 0}\|X_{0,m}\|_2 \leq \|X_0\|_2 < \infty$, (4.30) follows by the martingale convergence theorem in $\mathbb{L}^2$ (see for instance Corollary 2.2 in Hall and Heyde (1980)). This ends the proof of Proposition 4.1. $\square$

4.2 Construction of approximating sample covariance matrices associated with Gaussian random variables.

Let $(Z_k)_{k\in\mathbb{Z}}$ be a centered Gaussian process with real values, whose covariance function is given, for any $k,\ell\in\mathbb{Z}$, by

$$\mathrm{Cov}(Z_k,Z_\ell)=\mathrm{Cov}(X_k,X_\ell)\,. \tag{4.31}$$

For n a positive integer, we consider n independent copies of the Gaussian process $(Z_k)_{k\in\mathbb{Z}}$ that are in addition independent of $(X_k^{(i)})_{k\in\mathbb{Z},i\in\{1,\dots,n\}}$. We shall denote these copies by $(Z_k^{(i)})_{k\in\mathbb{Z}}$ for $i=1,\dots,n$. For any $i\in\{1,\dots,n\}$, define $\mathbf{Z}_i=\big(Z_1^{(i)},\dots,Z_N^{(i)}\big)$. Let $\mathcal{Z}_n=(\mathbf{Z}_1^T|\dots|\mathbf{Z}_n^T)$ be the matrix whose columns are the $\mathbf{Z}_i^T$'s and consider its associated sample covariance matrix

$$\mathbf{G}_n=\frac{1}{n}\mathcal{Z}_n\mathcal{Z}_n^T\,. \tag{4.32}$$

For $k_{N,m}$ given in (4.8), we define now the random vectors $\big(\mathbf{v}_\ell^{(i)}\big)_{\ell\in\{1,\dots,k_{N,m}\}}$ as follows. They are defined as the random vectors $\big(\mathbf{u}_\ell^{(i)}\big)_{\ell\in\{1,\dots,k_{N,m}\}}$ defined in (4.13) and (4.14), but by replacing each $\bar{X}_{k,m}^{(i)}$ by $Z_k^{(i)}$. For any $i\in\{1,\dots,n\}$, we then define the random vectors $\widetilde{\mathbf{Z}}^{(i)}$ of dimension N, as follows:

$$\widetilde{\mathbf{Z}}^{(i)}=\big(\mathbf{v}_\ell^{(i)}\,,\,\ell=1,\dots,k_{N,m}\big)\,. \tag{4.33}$$

Let now

$$\widetilde{\mathcal{Z}}_n=\big(\widetilde{\mathbf{Z}}^{(1)T}|\dots|\widetilde{\mathbf{Z}}^{(n)T}\big)\ \text{ and }\ \widetilde{\mathbf{G}}_n=\frac{1}{n}\widetilde{\mathcal{Z}}_n\widetilde{\mathcal{Z}}_n^T\,. \tag{4.34}$$

In what follows, we shall prove the following proposition.

Proposition 4.2 *For any $z\in\mathbb{C}^+$, the convergence* (4.6) *holds true with $\mathbf{G_n}$ and $\widetilde{\mathbf{G}}_n$ as defined in* (4.32) *and* (4.34) *respectively.*

To prove the proposition above, we start by noticing that, for any $z=u+iv\in\mathbb{C}^+$,

$$\begin{aligned}\big|S_{F^{\mathbf{G}_n}}(z)-S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big|&=\Big|\int\frac{1}{x-z}dF^{\mathbf{G}_n}(x)-\int\frac{1}{x-z}dF^{\widetilde{\mathbf{G}}_n}(x)\Big|\\&\le\Big|\int\frac{F^{\mathbf{G}_n}(x)-F^{\widetilde{\mathbf{G}}_n}(x)}{(x-z)^2}dx\Big|\le\frac{\pi\,\big\|F^{\mathbf{G}_n}-F^{\widetilde{\mathbf{G}}_n}\big\|_\infty}{v}\,.\end{aligned}$$

Hence, by Theorem A.44 in Bai and Silverstein (2010),

$$\Big|\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)-\mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big)\Big|\le\frac{\pi}{vN}\mathrm{rank}\big(\mathcal{Z}_n-\widetilde{\mathcal{Z}}_n\big)\,.$$

By definition of $\mathcal{Z}_n$ and $\widetilde{\mathcal{Z}}_n$, $\mathrm{rank}\big(\mathcal{Z}_n-\widetilde{\mathcal{Z}}_n\big)\le\mathrm{card}(\mathcal{R}_{N,m})$, where $\mathcal{R}_{N,m}$ is defined in (4.22). Therefore, using (4.25), we get that, for any $z=u+iv\in\mathbb{C}^+$,

$$\Big|\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)-\mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big)\Big|\le\frac{\pi}{vN}\Big(\frac{N}{m+1}+m^2\Big)\,,$$

which converges to zero by letting n first tend to infinity and after m. This ends the proof of Proposition 4.2. □

4.3 Approximation of $\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big)$ by $\mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big)$.

In this section, we shall prove the following proposition.

Proposition 4.3 *Under the assumptions of Theorem 2.1, for any $z \in \mathbb{C}^+$, the convergence (4.7) holds true with $\bar{\mathbf{B}}_n$ and $\widetilde{\mathbf{G}}_n$ as defined in (4.16) and (4.34) respectively.*

With this aim, we shall use the Lindeberg method that is based on telescoping sums. In order to develop it, we first give the following definition:

Definition 4.1 *Let x be a vector of $\mathbb{R}^{nN}$ with coordinates*

$$x = \big(x^{(1)}, \dots, x^{(n)}\big) \text{ where for any } i \in \{1, \dots, n\},\ x^{(i)} = \big(x_k^{(i)}\,,\ k \in \{1, \dots, N\}\big)\,.$$

Let $z \in \mathbb{C}^+$ and $f := f_z$ be the function defined from $\mathbb{R}^{nN}$ to $\mathbb{C}$ by

$$f(x) = \frac{1}{N}\operatorname{Tr}\big(A(x) - z\mathbf{I}\big)^{-1} \quad \text{where} \quad A(x) = \frac{1}{n}\sum_{k=1}^{n}(x^{(k)})^T x^{(k)}\,, \tag{4.35}$$

and $\mathbf{I}$ is the identity matrix.

The function f, as defined above, admits partial derivatives of all orders. Indeed, let u be one of the coordinates of the vector x and $A_u = A(x)$ the matrix-valued function of the scalar u. Then, setting $G_u = \big(A_u - z\mathbf{I}\big)^{-1}$ and differentiating both sides of the equality $G_u(A_u - z\mathbf{I}) = \mathbf{I}$, it follows that

$$\frac{dG}{du} = -G\frac{dA}{du}G\,, \tag{4.36}$$

(see the equality (17) in Chatterjee (2006)). Higher-order derivatives may be computed by applying repeatedly the above formula. Upper bounds for some partial derivarives up to the fourth order are given in Appendix.

Now, using Definition 4.1 and the notations (4.15) and (4.33), we get that, for any $z \in \mathbb{C}^+$,

$$\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big) = \mathbb{E}f\big(\bar{\mathbf{X}}^{(1)}, \dots, \bar{\mathbf{X}}^{(n)}\big) - \mathbb{E}f\big(\widetilde{\mathbf{Z}}^{(1)}, \dots, \widetilde{\mathbf{Z}}^{(n)}\big)\,. \tag{4.37}$$

To continue the development of the Lindeberg method, we introduce additional notations. For any $i \in \{1, \dots, n\}$ and $k_{N,m}$ given in (4.8), we define the random vectors $\big(\mathbf{U}_\ell^{(i)}\big)_{\ell \in \{1, \dots, k_{N,m}\}}$ of dimension nN as follows. For any $\ell \in \{1, \dots, k_{N,m}\}$,

$$\mathbf{U}_\ell^{(i)} = \Big(\mathbf{0}_{(i-1)N}\,,\ \mathbf{0}_{(\ell-1)(m^2+m)}\,,\ \mathbf{u}_\ell^{(i)}\,,\ \mathbf{0}_{r_\ell}, \mathbf{0}_{(n-i)N}\Big)\,, \tag{4.38}$$

where the $\mathbf{u}_\ell^{(i)}$'s are defined in (4.13) and (4.14), and

$$r_\ell = N - \ell(m^2 + m) \ \text{ for } \ell \in \{1, \dots, k_{N,m} - 1\}, \text{ and } \ r_{k_{N,m}} = 0\,. \tag{4.39}$$

Note that the vectors $\big(\mathbf{U}_\ell^{(i)}\big)_{1 \le i \le n, 1 \le \ell \le k_{N,m}}$ are mutually independent. Moreover, with the notations (4.38) and (4.15), the following relations hold. For any $i \in \{1, \dots, n\}$,

$$\sum_{\ell=1}^{k_{N,m}} \mathbf{U}_\ell^{(i)} = \Big(\mathbf{0}_{N(i-1)}\,,\ \bar{\mathbf{X}}^{(i)}\,,\ \mathbf{0}_{(n-i)N}\Big) \ \text{ and } \ \sum_{i=1}^{n}\sum_{\ell=1}^{k_{N,m}} \mathbf{U}_\ell^{(i)} = \Big(\bar{\mathbf{X}}^{(1)}, \dots, \bar{\mathbf{X}}^{(n)}\Big)\,, \tag{4.40}$$

where the $\bar{\mathbf{X}}^{(i)}$'s are defined in (4.15).

Now, for any $i \in \{1, \ldots, n\}$, we define the random vectors $\big(\mathbf{V}_\ell^{(i)}\big)_{\ell \in \{1,\ldots,k_{N,m}\}}$ of dimension nN, as follows: for any $\ell \in \{1, \ldots, k_{N,m}\}$,

$$\mathbf{V}_\ell^{(i)} = \Big(\mathbf{0}_{(i-1)N} \,,\, \mathbf{0}_{(\ell-1)(m^2+m)} \,,\, \mathbf{v}_\ell^{(i)} \,,\, \mathbf{0}_{r_\ell}, \mathbf{0}_{(n-i)N}\Big) \,, \tag{4.41}$$

where r_ℓ is defined in (4.39) and the $\mathbf{v}_\ell^{(i)}$'s are defined in Section 4.2. With the notations (4.41) and (4.33), the following relations hold: for any $i \in \{1, \ldots, n\}$,

$$\sum_{\ell=1}^{k_{N,m}} \mathbf{V}_\ell^{(i)} = \Big(\mathbf{0}_{N(i-1)} \,,\, \widetilde{\mathbf{Z}}^{(i)} \,,\, \mathbf{0}_{N(n-i)}\Big) \text{ and } \sum_{i=1}^{n}\sum_{\ell=1}^{k_{N,m}} \mathbf{V}_\ell^{(i)} = \Big(\widetilde{\mathbf{Z}}^{(1)} \,,\ldots,\, \widetilde{\mathbf{Z}}^{(n)}\Big) \,, \tag{4.42}$$

where the $\widetilde{\mathbf{Z}}^{(i)}$'s are defined in (4.33). We define now, for any $i \in \{1, \ldots, n\}$,

$$\mathbf{S}_i = \sum_{s=1}^{i}\sum_{\ell=1}^{k_{N,m}} \mathbf{U}_\ell^{(s)} \text{ and } \mathbf{T}_i = \sum_{s=i}^{n}\sum_{\ell=1}^{k_{N,m}} \mathbf{V}_\ell^{(s)} \,, \tag{4.43}$$

and any $s \in \{1, \ldots, k_{N,m}\}$,

$$\mathbf{S}_s^{(i)} = \sum_{\ell=1}^{s} \mathbf{U}_\ell^{(i)} \text{ and } \mathbf{T}_s^{(i)} = \sum_{\ell=s}^{k_{N,m}} \mathbf{V}_\ell^{(i)} \,. \tag{4.44}$$

In all the notations above, we use the convention that $\sum_{k=r}^{s} = 0$ if $r > s$. Therefore, starting from (4.37), considering the relations (4.40) and (4.42), and using the notations (4.43) and (4.44), we successively get

$$\begin{aligned}\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big) &= \sum_{i=1}^{n}\Big(\mathbb{E}f\big(\mathbf{S}_i + \mathbf{T}_{i+1}\big) - \mathbb{E}f\big(\mathbf{S}_{i-1} + \mathbf{T}_i\big)\Big) \\ &= \sum_{i=1}^{n}\sum_{s=1}^{k_{N,m}}\Big(\mathbb{E}f\big(\mathbf{S}_{i-1} + \mathbf{S}_s^{(i)} + \mathbf{T}_{s+1}^{(i)} + \mathbf{T}_{i+1}\big) - \mathbb{E}f\big(\mathbf{S}_{i-1} + \mathbf{S}_{s-1}^{(i)} + \mathbf{T}_s^{(i)} + \mathbf{T}_{i+1}\big)\Big) \,.\end{aligned}$$

Therefore, setting for any $i \in \{1, \ldots, n\}$ and any $s \in \{1, \ldots, k_{N,m}\}$,

$$\mathbf{W}_s^{(i)} = \mathbf{S}_{i-1} + \mathbf{S}_s^{(i)} + \mathbf{T}_{s+1}^{(i)} + \mathbf{T}_{i+1} \,, \tag{4.45}$$

and

$$\widetilde{\mathbf{W}}_s^{(i)} = \mathbf{S}_{i-1} + \mathbf{S}_{s-1}^{(i)} + \mathbf{T}_{s+1}^{(i)} + \mathbf{T}_{i+1} \,, \tag{4.46}$$

we are lead to

$$\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\widetilde{\mathbf{G}}_n}}(z)\big) = \sum_{i=1}^{n}\sum_{s=1}^{k_{N,m}}\Big(\mathbb{E}\big(\Delta_s^{(i)}(f)\big) - \mathbb{E}\big(\widetilde{\Delta}_s^{(i)}(f)\big)\Big) \,, \tag{4.47}$$

where

$$\Delta_s^{(i)}(f) = f\big(\mathbf{W}_s^{(i)}\big) - f\big(\widetilde{\mathbf{W}}_s^{(i)}\big) \text{ and } \widetilde{\Delta}_s^{(i)}(f) = f\big(\mathbf{W}_{s-1}^{(i)}\big) - f\big(\widetilde{\mathbf{W}}_s^{(i)}\big) \,.$$

In order to continue the multidimensional Lindeberg method, it is useful to introduce the following notations.

Definition 4.2 *Let d_1 and d_2 be two positive integers. Let $A = (a_1, \dots, a_{d_1})$ and $B = (b_1, \dots, b_{d_2})$ be two real valued row vectors of respective dimensions d_1 and d_2. We define $A \otimes B$ as being the transpose of the Kronecker product of A by B. Therefore*

$$A \otimes B = \begin{pmatrix} a_1 B^T \\ \vdots \\ a_{d_1} B^T \end{pmatrix} \in \mathbb{R}^{d_1 d_2} .$$

For any positive integer k, the k-th transpose Kronecker power $A^{\otimes k}$ is then defined inductively by: $A^{\otimes 1} = A^T$ and $A^{\otimes k} = A \bigotimes \big(A^{\otimes(k-1)}\big)^T$.

Notice that, here, $A \otimes B$ is not exactly the usual Kronecker product (or Tensor product) of A by B that rather produces a row vector. However, for later notation convenience, the above notation is useful.

Definition 4.3 *Let d be a positive integer. If ∇ denotes the differentiation operator given by $\nabla = \big(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d}\big)$ acting on the differentiable functions $h : \mathbb{R}^d \to \mathbb{R}$, we define, for any positive integer k, $\nabla^{\otimes k}$ in the same way as in Definition 4.2. If $h : \mathbb{R}^d \to \mathbb{R}$ is k-times differentiable, for any $x \in \mathbb{R}^d$, let $D^k h(x) = \nabla^{\otimes k} h(x)$, and for any row vector Y of $\mathbb{R}^d$, we define $D^k h(x).Y^{\otimes k}$ as the usual scalar product in $\mathbb{R}^{d^k}$ between $D^k h(x)$ and $Y^{\otimes k}$. We write Dh for $D^1 h$.*

Let $z = u + iv \in \mathbb{C}^+$. We start by analyzing the term $\mathbb{E}\big(\Delta_s^{(i)}(f)\big)$ in (4.47). By Taylor's integral formula,

$$\Big|\mathbb{E}\big(\Delta_s^{(i)}(f)\big) - \mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{U}_s^{(i)\,\otimes 1}\big) - \frac{1}{2}\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{U}_s^{(i)\,\otimes 2}\big)\Big| \\ \leq \Big|\mathbb{E}\int_0^1 \frac{(1-t)^2}{2} D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big).\mathbf{U}_s^{(i)\,\otimes 3} dt\Big| . \quad (4.48)$$

Let us analyze the right-hand term of (4.48). Recalling the definition (4.38) of the $\mathbf{U}_s^{(i)}$'s, for any $t \in [0,1]$,

$$\mathbb{E}\big|D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big).\mathbf{U}_s^{(i)\,\otimes 3}\big| \\ \leq \sum_{k \in I_s}\sum_{\ell \in I_s}\sum_{j \in I_s} \mathbb{E}\Big(\Big|\frac{\partial^3 f}{\partial x_k^{(i)}\partial x_\ell^{(i)}\partial x_j^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big)\bar{X}_{k,m}^{(i)}\bar{X}_{\ell,m}^{(i)}\bar{X}_{j,m}^{(i)}\Big|\Big) \\ \leq \sum_{k \in I_s}\sum_{\ell \in I_s}\sum_{j \in I_s} \Big\|\frac{\partial^3 f}{\partial x_k^{(i)}\partial x_\ell^{(i)}\partial x_j^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big)\Big\|_2 \big\|\bar{X}_{k,m}^{(i)}\bar{X}_{\ell,m}^{(i)}\bar{X}_{j,m}^{(i)}\big\|_2 ,$$

where I_s is defined in (4.12). Therefore, using (4.11), stationarity and (4.23), it follows that, for any $t \in [0,1]$,

$$\mathbb{E}\big|D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big).\mathbf{U}_s^{(i)\,\otimes 3}\big| \\ \leq 8M^2 \sum_{k \in I_s}\sum_{\ell \in I_s}\sum_{j \in I_s} \Big\|\frac{\partial^3 f}{\partial x_k^{(i)}\partial x_\ell^{(i)}\partial x_j^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big)\Big\|_2 \big\|X_0\big\|_2 .$$

Notice that by (4.43) and (4.44),

$$\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)} = \big(\bar{\mathbf{X}}^{(1)}, \dots, \bar{\mathbf{X}}^{(i-1)}, w^{(i)}(t), \widetilde{\mathbf{Z}}^{(i+1)}, \dots, \widetilde{\mathbf{Z}}^{(n)}\big) , \quad (4.49)$$

where $w^{(i)}(t)$ is the row vector of dimension N defined by

$$w^{(i)}(t) = \mathbf{S}_{s-1}^{(i)} + t\mathbf{U}_s^{(i)} + \mathbf{T}_{s+1}^{(i)} = \big(\mathbf{u}_1^{(i)}, \dots, \mathbf{u}_{s-1}^{(i)}, t\mathbf{u}_s^{(i)}, \mathbf{v}_{s+1}^{(i)}, \dots, \mathbf{v}_{k_{N,m}}^{(i)}\big) , \quad (4.50)$$

where the $\mathbf{u}_\ell^{(i)}$'s are defined in (4.13) and (4.14) whereas the $\mathbf{v}_\ell^{(i)}$'s are defined in Section 4.2. Therefore, by Lemma 5.1 of the Appendix, (4.11), and since $(Z_k^{(i)})_{k\in\mathbb{Z}}$ is distributed as the stationary sequence $(Z_k)_{k\in\mathbb{Z}}$, we infer that there exists a positive constant C_1 not depending on (n, M, m) and such that, for any $t \in [0,1]$,

$$\Big\| \frac{\partial^3 f}{\partial x_k^{(i)} \partial x_\ell^{(i)} \partial x_j^{(i)}} \big(\widetilde{\mathbf{W}}_s^{(i)} + t \mathbf{U}_s^{(i)} \big) \Big\|_2 \leq C_1 \Big(\frac{M + \|Z_0\|_2}{v^3 N^{1/2} n^2} + \frac{N^{1/2}(M^3 + \|Z_0\|_6^3)}{v^4 n^3} \Big) .$$

Now, since Z_0 is a Gaussian random variable, $\|Z_0\|_6^6 = 15 \|Z_0\|_2^6$. Moreover, by (4.31), $\|Z_0\|_2 = \|X_0\|_2$. Therefore, there exists a positive constant C_2 not depending on (n, M, m) and such that, for any $t \in [0,1]$,

$$\mathbb{E}\big| D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{U}_s^{(i)}\big) . \mathbf{U}_s^{(i)\otimes 3} \big| \leq \frac{C_2 m^6 (1 + M^3)}{v^3 (1 \wedge v) N^{1/2} n^2} . \tag{4.51}$$

On another hand, since for any $i \in \{1, \dots, n\}$ and any $s \in \{1, \dots, k_{N,m}\}$, $\mathbf{U}_s^{(i)}$ is a centered random vector independent of $\widetilde{\mathbf{W}}_s^{(i)}$, it follows that

$$\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{U}_s^{(i)\otimes 1}\big) = 0 \ \text{ and } \ \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{U}_s^{(i)\otimes 2}\big) = \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) . \tag{4.52}$$

Hence starting from (4.48), using (4.51), (4.52) and the fact that $m^2 k_{N,m} \leq N$, we derive that there exists a positive constant C_3 not depending on on (n, M, m) and such that

$$\sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \Big| \mathbb{E}\big(\Delta_s^{(i)}(f)\big) - \frac{1}{2} \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) \Big| \leq C_3 \frac{(1 + M^5) N^{1/2} m^4}{v^3 (1 \wedge v) n} . \tag{4.53}$$

We analyze now the "Gaussian part" in (4.47), namely: $\mathbb{E}\big(\widetilde{\Delta}_s^{(i)}(f)\big)$. By Taylor's integral formula,

$$\Big| \mathbb{E}\big(\widetilde{\Delta}_s^{(i)}(f)\big) - \mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big) - \frac{1}{2} \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big) \Big| \\ \leq \Big| \mathbb{E} \int_0^1 \frac{(1-t)^2}{2} D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)} + t\mathbf{V}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 3} dt \Big| .$$

Proceeding as to get (4.53), we then infer that there exists a positive constant C_4 not depending on (n, M, m) and such that

$$\sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \Big| \mathbb{E}\big(\widetilde{\Delta}_s^{(i)}(f)\big) - \mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big) - \frac{1}{2} \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big) \Big| \\ \leq C_4 \frac{(1 + M^3) N^{1/2} m^4}{v^3 (1 \wedge v) n} . \tag{4.54}$$

We analyze now the terms $\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big)$ in (4.54). Recalling the definition (4.41) of the $\mathbf{V}_s^{(i)}$'s, we write

$$\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big) = \sum_{j \in I_s} \mathbb{E} \left(\frac{\partial f}{\partial x_j^{(i)}} \big(\widetilde{\mathbf{W}}_s^{(i)}\big) Z_j^{(i)} \right) ,$$

where I_s is defined in (4.12). To handle the terms in the right-hand side, we shall use the so-called Stein's identity for Gaussian vectors (see, for instance, Lemma 1 in Liu (1994)), as done by Neumann (2011) in the context of dependent real random variables: for $G = (G_1, \dots, G_d)$ a centered Gaussian vector of $\mathbb{R}^d$ and any function $h : \mathbb{R}^d \to \mathbb{R}$ such that its partial derivatives exist almost everywhere and $\mathbb{E}\big|\frac{\partial h}{\partial x_i}(G)\big| < \infty$ for any $i = 1, \dots, d$, the following identity holds true:

$$\mathbb{E}\big(G_i\, h(G)\big) = \sum_{\ell=1}^{d} \mathbb{E}(G_i G_\ell)\mathbb{E}\Big(\frac{\partial h}{\partial x_\ell}(G)\Big) \quad \text{for any } i \in \{1, \dots, d\}\,. \tag{4.55}$$

Using (4.55) with $G = \big(\mathbf{T}_{s+1}^{(i)}, Z_j^{(i)}\big) \in \mathbb{R}^{nN} \times \mathbb{R}$, $h : \mathbb{R}^{nN} \times \mathbb{R} \to \mathbb{R}$ satisfying $h(x, y) = \frac{\partial f}{\partial x_j^{(i)}}(x)$ for any $(x, y) \in \mathbb{R}^{nN} \times \mathbb{R}$, and noticing that G is independent of $\widetilde{\mathbf{W}}_s^{(i)} - \mathbf{T}_{s+1}^{(i)}$, we infer that, for any $j \in I_s$,

$$\mathbb{E}\left(\frac{\partial f}{\partial x_j^{(i)}}(\widetilde{\mathbf{W}}_s^{(i)}) Z_j^{(i)}\right) = \sum_{\ell=s+1}^{k_{N,m}} \sum_{k \in I_\ell} \mathbb{E}\left(\frac{\partial^2 f}{\partial x_k^{(i)} \partial x_j^{(i)}}(\widetilde{\mathbf{W}}_s^{(i)})\right) \mathrm{Cov}(Z_k^{(i)}, Z_j^{(i)})\,.$$

Therefore,

$$\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big) = \sum_{\ell=s+1}^{k_{N,m}} \sum_{k \in I_\ell} \sum_{j \in I_s} \mathbb{E}\left(\frac{\partial^2 f}{\partial x_k^{(i)} \partial x_j^{(i)}}(\widetilde{\mathbf{W}}_s^{(i)})\right) \mathrm{Cov}(Z_k^{(i)}, Z_j^{(i)})\,.$$

From (4.49) and (4.50) (with $t = 0$) and Lemma 5.1 of the Appendix, we infer that there exists a positive constant C_5 not depending on (n, M, m) and such that, for any $k \in I_\ell$ and any $j \in I_s$,

$$\mathbb{E}\left(\frac{\partial^2 f}{\partial x_k^{(i)} \partial x_j^{(i)}}(\widetilde{\mathbf{W}}_s^{(i)})\right) \leq C_5\Big(\frac{1}{Nnv^2} + \frac{1}{n^2 v^3}\big(\|X_0\|_2^2 + \|Z_0\|_2^2\big)\Big) \leq C_5 \frac{1 + 2\|X_0\|_2^2}{nv^2(1 \wedge v)(N \wedge n)}\,. \tag{4.56}$$

Hence, using the fact that $\mathrm{Cov}(Z_k^{(i)}, Z_j^{(i)}) = \mathrm{Cov}(Z_k, Z_j)$ together with (4.31), we then derive that

$$\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big) \leq C_5 \frac{1 + 2\|X_0\|_2^2}{nv^2(1 \wedge v)(N \wedge n)} \sum_{\ell=s+1}^{k_{N,m}} \sum_{k \in I_\ell} \sum_{j \in I_s} \big|\mathrm{Cov}(X_k, X_j)\big|\,. \tag{4.57}$$

By stationarity,

$$\sum_{k \in I_\ell} \sum_{j \in I_s} \big|\mathrm{Cov}(X_k, X_j)\big| = \sum_{j=1}^{m^2} \sum_{k=1}^{m^2} \big|\mathrm{Cov}(X_0, X_{k-j+(\ell-s)(m^2+m)})\big| \leq m^2 \sum_{k \in \mathcal{E}_{m,\ell}} \big|\mathrm{Cov}(X_0, X_k)\big|\,,$$

where $\mathcal{E}_{m,\ell} := \{1 - m^2 + (\ell - s)(m^2 + m), \dots, m^2 - 1 + (\ell - s)(m^2 + m)\}$. Notice that since $m \geq 1$, $\mathcal{E}_{m,\ell} \cap \mathcal{E}_{m,\ell+2} = \emptyset$. Then, summing on ℓ, and using the fact that $k_{N,m}(m^2 + m) \leq N$, we get that, for any $s \geq 1$,

$$\sum_{\ell=s+1}^{k_{N,m}} \sum_{k \in \mathcal{E}_{m,\ell}} \big|\mathrm{Cov}(X_0, X_k)\big| \leq 2 \sum_{k=m+1}^{m^2+N-1} \big|\mathrm{Cov}(X_0, X_k)\big|\,.$$

So, overall, for any positive integer s,

$$\sum_{\ell=s+1}^{k_{N,m}} \sum_{k \in I_\ell} \sum_{j \in I_s} \big|\mathrm{Cov}(X_k, X_j)\big| \leq 2m^2 \sum_{k=m+1}^{m^2+N-1} \big|\mathrm{Cov}(X_0, X_k)\big|\,. \tag{4.58}$$

Therefore, starting from (4.57) and using that $m^2 k_{N,m} \leq N$, it follows that

$$\sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \big|\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big)\big| \leq 2C_5 \frac{(1 + 2\|X_0\|_2^2)(1 + c(n))}{v^2(1 \wedge v)} \sum_{k \geq m+1} \big|\mathrm{Cov}(X_0, X_k)\big|\,. \tag{4.59}$$

Since $\mathcal{F}_{-\infty} = \bigcap_{k\in\mathbb{Z}} \sigma(\xi_k)$ is trivial, for any $k \in \mathbb{Z}$, $\mathbb{E}(X_k|\mathcal{F}_{-\infty}) = \mathbb{E}(X_k) = 0$ a.s. Therefore, the following decomposition is valid: $X_k = \sum_{r=-\infty}^{k} P_r(X_k)$. Next, since $\mathbb{E}\big(P_i(X_0)P_j(X_k)\big) = 0$ if $i \neq j$, we get, by stationarity, that for any integer $k \geq 0$,

$$\big|\mathrm{Cov}(X_0, X_k)\big| = \Big| \sum_{r=-\infty}^{0} \mathbb{E}\Big(P_r(X_0)P_r(X_k)\Big)\Big| \leq \sum_{r=0}^{\infty} \|P_0(X_r)\|_2 \|P_0(X_{k+r})\|_2 \,, \tag{4.60}$$

implying that for any non-negative integer u,

$$\sum_{k\geq u} \big|\mathrm{Cov}(X_0, X_k)\big| \leq \sum_{r\geq 0} \|P_0(X_r)\|_2 \sum_{k\geq u} \|P_0(X_k)\|_2 \,. \tag{4.61}$$

Hence, starting from (4.59) and considering (4.61) together with the condition (2.3), we derive that there exists a positive constant C_6 not depending on (n, M, m) such that

$$\sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \big|\mathbb{E}\big(Df\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 1}\big)\big| \leq \frac{C_6(1+c(n))}{v^2(1\wedge v)} \sum_{k\geq m+1} \|P_0(X_k)\|_2 \,. \tag{4.62}$$

We analyze now the terms of second order in (4.54), namely: $\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big)$. Recalling the definition (4.41) of the $\mathbf{V}_s^{(i)}$'s, we first write that

$$\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big) = \sum_{j_1\in I_s} \sum_{j_2\in I_s} \mathbb{E}\left(\frac{\partial^2 f}{\partial x_{j_1}^{(i)} \partial x_{j_2}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big) Z_{j_1}^{(i)} Z_{j_2}^{(i)}\right) \,, \tag{4.63}$$

where I_s is defined in (4.12). Using now (4.55) with $G = \big(\mathbf{T}_{s+1}^{(i)}, Z_{j_1}^{(i)}, Z_{j_2}^{(i)}\big) \in \mathbb{R}^{nN} \times \mathbb{R} \times \mathbb{R}$, $h : \mathbb{R}^{nN} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying $h(x, y, z) = y \frac{\partial^2 f}{\partial x_{j_1}^{(i)} \partial x_{j_2}^{(i)}}(x)$ for any $(x, y, z) \in \mathbb{R}^{nN} \times \mathbb{R} \times \mathbb{R}$, and noticing that G is independent of $\widetilde{\mathbf{W}}_s^{(i)} - \mathbf{T}_{s+1}^{(i)}$, we infer that, for any j_1, j_2 belonging to I_s,

$$\begin{aligned} \mathbb{E}\left(\frac{\partial^2 f}{\partial x_{j_1}^{(i)} \partial x_{j_2}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big) Z_{j_1}^{(i)} Z_{j_2}^{(i)}\right) &= \mathbb{E}\left(\frac{\partial^2 f}{\partial x_{j_1}^{(i)} \partial x_{j_2}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\right) \mathbb{E}\big(Z_{j_1}^{(i)} Z_{j_2}^{(i)}\big) \\ &+ \sum_{k=s+1}^{k_{N,m}} \sum_{j_3\in I_k} \mathbb{E}\left(\frac{\partial^3 f}{\partial x_{j_3}^{(i)} \partial x_{j_1}^{(i)} \partial x_{j_2}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big) Z_{j_1}^{(i)}\right) \mathbb{E}\big(Z_{j_3}^{(i)} Z_{j_2}^{(i)}\big) \,. \end{aligned} \tag{4.64}$$

Therefore, starting from (4.63) and using (4.64) combined with the definitions 4.2 and 4.3, it follows that

$$\begin{aligned} &\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big) \\ &= \mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big) + \sum_{k=s+1}^{k_{N,m}} \mathbb{E}\Big(D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)} \otimes \mathbb{E}\big(\mathbf{V}_k^{(i)} \otimes \mathbf{V}_s^{(i)}\big)\Big) \,. \end{aligned} \tag{4.65}$$

Next, with similar arguments, we infer that

$$\begin{aligned} &\sum_{k=s+1}^{k_{N,m}} \mathbb{E}\Big(D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)} \otimes \mathbb{E}\big(\mathbf{V}_k^{(i)} \otimes \mathbf{V}_s^{(i)}\big)\Big) = \\ &\sum_{k=s+1}^{k_{N,m}} \sum_{\ell=s+1}^{k_{N,m}} \mathbb{E}\big(D^4 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{V}_\ell^{(i)} \otimes \mathbf{V}_s^{(i)}\big) \otimes \mathbb{E}\big(\mathbf{V}_k^{(i)} \otimes \mathbf{V}_s^{(i)}\big) \,. \end{aligned} \tag{4.66}$$

By the definition (4.41) of the $\mathbf{V}_\ell^{(i)}$'s, we first write that

$$
\begin{aligned}
&\mathbb{E}\big(D^4 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{V}_\ell^{(i)}\otimes\mathbf{V}_s^{(i)}\big)\otimes\mathbb{E}\big(\mathbf{V}_k^{(i)}\otimes\mathbf{V}_s^{(i)}\big)\\
&\quad=\sum_{j_1\in I_\ell}\sum_{j_2\in I_s}\sum_{j_3\in I_k}\sum_{j_4\in I_s}\mathbb{E}\left(\frac{\partial^4 f}{\partial x_{j_1}^{(i)}\partial x_{j_2}^{(i)}\partial x_{j_3}^{(i)}\partial x_{j_4}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\right)\mathrm{Cov}\big(Z_{j_1}^{(i)},Z_{j_2}^{(i)}\big)\mathrm{Cov}\big(Z_{j_3}^{(i)},Z_{j_4}^{(i)}\big)\\
&\quad=\sum_{j_1\in I_\ell}\sum_{j_2\in I_s}\sum_{j_3\in I_k}\sum_{j_4\in I_s}\mathbb{E}\left(\frac{\partial^4 f}{\partial x_{j_1}^{(i)}\partial x_{j_2}^{(i)}\partial x_{j_3}^{(i)}\partial x_{j_4}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\right)\mathrm{Cov}\big(X_{j_1},X_{j_2}\big)\mathrm{Cov}\big(X_{j_3},X_{j_4}\big)\,, \quad (4.67)
\end{aligned}
$$

where for the last line, we have used that $(Z_k^{(i)})_{k\in\mathbb{Z}}$ is distributed as $(Z_k)_{k\in\mathbb{Z}}$ together with (4.31). From (4.49) and (4.50) (with $t=0$), Lemma 5.1 of the Appendix, and the stationarity of the sequences $(\bar{X}_{k,m}^{(i)})_{k\in\mathbb{Z}}$ and $(Z_k^{(i)})_{k\in\mathbb{Z}}$, we infer that there exists a positive constant C_7 not depending on (n,M,m) such that

$$
\begin{aligned}
\mathbb{E}\left(\frac{\partial^4 f}{\partial x_{j_1}^{(i)}\partial x_{j_2}^{(i)}\partial x_{j_3}^{(i)}\partial x_{j_4}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\right)&\le C_7\Big(\frac{1}{Nn^2v^3}+\frac{1}{Nn^3v^4}\Big(\sum_{k=1}^N\|\bar{X}_{k,m}^{(i)}\|_2^2+\sum_{k=1}^N\|Z_k^{(i)}\|_2^2\Big)\\
&\qquad+\frac{1}{Nn^4v^5}\Big(\Big\|\sum_{k=1}^N\big(\bar{X}_{k,m}^{(i)}\big)^2\Big\|_2^2+\Big\|\sum_{k=1}^N\big(Z_k^{(i)}\big)^2\Big\|_2^2\Big)\Big)\\
&\le\frac{C_7}{n^2Nv^3(1\wedge v^2)}\left(1+\frac{N\big(\|\bar{X}_{0,m}\|_2^2+\|Z_0\|_2^2\big)}{n}+\frac{N^2\big(\|\bar{X}_{0,m}\|_4^4+\|Z_0\|_4^4\big)}{n^2}\right).
\end{aligned}
$$

By (4.11) and (4.23), $\|\bar{X}_{0,m}\|_4^4\le(2M)^2\|\bar{X}_{0,m}\|_2^2\le 16M^2\|X_0\|_2^2$. Moreover, Z_0 being a Gaussian random variable, $\|Z_0\|_4^4=3\|Z_0\|_2^4$. Hence, by (4.31), $\|Z_0\|_4^4=3\|X_0\|_2^4$ and $\|Z_0\|_2^2=\|X_0\|_2^2$. Therefore, there exists a positive constant C_8 not depending on (n,M,m) and such that

$$
\mathbb{E}\left(\frac{\partial^4 f}{\partial x_{j_1}^{(i)}\partial x_{j_2}^{(i)}\partial x_{j_3}^{(i)}\partial x_{j_4}^{(i)}}\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\right)\le\frac{C_8(1+M^2)(1+c^2(n))}{n^2Nv^3(1\wedge v^2)}\,. \quad (4.68)
$$

On the other hand, by using (4.58) and (4.61), we get that, for any positive integer s,

$$
\begin{aligned}
&\sum_{k=s+1}^{k_{N,m}}\sum_{\ell=s+1}^{k_{N,m}}\sum_{j_1\in I_\ell}\sum_{j_2\in I_s}\sum_{j_3\in I_k}\sum_{j_4\in I_s}\big|\mathrm{Cov}\big(X_{j_1},X_{j_2}\big)\mathrm{Cov}\big(X_{j_3},X_{j_4}\big)\big|\\
&\qquad\le 4m^4\Big(\sum_{r\ge0}\|P_0(X_r)\|_2\Big)^2\Big(\sum_{k\ge m+1}\|P_0(X_k)\|_2\Big)^2. \quad (4.69)
\end{aligned}
$$

Whence, starting from (4.66), using (4.67), and considering the upper bounds (4.68) and (4.69) together with the condition (2.3), we derive that there exists a positive constant C_9 not depending on (n,M,m) such that

$$
\sum_{k=s+1}^{k_{N,m}}\mathbb{E}\Big(D^3 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)}\otimes\mathbb{E}\big(\mathbf{V}_k^{(i)}\otimes\mathbf{V}_s^{(i)}\big)\Big)\le\frac{C_9(1+M^2)(1+c^2(n))m^4}{n^2Nv^3(1\wedge v^2)}\,. \quad (4.70)
$$

So, overall, starting from (4.65), considering (4.70) and using the fact that $m^2k_{N,m}\le N$, we derive that

$$
\begin{aligned}
&\Big|\sum_{i=1}^n\sum_{s=1}^{k_{N,m}}\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big).\mathbf{V}_s^{(i)\otimes 2}\big)-\sum_{i=1}^n\sum_{s=1}^{k_{N,m}}\mathbb{E}\big(D^2 f\big(\widetilde{\mathbf{W}}_s^{(i)}\big)\big).\mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big)\Big|\\
&\qquad\le\frac{C_9(1+M^2)(1+c^2(n))m^2}{nv^3(1\wedge v^2)}\,. \quad (4.71)
\end{aligned}
$$

Then starting from (4.47), and considering the upper bounds (4.53), (4.54), (4.62) and (4.71), we get that

$$\Big|\mathbb{E}\big(S_{F^{\bar{\mathbf{B}}_n}}(z)\big) - \mathbb{E}\big(S_{F^{\tilde{\mathbf{G}}_n}}(z)\big)\Big| \leq \frac{1}{2}\sum_{i=1}^{n}\sum_{s=1}^{k_{N,m}} \Big|\mathbb{E}\big(D^2 f(\widetilde{\mathbf{W}}_s^{(i)})\big).\Big(\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) - \mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big)\Big)\Big|$$
$$+ \frac{4C_{10}(1+M^5)N^{1/2}m^4}{v^3(1\wedge v)n} + \frac{C_{10}(1+M^2)(1+c^2(n))m^2}{nv^3(1\wedge v^2)} + \frac{C_{10}(1+c^2(n))}{v^2(1\wedge v)}\sum_{k\geq m+1}\|P_0(X_k)\|_2\,,$$

where $C_{10} = \max(C_3, C_4, C_6, C_7)$. Since $c(n) \to c \in (0,\infty)$, it follows that the second and third terms in the right-hand side of the above inequality tend to zero as n tends to infinity. On another hand, by the condition (2.3), $\lim_{m\to\infty}\sum_{k\geq m+1}\|P_0(X_k)\|_2 = 0$. Therefore, Proposition 4.3 will follow if we can prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{M\to\infty}\limsup_{n\to\infty}\sum_{i=1}^{n}\sum_{s=1}^{k_{N,m}} \Big|\mathbb{E}\big(D^2 f(\widetilde{\mathbf{W}}_s^{(i)})\big).\Big(\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) - \mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big)\Big)\Big| = 0\,. \tag{4.72}$$

Using the fact that $(Z_k^{(i)})_{k\in\mathbb{Z}}$ is distributed as $(Z_k)_{k\in\mathbb{Z}}$ together with (4.31) and that $(\bar{X}_{k,m}^{(i)})_{k\in\mathbb{Z}}$ is distributed as $(\bar{X}_{k,m})_{k\in\mathbb{Z}}$, we first write that

$$\mathbb{E}\big(D^2 f(\widetilde{\mathbf{W}}_s^{(i)})\big).\Big(\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) - \mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big)\Big)$$
$$= \sum_{k\in I_s}\sum_{\ell\in I_s}\mathbb{E}\left(\frac{\partial^2 f}{\partial x_k^{(i)}\partial x_\ell^{(i)}}(\widetilde{\mathbf{W}}_s^{(i)})\right)\Big(\mathrm{Cov}\big(\bar{X}_{k,m},\bar{X}_{\ell,m}\big) - \mathrm{Cov}\big(X_k, X_\ell\big)\Big)\,.$$

Hence, by using (4.56) and stationarity, we get that there exists a positive constant C_{11} not depending on (n, M, m) such that

$$\Big|\mathbb{E}\big(D^2 f(\widetilde{\mathbf{W}}_s^{(i)})\big).\Big(\mathbb{E}\big(\mathbf{U}_s^{(i)\otimes 2}\big) - \mathbb{E}\big(\mathbf{V}_s^{(i)\otimes 2}\big)\Big)\Big|$$
$$\leq \frac{C_{11}}{nv^2(1\wedge v)(N\wedge n)}\sum_{\ell=1}^{m^2}\sum_{k=0}^{m^2-\ell}\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_0,X_k\big)\big|\,. \tag{4.73}$$

To handle the right-hand side term, we first write that

$$\sum_{\ell=1}^{m^2}\sum_{k=0}^{m^2-\ell}\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_0,X_k\big)\big| \leq m^2\sum_{k=0}^{m^2}\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_{0,m},X_{k,m}\big)\big|$$
$$+ m^2\sum_{k=0}^{m^2}\big|\mathrm{Cov}\big(X_{0,m},X_{k,m}\big) - \mathrm{Cov}\big(X_0,X_k\big)\big|\,, \tag{4.74}$$

where $X_{0,m}$ and $X_{k,m}$ are defined in (4.28). Notice now that $\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) = \mathrm{Cov}\big(X_{0,m},X_{k,m}\big) = 0$ if $k > m$. Therefore,

$$\sum_{k=0}^{m^2}\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_{0,m},X_{k,m}\big)\big| = \sum_{k=0}^{m}\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_{0,m},X_{k,m}\big)\big|\,.$$

Next, using stationarity, the fact that the random variables are centered, (4.11) and (4.29), we get that

$$\big|\mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_{0,m},X_{k,m}\big)\big|$$
$$= \big|\mathrm{Cov}\big(\bar{X}_{0,m} - X_{0,m},\bar{X}_{k,m}\big) + \mathrm{Cov}\big(X_{0,m} - \bar{X}_{0,m},\bar{X}_{k,m} - X_{k,m}\big) + \mathrm{Cov}\big(\bar{X}_{0,m},\bar{X}_{k,m} - X_{k,m}\big)\big|$$
$$\leq 4M\|X_{0,m} - \bar{X}_{0,m}\|_1 + 4\|\big(|X_0| - M\big)_+\|_2^2\,.$$

As to get (4.29), notice that $\|X_{0,m} - \bar{X}_{0,m}\|_1 \leq 2\|\big(|X_0| - M\big)_+\|_1$. Moreover, $\big(|x| - M\big)_+ \leq 2|x|\mathbf{1}_{|x|\geq M}$ which in turn implies that $M\big(|x| - M\big)_+ \leq 2|x|^2\mathbf{1}_{|x|\geq M}$. So, overall,

$$\sum_{k=0}^{m^2} \big|\mathrm{Cov}\big(\bar{X}_{0,m}, \bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_{0,m}, X_{k,m}\big)\big| \leq 32\, m\mathbb{E}\big(X_0^2\mathbf{1}_{|X_0|\geq M}\big)\,. \tag{4.75}$$

We handle now the second term in the right-hand side of (4.74). Let $b(m)$ be an increasing sequence of positive integers such that $b(m) \to \infty$, $b(m) \leq [m/2]$, and

$$\lim_{m\to\infty} b(m)\big\|X_0 - X_{0,[m/2]}\big\|_2^2 = 0\,. \tag{4.76}$$

Notice that since (4.30) holds true, it is always possible to find such a sequence. Now, using (4.60),

$$\sum_{k=b(m)}^{m^2} \big|\mathrm{Cov}\big(X_{0,m}, X_{k,m}\big) - \mathrm{Cov}\big(X_0, X_k\big)\big|$$
$$\leq \sum_{k=b(m)}^{m^2}\sum_{r=0}^{\infty} \|P_0(X_{r,m})\|_2\|P_0(X_{k+r,m})\|_2 + \sum_{k=b(m)}^{m^2}\sum_{r=0}^{\infty} \|P_0(X_r)\|_2\|P_0(X_{k+r})\|_2\,. \tag{4.77}$$

Recalling the definition (4.28) of the $X_{j,m}$'s, we notice that $P_0(X_{j,m}) = 0$ if $j \geq m+1$. Now, for any $j \in \{0, \ldots, m\}$,

$$\begin{aligned}\mathbb{E}(X_{j,m}|\xi_0) &= \mathbb{E}(\mathbb{E}(X_j|\varepsilon_j, \ldots, \varepsilon_{j-m})|\xi_0) = \mathbb{E}(\mathbb{E}(X_j|\varepsilon_j, \ldots, \varepsilon_{j-m})|\varepsilon_0, \ldots, \varepsilon_{j-m})\\ &= \mathbb{E}(X_j|\varepsilon_0, \ldots, \varepsilon_{j-m}) = \mathbb{E}(\mathbb{E}(X_j|\xi_0)|\varepsilon_0, \ldots, \varepsilon_{j-m}) \quad \text{a.s.}\end{aligned}$$

Actually, the two last equalities follow from the tower lemma, whereas, for the second one, we have used the following well known fact with $\mathcal{G}_1 = \sigma(\varepsilon_0, \ldots, \varepsilon_{j-m})$, $\mathcal{G}_2 = \sigma(\varepsilon_k, k \leq j-m-1)$ and $Y = X_{j,m}$: if Y is an integrable random variable, and $\mathcal{G}_1$ and $\mathcal{G}_2$ are two σ-algebras such that $\sigma(Y) \vee \mathcal{G}_1$ is independent of $\mathcal{G}_2$, then

$$\mathbb{E}(Y|\mathcal{G}_1 \vee \mathcal{G}_2) = \mathbb{E}(Y|\mathcal{G}_1) \quad \text{a.s.} \tag{4.78}$$

Similarly, for any $j \in \{0, \ldots, m-1\}$,

$$\mathbb{E}(X_{j,m}|\xi_{-1}) = \mathbb{E}(X_j|\varepsilon_{-1}, \ldots, \varepsilon_{j-m}) = \mathbb{E}(\mathbb{E}(X_j|\xi_{-1})|\varepsilon_{-1}, \ldots, \varepsilon_{j-m}) \quad \text{a.s.}$$

Then using the equality (4.78) with $\mathcal{G}_1 = \sigma(\varepsilon_{-1}, \ldots, \varepsilon_{j-m})$ and $\mathcal{G}_2 = \sigma(\varepsilon_0)$, we get that, for any $j \in \{1, \ldots, m-1\}$,

$$\mathbb{E}(X_{j,m}|\xi_{-1}) = \mathbb{E}(\mathbb{E}(X_j|\xi_{-1})|\varepsilon_0, \ldots, \varepsilon_{j-m}) \quad \text{a.s.}$$

whereas $\mathbb{E}(X_{m,m}|\xi_{-1}) = 0$ a.s. So, finally, $\|P_0(X_{m,m})\|_2 = \|\mathbb{E}(X_m|\varepsilon_0)\|_2$, $\|P_0(X_{j,m})\|_2 = 0$ if $j \geq m+1$, and, for any $j \in \{1, \ldots, m-1\}$,

$$\begin{aligned}\|P_0(X_{j,m})\|_2 &= \|\mathbb{E}(X_{j,m}|\xi_0) - \mathbb{E}(X_{j,m}|\xi_{-1})\|_2\\ &= \|\mathbb{E}\big(\mathbb{E}(X_j|\xi_0) - \mathbb{E}(X_j|\xi_{-1})|\varepsilon_0, \ldots, \varepsilon_{j-m}\big)\|_2 \leq \|P_0(X_j)\|_2\,.\end{aligned}$$

Therefore, starting from (4.77), we infer that

$$\sum_{k=b(m)}^{m^2} \big|\mathrm{Cov}\big(X_{0,m}, X_{k,m}\big) - \mathrm{Cov}\big(X_0, X_k\big)\big|$$
$$\leq 2\|X_0\|_2\|\mathbb{E}(X_m|\varepsilon_0)\|_2 + 2\sum_{r=0}^{\infty}\|P_0(X_r)\|_2 \sum_{k\geq b(m)} \|P_0(X_k)\|_2\,. \tag{4.79}$$

On the other hand,

$$\sum_{k=0}^{b(m)} \big|\mathrm{Cov}\big(X_{0,m}, X_{k,m}\big) - \mathrm{Cov}\big(X_0, X_k\big)\big| \\ \le \sum_{k=0}^{b(m)} \big|\mathrm{Cov}\big(X_0 - X_{0,m}, X_{k,m}\big)\big| + \sum_{k=0}^{b(m)} \big|\mathrm{Cov}\big(X_0, X_k - X_{k,m}\big)\big| . \quad (4.80)$$

Since the random variables are centered, $\mathrm{Cov}\big(X_0 - X_{0,m}, X_{k,m}\big) = \mathbb{E}\big(X_{k,m}(X_0 - X_{0,m})\big)$. Since $X_{k,m}$ is $\sigma(\varepsilon_{k-m}, \dots, \varepsilon_k)$-measurable,

$$\mathbb{E}\big(X_{k,m}(X_0 - X_{0,m})\big) = \mathbb{E}\big(X_{k,m}\big(\mathbb{E}(X_0|\varepsilon_k, \dots, \varepsilon_{k-m}) - \mathbb{E}(X_{0,m}|\varepsilon_k, \dots, \varepsilon_{k-m}\big)\big) .$$

But, for any $k \in \{0, \dots, m\}$, by using the equality (4.78) with $\mathcal{G}_1 = \sigma(\varepsilon_0, \dots, \varepsilon_{k-m})$ and $\mathcal{G}_2 = \sigma(\varepsilon_k, \dots, \varepsilon_1)$, it follows that

$$\mathbb{E}(X_{0,m}|\varepsilon_k, \dots, \varepsilon_{k-m}\big) = \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m}) \ \text{ a.s.} \quad (4.81)$$

and

$$\mathbb{E}(X_0|\varepsilon_k, \dots, \varepsilon_{k-m}\big) = \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m}) \ \text{ a.s.}$$

Whence,

$$\sum_{k=0}^{b(m)} \big|\mathrm{Cov}\big(X_0 - X_{0,m}, X_{k,m}\big)\big| = 0 . \quad (4.82)$$

To handle the second term in the right-hand side of (4.80), we start by writing that

$$\mathrm{Cov}\big(X_0, X_k - X_{k,m}\big) = \mathrm{Cov}\big(X_0 - X_{0,m}, X_k - X_{k,m}\big) + \mathrm{Cov}\big(X_{0,m}, X_k - X_{k,m}\big) . \quad (4.83)$$

Using the fact that the random variables are centered together with stationarity, we get that

$$\big|\mathrm{Cov}\big(X_0 - X_{0,m}, X_k - X_{k,m}\big)\big| \le \|X_0 - X_{0,m}\|_2^2 . \quad (4.84)$$

On the other hand, noticing that $\mathbb{E}(X_k - X_{k,m}|\varepsilon_k, \dots, \varepsilon_{k-m}) = 0$, and using the fact that the random variables are centered, and stationarity, it follows that

$$\big|\mathrm{Cov}\big(X_{0,m}, X_k - X_{k,m}\big)\big| = \big|\mathbb{E}\big(\big(X_{0,m} - \mathbb{E}(X_{0,m}|\varepsilon_k, \dots, \varepsilon_{k-m})\big)\big(X_k - X_{k,m}\big)\big)\big| \\ \le \|X_{0,m} - \mathbb{E}(X_{0,m}|\varepsilon_k, \dots, \varepsilon_{k-m})\|_2 \|X_0 - X_{0,m}\|_2 . \quad (4.85)$$

Next, using (4.81), we get that, for any $k \in \{0, \dots, m\}$,

$$\|X_{0,m} - \mathbb{E}(X_{0,m}|\varepsilon_k, \dots, \varepsilon_{k-m})\|_2 = \|X_{0,m} - \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m})\|_2 \\ = \|\mathbb{E}\big(X_0 - \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m})|\varepsilon_0, \dots, \varepsilon_{-m}\big)\|_2 \le \|X_0 - \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m})\|_2 . \quad (4.86)$$

Therefore, starting from (4.85), taking into account (4.86) and the fact that

$$\max_{0 \le k \le [m/2]} \|X_0 - \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{k-m})\|_2 \le \|X_0 - \mathbb{E}(X_0|\varepsilon_0, \dots, \varepsilon_{-[m/2]})\|_2 ,$$

we get that

$$\max_{0 \le k \le [m/2]} \big|\mathrm{Cov}\big(X_{0,m}, X_k - X_{k,m}\big)\big| \le \|X_0 - X_{0,[m/2]}\|_2^2 . \quad (4.87)$$

Starting from (4.83), gathering (4.84) and (4.87), and using the fact that $b(m) \le [m/2]$, we then derive that

$$\sum_{k=0}^{b(m)} \big|\mathrm{Cov}\big(X_0, X_k - X_{k,m}\big)\big)\big| \le 2\, b(m) \|X_0 - X_{0,[m/2]}\|_2^2 ,$$

which combined with (4.80) and (4.82) implies that

$$\sum_{k=0}^{b(m)} \left|\mathrm{Cov}\big(X_{0,m}, X_{k,m}\big) - \mathrm{Cov}\big(X_0, X_k\big)\right| \le 2\, b(m) \|X_0 - X_{0,[m/2]}\|_2^2 \,. \tag{4.88}$$

So, overall, starting from (4.74), gathering the upper bounds (4.75), (4.79) and (4.88), and taking into account the condition (2.3), we get that that there exists a positive constant C_{12} not depending on (n, M, m) and such that

$$\sum_{\ell=1}^{m^2} \sum_{k=0}^{m^2-\ell} \left|\mathrm{Cov}\big(\bar{X}_{0,m}, \bar{X}_{k,m}\big) - \mathrm{Cov}\big(X_0, X_k\big)\right|$$
$$\le C_{12}\Big(m^3 \mathbb{E}\big(X_0^2 \mathbf{1}_{|X_0|\ge M}\big) + m^2 \|\mathbb{E}(X_m|\varepsilon_0)\|_2 + m^2 \sum_{k\ge b(m)} \|P_0(X_k)\|_2 + m^2\, b(m)\|X_0 - X_{0,[m/2]}\|_2^2\Big)\,. \tag{4.89}$$

Therefore, starting from (4.73), considering the upper bound (4.89), using the fact that $m^2 k_{N,m} \le N$ and that $\lim_{n\to\infty} c(n) = c$, it follows that there exists a positive constant C_{13} not depending on (M, m) and such that

$$\limsup_{n\to\infty} \sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \left| E\big(D^2 f(\widetilde{\mathbf{W}}_s^{(i)})\big) . \Big(\mathbb{E}\big(\mathbf{U}_s^{(i)\,\otimes 2}\big) - \mathbb{E}\big(\mathbf{V}_s^{(i)\,\otimes 2}\big)\Big)\right|$$
$$\le \frac{C_{13}}{v^2(1\wedge v)} \Big(m \mathbb{E}\big(X_0^2 \mathbf{1}_{|X_0|\ge M}\big) + \|\mathbb{E}(X_m|\varepsilon_0)\|_2 + \sum_{k\ge b(m)} \|P_0(X_k)\|_2 + b(m)\|X_0 - X_{0,[m/2]}\|_2^2\Big)\,. \tag{4.90}$$

Letting first M tend to infinity and using the fact that X_0 belongs to $\mathbb{L}^2$, the first term in the right-hand side is going to zero. Letting now m tend to infinity the third term vanishes by the condition (2.3), whereas the last one goes to zero by taking into account (4.76). To show that the second term goes to zero as m tends to infinity, we notice that, by stationarity, $\|\mathbb{E}(X_m|\varepsilon_0)\|_2 \le \|\mathbb{E}(X_m|\xi_0)\|_2 = \|\mathbb{E}(X_0|\xi_{-m})\|_2$. By the reverse martingale convergence theorem, setting $\mathcal{F}_{-\infty} = \bigcap_{k\in\mathbb{Z}} \sigma(\xi_k)$, $\lim_{m\to\infty} \mathbb{E}(X_0|\xi_{-m}) = \mathbb{E}(X_0|\mathcal{F}_{-\infty}) = 0$ a.s. (since $\mathcal{F}_{-\infty}$ is trivial and $\mathbb{E}(X_0) = 0$). So, since X_0 belongs to $\mathbb{L}^2$, $\lim_{m\to\infty} \|\mathbb{E}(X_m|\varepsilon_0)\|_2 = 0$. This ends the proof of (4.72) and then of Proposition 4.3. □

4.4 End of the proof of Theorem 2.1

According to Propositions 4.1, 4.2 and 4.3, the convergence (4.3) follows. Therefore, to end the proof of Theorem 2.1, it remains to show that (4.4) holds true with $\mathbf{G}_n$ defined in Section 4.2. This can be achieved by using Theorem 1.1 in Silverstein (1995) combined with arguments developed in the proof of Theorem 1 in Yao (2012) (see also Wang *et al.* (2011)). With this aim, we consider $(y_k)_{k\in\mathbb{Z}}$ a sequence of i.i.d. real valued random variables with law $\mathcal{N}(0,1)$, and n independent copies of $(y_k)_{k\in\mathbb{Z}}$ that we denote by $(y_k^{(1)})_{k\in\mathbb{Z}}, \dots, (y_k^{(n)})_{k\in\mathbb{Z}}$. For any $i \in \{1, \dots, n\}$, define $\mathbf{y}_i = \big(y_1^{(i)}, \dots, y_N^{(i)}\big)$. Let $\mathcal{Y}_n = (\mathbf{y}_1^T | \dots | \mathbf{y}_n^T)$ be the matrix whose columns are the $\mathbf{y}_i^T$'s and consider its associated sample covariance matrix $\mathbf{Y}_n = \frac{1}{n}\mathcal{Y}_n\mathcal{Y}_n^T$. Let $\gamma(k) = \mathrm{Cov}(X_0, X_k)$ and note that, by (4.31), $\gamma(k)$ is also equal to $\mathrm{Cov}(Z_0, Z_k) = \mathrm{Cov}(Z_0^{(i)}, Z_k^{(i)})$ for any $i \in \{1, \dots, n\}$. Set

$$\Gamma_N := \big(\gamma_{j,k}\big) = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(N-1) \\ \gamma(1) & \gamma(0) & & \gamma(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \cdots & \gamma(0) \end{pmatrix}.$$

Note that (Γ_N) is bounded in spectral norm. Indeed, by the Gerschgorin theorem, the largest eigenvalue of Γ_N is not larger than $\gamma(0)+2\sum_{k\geq 1}|\gamma(k)|$ which, according to Remark 2.2, is finite. Note also that the vector $(\mathbf{Z}_1,\dots,\mathbf{Z}_n)$ has the same distribution as $\big(\mathbf{y}_1\Gamma_N^{1/2},\dots,\mathbf{y}_n\Gamma_N^{1/2}\big)$ where $\Gamma_N^{1/2}$ is the symmetric non-negative square root of Γ_N and the $\mathbf{Z}_i$'s are defined in Section 4.2. Therefore, for any $z\in\mathbb{C}^+$, $\mathbb{E}\big(S_{F^{\mathbf{G}_n}}(z)\big)=\mathbb{E}\big(S_{F^{\mathbf{A}_n}}(z)\big)$ where $\mathbf{A}_n=\Gamma_N^{1/2}\mathbf{Y}_n\Gamma_N^{1/2}$. The proof of (4.4) is then reduced to prove that, for any $z\in\mathbb{C}^+$,

$$\lim_{n\to\infty}\mathbb{E}\big(S_{F^{\mathbf{A}_n}}(z)\big)=S(z)\,, \tag{4.91}$$

where S is defined in (2.4). According to Theorem 1.1 in Silverstein (1995), if one can show that

$$F^{\Gamma_N} \text{ converges to a probability distribution } H, \tag{4.92}$$

then (4.91) holds with S satisfying the equation (1.4) in Silverstein (1995). Due to the Toeplitz form of Γ_N and to the fact that $\sum_{k\geq 0}|\gamma(k)|<\infty$ (see Remark 2.2), the convergence (4.92) can be proved by taking into account the arguments developed in the proof of Theorem 1 of Yao (2012). Indeed, the fundamental eigenvalue distribution theorem of Szegö for Toeplitz forms allows to assert that the empirical spectral distribution of Γ_N converges weakly to a non random distribution H that is defined via the spectral density of $(X_k)_{k\in\mathbb{Z}}$ (see Relations (12) and (13) in Yao (2012)). To end the proof, it suffices to notice that the relation (1.4) in Silverstein (1995) combined with the relation (13) in Yao (2012) leads to (2.4). $\square$

5 Appendix

In this section, we give some upper bounds for the partial derivatives of f defined in (4.35).

Lemma 5.1 *Let x be a vector of $\mathbb{R}^{nN}$ with coordinates*

$$x=\big(x^{(1)},\dots,x^{(n)}\big) \text{ where for any } i\in\{1,\dots,n\},\ x^{(i)}=\big(x_k^{(i)}\,,\,k\in\{1,\dots,N\}\big)\,.$$

Let $z=u+\sqrt{-1}v\in\mathbb{C}^+$ and $f:=f_z$ be the function defined in (4.35)*. Then, for any $i\in\{1,\dots,n\}$ and any $j,k,\ell,m\in\{1,\dots,N\}$, the following inequalities hold true:*

$$\left|\frac{\partial^2 f}{\partial x_m^{(i)}\partial x_j^{(i)}}(x)\right|\leq\frac{8}{v^3n^2N}\sum_{r=1}^N|x_r^{(i)}|^2+\frac{2}{v^2nN}\,,$$

$$\left|\frac{\partial^3 f}{\partial x_\ell^{(i)}\partial x_m^{(i)}\partial x_j^{(i)}}(x)\right|\leq\frac{48}{v^4n^3N}\left(\sum_{r=1}^N|x_r^{(i)}|^2\right)^{3/2}+\frac{24}{v^3n^2N}\left(\sum_{r=1}^N|x_r^{(i)}|^2\right)^{1/2}\,,$$

and

$$\left|\frac{\partial^4 f}{\partial x_k^{(i)}\partial x_\ell^{(i)}\partial x_m^{(i)}\partial x_j^{(i)}}(x)\right|\leq\frac{24\times 16}{v^5n^4N}\left(\sum_{r=1}^N|x_r^{(i)}|^2\right)^{2}+\frac{36\times 8}{v^4n^3N}\sum_{r=1}^N|x_r^{(i)}|^2+\frac{24}{v^3n^2N}\,.$$

Proof. Recall that $f(x)=\frac{1}{N}\mathrm{Tr}\big(A(x)-z\mathbf{I}\big)^{-1}$ where $A(x)=\frac{1}{n}\sum_{k=1}^n(x^{(k)})^Tx^{(k)}$. To prove the lemma, we shall proceed as in Chatterjee (2006) (see the proof of its Theorem 1.3) but with some modifications since his computations are made in case where $A(x)$ is a Wigner matrix of order N.

Let $i\in\{1,\dots,n\}$ and consider for any $j,k\in\{1,\dots,N\}$, the notations ∂_j instead of $\partial/\partial x_j^{(i)}$, ∂^2_{jk} instead of $\partial^2/\partial x_j^{(i)}\partial x_k^{(i)}$ and so on. We shall also write A instead of $A(x)$, f instead of $f(x)$, and define $G=\big(A-z\mathbf{I}\big)^{-1}$.

Note that $\partial_j A$ is the matrix with $n^{-1}\big(x_1^{(i)},\dots,x_{j-1}^{(i)},2x_j^{(i)},x_{j+1}^{(i)},\dots,x_N^{(i)}\big)$ as the j^{th} row, its transpose as the j^{th} column, and zero otherwise. Thus, the Hilbert-Schmidt norm of $\partial_j A$ is bounded as follows:

$$\|\partial_j A\|_2 = \frac{1}{n}\Big(2\sum_{k=1\,,k\neq j}^{N}|x_k^{(i)}|^2+4|x_j^{(i)}|^2\Big)^{1/2} \le \frac{2}{n}\Big(\sum_{k=1}^{N}|x_k^{(i)}|^2\Big)^{1/2}. \tag{5.1}$$

Now, for any $m,j\in\{1,\dots,N\}$ such that $m\neq j$, $\partial^2_{mj}A$ has only two non-zero entries which are equal to $1/n$, whereas if $m=j$, it has only one non-zero entry which is equal to $2/n$. Hence,

$$\|\partial^2_{mj}A\|_2 \le \frac{2}{n}. \tag{5.2}$$

Finally, note that $\partial^3_{lmj}A\equiv 0$ for any $j,m,l\in\{1,\dots,N\}$.

Now, by using (4.36), it follows that, for any $j\in\{1,\dots,N\}$,

$$\partial_j f = -\frac{1}{N}\mathrm{Tr}(G(\partial_j A)G). \tag{5.3}$$

In what follows, the notations $\sum_{\{j',m'\}=\{j,m\}}$, $\sum_{\{j',m',\ell'\}=\{j,m,\ell\}}$ and $\sum_{\{j',m',\ell',k'\}=\{j,m,\ell,k\}}$ mean respectively the sum over all permutations of $\{j,m\}$, of $\{j,m,\ell\}$ and of $\{j,m,\ell,k\}$. Therefore the first sum consists of 2 terms, the second one of 6 terms and the last one of 24 terms. Starting from (5.3) and applying repeatedly (4.36), we then derive the following cumbersome formulas for the partial derivatives up to the order four: for any $j,m,\ell,k\in\{1,\dots,N\}$,

$$\partial^2_{mj}f = \frac{1}{N}\sum_{\{j',m'\}=\{j,m\}}\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial_{m'}A)G\big) - \frac{1}{N}\mathrm{Tr}\big(G(\partial^2_{mj}A)G\big), \tag{5.4}$$

$$\begin{aligned}\partial^3_{\ell mj}f = &-\frac{1}{N}\sum_{\{j',m',\ell'\}=\{j,m,\ell\}}\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial_{m'}A)G(\partial_{\ell'}A)G\big)\\ &+\frac{1}{N}\sum_{\{j',m'\}=\{j,m\}}\mathrm{Tr}\Big(G(\partial^2_{\ell j'}A)G(\partial_{m'}A)G+G(\partial_{j'}A)G(\partial^2_{\ell m'}A)G\Big)\\ &+\frac{1}{N}\mathrm{Tr}\big(G(\partial_\ell A)G(\partial^2_{mj}A)G\big)+\frac{1}{N}\mathrm{Tr}\big(G(\partial^2_{mj}A)G(\partial_\ell A)G\big),\end{aligned} \tag{5.5}$$

and

$$\partial^4_{k\ell mj}f := I_1+I_2+I_3+I_4+I_5+I_6, \tag{5.6}$$

where

$$I_1 = \frac{1}{N}\sum_{\{j',m',\ell',k'\}=\{j,m,\ell,k\}}\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial_{m'}A)G(\partial_{\ell'}A)G(\partial_{k'}A)G\big),$$

$$\begin{aligned}I_2 = -\frac{1}{N}\sum_{\{j',m',\ell'\}=\{j,m,\ell\}}\Big(&\mathrm{Tr}\big(G(\partial^2_{kj'}A)G(\partial_{m'}A)G(\partial_{\ell'}A)G\big)+\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial^2_{km'}A)G(\partial_{\ell'}A)G\big)\\ &+\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial_{m'}A)G(\partial^2_{k\ell'}A)G\big)\Big),\end{aligned}$$

$$\begin{aligned}I_3 = &-\frac{1}{N}\sum_{\{j',m'\}=\{j,m\}}\Big(\mathrm{Tr}\big(G(\partial^2_{\ell j'}A)G(\partial_k A)G(\partial_{m'}A)G\big)+\mathrm{Tr}\big(G(\partial^2_{\ell j'}A)G(\partial_{m'}A)G(\partial_k A)G\big)\Big)\\ &-\frac{1}{N}\sum_{\{j',m'\}=\{j,m\}}\Big(\mathrm{Tr}\big(G(\partial_k A)G(\partial^2_{\ell j'}A)G(\partial_{m'}A)G\big)+\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial^2_{\ell m'}A)G(\partial_k A)G\big)\Big)\\ &-\frac{1}{N}\sum_{\{j',m'\}=\{j,m\}}\Big(\mathrm{Tr}\big(G(\partial_k A)G(\partial_{j'}A)G(\partial^2_{\ell m'}A)G\big)+\mathrm{Tr}\big(G(\partial_{j'}A)G(\partial_k A)G(\partial^2_{\ell m'}A)G\big)\Big),\end{aligned}$$

$$I_4 = -\frac{1}{N} \sum_{\{k',\ell'\}=\{k,\ell\}} \Big(\mathrm{Tr}\big(G(\partial^2_{mj}A)G(\partial_{k'}A)G(\partial_{\ell'}A)G\big) + \mathrm{Tr}\big(G(\partial_{k'}A)G(\partial^2_{mj}A)G(\partial_{\ell'}A)G\big) + \mathrm{Tr}\big(G(\partial_{k'}A)G(\partial_{\ell'}A)G(\partial^2_{mj}A)G\big)\Big)\,,$$

$$I_5 = \frac{1}{N} \sum_{\{k',\ell'\}=\{k,\ell\}} \sum_{\{j',m'\}=\{j,m\}} \mathrm{Tr}\big(G(\partial^2_{\ell' j'}A)G(\partial^2_{k'm'}A)G\big)\,,$$

and

$$I_6 = \frac{1}{N}\mathrm{Tr}\big(G(\partial^2_{mj}A)G(\partial^2_{k\ell}A)G\big) + \frac{1}{N}\mathrm{Tr}\big(G(\partial^2_{k\ell}A)G(\partial^2_{mj}A)G\big)\,.$$

We start by giving an upper bound for $\partial^2_{mj} f$. Since the eigenvalues of G^2 are all bounded by v^{-2}, then so are its entries. Then, as $\mathrm{Tr}(G(\partial^2_{mj}A)G) = \mathrm{Tr}((\partial^2_{mj}A)G^2)$, it follows that

$$|\mathrm{Tr}(G(\partial^2_{mj}A)G)| = |\mathrm{Tr}((\partial^2_{mj}A)G^2)| \le 2v^{-2}n^{-1}\,. \tag{5.7}$$

Next, to give an upper bound for $|\mathrm{Tr}\big(G(\partial_j A)G(\partial_m A)G\big)|$, it is useful to recall some properties of the Hilbert-Schmidt norm: Let $B = (b_{ij})_{1\le i,j\le N}$ and $C = (c_{ij})_{1\le i,j\le N}$ be two $N\times N$ complex matrices in $\mathcal{L}_2$, the set of Hilbert-Schmidt operators. Then

(a)- $|\mathrm{Tr}(BC)| \le \|B\|_2\|C\|_2$.

(b)- If B admits a spectral decomposition with eigenvalues $\lambda_1,\dots,\lambda_N$, then $\max\{\|BC\|_2, \|CB\|_2\} \le \max_{1\le i\le N}|\lambda_i|.\|C\|_2$.

(See e.g. Wilkinson (1965) pages 55-58, for a proof of these facts).

Using the properties of the Hilbert-Schmidt norm recalled above, the fact that the eigenvalues of G are all bounded by v^{-1}, and (5.1), we then derive that

$$\begin{aligned}|\mathrm{Tr}(G(\partial_j A)G(\partial_m A)G)| &\le \|G(\partial_j A)G\|_2.\|(\partial_m A)G\|_2 \le \|G\|.\|(\partial_j A)G\|_2.\|\partial_m A\|_2.\|G\| \\ &\le \|G\|^3.\|\partial_j A\|_2.\|\partial_m A\|_2 \le \frac{4}{v^3n^2}\sum_{k=1}^N |x_k^{(i)}|^2\,. \end{aligned}\tag{5.8}$$

Starting from (5.4) and considering (5.7) and (5.8), the first inequality of Lemma 5.1 follows.

Next, using again the above properties (a) and (b), the fact that the eigenvalues of G are all bounded by v^{-1}, (5.1) and (5.2), we get that

$$\begin{aligned}|\mathrm{Tr}(G(\partial_j A)G(\partial_m A)G(\partial_\ell A)G)| &\le \|G(\partial_j A)G(\partial_m A)G\|_2.\|(\partial_\ell A)G\|_2 \\ &\le \|G(\partial_j A)G(\partial_m A)\|_2.\|G\|^2.\|\partial_\ell A\|_2 \le \|G(\partial_j A)\|_2.\|G(\partial_m A)\|_2.\|G\|^2.\|\partial_\ell A\|_2 \\ &\le \|G\|^4.\|\partial_j A\|_2.\|\partial_m A\|_2.\|\partial_\ell A\|_2 \le \frac{8}{v^4n^3}\Big(\sum_{k=1}^N |x_k^{(i)}|^2\Big)^{3/2}\,,\end{aligned}\tag{5.9}$$

and

$$\begin{aligned}|\mathrm{Tr}(G(\partial^2_{\ell j}A)G(\partial_m A)G)| &\le \|G(\partial^2_{\ell j}A)G\|_2.\|(\partial_m A)G\|_2 \le \|G\|^2\|G(\partial^2_{\ell j}A)\|_2.\|\partial_m A\|_2 \\ &\le \|G\|^3.\|\partial^2_{\ell j}A\|_2.\|\partial_m A\|_2 \le \frac{4}{v^3n^2}\Big(\sum_{k=1}^N |x_k^{(i)}|^2\Big)^{1/2}\,.\end{aligned}\tag{5.10}$$

The same last bound is obviously valid for $|\mathrm{Tr}(G(\partial_m A)G(\partial^2_{\ell j}A)G)|$. Hence, starting from (5.5) and considering (5.9) and (5.10), the second inequality of Lemma 5.1 follows.

It remains to prove the third inequality of Lemma 5.1. Using again the above properties (a) and (b), the fact that the eigenvalues of G are all bounded by v^{-1}, (5.1) and (5.2), we infer that

$$|\mathrm{Tr}(G(\partial_j A)G(\partial_m A)G(\partial_\ell A)G(\partial_k A)G)| \le \frac{16}{v^5n^4}\Big(\sum_{k=1}^N |x_k^{(i)}|^2\Big)^2\,, \tag{5.11}$$

$$|\mathrm{Tr}(G(\partial^2_{\ell j}A)G(\partial_m A)G(\partial_k A)G)| \leq \frac{8}{v^4 n^3}\sum_{k=1}^{N}\big|x^{(i)}_k\big|^2\,, \tag{5.12}$$

and

$$|\mathrm{Tr}(G(\partial^2_{\ell j}A)G(\partial^2_{mk}A)G)| \leq \frac{4}{v^3 n^2}\,. \tag{5.13}$$

Clearly the bound (5.12) is also valid for the quantities $|\mathrm{Tr}(G(\partial_m A)G(\partial^2_{\ell j}A)G(\partial_k A)G)|$ and $|\mathrm{Tr}(G(\partial_m A)G(\partial_k A)G(\partial^2_{\ell j}A)G)|$. So, overall, starting from (5.6) and considering (5.11), (5.12) and (5.13), the third inequality of Lemma 5.1 follows. □

Acknowledgements. The authors would like to thank the referee for carefully reading the manuscript and for numerous suggestions which improved the presentation of this paper. The authors are also indebted to Djalil Chafaï for helpful discussions.

References

[1] Bai, Z. and Silverstein, J. W. (2010). Spectral analysis of large dimensional random matrices. Second edition. Springer Series in Statistics. Springer, New York.

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

[3] Chafaï, D., Guédon, O., Lecué, G. and Pajor, A. (2012). Interactions between compressed sensing, random matrices, and high dimensional geometry. To appear in Panoramas et Synthèses 38, Société Mathématique de France (SMF).

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

[5] Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* **16**, 3-22.

[6] Hall, P. and Heyde, C. C. (1980). Martingale limit theory and its application. Probability and Mathematical Statistics. Academic Press, New York-London.

[7] Horn, R. A. and Johnson, C. R. (1985). Matrix analysis. Cambridge University Press, Cambridge.

[8] Ibragimov, I. A. and Linnik, Yu. V. (1971). Independent and stationary sequences of random variables. Translation from the Russian edited by J. F. C. Kingman. Wolters-Noordhoff Publishing, Groningen.

[9] Liu, J.S. (1994). Siegel's formula via Stein's identities. *Statist. Probab. Lett.* **21**, 247-251.

[10] Marčenko, V. and Pastur, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.* **72**, 507-536.

[11] Merlevède, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariance principles for stationary sequences. *Probab. Surv.* **3**, 1-36.

[12] Neumann, M. (2011). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM Probab. Stat., published on line.*

[13] Pan, G. (2010). Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. *J. Multivariate Anal.* **101**, 1330-1338.

[14] Peligrad, M. and Utev, S. (2006). Central limit theorem for stationary linear processes. *Ann. Probab.* **34**, 1608-1622.

[15] Pfaffel, O. and Schlemm, E. (2011). Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.* **31**, 313-329.

[16] Priestley, M. B. (1988). Nonlinear and Nonstationary Time Series Analysis. Academic Press.

[17] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55**, 331-339.

[18] Silverstein, J. W. and Bai. Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.* **54**, 175-192.

[19] Wang, C., Jin, B. and Miao, B. (2011). On limiting spectral distribution of large sample covariance matrices by VARMA(p, q). *J. Time Series Anal.* **32**, 539-546.

[20] Wilkinson, J. H. (1965). The Algebraic Eigenvalue Problem. Clarendon Press, Oxford.

[21] Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102**, 14150-14154.

[22] Wu, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4**, 207-226.

[23] Yao, J. (2012). A note on a Marčenko-Pastur type theorem for time series. *Statist. Probab. Lett.* **82**, 22-28.

[24] Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20**, 50-68.