THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

École Doctorale Mathématiques et Sciences et Technologie de l'Information et de la Communication

Discipline : Mathématiques

Présentée par

Marwa Banna

Distribution spectrale limite pour des matrices à entrées corrélées et inégalité de type Bernstein

---

Directeurs de thèse : **Florence Merlevède** et **Emmanuel Rio**

Soutenue le 25 Septembre 2015
Devant le jury composé de

M. Włodzimierz Bryc	University of Cincinnati	Rapporteur
M. Bernard Delyon	Université de Rennes 1	Rapporteur
M. Olivier Guédon	Univ. Paris-Est Marne-la-Vallée	Examinateur
M. Walid Hachem	Télécom ParisTech	Examinateur
Mme. Florence Merlevède	Univ. Paris-Est Marne-la-Vallée	Directrice de thèse
M. Jamal Najim	Univ. Paris-Est Marne-la-Vallée	Examinateur
M. Emmanuel Rio	Univ. de Versailles-Saint-Quentin	Directeur de thèse

Thèse préparée au
Laboratoire LAMA CNRS UMR 8050
Université de Paris-Est Marne-la-Vallée
5, boulevard Descartes, Champs-sur-Marne
77454 Marne-la-Vallée cedex 2, France

Résumé

Cette thèse porte essentiellement sur l'étude de la distribution spectrale limite de grandes matrices aléatoires dont les entrées sont corrélées et traite également d'inégalités de déviation pour la plus grande valeur propre d'une somme de matrices aléatoires auto-adjointes et géométriquement absolument réguliers.

On s'intéresse au comportement asymptotique de grandes matrices de covariances et de matrices de type Wigner dont les entrées sont des fonctionnelles d'une suite de variables aléatoires à valeurs réelles indépendantes et de même loi. On montre que dans ce contexte la distribution spectrale empirique des matrices peut être obtenue en analysant une matrice gaussienne ayant la même structure de covariance. Cette approche est valide que ce soit pour des processus à mémoire courte ou pour des processus exhibant de la mémoire longue, et on montre ainsi un résultat d'universalité concernant le comportement asymptotique du spectre de ces matrices.

Notre approche consiste en un mélange de la méthode de Lindeberg par blocs et d'une technique d'interpolation Gaussienne. Une nouvelle inégalité de concentration pour la transformée de Stieltjes pour des matrices symétriques ayant des lignes m-dépendantes est établie. Notre méthode permet d'obtenir, sous de faibles conditions, l'équation intégrale satisfaite par la transformée de Stieltjes de la distribution spectrale limite. Ce résultat s'applique à des matrices associées à des fonctions de processus linéaires, à des modèles ARCH ainsi qu'à des modèles non-linéaires de type Volterra.

On traite également le cas des matrices de Gram dont les entrées sont des fonctionnelles d'un processus absolument régulier (i.e. β-mélangeant). On établit une inégalité de concentration qui nous permet de montrer, sous une condition de décroissance arithmétique des coefficients de β-mélange, que la transformée de Stieltjes se concentre autour de sa moyenne. On réduit ensuite le problème à l'étude d'une matrice gaussienne ayant une structure de covariance similaire via la méthode de Lindeberg par blocs. Des applications à des chaînes de Markov stationnaires et Harris récurrentes ainsi qu'à des systèmes dynamiques sont données.

Dans le dernier chapitre de cette thèse, on étudie des inégalités de déviation pour la plus grande valeur propre d'une somme de matrices aléatoires auto-adjointes. Plus précisément, on établit une inégalité de type Bernstein pour la plus grande valeur propre de la somme de matrices auto-ajointes, centrées et géométriquement β-mélangeantes dont la plus grande valeur propre est bornée. Ceci étend d'une part le résultat de Merlevède et al. (2009) à un cadre matriciel et généralise d'autre part, à un facteur logarithmique près, les résultats de Tropp (2012) pour des sommes de matrices indépendantes.

Abstract

In this thesis, we investigate mainly the limiting spectral distribution of random matrices having correlated entries and prove as well a Bernstein-type inequality for the largest eigenvalue of the sum of self-adjoint random matrices that are geometrically absolutely regular.

We are interested in the asymptotic spectral behavior of sample covariance matrices and Wigner-type matrices having correlated entries that are functions of independent random variables. We show that the limiting spectral distribution can be obtained by analyzing a Gaussian matrix having the same covariance structure. This approximation approach is valid for both short and long range dependent stationary random processes just having moments of second order.

Our approach is based on a blend of a blocking procedure, Lindeberg's method and the Gaussian interpolation technique. We also develop new tools including a concentration inequality for the spectral measure for matrices having K-dependent rows. This method permits to derive, under mild conditions, an integral equation of the Stieltjes transform of the limiting spectral distribution. Applications to matrices whose entries consist of functions of linear processes, ARCH processes or non-linear Volterra-type processes are also given.

We also investigate the asymptotic behavior of Gram matrices having correlated entries that are functions of an absolutely regular random process. We give a concentration inequality of the Stieltjes transform and prove that, under an arithmetical decay condition on the absolute regular coefficients, it is almost surely concentrated around its expectation. The study is then reduced to Gaussian matrices, with a close covariance structure, proving then the universality of the limiting spectral distribution. Applications to stationary Harris recurrent Markov chains and to dynamical systems are also given.

In the last chapter, we prove a Bernstein type inequality for the largest eigenvalue of the sum of self-adjoint centered and geometrically absolutely regular random matrices with bounded largest eigenvalue. This inequality is an extension to the matrix setting of the Bernstein-type inequality obtained by Merlevède et al. (2009) and a generalization, up to a logarithmic term, of Tropp's inequality (2012) by relaxing the independence hypothesis.

Notations

Acronyms

a.s.	almost surely
i.i.d.	independent and identically distributed
LSD	limiting spectral distribution

Linear Algebra

$\mathbf{I}_d, \mathbb{I}_d$	identity matrix of order d
A^T	transpose of a matrix A
$\mathrm{Tr}(A)$, $\mathrm{Rank}(A)$	trace and rank of a matrix A
$\|A\|$, $\|A\|_2$	spectral and Frobenius norms of a matrix A
$\mathbf{0}_{p,q}$	zero matrix of dimension $p \times q$
$\mathbf{0}_p$	zero vector of dimension p
$\|X\|_p$	$\mathbb{L}^p$ norm of a vector X
$\preceq$, $\succeq$, $\prec$, $\succ$	matrix inequalities: $A \preceq B$ means that $B - A$ is positive semi-definite, whereas $A \prec B$ means that $B - A$ is positive definite

Probability Theory

$(\Omega, \mathcal{F}, \mathcal{P})$	probability space with σ-algebra $\mathcal{F}$ and measure $\mathcal{P}$
$\mathbb{P}$	probability
$\sigma(X)$	σ-algebra generated by X
$\mathbb{E}(X)$, $\mathrm{Var}(X)$	expectation and variance of X
$\mathrm{Cov}(X, Y)$	covariance of X and Y
$\|X\|_p$	$\mathbb{L}^p$ norm of X
$X \ =^{\mathcal{D}} \ Y$	X and Y have the same distribution
$\xrightarrow{w}$	weak convergence of measures
δ_x	the Dirac measure at point x

Analysis

$\mathbb{Z}, \mathbb{R}, \mathbb{C}$	set of integers, real and complex numbers
$\mathfrak{Re}(z)$, $\mathfrak{Im}(\mathfrak{z})$	real and imaginary parts of z
$\mathbb{C}_+$	$\{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}$
$\mathbf{1}_A$	indicator function of A
$[x]$	integer part of x
$\lceil x \rceil$	the smallest integer which is larger or equal to x
$x \wedge y$, $x \vee y$	$\min\{x, y\}$ and $\max\{x, y\}$
$\limsup_{n\to\infty} x_n$	limit superior of a sequence $(x_n)_n$

Contents

Chapter 1

Introduction

The major part of the thesis is devoted to the study of the asymptotic spectral behavior of random matrices. Letting $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a sequence of $N = N(n)$-dimensional real-valued random vectors, an object of investigation will be the $N \times N$ associated *sample covariance matrix* $\mathbf{B}_n$ given by

$$\mathbf{B}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T = \frac{1}{n}\mathcal{X}_n \mathcal{X}_n^T ,$$

where $\mathcal{X}_n = (X_{i,j})_{ij}$ is the $N \times n$ matrix having $\mathbf{X}_1, \dots, \mathbf{X}_n$ as columns.

The interest in describing the spectral properties of $\mathbf{B}_n$ has emerged from multivariate statistical inference since many test statistics can be expressed in terms of functionals of their eigenvalues. This goes back to Wishart [77] in 1928, who considered sample covariance matrices with independent Gaussian entries. However, it took several years for a concrete mathematical theory of the spectrum of random matrices to begin emerging.

Letting $(X_{i,j})_{i,j}$ be an array of random variables, another object of investigation will be the following $n \times n$ *symmetric random matrix* $\mathbb{X}_n$ defined by:

$$\mathbb{X}_n := \frac{1}{\sqrt{n}} \begin{cases} X_{i,j} & \text{if } 1 \leq j \leq i \leq n \\ X_{j,i} & \text{if } 1 \leq i < j \leq n \,. \end{cases} \tag{1.1}$$

Motivated by physical applications, that were mainly due to Wigner, Dyson and Mehta in the 1950s, symmetric matrices started as well attracting attention in various fields in physics. For instance, in nuclear physics, the spectrum of large size Hamiltonians of big nuclei was regarded via that of a symmetric random matrix $\mathbb{X}_n$ with Gaussian entries. Being applied as statistical models for heavy-nuclei atoms, such matrices, known as Wigner matrices, were since widely studied.

Random matrix theory has then become a major tool in many fields, including number theory, combinatorics, quantum physics, signal processing, wireless communications, multivariate statistical analysis, finance, ...etc. It has been used as an indirect method for solving complicated problems arising from physical or mathematical systems. For this reason, it is said that random matrix theory owes its existence to its applications.

Moreover, it connects several mathematical branches by using tools from different domains including: classical analysis, graph theory, combinatorial analysis, orthogonal polynomials, free probability theory, ...etc.

Consequently, random matrix theory has become a very active mathematical domain and this lead to the appearance of several major monographs in this field [3, 5, 56, 68].

The major part of this thesis is devoted to the study of high-dimensional sample covariance and Wigner-type matrices. We shall namely study the global asymptotic behavior of their eigenvalues and focus on the identification and universality of the limiting spectral distribution. We shall investigate as well deviation inequalities of Bernstein type for the largest eigenvalue of the sum of self-adjoint matrices that are geometrically absolutely regular.

1.1 Limiting spectral distribution

We start by giving the following motivation: Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. centered random vectors with fixed dimension N and covariance matrix $\Sigma := \mathbb{E}(\mathbf{X}_1\mathbf{X}_1^T) = \ldots = \mathbb{E}(\mathbf{X}_n\mathbf{X}_n^T)$. We have by the law of large numbers

$$\lim_{n\to+\infty} \mathbf{B}_n = \mathbb{E}(\mathbf{B}_n) = \Sigma \quad \text{almost surely.}$$

A natural question is then to ask: how would $\mathbf{B}_n$ behave when both N and n tend to infinity?

We shall see, in the sequel, that when the dimension N tends to infinity with n, the spectrum of $\mathbf{B}_n$ will tend to something completely deterministic.

In order to describe the global distribution of the eigenvalues, it is convenient to introduce the empirical spectral measure and the empirical spectral distribution function:

Definition 1.1. *For a square matrix A of order N with real eigenvalues $(\lambda_k)_{1\leq k\leq N}$, the empirical spectral measure and distribution function are respectively defined by*

$$\mu_A = \frac{1}{N}\sum_{k=1}^{N}\delta_{\lambda_k} \text{ and } F^A(x) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_{\{\lambda_k\leq x\}}\,,$$

where δ_x denotes the Dirac measure at point x.

μ_A is a normalized counting measure of the eigenvalues of A. It is simply a discrete random probability measure that gives a global description of the behavior of the spectrum of A.

A typical object of interest is the study of the limit of the empirical measure when N and n tend to infinity at the same order. The first result on the limiting spectral distribution for sample covariance matrices was due to Marčenko and Pastur [44] in 1967 who proved the convergence of the empirical spectral measure to the deterministic Marčenko-Pastur law; named after them.

Theorem 1.2. *(Marčenko-Pastur theorem, [44]) Let $(X_{i,j})_{i,j}$ be an array of i.i.d. centered random variables with common variance 1. Provided that $\lim_{n\to\infty} N/n = c \in (0,\infty)$, then, almost surely, $\mu_{\mathbf{B}_n}$ converges weakly to the Marčenko-Pastur law defined by*

$$\mu_{MP}(dx) = \left(1-\frac{1}{c}\right)_+\delta_0 + \frac{1}{2\pi cx}\sqrt{(b_c-x)(x-a_c)}\mathbf{1}_{[a_c,b_c]}(x)dx\,,$$

where $a_c = (1-\sqrt{c})^2$, $b_c = (1+\sqrt{c})^2$ and $(x)_+ = \max(x,0)$.

The original Marčenko-Pastur theorem is stated for random variables having moment of fourth order; we refer to Yin [81] for the proof under moments of second order only.

The above theorem can be seen as an analogue of the law of large numbers in the sense that, almost surely, a random average converges to a deterministic quantity. It also reminds us of the central limit theorem in the sense that the limiting law is universal and depends on the distribution of the matrix entries only through their common variance.

Another way to describe the limiting spectral distribution is by identifying its Stieltjes transform.

Definition 1.3. *The Stieltjes transform of a non-negative measure μ on $\mathbb{R}$ with finite total mass is defined for any $z \in \mathbb{C}_+$ by*

$$S_\mu(z) := \int \frac{1}{x-z}\,\mu(dx)\,,$$

where we denote by $\mathbb{C}_+$ the set of complex numbers with positive imaginary part.

A very important property of the Stieltjes transform is that it characterizes the measure μ via the following inversion formula: if a and b are two points of continuity of μ, i.e. $\mu(\{a\}) = \mu(\{b\}) = 0$, then

$$\mu(]a,b[) = \lim_{y \searrow 0} \frac{1}{\pi} \int_a^b \mathfrak{Im} S_\mu(x+iy)\,dx\,.$$

We also note that, for an $N \times N$ Hermitian matrix A, the Stieltjes transform of μ_A is given for each $z = u + iv \in \mathbb{C}_+$ by

$$S_A(z) := S_{\mu_A}(z) = \int \frac{1}{x-z}\mu_A(dx) = \frac{1}{N}\mathrm{Tr}(A - z\mathbf{I})^{-1}\,,$$

where $\mathbf{I}$ is the identity matrix. We shall refer to S_A as the Stieltjes transform of the matrix A.

For a sequence of matrices A_n, the weak convergence of μ_{A_n} to a probability measure μ is equivalent to the point-wise convergence in $\mathbb{C}_+$ of $S_{A_n}(z)$ to $S_\mu(z)$:

$$\left(\mu_{A_n} \xrightarrow[n\to\infty]{w} \mu\right) \Leftrightarrow \left(\forall z \in \mathbb{C}_+,\, S_{A_n}(z) \xrightarrow[n\to\infty]{} S_\mu(z)\right).$$

For instance, Theorem 1.2 can be proved by showing that, for any $z \in \mathbb{C}_+$, $S_{\mathbf{B}_n}(z)$ converges almost surely to the Stieltjes transform $S_{\mu_{MP}}(z)$ of the Marčenko-Pastur law satisfying the following equation:

$$S_{\mu_{MP}}(z) = \frac{1}{-z + 1 - c - czS_{\mu_{MP}}(z)}\,.$$

The Stieltjes transform turns out to be a well-adapted tool for the asymptotic study of empirical measures and its introduction to random matrix theory gave birth to the well-known resolvent method, also called the Stieltjes transform method.

1.1.1 A brief literature review on covariance matrices with correlated entries

Since Marčenko-Pastur's pioneering paper [44], there has been a large amount of work aiming to relax the independence structure between the entries of $\mathcal{X}_n$. The literature is rich with results on this issue but we shall only mention certain ones that are somehow related to this thesis.

We start by the model studied initially by Yin [81] and then by Silverstein [64], who considered a linear transformation of independent random variables which leads to the study of the empirical spectral distribution of random matrices of the form:

$$\mathbf{B}_n = \frac{1}{n}\Gamma_N^{1/2}\mathcal{X}_n\mathcal{X}_n^T\Gamma_N^{1/2}\,. \tag{1.2}$$

More precisely, in the latter paper, the following theorem is proved:

Theorem 1.4. *(Theorem 1.1, [64]) Let $\mathbf{B}_n$ be the matrix defined in (1.2). Assume that:*

- $\lim_{n\to\infty} N/n = c \in (0,\infty)$,
- *$\mathcal{X}_n$ is an $N \times n$ matrix whose entries are i.i.d. centered random variables with common variance 1,*
- *Γ_N is an $N \times N$ positive semi-definite Hermitian random matrix such that F^{Γ_N} converges almost surely in distribution to a deterministic distribution H on $[0,\infty)$ as $N \to \infty$,*
- *Γ_N and $\mathcal{X}_n$ are independent.*

Then, almost surely, $\mu_{\mathbf{B}_n}$ converges weakly to a deterministic probability measure μ whose Stieltjes transform $S = S(z)$ satisfies for any $z \in \mathbb{C}_+$ the equation

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

The above equation is uniquely solvable in the class of analytic functions S in $\mathbb{C}_+$ satisfying: $-\frac{1-c}{z} + cS \in \mathbb{C}_+$.

For further investigations on the model mentioned above, one can check Silverstein and Bai [65] and Pan [54].

Another models of sample covariance matrices with correlated entries, in which the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent, have been later considered. For example, Hachem *et al* [33] consider the case where the entries are modeled by a short memory linear process of infinite range having independent Gaussian innovations.

Later, Bai and Zhou [6] derive the LSD of $\mathbf{B}_n$ by assuming a more general dependence structure:

Theorem 1.5. *(Theorem 1.1, [6]) Assume that the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent and that*

i. For all i, $\mathbb{E}(X_{ki}X_{\ell i}) = \gamma_{k,\ell}$ and for any deterministic matrix $N \times N$, $R = (r_{k\ell})$, with bounded spectral norm

$$\mathbb{E}\Big|\mathbf{X}_i^T R \mathbf{X}_i - \mathrm{Tr}(R\Gamma_N)\Big|^2 = o(n^2) \quad \text{where } \Gamma_N = (\gamma_{k,\ell})$$

ii. $\lim_{n\to\infty} N/n = c \in (0, \infty)$

iii. The spectral norm of Γ_N is uniformly bounded and μ_{Γ_N} converges in law to μ_H.

Then, almost surely, $\mu_{\mathbf{B}_n}$ converges in law to a non random probability measure whose Stieltjes transform $S = S(z)$ satisfies the equation : for all $z \in \mathbb{C}^+$

$$z = -\frac{1}{\underline{S}} + c \int \frac{t}{1 + \underline{S}t} d\mu_H(t) \, ,$$

where $\underline{S}(z) := -(1-c)/z + cS(z)$.

For the purpose of applications, Bai and Zhou prove, in Corollary 1.1 of [6], that Assumption (i.) is verified as soon as

- $n^{-1} \max_{k \neq \ell} \mathbb{E}\Big(X_{ki}X_{\ell i} - \gamma_{k,\ell}\Big)^2 \to 0$ uniformly in $i \leq n$
- $n^{-2} \sum_{\Lambda} \Big(\mathbb{E}(X_{ki}X_{\ell i} - \gamma_{k,\ell})(X_{ik'}X_{i\ell'} - \gamma_{k',\ell'})\Big)^2 \to 0$ uniformly in $i \leq n$

where

$$\Lambda = \{(k,\ell,k',\ell') : 1 \leq k,\ell,k',\ell' \leq p\} \backslash \{(k,\ell,k',\ell') : k = k' \neq \ell = \ell' \text{ or } k = \ell' \neq k' = \ell\} .$$

They also give possible applications of their result and establish the limiting spectral distribution for Spearman's rank correlation matrices, sample covariance matrices for finite populations and sample covariance matrices generated by causal AR(1) models.

Another application of Bai and Zhou's result is the following: let $(\varepsilon_k)_k$ be a sequence of i.i.d. centered random variables with common variance 1 and let $(X_k)_k$ be the linear process defined by

$$X_k = \sum_{j=0}^{\infty} a_j \varepsilon_{k-j} ,$$

with $(a_k)_k$ being a linear filter of real numbers. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be *independent copies* of the N-dimensional vector $(X_1, \ldots, X_N)^T$ and consider the associated sample covariance matrix $\mathbf{B}_n$.

For this model, Yao [80] then shows that the hypotheses of Theorem 1.5 are satisfied if:

- $\lim_{n\to\infty} N/n = c \in (0,\infty)$,
- the error process has a fourth moment: $\mathbb{E}\varepsilon_1^4 < \infty$,
- the linear filter $(a_k)_k$ is absolutely summable, $\sum_{k=0}^{\infty} |a_k| < \infty$,

and proves that, almost surely, $\mu_{\mathbf{B}_n}$ converges weakly to a non-random probability measure μ whose Stieltjes transform $S = S(z)$ satisfies for any $z \in \mathbb{C}_+$ the equation

$$z = -\frac{1}{\underline{S}} + \frac{c}{2\pi} \int_0^{2\pi} \frac{1}{\underline{S} + \Big(2\pi f(\lambda)\Big)^{-1}} \, d\lambda \,, \tag{1.3}$$

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

$$f(\lambda) = \frac{1}{2\pi}\sum_k \operatorname{Cov}(X_0, X_k)\, \mathrm{e}^{\mathrm{i}k\lambda} \qquad \text{for } \lambda \in [0, 2\pi[.$$

Still in the context of the linear model described above, Pfaffel and Schlemm [56] relax the equidistribution assumption on the innovations and derive the limiting spectral distribution of $\mathbf{B}_n$. They use a different approach than the one considered in [6] and [80] but they still assume the finiteness of the *fourth* moments of the innovations plus a polynomial decay of the coefficients of the underlying linear process.

We also mention that Pan et al. [55] relax the moment conditions and derive the limiting spectral distribution by just assuming the finiteness of moments of second order. This result will be a consequence of our Theorem 2.2 and shall be given in Section 2.2.1.

We finally note that Assumption (i.) is also satisfied when considering Gaussian vectors or isotropic vectors with log-concave distribution (see [53]); however, it is hard to be verified for nonlinear time series, as ARCH models, without assuming conditions on the rate of convergence of the mixing coefficients of the underlying process.

1.1.2 Sample covariance matrices associated with functions of i.i.d. random variables

An object of investigation of this thesis will be the asymptotic spectral behavior of sample covariance matrices $\mathbf{B}_n$ associated with functions of i.i.d. random variables. Mainly, we shall suppose that the entries of $\mathcal{X}_n = (X_{k,\ell})_{k\ell}$ consist of one of the following forms of stationary processes:

Let $(\xi_{i,j})_{(i,j)\in\mathbb{Z}^2}$ be an array of i.i.d. real-valued random variables and let $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be the stationary process defined by

$$X_{k,\ell} = g(\xi_{k-i,\ell}\ ;\ i \in \mathbb{Z}), \tag{1.4}$$

or by

$$X_{k,\ell} = g(\xi_{k-i,\ell-j}\ ;\ (i,j) \in \mathbb{Z}^2), \tag{1.5}$$

where g is real-valued measurable function such that

$$\mathbb{E}(X_{k,\ell}) = 0 \quad \text{and} \quad \mathbb{E}(X_{k,\ell}^2) < \infty\,.$$

This framework is very general and includes widely used linear and non-linear processes. We mention, for instance, functions of linear processes, ARCH models and non-linear Volterra models as possible examples of stationary processes of the above forms. We also refer to the papers [78, 79] by Wu for more applications.

Following Priestley [61] and Wu [78], $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ can be viewed as a physical system with the $\xi_{i,j}$'s being the input, $X_{k,\ell}$ the output and g the transform or data-generating mechanism.

We are interested in studying the asymptotic behavior of $\mathbf{B}_n$ when both n and N tend to infinity and are such that $\lim_{n\to\infty} N/n = c \in (0,\infty)$. With this aim, we follow a different approach consisting of approximating $\mathbf{B}_n$ with a sample covariance matrix $\mathbf{G}_n$, associated with a Gaussian process $(Z_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ having the same covariance structure as $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$, and then using the Gaussian structure of $\mathbf{G}_n$, we establish the limiting distribution.

This shall be done by comparing the Stieltjes transform of $\mathbf{B}_n$ by the expectation of that of $\mathbb{G}_n$. Indeed, if we prove that for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} |S_{\mathbf{B}_n}(z) - \mathbb{E}(S_{\mathbf{G}_n}(z))| = 0 \quad \text{a.s.}$$

then the study is reduced to proving the convergence of $\mathbb{E}(S_{\mathbf{G}_n}(z))$ to the Stieltjes transform of a non-random probability measure, say μ.

We note that if the $X_{k,\ell}$'s are defined as in (1.4) then the columns of $\mathcal{X}_n$ are independent and it follows, in this case, by Guntuboyina and Leeb's concentration inequality [32] of the spectral measure that for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} |S_{\mathbf{B}_n}(z) - \mathbb{E}(S_{\mathbf{B}_n}(z))| = 0 \quad \text{a.s.} \tag{1.6}$$

which reduces the study to proving that for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} |\mathbb{E}(S_{\mathbf{B}_n}(z)) - \mathbb{E}(S_{\mathbf{G}_n}(z))| = 0\,. \tag{1.7}$$

However, in the case where the matrix entries consist of the stationary process defined in (1.5), the independence structure between the rows or the columns of $\mathcal{X}_n$ is no longer valid. Thus, the concentration inequality given by Guntuboyina and Leeb does not apply anymore.

The convergence (1.6) can be however proved by approximating first $\mathcal{X}_n$ with an m_n-dependent block matrix and then using a concentration inequality for the spectral measure of matrices having m_n-dependent columns (see Section 3.7).

As we shall see in Theorems 2.1 and 3.5, the convergence (1.7) always holds without any conditions on the covariance structure of $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$. This shows a universality scheme for the limiting spectral distribution of $\mathbf{B}_n$, as soon as the $X_{k,\ell}$'s have the dependence structure (1.4) or (1.5), without demanding any rate of convergence to zero of the correlation between the matrix entries.

The convergence (1.7) can be achieved via a Lindeberg method by blocks as described in Sections 2.3.2.1 and 3.8. The Lindeberg method consists in writing the difference of the expectation of the Stieltjes transforms of $\mathbf{B}_n$ and $\mathbf{G}_n$ as a telescoping sum and using Taylor expansions. This method can be used in the context of random matrices since the Stieltjes transform admits partial derivatives of all orders as shown in Sections A.1 and A.2.

In the traditional Lindeberg method, the telescoping sums consist of replacing the random variables involved in the partial sum, one at a time, by Gaussian random variables. While here, we shall replace blocks of entries, one at a time, by Gaussian blocks having the same covariance structure.

The Lindeberg method is popular with these types of problems. It is known to be an efficient tool to derive limit theorems and, up to our knowledge, it has been used for the first time in the context of random matrices by Chatterjee [20] who treated random matrices with exchangeable entries and established their limiting spectral distribution.

As a conclusion, $S_{\mathbf{B}_n}$ converges almost surely to the Stiletjes transform S of a non-random probability measure as soon as $\mathbb{E}(S_{\mathbf{G}_n}(z))$ converges to S.

1.1.3 Gram matrices associated with functions of β-mixing random variables

Assuming that the $\mathbf{X}_k$'s are independent copies of the vector $\mathbf{X} = (X_1, \ldots, X_N)^T$ can be viewed as repeating independently an N-dimensional process n times to obtain the $\mathbf{X}_k$'s. However, in practice it is not always possible to observe a high dimensional process several times. In the case where only one observation of length Nn can be recorded, it seems reasonable to partition it into n dependent observations of length N, and to treat them as n *dependent* observations. In other words, it seems reasonable to consider the $N \times n$ matrix $\mathcal{X}_n$ defined by

$$\mathcal{X}_n = \begin{pmatrix} X_1 & X_{N+1} & \cdots & X_{(n-1)N+1} \\ X_2 & X_{N+2} & \cdots & X_{(n-1)N+2} \\ \vdots & \vdots & & \vdots \\ X_N & X_{2N} & \cdots & X_{nN} \end{pmatrix}$$

and study the asymptotic behavior of its associated Gram matrix $\mathbf{B}_n$ given by

$$\mathbf{B}_n = \frac{1}{n}\mathcal{X}_n\mathcal{X}_n^T = \frac{1}{n}\sum_{k=1}^{n}\mathbf{X}_k\mathbf{X}_k^T$$

where for any $k = 1, \ldots, n$, $\mathbf{X}_k = (X_{(k-1)N+1}, \ldots, X_{kN})^T$.

Up to our knowledge this was first done by Pfaffel and Schlemm [60] who showed that this approach is valid and leads to the correct asymptotic eigenvalue distribution of the sample covariance matrix if the components of the underlying process are modeled as short memory linear filters of independent random variables. Assuming that the innovations have finite fourth moments and that the coefficients of the linear filter decay arithmetically, they prove that Stieltjes transform of the limiting spectral distribution of $\mathbf{B}_n$ satisfies (1.3).

In chapter 4, we shall relax the dependence structure of this matrix by supposing that its entries consist of functions of absolutely regular random variables. Before fully introducing the model, let us recall the definition of the absolute regular or β-mixing coefficients:

Definition 1.6. *(Rozanov and Volkonskii [72]) The absolutely regular or the β-mixing coefficient between two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined by*

$$\beta(\mathcal{A},\mathcal{B}) = \frac{1}{2}\sup\left\{\sum_{i\in I}\sum_{j\in J}\Big|\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\Big|\right\},$$

where the supremum is taken over all finite partitions $(A_i)_{i\in I}$ and $(B_j)_{j\in J}$ that are respectively $\mathcal{A}$ and $\mathcal{B}$ measurable.

The coefficients $(\beta_n)_{n\geq 0}$ of a sequence $(\varepsilon_i)_{i\in\mathbb{Z}}$ are defined by

$$\beta_0 = 1 \quad \text{and} \quad \beta_n = \sup_{k\in\mathbb{Z}}\beta\Big(\sigma(\varepsilon_\ell\,,\ \ell\leq k)\ ,\ (\varepsilon_{\ell+n}\,,\ \ell\geq k)\Big) \quad \text{for} \quad n\geq 1 \tag{1.8}$$

and $(\varepsilon_i)_{i\in\mathbb{Z}}$ is then said to be absolutely regular or β-mixing if $\beta_n\to 0$ as $n\to\infty$.

We note that absolutely regular processes exist widely. For example, a strictly stationary Markov process is β-mixing if and only if it is an aperiodic recurrent Harris chain. Moreover, many common time series models are β-mixing and the rates of decay of the associated β_k coefficients are known given the parameters of the process. Among the processes for which such knowledge is available are ARMA models [49] and certain dynamical systems and Markov processes. One can also check [25] for an overview of such results.

We shall consider a more general framework than functions of i.i.d. random variables and define the non-causal stationary process $(X_k)_{k\in\mathbb{Z}}$ as follows: for any $k\in\mathbb{Z}$ let

$$X_k = g(\ldots,\varepsilon_{k-1},\varepsilon_k,\varepsilon_{k+1},\ldots)\,, \tag{1.9}$$

where $(\varepsilon_i)_{i\in\mathbb{Z}}$ is an absolutely regular stationary process and g a measurable function from $\mathbb{Z}$ to $\mathbb{R}$ such that

$$\mathbb{E}(X_k) = 0 \quad \text{and} \quad \mathbb{E}(X_k^2) < \infty\,.$$

The interest is again to study the limiting spectral distribution of the sample covariance matrix $\mathbf{B}_n$ associated with $(X_k)_{k\in\mathbb{Z}}$ when N and n tend to infinity and are such that $\lim_{n\to\infty} N/n = c\in(0,\infty)$.

The first step consists of proving that, under the following arithmetical decay condition on the β-mixing coefficients:

$$\sum_{n\geq 1} \frac{\log(n)^{\frac{3\alpha}{2}}}{\sqrt{n}} \beta_n < \infty \quad \text{for some } \alpha > 1\,,$$

the Stieltjes transform is concentrated almost surely around its expectation as n tends to infinity. This shall be achieved by proving, with the help of Berbee's coupling lemma [11] (see Lemma A.16), a concentration inequality of the empirical spectral measures.

The study is then reduced to proving that the expectation of the Stieltjes transform converges to that of a non-random probability measure. This can be achieved by approximating it with the expectation of the Stieltjes transform of a Gaussian matrix having a close covariance structure. We shall namely prove for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} |\mathbb{E}(S_{\mathbf{B}_n}(z)) - \mathbb{E}(S_{\mathbf{G}_n}(z))| = 0\,, \tag{1.10}$$

with $\mathbf{G}_n$ being the sample covariance matrix given by

$$\mathbf{G}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{Z}_k \mathbf{Z}_k^T$$

and $\mathbf{Z}_1, \dots \mathbf{Z}_n$ being independent copies of the Gaussian vector $\mathbf{Z} = (Z_1, \dots, Z_N)^T$ where $(Z_k)_k$ is a Gaussian process having the same covariance structure as $(X_k)_k$.

The above approximation is again proved via the Lindeberg method by blocks and is done without requiring any rate of convergence to zero of the correlation between the entries nor of the β-mixing coefficients.

Therefore, provided that the β-coefficients satisfy the above arithmetical decay condition, we prove that $\mathbf{B}_n$, being the sum of dependent rank-one matrices, has the same asymptotic spectral behavior as a Gram matrix, being the sum of *independent Gaussian* rank-one matrices.

Finally, provided that the spectral density of $(X_k)_k$ exists, we prove that almost surely, $\mu_{\mathbf{B}_n}$ converges weakly to the non-random limiting probability measure whose Stieltjes transform satisfies equation (1.3).

The first chapters of this thesis are devoted to the study of the asymptotic global behavior of eigenvalues of different models of matrices with correlated entries, while the last chapter is devoted to the study of deviation inequalities for the largest eigenvalue of the sum of weakly dependent self-adjoint random matrices.

1.2 Bernstein type inequality for dependent random matrices

As we have mentioned, the analysis of the spectrum of large matrices has known significant development recently due to its important role in several domains. Another important question is to study the fluctuations of a Hermitian matrix $\mathbb{X}$ from its expectation measured by its largest eigenvalue. Matrix concentration inequalities give probabilistic bounds for such fluctuations and provide effective methods for studying several models.

For a family $(\mathbb{X}_i)_{i\geq 1}$ of $d \times d$ self-adjoint centered random matrices, it is quite interesting to give, for any $x > 0$, upper bounds of the probability

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) \geq x\Big),$$

where $\lambda_{\max}$ denotes the maximum eigenvalue of $\sum_{i=1}^{n}\mathbb{X}_i$. In the scalar case, that is for $d = 1$, this is the probability that the sum of random variables trespasses a certain positive number x.

There are several kinds of inequalities providing exponential bounds for the probability of large deviations of a sum of random variables with bounded increments. For instance, the Bernstein inequality permits to estimate such probability by a monotone decreasing exponential function in terms of the variance of the sum's increments.

The starting point to get such exponential bounds is the following Chernoff bound: denoting by $(X_i)_i$ a sequence of real-valued random variables, we have for any $x > 0$

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \geq x\Big) \leq \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot \mathbb{E}\exp\Big(t\sum_{i=1}^{n} X_i\Big)\Big\}. \tag{1.11}$$

The Laplace transform method, which is due to Bernstein in the scalar case, is generalized to the sum of independent Hermitian random matrices by Ahlswede and

Winter. They prove in the Appendix of [2] that the usual Chernoff bound has the following counterpart in the matrix setting:

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) \geq x\Big) \leq \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^{n}\mathbb{X}_i\Big)\Big\}. \tag{1.12}$$

We note that the Laplace transform of the sum of random variables appearing in (1.11) is replaced by the trace of the Laplace transform of the sum of random matrices in (1.12). Obviously, the main problem now is to give a suitable bound for

$$L_n(t) := \mathbb{E}\mathrm{Tr}\exp(t\sum_{i=1}^{n}\mathbb{X}_i).$$

As matrices do not commute, many tools, available in the scalar setting to get an upper bound of $L_n(t)$, cannot be straightforward extended. We give in Section A.4 some preliminary materials on some operator functions and some tools used in the matrix setting.

In the *independent* case, Ahlswede and Winter [2] prove by applying the Golden-Thompson inequality [29, 69], Lemma A.12, that for any $t > 0$

$$\begin{aligned}
L_n(t) &\leq \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\mathbb{X}_n}\cdot\mathrm{e}^{t\sum_{i=1}^{n-1}\mathbb{X}_i}\Big) = \mathrm{Tr}\Big(\mathbb{E}(\mathrm{e}^{t\mathbb{X}_n})\cdot\mathbb{E}\Big(\mathrm{e}^{t\sum_{i=1}^{n-1}\mathbb{X}_i}\Big)\Big)\\
&\leq \lambda_{\max}(\mathbb{E}\mathrm{e}^{t\mathbb{X}_n})\cdot\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{i=1}^{n-1}\mathbb{X}_i}\Big)\\
&\leq d\cdot\prod_{i=1}^{n}\lambda_{\max}(\mathbb{E}\mathrm{e}^{t\mathbb{X}_i}),
\end{aligned}$$

where we note that the equality in the first line follows by the independence of the $\mathbb{X}_i$'s.

Following an approach based on Lieb's concavity theorem (Theorem 6, [42]), Tropp improves, in [70], the above bound and gets for any $t > 0$

$$L_n(t) = \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^{n}\mathbb{X}_i\Big) \leq \mathrm{Tr}\exp\Big(\sum_{i=1}^{n}\log\mathbb{E}\mathrm{e}^{t\mathbb{X}_i}\Big). \tag{1.13}$$

This bound, combined with another one on $\mathbb{E}\mathrm{e}^{t\mathbb{X}_i}$, allows him to prove the following Bernstein type inequality for *independent* self-adjoint matrices:

Theorem 1.7. *(Theorem 6.3, [70]) Consider a family $\{X_i\}_i$ of independent self-adjoint random matrices with dimension d. Assume that each matrix satisfies*

$$\mathbb{E}\mathbb{X}_i = 0 \quad \text{and} \quad \lambda_{\max}(\mathbb{X}_i) \leq M \ \text{a.s.}$$

Then for any $x > 0$,

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^{n} \mathbb{X}_i\Big) \geq x\Big) \leq d \cdot \exp\Big(-\frac{x^2/2}{\sigma^2 + xM/3}\Big),$$

where $\sigma^2 := \lambda_{\max}\Big(\sum_{i=1}^{n} \mathbb{E}(\mathbb{X}_i^2)\Big)$.

Let us mention that extensions of the so-called Hoeffding-Azuma inequality for matrix martingales and of the so-called McDiarmid bounded difference inequality for matrix-valued functions of independent random variables are also given in [70].

Taking another direction, Mackey et al. [43] extend to the matrix setting Chatterjee's technique for developing scalar concentration inequalities via Stein's method of exchangeable pairs [19, 21], and established Bernstein and Hoeffding inequalities as well as other concentration inequalities. Following this approach, Paulin et al. [57] established a so-called McDiarmid inequality for matrix-valued functions of dependent random variables under conditions on the associated Dobrushin interdependence matrix.

1.2.1 Geometrically absolutely regular matrices

We shall extend the above Bernstein-type inequality for a class of dependent matrices. We note that in this case, the first step of Ahlswede and Winter's iterative procedure as well as Tropp's concave trace function method fail. Therefore additional transformations on the Laplace transform have to be made.

Even in the scalar dependent case, obtaining sharp Bernstein-type inequalities is a challenging problem and a dependence structure of the underlying process has obviously to be precise. Consequently, obtaining such an inequality for the largest eigenvalue of the sum of n self-adjoint dependent random matrices can be more challenging and technical due to the difficulties arising from both the dependence and the non-commutative structure.

We obtain, in this thesis, a Bernstein-type inequality for the largest eigenvalue of partial sums associated with self-adjoint *geometrically absolutely regular* random matrices. We note that this kind of dependence cannot be compared to the dependence structure imposed in [43] or [57].

We say that a sequence $(\mathbb{X}_i)_i$ of $d \times d$ matrices is *geormetrically* absolutely regular if there exists a positive constant c such that for any $k \geq 1$

$$\beta_k = \sup_j \beta(\sigma(\mathbb{X}_i\,,\, i \leq j), \sigma(\mathbb{X}_i\,,\, i \geq j+k)) \leq \mathrm{e}^{-c(k-1)} \tag{1.14}$$

with β being the absolute regular mixing coefficient given in Definition 1.6.

We note that the absolute regular coefficients can be computed in many situations. We refer to the work by Doob [24] for sufficient conditions on Markov chains to be geometrically absolutely regular or by Mokkadem [50] for mild conditions ensuring ARMA vector processes to be also geometrically β-mixing.

Clearly, the dependence between two ensembles of matrices depends on the gap separating the σ-algebras generated by these ensembles. For geometrically β-mixing matrices, this dependence decreases exponentially with the gap separating them.

In Chapter 5, we prove that if $(\mathbb{X}_i)_i$ is a sequence of $d \times d$ Hermitian matrices satisfying (1.14) and such that

$$\mathbb{E}(\mathbb{X}_i) = \mathbf{0} \quad \text{and} \quad \lambda_{\max}(\mathbb{X}_i) \leq 1 \quad \text{a.s.}$$

then for any $x > 0$

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^n \mathbb{X}_i\Big) \geq x\Big) \leq d \exp\Big(-\frac{Cx^2}{v^2 n + c^{-1} + x(\log n)^2}\Big),$$

where C is a universal constant and v^2 is given by

$$v^2 = \sup_{K \subseteq \{1,\dots,n\}} \frac{1}{\mathrm{Card} K} \lambda_{\max}\Big(\mathbb{E}\Big(\sum_{i \in K} \mathbb{X}_i\Big)^2\Big).$$

The full announcement of the above inequality is given in Theorem 5.1.

We note that for $d = 1$, we re-obtain the best Bernstein-type inequality so far, for geometrically absolutely regular random variables, proved by Merlevède et al. [46].

Therefore, our inequality can be viewed as an extension to the matrix setting of the Bernstein-type inequality obtained by Merlevède et al. and as a generalization, up to a logarithmic term, of Theorem 1.7 by Tropp from *independent* to geometrically absolutely regular matrices.

We note that an extra *logarithmic* factor appearing in our inequality, with respect to the independent case, cannot be avoided even in the scalar case. Indeed, Adamczak proves in Theorem 6 and Section 3.3 of [1] a Bernstein-type inequality for the partial sum associated with bounded functions of a geometrically ergodic Harris recurrent Markov chain. He shows that even in this context where it is possible to go back to the independent setting by creating random i.i.d. cycles, a *logarithmic* factor cannot be avoided.

1.2.2 Strategy of the proof

The starting point is still, as for independent matrices, the matrix Chernoff bound (1.12) which remains valid in the dependent case. However, the procedures used in [2] and [70] fail for dependent matrices from the very first step and thus another approach should be followed.

The first step will be creating gaps between the matrices considered with the aim of decoupling them in order to break their dependence structure. As done by Merlevède et al. [46, 47], we shall partition the n matrices in blocks indexed by a Cantor type set, say K_n plus a remainder:

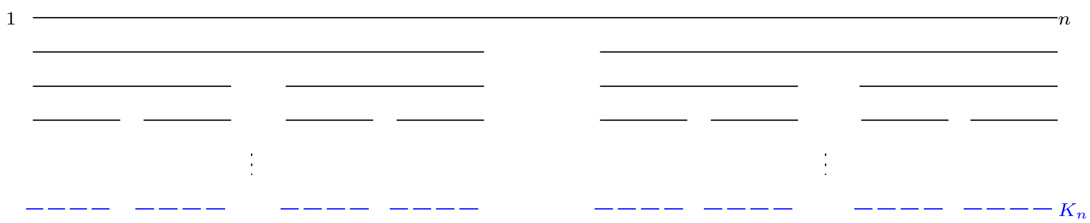

Figure 1.1 – Construction of the Cantor-type set K_n

The above figure gives an idea on how the Cantor-type set K_n is constructed. The first line of Figure 1.1 represents the set of indices $\{1,\dots,n\}$, whereas the last one represents K_n which consists of the union of disjoint subsets of consecutive integers represented by the blue dashes. This construction is done in an iterative way and will be explained in details in Section 5.3.2 of the chapter 5.

The main step is then to control the log-Laplace transform of the partial sum on K_n. By doing so, we are only considering blocks of random matrices separated by gaps of certain width. The larger the gap between two blocks is, the less dependent the associated matrices are. The remaining blocks of matrices, i.e. those indexed by $i \in \{1,\dots,n\}\backslash K_n$, are re-indexed and a new Cantor type set is similarly constructed. We repeat the same procedure until we cover all the matrices.

Having controlled the matrix log-Laplace transform of the partial sum on each Cantor-type set, the log-Laplace transform of the total partial sum is then handled with the help of the following Lemma:

Lemma 1.8. *Let $\mathbb{U}_0, \mathbb{U}_1, \dots$ be a sequence of $d\times d$ self-adjoint random matrices. Assume that there exist positive constants $\sigma_0, \sigma_1, \dots$ and $\kappa_0, \kappa_1, \dots$ such that, for any $i \geq 0$ and any t in $[0, 1/\kappa_i[$,*

$$\log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\,\mathbb{U}_i}\Big) \leq C_d + (\sigma_i t)^2/(1-\kappa_i t)\,,$$

where C_d is a positive constant depending only on d. Then, for any positive integer m and any t in $[0, 1/(\kappa_0+\kappa_1+\cdots+\kappa_m)[$,

$$\log \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k=0}^{m}\mathbb{U}_k\Big) \leq C_d + (\sigma t)^2/(1-\kappa t),$$

where $\sigma = \sigma_0+\sigma_1+\cdots+\sigma_m$ and $\kappa = \kappa_0+\kappa_1+\cdots+\kappa_m$.

This lemma, which is due to Merlevède et al. in the scalar case, provides bounds for the log-Laplace transform of any sum of self-adjoint random matrices and thus allows us to control the total sum after controlling the partial sum on each Cantor set separately. So obviously, the main step is to obtain suitable upper bounds of the log-Laplace transform of the partial sum on certain Cantor-type sets:

$$\log \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k\in K_A}\mathbb{X}_k\Big)$$

where K_A denotes a Cantor-like set constructed from $\{1, \dots, A\}$.

To benefit from some ideas developed in [2] or [70], we shall rather bound the matrix log-Laplace transform by that of the sum of certain *independent* and self-adjoint random matrices plus a small error term. Lemma 5.7 is in this direction and can be viewed as a decoupling lemma for the Laplace transform in the matrix setting.

As we shall see, a well-adapted dependence structure allowing such a procedure is the absolute regularity structure. This structure allows, by Berbee's coupling lemma stated in Lemma A.17 a "good coupling" in terms of the absolute regular coefficients even when the variables take values in a high dimensional space. In fact, working with $d \times d$ random matrices can be viewed as working with random vectors of dimension d^2.

For such dependence structures, the approach followed by Merlevède et al. [46, 47] is well-performing; however, its extension to the matrix setting is not straight forward because several tools used in the scalar case are no longer valid in the non-commutative case.

For instance, this dependence structure allows the following control in the scalar case: consider any index sets Q and Q' of natural numbers separated by a gap of width at least n; i.e., there exists p such that $Q \subset [1, p]$ and $Q' \subset [n + p, \infty)$. Then for any $t > 0$,

$$\begin{aligned} &\mathbb{E}\Big(\mathrm{e}^{t\sum_{i\in Q} X_i}\mathrm{e}^{t\sum_{i\in Q'} X_i}\Big) \\ &\qquad \leq \mathbb{E}\Big(\mathrm{e}^{t\sum_{i\in Q} X_i}\Big)\mathbb{E}\Big(\mathrm{e}^{t\sum_{i\in Q'} X_i}\Big) + \varepsilon(n)\Big\|\mathrm{e}^{t\sum_{i\in Q} X_i}\Big\|_\infty\Big\|\mathrm{e}^{t\sum_{i\in Q'} X_i}\Big\|_\infty, \end{aligned} \tag{1.15}$$

where $\varepsilon(n)$ is a sequence of positive real numbers depending on the dependence coefficients. The binary tree structure of the Cantor-type sets allows iterating the decorrelation procedure mentioned above to suitably handle the log-Laplace transform of the partial sum of real-valued random variables on each of the Cantor-type sets.

Iterating a procedure as (1.15) in the matrix setting cannot lead to suitable exponential inequalities essentially due to the fact that the extension of the Golden-Thompson inequality [29, 69], Lemma A.12, to three or more Hermitian matrices fails. This can add more difficulty to the non-commutative case and can complicate the proof because more coupling arguments and computations are required.

The decoupling lemma 5.7 associated with additional coupling arguments will then allow us to prove our key Proposition 5.6 giving a bound for the Laplace transform of the partial sum, indexed by Cantor-type set, of self-adjoint random matrices.

1.3 Organisation of the thesis and references

This thesis is organized as follows: in Chapter 2 we consider sample covariance matrices associated with functions of i.i.d. random variables and explain precisely the approach followed for deriving the limiting spectral distribution. The main steps of this approach were initiated in the following paper:

- M. Banna and F. Merlevède. Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *Journal of Theoretical Probability*, pages 1–39, 2013.

In Chapter 3, we study the asymptotic behavior of symmetric and Gram matrices whose entries, being functions of i.i.d. random variables, are correlated across both rows and columns. We investigate again, in Chapter 4, the limiting spectral distribution but by considering this time Gram matrices associated with functions of β-mixing random variables. The results of Chapters 3 and 4 are respectively contained in the following papers:

- M. Banna, F. Merlevède, and M. Peligrad. On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Processes and their Applications*, 125(7):2700–2726, 2015.

- M. Banna. Limiting spectral distribution of gram matrices associated with functionals of β-mixing processes. *Journal of Mathematical Analysis and Applications*, 433(1):416 – 433, 2016.

In Chapter 5, we give a Bernstein-type inequality for the largest eigenvalue of the sum of geometrically absolutely regular matrices. This is the result of the recently submitted paper:

- M. Banna, F. Merlevède, and P. Youssef. Bernstein type inequality for a class of dependent random matrices. *arXiv preprint arXiv:1504.05834*, 2015.

Finally, we collect, in the Appendix, some technical and preliminary lemmas used throughout the thesis.

Chapter 2

Matrices associated with functions of i.i.d. variables

We consider, in this chapter, the following sample covariance matrix

$$\mathbf{B}_n = \frac{1}{n}\sum_{k=1}^{n}\mathbf{X}_k\mathbf{X}_k^T \tag{2.1}$$

and suppose that $\mathbf{X}_k$'s are independent copies of the N-dimensional vector $(X_1, \dots, X_N)$ where $(X_k)_{k\in\mathbb{Z}}$ is a stationary process 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 $k \in \mathbb{Z}$,

$$X_k = g(\xi_k) \quad \text{with} \quad \xi_k := (\dots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \dots) \tag{2.2}$$

is a proper random variable and

$$\mathbb{E}(g(\xi_k)) = 0 \quad \text{and} \quad \|g(\xi_k)\|_2 < \infty\,.$$

We shall investigate the limiting spectral distribution of $\mathbf{B}_n$ via an approach based on a blend of a blocking procedure, Lindeberg's method and the Gaussian interpolation technique. The main idea in our approach is to approximate $\mathbf{B}_n$ with a Gaussian sample covariance matrix $\mathbf{G}_n$ having the same covariance structure as $\mathbf{B}_n$ and then use the Gaussian structure of $\mathbf{G}_n$ to establish the limiting distribution.

We note that even though this model is a particular case of the models studied in Chapters 3 and 4, we shall treat it separately with the aim of giving a clearer proof enabling us to shed light on the main steps of the approximations and computations involved in this approach. Moreover, some of this chapter's contents will be useful materials for Chapter 4.

2.1 Main result

Let us start by describing precisely our the model. For a positive integer n, we consider n independent copies of the sequence $(\varepsilon_k)_{k\in\mathbb{Z}}$ which we denote by $(\varepsilon_k^{(i)})_{k\in\mathbb{Z}}$ for $i = 1, \ldots, n$. Setting

$$\xi_k^{(i)} = \left(\ldots, \varepsilon_{k-1}^{(i)}, \varepsilon_k^{(i)}, \varepsilon_{k+1}^{(i)}, \ldots\right) \quad \text{and} \quad X_k^{(i)} = g(\xi_k^{(i)}) \,,$$

it follows that $(X_k^{(1)})_k, \ldots, (X_k^{(n)})_k$ are n independent copies of the process $(X_k)_{k\in\mathbb{Z}}$. Let $N = N(n)$ be a sequence of positive integers and define for any $i = 1, \ldots, n$, the vector

$$\mathbf{X}_i = (X_1^{(i)}, \ldots, X_N^{(i)})^T \,.$$

Let $\mathbf{G}_n$ be the sample covariance matrix associated with a Gaussian process $(Z_k)_{k\in\mathbb{Z}}$ having the same covariance structure as $(X_k)_{k\in\mathbb{Z}}$. Namely, for any $k, \ell \in \mathbb{Z}$,

$$\mathrm{Cov}(Z_k, Z_\ell) = \mathrm{Cov}(X_k, X_\ell) \,. \tag{2.3}$$

For $i = 1, \ldots, n$, we denote by $(Z_k^{(i)})_{k\in\mathbb{Z}}$ an independent copy of $(Z_k)_k$ that is also independent of $(X_k)_k$ and we define the $N \times N$ sample covariance matrix $\mathbf{G}_n$ by

$$\mathbf{G}_n = \frac{1}{n} \mathcal{Z}_n \mathcal{Z}_n^T = \frac{1}{n} \sum_{k=1}^{n} \mathbf{Z}_i \mathbf{Z}_i^T \,, \tag{2.4}$$

where for any $i = 1, \ldots, n$, $\mathbf{Z}_i = (Z_1^{(i)}, \ldots, Z_N^{(i)})^T$ and $\mathcal{Z}_n$ is the matrix with $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ as columns.

We state now our main result.

Theorem 2.1. *Let* $\mathbf{B}_n$ *and* $\mathbf{G}_n$ *be the matrices defined in* (2.1) *and* (2.4) *respectively. Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *then for any* $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} |S_{\mathbf{B}_n}(z) - \mathbb{E}(S_{\mathbf{G}_n}(z))| = 0 \quad a.s.$$

This approximation allows us to reduce the study of empirical spectral measure of $\mathbf{B}_n$ to the expectation of that of a Gaussian sample covariance matrix with the same covariance structure without requiring any rate of convergence to zero of the correlation between the entries.

Theorem 2.2. *Let* $\mathbf{B}_n$ *be the matrix defined in* (2.1) *and associated with* $(X_k)_{k\in\mathbb{Z}}$. *Let* $\gamma_k := \mathbb{E}(X_0 X_k)$ *and assume that*

$$\sum_{k\geq 0} |\gamma_k| < \infty \, . \tag{2.5}$$

Provided that $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *then , almost surely,* $\mu_{\mathbf{B}_n}$ *converges weakly to a probability measure* μ *whose Stieltjes transform* $S = S(z)$ $(z \in \mathbb{C}^+)$ *satisfies the equation*

$$z = -\frac{1}{\underline{S}} + \frac{c}{2\pi}\int_0^{2\pi} \frac{1}{\underline{S} + \big(2\pi f(\lambda)\big)^{-1}} d\lambda \, , \tag{2.6}$$

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

The proof of this Theorem will be postponed to Section 2.4.

Remark 2.3. *The spectral density function* f *of* $(X_k)_{k\in\mathbb{Z}}$ *is the discrete Fourier transform of the autocovariance function. Under the condition* (2.5), *the spectral density* f *of* $(X_k)_k$ *exists, is continuous and bounded on* $[0, 2\pi)$. *Moreover, Yao proves in Proposition 1 of [80] that the limiting spectral distribution is in this case compactly supported.*

Remark 2.4. *Our Theorem 2.2 is stated for the case of short memory dependent sequences. We note that it remains valid for long memory dependent sequences* adapted to the natural filtration. *We refer to Section 3.5 in [45] for the proof of the latter case.*

2.2 Applications

In this section, we consider some examples of stationary processes that can be represented as a function of i.i.d. random variables and give sufficient conditions under which (2.5) is satisfied.

2.2.1 Linear processes

We shall start with an example of linear filters. Let $(\varepsilon_k)_k$ be a sequence of i.i.d. centered random variables and let

$$X_k = \sum_{\ell \in \mathbb{Z}} a_\ell \varepsilon_{k-\ell} \,, \tag{2.7}$$

with $(a_\ell)_\ell$ being a linear filter or simply a sequence of real numbers. The so-called non-causal linear process $(X_k)_{k\in\mathbb{Z}}$ is widely used in a variety of applied fields. It is properly defined for any square summable sequence $(a_\ell)_{\ell\in\mathbb{Z}}$ if and only if the stationary sequence of innovations $(\varepsilon_k)_k$ has a bounded spectral density. In general, the covariances of $(X_k)_k$ might not be summable so that the linear process might exhibit long range dependence. Obviously, X_k can be written under the form (2.2) and Theorem 2.1 follows even for the case of long range dependence.

We note now that for this linear process,

$$\gamma_k = \|\varepsilon_0\|_2^2 \sum_{\ell\in\mathbb{Z}} a_\ell a_{k-\ell} \,,$$

and thus we infer that (2.5) is satisfied if

$$\|\varepsilon_0\|_2 < \infty \quad \text{and} \quad \sum_{\ell\in\mathbb{Z}} |a_\ell| < \infty \,.$$

Our Theorem 2.2 then holds improving Theorem 2.5 of Bai and Zhou [6] and Theorem 1 of Yao [80], that require ε_0 to be in $\mathbb{L}^4$ and gives the result of Theorem 1 of Pan et al. [55].

2.2.2 Functions of linear processes

We focus now on functions of real-valued linear processes. Define

$$X_k = h\Big(\sum_{\ell\in\mathbb{Z}} a_\ell \varepsilon_{k-\ell}\Big) - \mathbb{E}\Big(h\Big(\sum_{\ell\in\mathbb{Z}} a_\ell \varepsilon_{k-\ell}\Big)\Big)\,, \tag{2.8}$$

where $(a_\ell)_{\ell\in\mathbb{Z}}$ 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, under which (2.5) is satisfied.

With aim we shall, we introduce the projection operator: for any k and j belonging to $\mathbb{Z}$, let

$$P_j(X_k) = \mathbb{E}(X_k|\mathcal{F}_j) - \mathbb{E}(X_k|\mathcal{F}_{j-1})\,,$$

where $\mathcal{F}_j = \sigma(\varepsilon_i\,, i \le j)$ with the convention that $\mathcal{F}_{-\infty} = \bigcap_{k\in\mathbb{Z}} \mathcal{F}_k$ and $\mathcal{F}_{+\infty} = \bigvee_{k\in\mathbb{Z}} \mathcal{F}_k$. As we shall see in the sequel, the quantity $\|P_0(X_k)\|_2$ can be computed in many situations including non-linear models and for such cases we shall rather consider the following condition:

$$\sum_{k\in\mathbb{Z}} \|P_0(X_k)\|_2 < \infty\,. \tag{2.9}$$

We note that this condition is known in the literature as the Hannan-Heyde condition and is well adapted to the study of time series. Moreover, it implies the absolute summability of the covariance; condition (2.5). To see this fact, we start by noting that since $\mathcal{F}_{-\infty} = \bigcap_{k\in\mathbb{Z}} \mathcal{F}_k$ is trivial then 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:

$$X_k = \sum_{j\in\mathbb{Z}} P_j(X_k)\,.$$

Since $\mathbb{E}(P_i(X_0)P_j(X_k)) = 0$ if $i \neq j$, then, by stationarity, we get for any integer $k \ge 0$,

$$|\mathbb{E}(X_0X_k)| = \Big|\sum_{j\in\mathbb{Z}} \mathbb{E}\Big(P_j(X_0)P_j(X_k)\Big)\Big| \le \sum_{j\in\mathbb{Z}} \|P_0(X_j)\|_2 \|P_0(X_{k+j})\|_2\,.$$

Taking the sum over $k \in \mathbb{Z}$, we get that

$$\sum_{k\in\mathbb{Z}} |\mathbb{E}(X_0 X_k)| \le \Big(\sum_{k\in\mathbb{Z}} \|P_0(X_k)\|_2\Big)^2 .$$

The Hannan-Heyde condition is known to be sufficient for the validity of the central limit theorem for the partial sums (normalized by $\sqrt{n}$) associated with an adapted regular stationary process in $\mathbb{L}^2$. It is also essentially optimal for the absolute summability of the covariances. Indeed, for a causal linear process with non-negative coefficients and generated by a sequence of i.i.d. centered random variables in $\mathbb{L}^2$, both conditions (2.5) and (2.9) are equivalent to the summability of the coefficients.

Remark 2.5. *Let us mention that, by Remark 3.3 of [23] , the following conditions are together sufficient for the validity of* (2.9)*:*

$$\sum_{k\ge1} \frac{1}{\sqrt{k}} \|\mathbb{E}(X_k|\mathcal{F}_0)\|_2 < \infty \quad \textit{and} \quad \sum_{k\ge1} \frac{1}{\sqrt{k}} \|X_{-k} - \mathbb{E}(X_{-k}|\mathcal{F}_0)\|_2 < \infty . \tag{2.10}$$

We specify two different classes of models for which our Theorem 2.2 applies and we give sufficient conditions for (2.5) to be satisfied. Other classes of models, including non-linear time series such as iterative Lipschitz models, can be found in Wu [79].

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

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

Corollary 2.6. *Assume that*

$$\sum_{k\in\mathbb{Z}} \|w_h(|a_k\varepsilon_0|)\|_2 < \infty , \tag{2.11}$$

or that

$$\sum_{k\ge1} \frac{1}{k^{1/2}} \Big\|w_h\Big(\sum_{\ell\ge k} |a_\ell||\varepsilon_{k-\ell}|\Big)\Big\|_2 < \infty \quad \textit{and} \quad \sum_{k\ge1} \frac{1}{k^{1/2}} \Big\|w_h\Big(\sum_{\ell\le -k} |a_\ell||\varepsilon_{-k-\ell}|\Big)\Big\|_2 < \infty . \tag{2.12}$$

Then, provided that $c(n) = N/n \to c \in (0,\infty)$, the conclusion of Theorem 2.2 holds *for $\mu_{\mathbf{B}_n}$ where $\mathbf{B}_n$ is the sample covariance matrix defined in (2.1) and associated with $(X_k)_{k\in\mathbf{Z}}$ defined in (2.8).*

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

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

then condition (2.11) is satisfied and the conclusion of Corollary 2.6 holds.

Example 2. Assume $\|\varepsilon_0\|_\infty \le M$ where M is a finite positive constant, and that $|a_k| \le C\rho^{|k|}$ where $\rho \in (0,1)$ and C is a finite positive constant, then the condition (2.12) is satisfied and the conclusion of Corollary 2.6 holds as soon as

$$\sum_{k\ge 1} \frac{1}{\sqrt{k}} w_h\Big(MC\frac{\rho^k}{1-\rho}\Big) < \infty\,.$$

Using the usual comparison between series and integrals, it follows that the latter condition is equivalent to

$$\int_0^1 \frac{w_h(t)}{t\sqrt{|\log t|}}dt < \infty\,. \tag{2.13}$$

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

Proof of Corollary 2.6. To prove the corollary, it suffices to show that the condition (2.5) is satisfied as soon as (2.11) or (2.12) 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 $k \in \mathbb{Z}$,

$$\begin{aligned}\|P_0(X_k)\|_2 &= \Big\|\mathbb{E}_\varepsilon\Big(h\Big(\sum_{\ell\le k-1} a_\ell\varepsilon^*_{k-\ell} + \sum_{\ell\ge k} a_\ell\varepsilon_{k-\ell}\Big) - h\Big(\sum_{\ell\le k} a_\ell\varepsilon^*_{k-\ell} + \sum_{\ell\ge k+1} a_\ell\varepsilon_{k-\ell}\Big)\Big)\Big\|_2 .\\ &\le \Big\|w_h(|a_k(\varepsilon_0 - \varepsilon_0^*)|)\Big\|_2\end{aligned}$$

Next, by the subadditivity of $w_h(\cdot)$,

$$w_h(|a_k(\varepsilon_0 - \varepsilon_0^*)|) \le w_h(|a_k\varepsilon_0|) + w_h(|a_k\varepsilon_0^*|)\,.$$

Whence, $\|P_0(X_k)\|_2 \leq 2\|w_h(|a_k\varepsilon_0|)\|_2$. This proves that the condition (2.9) is satisfied under (2.11).

We prove now that if (2.12) holds then so does condition (2.9). According to Remark 2.5, it suffices to prove that the conditions in (2.10) are satisfied. With the same notations as before, we have that, for any $k \geq 1$,

$$\mathbb{E}(X_k|\mathcal{F}_0) = \mathbb{E}_\varepsilon\Big(h\Big(\sum_{\ell\leq k-1} a_\ell\varepsilon^*_{k-\ell} + \sum_{\ell\geq k} a_\ell\varepsilon_{k-\ell}\Big) - h\Big(\sum_{\ell\in\mathbb{Z}} a_\ell\varepsilon^*_{k-\ell}\Big)\Big).$$

Hence, for any non-negative integer k,

$$\|\mathbb{E}(X_k|\mathcal{F}_0)\|_2 \leq \Big\|w_h\Big(\sum_{\ell\geq k} |a_\ell(\varepsilon_{k-\ell} - \varepsilon^*_{k-\ell})|\Big)\Big\|_2 \leq 2\Big\|w_h\Big(\sum_{\ell\geq k} |a_\ell||\varepsilon_{k-\ell}|\Big)\Big\|_2,$$

where we have used the subadditivity of $w_h(\cdot)$ for the last inequality. Similarly, we prove that

$$\|X_{-k} - \mathbb{E}(X_{-k}|\mathcal{F}_0)\|_2 \leq 2\Big\|w_h\Big(\sum_{\ell\leq -k} |a_\ell||\varepsilon_{-k-\ell}|\Big)\Big\|_2.$$

These inequalities entail that the conditions in (2.10) hold as soon as those in (2.12) do. This ends the proof of Corollary 2.6.

$\square$

2.3 The proof of the universality result

We give in this section the main steps for proving Theorem 2.1 and we explain how the dependence structure in each column is broken allowing us to approximate the matrix with a Gaussian one via the Lindeberg method by blocks.

Since the columns of $\mathcal{X}_n$ are independent, it follows by Guntuboyina and Leeb's concentration inequality [32] of the spectral measure or by Step 1 of the proof of Theorem 1.1 in Bai and Zhou [6] that for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|S_{\mathbf{B}_n}(z) - \mathbb{E}(S_{\mathbf{B}_n}(z))\Big| = 0 \quad \text{a.s.}$$

Hence, our aim is to prove that for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left|\mathbb{E}(S_{\mathbf{B}_n}(z)) - \mathbb{E}(S_{\mathbf{G}_n}(z))\right| = 0 \quad \text{a.s.}$$

The first step is to "break" the dependence structure of the coordinates of each column $\mathbf{X}_i$ of $\mathcal{X}_n$. With this aim, we introduce a parameter m and construct mutually independent columns $\bar{\mathbf{X}}_{1,m}, \dots, \bar{\mathbf{X}}_{n,m}$ whose entries consist of $2m$-dependent random variables bounded by a positive number M and separated by blocks of zero entries of dimension $3m$. The aim of replacing certain entries by zeros is to create gaps of length $3m$ between some entries and benefit from their $2m$-dependent structure. $\bar{\mathbf{X}}_{1,m}, \dots, \bar{\mathbf{X}}_{n,m}$ will then consist of relatively big blocks of non-zero entries separated by small blocks of zero entries of dimension $3m$. We note that the non-zero blocks of entries are mutually independent since the gap between any two of them is at least $2m$. We shall then approximate $\mathbf{B}_n$ with the sample covariance matrix

$$\bar{\mathbf{B}}_{n,m} := \frac{1}{n}\sum_{i=1}^{n} \bar{\mathbf{X}}_{i,m}\bar{\mathbf{X}}_{i,m}^T.$$

This approximation will be done in such a way that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}\Big(S_{\mathbf{B}_n}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big)\right| = 0\,. \tag{2.14}$$

We shall then construct a Gaussian sample covariance matrix

$$\bar{\mathbf{G}}_{n,m} := \frac{1}{n}\sum_{i=1}^{n} \bar{\mathbf{Z}}_{i,m}\bar{\mathbf{Z}}_{i,m}^T$$

having the same block structure as $\bar{\mathbf{B}}_{n,m}$ and associated with a sequence of *Gaussian* random variables having the same covariance structure as the $2m$-dependent sequence constructed at the first step. $\bar{\mathbf{B}}_{n,m}$ is then approximated with $\bar{\mathbf{G}}_{n,m}$ via the Lindeberg method which consists of replacing at each time a non-zero block by its corresponding Gaussian block that has eventually the same covariance structure. This method allows us to prove for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left| \mathbb{E}(S_{\bar{\mathbf{B}}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) \right| = 0 \,. \tag{2.15}$$

In view of (2.14) and (2.15), Theorem 2.1 will then follow if we can prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left| \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n}}(z)) \right| = 0 \,. \tag{2.16}$$

This will be achieved via a well-adapted Gaussian interpolation technique as described in Section 2.3.2

2.3.1 Breaking the dependence structure of the matrix

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

$$k_{N,m} = \left[\frac{N}{m^2+3m} \right] , \tag{2.17}$$

where we recall that $[\,\cdot\,]$ denotes the integer part. We shall partition $\{1, \dots, N\}$ by writing it as the 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 $\ell \in \{1, \dots, k_{N,m}\}$,

$$\begin{aligned} I_\ell &:= \Big[(\ell-1)(m^2+3m)+1 \,,\; (\ell-1)(m^2+3m)+m^2 \Big] \cap \mathbb{N}, \\ J_\ell &:= \Big[(\ell-1)(m^2+3m)+m^2+1 \,,\; \ell(m^2+3m) \Big] \cap \mathbb{N} \,, \end{aligned} \tag{2.18}$$

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

$$I_{k_{N,m}+1} = \Big[k_{N,m}(m^2+3m)+1 \,,\; N \Big] \cap \mathbb{N} \,,$$

and $J_{k_{N,m}+1} = \emptyset$. Note that $I_{k_{N,m}+1} = \emptyset$ if $k_{N,m}(m^2+3m) = N$.

Truncation and construction of a $2m$-dependent sequence

Let M be a fixed positive number that depends neither on N, n nor 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 $i \in \{1, \ldots, n\}$, let

$$\widetilde{X}^{(i)}_{k,m} := \widetilde{X}^{(i)}_{k,M,m} = \mathbb{E}\Big(\varphi_M(X^{(i)}_k)|\varepsilon^{(i)}_{k-m}, \ldots, \varepsilon^{(i)}_{k+m}\Big)$$

and set

$$\bar{X}^{(i)}_{k,m} = \widetilde{X}^{(i)}_{k,m} - \mathbb{E}\Big(\widetilde{X}^{(i)}_{k,m}\Big)\,. \tag{2.19}$$

Notice that $(\bar{X}^{(1)}_{k,m})_k, \ldots, (\bar{X}^{(n)}_{k,m})_k$ are n independent copies of the centered and stationary sequence $(\bar{X}_{k,m})_k$ defined for any $k \in \mathbb{Z}$ by

$$\bar{X}_{k,m} = \widetilde{X}_{k,m} - \mathbb{E}\Big(\widetilde{X}_{k,m}\Big) \quad \text{where} \quad \widetilde{X}_{k,m} = \mathbb{E}(\varphi_M(X_k)|\varepsilon_{k-m}, \ldots, \varepsilon_{k+m}). \tag{2.20}$$

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

$$\|\bar{X}^{(i)}_{k,m}\|_\infty = \|\bar{X}_{k,m}\|_\infty \le 2M\,. \tag{2.21}$$

Moreover, for any $i \in \{1, \ldots, n\}$, we note that $(\bar{X}^{(i)}_{k,m})_{k\in\mathbb{Z}}$ forms a $2m$-dependent sequence in the sense that $\bar{X}^{(i)}_{k,m}$ and $\bar{X}^{(i)}_{k',m}$ are independent if $|k-k'| > 2m$.

Construction of columns consisting of independent blocks

For any $i = 1, \ldots, n$, we let $\mathbf{u}_{i,\ell}$ be the row random vectors defined for any $\ell = 1, \ldots, k_{N,m} - 1$ by

$$\mathbf{u}_{i,\ell} = \Big((\bar{X}^{(i)}_{k,m})_{k\in I_\ell}, \mathbf{0}_{3m}\Big) \tag{2.22}$$

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

$$\mathbf{u}_{i,k_{N,m}} = \Big((\bar{X}^{(i)}_{k,m})_{k\in I_{k_{N,m}}}, \mathbf{0}_r\Big) \tag{2.23}$$

where $r = 3m + N - k_{N,m}(m^2+3m)$. We note that the vectors in (2.22) are of dimension $m^2 + 3m$ whereas that in (2.23) is of dimension $N - (k_{N,m} - 1)(m^2 + 3m)$. We also note that for any $i = 1, \ldots, n$ and $\ell = 1, \ldots, k_{N,m}$, the random vectors $\mathbf{u}_{i,\ell}$ are mutually independent.

For any $i \in \{1, \ldots, n\}$, we define now the random vectors $\bar{\mathbf{X}}_{i,m}$ of dimension N by setting

$$\bar{\mathbf{X}}_{i,m} = \Big(\mathbf{u}_{i,\ell}\,,\ \ell = 1,\dots,k_{N,m}\Big)^T \tag{2.24}$$

and we let $\bar{\mathcal{X}}_{n,m}$ be the matrix whose columns consist of the $\bar{\mathbf{X}}_{i,m}$'s. Finally, we define the associated sample covariance matrix

$$\bar{\mathbf{B}}_{n,m} := \frac{1}{n}\bar{\mathcal{X}}_n\bar{\mathcal{X}}_n^T = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{X}}_{i,m}\bar{\mathbf{X}}_{i,m}^T\,. \tag{2.25}$$

Approximation with the associated sample covariance matrix

In what follows, we shall prove the following proposition.

Proposition 2.7. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$*, then for any* $z \in \mathbb{C}^+$*,*

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

Proof. By Proposition A.1 and Cauchy-Swharz's inequality, it follows that

$$\begin{aligned}&\Big|\mathbb{E}\Big(S_{\mathbf{B}_n}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}_{\mathbf{n},\mathbf{m}}}(z)\Big)\Big| \\ &\qquad \le \frac{\sqrt{2}}{v^2}\Big\|\frac{1}{N}\mathrm{Tr}(\mathbf{B}_n + \bar{\mathbf{B}}_{n,m})\Big\|_1^{1/2}\Big\|\frac{1}{Nn}\mathrm{Tr}(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})^T\Big\|_1^{1/2}. \end{aligned}\tag{2.26}$$

By the definition of $\mathbf{B}_n$ and the fact that for each i, $(X_k^{(i)})_{k\in\mathbb{Z}}$ is an independent copy of the stationary sequence $(X_k)_{k\in\mathbb{Z}}$, we infer that

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

Now, setting

$$\mathcal{I}_{N,m} = \bigcup_{\ell=1}^{k_{N,m}} I_\ell \quad\text{and}\quad \mathcal{R}_{N,m} = \{1,\dots,N\}\backslash\mathcal{I}_{N,m}\,, \tag{2.27}$$

then by using the stationarity of the sequence $(\bar{X}_{k,m}^{(i)})_{k\in\mathbb{Z}}$ and the fact that $\mathrm{Card}(\mathcal{I}_{N,m}) =$

$m^2 k_{N,m} \le N$, we get

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

Noting that,

$$\|\bar{X}_{0,m}\|_2 \le 2\|\widetilde{X}_{0,m}\|_2 \le 2\|\varphi_M(X_0)\|_2 \le 2\|X_0\|_2\,, \tag{2.28}$$

we infer that $N^{-1}\mathbb{E}|\mathrm{Tr}(\bar{\mathbf{B}}_{n,m})| \le 4\|X_0\|_2^2$. Now, by the definition of $\mathcal{X}_n$ and $\bar{\mathcal{X}}_{n,m}$,

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

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

$$\mathrm{Card}(\mathcal{R}_{N,m}) = N - m^2 k_{N,m} \le \frac{3N}{m+3} + m^2\,, \tag{2.29}$$

we get that

$$\begin{aligned}\frac{1}{Nn}&\mathbb{E}|\mathrm{Tr}(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})^T| \\ &\le \|X_0 - \bar{X}_{0,m}\|_2^2 + (3(m+3)^{-1} + m^2 N^{-1})\|X_0\|_2^2\,.\end{aligned}$$

Starting from (2.26), considering the above upper bounds, we derive that there exists a positive constant C not depending on (m, M) such that

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

Therefore, Proposition 2.7 will follow if we prove that

$$\lim_{m\to\infty}\limsup_{M\to\infty}\|X_0 - \bar{X}_{0,m}\|_2 = 0\,. \tag{2.30}$$

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

$k \in \mathbb{Z}$,

$$X_{k,m} = \mathbb{E}\Big(X_k | \varepsilon_{k-m}, \ldots, \varepsilon_{k+m}\Big) .$$

With the above notation, we write that

$$\|X_0 - \bar{X}_{0,m}\|_2 \leq \|X_0 - X_{0,m}\|_2 + \|X_{0,m} - \bar{X}_{0,m}\|_2 . \tag{2.31}$$

As X_0 is centered then so is $X_{0,m}$ and thus recalling the definition (2.20) of $\bar{X}_{0,m}$, we write

$$\|X_{0,m} - \bar{X}_{0,m}\|_2 = \|X_{0,m} - \mathbb{E}(X_{0,m}) - \bar{X}_{0,m}\|_2 = \|X_{0,m} - \widetilde{X}_{0,m} - \mathbb{E}(X_{0,m} - \widetilde{X}_{0,m})\|_2 .$$

Therefore, it follows that

$$\|X_{0,m} - \bar{X}_{0,m}\|_2 \leq 2\|X_{0,m} - \widetilde{X}_{0,m}\|_2 \leq 2\|X_0 - \varphi_M(X_0)\|_2 = 2\|(|X_0| - M)_+\|_2 . \tag{2.32}$$

Since X_0 belongs to $\mathbb{L}^2$, $\lim_{M\to\infty} \|(|X_0| - M)_+\|_2 = 0$. Therefore, to prove (2.30) and thus Proposition 2.7, it suffices to prove that

$$\lim_{m\to\infty} \|X_0 - X_{0,m}\|_2 = 0 . \tag{2.33}$$

Since $(X_{0,m})_{m\geq 0}$ is a martingale with respect to the increasing filtration $(\mathcal{G}_m)_{m\geq 0}$ defined by $\mathcal{G}_m = \sigma(\varepsilon_{-m}, \ldots, \varepsilon_m)$, and is such that

$$\sup_{m\geq 0} \|X_{0,m}\|_2 \leq \|X_0\|_2 < \infty ,$$

(2.33) follows by the martingale convergence theorem in $\mathbb{L}^2$ (see for instance Corollary 2.2 in Hall and Heyde [34]). This ends the proof of Proposition 2.7.

$\square$

2.3.2 Approximation with Gaussian sample covariance matrices via Lindeberg's method

After having broken the dependence structure of the initial matrix $\mathbf{B}_n$ and reduced the study to the matrix $\bar{\mathbf{B}}_{n,m}$ consisting of independent blocks, it is time to prove the universality of limiting distribution by the approximation with a sample covariance matrix having the same block structure as $\bar{\mathbf{B}}_{n,m}$ and associated with a Gaussian process having the same covariance structure as $(\bar{X}_{k,m})_{k\in\mathbb{Z}}$.

Construction of the Gaussian matrices

We shall first consider a sequence $(Z_{k,m})_{k\in\mathbb{Z}}$ of centered Gaussian random variables such that for any $k,\ell\in\mathbb{Z}$,

$$\operatorname{Cov}(Z_{k,m}\,,Z_{\ell,m})=\operatorname{Cov}(\bar{X}_{k,m}\,,\bar{X}_{\ell,m})\,. \tag{2.34}$$

For $i=1,\dots,n$, we then let $(Z^{(i)}_{k,m})_k$ be an independent copy of $(Z_{k,m})_k$ and we define the $N\times n$ matrix $\mathcal{Z}_{n,m}=((\mathcal{Z}_{n,m})_{k,i})=(Z^{(i)}_{k,m})$. We finally define the associated sample covariance matrix

$$\mathbf{G}_{n,m}=\frac{1}{n}\mathcal{Z}_{n,m}\mathcal{Z}^T_{n,m}\,. \tag{2.35}$$

We define now for any $\ell=1,\dots,k_{N,m}$ the random vectors $\mathbf{v}_{i,\ell}$ in the same way as the $\mathbf{u}_{i,\ell}$'s in (2.22) and (2.23) but by replacing each $\bar{X}^{(i)}_{k,m}$ by $Z^{(i)}_{k,m}$. For any $i\in\{1,\dots,n\}$, we then define the random vectors $\bar{\mathbf{Z}}_{i,m}$ of dimension N, as follows:

$$\bar{\mathbf{Z}}_{i,m}=(\mathbf{v}_{i,\ell}\,,\,\ell=1,\dots,k_{N,m})^T\,. \tag{2.36}$$

Let now

$$\bar{\mathcal{Z}}_{n,m}=\Big(\bar{\mathbf{Z}}_{1,m}|\dots|\bar{\mathbf{Z}}_{n,m}\Big)\quad\text{and}\quad\bar{\mathbf{G}}_{n,m}=\frac{1}{n}\bar{\mathcal{Z}}_{n,m}\bar{\mathcal{Z}}^T_{n,m}\,.$$

We note now that the matrix $\bar{\mathcal{Z}}_{n,m}$ is constructed from $\mathcal{Z}_{n,m}$ by replacing some of its entries by zeros so that it has the same block structure as $\bar{\mathcal{X}}_{n,m}$. This common block and covariance structure between $\bar{\mathcal{X}}_{n,m}$ and $\bar{\mathcal{Z}}_{n,m}$ will allow us to prove, via the Lindeberg method, the following convergence: for any $z\in\mathbb{C}^+$,

$$\lim_{n\to+\infty}\Big|\mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big)-\mathbb{E}\Big(S_{\bar{\mathbf{G}}_{n,m}}(z)\Big)\Big|=0\,.$$

Having proved the above convergence, the study is reduced to approximating the Gaussian matrices. As $\mathcal{Z}_{n,m}$ and $\bar{\mathcal{Z}}_{n,m}$ have the same entries up to a relatively small number of zero-rows, the Rank theorem (Theorem A.44 in [5]) allows us to prove the following convergence: for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left| \mathbb{E}\Big(S_{\bar{\mathbf{G}}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big) \right| = 0.$$

In order to complete the proof of Theorem 2.1, it remains to prove for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left| \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big) \right| = 0\,.$$

We note that these two matrices are associated with Gaussian processes having different covariance structures. The technique used for proving the above convergence is based on the Gaussian interpolation technique followed by a suitable control of the covariance.

2.3.2.1 Lindeberg method by blocks

In this section, we shall approximate $\bar{\mathbf{B}}_{n,m}$ with the Gaussian matrix $\bar{\mathbf{G}}_{n,m}$.

Proposition 2.8. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *then for any* $z \in \mathbb{C}^+$*,*

$$\lim_{n\to+\infty} \left| \mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{G}}_{n,m}}(z)\Big) \right| = 0\,.$$

The Stieltjes transform can be expressed as a function f of the matrix entries admitting partial derivatives of all orders. The above convergence shall be proved via the Lindeberg method by blocks which consists of writing the above difference as a telescoping sum and replacing at each time a non-zero block of entries by its corresponding Gaussian one.

In order to develop this, we first give the following definition of the function $f := f_z$ allowing us to write, for any $z \in \mathbb{C}_+$, the Stieltjes transform of a sample covariance matrix in terms of its entries.

Definition 2.9. *Let* $\mathbf{x}$ *be a vector of* $\mathbb{R}^{nN}$ *with coordinates*

$$\mathbf{x} = \Big(\mathbf{x}_1^T, \dots, \mathbf{x}_n^T\Big) \text{ where for any } i \in \{1,\dots,n\},\ \mathbf{x}_i = \Big(x_k^{(i)}\,,\, k \in \{1,\dots,N\}\Big)^T.$$

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

$$f(\mathbf{x}) = \frac{1}{N}\mathrm{Tr}\Big(A(\mathbf{x}) - z\mathbf{I}\Big)^{-1} \quad \textit{where} \quad A(\mathbf{x}) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k\mathbf{x}_k^T\,, \tag{2.37}$$

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 $A_u = A(\mathbf{x})$ the matrix-valued function of the scalar u. Then, setting $G_u = (A_u - z\mathbf{I})^{-1}$ and differentiating both sides of the equality $G_u(A_u - z\mathbf{I}) = \mathbf{I}$, it follows that

$$\frac{dG}{du} = -G\frac{dA}{du}G \tag{2.38}$$

(see the equality (17) in Chatterjee [20]). Higher-order derivatives may be computed by applying repeatedly the above formula. Upper bounds for the partial derivatives of f up to the third order are given in Appendix A.1.

Proof. Using Definition 2.9 and the notations (2.24) and (2.36), we get that, for any $z \in \mathbb{C}^+$,

$$\mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{G}}_{n,m}}(z)\Big) = \mathbb{E}f\Big(\bar{\mathbf{X}}_{1,m}^T, \dots, \bar{\mathbf{X}}_{n,m}^T\Big) - \mathbb{E}f\Big(\bar{\mathbf{Z}}_{1,m}^T, \dots, \bar{\mathbf{Z}}_{n,m}^T\Big)\,. \tag{2.39}$$

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

$$\mathbf{U}_{i,\ell} = \Big(\mathbf{0}_{(i-1)N}\,,\, \mathbf{0}_{(\ell-1)(m^2+3m)}\,,\, \mathbf{u}_{i,\ell}\,,\, \mathbf{0}_{r_\ell}, \mathbf{0}_{(n-i)N}\Big)\,, \tag{2.40}$$

where the $\mathbf{u}_{i,\ell}$'s are defined in (2.22) and (2.23), and

$$r_\ell = N - \ell(m^2 + 3m) \ \text{ for } \ell \in \{1, \dots, k_{N,m} - 1\}, \text{ and } \ r_{k_{N,m}} = 0\,. \tag{2.41}$$

Note that the vectors $(\mathbf{U}_{i,\ell})_{1 \le i \le n, 1 \le \ell \le k_{N,m}}$ are mutually independent. Moreover, with the notations (2.24) and (2.40), the following relations hold: for any $i \in \{1, \dots, n\}$,

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

and

$$\sum_{i=1}^{n} \sum_{\ell=1}^{k_{N,m}} \mathbf{U}_{i,\ell} = \Big(\bar{\mathbf{X}}_{1,m}^T \, , \dots , \, \bar{\mathbf{X}}_{n,m}^T \Big) \, ,$$

where the $\bar{\mathbf{X}}_i$'s are defined in (2.24). Now, for any $i \in \{1, \dots, n\}$, we define the random vectors $(\mathbf{V}_{i,\ell})_{\ell \in \{1,\dots,k_{N,m}\}}$ of dimension nN, as follows: for any $\ell \in \{1, \dots, k_{N,m}\}$,

$$\mathbf{V}_{i,\ell} = \Big(\mathbf{0}_{(i-1)N} \, , \, \mathbf{0}_{(\ell-1)(m^2+3m)} \, , \, \mathbf{v}_{i,\ell} \, , \, \mathbf{0}_{r_\ell} , \mathbf{0}_{(n-i)N} \Big) \, ,$$

where r_ℓ is precised in (2.41). With the above notations, the following relations hold: for any $i \in \{1, \dots, n\}$,

$$\sum_{\ell=1}^{k_{N,m}} \mathbf{V}_{i,\ell} = \Big(\mathbf{0}_{N(i-1)} \, , \, \bar{\mathbf{Z}}_{i,m}^T \, , \, \mathbf{0}_{N(n-i)} \Big)$$

and

$$\sum_{i=1}^{n} \sum_{\ell=1}^{k_{N,m}} \mathbf{V}_{i,\ell} = \Big(\bar{\mathbf{Z}}_{1,m}^T \, , \dots , \, \bar{\mathbf{Z}}_{n,m}^T \Big) \, .$$

We define now, for any $i \in \{1, \dots, n\}$,

$$\mathbf{S}_i = \sum_{s=1}^{i} \sum_{\ell=1}^{k_{N,m}} \mathbf{U}_{s,\ell} \text{ and } \mathbf{T}_i = \sum_{s=i}^{n} \sum_{\ell=1}^{k_{N,m}} \mathbf{V}_{s,\ell} \, , \tag{2.42}$$

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

$$\mathbf{S}_{i,s} = \sum_{\ell=1}^{s} \mathbf{U}_{i,\ell} \text{ and } \mathbf{T}_{i,s} = \sum_{\ell=s}^{k_{N,m}} \mathbf{V}_{i,\ell} \, . \tag{2.43}$$

In all the notations above, we use the convention that $\sum_{k=r}^{s} = 0$ if $r > s$. Therefore, starting from (2.39), considering the above relations and notations, we get

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

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

$$
\mathbf{W}_{i,s} = \mathbf{S}_{i-1} + \mathbf{S}_{i,s} + \mathbf{T}_{i,s+1} + \mathbf{T}_{i+1} , \tag{2.44}
$$

and

$$
\widetilde{\mathbf{W}}_{i,s} = \mathbf{S}_{i-1} + \mathbf{S}_{i,s-1} + \mathbf{T}_{i,s+1} + \mathbf{T}_{i+1} , \tag{2.45}
$$

we are lead to

$$
\mathbb{E}(S_{\bar{\mathbf{B}}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) = \sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \Big(\mathbb{E}(\Delta_{i,s}(f)) - \mathbb{E}(\widetilde{\Delta}_{i,s}(f)) \Big) , \tag{2.46}
$$

where

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

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

Definition 2.10. *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 $A^{\otimes k} = A \bigotimes \left(A^{\otimes (k-1)}\right)^T$.

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 2.11. *Let d be a positive integer. If ∇ denotes the differentiation operator given by $\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d}\right)$ 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 2.10. 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 $z = u + iv \in \mathbb{C}^+$ and let us control the right-hand side of (2.46). We have by Taylor's integral formula

$$\begin{aligned}\Delta_{i,s}(f) \;=\; & Df\big(\widetilde{\mathbf{W}}_{i,s}\big).\mathbf{U}_{i,s}^{\otimes 1} \;+\; \frac{1}{2}D^2 f\big(\widetilde{\mathbf{W}}_{i,s}\big).\mathbf{U}_{i,s}^{\otimes 2} \\ & \qquad + \int_0^1 \frac{(1-t)^2}{2} D^3 f\big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\big).\mathbf{U}_{i,s}^{\otimes 3} dt\end{aligned}$$

and

$$\begin{aligned}\widetilde{\Delta}_{i,s}(f) \;=\; & Df\big(\widetilde{\mathbf{W}}_{i,s}\big).\mathbf{V}_{i,s}^{\otimes 1} \;+\; \frac{1}{2}D^2 f\big(\widetilde{\mathbf{W}}_{i,s}\big).\mathbf{V}_{i,s}^{\otimes 2} \\ & \qquad + \int_0^1 \frac{(1-t)^2}{2} D^3 f\big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{V}_{i,s}\big).\mathbf{V}_{i,s}^{\otimes 3} dt\,,\end{aligned}$$

Since for any $i \in \{1, \ldots, n\}$ and any $s \in \{1, \ldots, k_{N,m}\}$, $\mathbf{U}_{i,s}$ and $\mathbf{V}_{i,s}$ are two centered random vectors independent of $\widetilde{\mathbf{W}}_{i,s}$, it follows that

$$\mathbb{E}\Big(Df(\widetilde{\mathbf{W}}_{i,s}).\mathbf{U}_{i,s}^{\otimes 1}\Big) = \mathbb{E}\Big(Df(\widetilde{\mathbf{W}}_{i,s})\Big).\,\mathbb{E}(\mathbf{U}_{i,s}^{\otimes 1}) = 0$$

and

$$\mathbb{E}\Big(Df(\widetilde{\mathbf{W}}_{i,s}).\mathbf{V}_{i,s}^{\otimes 1}\Big) = \mathbb{E}\Big(Df(\widetilde{\mathbf{W}}_{i,s})\Big).\,\mathbb{E}(\mathbf{V}_{i,s}^{\otimes 1}) = 0\,.$$

Since in addition, $\mathbb{E}(\mathbf{U}_{i,s}^{\otimes 2}) = \mathbb{E}(\mathbf{V}_{i,s}^{\otimes 2})$, it also follows that

$$\begin{aligned}\mathbb{E}\Big(D^2 f(\widetilde{\mathbf{W}}_{i,s}) .\, \mathbf{U}_{i,s}^{\otimes 2}\Big) &= \mathbb{E}\Big(D^2 f(\widetilde{\mathbf{W}}_{i,s})\Big) .\, \mathbb{E}\Big(\mathbf{U}_{i,s}^{\otimes 2}\Big) \\ &= \mathbb{E}\Big(D^2 f(\widetilde{\mathbf{W}}_{i,s})\Big) .\, \mathbb{E}\Big(\mathbf{V}_{i,s}^{\otimes 2}\Big) = \mathbb{E}\Big(D^2 f(\widetilde{\mathbf{W}}_{i,s}) .\, \mathbf{V}_s^{(i)\,\otimes 2}\Big) .\end{aligned}$$

Therefore, considering the above Taylor expansions and taking the expectation, we get

$$\begin{aligned}\mathbb{E}(\Delta_{i,s}(f)) - \mathbb{E}(\widetilde{\Delta}_{i,s}(f)) = \mathbb{E}\int_0^1 \frac{(1-t)^2}{2} D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big) . \mathbf{U}_{i,s}^{\otimes 3}\, dt \\ + \mathbb{E}\int_0^1 \frac{(1-t)^2}{2} D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{V}_{i,s}\Big) . \mathbf{V}_{i,s}^{\otimes 3}\, dt . \end{aligned} \tag{2.47}$$

Let us analyze the first term of the right-hand side of (2.47). Recalling the definition (2.40) of the $\mathbf{U}_{i,s}$'s, we get for any $t \in [0,1]$,

$$\begin{aligned}&\mathbb{E}\Big|D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big) . \mathbf{U}_{i,s}^{\otimes 3}\Big| \\ &\quad \le \sum_{k\in I_s}\sum_{\ell\in I_s}\sum_{j\in I_s} \mathbb{E}\Big| \frac{\partial^3 f}{\partial x_k^{(i)} \partial x_\ell^{(i)} \partial x_j^{(i)}}\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big) \bar{X}_{k,m}^{(i)} \bar{X}_{\ell,m}^{(i)} \bar{X}_{j,m}^{(i)} \Big| \\ &\qquad \le \sum_{k\in I_s}\sum_{\ell\in I_s}\sum_{j\in I_s} \Big\| \frac{\partial^3 f}{\partial x_k^{(i)} \partial x_\ell^{(i)} \partial x_j^{(i)}}\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big)\Big\|_2 \|\bar{X}_{k,m}^{(i)} \bar{X}_{\ell,m}^{(i)} \bar{X}_{j,m}^{(i)}\|_2 ,\end{aligned}$$

where I_s is defined in (2.18). Therefore, using (2.21), the stationarity and (2.28), it follows that, for any $t \in [0,1]$,

$$\begin{aligned}&\mathbb{E}\Big|D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big) . \mathbf{U}_{i,s}^{\otimes 3}\Big| \\ &\qquad \le 8M^2 \|X_0\|_2 \sum_{k\in I_s}\sum_{\ell\in I_s}\sum_{j\in I_s} \Big\| \frac{\partial^3 f}{\partial x_k^{(i)} \partial x_\ell^{(i)} \partial x_j^{(i)}}\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big)\Big\|_2 .\end{aligned}$$

Notice that by (2.42) and (2.43),

$$\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s} = \Big(\bar{\mathbf{X}}_{1,m}^T, \dots, \bar{\mathbf{X}}_{i-1,m}^T, w_i(t), \bar{\mathbf{Z}}_{i+1,m}^T, \dots, \bar{\mathbf{Z}}_{n,m}^T\Big) ,$$

where $w_i(t)$ is the row vector of dimension N defined by

$$\begin{aligned} w_i(t) &= \mathbf{S}_{i,s-1} + t\mathbf{U}_{i,s} + \mathbf{T}_{i,s+1} \\ &= (\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,s-1}, t\mathbf{u}_{i,s}, \mathbf{v}_{i,s+1}, \dots, \mathbf{v}_{i,k_{N,m}}) . \end{aligned}$$

Therefore, by Lemma A.2 of the Appendix, (2.21) and the fact that $(Z^{(i)}_{k,m})_{k\in\mathbb{Z}}$ is distributed as the stationary sequence $(Z_{k,m})_{k\in\mathbb{Z}}$, we infer that there exists a positive constant C not depending on (n, M, m) and such that, for any $t \in [0, 1]$,

$$\Big\| \frac{\partial^3 f}{\partial x^{(i)}_k \partial x^{(i)}_\ell \partial x^{(i)}_j} \big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\big) \Big\|_2 \leq C \Big(\frac{M + \|Z_{0,m}\|_2}{v^3 N^{1/2} n^2} + \frac{N^{1/2}(M^3 + \|Z_{0,m}\|_6^3)}{v^4 n^3} \Big) .$$

Now, since $Z_{0,m}$ is a Gaussian random variable, $\|Z_{0,m}\|_6^6 = 15\|Z_{0,m}\|_2^6$. Moreover, by (2.3), $\|Z_{0,m}\|_2 = \|\bar{X}_{0,m}\|_2 \leq 2\|X_0\|_2$. Therefore, there exists a positive constant C not depending on (n, M, m) and such that, for any $t \in [0, 1]$,

$$\mathbb{E}\Big| D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{U}_{i,s}\Big).\mathbf{U}^{\otimes 3}_{i,s}\Big| \leq \frac{Cm^6(1 + M^3)}{v^3(1 \wedge v)N^{1/2}n^2} .$$

We similarly analyze the "Gaussian part" in the right-hand side of (2.47) and we get

$$\mathbb{E}\Big| D^3 f\Big(\widetilde{\mathbf{W}}_{i,s} + t\mathbf{V}_{i,s}\Big).\mathbf{V}^{\otimes 3}_{i,s}\Big| \leq \frac{Cm^6(1 + M^3)}{v^3(1 \wedge v)N^{1/2}n^2} .$$

By the above bounds and the fact that $m^2 k_{N,m} \leq N$, we derive that there exists a positive constant C not depending on on (n, M, m) such that

$$\sum_{i=1}^{n} \sum_{s=1}^{k_{N,m}} \Big| \mathbb{E}\Big(\Delta_{i,s}(f)\Big) - \mathbb{E}\Big(\tilde{\Delta}_{i,s}(f)\Big)\Big| \leq C \frac{(1 + M^5)N^{1/2}m^4}{v^3(1 \wedge v)n} .$$

Starting from (2.46) and considering the fact that $\lim_{n\to\infty} N/n = c \in (0, \infty)$, we get for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} \Big| \mathbb{E}(S_{\bar{\mathbf{B}}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z))\Big| = 0 .$$

$\square$

2.3.2.2 Approximation of the Gaussian matrices

As mentioned before, we shall now approximate the intermediate Gaussian matrices by proving the following convergence:

Proposition 2.12. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *then for any* $z \in \mathbb{C}^+$*,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left| \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{G}}_{n,m}}(z)\Big) \right| = 0.$$

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

$$\begin{aligned} \left| \mathbb{E}(S_{\mathbf{G}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) \right| &\leq \mathbb{E}\left| \int \frac{1}{x-z} dF^{\mathbf{G}_{n,m}}(x) - \int \frac{1}{x-z} dF^{\bar{\mathbf{G}}_{n,m}}(x) \right| \\ &\leq \mathbb{E}\left| \int \frac{F^{\mathbf{G}_{n,m}}(x) - F^{\bar{\mathbf{G}}_{n,m}}}{(x-z)^2} dx \right| \\ &\leq \frac{\pi}{v} \, \| F^{\mathbf{G}_{n,m}} - F^{\bar{\mathbf{G}}_{n,m}} \|_\infty \,. \end{aligned}$$

Hence, by Theorem A.44 by Bai and Silverstein [5],

$$\left| \mathbb{E}(S_{\mathbf{G}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) \right| \leq \frac{\pi}{vN} \operatorname{Rank}\Big(\mathcal{Z}_{n,m} - \bar{\mathcal{Z}}_{n,m} \Big) \,.$$

By definition of $\mathcal{Z}_{n,m}$ and $\bar{\mathcal{Z}}_{n,m}$, $\operatorname{Rank}\Big(\mathcal{Z}_{n,m} - \bar{\mathcal{Z}}_{n,m}\Big) \leq \operatorname{Card}(\mathcal{R}_{N,m})$, where $\mathcal{R}_{N,m}$ is defined in (2.27). Therefore, using (2.29), we get that, for any $z = u + iv \in \mathbb{C}^+$,

$$\left| \mathbb{E}(S_{\mathbf{G}_{n,m}}(z)) - \mathbb{E}(S_{\bar{\mathbf{G}}_{n,m}}(z)) \right| \leq \frac{\pi}{vN} \left(\frac{3N}{m+3} + m^2 \right),$$

which converges to zero by letting n tend to infinity and then m. This ends the proof of Proposition 2.12.

□

In order to end the proof, it remains to approximate the Gaussian sample covariance matrices $\mathbf{G}_n$ and $\mathbf{G}_{n,m}$. We note that these matrices have different covariance structure however they have the same block structure and do not contain zero entries.

We end the proof by proving the following convergence:

Proposition 2.13. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *then for any* $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big)\right| = 0.$$

Proof. To prove this theorem, we shall rather consider the following symmetric matrices: let $n' = N + n$ and let $\mathbb{G}_{n'}$ and $\mathbb{G}_{n',m}$ be the symmetric matrices of order n' defined by

$$\mathbb{G}_{n'} = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0}_{n,n} & \mathcal{Z}_n^T \\ \mathcal{Z}_n & \mathbf{0}_{N,N} \end{pmatrix} \quad \text{and} \quad \mathbb{G}_{n',m} = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0}_{n,n} & \mathcal{Z}_{n,m}^T \\ \mathcal{Z}_{n,m} & \mathbf{0}_{N,N} \end{pmatrix}.$$

Now as the non-zero eigenvalues of $\mathbb{G}_{n'}$ are plus and minus the singular values of $\mathcal{Z}_n$, it easily follows that for any $z \in \mathbb{C}_+$,

$$z S_{\mathbb{G}^2_{n'}}(z^2) = S_{\mathbb{G}_{n'}}(z).$$

We note that the eigenvalues of $\mathbb{G}^2_{n'}$ are the eigenvalues of $n^{-1}\mathcal{Z}_n^T\mathcal{Z}_n$ together with those of $n^{-1}\mathcal{Z}_n\mathcal{Z}_n^T$. Since these two latter matrices have the same non-zero eigenvalues, the following relation holds: for any $z \in \mathbb{C}^+$

$$S_{\mathbf{G}_n}(z) = z^{-1/2}\frac{n}{2N} S_{\mathbb{G}_{n'}}(z^{1/2}) + \frac{n-N}{2Nz}$$

(see, for instance, page 549 in Rashidi Far *et al* [62]). Similarly, the same relation also holds for the matrices $\mathbf{G}_{n,m}$ and $\mathbb{G}_{n',m}$ and

$$S_{\mathbf{G}_{n,m}}(z) = z^{-1/2}\frac{n}{2N} S_{\mathbb{G}_{n',m}}(z^{1/2}) + \frac{n-N}{2Nz}.$$

Since $n'/N \to 1 + c^{-1}$, it is then equivalent to prove for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \left|\mathbb{E}\Big(S_{\mathbb{G}_{n'}}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_{n',m}}(z)\Big)\right| = 0\,.$$

Noting that $(\mathbb{G}_{n'})_{k,\ell} = n^{-1/2} Z^{(\ell)}_{k-n} \mathbf{1}_{k>n}\mathbf{1}_{\ell\leq n}$ and $(\mathbb{G}_{n',m})_{k,\ell} = n^{-1/2} Z^{(\ell)}_{k-n,m} \mathbf{1}_{k>n}\mathbf{1}_{\ell\leq n}$ if $1 \leq \ell \leq k \leq n'$ and keeping in mind the independence structure between the columns, we apply Lemma A.5 of the Appendix and we get for any $z \in \mathbb{C}^+$,

$$
\begin{aligned}
&\mathbb{E}(S_{\mathbb{G}_{n'}}(z)) - \mathbb{E}(S_{\mathbb{G}_{n',m}}(z)) \\
&= \frac{n'}{2n}\sum_{j=1}^{n}\sum_{k,\ell=n+1}^{n+N}\int_0^1 \Big(\mathbb{E}(Z_{k-n}Z_{\ell-n}) - \mathbb{E}(Z_{k-n,m}Z_{\ell-n,m})\Big)\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k,j}\partial x_{\ell,j}}(\mathbf{g}(t))\Big)dt \\
&= \frac{N+n}{2n}\sum_{j=1}^{n}\sum_{k,\ell=1}^{N}\int_0^1 \Big(\mathbb{E}(Z_kZ_\ell) - \mathbb{E}(Z_{k,m}Z_{\ell,m})\Big)\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)dt \qquad (2.48)
\end{aligned}
$$

where, for $t \in [0,1]$,

$$
\mathbf{g}(t) = \sqrt{N+n}\Big(\sqrt{t}(\mathbb{G}_{n'})_{k,\ell} + \sqrt{1-t}(\mathbb{G}_{n',m})_{k,\ell}\Big)_{1\le \ell\le k\le n'}
$$

and f is the function that allows us to write the Stieltjes transform of a symmetric matrix in terms of its entries. For a precise definition of f for the case of symmetric matrices, see (A.17) in the Appendix.

Then, by using (2.3) and (2.34), we write the following decomposition

$$
\begin{aligned}
\mathbb{E}(Z_kZ_\ell) - \mathbb{E}(Z_{k,m}Z_{\ell,m}) &= \mathbb{E}(X_kX_\ell) - \mathbb{E}(\bar{X}_{k,m}\bar{X}_{\ell,m}) \\
&= \mathbb{E}\Big(X_k(X_\ell - \bar{X}_{\ell,m})\Big) + \mathbb{E}\Big(\bar{X}_{\ell,m}(X_k - \bar{X}_{k,m})\Big) \qquad (2.49)
\end{aligned}
$$

We shall decompose the right-hand side of (2.48) into two sums according to the decomposition (2.49) and treat them separately. Let us prove that there exists a constant C not depending on n and t such that

$$
\Big|\sum_{j=1}^{n}\sum_{k,\ell=1}^{N}\mathbb{E}(X_k(X_\ell - \bar{X}_{\ell,m}))\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)\Big| \le C\frac{N}{N+n}\|X_0\|_2\|X_0 - X_{0,m}\|_2 \,.
$$

To do this, we note first that without loss of generality $\mathbf{g}(t)$ can be taken independent of $(X_{k,\ell})$ and then

$$
\mathbb{E}(X_k(X_\ell-\bar{X}_{\ell,m}))\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big) = \mathbb{E}\Big(X_k(X_\ell-\bar{X}_{\ell,m})\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)\,.
$$

On the other hand, Lemma A.4 by Merlevède and Peligrad [45] allow us to control the second order partial derivative of f in the following manner:

$$\sum_{j=1}^{n}\Big|\sum_{k,\ell=1}^{N} a_k b_\ell \frac{\partial^2}{\partial x_{k+n,j}\partial x_{\ell+n,j}} f(\mathbf{g}(t))\Big| \leq \frac{C}{N+n}\Big(\sum_{k=1}^{N} a_k^2 \sum_{\ell=1}^{N} b_\ell^2\Big)^{1/2},$$

where $(a_k)_k$ and $(b_k)_k$ are two sequences of real numbers and C is a universal constant depending only on the imaginary part of z. Applying the above inequality with $a_k = X_k$ and $b_\ell = X_\ell - \bar{X}_{\ell,m}$, we get for any $z \in \mathbb{C}^+$,

$$\Big|\sum_{j=1}^{n}\sum_{k,\ell=1}^{N} X_k(X_\ell - \bar{X}_{\ell,m})\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)\Big| \\ \leq \frac{C}{N+n}\Big(\sum_{k=1}^{n} X_k^2 \sum_{\ell=1}^{N}(X_\ell - \bar{X}_{\ell,m})^2\Big)^{1/2}$$

Thus by Cauchy-Schwarz's inequality and then by the stationarity of the variables, we infer for any $z \in \mathbb{C}_+$ and any $t \in [0,1]$ that

$$\Big|\sum_{j=1}^{n}\sum_{k,\ell=1}^{N} \mathbb{E}\Big(X_k(X_\ell - \bar{X}_{\ell,m})\Big)\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)\Big| \\ \leq \frac{C}{N+n}\Big(\sum_{k=1}^{N}\mathbb{E}(X_k^2)\sum_{\ell=1}^{N}\mathbb{E}(X_\ell - \bar{X}_{\ell,m})^2\Big)^{1/2} \\ = C\frac{N}{N+n}\|X_0\|_2\,\|X_0 - \bar{X}_{0,m}\|_2\,.$$

We similarly prove that

$$\Big|\sum_{j=1}^{n}\sum_{k,\ell=1}^{N} \mathbb{E}\Big(\bar{X}_{\ell,m}(X_k - \bar{X}_{k,m})\Big)\mathbb{E}\Big(\frac{\partial^2 f}{\partial x_{k+n,j}\partial x_{\ell+n,j}}(\mathbf{g}(t))\Big)\Big| \leq C\frac{N}{N+n}\|X_0\|_2\|X_0 - \bar{X}_{0,m}\|_2\,.$$

Therefore, starting from (2.48), considering the decomposition (2.49) and the above upper bounds, we infer that

$$\Big|\mathbb{E}(S_{\mathbb{G}_{n'}}(z)) - \mathbb{E}(S_{\mathbb{G}_{n',m}}(z))\Big| \leq C\frac{N}{n}\|X_0\|_2\,\|X_0 - \bar{X}_{0,m}\|_2\,.$$

Recalling that $N/n \to c \in (0, \infty)$ and taking the limit on n we get that

$$\limsup_{n\to\infty} \left| \mathbb{E}(S_{\mathbb{G}_{n'}}(z)) - \mathbb{E}(S_{\mathbb{G}_{n',m}}(z)) \right| \le C \|X_0\|_2 \, \|X_0 - \bar{X}_{0,m}\|_2 \, .$$

We end the proof by noting that by relations (2.31), (2.32) and (2.33) we have

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

$\square$

2.4 The limiting Spectral distribution

In this section, we give a proof of Theorem 2.2 which according to Theorem 2.1 follows if we prove that for any $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} \mathbb{E} S_{\mathbf{G}_n}(z) = S(z) \, . \tag{2.50}$$

with S satisfying Equation (2.6). This can be achieved by using Theorem 1.4 by Silverstein combined with arguments developed in the proof of Theorem 1 of Yao [80] (see also [73]).

With this aim, we consider a sequence $(y_k)_{k\in\mathbb{Z}}$ of i.i.d. real valued random variables with law $\mathcal{N}(0,1)$, and then consider 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 $i \in \{1, \dots, n\}$, we define the row random vector

$$\mathbf{y}_i = \left(y_1^{(i)}, \dots, y_N^{(i)} \right) ,$$

and let $\mathcal{Y}_n$ be the $N \times n$ matrix whose columns are the $\mathbf{y}_i^T$'s. Finally, we consider its associated sample covariance matrix

$$\mathbf{Y}_n = \frac{1}{n} \mathcal{Y}_n \mathcal{Y}_n^T = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^T \, .$$

Set

$$\Gamma_N := \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & & \gamma_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \cdots & \gamma_0 \end{pmatrix},$$

where we recall that $\gamma_k = \mathrm{Cov}(X_0, X_k)$ and that, by (2.3), we have for any $i \in \{1, \ldots, n\}$

$$\gamma_k = \mathrm{Cov}(Z_0, Z_k) = \mathrm{Cov}(Z_0^{(i)}, Z_k^{(i)}) .$$

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 is finite.

Note also that the vector $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ has the same distribution as

$$\Big(\mathbf{y}_1 \Gamma_N^{1/2}, \ldots, \mathbf{y}_n \Gamma_N^{1/2}\Big)$$

where $\Gamma_N^{1/2}$ is the symmetric non-negative square root of Γ_N. Therefore, for any $z \in \mathbb{C}^+$,

$$\mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big) = \mathbb{E}\Big(S_{\mathbf{A}_n}(z)\Big) ,$$

where $\mathbf{A}_n = \Gamma_N^{1/2} \mathbf{Y}_n \Gamma_N^{1/2}$. The proof of (2.50) is then reduced to prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \mathbb{E}\Big(S_{\mathbf{A}_n}(z)\Big) = S(z) . \tag{2.51}$$

According to Theorem 1.4 by Silverstein [64], if one can show that

$$F^{\Gamma_N} \text{ converges to a probability distribution } H, \tag{2.52}$$

then (2.51) holds with $S = S(z)$ satisfying the equation:

$$S = \int \frac{1}{-z + \lambda(1 - c - czS)} \, \mathrm{d}H(\lambda) .$$

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

$$z = -\frac{1}{\underline{S}} + c \int \frac{\lambda}{1 + \lambda \underline{S}} \, \mathrm{d}H(\lambda) .$$

We note that $\underline{S} := \underline{S}(z)$ is a Stieltjes transform and that $\mathfrak{Im}(\underline{S}) > 0$. One can check Equations (1.1)-(1.4) in [64] for more details.

Due to the Toeplitz form of Γ_N and to the fact that $\sum_{k\geq 0} |\gamma_k| < \infty$ then the fundamental eigenvalue distribution theorem of Szegö for Toeplitz forms [31] allows to assert that (2.52) holds and that the empirical spectral distribution of Γ_N converges weakly to a non-random distribution H that is defined via the spectral density of $(X_k)_{k\in\mathbb{Z}}$. More precisely, for any continuous and bounded function φ, we have

$$\int \varphi(\lambda)\, \mathrm{d}H(\lambda) = \frac{1}{2\pi}\int_0^{2\pi} \varphi(2\pi f(\lambda))\mathrm{d}\lambda\,.$$

To end the proof, it suffices to notice that the function

$$\varphi(\lambda) := \frac{\lambda}{1+\lambda\underline{S}}$$

is continuous and bounded by $1/\mathfrak{Im}(\underline{S})$ and then combine the above relations to get (2.6).

$\square$

Chapter 3

Symmetric matrices with correlated entries

The limiting spectral distribution of *symmetric* matrices with correlated entries has received as well a lot of attention in the last two decades. The starting point consists of deep results for symmetric matrices with correlated Gaussian entries by Khorunzhy and Pastur [40], Boutet de Monvel and Khorunzhy [14], Chakrabarty *et al* [18] among others.

There is also a sustained effort for studying linear filters of independent random variables as entries of symmetric matrices. For instance, Anderson and Zeitouni [3] treat symmetric matrices with entries that are linear processes of finite range having independent innovations. They find the limiting spectral distribution assuming that distant above-diagonal entries are independent but nearby entries may be correlated.

In this chapter, we consider symmetric random matrices whose entries are functions of i.i.d. real-valued random variables. Our main goal is to reduce the study of the limiting spectral distribution to the same problem for a Gaussian matrix having the same covariance structure as the underlying process. In this way we prove the universality and we are able to formulate various limiting results for large classes of matrices. We also treat large sample covariance matrices with correlated entries, known under the name of Gram matrices.

This chapter is organized in the following way. Sections 3.1 and 3.2 contain respectively the main results for symmetric matrices and sample covariance matrices. As an intermediate step, we also treat in Section 3.4, matrices associated with K-dependent random fields, results that have interest in themselves. Applications to matrices whose entries are either linear random fields or nonlinear random fields as Volterra-type processes are given in Section 3.3. The main proofs are included in Sections 3.5 and 3.6, however Section 3.8 is devoted for the Lindeberg method by blocks. In Section 3.7 we prove a concentration inequality of the spectral measure for row-wise K-dependent random matrices.

3.1 Symmetric matrices with correlated entries

Let $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be an array of real-valued random variables, and consider its associated symmetric random matrix $\mathbf{X}_n$ of order n defined by

$$(\mathbf{X}_n)_{i,j} = \begin{cases} X_{i,j} & \text{if } 1 \leq j \leq i \leq n \\ X_{j,i} & \text{if } 1 \leq i < j \leq n \, . \end{cases} \tag{3.1}$$

Define then

$$\mathbb{X}_n := \frac{1}{\sqrt{n}}\mathbf{X}_n \, . \tag{3.2}$$

We shall study the limiting spectral distribution of the symmetric matrix $\mathbb{X}_n$ when the process $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ has the following dependence structure: for any $(k,\ell)\in\mathbb{Z}^2$,

$$X_{k,\ell} = g(\xi_{k-i,\ell-j} \ ; \ (i,j)\in\mathbb{Z}^2) \, , \tag{3.3}$$

where $(\xi_{i,j})_{(i,j)\in\mathbb{Z}^2}$ is an array of i.i.d. real-valued random variables given on a common probability space $(\Omega,\mathcal{K},\mathbb{P})$, and g is a measurable function from $\mathbb{R}^{\mathbb{Z}^2}$ to $\mathbb{R}$ such that

$$\mathbb{E}(X_{0,0}) = 0 \quad \text{and} \quad \|X_{0,0}\|_2 < \infty \, .$$

A representation as (3.3) includes linear as well as many widely used nonlinear random fields models as special cases. Moreover, the entries on and below the diagonal are dependent across both rows and columns.

We shall prove a universality scheme for the random matrix $\mathbb{X}_n$ as soon as the entries of the symmetric matrix $\sqrt{n}\mathbb{X}_n$ have the dependence structure (3.3). It is noteworthy to indicate that this result will not require rate of convergence to zero of the correlation between the entries.

With this aim, we shall first let $(G_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be a real-valued centered Gaussian random field, with covariance function given by

$$\mathbb{E}(G_{k,\ell}G_{i,j}) = \mathbb{E}(X_{k,\ell}X_{i,j}) \quad \text{for any } (k,\ell) \text{ and } (i,j) \text{ in } \mathbb{Z}^2 . \tag{3.4}$$

Let then $\mathbf{G}_n$ be the symmetric random matrix defined by

$$(\mathbf{G}_n)_{i,j} = \begin{cases} G_{i,j} & \text{if } 1 \leq j \leq i \leq n \\ G_{j,i} & \text{if } 1 \leq i < j \leq n \end{cases}$$

and define its normalized matrix

$$\mathbb{G}_n := \frac{1}{\sqrt{n}}\mathbf{G}_n. \tag{3.5}$$

Theorem 3.1. *Let $\mathbb{X}_n$ and $\mathbb{G}_n$ be the symmetric matrices defined in (3.2) and (3.5) respectively. Then, for any $z \in \mathbb{C}^+$,*

$$\lim_{n\to\infty} \Big| S_{\mathbb{X}_n}(z) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big)\Big| = 0 \quad \textit{almost surely.}$$

The importance of Theorem 3.1 is that it reduces the study of the limiting spectral distribution function of a symmetric matrix whose entries are functions of i.i.d. random variables to studying the same problem for a Gaussian matrix with the same covariance structure.

We recall now that the Lévy distance between two distribution functions F and G defined by

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

and note that a sequence of distribution functions $F_n(x)$ converges to a distribution function $F(x)$ at all continuity points x of F if and only if $L(F_n, F) \to 0$.

We give the following corollary that is a direct consequence of our Theorem 3.1 together with Theorem B.9 in Bai and Silverstein [5]. One can also check the arguments on page 38 in [5], based on Vitali's convergence theorem.

Corollary 3.2. *Assume that $\mathbb{X}_n$ and $\mathbb{G}_n$ are as in Theorem 3.1. Furthermore, assume there exists a distribution function F such that*

$$\mathbb{E}\Big(F^{\mathbb{G}_n}(t)\Big) \to F(t) \text{ for all continuity points } t \in \mathbb{R} \text{ of } F\,.$$

Then

$$\mathbb{P}(L(F^{\mathbb{X}_n(\omega)}, F) \to 0) = 1\,. \tag{3.6}$$

For instance, Corollary 3.2 above combined with the proof of Theorem 2 in Khorunzhy and Pastur [40] concerning the asymptotic spectral behavior of certain ensembles with correlated Gaussian entries (see also Theorem 17.2.1 in [56]), gives the following:

Theorem 3.3. *Let $\mathbb{X}_n$ be the symmetric matrix defined in (3.2). For any $(k,\ell) \in \mathbb{Z}^2$, let $\gamma_{k,\ell} = \mathbb{E}(X_{0,0}X_{k,\ell})$. Assume that*

$$\sum_{k,\ell\in\mathbb{Z}} |\gamma_{k,\ell}| < \infty\,, \tag{3.7}$$

and that the following holds: for any $(k,\ell) \in \mathbb{Z}^2$,

$$\gamma_{k,\ell} = \gamma_{\ell,k}\,, \tag{3.8}$$

Then (3.6) holds, where F is a non-random distribution function whose Stieltjes transform $S(z)$ is uniquely defined by the relations:

$$S(z) = \int_0^1 h(x,z)dx\,, \tag{3.9}$$

where $h(x,z)$ is a solution to the equation

$$h(x,z) = \Big(-z + \int_0^1 f(x,y)h(y,z)dy\Big)^{-1} \quad \text{with} \quad f(x,y) = \sum_{k,j\in\mathbb{Z}} \gamma_{k,j}\mathrm{e}^{-2\pi\mathrm{i}(kx+jy)}\,. \tag{3.10}$$

Equation (3.10) is uniquely solvable in the class $\mathcal{F}$ of functions $h(x,z)$ with domain $(x,z) \in [0,1] \otimes \mathbb{C}\backslash\mathbb{R}$, which are analytic with respect to z for each fixed x, continuous with respect to x for each fixed z and satisfying the conditions: $\lim_{v\to\infty} v \operatorname{Im} h(x, \mathrm{i}v) \leq 1$ and $\operatorname{Im}(z) \operatorname{Im} h(x,z) > 0$.

Remark 3.4. *If condition (3.8) of Theorem 3.3 is replaced by: $\gamma_{\ell,k} = V(\ell)V(k)$ where V is an even function, then its conclusion can be given in the following alternative way: the convergence (3.6) holds where F is a non-random distribution function whose Stieltjes transform $S(z)$ is given by the relation*

$$S(z) = \int_0^\infty \frac{d\upsilon(\lambda)}{-z - \lambda h(z)}$$

where $\upsilon(t) = \lambda\{x \in [0,1]; f(x) < t\}$, λ is the Lebesgue measure, for $x \in [0,1]$,

$$f(x) = \sum_{k\in\mathbb{Z}} V(k) \mathrm{e}^{2\pi \mathrm{i} kx}$$

and $h(z)$ is solution to the equation

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

This equation is uniquely solvable in the class of analytic functions in $\mathbb{C}\backslash\mathbb{R}$ satisfying the conditions: $\lim_{x\to\infty} x h(\mathrm{i}x) < \infty$ and $\operatorname{Im}\big(h(z)\big)\operatorname{Im}(z) > 0$ for $z \in \mathbb{C}\backslash\mathbb{R}$. (See Boutet de Monvel and Khorunzhy [14]).

3.2 Gram matrices with correlated entries

Adapting the proof of Theorem 3.1, we can also obtain a universality scheme for large sample covariance matrices associated with a process $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ having the representation (3.3). So, all along this section $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ is assumed to be a random field having the representation (3.3). To define the Gram matrices associated with this random field, we consider two positive integers N and p and let $\mathcal{X}_{N,p}$ be the $N \times p$ matrix defined by

$$\mathcal{X}_{N,p} = \big(X_{i,j}\big)_{1\leq i\leq N, 1\leq j\leq p}\,. \tag{3.11}$$

Define now the symmetric matrix $\mathbb{B}_N$ of order N by

$$\mathbb{B}_N = \frac{1}{p}\mathcal{X}_{N,p}\mathcal{X}_{N,p}^T := \frac{1}{p}\sum_{k=1}^{p}\mathbf{r}_k\mathbf{r}_k^T\,, \tag{3.12}$$

where $\mathbf{r}_k = (X_{1,k},\cdots,X_{N,k})^T$ is the k-th column of $\mathcal{X}_{N,p}$. We note again that the entries of $\mathcal{X}_{N,p}$ are dependent across both rows and columns. The matrix $\mathbb{B}_N$ shall be referred to as the sample covariance matrix or also the Gram matrix associated with the process $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$.

We shall approximate the Stieltjes transform of $\mathbb{B}_N$ by that of the Gram matrix

$$\mathbb{H}_N = \frac{1}{p}\mathcal{G}_{N,p}\mathcal{G}_{N,p}^T \tag{3.13}$$

associated with a real-valued centered Gaussian random field $(G_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$, with covariance function given by (3.4).

Theorem 3.5. *Let $\mathbb{B}_N$ be defined by (3.12) and $\mathbb{H}_N$ by (3.13) Then, provided that $N,p\to\infty$ such that $N/p\to c\in(0,\infty)$, for any $z\in\mathbb{C}^+$,*

$$\lim_{n\to\infty}\Big|S_{\mathbb{B}_N}(z)-\mathbb{E}\big(S_{\mathbb{H}_N}(z)\big)\Big| = 0 \quad \textit{almost surely.} \tag{3.14}$$

Therefore, if $N,p\to\infty$ such that $N/p\to c\in(0,\infty)$ and if there exists a distribution function F such that

$$\mathbb{E}\Big(F^{\mathbb{H}_N}(t)\Big)\to F(t) \textit{ for all continuity points } t\in\mathbb{R} \textit{ of } F$$

then

$$\mathbb{P}(L(F^{\mathbb{B}_N(\omega)},F)\to 0) = 1\,. \tag{3.15}$$

Proof. To prove this theorem, we shall rather consider the following symmetric matrix: let $n = N+p$ and consider the symmetric matrix $\mathbb{X}_n$ of order n defined by

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

Let $\mathbb{Y}_n$ be defined as $\mathbb{X}_n$ but with $\mathcal{G}_{N,p}$ replacing $\mathcal{X}_{N,p}$. By the same arguments given in Proposition 2.13, the following relations hold: for any $z \in \mathbb{C}^+$

$$S_{\mathbb{B}_N}(z) = z^{-1/2}\frac{n}{2N}S_{\mathbb{X}_n}(z^{1/2}) + \frac{p-N}{2Nz}$$

and

$$S_{\mathbb{H}_N}(z) = z^{-1/2}\frac{n}{2N}S_{\mathbb{Y}_n}(z^{1/2}) + \frac{p-N}{2Nz}$$

(see, for instance, page 549 in Rashidi Far *et al* [62] for arguments leading to the above relations). Thus, we infer that since $n/N \to 1 + c^{-1}$ then it is equivalent to prove that for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left| S_{\mathbb{X}_n}(z) - \mathbb{E}\Big(S_{\mathbb{Y}_n}(z)\Big) \right| = 0 \ \text{ almost surely.} \tag{3.16}$$

The proof of (3.16) shall be omitted because it is a simple adaptation of that of Theorem 3.1. Indeed, following the lines of the proof of Theorem 3.1, we can infer that its conclusion still holds even when the stationarity of entries of $\mathbb{X}_n$ and $\mathbb{G}_n$ is slightly relaxed as above.

□

Theorem 3.5 together with Theorem 2.1 in Boutet de Monvel *et al* [15] allow then to derive the limiting spectral distribution of large sample covariance matrices associated with a process $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ having the representation (3.3) and satisfying a short range dependence condition.

Theorem 3.6. *Let $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be a real-valued stationary random field given by (3.3). Assume that (3.7) holds. Then, provided that $N, p \to \infty$ such that $N/p \to c \in (0,\infty)$, $\mathbb{P}(L(F^{\mathbb{B}_N(\omega)}, F) \to 0) = 1$ where F is a non-random distribution function whose Stieltjes transform $S(z)$, $z \in \mathbb{C}^+$ is uniquely defined by the relations:*

$$S(z) = \int_0^1 h(x,z)dx\,,$$

where $h(x,z)$ is a solution to the equation

$$h(x,z) = \Big(-z + \int_0^1 \frac{f(x,s)}{1 + c\int_0^1 f(u,s)h(u,z)du} ds \Big)^{-1}, \tag{3.17}$$

with $f(x,y)$ given in (3.10)*.*

Equation (3.17) *is uniquely solvable in the class $\mathcal{F}$ of functions $h(x,z)$ as described after the statement of Theorem* (3.3)*.*

We refer to the paper by Boutet de Monvel *et al* [15] regarding discussions on the smoothness and boundedness of the limiting density of states. Note that condition (3.7) is required in the statement of Theorem 3.6 only because all the estimates in the proof of Theorem 2.1 in [15] require it. However using arguments as developed in the paper by Chakrabarty *et al* [18], it can be proved that if the process $(X_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ admits a spectral density then there exists a non-random distribution function F such that $\mathbb{P}(L(F^{\mathbb{B}_N(\omega)}, F) \to 0) = 1$ (if $N/p \to c \in (0,\infty)$). Unfortunately the arguments developed in [18] do not allow, in general, to exhibit the limiting equation (3.17) which gives a lot of information on the limiting spectral distribution.

Notice however that if we add the assumption that the lines (resp. the columns) of $\mathcal{X}_{N,p}$ are non correlated (corresponding to the semantically (resp. spatially) "patterns" studied in Section 3 of [15]), condition (3.7) is not needed to exhibit the limiting equation of the Stieltjes transform. Indeed, in this situation, the lines (resp. the columns) of $\mathcal{G}_{N,p}$ become then independent and the result of Merlevède and Peligrad [45] about the limiting spectral distribution of Gram random matrices associated to independent copies of a stationary process applies. Proving, however, Theorem 3.6 in its full generality and without requiring condition (3.7) to hold, remains an open question.

Proof. In view of the convergence (3.14), it suffices to show that when $N, p \to \infty$ such that $N/p \to c \in (0,\infty)$, then for any $z \in \mathbb{C}^+$, $\mathbb{E}\Big(S_{\mathbb{H}_N}(z)\Big)$ converges to

$$S(z) = \int_0^1 h(x,z)dx$$

where $h(x,z)$ is a solution to the equation (3.17). This follows by applying Theorem 2.1 in Boutet de Monvel *et al* [15] . Indeed setting $\tilde{\mathbb{H}}_N = \frac{p}{N}\mathbb{H}_N$, this theorem asserts that if (3.7) holds then, when $N, p \to \infty$ such that $N/p \to c \in (0,\infty)$, $\mathbb{E}\Big(S_{\tilde{\mathbb{H}}_N}(z)\Big)$ converges to

$$m(z) = \int_0^1 v(x,z)dx\,,$$

for any $z \in \mathbb{C}^+$, where $v(x,z)$ is a solution to the equation

$$v(x,z) = \Big(-z + c^{-1} \int_0^1 \frac{f(x,s)}{1 + \int_0^1 f(u,s)v(u,z)du} ds \Big)^{-1} .$$

This implies that $\mathbb{E}(S_{\mathbb{H}_N}(z))$ converges to $S(z)$ as defined in the theorem since the following relation holds: $S(z) = c^{-1} m(z/c)$.

□

3.3 Examples

All along this section, $(\xi_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ will designate a double indexed sequence of i.i.d. real-valued random variables defined on a common probability space, centered and in $\mathbb{L}^2$.

3.3.1 Linear processes

Let $(a_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be a double indexed sequence of numbers such that

$$\sum_{k,\ell\in\mathbb{Z}} |a_{k,\ell}| < \infty . \tag{3.18}$$

Let then $(X_{i,j})_{(i,j)\in\mathbb{Z}^2}$ be the linear random field in $\mathbb{L}^2$ defined by: for any $(i,j) \in \mathbb{Z}^2$,

$$X_{i,j} = \sum_{k,\ell\in\mathbb{Z}} a_{k,\ell}\xi_{k+i,\ell+j} . \tag{3.19}$$

Corollary 3.2 (resp. Theorem 3.5) then applies to the matrix $\mathbb{X}_n$ (resp. $\mathbb{B}_N$) associated with the linear random field $(X_{i,j})_{(i,j)\in\mathbb{Z}^2}$ given in (3.2).

For the case of short dependence, based on our Theorem 3.6, we can describe the limit of the empirical spectral distribution of the Gram matrix associated with a linear random field.

Corollary 3.7. *Assume that $X_{i,j}$ is defined by (3.19) and that condition (3.18) is satisfied. Let N and p be positive integers, such that $N, p \to \infty$, $N/p \to c \in (0,\infty)$ and* *let*

$$\mathcal{X}_{N,p} = \big(X_{i,j}\big)_{1\le i\le N, 1\le j\le p} \quad \textit{and} \quad \mathbb{B}_N = N^{-1}\mathcal{X}_{N,p}\mathcal{X}_{N,p}^T .$$

Then the convergence (3.15) holds for $F^{\mathbb{B}_N}$, where F is a non-random distribution function whose Stieltjes transform satisfies the relations given in Theorem 3.6 with

$$\gamma_{k,j} = \|\xi_{0,0}\|_2^2 \sum_{u,v\in\mathbb{Z}} a_{u,v}a_{u+k,v+j} .$$

Concerning the Wigner-type matrix $\mathbb{X}_n$, we obtain by Remark 3.4, the following corollary, describing the limit in a particular case.

Corollary 3.8. *Let $(a_n)_{n\in\mathbb{Z}}$ be a sequence of numbers such that*

$$\sum_{k\in\mathbb{Z}} |a_k| < \infty$$

and define

$$X_{i,j} = \sum_{k,\ell\in\mathbb{Z}} a_k a_\ell \xi_{k+i,\ell+j}$$

for any $(i,j)\in\mathbb{Z}^2$. Consider the symmetric matrix $\mathbb{X}_n$ associated with $(X_{i,j})_{(i,j)\in\mathbb{Z}^2}$ and defined by (3.2). Then (3.6) holds, where F is a non-random distribution function whose Stieltjes transform satisfies the relation given in Remark 3.4 with

$$f(x) = \|\xi_{0,0}\|_2 \sum_{k\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} a_j a_{j+k} \mathrm{e}^{2\pi \mathrm{i} kx} .$$

3.3.2 Volterra-type processes

Other classes of stationary random fields having the representation (3.3) are Volterra-type processes which play an important role in the nonlinear system theory. For any $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$, define a second-order Volterra expansion as follows:

$$X_{\mathbf{k}} = \sum_{\mathbf{u}\in\mathbb{Z}^2} a_{\mathbf{u}}\xi_{\mathbf{k}-\mathbf{u}} + \sum_{\mathbf{u},\mathbf{v}\in\mathbb{Z}^2} b_{\mathbf{u},\mathbf{v}}\xi_{\mathbf{k}-\mathbf{u}}\xi_{\mathbf{k}-\mathbf{v}} , \tag{3.20}$$

where $a_{\mathbf{u}}$ and $b_{\mathbf{u},\mathbf{v}}$ are real numbers satisfying

$$b_{\mathbf{u},\mathbf{v}} = 0 \text{ if } \mathbf{u} = \mathbf{v}\,, \sum_{\mathbf{u}\in\mathbb{Z}^2} a_{\mathbf{u}}^2 < \infty \quad \text{and} \quad \sum_{\mathbf{u},\mathbf{v}\in\mathbb{Z}^2} b_{\mathbf{u},\mathbf{v}}^2 < \infty\,. \tag{3.21}$$

Under the above conditions, the random field $X_{\mathbf{k}}$ exists, is centered and in $\mathbb{L}^2$, and Corollary 3.2 (resp. Theorem 3.5) applies to the matrix $\mathbb{X}_n$ (resp. $\mathbb{B}_N$) associated with the Volterra-type random field. Further generalization to arbitrary finite order Volterra expansion is straightforward.

If we reinforced condition (3.21), we derive the following result concerning the limit of the empirical spectral distribution of the Gram matrix associated with the Volterra-type process:

Corollary 3.9. *Assume that $(X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^2}$ is defined by (3.20) and that the following additional condition is assumed:*

$$\sum_{\mathbf{u}\in\mathbb{Z}^2} |a_{\mathbf{u}}| < \infty\ , \ \sum_{\mathbf{v}\in\mathbb{Z}^2}\Big(\sum_{\mathbf{u}\in\mathbb{Z}^2} b_{\mathbf{u},\mathbf{v}}^2\Big)^{1/2} < \infty \quad \textit{and} \quad \sum_{\mathbf{v}\in\mathbb{Z}^2}\Big(\sum_{\mathbf{u}\in\mathbb{Z}^2} b_{\mathbf{v},\mathbf{u}}^2\Big)^{1/2} < \infty\,. \tag{3.22}$$

Let N and p be positive integers, such that $N, p \to \infty$, $N/p \to c \in (0,\infty)$. Let

$$\mathcal{X}_{N,p} = \big(X_{i,j}\big)_{1\le i\le N, 1\le j\le p} \quad \textit{and} \quad \mathbb{B}_N = N^{-1}\mathcal{X}_{N,p}\mathcal{X}_{N,p}^T\,.$$

Then (3.15) holds for $F^{\mathbb{B}_N}$, where F is a non-random distribution function whose Stieltjes transform satisfies the relations given in Theorem 3.6 with

$$\gamma_{\mathbf{k}} = \|\xi_{0,0}\|_2^2 \sum_{\mathbf{u}\in\mathbb{Z}^2} a_{\mathbf{u}} a_{\mathbf{u}+\mathbf{k}} + \|\xi_{0,0}\|_2^4 \sum_{\mathbf{u},\mathbf{v}\in\mathbb{Z}^2} b_{\mathbf{u},\mathbf{v}}(b_{\mathbf{u}+\mathbf{k},\mathbf{v}+\mathbf{k}} + b_{\mathbf{v}+\mathbf{k},\mathbf{u}+\mathbf{k}}) \ \textit{ for any } \mathbf{k}\in\mathbb{Z}^2\,. \tag{3.23}$$

If we impose additional symmetric conditions to the coefficients $a_{\mathbf{u}}$ and $b_{\mathbf{u},\mathbf{v}}$ defining the Volterra random field (3.20), we can derive the limiting spectral distribution of its associated symmetric matrix $\mathbb{X}_n$ defined by (3.2). Indeed if for any $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $\mathbb{Z}^2$,

$$a_{\mathbf{u}} = a_{u_1} a_{u_2}\,,\ b_{\mathbf{u},\mathbf{v}} = b_{u_1,v_1} b_{u_2,v_2}\,, \tag{3.24}$$

where the a_i and $b_{i,j}$ are real numbers satisfying

$$b_{i,j} = 0 \text{ if } i = j\,,\ \sum_{i\in\mathbb{Z}} |a_i| < \infty \ \text{ and } \sum_{(i,j)\in\mathbb{Z}^2} |b_{i,j}| < \infty\,, \tag{3.25}$$

then $(\gamma_{k,\ell})$ satisfies (3.7) and (3.8) . Hence, an application of Theorem 3.3 leads to the following result.

Corollary 3.10. *Assume that $(X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^2}$ is defined by (3.20) and that conditions (3.24) and (3.25) are satisfied. Define the symmetric matrix $\mathbb{X}_n$ by (3.2). Then (3.6) holds, where F is a nonrandom distribution function whose Stieltjes transform is uniquely defined by the relations given in Theorem 3.3 with*

$$\gamma_{s,t} = A(s)A(t) + B_1(s)B_1(t) + B_2(s)B_2(t)$$

where

$$A(t) = \|\xi_{0,0}\|_2 \sum_{i\in\mathbb{Z}} a_i a_{i+t}\ , \quad B_1(t) = \|\xi_{0,0}\|_2^2 \sum_{(i,r)\in\mathbb{Z}^2} b_{i,r} b_{i+t,r+t}$$

and

$$B_2(t) = \|\xi_{0,0}\|_2^2 \sum_{(i,r)\in\mathbb{Z}^2} b_{i,r} b_{r+t,i+t}\,.$$

3.4 Symmetric matrices with K-dependent entries

The proof of the main result, Theorem 3.1, shall be based on an approximation of the underlying symmetric matrix by another symmetric matrix whose entries are $2m$-dependent. We shall first prove a universality scheme for symmetric matrices with K-dependent entries and we note that this result has an interest in itself.

So the interest of this section will be proving a universality scheme for the limiting spectral distribution of symmetric matrices $\mathbf{X}_n = [X_{k,\ell}]_{k,\ell=1}^n$ normalized by $\sqrt{n}$ when the entries are real-valued random variables defined on a common probability space and satisfy a K-dependence condition (see Assumption $\mathbf{A_3}$). As we shall see later, Theorem 3.11 below will be a key step to prove Theorem 3.1 .

Let us start by introducing some assumptions concerning the entries $(X_{k,\ell}, 1 \le \ell \le k \le n)$.

$\mathbf{A_1}$ For all positive integers n, $\mathbb{E}(X_{k,\ell}) = 0$ for all $1 \leq \ell \leq k \leq n$, and

$$\frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{k} \mathbb{E}(|X_{k,\ell}|^2) \leq C < \infty \, .$$

$\mathbf{A_2}$ For any $\tau > 0$,

$$L_n(\tau) := \frac{1}{n^2} \sum_{k=1}^{n} \sum_{\ell=1}^{k} \mathbb{E}(|X_{k,\ell}|^2 \mathbf{1}_{|X_{k,\ell}| > \tau\sqrt{n}}) \to_{n\to\infty} 0 \, .$$

$\mathbf{A_3}$ There exists a positive integer K such that for all positive integers n, the following holds: for all nonempty subsets

$$A, B \subset \{(k,\ell) \in \{1, \dots, n\}^2 \,|\, 1 \leq \ell \leq k \leq n\}$$

such that

$$\min_{(i,j)\in A} \min_{(k,\ell)\in B} \max(|i-k|, |j-\ell|) > K$$

the σ-fields

$$\sigma\Big(X_{i,j} \, , \, (i,j) \in A\Big) \quad \text{and} \quad \sigma\Big(X_{k,\ell} \, , \, (k,\ell) \in B\Big)$$

are independent.

Condition $\mathbf{A}_3$ states that variables with index sets which are at a distance larger than K are independent.

In Theorem 3.11 below, we then obtain a universality result for symmetric matrices whose entries are K-dependent and satisfy $\mathbf{A}_1$ and the traditional Lindeberg's condition $\mathbf{A}_2$. Note that $\mathbf{A}_2$ is known to be a necessary and sufficient condition for the empirical spectral distribution of $n^{-1/2}\mathbf{X}_n$ to converge almost surely to the semi-circle law when the entries $X_{i,j}$ are independent, centered and with common variance not depending on n (see Theorem 9.4.1 in Girko [28]).

Theorem 3.11. *Let* $\mathbf{X}_n = [X_{k,\ell}]_{k,\ell=1}^n$ *be a symmetric matrix of order* n *whose entries* $(X_{k,\ell}, 1 \leq \ell \leq k \leq n)$ *are real-valued random variables satisfying conditions* $\mathbf{A_1}$, $\mathbf{A_2}$ *and*

$\mathbf{A_3}$. *Let* $\mathbf{G}_n = [G_{i,j}]_{i,j=1}^n$ *be a symmetric matrix of order* n *whose entries* $(G_{k,\ell})_{1 \leq \ell \leq k \leq n}$ *are real-valued centered Gaussian random variables with covariance function given by*

$$\mathbb{E}(G_{k,\ell}G_{i,j}) = \mathbb{E}(X_{k,\ell}X_{i,j})\,. \tag{3.26}$$

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

$$\lim_{n\to\infty} \left| S_{\mathbb{X}_n}(z) - \mathbb{E}(S_{\mathbb{G}_n}(z)) \right| = 0 \quad \textit{almost surely,} \tag{3.27}$$

where $\mathbb{X}_n = n^{-1/2}\mathbf{X}_n$ *and* $\mathbb{G}_n = n^{-1/2}\mathbf{G}_n$.

The proof of this result, which is based on a blend of blocking procedure and Lindeberg's method, will be given in the last section of this chapter.

As we mentioned at the beginning of the section, this theorem will be an intermediate result allowing us to prove that, in the stationary and non triangular setting, the K-dependence condition can be relaxed and more general models for the entries can be considered.

However, the above theorem has also interest in itself. For instance, for the matrices with real entries, this theorem makes it possible to weaken the conditions of Theorems 2.5 and 2.6 in Anderson and Zeitouni [3]. More precisely, due to our Theorem 3.11, their assumption 2.2.1 (Ib) can be weakened from the boundedness of all moments to the boundedness of moments of order 2 only plus $\mathbf{A}_2$.

Furthermore their result can be strengthened by replacing the convergence in probability by the almost sure convergence. Indeed, our Theorem 3.11 shows that if their assumption 2.2.1 (Ib) is replaced by $\mathbf{A}_1$ plus $\mathbf{A}_2$, then to study the limiting spectral distribution we can actually assume without loss of generality that the entries come from a Gaussian random field with the same covariance structure as the initial entries. If the $X_{k,\ell}$ are Gaussian random variables then the boundedness of all moments means the boundedness of moments of order 2.

3.5 Proof of the universality result, Theorem 3.1

The proof of this theorem shall be divided into three major steps. With the aim of breaking the dependence structure of $\mathbb{X}_n$, we shall first approximate its Stieltjes transform by that of a symmetric matrix $\mathbb{X}_n^{(m)}$ whose entries $X_{k,\ell}^{(m)}$ form a $2m$-dependent random field with m being a sequence of integers tending to infinity after n. Then applying Theorem 3.11, we approximate $\mathbb{X}_n^{(m)}$ by a symmetric matrix $\mathbb{G}_n^{(m)}$ associated with a $2m$-dependent Gaussian filed having the same covariance structure as $(X_{k,\ell}^{(m)})_{k,\ell}$. Once this is done it remains to approximate the Gaussian matrices $\mathbb{G}_n$ and $\mathbb{G}_n^{(m)}$. This approximation is based on the Gaussian interpolation technique and a suitable control of the partial derivatives of the Stieltjes transform.

Approximation by a matrix having $2m$-dependent entries

For m a positive integer (fixed for the moment) and for any (u,v) in $\mathbb{Z}^2$ define

$$X_{u,v}^{(m)} = \mathbb{E}\Big(X_{u,v}|\mathcal{F}_{u,v}^{(m)}\Big)\,, \tag{3.28}$$

where $\mathcal{F}_{u,v}^{(m)} := \sigma(\xi_{i,j}\,;\, u-m \le i \le u+m,\ v-m \le j \le v+m)$.

Let $\mathbf{X}_n^{(m)}$ be the symmetric random matrix of order n associated with $(X_{u,v}^{(m)})_{(u,v)\in\mathbb{Z}^2}$ and defined by

$$(\mathbf{X}_n^{(m)})_{i,j} = \begin{cases} X_{i,j}^{(m)} & \text{if } 1 \le j \le i \le n \\ X_{j,i}^{(m)} & \text{if } 1 \le i < j \le n\ . \end{cases}$$

Let $\mathbb{X}_n^{(m)}$ be the normalized symmetric matrix given by

$$\mathbb{X}_n^{(m)} = n^{-1/2}\mathbf{X}_n^{(m)}\,. \tag{3.29}$$

We first show that, for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\left|S_{\mathbb{X}_n}(z) - S_{\mathbb{X}_n^{(m)}}(z)\right| = 0 \ \text{ a.s.} \tag{3.30}$$

We have by Lemma A.3 of the Appendix, proved by Götze *et al.* [30], that

$$\left|S_{\mathbb{X}_n}(z) - S_{\mathbb{X}_n^{(m)}}(z)\right|^2 \leq \frac{1}{n^2 v^4} \mathrm{Tr}\Big(\big(\mathbf{X}_n - \mathbf{X}_n^{(m)}\big)^2\Big)\,,$$

where $v = \mathrm{Im}(z)$. Hence

$$\left|S_{\mathbb{X}_n}(z) - S_{\mathbb{X}_n^{(m)}}(z)\right|^2 \leq \frac{2}{n^2 v^4} \sum_{1\leq \ell \leq k \leq n} \big(X_{k,\ell} - X_{k,\ell}^{(m)}\big)^2\,.$$

Since the shift is ergodic with respect to the measure generated by a sequence of i.i.d. random variables and the sets of summations are on regular sets, the ergodic theorem entails that

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\leq k,\ell \leq n} \big(X_{k,\ell} - X_{k,\ell}^{(m)}\big)^2 = \mathbb{E}\big(X_{0,0} - X_{0,0}^{(m)}\big)^2 \quad \text{a.s. and in } \mathbb{L}^1\,.$$

Therefore

$$\limsup_{n\to\infty} \left|S_{\mathbb{X}_n}(z) - S_{\mathbb{X}_n^{(m)}}(z)\right|^2 \leq 2v^{-4} \|X_{0,0} - X_{0,0}^{(m)}\|_2^2 \ \text{a.s.} \tag{3.31}$$

Now, by the martingale convergence theorem

$$\|X_{0,0} - X_{0,0}^{(m)}\|_2 \to 0 \text{ as } m \to \infty\,, \tag{3.32}$$

which combined with (3.31) proves (3.30).

Approximation by a matrix with $2m$-dependent Gaussian entries

Let now $(G_{k,\ell}^{(m)})_{(k,\ell)\in\mathbb{Z}^2}$ be a real-valued centered Gaussian random field, with covariance function given by

$$\mathbb{E}(G_{k,\ell}^{(m)} G_{i,j}^{(m)}) = \mathbb{E}(X_{k,\ell}^{(m)} X_{i,j}^{(m)}) \quad \text{for any } (k,\ell) \text{ and } (i,j) \text{ in } \mathbb{Z}^2\,. \tag{3.33}$$

Note that the process $(G_{k,\ell}^{(m)})_{(k,\ell)\in\mathbb{Z}^2}$ is then in particular $2m$-dependent. Let now $\mathbf{G}_n^{(m)}$

be the symmetric random matrix of order n defined by

$$(\mathbf{G}_n^{(m)})_{i,j} = \begin{cases} G_{i,j}^{(m)} & \text{if } 1 \le j \le i \le n \\ G_{j,i}^{(m)} & \text{if } 1 \le i < j \le n \, , \end{cases}$$

and set $\mathbb{G}_n^{(m)} = \mathbf{G}_n^{(m)}/\sqrt{n}$. We shall prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left| S_{\mathbb{X}_n^{(m)}}(z) - \mathbb{E}\Big(S_{\mathbb{G}_n^{(m)}}(z)\Big) \right| = 0, \text{ almost surely.} \tag{3.34}$$

With this aim, we shall apply Theorem 3.11 after showing in what follows that $(X_{k,\ell}^{(m)}, 1 \le \ell \le k \le n)$ satisfies its assumptions.

Note that the sigma-algebras $\mathcal{F}_{u,v}^{(m)} := \sigma(\xi_{i,j} \, ; \, u-m \le i \le u+m, \; v-m \le j \le v+m)$ and $\mathcal{F}_{k,\ell}^{(m)}$ are independent as soon as $|u-k| > 2m$ or $|v-\ell| > 2m$. From this consideration, we then infer that $(X_{k,\ell}^{(m)}, 1 \le \ell \le k \le n)$ satisfies the assumption $\mathbf{A}_3$ of Section 3.4 with $K = 2m$.

On another hand, since $X_{k,\ell}$ is a centered random variable then so is $X_{k,\ell}^{(m)}$. Moreover,

$$\|X_{k,\ell}^{(m)}\|_2 \le \|X_{k,\ell}\|_2 = \|X_{1,1}\|_2 \, .$$

Hence $(X_{k,\ell}^{(m)}, 1 \le \ell \le k \le n)$ satisfies the assumption $\mathbf{A}_1$ of Section 3.4.

We prove now that the assumption $\mathbf{A}_2$ of Section 3.4 holds. With this aim, we first notice that, by Jensen's inequality and stationarity, for any $\tau > 0$,

$$\mathbb{E}((X_{k,\ell}^{(m)})^2 \mathbf{1}_{|X_{k,\ell}^{(m)}| > \tau\sqrt{n}}) \le \mathbb{E}(X_{1,1}^2 \mathbf{1}_{|X_{1,1}^{(m)}| > \tau\sqrt{n}}) \, .$$

Notice now that if X is a real-valued random variable and $\mathcal{F}$ is a sigma-algebra, then for any $\varepsilon > 0$, we have by Lemma 6.3 of [22] that

$$\mathbb{E}\Big(X^2 \mathbf{1}_{|\mathbb{E}(X|\mathcal{F})| > 2\varepsilon}\Big) \le 2\,\mathbb{E}\Big(X^2 \mathbf{1}_{|X| > \varepsilon}\Big) \, .$$

Therefore,

$$\mathbb{E}((X_{k,\ell}^{(m)})^2 \mathbf{1}_{|X_{k,\ell}^{(m)}| > \tau\sqrt{n}}) \le 2\mathbb{E}(X_{1,1}^2 \mathbf{1}_{|X_{1,1}| > \tau\sqrt{n}/2})$$

which proves that $(X_{k,\ell}^{(m)}\,,\,1 \le \ell \le k \le n)$ satisfies $\mathbf{A}_2$ because $\mathbb{E}(X_{1,1}^2) < \infty$.

Since $(X_{k,\ell}^{(m)}\,,\,1 \le \ell \le k \le n)$ satisfies the assumptions $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ of Section 3.4, applying Theorem 3.11, (3.34) follows.

Approximation of the Gaussian matrices

According to (3.30) and (3.34), the theorem will follow if we prove the following convergence: for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\Big|\mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n^{(m)}}(z)\Big)\Big| = 0\,. \tag{3.35}$$

With this aim, we apply Lemma A.5 from Section A.2 which gives

$$\begin{aligned}\mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) &- \mathbb{E}\Big(S_{\mathbb{G}_n^{(m)}}(z)\Big)\\ &= \frac{1}{2}\sum_{1\le \ell\le k\le n}\ \sum_{1\le j\le i\le n}\int_0^1\Big(\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(G_{k,\ell}^{(m)}G_{i,j}^{(m)})\Big)\mathbb{E}\Big(\partial_{k\ell}\partial_{ij}f(\mathbf{g}(t))\Big)\,,\end{aligned}$$

where f is defined in (A.17) and , for $t \in [0,1]$,

$$\mathbf{g}(t) = (\sqrt{t}G_{k,\ell} + \sqrt{1-t}G_{k,\ell}^{(m)})_{1\le \ell\le k\le n}\,.$$

We shall prove that, for any t in $[0,1]$,

$$\begin{aligned}\Big|\sum_{1\le \ell\le k\le n}\ \sum_{1\le j\le i\le n}\Big(\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(G_{k,\ell}^{(m)}G_{i,j}^{(m)})\Big)\mathbb{E}\Big(\partial_{k\ell}\partial_{ij}f(\mathbf{g}(t))\Big)\Big|&\\ \le C\|X_{0,0}^{(m)} - X_{0,0}\|_2\|X_{0,0}\|_2\,.&\end{aligned} \tag{3.36}$$

where C is a constant not depending on n and t. Integrating on $[0,1]$ and then taking into account that $\|X_{0,0} - X_{0,0}^{(m)}\|_2^2 \to 0$ as $m \to \infty$, (3.35) follows by letting n and then m tend to infinity.

To prove (3.36), using (3.33) and (3.4), we write now the following decomposition:

$$\begin{aligned}\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(G_{k,\ell}^{(m)}G_{i,j}^{(m)}) &= \mathbb{E}(X_{k,\ell}X_{i,j}) - \mathbb{E}(X_{k,\ell}^{(m)}X_{i,j}^{(m)}) \\ &= \mathbb{E}(X_{k,\ell}(X_{i,j} - X_{i,j}^{(m)})) - \mathbb{E}((X_{k,\ell}^{(m)} - X_{k,\ell})X_{i,j}^{(m)})\,. \end{aligned} \tag{3.37}$$

We shall decompose the sum on the left-hand side of (3.36) in two sums according to the decomposition (3.37) and analyze them separately as done in Proposition 2.13.

Next by Lemma A.4 from from Section A.2 applied with $a_{k,\ell} = (X_{k,\ell}^{(m)} - X_{k,\ell})$ and $b_{k,\ell} = X_{k,\ell}^{(m)}$ gives: for any $z = u + iv \in \mathbb{C}^+$,

$$\begin{aligned}&\Big| \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} ((X_{k,\ell}^{(m)} - X_{k,\ell})X_{i,j}^{(m)})\Big(\partial_{k\ell}\partial_{ij} f(\mathbf{g}(t))\Big)\Big| \\ &\qquad\le \frac{2}{v^3 n^2}\Big(\sum_{1\le \ell\le k\le n}(X_{k,\ell}^{(m)} - X_{k,\ell})^2\Big)^{1/2}\Big(\sum_{1\le j\le i\le n}(X_{i,j}^{(m)})^2\Big)^{1/2}.\end{aligned}$$

Therefore, by Cauchy-Schwarz's inequality, we derive

$$\begin{aligned}&\Big| \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} \mathbb{E}((X_{k,\ell}^{(m)} - X_{k,\ell})X_{i,j}^{(m)})\mathbb{E}\left(\partial_{k\ell}\partial_{ij} f(\mathbf{g}(t))\right)\Big| \\ &\qquad\le \frac{2}{v^3 n^2}\Big(\sum_{1\le \ell\le k\le n}\mathbb{E}(X_{k,\ell}^{(m)} - X_{k,\ell})^2 \sum_{1\le j\le i\le n}\mathbb{E}(X_{i,j}^{(m)})^2\Big)^{1/2}.\end{aligned}$$

Using stationarity it follows that, for any $z = u + iv \in \mathbb{C}^+$ and any t in $[0, 1]$,

$$\begin{aligned}&\Big| \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} \mathbb{E}((X_{k,\ell}^{(m)} - X_{k,\ell})X_{i,j}^{(m)})\mathbb{E}\Big(\partial_{k\ell}\partial_{ij} f(\mathbf{g}(t))\Big)\Big| \\ &\qquad\le 2v^{-3}\|X_{0,0}^{(m)} - X_{0,0}\|_2\|X_{0,0}\|_2\,.\end{aligned}$$

Similarly, we can prove that for any $z = u + iv \in \mathbb{C}^+$ and any t in $[0, 1]$,

$$\begin{aligned}&\Big| \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} \mathbb{E}(X_{k,\ell}(X_{i,j} - X_{i,j}^{(m)}))\mathbb{E}\Big(\partial_{k\ell}\partial_{ij} f(\mathbf{g}(t))\Big)\Big| \\ &\qquad\le 2v^{-3}\|X_{0,0}^{(m)} - X_{0,0}\|_2\|X_{0,0}\|_2\,.\end{aligned}$$

This leads to (3.36) and then ends the proof of the theorem.

□

3.6 Proof of Theorem 3.3

In order to establish Theorem 3.3, it suffices to apply Theorem 3.1 and to derive the limit of $\mathbb{E}(S_{\mathbb{G}_n}(z))$ for any $z \in \mathbb{C}^+$, where $\mathbb{G}_n$ is the symmetric matrix defined in Theorem 3.1. With this aim, we apply Proposition A.10 given in the Appendix.

Proposition A.10 is a modification of Theorem 2 in Khorunzhy and Pastur [40] (see also Theorem 17.2.1 in [56]) since in our case, we cannot use directly the conclusion of their theorem: we are not exactly in the situation described there. Their symmetric matrix is defined via a *symmetric* real-valued centered Gaussian random field $(W_{k,\ell})_{k,\ell}$ satisfying the following property:

$$W_{k,\ell} = W_{\ell,k} \quad \text{for any } (k,\ell) \in \mathbb{Z}^2$$

and also (2.8) in [40].

In our situation, and if (3.8) is assumed, the entries $(g_{k,\ell})_{1 \le k,\ell \le n}$ of $n^{1/2}\mathbb{G}_n$ have the following covariances

$$\mathbb{E}(g_{i,j}g_{k,\ell}) = \gamma_{i-k,j-\ell}(\mathbf{1}_{i\ge j,k\ge \ell} + \mathbf{1}_{j>i,\ell>k}) + \gamma_{i-\ell,j-k}(\mathbf{1}_{i\ge j,\ell>k} + \mathbf{1}_{j>i,k\ge\ell})\,, \qquad (3.38)$$

since by (3.4) and stationarity

$$g_{k,\ell} = G_{\max(k,\ell),\min(k,\ell)} \ \text{ and } \ \mathbb{E}(G_{i,j}, G_{k,\ell}) = \gamma_{k-i,\ell-j}\,.$$

Hence, because of the indicator functions appearing in (3.38), our covariances do not satisfy the condition (2.8) in [40]. However, the conclusion of Theorem 2 in [40] also holds for $S_{\mathbb{G}_n}(z)$ provided that (3.7) and (3.8) are satisfied. We did not find any reference where the assertion above is mentioned so Proposition A.10 is proved with this aim. □

3.7 Concentration of the spectral measure

Next proposition is a generalization to row-wise K-dependent random matrices of Theorem 1 (ii) of Guntuboyina and Leeb [32].

Proposition 3.12. *Let $(X_{k,\ell}^{(n)})_{1\le \ell\le k\le n}$ be an array of complex-valued random variables defined on a common probability space. Assume that there exists a positive integer K such that for any integer $u\in[1,n-K]$, the σ-fields*

$$\sigma\Big(X_{i,j}^{(n)}\,,\,1\le j\le i\le u\Big)\quad\text{and}\quad\sigma\Big(X_{k,\ell}^{(n)}\,,\;1\le \ell\le k\,,\,u+K+1\le k\le n\Big)$$

are independent. Define the symmetric matrix $\mathbb{X}_n$ by (3.2). Then for every measurable function $f:\mathbb{R}\to\mathbb{R}$ of bounded variation, any $n\ge K$ and any $r\ge 0$,

$$\mathbb{P}\Big(\Big|\int f d\nu_{\mathbb{X}_n}-\mathbb{E}\int f d\nu_{\mathbb{X}_n}\Big|\ge r\Big)\le 2\exp\Big(-\frac{nr^2}{160KV_f^2}\Big)\,, \tag{3.39}$$

where V_f is the variation of the function f.

Application to the Stieltjes transform. Assume that the assumptions of Proposition 3.12 hold. Let $z=u+\mathrm{i}v\in\mathbb{C}^+$ and note that

$$S_{\mathbb{X}_n}(z)=\int\frac{1}{x-z}d\nu_{\mathbb{X}_n}(x)=\int f_1(x)d\nu_{\mathbb{X}_n}(x)+\mathrm{i}\int f_2(x)d\nu_{\mathbb{X}_n}(x)\,,$$

where $f_1(x)=\frac{x-u}{(x-u)^2+v^2}$ and $f_2(x)=\frac{v}{(x-u)^2+v^2}$. Now

$$V_{f_1}=\|f_1'\|_1=\frac{2}{v}\ \text{and}\ V_{f_2}=\|f_2'\|_1=\frac{2}{v}\,.$$

Therefore, by applying Proposition 3.12 to f_1 and f_2, we get that for any $n\ge K$ and any $r\ge 0$,

$$\mathbb{P}\Big(\Big|S_{\mathbb{X}_n}(z)-\mathbb{E}S_{\mathbb{X}_n}(z)\Big|\ge r\Big)\le 4\exp\Big(-\frac{nr^2v^2}{2560K}\Big)\,. \tag{3.40}$$

Proof of Proposition 3.12. It is convenient to start by considering the map A which "constructs" symmetric matrices of order n as in (3.2). To define it, let $N=n(n+1)/2$

and write elements of $\mathbb{R}^N$ as $\mathbf{x} = (r_1, \dots, r_n)$ where $r_i = (x_{i,j})_{1\le j\le i}$. For any $\mathbf{x} \in \mathbb{R}^N$, let $A(\mathbf{x}) = A(r_1, \dots, r_n)$ be the matrix defined by

$$(A(\mathbf{x}))_{ij} = \frac{1}{\sqrt{n}} \begin{cases} x_{i,j} = (r_i)_j & \text{if } i \ge j \\ x_{j,i} = (r_j)_i & \text{if } i < j\,. \end{cases} \tag{3.41}$$

For $1 \le i \le n$, let $R_i = (X_{i,j}^{(n)})_{1\le j\le i}$. By definition, we have that

$$\mathbb{X}_n = A(R_1, \dots, R_n)\,.$$

Let then h be the function from $\mathbb{C}^N$ to $\mathbb{R}$ defined by

$$h(R_1, \dots, R_n) = \int f d\nu_{A(R_1,\dots,R_n)}\,.$$

Let $n \ge K$. Denoting by $\mathcal{F}_k = \sigma(R_1, \dots, R_k)$ for $k \ge 1$, and by $\mathcal{F}_0 = \{\emptyset, \Omega\}$, we then write the following martingale decomposition:

$$\begin{aligned}\int f d\nu_{\mathbb{X}_n} - \mathbb{E}\int f d\nu_{\mathbb{X}_n} &= h(R_1, \dots, R_n) - \mathbb{E}h(R_1, \dots, R_n) \\ &= \sum_{i=1}^{[n/K]} \Big(\mathbb{E}\Big(h(R_1, \dots, R_n)|\mathcal{F}_{iK}\Big) - \mathbb{E}\Big(h(R_1, \dots, R_n)|\mathcal{F}_{(i-1)K}\Big)\Big) \\ &\qquad + \mathbb{E}\Big(h(R_1, \dots, R_n)|\mathcal{F}_n\Big) - \mathbb{E}\Big(h(R_1, \dots, R_n)|\mathcal{F}_{K[n/K]}\Big) \\ &:= \sum_{i=1}^{[n/K]+1} d_{i,n}\,.\end{aligned}$$

Let

$$\mathbf{R}_n = (R_1, \dots, R_n) \quad \text{and} \quad \mathbf{R}_n^{k,\ell} = (R_1, \dots, R_k, 0, \dots, 0, R_{\ell+1}, \dots, R_n)\,.$$

Note now that, for any $i \in \{1, \dots, [n/K]\}$,

$$\mathbb{E}\Big(h(\mathbf{R}_n^{(i-1)K,(i+1)K})|\mathcal{F}_{iK}\Big) = \mathbb{E}\Big(h(\mathbf{R}_n^{(i-1)K,(i+1)K})|\mathcal{F}_{(i-1)K}\Big)\,. \tag{3.42}$$

To see this it suffices to apply Lemma A.6 with

$$X = (R_{(i+1)K+1}, \dots, R_n)\,, \quad Y = (R_1, \dots, R_{(i-1)K})$$

and

$$Z = (R_1, \dots, R_{iK}) .$$

Therefore, by taking into account (3.42), we get that, for any $i \in \{1, \dots, [n/K]\}$,

$$\begin{aligned} &\mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{iK}\Big) - \mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{(i-1)K}\Big) \\ &= \mathbb{E}\Big(h(\mathbf{R}_n) - h(\mathbf{R}_n^{(i-1)K,(i+1)K})|\mathcal{F}_{iK}\Big) - \mathbb{E}\Big(h(\mathbf{R}_n) - h(\mathbf{R}_n^{(i-1)K,(i+1)K})|\mathcal{F}_{(i-1)K}\Big) . \end{aligned} \tag{3.43}$$

We write now that

$$\begin{aligned} &h(\mathbf{R}_n) - h(\mathbf{R}_n^{(i-1)K,(i+1)K}) \\ &= \sum_{j=iK+1}^{(i+1)K} \Big(h(\mathbf{R}_n^{iK,j-1}) - h(\mathbf{R}_n^{iK,j})\Big) + \sum_{j=(i-1)K+1}^{iK} \Big(h(\mathbf{R}_n^{j,(i+1)K}) - h(\mathbf{R}_n^{j-1,(i+1)K})\Big) , \end{aligned} \tag{3.44}$$

since $\mathbf{R}_n = \mathbf{R}_n^{iK,iK}$. But if $\mathbb{Y}_n$ and $\mathbb{Z}_n$ are two symmetric matrices of size n, then

$$\Big| \int f d\nu_{\mathbb{Y}_n} - \int f d\nu_{\mathbb{Z}_n} \Big| \le V_f \| F^{\mathbb{Y}_n} - F^{\mathbb{Z}_n} \|_\infty$$

(see for instance the proof of Theorem 6 in [32]). Hence, from Theorem A.43 in Bai and Silverstein [5],

$$\Big| \int f d\nu_{\mathbb{Y}_n} - \int f d\nu_{\mathbb{Z}_n} \Big| \le \frac{V_f}{n} \operatorname{Rank}\Big(\mathbb{Y}_n - \mathbb{Z}_n\Big) .$$

With our notations, this last inequality implies that for any $0 \le k \le \ell \le n$ and $0 \le i \le j \le n$

$$\Big| h\Big(\mathbf{R}_n^{k,\ell}\Big) - h\Big(\mathbf{R}_n^{i,j}\Big) \Big| \le \frac{V_f}{n} \operatorname{Rank}\Big(A(\mathbf{R}_n^{k,\ell}) - A(\mathbf{R}_n^{i,j})\Big) . \tag{3.45}$$

Starting from (3.44) and using (3.45) together with

$$\operatorname{Rank}\Big(A(\mathbf{R}_n^{iK,j-1}) - A(\mathbf{R}_n^{iK,j})\Big) \le 2$$

and

$$\operatorname{Rank}\Big(A(\mathbf{R}_n^{j,(i+1)K}) - A(\mathbf{R}_n^{j-1,(i+1)K})\Big) \le 2 ,$$

we get that

$$\Big| h(\mathbf{R}_n) - h\Big(\mathbf{R}_n^{(i-1)K,(i+1)K}\Big) \Big| \le \frac{4K}{n} V_f . \tag{3.46}$$

Starting from (3.43) and using (3.46), it follows that, for any $i \in \{1, \dots, [n/K]\}$,

$$\left|\mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{iK}\Big) - \mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{(i-1)K}\Big)\right| \leq \frac{8K}{n} V_f$$

On another hand, since $\mathbf{R}_n^{K[n/K],n}$ is $\mathcal{F}_{K[n/K]}$-measurable,

$$\begin{aligned}&\mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_n\Big) - \mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{K[n/K]}\Big) \\ &\qquad = \mathbb{E}\Big(h(\mathbf{R}_n) - h(\mathbf{R}_n^{K[n/K],n})|\mathcal{F}_n\Big) - \mathbb{E}\Big(h(\mathbf{R}_n) - h(\mathbf{R}_n^{K[n/K],n})|\mathcal{F}_{K[n/K]}\Big)\,.\end{aligned}$$

Now

$$h(\mathbf{R}_n) - h(\mathbf{R}_n^{K[n/K],n}) = \sum_{j=K[n/K]+1}^{n} \Big(h(\mathbf{R}_n^{j,n}) - h(\mathbf{R}_n^{j-1,n})\Big)\,.$$

So, proceeding as before, we infer that

$$\left|\mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_n\Big) - \mathbb{E}\Big(h(\mathbf{R}_n)|\mathcal{F}_{K[n/K]}\Big)\right| \leq \frac{4K}{n} V_f\,.$$

So, overall we derive that for any $i \in \{1, \dots, [n/K]\}$

$$\|d_{i,n}\|_\infty \leq \frac{8K}{n} V_f \quad \text{and} \quad \|d_{[n/K]+1,n}\|_\infty \leq \frac{4K}{n} V_f\,.$$

Therefore, the proposition follows by applying the Azuma-Hoeffding inequality for martingales.

□

3.8 Proof of Theorem 3.11 via the Lindeberg method by blocks

By using inequality (3.40) together with the Borel-Cantelli lemma, it follows that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left|S_{\mathbb{X}_n}(z) - \mathbb{E}\Big(S_{\mathbb{X}_n}(z)\Big)\right| = 0 \ \text{ almost surely.}$$

To prove the almost sure convergence (3.27), it suffices then to prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left| \mathbb{E}\Big(S_{\mathbb{X}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) \right| = 0\,. \tag{3.47}$$

We start by truncating the entries of the matrix $\mathbb{X}_n$. Since $\mathbf{A_2}$ holds, we can consider a decreasing sequence of positive numbers τ_n such that, as $n \to \infty$,

$$\tau_n \to 0\ ,\ \ L_n(\tau_n) \to 0 \ \text{ and } \ \tau_n \sqrt{n} \to \infty\,. \tag{3.48}$$

Let $\bar{\mathbf{X}}_n = [\bar{X}_{k,\ell}]_{k,\ell=1}^n$ be the symmetric matrix of order n whose entries are given by:

$$\bar{X}_{k,\ell} = X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\le \tau_n\sqrt{n}} - \mathbb{E}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\le \tau_n\sqrt{n}}\Big)$$

and define $\bar{\mathbb{X}}_n := n^{-1/2}\bar{\mathbf{X}}_n$. Using (3.48), it has been proved in Section 2.1 of [30] that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left| \mathbb{E}\Big(S_{\mathbb{X}_n}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) \right| = 0\,.$$

Therefore, to prove (3.47) (and then the theorem), it suffices to show that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left| \mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) \right| = 0\,. \tag{3.49}$$

The proof of (3.49) is then divided in three steps. The first step consists of replacing in (3.49), the matrix $\mathbb{G}_n$ by a symmetric matrix $\bar{\mathbb{G}}_n$ of order n whose entries are real-valued Gaussian random variables with the same covariance structure as the entries of $\bar{\mathbb{X}}_n$. The second step consists of "approximating" $\bar{\mathbb{X}}_n$ and $\bar{\mathbb{G}}_n$ by matrices with "big square independent blocks" containing the entries spaced by "small blocks" around them containing only zeros as entries. Due to the assumption $\mathbf{A_3}$, the random variables contained in two different big blocks will be independent. The third and last step consists of proving the mean convergence (3.49) but with $\bar{\mathbb{X}}_n$ and $\mathbb{G}_n$ replaced by their approximating matrices with independent blocks. This step will be achieved with the help of the Lindeberg method.

Step 1. Let $\bar{\mathbf{G}}_n = [\bar{G}_{k,\ell}]_{k,\ell=1}^n$ be the symmetric matrix of order n whose entries $(\bar{G}_{k,\ell}\,,\, 1 \leq \ell \leq k \leq n)$ are real-valued centered Gaussian random variables with the following covariance structure: for any $1 \leq \ell \leq k \leq n$ and any $1 \leq j \leq i \leq n$,

$$\mathbb{E}(\bar{G}_{k,\ell}\bar{G}_{i,j}) = \mathbb{E}(\bar{X}_{k,\ell}\bar{X}_{i,j})\,. \tag{3.50}$$

There is no loss of generality by assuming in the rest of the proof that the σ-fields $\sigma(\bar{G}_{k,\ell}\,,\, 1 \leq \ell \leq k \leq n)$ and $\sigma(X_{k,\ell}\,,\, 1 \leq \ell \leq k \leq n)$ are independent.

Letting $\bar{\mathbb{G}}_n = \frac{1}{\sqrt{n}}\bar{\mathbf{G}}_n$, we shall prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big)\Big| = 0\,. \tag{3.51}$$

Applying Lemma A.5, we get

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big)\Big| \leq \frac{2}{v^3 n^2} \sum_{1\leq \ell\leq k\leq n} \sum_{1\leq j\leq i\leq n} |\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(\bar{G}_{k,\ell}\bar{G}_{i,j})|\,, \tag{3.52}$$

where $v = \mathrm{Im}(z)$. Recall now that

$$\mathbb{E}(G_{k,\ell}G_{i,j}) = \mathbb{E}(X_{k,\ell}X_{i,j}) \quad \text{and} \quad \mathbb{E}(\bar{G}_{k,\ell}\bar{G}_{i,j}) = \mathbb{E}(\bar{X}_{k,\ell}\bar{X}_{i,j})\,.$$

Hence, setting $b_n = \tau_n\sqrt{n}$, we have

$$\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(\bar{G}_{k,\ell}\bar{G}_{i,j}) = \mathrm{Cov}\Big(X_{k,\ell}, X_{i,j}\Big) - \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\leq b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\leq b_n}\Big)\,.$$

Note that

$$\begin{aligned}
&\mathrm{Cov}\Big(X_{k,\ell}, X_{i,j}\Big) - \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\leq b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\leq b_n}\Big)\\
&\quad = \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\leq b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|> b_n}\Big) + \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|> b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|> b_n}\Big)\\
&\qquad\qquad + \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|> b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\leq b_n}\Big)
\end{aligned}$$

implying, by Cauchy-Schwarz's inequality, that

$$\begin{aligned}&\Big|\mathrm{Cov}\Big(X_{k,\ell}, X_{i,j}\Big) - \mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\le b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\le b_n}\Big)\Big| \\ &\qquad\le 2b_n\mathbb{E}\Big(|X_{i,j}|\mathbf{1}_{|X_{i,j}|>b_n}\Big) + 2b_n\mathbb{E}\Big(|X_{k,\ell}|\mathbf{1}_{|X_{k,\ell}|>b_n}\Big) \\ &\qquad\quad + 2\|X_{i,j}\mathbf{1}_{|X_{i,j}|>b_n}\|_2\|X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|>b_n}\|_2 \\ &\qquad\le 3\,\mathbb{E}\Big(|X_{i,j}|^2\mathbf{1}_{|X_{i,j}|>b_n}\Big) + 3\,\mathbb{E}\Big(|X_{k,\ell}|^2\mathbf{1}_{|X_{k,\ell}|>b_n}\Big)\,.\end{aligned}$$

Note also that, by assumption $\mathbf{A}_3$,

$$\begin{aligned}&\Big|\mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\le b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\le b_n}\Big) - \mathrm{Cov}\Big(X_{k,\ell}, X_{i,j}\Big)\Big| \\ &\quad = \mathbf{1}_{i\in[k-K,k+K]}\mathbf{1}_{j\in[\ell-K,\ell+K]}\Big|\mathrm{Cov}\Big(X_{k,\ell}\mathbf{1}_{|X_{k,\ell}|\le b_n}, X_{i,j}\mathbf{1}_{|X_{i,j}|\le b_n}\Big) - \mathrm{Cov}\Big(X_{k,\ell}, X_{i,j}\Big)\Big|\,.\end{aligned}$$

So, overall,

$$\begin{aligned}&\Big|\mathbb{E}(G_{k,\ell}G_{i,j}) - \mathbb{E}(\bar{G}_{k,\ell}\bar{G}_{i,j})\Big| \\ &\qquad\le 3\,\mathbb{E}\Big(|X_{i,j}|^2\mathbf{1}_{|X_{i,j}|>b_n}\Big)\mathbf{1}_{k\in[i-K,i+K]}\mathbf{1}_{\ell\in[j-K,j+K]} \\ &\qquad\quad + 3\,\mathbb{E}\Big(|X_{k,\ell}|^2\mathbf{1}_{|X_{k,\ell}|>b_n}\Big)\mathbf{1}_{i\in[k-K,k+K]}\mathbf{1}_{j\in[\ell-K,\ell+K]}\,.\end{aligned}$$

Hence, starting from (3.52) and taking into account the above inequality, we derive that

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big)\Big| \le \frac{12}{n^2v^3}(2K+1)^2\sum_{k=1}^{n}\sum_{\ell=1}^{k}\mathbb{E}\Big(|X_{k,\ell}|^2\mathbf{1}_{|X_{k,\ell}|>b_n}\Big)\,.$$

which converges to zero as n tends to infinity, by assumption $\mathbf{A}_2$. This ends the proof of (3.51).

Step 2: Reduction to matrices with independent blocks.

Let $p := p_n$ such that $p_n \to \infty$, $p_n/n \to 0$, and $\tau_n p_n^4 \to 0$. Clearly we can take $p_n > K$, $p_n + K \le n/3$, and set

$$q = q_n = \left[\frac{n}{p+K}\right] - 1\,. \tag{3.53}$$

Let

$$I_\ell = \Big[\ell(p+K)+1\,,\,\ell(p+K)+p\Big] \cap \mathbb{N}$$

and

$$\mathcal{E}_{k,\ell} = \left\{(u,v) \in I_k \times I_{\ell-1}\right\}. \tag{3.54}$$

For any $k = 1,\dots,q_n$ and any $\ell = 1,\dots,k$, we define now the real matrices $B_{k,\ell}$ of size $p \times p$ whose entries consist of all the $\bar{X}_{u,v}$ for $(u,v) \in \mathcal{E}_{k,\ell}$. More precisely,

$$B_{k,\ell} := \Big\{b_{k,\ell}(i,j)\Big\}_{i,j=1}^{p} \quad \text{where } b_{k,\ell}(i,j) = \bar{X}_{k(p+K)+i,(\ell-1)(p+K)+j}\,. \tag{3.55}$$

Similarly, we define the real matrices $B^*_{k,\ell}$ of size $p \times p$ whose entries consist of all the $\bar{G}_{u,v}$ for $(u,v) \in \mathcal{E}_{k,\ell}$. Therefore,

$$B^*_{k,\ell} := \Big\{b^*_{k,\ell}(i,j)\Big\}_{i,j=1}^{p} \quad \text{where } b^*_{k,\ell}(i,j) = \bar{G}_{k(p+K)+i,(\ell-1)(p+K)+j}\,. \tag{3.56}$$

Using the blocks $B_{k,\ell}$ we construct now an $n \times n$ matrix, $\widetilde{\mathbf{X}}_n$, by inserting 0's. Actually we start from the matrix $\overline{\mathbf{X}}_n$, keep the blocks $B_{k,\ell}$ and $B^T_{k,\ell}$ and replace all the other variables by 0's. For the sake of clarity we describe the south-western part of the matrix $\widetilde{\mathbf{X}}_n$ below, the other part being constructed by symmetry.

$$\widetilde{\mathbf{X}}_n := \begin{pmatrix}
\mathbf{0}_{p,p} & & & & & \dots & & & \\
\mathbf{0}_{K,p} & \mathbf{0}_{K,K} & & & & \dots & & & \\
B_{1,1} & \mathbf{0}_{p,K} & \mathbf{0}_{p,p} & & & \dots & & & \\
\mathbf{0}_{K,p} & \mathbf{0}_{K,K} & \mathbf{0}_{K,p} & \mathbf{0}_{K,K} & & \dots & & & \\
B_{2,1} & \mathbf{0}_{p,K} & B_{2,2} & \mathbf{0}_{p,K} & \mathbf{0}_{p,p} & \dots & & & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
B_{q-1,1} & \mathbf{0}_{p,K} & B_{q-1,2} & \mathbf{0}_{p,K} & B_{q-1,3} & \dots & & & \\
\mathbf{0}_{K,p} & \mathbf{0}_{K,K} & \mathbf{0}_{K,p} & \mathbf{0}_{K,K} & \mathbf{0}_{K,p} & \dots & \mathbf{0}_{K,K} & & \\
B_{q,1} & \mathbf{0}_{p,K} & B_{q,2} & \mathbf{0}_{p,K} & B_{q,3} & \dots & B_{q,q} & \mathbf{0}_{p,p} & \\
\mathbf{0}_{m,p} & \mathbf{0}_{m,K} & \mathbf{0}_{m,p} & \mathbf{0}_{m,K} & \mathbf{0}_{m,p} & \dots & \mathbf{0}_{m,K} & \mathbf{0}_{m,p} & \mathbf{0}_{m,m}
\end{pmatrix}, \tag{3.57}$$

where $m = m_n = n - q(p+K) - p$. $\widetilde{\mathbf{G}}_n$ is constructed as $\widetilde{\mathbf{X}}_n$ with the $B^*_{k,\ell}$ in place of the $B_{k,\ell}$.

In what follows, we shall prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n}(z)\Big)\Big| = 0 \quad \text{and} \quad \lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{G}}_n}(z)\Big)\Big| = 0\,, \qquad (3.58)$$

with $\widetilde{\mathbb{X}}_n := n^{-1/2}\widetilde{\mathbf{X}}_n$, $\widetilde{\mathbb{G}}_n := n^{-1/2}\widetilde{\mathbf{G}}_n$.

To prove it we first introduce two other symmetric $n\times n$ matrices $\widehat{\mathbf{X}}_n = [\widehat{X}_{k,\ell}]_{k,\ell=1}^n$ and $\widehat{\mathbf{G}}_n = [\widehat{G}_{k,\ell}]_{k,\ell=1}^n$ constructed from $\bar{\mathbf{X}}_n$ and $\bar{\mathbf{G}}_n$ respectively, by replacing the entries by zeros in square blocks of size p around the diagonal. More precisely, for any $1 \le i,j \le n$,

$$\widehat{X}_{i,j} = 0 \ \text{ if } (i,j) \in \cup_{\ell=0}^{q_n} \mathcal{E}_{\ell,\ell+1} \ \text{ and } \ \widehat{X}_{i,j} = \bar{X}_{i,j} \ \text{ otherwise}$$

and

$$\widehat{G}_{i,j} = 0 \ \text{ if } (i,j) \in \cup_{\ell=0}^{q_n} \mathcal{E}_{\ell,\ell+1} \ \text{ and } \ \widehat{G}_{i,j} = \bar{G}_{i,j} \ \text{ otherwise}$$

where we recall that the sets $\mathcal{E}_{\ell,\ell+1}$ have been defined in (3.54). Denote now $\widehat{\mathbb{X}}_n = \frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n$ and $\widehat{\mathbb{G}}_n = \frac{1}{\sqrt{n}}\widehat{\mathbf{G}}_n$.

By Lemma A.3, we get, for any $z = u + \mathrm{i}v \in \mathbb{C}^+$, that

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{X}}_n}(z)\Big)\Big|^2 \le \mathbb{E}\Big(\Big|S_{\bar{\mathbb{X}}_n}(z) - S_{\widehat{\mathbb{X}}_n}(z)\Big|^2\Big) \le \frac{1}{n^2v^4}\mathbb{E}\Big(\mathrm{Tr}\Big((\bar{\mathbf{X}}_n - \widehat{\mathbf{X}}_n)^2\Big)\Big)\,.$$

Therefore,

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{X}}_n}(z)\Big)\Big|^2 \le \frac{1}{n^2v^4}\sum_{\ell=0}^{q_n}\sum_{(i,j)\in\mathcal{E}_{\ell,\ell+1}} \mathbb{E}(|\bar{X}_{i,j}|^2)\,.$$

But $\|\bar{X}_{i,j}\|_\infty \le 2\tau_n\sqrt{n}$. Hence,

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{X}}_n}(z)\Big)\Big|^2 \le \frac{4}{n^2v^4}(q_n+1)p_n^2\tau_n^2 n \le \frac{4}{v^4}\tau_n^2 p_n\,.$$

By our selection of p_n, we obviously have that $\tau_n^2 p_n \to 0$ as $n \to \infty$. It follows that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{X}}_n}(z)\Big)\Big| = 0\,. \qquad (3.59)$$

With similar arguments, we get that, for any $z = u + \mathrm{i}v \in \mathbb{C}^+$,

$$\Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{G}}_n}(z)\Big)\Big|^2 \leq \frac{1}{n^2 v^4} \sum_{\ell=0}^{q_n} \sum_{(i,j)\in\mathcal{E}_{\ell,\ell}} \mathbb{E}(|\bar{G}_{i,j}|^2)\,.$$

But $\|\bar{G}_{i,j}\|_2 = \|\bar{X}_{i,j}\|_2$. So, as before, we derive that for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widehat{\mathbb{G}}_n}(z)\Big)\Big| = 0\,. \tag{3.60}$$

From (3.59) and (3.60), the mean convergence (3.58) follows if we prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\widehat{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n}(z)\Big)\Big| = 0 \quad \text{and} \quad \lim_{n\to\infty} \Big|\mathbb{E}\Big(S_{\widehat{\mathbb{G}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{G}}_n}(z)\Big)\Big| = 0\,, \tag{3.61}$$

Integrating by parts and applying Theorem A.43 in Bai and Silverstein [5], we get as in Proposition 2.12 that

$$\Big|S_{\widehat{\mathbb{X}}_n}(z) - S_{\widetilde{\mathbb{X}}_n}(z)\Big| \leq \frac{\pi}{vn}\mathrm{Rank}\Big(\widehat{\mathbb{X}}_n - \widetilde{\mathbb{X}}_n\Big)\,.$$

But, by counting the numbers of rows and of columns with entries that can be different from zero, we infer that

$$\mathrm{Rank}\Big(\widehat{\mathbb{X}}_n - \widetilde{\mathbb{X}}_n\Big) \leq 2(q_n K + m_n) \leq 2(np^{-1}K + p + 2K)\,.$$

Therefore,

$$\Big|S_{\widehat{\mathbb{X}}_n}(z) - S_{\widetilde{\mathbb{X}}_n}(z)\Big| \leq \frac{2\pi}{v}(Kp^{-1} + pn^{-1} + 2Kn^{-1})\,.$$

With similar arguments, we get

$$\Big|S_{\widehat{\mathbb{G}}_n}(z) - S_{\widetilde{\mathbb{G}}_n}(z)\Big| \leq \frac{2\pi}{v}(Kp^{-1} + pn^{-1} + 2Kn^{-1})\,.$$

Since $p = p_n \to \infty$ and $p_n/n \to 0$, as $n \to \infty$, (3.61) (and then (3.58)) follows from the two above inequalities. Therefore, to prove that the mean convergence (3.49) holds, it suffices to prove that, for any $z \in \mathbb{C}^+$,

$$\lim_{n\to\infty} \left|\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{G}}_n}(z)\Big)\right| = 0\,. \tag{3.62}$$

This is done in the next step.

Step 3: Lindeberg method.

To prove (3.62), we shall use the Lindeberg method. Recall that the σ-fields $\sigma(\bar{G}_{k,\ell}\,,\, 1 \leq \ell \leq k \leq n)$ and $\sigma(X_{k,\ell}\,,\, 1 \leq \ell \leq k \leq n)$ are assumed to be independent. Furthermore, by the hypothesis $\mathbf{A}_3$, all the blocks $(B_{k,\ell})$ and $(B^*_{k,\ell})$ $1 \leq \ell \leq k \leq q$ are independent.

We shall replace one by one the blocks $B_{k,\ell}$ by the "Gaussian" ones $B^*_{k,\ell}$ with the same covariance structure. So, starting from the matrix $\widetilde{\mathbf{X}}_n = \widetilde{\mathbf{X}}_n(0)$, the first step is to replace its block B_{q_n,q_n} by $B^*_{q_n,q_n}$, this gives a new matrix. Note that, at the same time, $B^T_{q_n,q_n}$ will also be replaced by $(B^*_{q_n,q_n})^T$. We denote this matrix by $\widetilde{\mathbf{X}}_n(1)$ and re-denote the block replaced by $B(1)$ and the new one by $B^*(1)$.

At the second step, we replace, in the new matrix $\widetilde{\mathbf{X}}_n(1)$, the block $B(2) := B_{q_n,q_n-1}$ by $B^*(2) := B^*_{q_n,q_n-1}$, and call the new matrix $\widetilde{\mathbf{X}}_n(2)$ and so on. Therefore, after the q_n-th step, in the matrix $\widetilde{\mathbf{X}}_n$ we have replaced the blocks $B(q_n - \ell + 1) = B_{q_n,\ell}\,,\ \ell = 1, \ldots, q_n$ (and their transposed) by the blocks $B^*(q_n - \ell + 1) = B^*_{q_n,\ell}\,,\ \ell = 1, \ldots, q_n$ (and their transposed) respectively. This matrix is denoted by $\widetilde{\mathbf{X}}_n(q_n)$.

Next, the $q_n + 1$-th step will consist of replacing the block $B(q_n + 1) = B_{q_n-1,q_n-1}$ by $B^*_{q_n-1,q_n-1}$ and obtain the matrix $\widetilde{\mathbf{X}}_n(q_n + 1)$. So finally after $q_n(q_n + 1)/2$ steps, we have replaced all the blocks $B_{k,\ell}$ and $B^T_{k,\ell}$ of the matrix $\widetilde{\mathbb{X}}_n$ to obtain at the end the matrix $\widetilde{\mathbf{X}}_n(q_n(q_n + 1)/2) = \widetilde{\mathbb{G}}_n$.

Therefore we have

$$\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{G}}_n}(z)\Big) = \sum_{k=1}^{k_n} \Big(\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k-1)}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k)}(z)\Big)\Big)\,. \tag{3.63}$$

where $k_n = q_n(q_n + 1)/2$.

Let k in $\{1, \ldots, k_n\}$ and notice that $\widetilde{\mathbb{X}}_n(k-1)$ and $\widetilde{\mathbb{X}}_n(k)$ differ only by the variables in the block $B(k)$ replaced at the step k. Define then the vector $\mathbf{X}$ of $\mathbb{R}^{p^2}$ consisting of all the entries of $B(k)$, the vector $\mathbf{Y}$ of $\mathbb{R}^{p^2}$ consisting of all the entries of $B^*(k)$ (in the same order we have defined the coordinates of $\mathbf{X}$). Denote by $\mathbf{Z}$ the vector of $\mathbb{R}^{N-p^2}$ (where $N = n(n+1)/2$) consisting of all the entries on and below the diagonal of $\widetilde{\mathbf{X}}_n(k-1)$ except the ones that are in the block matrix $B(k)$. More precisely if (u,v) are such that $B(k) = B_{u,v}$, then

$$\mathbf{X} = \Big((b_{u,v}(i,j))_{j=1,\dots,p} \,,\; i = 1,\dots,p \Big)$$

and

$$\mathbf{Y} = \Big((b^*_{u,v}(i,j))_{j=1,\dots,p} \,,\; i = 1,\dots,p \Big)$$

where $b_{u,v}(i,j)$ and $b^*_{u,v}(i,j)$ are defined in (3.55) and (3.56) respectively. In addition,

$$\mathbf{Z} = \Big((\widetilde{\mathbf{X}}_n(k-1))_{i,j} \,:\, 1 \le j \le i \le n \,,\, (i,j) \notin \mathcal{E}_{u,v} \Big) \,,$$

where $\mathcal{E}_{u,v}$ is defined in (3.54). The notations above allow to write

$$\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k-1)}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k)}(z)\Big) = \mathbb{E} f(\pi(\mathbf{X},\mathbf{Z})) - \mathbb{E} f(\pi(\mathbf{Y},\mathbf{Z})) \,,$$

where f is the function from $\mathbb{R}^N$ to $\mathbb{C}$ defined by (A.17) and $\pi : \mathbb{R}^N \to \mathbb{R}^N$ is a certain permutation. Note that, by our hypothesis $\mathbf{A}_3$ and our construction, the vectors $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are independent. Moreover $\mathbf{X}$ and $\mathbf{Y}$ are centered at expectation, have the same covariance structure and finite moments of order 3. Applying then Lemma A.7 of the Appendix and taking into account (A.18), we derive that, for a constant C depending only on $\mathrm{Im}(z)$,

$$\Big|\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k-1)}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k)}\Big)\Big| \le \frac{Cp^4}{n^{5/2}} \sum_{(i,j)\in\mathcal{E}_{u,v}} \Big(\mathbb{E}(|\bar{X}_{i,j}|^3) + \mathbb{E}(|\bar{G}_{i,j}|^3)\Big) \,.$$

So overall,

$$\sum_{k=1}^{k_n} |\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k-1)}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n(k)}\Big)| \le \frac{Cp^4}{n^{5/2}} \sum_{1\le\ell\le k\le q} \sum_{(i,j)\in\mathcal{E}_{k,\ell}} \Big(\mathbb{E}(|\bar{X}_{i,j}|^3) + \mathbb{E}(|\bar{G}_{i,j}|^3)\Big) \,. \tag{3.64}$$

By taking into account that

$$\mathbb{E}(|\bar{X}_{i,j}|^3) \leq 2\tau_n\sqrt{n}\mathbb{E}(|X_{i,j}|^2)$$

and also

$$\mathbb{E}(|\bar{G}_{i,j}|^3) \leq 2\left(\mathbb{E}(|\bar{G}_{i,j}|^2)\right)^{3/2} = 2\left(\mathbb{E}(|\bar{X}_{i,j}|^2)\right)^{3/2} \leq 4\tau_n\sqrt{n}\mathbb{E}(|X_{i,j}|^2)\,,$$

it follows from (3.63) and (3.64) that, for a constant C' depending only on $\mathrm{Im}(z)$,

$$\left|\mathbb{E}\Big(S_{\widetilde{\mathbb{X}}_n}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbb{G}}_n}(z)\Big)\right| \leq \frac{C'p^4}{n^2}\tau_n \sum_{1\leq j\leq i\leq n} \mathbb{E}(|X_{i,j}|^2)$$

which converges to 0 by $\mathbf{A}_1$ and the selection of p_n. This ends the proof of (3.62) and then of the theorem. □

Chapter 4

Matrices associated with functions of β-mixing processes

In this chapter, we study the model of Gram matrices considered in [60] as well as sample covariance matrices. But we shall consider a more general setting in which the matrix entries come from a non-causal stationary process $(X_k)_{k\in\mathbb{Z}}$ defined as follows: let $(\varepsilon_i)_{i\in\mathbb{Z}}$ an absolutely regular process with β-mixing sequence $(\beta_k)_{k\geq 0}$ and let $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function such that X_k, which is defined for any $k \in \mathbb{Z}$ by

$$X_k = g(\xi_k) \quad \text{with} \quad \xi_k = (\dots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \dots), \tag{4.1}$$

is a proper centered random variable having finite moment of second order; that is,

$$\mathbb{E}(X_k) = 0 \quad \text{and} \quad \|X_k\|_2 < \infty\,.$$

We first give a concentration inequality of the spectral measure allowing us to prove that, under an arithmetical decay condition on the β-mixing coefficients, the Stieltjes transform is concentrated almost surely around its expectation as n tends to infinity. This is a main step allowing us to replace the Stieltjes transform by its expectation.

Having this is done, the problem is reduced to proving that the latter converges to Stieltjes transform of a non-random probability measure. This will be achieved in Theorem 4.2 which allows approximating the expectation of the Stieltjes transform by that of a Gaussian matrix having a close covariance structure. Finally, provided that the spectral density of $(X_k)_k$ exists, we give in Theorem 4.3 the equation satisfied by the Stieltjes transform of the limiting distribution.

This chapter shall be organized as follows: In Section 4.1, we specify the models studied and state the limiting results. The proofs shall be deferred to Sections 4.3 and 4.4, whereas applications to examples of Markov chains and dynamical systems shall be introduced in Section 4.2.

4.1 Main results

Let $N := N(n)$ be a sequence of positive integers and consider the $N \times n$ random matrix $\mathcal{X}_n$ defined by

$$\mathcal{X}_n = ((\mathcal{X}_n)_{i,j}) = (X_{(j-1)N+i}) = \begin{pmatrix} X_1 & X_{N+1} & \cdots & X_{(n-1)N+1} \\ X_2 & X_{N+2} & \cdots & X_{(n-1)N+2} \\ \vdots & \vdots & & \vdots \\ X_N & X_{2N} & \cdots & X_{nN} \end{pmatrix} \in \mathcal{M}_{N\times n}(\mathbb{R}) \tag{4.2}$$

and note that its entries are dependent across both rows and columns. Let $\mathbf{B}_n$ be its corresponding Gram matrix given by

$$\mathbf{B}_n = \frac{1}{n}\mathcal{X}_n\mathcal{X}_n^T. \tag{4.3}$$

In what follows, $\mathbf{B}_n$ will be referred to as the Gram matrix associated with $(X_k)_{k\in\mathbb{Z}}$. We note that it can be written as the sum of dependent rank one matrices. Namely,

$$\mathbf{B}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T\,,$$

where for any $i = 1, \ldots, n$, $\mathbf{X}_i = (X_{(i-1)N+1}, \ldots, X_{iN})^T$.

We are interested in the study of the limiting distribution of the empirical spectral measure $\mu_{\mathbf{B}_n}$ defined by

$$\mu_{\mathbf{B}_n}(x) = \frac{1}{N}\sum_{k=1}^{N}\delta_{\lambda_k}\,,$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\mathbf{B}_n$. The main purpose of this chapter is to break the dependence structure of the matrix and use then approximation techniques from Chapter 2.

As a first step, we start by showing that if the β-mixing coefficients decay arithmetically then the Stieltjes transform of $\mathbf{B}_n$ concentrates almost surely around its expectation as n tends to infinity .

Theorem 4.1. *Let $\mathbf{B}_n$ be the matrix defined in (4.3) and associated with $(X_k)_{k\in\mathbb{Z}}$ defined in (4.1). If $\lim_{n\to\infty} N/n = c \in (0,\infty)$ and*

$$\sum_{n\geq 1} \log(n)^{\frac{3\alpha}{2}} n^{-\frac{1}{2}}\beta_n < \infty \quad \textit{for some } \alpha > 1\,, \tag{4.4}$$

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

$$S_{\mathbf{B}_n}(z) - \mathbb{E}(S_{\mathbf{B}_n}(z)) \to 0 \textit{ almost surely, as } n \to +\infty\,.$$

As the $\mathbf{X}_i$'s are dependent, classical arguments as those in Theorem 1 *(ii)* of Guntuboyina and Leeb [32] or on page 34 of [5] are not sufficient to prove the above convergence. In fact, we shall use maximal coupling for absolutely regular sequences in order to break the dependence structure between the columns allowing us to prove a concentration inequality of the associated empirical spectral measure. The proof of Theorem 4.1 shall be postponed to Section 4.3.

In the following Theorem, we shall approximate the expectation of the Stieltjes transform of $\mathbf{B}_n$ with that of a sample covariance matrix $\mathbf{G}_n$ which is the sum of i.i.d. rank one matrices associated with a Gaussian process $(Z_k)_{k\in\mathbb{Z}}$ having the same covariance structure as $(X_k)_{k\in\mathbb{Z}}$. Namely, for any $k, \ell \in \mathbb{Z}$,

$$\mathrm{Cov}(Z_k, Z_\ell) = \mathrm{Cov}(X_k, X_\ell)\,. \tag{4.5}$$

Denoting, for $i = 1, \ldots, n$, by $(Z_k^{(i)})_{k\in\mathbb{Z}}$ an independent copy of $(Z_k)_k$ that is also independent of $(X_k)_k$, we then define the $N \times N$ sample covariance matrix $\mathbf{G}_n$ by

$$\mathbf{G}_n = \frac{1}{n}\mathcal{Z}_n\mathcal{Z}_n^T = \frac{1}{n}\sum_{k=1}^{n}\mathbf{Z}_i\mathbf{Z}_i^T , \tag{4.6}$$

where for any $i = 1, \dots, n$, $\mathbf{Z}_i = (Z_1^{(i)}, \dots, Z_N^{(i)})^T$ and $\mathcal{Z}_n$ is the matrix whose columns are the $\mathbf{Z}_i$'s. Namely, $\mathcal{Z}_n := ((\mathcal{Z}_n)_{u,v}) = (Z_u^{(v)})$.

Theorem 4.2. *Let $\mathbf{B}_n$ and $\mathbf{G}_n$ be the matrices defined in (4.3) and (4.6) respectively. Provided that $\lim_{n\to\infty} N/n = c \in (0,\infty)$ and $\lim_{n\to\infty}\beta_n = 0$, then for any $z \in \mathbb{C}_+$,*

$$\lim_{n\to\infty} |\mathbb{E}(S_{\mathbf{B}_n}(z)) - \mathbb{E}(S_{\mathbf{G}_n}(z))| = 0.$$

The above Theorem allows us to reduce the study of the expectation of the LSD of $\mathbf{B}_n$ to that of a Gram matrix, being the sum of independent rank-one matrices associated with a Gaussian process, without requiring any rate of convergence to zero of the correlation between the entries nor of the β-mixing coefficients.

Theorem 4.3. *Let $\mathbf{B}_n$ and $\mathbf{G}_n$ be the matrices defined in (4.3) and (4.6) respectively. Provided that $\lim_{n\to\infty} N/n = c \in (0,\infty)$ and that (4.4) is satisfied, then for any $z \in \mathbb{C}_+$,*

$$\lim_{n\to\infty} |S_{\mathbf{B}_n}(z) - S_{\mathbf{G}_n}(z)| = 0 \quad a.s. \tag{4.7}$$

Moreover, if $(X_k)_{k\in\mathbb{Z}}$ admits a spectral density f, then with probability one, $\mu_{\mathbf{B}_n}$ converges weakly to a probability measure μ whose Stieltjes transform $S = S(z)$ ($z \in \mathbb{C}^+$) satisfies the equation

$$z = -\frac{1}{\underline{S}} + \frac{c}{2\pi}\int_0^{2\pi} \frac{1}{\underline{S} + \big(2\pi f(\lambda)\big)^{-1}} d\lambda , \tag{4.8}$$

where $\underline{S}(z) := -(1-c)/z + cS(z)$.

Check, for instance, Remarks 2.3 and 2.4.

Now, we shall consider the case where the Gram matrix is the sum of independent rank-one matrices whose entries are functionals of absolutely regular sequences. More precisely, we consider the sample covariance matrix

$$\mathbf{A}_n := \frac{1}{n}\sum_{k=1}^{n} \widetilde{\mathbf{X}}_k \widetilde{\mathbf{X}}_k^T, \tag{4.9}$$

with the $\widetilde{\mathbf{X}}_k$'s being independent copies of $(X_1, \ldots, X_N)^T$ and $(X_k)_{k\in\mathbb{Z}}$ the process defined in (4.1).

Theorem 4.4. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *and* $\lim_{n\to\infty} \beta_n = 0$*, then for any* $z \in \mathbb{C}$

$$\lim_{n\to\infty} |S_{\mathbf{A}_n}(z) - S_{\mathbf{G}_n}(z)| = 0 \quad a.s. \tag{4.10}$$

Moreover, if $(X_k)_{k\in\mathbb{Z}}$ *admits a spectral density* f*, then, with probability one,* $\mu_{\mathbf{A}_n}$ *converges weakly to a probability measure whose Stieltjes transform* $S = S(z)$ $(z \in \mathbb{C}^+)$ *satisfies equation* (4.8)*.*

Remark 4.5. *Since the* $\widetilde{\mathbf{X}}_k$*'s are mutually independent, then, by Theorem 1 (ii) in [32] or the arguments on page 34 in [5], we can approximate directly* $S_{\mathbf{A}_n}(z)$ *by its expectation and there is no need of any coupling arguments as in Theorem 4.1 and thus of the arithmetic decay condition* (4.4) *on the absolutely regular coefficients. So it suffices to prove that for any* $z \in \mathbb{C}_+$*,*

$$\lim_{n\to\infty} \left|\mathbb{E}\Big(S_{\mathbf{A}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big)\right| = 0$$

which can be exactly done as in Theorem 4.2 after simple modifications of indices. For this reason, its proof shall be omitted.

4.2 Applications

In this section we shall apply the results of Section 4.1 to a Harris recurrent Markov chain and some uniformly expanding maps in dynamical systems.

Harris recurrent Markov chain

We start by recalling that a stationary Markov chain is aperiodic and Harris recurrent if and only if it satisfies absolute regularity, i.e. $\beta_n \to 0$ as $n \to \infty$. For more details, one can see [16, 52].

The following example is a symmetrized version of the Harris recurrent Markov chain defined by Doukhan *et al.* [26]. Let $(\varepsilon_n)_{n\in\mathbb{Z}}$ be a stationary Markov chain taking values in $E = [-1, 1]$ and let K be its Markov kernel defined by

$$K(x, .) = (1 - |x|)\delta_x + |x|\nu \,,$$

with ν being a symmetric atomless law on E and δ_x denoting the Dirac measure at point x.

Assume that $\theta = \int_E |x|^{-1}\nu(dx) < \infty$ then $(\varepsilon_n)_{n\in\mathbb{Z}}$ is positively recurrent and the unique invariant measure π is given by

$$\pi(dx) = \theta^{-1}|x|^{-1}\nu(dx)\,.$$

We shall assume in what follows that ν satisfies for any $x \in [0, 1]$,

$$\frac{d\nu}{dx}(x) \leq c\, x^a \quad \text{for some } a\,,\ c > 0\,. \tag{4.11}$$

Now, let g be a measurable function defined on E such that

$$X_k = g(\varepsilon_k) \tag{4.12}$$

is a centered random variable having a finite second moment.

Corollary 4.6. *Let $(X_k)_{k\in\mathbb{Z}}$ be defined in (4.12). Assume that ν satisfies (4.11) and that for any $x \in E$,*

$$g(-x) = -g(x) \quad \textit{and} \quad |g(x)| \leq C|x|^{1/2}$$

with C being a positive constant. Then, provided that $N/n \to c \in (0, \infty)$, the conclusions of Theorems 4.2 and 4.4 hold and the limiting measure μ has a compact support. In addition, if (4.11) holds with $a > 1/2$ then Theorems 4.1 and 4.3 follow as well.

Proof. Doukhan *et al.* prove in Section 4 of [26] that if (4.11) is satisfied then $(\varepsilon_k)_{k\in\mathbb{Z}}$ is an absolutely regular sequence with

$$\beta_n = O(n^{-a}) \quad \text{as} \quad n \to \infty\,.$$

Thereby, Theorems 4.2 and 4.4 follow. Now, noting that g is an odd function we have

$$\mathbb{E}(g(\varepsilon_k)|\varepsilon_0) = (1-|\varepsilon_0|)^k g(\varepsilon_0) \quad \text{a.s.}$$

Therefore, by the assumption on g and (4.11), we get for any $k \geq 0$,

$$\begin{aligned}\gamma_k := \mathbb{E}(X_0X_k) &= \mathbb{E}\Big(g(\varepsilon_0)\mathbb{E}(g(\varepsilon_k)|\varepsilon_0)\Big)\\ &= \theta^{-1}\int_E g^2(x)(1-|x|)^k|x|^{-1}\nu(dx)\\ &\leq c\,C^2\,\theta^{-1}\int_E \frac{x^{a+1}}{|x|}(1-|x|)^k dx\,.\end{aligned}$$

By the properties of the Beta and Gamma functions, $|\gamma_k| = O(k^{-(a+1)})$ which implies $\sum_k |\gamma_k| < \infty$ and thus the spectral density f is continuous and bounded over $[0,2\pi)$ and the limiting measure μ has a compact support (see Remark 2.3).

However, if in addition $a > 1/2$ then (4.4) is also satisfied and Theorems 4.1 and 4.3 follow as well.

Uniformly expanding maps

Functionals of absolutely regular sequences occur naturally as orbits of chaotic dynamical systems. For instance, for uniformly expanding maps $T : [0,1] \to [0,1]$ with absolutely continuous invariant measure ν, one can write

$$T^k = g(\varepsilon_k, \varepsilon_{k+1}, \ldots)$$

for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ where $(\varepsilon_k)_{k\geq 0}$ is an absolutely regular sequence. We refer to Section 2 of [36] for more details and for a precise definition of such maps (see also Example 1.4 in [13]).

Hofbauer and Keller prove in Theorem 4 of [36] that the mixing rate of $(\varepsilon_k)_{k\geq 0}$

decreases exponentially, i.e.

$$\beta_k \leq Ce^{-\lambda k}\,, \quad \text{for some } C,\, \lambda > 0\,, \tag{4.13}$$

and thus (4.4) holds. Setting for any $k \geq 0$,

$$X_k = h \circ T^k - \nu(f), \tag{4.14}$$

where $h : [0,1] \to \mathbb{R}$ is a continuous Hölder function, the Theorems in Section 4.1 hold for the associated matrices $\mathbf{B}_n$ and $\mathbf{A}_n$. Moreover, Hofbauer and Keller prove in Theorem 5 of [36] that $\sum_k |\mathrm{Cov}(X_0, X_k)| < \infty$ which implies that the spectral density f exists, is continuous and bounded on $[0, 2\pi)$ and that the limiting measure μ is compactly supported.

4.3 Concentration of the spectral measure

In this Section, we give a proof of Theorem 4.1 with the help of a concentration inequality of the spectral measure. For absolutely regular sequences

Let m be a positive integer (fixed for the moment) such that $m \leq \sqrt{N}/2$ and let $(X_{k,m})_{k\in\mathbb{Z}}$ be the sequence defined for any $k \in \mathbb{Z}$ by,

$$X_{k,m} = \mathbb{E}(X_k|\varepsilon_{k-m}, \ldots, \varepsilon_{k+m}) := g_k(\varepsilon_{k-m}, \ldots, \varepsilon_{k+m})\,, \tag{4.15}$$

where g_k is a measurable function from $\mathbb{R}^{2m+1}$ to $\mathbb{R}$. Consider the $N \times n$ matrix $\mathcal{X}_{n,m} = ((\mathcal{X}_{n,m})_{i,j}) = (X_{(j-1)N+i,\,m})$ and finally set

$$\mathbf{B}_{n,m} = \frac{1}{n}\mathcal{X}_{n,m}\mathcal{X}_{n,m}^T. \tag{4.16}$$

The proof will be done in two principal steps. First, we shall prove that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left|S_{\mathbf{B}_n}(z) - S_{\mathbf{B}_{n,m}}(z)\right| = 0 \quad \text{a.s.} \tag{4.17}$$

and then

$$\lim_{n\to\infty} \left|S_{\mathbf{B}_{n,m}}(z) - \mathbb{E}\big(S_{\mathbf{B}_{n,m}}(z)\big)\right| = 0 \quad \text{a.s.} \tag{4.18}$$

We note that by Lemma A.1 we have, for any $z = u + iv \in \mathbb{C}^+$,

$$\left|S_{\mathbf{B}_n}(z) - S_{\mathbf{B}_{n,m}}(z)\right|^2 \leq \frac{2}{v^4}\left(\frac{1}{N}\operatorname{Tr}(\mathbf{B}_n + \mathbf{B}_{n,m})\right)\left(\frac{1}{Nn}\operatorname{Tr}(\mathcal{X}_n - \mathcal{X}_{n,m})(\mathcal{X}_n - \mathcal{X}_{n,m})^T\right). \tag{4.19}$$

Recall that mixing implies ergodicity and note that as $(\varepsilon_k)_{k\in\mathbb{Z}}$ is an ergodic sequence of real-valued random variables then $(X_k)_{k\in\mathbb{Z}}$ is also so. Therefore, by the ergodic theorem,

$$\lim_{n\to+\infty}\frac{1}{N}\operatorname{Tr}(\mathbf{B}_n) = \lim_{n\to+\infty}\frac{1}{Nn}\sum_{k=1}^{Nn} X_k^2 = \mathbb{E}(X_0^2) \quad \text{a.s.} \tag{4.20}$$

Similarly,

$$\lim_{n\to\infty}\frac{1}{N}\operatorname{Tr}(\mathbf{B}_{n,m}) = \mathbb{E}\Big(X_{0,m}^2\Big) \quad \text{a.s.} \tag{4.21}$$

Starting from (4.19) and noticing that $\mathbb{E}(X_{0,m}^2) \leq \mathbb{E}(X_0^2)$, it follows that (4.17) holds if we prove

$$\lim_{m\to+\infty}\limsup_{n\to+\infty}\left|\frac{1}{Nn}\operatorname{Tr}(\mathcal{X}_n - \mathcal{X}_{n,m})(\mathcal{X}_n - \mathcal{X}_{n,m})^T\right| = 0 \quad \text{a.s.} \tag{4.22}$$

By the construction of $\mathcal{X}_n$ and $\mathcal{X}_{n,m}$ and again the ergodic theorem, we get

$$\lim_{n\to\infty}\frac{1}{Nn}\operatorname{Tr}(\mathcal{X}_n - \mathcal{X}_{n,m})(\mathcal{X}_n - \mathcal{X}_{n,m})^T$$
$$= \lim_{n\to\infty}\frac{1}{Nn}\sum_{k=1}^{Nn}(X_k - X_{k,m})^2 = \mathbb{E}\Big(X_0 - X_{0,m}\Big)^2 \quad \text{a.s.}$$

(4.22) follows by applying the usual martingale convergence theorem in $\mathbb{L}^2$, from which we infer that

$$\lim_{m\to+\infty}\|X_0 - \mathbb{E}(X_0|\varepsilon_{-m},\ldots,\varepsilon_m)\|_2 = 0\,.$$

We turn now to the proof of (4.18). With this aim, we shall prove that for any $z = u + iv$ and $x > 0$,

$$\mathbb{P}\Big(\left|S_{\mathbf{B}_{n,m}}(z) - \mathbb{E}S_{\mathbf{B}_{n,m}}(z)\right| > 4x\Big)$$
$$\leq 4\exp\left\{-\frac{x^2\,v^2\,N^2(\log n)^\alpha}{256\,n^2}\right\} + \frac{32\,n^2(\log n)^\alpha}{x^2\,v^2\,N^2}\beta_{\left[\frac{n}{(\log n)^\alpha}\right]N}\,, \tag{4.23}$$

for some $\alpha > 1$. Noting that

$$\sum_{n\geq 2} (\log n)^{\alpha}\, \beta_{\left[\frac{n^2}{(\log n)^{\alpha}}\right]} < +\infty \text{ is equivalent to (4.4)}$$

and applying Borel-Cantelli Lemma, (4.18) follows from (4.4) and the fact that $\lim_{n\to\infty} N/n = c$. Now, to prove (4.23), we start by noting that

$$\begin{aligned}\mathbb{P}\Big(\Big|S_{\mathbf{B}_{n,m}}(z) - \mathbb{E}\left(S_{\mathbf{B}_{n,m}}(z)\right)\Big| > 4x\Big) \\ \leq \mathbb{P}\Big(\Big|\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big)\Big| > 2x\Big) \\ + \mathbb{P}\Big(\Big|\mathfrak{Im}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Im}\Big(S_{\mathbf{B}_{n,m}}(z)\Big)\Big| > 2x\Big)\end{aligned}$$

For a row vector $\mathbf{x} \in \mathbb{R}^{Nn}$, we partition it into n elements of dimension N and write $\mathbf{x} = \big(x_1, \ldots, x_n\big)$ where $x_1, \ldots, x_n$ are row vectors of $\mathbb{R}^N$. Now, let $A(\mathbf{x})$ and $B(\mathbf{x})$ be respectively the $N \times n$ and $N \times N$ matrices defined by

$$A(\mathbf{x}) = \left(x_1^T | \ldots | x_n^T\right) \quad \text{and} \quad B(\mathbf{x}) = \frac{1}{n} A(\mathbf{x}) A(\mathbf{x})^T.$$

Also, let $h_1 := h_{1,z}$ and $h_2 := h_{2,z}$ be the functions defined from $\mathbb{R}^{Nn}$ into $\mathbb{R}$ by

$$h_1(\mathbf{x}) = \int f_{1,z}\, d\mu_{B(\mathbf{x})} \quad \text{and} \quad h_2(\mathbf{x}) = \int f_{2,z}\, d\mu_{B(\mathbf{x})},$$

where

$$f_{1,z}(\lambda) = \frac{\lambda - u}{(\lambda - u)^2 + v^2} \quad \text{and} \quad f_{2,z}(\lambda) = \frac{v}{(\lambda - u)^2 + v^2}$$

and note that $S_{B(\mathbf{x})}(z) = h_1(\mathbf{x}) + i h_2(\mathbf{x})$. Now, denoting by $\mathbf{X}_{1,m}^T, \ldots, \mathbf{X}_{n,m}^T$ the columns of $\mathcal{X}_{n,m}$ and setting $\mathbf{A}$ to be the row random vector of $\mathbb{R}^{Nn}$ given by

$$\mathbf{A} = (\mathbf{X}_{1,m}, \ldots, \mathbf{X}_{n,m}),$$

we note that $B(\mathbf{A}) = \mathbf{B}_{n,m}$ and $h_1(\mathbf{A}) = \mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big)$.

Moreover, letting q be a positive integer less than n, we set $\mathcal{F}_i = \sigma(\varepsilon_k,\ k \leq iN + m)$ for $1 \leq i \leq [n/q]q$ with the convention that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and that $\mathcal{F}_s = \mathcal{F}_n$ for any $s \in$

$\{[n/q]q, \ldots, n\}$. Noting that $\mathbf{X}_{1,m}, \ldots, \mathbf{X}_{i,m}$ are $\mathcal{F}_i$-measurable, we write the following decomposition:

$$\begin{aligned}
&\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) \\
&\qquad = h_1(\mathbf{X}_{1,m}, \ldots, \mathbf{X}_{n,m}) - \mathbb{E}h_1(\mathbf{X}_{1,m}, \ldots, \mathbf{X}_{n,m}) \\
&\qquad\qquad = \sum_{i=1}^{[n/q]} \Big(\mathbb{E}(h_1(\mathbf{A})|\mathcal{F}_{iq}) - \mathbb{E}(h_1(\mathbf{A})|\mathcal{F}_{(i-1)q})\Big).
\end{aligned}$$

Now, let $(\mathbf{A}_i)_i$ be the vectors of $\mathbb{R}^{Nn}$ defined for any $i \in \{1, \ldots, [n/q]-1\}$ by

$$\mathbf{A}_i = \Big(\mathbf{X}_{1,m}\,, \ldots, \mathbf{X}_{(i-1)q,m}\,, \underbrace{\mathbf{0}_N, \ldots, \mathbf{0}_N}_{2q\,\text{times}}, \mathbf{X}_{(i+1)q+1,m}\,, \ldots, \mathbf{X}_{n,m}\Big),$$

and for $i = [n/q]$ by

$$\mathbf{A}_{[\frac{n}{q}]} = \Big(\mathbf{X}_{1,m}\,, \ldots, \mathbf{X}_{([n/q]-1)q,m}\,, \mathbf{0}_N, \ldots, \mathbf{0}_N\Big).$$

Noting that $\mathbb{E}\Big(h_1(\mathbf{A}_{[n/q]})|\mathcal{F}_n\Big) = \mathbb{E}(h_1\Big(\mathbf{A}_{[n/q]})|\mathcal{F}_{([n/q]-1)q}\Big)$, we write

$$\begin{aligned}
&\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) \\
&\quad = \sum_{i=1}^{[n/q]} \Big(\mathbb{E}\Big(h_1(\mathbf{A}) - h_1(\mathbf{A}_i)|\mathcal{F}_{iq}\Big) - \mathbb{E}\Big(h_1(\mathbf{A}) - h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q}\Big)\Big) \\
&\qquad\quad + \sum_{i=1}^{[n/q]-1} \Big(\mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}\Big) - \mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q}\Big)\Big) \\
&\qquad\qquad := M_{[n/q],\,q} + \sum_{i=1}^{[n/q]-1} \Big(\mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}\Big) - \mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q}\Big)\Big).
\end{aligned}$$

Thus, we get

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Re}\Big(S_{\mathbf{B}_{n,m}}(z)\Big)\Big| > 2x\Big) \\
&\quad \leq \mathbb{P}\left(|M_{[n/q],\,q}| > x\right) + \mathbb{P}\Big(\Big|\sum_{i=1}^{[n/q]-1} \Big(\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}) - \mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q})\Big)\Big| > x\Big).
\end{aligned} \tag{4.24}$$

Note that $(M_{k,q})_k$ is a centered martingale with respect to the filtration $(\mathcal{G}_{k,q})_k$ defined by $\mathcal{G}_{k,q} = \mathcal{F}_{kq}$. Moreover, for any $k \in \{1, \dots, [n/q]\}$,

$$\begin{aligned}\|M_{k,q} - M_{k-1,q}\|_\infty &= \|\mathbb{E}(h_1(\mathbf{A}) - h_1(\mathbf{A}_k)|\mathcal{F}_{kq}) - \mathbb{E}(h_1(\mathbf{A}) - h_1(\mathbf{A}_k)|\mathcal{F}_{(k-1)q})\|_\infty \\ &\leq 2\|h_1(\mathbf{A}) - h_1(\mathbf{A}_k)\|_\infty\end{aligned}$$

Noting that $\|f'_{1,z}\|_1 = 2/v$ then by integrating by parts, we get

$$\begin{aligned}|h_1(\mathbf{A}) - h_1(\mathbf{A}_k)| &= \left|\int f_{1,z} d\mu_{B(\mathbf{A})} - \int f_{1,z} d\mu_{B(\mathbf{A}_k)}\right| \\ &\leq \|f'_{1,z}\|_1 \|F^{B(\mathbf{A})} - F^{B(\mathbf{A}_k)}\|_\infty \\ &\leq \frac{2}{vN} \mathrm{Rank}\Big(A(\mathbf{A}) - A(\mathbf{A}_k)\Big),\end{aligned} \tag{4.25}$$

where the second inequality follows from Theorem A.44 in [5]. As for any $k \in \{1, \dots, [n/q]-1\}$,

$$\mathrm{Rank}\Big(A(\mathbf{A}) - A(\mathbf{A}_k)\Big) \leq 2q \quad \text{and} \quad \mathrm{Rank}\Big(A(\mathbf{A}) - A(\mathbf{A}_{[n/q]})\Big) \leq q\,,$$

then overall we derive that almost surely

$$\|M_{k,q} - M_{k-1,q}\|_\infty \leq \frac{8q}{vN} \quad \text{and} \quad \|M_{[n/q],q} - M_{[n/q]-1,q}\|_\infty \leq \frac{4q}{vN}$$

and hence applying the Azuma-Hoeffding inequality for martingales we get for any $x > 0$,

$$\mathbb{P}\left(|M_{[n/q],\,q}| > x\right) \leq 2\exp\left(-\frac{x^2 v^2 N^2}{128\, q\, n}\right). \tag{4.26}$$

Now to control the second term of the right-hand side of (4.24), we have, by Markov's inequality and orthogonality, for any $x > 0$,

$$\begin{aligned}&\mathbb{P}\Big(\Big|\sum_{i=1}^{[n/q]-1}\Big(\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}) - \mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q})\Big)\Big| > x\Big) \\ &\qquad\qquad \leq \frac{1}{x^2}\sum_{i=1}^{[n/q]-1}\|\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}) - \mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q})\|_2^2.\end{aligned} \tag{4.27}$$

Let i be a fixed integer in $\{1,\ldots,[n/q]-1\}$. To give an upper bound for $\|\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{iq})-\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q})\|_2$, we first notice that by Berbee's maximal coupling lemma [11], one can construct a sequence $(\varepsilon'_k)_{k\in\mathbb{Z}}$ distributed as $(\varepsilon_k)_{k\in\mathbb{Z}}$ and independent of $\mathcal{F}_{iq}=\sigma(\varepsilon_k,\ k\le iqN+m)$ such that

$$\mathbb{P}(\varepsilon'_k\neq\varepsilon_k,\ \text{for some } k\ge \ell+iqN+m)=\beta_\ell. \tag{4.28}$$

Let $(X'_{k,m})_{k\ge1}$ be the sequence defined for any $k\ge1$ by

$$X'_{k,m}=g_k(\varepsilon'_{k-m},\ldots,\varepsilon'_{k+m})\,,$$

with g_k being defined in (4.15) and let $\mathbf{X}'_{i,m}$ be the row vector of $\mathbb{R}^N$ defined by

$$\mathbf{X}'_{i,m}=(X'_{(i-1)N+1,m},\ldots,X'_{iN,m})\,.$$

Finally, we define for any $i\in\{1,\ldots,[n/q]-1\}$ the row random vector $\mathbf{A}'_i$ of $\mathbb{R}^{Nn}$ by

$$\mathbf{A}'_i=\Big(\mathbf{X}_{1,m}\,,\ldots,\mathbf{X}_{(i-1)q,m}\,,\underbrace{\mathbf{0}_N,\ldots,\mathbf{0}_N}_{2q\,\text{times}},\ \mathbf{X}'_{(i+1)q+1,m}\,,\ldots,\mathbf{X}'_{n,m}\Big).$$

As $\mathbf{X}'_{(i+1)q+1,m},\ldots,\mathbf{X}'_{n,m}$ are independent of $\mathcal{F}_{iq}$ then

$$\mathbb{E}(h_1(\mathbf{A}'_i)|\mathcal{F}_{iq})=\mathbb{E}(h_1(\mathbf{A}'_i)|\mathcal{F}_{(i-1)q})\,.$$

Thus we write

$$\begin{aligned}\mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}\Big)&-\mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q}\Big)\\&=\mathbb{E}\Big(h_1(\mathbf{A}_i)-h_1(\mathbf{A}'_i)|\mathcal{F}_{iq}\Big)-\mathbb{E}\Big(h_1(\mathbf{A}_i)-h_1(\mathbf{A}'_i)|\mathcal{F}_{(i-1)q}\Big).\end{aligned}$$

and infer that

$$\begin{aligned}&\|\mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}\Big) - \mathbb{E}\Big(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q}\Big)\|_2 \\ &\qquad \le \|\mathbb{E}\Big(h_1(\mathbf{A}_i) - h_1(\mathbf{A}'_i)|\mathcal{F}_{iq}\Big)\|_2 + \|\mathbb{E}\Big(h_1(\mathbf{A}_i) - h_1(\mathbf{A}'_i)|\mathcal{F}_{(i-1)q}\Big)\|_2 \\ &\qquad\qquad \le 2\|h_1(\mathbf{A}_i) - h_1(\mathbf{A}'_i)\|_2\,.\end{aligned}$$

Similarly as in (4.25), we have

$$\begin{aligned}\Big|h_1(\mathbf{A}_i) - h_1(\mathbf{A}'_i)\Big| &\le \frac{2}{vN}\mathrm{Rank}(A(\mathbf{A}_i) - A(\mathbf{A}'_i)) \\ &\le \frac{2}{vN}\sum_{\ell=(i+1)q+1}^{n} \mathbf{1}_{\{\mathbf{X}'_{\ell,m}\neq\mathbf{X}_{\ell,m}\}} \\ &\le \frac{2n}{vN}\,\mathbf{1}_{\{\varepsilon'_k\neq\varepsilon_k,\ \text{for some}\ k\ge (i+1)qN+1-m\}}\,.\end{aligned}$$

Hence by (4.28), we infer that

$$\|h_1(\mathbf{A}_i) - h_1(\mathbf{A}'_i)\|_2^2 \le \frac{4n^2}{v^2N^2}\,\beta_{qN+1-2m} \le \frac{4n^2}{v^2N^2}\,\beta_{(q-1)N}, \tag{4.29}$$

Starting from (4.27) and taking into account the above upper bounds, it follows that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{[n/q]-1}\Big(\mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{iq}) - \mathbb{E}(h_1(\mathbf{A}_i)|\mathcal{F}_{(i-1)q})\Big)\Big| > x\Big) \le \frac{16n^3}{x^2\,v^2q\,N^2}\beta_{(q-1)N}. \tag{4.30}$$

Therefore, considering (4.24) and gathering the upper bounds in (4.26) and (4.30), we get

$$\begin{aligned}&\mathbb{P}\Big(\Big|\mathfrak{Re}(S_{\mathbf{B}_{n,m}}(z)) - \mathbb{E}\,\mathfrak{Re}(S_{\mathbf{B}_{n,m}}(z))\Big| > 2x\Big) \\ &\qquad\qquad \le 2\exp\left(-\frac{x^2v^2N^2}{128\,q\,n}\right) + \frac{16n^3}{x^2\,v^2q\,N^2}\beta_{(q-1)N}.\end{aligned}$$

Finally, noting that $\mathbb{P}\Big(\Big|\mathfrak{Im}\Big(S_{\mathbf{B}_{n,m}}(z)\Big) - \mathbb{E}\,\mathfrak{Im}\Big(S_{\mathbf{B}_{n,m}}(z)\Big)\Big| > 2x\Big)$ also admits the same upper bound and choosing $q = [n/(\log n)^\alpha] + 1$, (4.23) follows. This ends the proof of Theorem 4.1.

4.4 Proof of Theorem 4.2

The proof, being technical, will be divided into three major steps (Sections 4.4.1 to 4.4.3).

4.4.1 A first approximation

Let m be a fixed positive integer and set $p := p(m) = a_m m$ with $(a_n)_{n\geq 1}$ being a sequence of positive integers such that $\lim_{n\to\infty} a_n = \infty$. Setting

$$k_N = \left[\frac{N}{p+3m}\right],$$

we write the subset $\{1, \dots, Nn\}$ as a union of disjoint subsets of $\mathbb{N}$ as follows:

$$[1, Nn] \cap \mathbb{N} = \bigcup_{i=1}^{n} [(i-1)N + 1, iN] \cap \mathbb{N} = \bigcup_{i=1}^{n} \bigcup_{\ell=1}^{k_N+1} I_\ell^i \cup J_\ell^i,$$

where, for $i \in \{1, \dots, n\}$ and $\ell \in \{1, \dots, k_N\}$,

$$\begin{aligned} I_\ell^i &:= \Big[(i-1)N + (\ell-1)(p+3m) + 1\,,\ (i-1)N + (\ell-1)(p+3m) + p\Big] \cap \mathbb{N}, \\ J_\ell^i &:= \Big[(i-1)N + (\ell-1)(p+3m) + p + 1\,,\ (i-1)N + \ell(p+3m)\Big] \cap \mathbb{N}, \end{aligned}$$

and, for $\ell = k_N + 1$, $I_{k_N+1}^i = \emptyset$ and

$$J_{k_N+1}^i = \Big[(i-1)N + k_N(p+3m) + 1\,,\ iN\Big] \cap \mathbb{N}.$$

Note that for all $i \in \{1, \dots, n\}$, $J_{k_N+1}^i = \emptyset$ if $k_N(p+3m) = N$.

Now, let M be a fixed positive number not depending on (n, m) and let φ_M be the function defined by $\varphi_M(x) = (x \wedge M) \vee (-M)$. Setting

$$B_{i,\ell} = (\varepsilon_{(i-1)N+(\ell-1)(p+3m)+1-m}, \dots, \varepsilon_{(i-1)N+(\ell-1)(p+3m)+p+m}), \tag{4.31}$$

we define the sequences $(\widetilde{X}_{k,m,M})_{k\geq 1}$ and $(\bar{X}_{k,m,M})_{k\geq 1}$ as follows:

$$\widetilde{X}_{k,m,M} = \begin{cases} \mathbb{E}(\varphi_M(X_k)|B_{i,\ell}) & \text{if} \quad k \in I_\ell^i \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\bar{X}_{k,m,M} = \widetilde{X}_{k,m,M} - \mathbb{E}(\widetilde{X}_{k,m,M}). \tag{4.32}$$

To soothe the notations, we shall write $\widetilde{X}_{k,m}$ and $\bar{X}_{k,m}$ instead of $\widetilde{X}_{k,m,M}$ and $\bar{X}_{k,m,M}$ respectively. Note that for any $k \geq 1$,

$$\|\bar{X}_{k,m}\|_2 \leq 2\|\widetilde{X}_{k,m}\|_2 = 2\|\mathbb{E}(\varphi_M(X_k)|B_{i,\ell})\|_2 \tag{4.33}$$
$$\leq 2\|\varphi_M(X_k)\|_2 \leq 2\|X_k\|_2 = 2\|X_0\|_2, \tag{4.34}$$

and

$$\|\bar{X}_{k,m}\|_\infty \leq 2\|\widetilde{X}_{k,m}\|_\infty \leq 2M, \tag{4.35}$$

where the last equality in (4.33) follows from the stationarity of $(X_k)_k$. As $\bar{X}_{k,m}$ is $\sigma(B_{i,\ell})$-measurable then it can be written as a measurable function h_k of $B_{i,\ell}$, i.e.

$$\bar{X}_{k,m} = h_k(B_{i,\ell}). \tag{4.36}$$

Finally, let $\bar{\mathcal{X}}_{n,m} = ((\bar{\mathcal{X}}_{n,m})_{i,j}) = (\bar{X}_{(j-1)N+i,\,m})$ and set

$$\bar{\mathbf{B}}_{n,m} = \frac{1}{n}\bar{\mathcal{X}}_{n,m}\bar{\mathcal{X}}_{n,m}^T. \tag{4.37}$$

We shall approximate $\mathbf{B}_n$ by $\bar{\mathbf{B}}_{n,m}$ by applying the following proposition:

Proposition 4.7. *Let $\mathbf{B}_n$ and $\bar{\mathbf{B}}_{n,m}$ be the matrices defined in (4.3) and (4.37) respectively then if $\lim_{n\to\infty} N/n = c \in (0,\infty)$, we have for any $z \in \mathbb{C}^+$,*

$$\lim_{m\to+\infty} \limsup_{M\to+\infty} \limsup_{n\to+\infty} \Big|\mathbb{E}\Big(S_{\mathbf{B}_n}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big)\Big| = 0. \tag{4.38}$$

Proof. By Lemma A.1and Cauchy-Schwarz's inequality, it follows that

$$\begin{aligned}&\Big|\mathbb{E}\Big(S_{\mathbf{B}_n}(z)\Big)-\mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big)\Big|\\&\qquad\leq\frac{\sqrt{2}}{v^2}\left\|\frac{1}{N}\mathrm{Tr}(\mathbf{B}_n+\bar{\mathbf{B}}_{n,m})\right\|_1^{1/2}\left\|\frac{1}{Nn}\mathrm{Tr}(\mathcal{X}_n-\bar{\mathcal{X}}_{n,m})(\mathcal{X}_n-\bar{\mathcal{X}}_{n,m})^T\right\|_1^{1/2}.\end{aligned}\tag{4.39}$$

By the definition of $\mathbf{B}_n$, $N^{-1}\mathbb{E}\Big|\mathrm{Tr}(\mathbf{B}_n)\Big| = \|X_0\|_2^2$. Similarly and due to the fact that $pk_N \leq N$ and (4.33),

$$\frac{1}{N}\mathbb{E}\Big|\mathrm{Tr}(\bar{\mathbf{B}}_{n,m})\Big| = \frac{1}{Nn}\sum_{i=1}^{n}\sum_{\ell=1}^{k_N}\sum_{k\in I_\ell^i}\|\bar{X}_{k,m}\|_2^2 \leq 4\|X_0\|_2^2\,.\tag{4.40}$$

Moreover, by the construction of $\mathcal{X}_n$ and $\bar{\mathcal{X}}_{n,m}$, we have

$$\begin{aligned}&\frac{1}{Nn}\mathbb{E}|\mathrm{Tr}(\mathcal{X}_n-\bar{\mathcal{X}}_{n,m})(\mathcal{X}_n-\bar{\mathcal{X}}_{n,m})^T|\\&\qquad=\frac{1}{Nn}\sum_{i=1}^{n}\sum_{\ell=1}^{k_N}\sum_{k\in I_\ell^i}\|X_k-\bar{X}_{k,m}\|_2^2+\frac{1}{Nn}\sum_{i=1}^{n}\sum_{\ell=1}^{k_N+1}\sum_{k\in J_\ell^i}\|X_k\|_2^2\,.\end{aligned}$$

Now, since X_k is centered, we write for $k \in I_\ell^i$,

$$\begin{aligned}\|X_k-\bar{X}_{k,m}\|_2&=\|X_k-\widetilde{X}_{k,m}-\mathbb{E}(X_k-\widetilde{X}_{k,m})\|_2\leq 2\|X_k-\widetilde{X}_{k,m}\|_2\\&\leq 2\|X_k-\mathbb{E}(X_k|B_{i,\ell})\|_2+2\|\widetilde{X}_{k,m}-\mathbb{E}(X_k|B_{i,\ell})\|_2\,.\end{aligned}\tag{4.41}$$

Analyzing the second term of the last inequality, we get

$$\begin{aligned}\|\widetilde{X}_{k,m}-\mathbb{E}(X_k|B_{i,\ell})\|_2&=\|\mathbb{E}(X_k-\varphi_M(X_k)|B_{i,\ell})\|_2\\&\leq\|X_k-\varphi_M(X_k)\|_2\\&=\|(|X_0|-M)_+\|_2.\end{aligned}\tag{4.42}$$

As X_0 belongs to $\mathbb{L}^2$, then $\lim_{M\to+\infty}\|(|X_0|-M)_+\|_2=0$. Now, we note that for $k\in I_\ell^i$,

$\sigma(\varepsilon_{k-m},\dots,\varepsilon_{k+m}) \subset \sigma(B_{i,\ell})$ which implies that

$$\begin{aligned}\|X_k - \mathbb{E}(X_k|B_{i,\ell})\|_2 &\leq \|X_k - \mathbb{E}(X_k|\varepsilon_{k-m},\dots,\varepsilon_{k+m})\|_2 \\ &= \|X_0 - \mathbb{E}(X_0|\varepsilon_{-m},\dots,\varepsilon_m)\|_2 \\ &= \|X_0 - X_{0,m}\|_2 \,, \end{aligned} \tag{4.43}$$

where the first equality is due to the stationarity. Therefore, by (4.42), (4.43), the fact that $pk_N \leq N$ and

$$\mathrm{Card}\Big(\bigcup_{i=1}^{n}\bigcup_{\ell=1}^{k_N+1} J_\ell^i\Big) \leq Nn - npk_N \,,$$

we infer that

$$\begin{aligned}&\frac{1}{Nn}\mathbb{E}|\mathrm{Tr}(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})(\mathcal{X}_n - \bar{\mathcal{X}}_{n,m})^T| \\ &\qquad \leq 8\|X_0 - X_{0,m}\|_2^2 + 8\|(|X_0| - M)_+\|_2^2 + (3(a_m+3)^{-1} + a_m m N^{-1})\|X_0\|_2^2.\end{aligned}$$

Thus starting from (4.39), considering the above upper bounds, we derive that there exists a positive constant C not depending on (n,m,M) such that

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\Big|\mathbb{E}\Big(S_{\mathbf{B}_n}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big)\Big| \leq \frac{C}{v^2}\Big(\|X_0 - X_{0,m}\|_2^2 + \frac{3}{a_m}\Big)^{1/2}.$$

Taking the limit on m, Proposition 4.7 follows by applying the martingale convergence theorem in $\mathbb{L}^2$ and that fact that a_m converges to infinity. □

4.4.2 Approximation by a Gram matrix with independent blocks

By Berbee's classical coupling lemma [11], one can construct by induction a sequence of random variables $(\varepsilon_k^*)_{k\geq 1}$ such that:

- For any $1 \leq i \leq n$ and $1 \leq \ell \leq k_N$,

$$B_{i,\ell}^* = (\varepsilon^*_{(i-1)N+(\ell-1)(p+3m)+1-m},\dots,\varepsilon^*_{(i-1)N+(\ell-1)(p+3m)+p+m})$$

  has the same distribution as $B_{i,\ell}$ defined in (4.31).

- The array $(B^*_{i,\ell})_{1\le i\le n,\,1\le \ell\le k_N}$ is i.i.d.
- For any $1 \le i \le n$ and $1 \le \ell \le k_N$, $\mathbb{P}(B_{i,\ell} \neq B^*_{i,\ell}) \le \beta_m$.

(see page 484 of [71] for more details concerning the construction of the array $(B^*_{i,\ell})_{i,\ell\ge 1}$).
We define now the sequence $(\bar X^*_{k,m})_{k\ge 1}$ as follows:

$$\bar X^*_{k,m} = h_k(B^*_{i,\ell}) \quad \text{if } k \in I^i_\ell , \tag{4.44}$$

where the functions h_k are defined in (4.36). We construct the $N \times n$ random matrix $\bar{\mathcal{X}}^*_{n,m} = ((\bar{\mathcal{X}}^*_{n,m})_{i,j}) = (\bar X^*_{(j-1)N+i,\,m})$.
Note that the block of entries $(\bar X^*_{k,m},\ k \in I^i_\ell)$ is independent of $(\bar X^*_{k,m},\ k \in I^{i'}_{\ell'})$ if $(i,\ell) \neq (i',\ell')$. Thus, $\bar{\mathcal{X}}^*_{n,m}$ has independent blocks of dimension p separated by null blocks whose dimension is at least $3m$. Setting

$$\bar{\mathbf{B}}^*_{n,m} := \frac{1}{n}\bar{\mathcal{X}}^*_{n,m}\bar{\mathcal{X}}^{*\,T}_{n,m} , \tag{4.45}$$

we approximate $\bar{\mathbf{B}}_{n,m}$ by the Gram matrix $\bar{\mathbf{B}}^*_{n,m}$ as shown in the following proposition.

Proposition 4.8. *Let $\bar{\mathbf{B}}_{n,m}$ and $\bar{\mathbf{B}}^*_{n,m}$ be defined in (4.37) and (4.45) respectively. Assuming that $\lim_{n\to\infty} N/n = c \in (0,\infty)$ and $\lim_{n\to\infty}\beta_n = 0$ then for any $z \in \mathbb{C}^+$,*

$$\lim_{m\to+\infty}\limsup_{M\to+\infty}\limsup_{n\to+\infty}\left|\mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big)\right| = 0. \tag{4.46}$$

Proof. By Lemma A.1 and Cauchy-Schwarz's inequality, it follows that

$$\begin{aligned}&\left|\mathbb{E}\Big(S_{\bar{\mathbf{B}}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big)\right| \\ &\qquad \le \frac{\sqrt{2}}{v^2}\left\|\frac{1}{N}\mathrm{Tr}(\bar{\mathbf{B}}_{n,m}+\bar{\mathbf{B}}^*_{n,m})\right\|_1^{1/2}\left\|\frac{1}{Nn}\mathrm{Tr}(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})^T\right\|_1^{1/2}\end{aligned} \tag{4.47}$$

Notice that

$$\|\bar X^*_{k,m}\|_2 = \|h_k(B^*_{i,\ell})\|_2 = \|h_k(B_{i,\ell})\|_2 = \|\bar X_{k,m}\|_2 \le 2\|X_0\|_2 ,$$

where the second equality follows from the fact that $B^*_{i,\ell}$ is distributed as $B_{i,\ell}$ whereas the last inequality follows from (4.33). Thus, we get from the definition of $\bar{\mathbf{B}}^*_{n,m}$ and the fact that $pk_N \le N$,

$$\frac{1}{N}\mathbb{E}\Big|\mathrm{Tr}(\bar{\mathbf{B}}^*_{n,m})\Big| = \frac{1}{Nn}\sum_{i=1}^{n}\sum_{\ell=1}^{k_N}\sum_{k\in I^i_\ell}\|\bar{X}^*_{k,m}\|_2^2 \le 4\|X_0\|_2^2. \tag{4.48}$$

Considering (4.47), (4.40) and (4.48), we infer that Proposition 4.8 follows once we prove that

$$\lim_{m\to+\infty}\limsup_{M\to+\infty}\limsup_{n\to+\infty}\frac{1}{Nn}\mathbb{E}\Big|\mathrm{Tr}(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})^T\Big| = 0\,. \tag{4.49}$$

By the construction of $\bar{\mathcal{X}}_{n,m}$ and $\bar{\mathcal{X}}^*_{n,m}$, we write

$$\frac{1}{Nn}\mathbb{E}\Big|\mathrm{Tr}(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})(\bar{\mathcal{X}}_{n,m}-\bar{\mathcal{X}}^*_{n,m})^T\Big| = \frac{1}{Nn}\sum_{i=1}^{n}\sum_{\ell=1}^{k_N}\sum_{k\in I^i_\ell}\|\bar{X}_{k,m}-\bar{X}^*_{k,m}\|_2^2. \tag{4.50}$$

Now, let L be a fixed positive real number strictly less than M and not depending on (n,m,M). To control the term $\|\bar{X}_{k,m}-\bar{X}^*_{k,m}\|_2^2$, we write for $k \in I^i_\ell$,

$$\begin{aligned}
\|\bar{X}_{k,m}-\bar{X}^*_{k,m}\|_2^2 &= \|(h_k(B_{i,\ell})-h_k(B^*_{i,\ell}))\mathbf{1}_{B_{i,\ell}\neq B^*_{i,\ell}}\|_2^2\\
&\le 4\|h_k(B_{i,\ell})\mathbf{1}_{B_{i,\ell}\neq B^*_{i,\ell}}\|_2^2 = 4\|\bar{X}_{k,m}\mathbf{1}_{B_{i,\ell}\neq B^*_{i,\ell}}\|_2^2\\
&\le 12\|\bar{X}_{k,m}-\mathbb{E}(X_k|B_{i,\ell})\|_2^2 + 12\|\mathbb{E}(X_k|B_{i,\ell})-\mathbb{E}(\varphi_L(X_k)|B_{i,\ell})\|_2^2\\
&\qquad + 12\|\mathbb{E}(\varphi_L(X_k)|B_{i,\ell})\mathbf{1}_{B_{i,\ell}\neq B^*_{i,\ell}}\|_2^2\,.
\end{aligned}$$

Since $\mathbb{P}(B_{i,\ell}\neq B^*_{i,\ell}) \le \beta_m$ and $\varphi_L(X_k)$ is bounded by L, we get

$$\|\mathbb{E}(\varphi_L(X_k)|B_{i,\ell})\mathbf{1}_{B_{i,\ell}\neq B^*_{i,\ell}}\|_2^2 \le L^2\beta_m\,.$$

Moreover, it follows from the fact that X_k is centered and (4.42) that

$$\|\bar{X}_{k,m}-\mathbb{E}(X_k|B_{i,\ell})\|_2^2 \le 4\|\widetilde{X}_{k,m}-\mathbb{E}(X_k|B_{i,\ell})\|_2^2 \le 4\|(|X_0|-M)_+\|_2^2$$

and

$$\|\mathbb{E}(X_k|B_{i,\ell})-\mathbb{E}(\varphi_L(X_k)|B_{i,\ell})\|_2^2 \le \|X_k-\varphi_L(X_k)\|_2^2 = \|(|X_0|-L)_+\|_2^2.$$

Hence gathering the above upper bounds we get

$$\|\bar{X}_{k,m} - \bar{X}^*_{k,m}\|_2^2 \leq 48\|(|X_0| - M)_+\|_2^2 + 12\|(|X_0| - L)_+\|_2^2 + 12L^2\beta_m. \tag{4.51}$$

As $pk_N \leq N$, the right-hand-side of (4.50) converges to zero by letting first M, then m and finally L tend to infinity. Therefore, (4.49) and thus the proposition follow. □

4.4.3 Approximation with a Gaussian matrix

In order to complete the proof of the theorem, it suffices, in view of (4.38) and (4.46), to prove the following convergence: for any $z \in \mathbb{C}^+$,

$$\lim_{m\to+\infty} \limsup_{n\to+\infty} \Big|\mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big)\Big| = 0. \tag{4.52}$$

With this aim, we shall first consider a sequence $(Z_{k,m})_{k\in\mathbb{Z}}$ such that for any $k, \ell \in \{1, \ldots, N\}$,

$$\mathrm{Cov}(Z_{k,m}\,, Z_{\ell,m}) = \mathrm{Cov}(\bar{X}^*_{k,m}\,, \bar{X}^*_{\ell,m}) \tag{4.53}$$

and let $(Z^{(i)}_{k,m})_k$, $i = 1, \ldots, n$, be n independent copies of $(Z_{k,m})_k$. We then define the $N \times n$ matrix $\mathcal{Z}_{n,m} = ((\mathcal{Z}_{n,m})_{u,v}) = (Z^{(v)}_{u,m})$ and finally set

$$\mathbf{G}_{n,m} = \frac{1}{n}\mathcal{Z}_{n,m}\mathcal{Z}^T_{n,m}\,. \tag{4.54}$$

Now, we shall construct a matrix $\widetilde{\mathcal{Z}}_{n,m}$ having the same block structure as the matrix $\bar{\mathcal{X}}^*_{n,m}$. With this aim, we let for $\ell = 1, \ldots, k_N$,

$$I_\ell = \{(\ell-1)(p+3m)+1, \ldots, (\ell-1)(p+3m)+p\}$$

and let $\widetilde{Z}^{(i)}_{k,m}$ be defined for any $1 \leq i \leq n$ and $1 \leq k \leq N$ by

$$\widetilde{Z}^{(i)}_{k,m} = \begin{cases} Z^{(i)}_{k,m} & \text{if} \quad k \in I_\ell \ \text{ for some } \ell \in \{1, \ldots, k_N\} \\ 0 & \text{otherwise} \ . \end{cases}$$

We define now the $N \times n$ matrix $\widetilde{\mathcal{Z}}_{n,m} = ((\widetilde{\mathcal{Z}}_{n,m})_{u,v}) = (\widetilde{Z}^{(v)}_{u,m})$ and we note that $\widetilde{\mathcal{Z}}_{n,m}$, as $\bar{\mathcal{X}}^*_{n,m}$, consists of independent blocks of dimension p separated by null blocks whose dimension is at least $3m$. We finally set

$$\widetilde{\mathbf{G}}_{n,m} = \frac{1}{n}\widetilde{\mathcal{Z}}_{n,m}\widetilde{\mathcal{Z}}^T_{n,m}\,. \tag{4.55}$$

Provided that $\lim_{n\to\infty} n/N = c \in (0,\infty)$, we have by Proposition 4.2 in [8] that for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty}\limsup_{n\to\infty}\left|\mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbf{G}}_{n,m}}(z)\Big)\right| = 0.$$

In order to prove (4.52), we shall prove for any $z \in \mathbb{C}^+$,

$$\lim_{n\to+\infty}\left|\mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbf{G}}_{n,m}}(z)\Big)\right| = 0$$

and then

$$\lim_{m\to\infty}\limsup_{M\to\infty}\limsup_{n\to\infty}\left|\mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big)\right| = 0\,.$$

The technique followed to prove the first convergence is based on Lindeberg's method by blocks, whereas, for the second, it is based on the Gaussian interpolation technique.

Proposition 4.9. *Provided that $N/n \to c \in (0,\infty)$, then for any $z = x+iy \in \mathbb{C}^+$,*

$$\lim_{n\to+\infty}\left|\mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbf{G}}_{n,m}}(z)\Big)\right| = 0. \tag{4.56}$$

We don't give a proof of this proposition because it is the same as that of Proposition 2.8. Indeed, $\bar{\mathcal{X}}^*_{n,m}$ has a similar structure as the matrix $\bar{\mathcal{X}}_{n,m}$ considered in the Chapter 2 and consists as well of independent blocks separated by blocks of zero entries. Therefore, following the lines of Proposition 2.8 gives for any $z = x+iy \in \mathbb{C}_+$,

$$|\mathbb{E}\Big(S_{\bar{\mathbf{B}}^*_{n,m}}(z)\Big) - \mathbb{E}\Big(S_{\widetilde{\mathbf{G}}_{n,m}}(z)\Big)| \le \frac{Cp^2(1+M^3)N^{1/2}}{y^3(1\wedge y)n}$$

which converges to 0 as n tends to infinity since $\lim_{n\to\infty} N/n = c \in (0,\infty)$.

□

To end the proof of Theorem 4.2, it remains to prove the following convergence.

Proposition 4.10. *Provided that* $\lim_{n\to\infty} N/n = c \in (0,\infty)$ *and* $\lim_{n\to\infty} \beta_n = 0$ *then for any* $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \Big| \mathbb{E}\Big(S_{\mathbf{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbf{G}_{n,m}}(z)\Big) \Big| = 0.$$

Proof. Let $n' = N + n$ and $\mathbb{G}_{n'}$ and $\mathbb{G}_{n',m}$ be the symmetric matrices of order n' defined by

$$\mathbb{G}_{n'} = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0}_{n,n} & \mathcal{Z}_n^T \\ \mathcal{Z}_n & \mathbf{0}_{N,N} \end{pmatrix} \quad \text{and} \quad \mathbb{G}_{n',m} = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0}_{n,n} & \mathcal{Z}_{n,m}^T \\ \mathcal{Z}_{n,m} & \mathbf{0}_{N,N} \end{pmatrix} .$$

We recall that since $n'/N \to 1 + c^{-1}$ then it is equivalent to prove for any $z \in \mathbb{C}^+$,

$$\lim_{m\to\infty} \limsup_{M\to\infty} \limsup_{n\to\infty} \Big| \mathbb{E}\Big(S_{\mathbb{G}_{n'}}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{G}_{n',m}}(z)\Big) \Big| = 0\,.$$

Following the lines of Proposition 2.13, we infer that for any $z \in \mathbb{C}^+$,

$$\Big| \mathbb{E}(S_{\mathbb{G}_{n'}}(z)) - \mathbb{E}(S_{\mathbb{G}_{n',m}}(z)) \Big|^2 \le \frac{C}{n^2} \sum_{k=1}^N \Big(\|X_k\|_2^2 + \|\bar{X}^*_{k,m}\|_2^2 \Big) \sum_{\ell=1}^N \|X_\ell - \bar{X}^*_{\ell,m}\|_2^2\,.$$

In view of (4.41), (4.42), (4.43) and (4.51), we get

$$\begin{aligned} \|X_\ell - \bar{X}^*_{\ell,m}\|^2 &\le 2\|X_\ell - \bar{X}_{\ell,m}\|^2 + 2\|\bar{X}_{\ell,m} - \bar{X}^*_{\ell,m}\|^2 \\ &\le 16\|X_0 - X_{0,m}\|_2^2 + 112\|(|X_0| - M)_+\|_2^2 \\ &\qquad + 24\|(|X_0| - L)_+\|_2^2 + 24L^2\beta_m\,, \end{aligned}$$

where L is a fixed positive real number strictly less than M and not depending on (n, m, M). Taking into account the stationarity of $(X_k)_k$ and (4.33), we then infer that

$$\begin{aligned} \Big| \mathbb{E}(S_{\mathbb{G}_{n'}}(z)) - \mathbb{E}(S_{\mathbb{G}_{n',m}}(z)) \Big|^2 \le \frac{CN^2}{n^2} \Big(&\|X_0 - X_{0,m}\|_2^2 + \|(|X_0| - M)_+\|_2^2 \\ &+ \|(|X_0| - L)_+\|_2^2 + L^2\beta_m \Big)\,, \end{aligned}$$

which converges to zero by letting first n, then M followed by m and finally L tend to infinity. This ends the proof of the proposition.

$\square$

Chapter 5

Bernstein Type Inequality for Dependent Matrices

In this chapter, we prove a Bernstein type inequality for the sum of self-adjoint centered and geometrically absolutely regular random matrices with bounded largest eigenvalue.

This inequality can be viewed as an extension to the matrix setting of the Bernstein-type inequality obtained by Merlevède et al. [46] in the context of real-valued bounded random variables that are geometrically absolutely regular. It can be also viewed as a generalization, up to a logarithmic term, of the Bernstein-type inequality obtained by Tropp [70] from *independent* to geometrically absolutely regular matrices.

This chapter shall be organized as follows: we shall start by describing the matrices and the dependence structure and then announce our Bernstein-type inequality (Theorem 5.1). A subsection shall be also devoted to some examples of matrix models where this Bernstein-type inequality applies.

Moreover, Section A.4 of the Appendix will be devoted to a brief introduction on some operator functions in addition to some tools and lemmas that will be useful to the proof given in Section 5.3.

5.1 A Bernstein-type inequality for geometrically β-mixing matrices

For any $d \times d$ matrix $\mathbb{X} = [(\mathbb{X})_{i,j}]_{i,j=1}^d$ whose entries belong to $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, we associate its corresponding vector $\mathbf{X}$ in $\mathbb{K}^{d^2}$ whose coordinates are the entries of $\mathbb{X}$ i.e.

$$\mathbf{X} = \Big((\mathbb{X})_{i,j} \, , 1 \leq i \leq d \Big)_{1 \leq j \leq d} \, .$$

Therefore $\mathbf{X} = \Big(X_i \, , 1 \leq i \leq d^2 \Big)$ where

$$X_i = (\mathbb{X})_{i-(j-1)d,j} \ \text{ for } (j-1)d + 1 \leq i \leq jd \, ,$$

and $\mathbf{X}$ will be called the vector associated with $\mathbb{X}$. Reciprocally, given $\mathbf{X} = \Big(X_\ell \, , 1 \leq \ell \leq d^2 \Big)$ in $\mathbb{K}^{d^2}$ we shall associate a $d \times d$ matrix $\mathbb{X}$ by setting

$$\mathbb{X} = \Big[(\mathbb{X})_{i,j} \Big]_{i,j=1}^n \text{ where } (\mathbb{X})_{i,j} = X_{i+(j-1)d} \, .$$

The matrix $\mathbb{X}$ will be referred to as the matrix associated with $\mathbf{X}$.

All along this chapter, we consider a family $(\mathbb{X}_i)_{i \geq 1}$ of $d \times d$ self-adjoint random matrices whose entries are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in $\mathbb{K}$, and that are geometrically absolutely regular in the following sense. Let

$$\beta_0 = 1 \ \text{ and } \ \beta_k = \sup_{j \geq 1} \beta(\sigma(\mathbf{X}_i \, , \, i \leq j), \sigma(\mathbf{X}_i \, , \, i \geq j+k)) \, , \text{ for any } k \geq 1, \tag{5.1}$$

where $\beta(\mathcal{A}, \mathcal{B})$ is the absolute regular coefficient introduced in Definition 1.6. The β_k's are usually called the coefficients of absolute regularity of the sequence of vectors $(\mathbf{X}_i)_{i \geq 1}$ and we shall assume in this chapter that they decrease geometrically in the sense that there exists $c > 0$ such that for any integer $k \geq 1$,

$$\beta_k = \sup_{j \geq 1} \beta(\sigma(\mathbf{X}_i \, , \, i \leq j), \sigma(\mathbf{X}_i \, , \, i \geq j+k)) \leq \mathrm{e}^{-c(k-1)} \, . \tag{5.2}$$

We shall assume that the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is rich enough to contain a sequence $(\epsilon_i)_{i \in \mathbb{Z}} = (\delta_i, \eta_i)_{i \in \mathbb{Z}}$ of i.i.d. random variables with uniform distribution over $[0,1]^2$, independent of $(\mathbf{X}_i)_{i\geq 0}$.

Our aim is to give a probabilistic bound for $\lambda_{\max}(\sum_{i=1}^n \mathbb{X}_i)$ with $(\mathbb{X}_i)_i$ being a sequence of dependent centered self-adjoint random matrices with a uniformly bounded largest eigenvalue and satisfying (5.2). With this aim, we use the matrix Chernoff bound (1.12) and control the matrix Laplace transform

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^n \mathbb{X}_i\Big)$$

to get the following Bernstein type inequality. Let

$$\gamma(c,n) = \frac{\log n}{\log 2}\max\Big(2, \frac{32\log n}{c\log 2}\Big). \tag{5.3}$$

and

$$v^2 = \sup_{K\subseteq\{1,\dots,n\}} \frac{1}{\mathrm{Card}K}\lambda_{\max}\Big(\mathbb{E}\Big(\sum_{i\in K}\mathbb{X}_i\Big)^2\Big) \tag{5.4}$$

with the suprimum being taken over all the non-empty subsets $K\subseteq\{1,\dots,n\}$.

Theorem 5.1. *Let $(\mathbb{X}_i)_{i\geq 1}$ be a family of self-adjoint random matrices of size d. Assume that (5.2) holds and that there exists a positive constant M such that for any $i\geq 1$,*

$$\mathbb{E}(\mathbb{X}_i) = \mathbf{0} \quad \textit{and} \quad \lambda_{\max}(\mathbb{X}_i)\leq M \quad \textit{almost surely.} \tag{5.5}$$

Then for any t such that $0 < tM < 1/\gamma(c,n)$, we have

$$\log\mathbb{E}\mathrm{Tr}\Big(\exp\Big(t\sum_{i=1}^n\mathbb{X}_i\Big)\Big) \leq \log d + \frac{t^2 n\big(15v + 2M/\sqrt{cn}\big)^2}{1 - tM\gamma(c,n)}, \tag{5.6}$$

where $\gamma(c,n)$ and v are defined in (5.3) and (5.4). In terms of probabilities, there exists a universal positive constant C such that for any $x>0$ and any integer $n\geq 2$,

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^n\mathbb{X}_i\Big)\geq x\Big) \leq d\exp\Big(-\frac{Cx^2}{v^2 n + c^{-1}M^2 + xM\gamma(c,n)}\Big).$$

The bound (5.6) on the matrix log-Laplace allows us to control the expectation of the maximum eigenvalue:

Corollary 5.2. *Under the conditions of Theorem 5.1,*

$$\mathbb{E}\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) \leq 30v\sqrt{n\log d} + 4Mc^{-1/2}\sqrt{\log d} + M\gamma(c,n)\log d$$

where $\gamma(c,n)$ and v are defined in (5.3) and (5.4) respectively.

Proof. We proceed as in Section 4.2.4 of [43] and get for any $0 < t < \frac{1}{M\gamma(c,n)}$

$$\begin{aligned}\mathbb{E}\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) &= \frac{1}{t}\mathbb{E}\log \mathrm{e}^{\lambda_{\max}\left(t\sum_{i=1}^{n}\mathbb{X}_i\right)} = \frac{1}{t}\mathbb{E}\log\lambda_{\max}\Big(\mathrm{e}^{t\sum_{i=1}^{n}\mathbb{X}_i}\Big)\\ &\leq \frac{1}{t}\mathbb{E}\log\operatorname{Tr}\exp\Big(t\sum_{i=1}^{n}\mathbb{X}_i\Big)\\ &\leq \frac{1}{t}\log\mathbb{E}\operatorname{Tr}\exp\Big(t\sum_{i=1}^{n}\mathbb{X}_i\Big).\end{aligned}$$

The first inequality follows from the fact that the matrix exponential is positive definite and thus its largest eigenvalue is bounded by the trace, whereas the second one follows from Jensen's inequality and the concavity of the matrix logarithm. Considering the bound on the log-Laplace transform in (5.6) and taking the infimum on t, we get

$$\mathbb{E}\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) \leq \inf_{0<tM<1/\gamma(c,n)}\frac{1}{t}\Big(\log d + \frac{t^2 n(15v + 2M/\sqrt{cn})^2}{1 - tM\gamma(c,n)}\Big).$$

Thanks to a personal communication with professor W. Bryc, we notice that the problem can be reduced to

$$\inf_{0<t<1}\Big(\frac{b\log d}{t} + \frac{at}{b(1-t)}\Big)$$

where $a = n(15v + 2M/\sqrt{cn})^2$ and $b = M\gamma(c,n)$. The latter function is convex on $t \in (0,1)$ and thus the minimum is attained at the unique zero of the derivative giving the desired upper bound. ☐

In the following section, we give examples of matrices for which Theorem 5.1 can be applied.

5.2 Applications

Let $(\tau_k)_k$ be a stationary sequence of real-valued random variables such that $\|\tau_1\|_\infty \le 1$ a.s. Consider a family $(\mathbb{Y}_k)_k$ of $d \times d$ independent self-adjoint random matrices which is independent of $(\tau_k)_k$. For any $i = 1, \dots, n$, let $\mathbb{X}_i = \tau_i \mathbb{Y}_i$ and note that in this case

$$\beta_k = \beta(\sigma(\tau_i\,,\, i \le 0), \sigma(\tau_i\,,\, i \ge k))\,.$$

Corollary 5.3. *Assume that there exists a positive constant c such that $\beta_k \le e^{-c(k-1)}$ for any $k \ge 1$ and suppose that each random matrix $\mathbb{Y}_k$ satisfies*

$$\mathbb{E}\mathbb{Y}_k = \mathbf{0} \quad , \quad \lambda_{\max}(\mathbb{Y}_k) \le M \quad \textit{and} \quad \lambda_{\min}(\mathbb{Y}_k) \ge -M \quad \textit{almost surely.}$$

Then for any $t > 0$ and any integer $n \ge 2$,

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{k=1}^n \tau_k \mathbb{Y}_k\Big) \ge t\Big) \le d \exp\Big(- \frac{Ct^2}{nM^2\mathbb{E}(\tau_0^2) + M^2 + tM(\log n)^2}\Big)\,,$$

where C is a positive constant depending only on c.

Proof. The above corollary follows by noting that for any $K \subseteq \{1, \dots, n\}$

$$\Sigma_K := \mathbb{E}\Big(\sum_{k\in K} \tau_k \mathbb{Y}_k\Big)^2 = \sum_{k \in K} \mathbb{E}(\tau_k^2)\mathbb{E}(\mathbb{Y}_k^2) = \mathbb{E}(\tau_0^2) \sum_{k\in K} \mathbb{E}(\mathbb{Y}_k^2),$$

which, by Weyl's inequality, implies that $\lambda_{\max}(\Sigma_K) \le M^2 \mathrm{Card}(K)\mathbb{E}(\tau_0^2)$. Therefore, we infer that $v^2 \le M^2\mathbb{E}(\tau_0^2)$. □

We consider now another model of matrices. Let $(X_k)_{k\in\mathbb{Z}}$ be a geometrically absolutely regular sequence of real-valued centered random variables. That is, there exists a positive constant c_0 such that for any $k \ge 1$,

$$\sup_{\ell\in\mathbb{Z}} \beta\Big(\sigma(X_i\,,\, i \le \ell), \sigma(X_i\,,\, i \ge k+\ell)\Big) \le e^{-c_0(k-1)}\,. \tag{5.7}$$

For any $i = 1, \dots, n$, let $\mathbb{X}_i$ be the $d \times d$ random matrix defined by

$$\mathbb{X}_i = \mathbf{C}_i \mathbf{C}_i^T - \mathbb{E}(\mathbf{C}_i\mathbf{C}_i^T)$$

where $\mathbf{C}_i = (X_{(i-1)d+1}, \ldots, X_{id})^T$. Note that $\sum_{i=1}^n \mathbb{X}_i = \mathbb{X}\mathbb{X}^T - \mathbb{E}(\mathbb{X}\mathbb{X}^T)$, where $\mathbb{X}$ is the $d \times n$ matrix given by

$$\mathbb{X} = \begin{pmatrix} X_1 & X_{d+1} & \cdots & X_{(n-1)d+1} \\ X_2 & X_{d+2} & \cdots & X_{(n-1)d+2} \\ \vdots & \vdots & & \vdots \\ X_d & X_{2d} & \cdots & X_{nd} \end{pmatrix}.$$

We also note that in this case, for any $k \geq 1$,

$$\begin{aligned}\beta_k &= \sup_{\ell \in \mathbb{Z}} \beta\Big(\sigma(\mathbf{C}_i, i \leq \ell), \sigma(\mathbf{C}_i\,,\, i \geq \ell + k)\Big) \\ &= \sup_{\ell \in \mathbb{Z}} \beta\Big(\sigma(X_i, i \leq \ell d), \sigma(X_i\,,\, i \geq (k+\ell-1)d+1)\Big) \\ &\leq e^{-c_0 d(k-1)}\,.\end{aligned}$$

Corollary 5.4. *Assume that $(X_k)_k$ satisfies (5.7). Suppose in addition that there exists a positive constant M satisfying $\sup_k \|X_k\|_\infty \leq M$ a.s. Then, for any $x > 0$ and any integer $n \geq 2$*

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^n \mathbb{X}_i\Big) \geq x\Big) \leq d \exp\Big(-\frac{Cx^2}{ndM^4 + dM^4 + xM^2(d\log n + \log^2 n)}\Big),$$

where C is a positive constant depending only on c_0.

Proof. As $\mathbf{0} \preceq \mathbf{C}_i\mathbf{C}_i^T$ for any $i \in \{1, \ldots, n\}$, it follows by Weyl's inequality that almost surely,

$$\lambda_{\max}(\mathbb{X}_i) \leq \lambda_{\max}(\mathbf{C}_i\mathbf{C}_i^T) \leq \mathrm{Tr}(\mathbf{C}_i\mathbf{C}_i^T)\,.$$

Thus we infer that $\lambda_{\max}(\mathbb{X}_i) \leq dM^2$ a.s. To get the desired result, it remains to control v^2. We have for any $K \subseteq \{1, \ldots, N\}$,

$$\Sigma_K := \mathbb{E}\Big(\sum_{i \in K} \mathbb{X}_i\Big)^2 = \sum_{i,j \in K} \mathrm{Cov}(\mathbf{C}_i\mathbf{C}_i^T, \mathbf{C}_j\mathbf{C}_j^T)$$

and we note that the $(k,\ell)^{th}$ component of Σ_K is

$$(\Sigma_K)_{k,\ell} = \Big[\mathbb{E}\Big(\sum_{i\in K}\mathbb{X}_i\Big)^2\Big]_{k,\ell} = \sum_{i,j\in K}\sum_{s=1}^{d}\mathrm{Cov}\Big(X_{(i-1)d+k}\,X_{(i-1)d+s}\,,X_{(j-1)d+s}\,X_{(j-1)d+\ell}\Big)\,.$$

Therefore we infer by Gerschgorin's theorem that

$$\begin{aligned}\Big|\lambda_{\max}\Big(\Sigma_K\Big)\Big| &\le \sup_k \sum_{\ell=1}^{d}|(\Sigma_K)_{k,\ell}| \\ &\le \sup_k \sum_{i,j\in K}\sum_{\ell=1}^{d}\sum_{s=1}^{d}\Big|\mathrm{Cov}\Big(X_{(i-1)d+k}\,X_{(i-1)d+s}\,,X_{(j-1)d+s}\,X_{(j-1)d+\ell}\Big)\Big|\,.\end{aligned}$$

This is controlled by Ibragimov's covariance inequality:

Lemma 5.5. *([38]) Let X and Y be two almost surely bounded random variables. Then*

$$|\mathrm{Cov}(X,Y)| \le 2\|X\|_\infty\|Y\|_\infty\alpha(\sigma(X),\sigma(Y))\,.$$

As for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, $2\alpha(\mathcal{A},\mathcal{B}) \le \beta(\mathcal{A},\mathcal{B})$ then

$$|\mathrm{Cov}(X,Y)| \le \|X\|_\infty\|Y\|_\infty\beta(\sigma(X),\sigma(Y))\,.$$

After tedious computations involving the above inequality, we infer that $v^2 \le c_1 dM^4$ where c_1 is a positive constant depending only on c_0. Applying Theorem 5.1 with these upper bounds ends the proof. □

5.3 Proof of the Bernstein-type inequality

The proof of Theorem 5.1 being very technical, is divided into several steps. In Section 5.3.1, we give the main ingredient to prove our Bernstein-type inequality, namely: a bound for the Laplace transform of the partial sum, indexed by a suitable Cantor-type set, of the self-adjoint random matrices under consideration (see Proposition 5.6 and Section 5.3.2 for the construction of this suitable Cantor-set). As quoted in the introduction, this key result is based on a decoupling lemma which is stated in Section 5.3.3. The proof of Theorem 5.1 is completed in Section 5.3.5.

5.3.1 A key result

The next proposition is the main ingredient to prove Theorem 5.1. For a given positive integer A, it is based on a suitable construction of a subset K_A of $\{1,\dots,A\}$ for which it is possible to give a good upper bound for the Laplace transform of $\sum_{i\in K_A}\mathbb{X}_i$. Its proof is based on the decoupling Lemma 5.7 below that allows comparing $\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i\in K_A}\mathbb{X}_i\right)$ with the same quantity but replacing $\sum_{i\in K_A}\mathbb{X}_i$ with a sum of *independent* blocks of matrices.

Proposition 5.6. *Let $(\mathbb{X}_i)_{i\geq 1}$ be as in Theorem 5.1. Let A be a positive integer larger than 2. Then there exists a subset K_A of $\{1,\dots,A\}$ with $\mathrm{Card}(K_A)\geq A/2$, such that for any positive t such that $tM\leq \min\left(\frac{1}{2},\frac{c\log 2}{32\log A}\right)$,*

$$\log \mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i\in K_A}\mathbb{X}_i\right)\leq \log d + 4\times 3.1t^2Av^2+\frac{9(tM)^2}{c}\exp\left(-\frac{3c}{32tM}\right), \tag{5.8}$$

where v^2 is defined in (5.4).

The proof of this proposition is divided into several steps. First, we shall construct a Cantor-type set K_A from $\{1,\dots,A\}$ and then we prove a fundamental decoupling lemma of the matrix log-Laplace before giving the full proof of Proposition 5.6.

5.3.2 Construction of a Cantor-like subset K_A

As in [46] and [47], the set K_A will be a finite union of 2^ℓ disjoint sets of consecutive integers with same cardinality spaced according to a recursive 'Cantor"-like construction. Let

$$\delta=\frac{\log 2}{2\log A}\quad\text{and}\quad \ell:=\ell_A=\sup\left\{k\in\mathbb{N}^*\ :\ \frac{A\delta(1-\delta)^{k-1}}{2^k}\geq 2\right\}.$$

Note that $\ell\leq \log A/\log 2$ and $\delta\leq 1/2$ (since $A\geq 2$). Let $n_0=A$ and for any $j\in\{1,\dots,\ell\}$,

$$n_j=\left\lceil\frac{A(1-\delta)^j}{2^j}\right\rceil\quad\text{and}\quad d_{j-1}=n_{j-1}-2n_j\,. \tag{5.9}$$

For any non-negative x, we denote by $\lceil x \rceil$ the smallest integer which is larger or equal to x and we note that $x \leq \lceil x \rceil < x+1$. Note that for any $j \in \{0, \dots, \ell - 1\}$,

$$d_j \geq \frac{A\delta(1-\delta)^j}{2^j} - 2 \geq \frac{A\delta(1-\delta)^j}{2^{j+1}}\,, \tag{5.10}$$

where the last inequality follows from the definition of ℓ. Moreover,

$$n_\ell \leq \frac{A(1-\delta)^\ell}{2^\ell} + 1 \leq \frac{A(1-\delta)^\ell}{2^{\ell-1}}\,, \tag{5.11}$$

where the last inequality follows from the fact that $\dfrac{A\delta(1-\delta)^{\ell-1}}{2^\ell} \times \dfrac{1-\delta}{\delta} \geq 2$ by the definition of ℓ and the fact that $\delta \leq 1/2$.

To construct K_A we proceed as follows. At the first step, we divide the set $\{1, \dots, A\}$ into three disjoint subsets of consecutive integers: $I_{1,1}$, $I^*_{0,1}$ and $I_{1,2}$. These subsets are such that $\mathrm{Card}(I_{1,1}) = \mathrm{Card}(I_{1,2}) = n_1$ and $\mathrm{Card}(I^*_{0,1}) = d_0$. At the second step, we divide $I_{1,1}$ and $I_{1,2}$ into three disjoint subsets of consecutive integers as follows: for $i = 1, 2$,

$$I_{1,i} = I_{2,2i-1} \cup I^*_{1,i} \cup I_{2,2i}$$

where $\mathrm{Card}(I_{2,2i-1}) = \mathrm{Card}(I_{2,2i}) = n_2$ and $\mathrm{Card}(I^*_{1,i}) = d_1$. Iterating this process, we have constructed, after j steps ($1 \leq j \leq \ell_A$), a family $\{I_{j,i}\,, i = 1, \dots, 2^j\}$ of 2^j sets of consecutive integers each of cardinal n_j such that for any $i = 1, \dots, 2^{j-1}$,

$$I_{j-1,i} = I_{j,2i-1} \cup I^*_{j-1,i} \cup I_{j,2i}\,,$$

where $I^*_{j-1,i} = \{b_{j,2i-1} + 1, \dots, a_{j,2i} - 1\}$, $a_{j,i} = \min\{k \in I_{j,i}\}$ and $b_{j,i} = \max\{k \in I_{j,i}\}$.

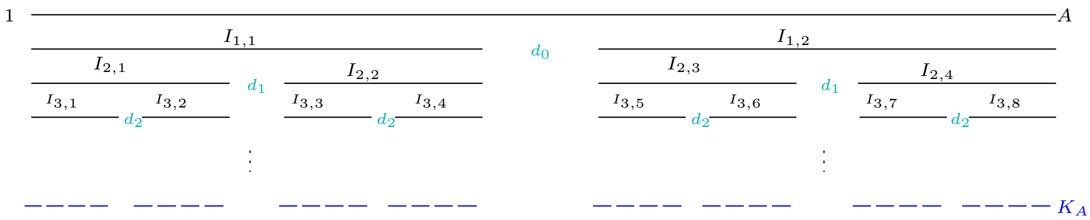

Figure 5.1 – Construction of the Cantor-type set K_A

After ℓ steps we have then constructed 2^ℓ sets of consecutive integers, $I_{\ell,i}$, $i = 1, \dots, 2^\ell$, each of cardinal n_ℓ such that $I_{\ell,2i-1}$ and $I_{\ell,2i}$ are spaced by $d_{\ell-1}$ integers. The set of consecutive integers K_A is then defined by

$$K_A = \bigcup_{k=1}^{2^\ell} I_{\ell,k} \,.$$

Note that

$$\{1, \dots, A\} = K_A \cup (\cup_{j=0}^{\ell-1} \cup_{i=1}^{2^j} I^*_{j,i})$$

Therefore

$$\mathrm{Card}(\{1, \dots, A\} \setminus K_A) = \sum_{j=0}^{\ell-1} \sum_{i=1}^{2^j} \mathrm{Card}(I^*_{j,i}) = \sum_{j=0}^{\ell-1} 2^j d_j = A - 2^\ell n_\ell \,.$$

But

$$A - 2^\ell n_\ell \le A\Big(1 - (1-\delta)^\ell\Big) = A\delta \sum_{j=0}^{\ell-1} (1-\delta)^j \le A\delta\ell \le \frac{A}{2} \,. \tag{5.12}$$

Therefore $A \ge \mathrm{Card}(K_A) \ge A/2$.

In the rest of the proof, the following notation will be also useful: for any $k \in \{0, 1, \dots, \ell\}$ and any $j \in \{1, \dots, 2^k\}$, let

$$K_{k,j} := K_{A,k,j} = \bigcup_{i=(j-1)2^{\ell-k}+1}^{j2^{\ell-k}} I_{\ell,i} \tag{5.13}$$

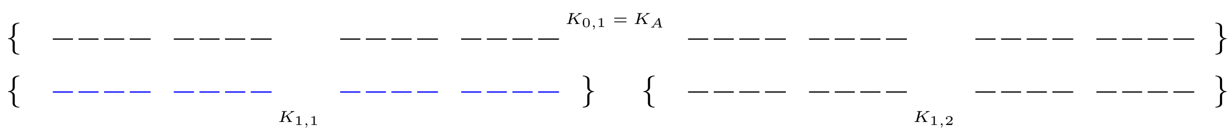

Figure 5.2 – Illustration of $K_{0,1}$, $K_{1,1}$ and $K_{1,2}$

Therefore $K_{0,1} = K_A$ and, for any $j \in \{1, \dots, 2^\ell\}$, $K_{\ell,j} = I_{\ell,j}$. Moreover, for any $k \in \{1, \dots, \ell\}$ and any $j \in \{1, \dots, 2^{k-1}\}$, there are exactly d_{k-1} integers between $K_{k,2j-1}$ and $K_{k,2j}$.

We include the following figure for a better illustration of the notation that we adopt all along the chapter.

5.3.3 A fundamental decoupling lemma

We start by introducing some notations that we adopt all along this chapter, then we state the decoupling Lemma 5.7 below which is fundamental to prove Proposition 5.6.

For any positive integer A, let K_A be defined as in Subsection 5.3.2. For any integer $m \in \{0, \dots, \ell\}$, $(\mathbf{V}_j^{(m)})_{1 \leq j \leq 2^m}$ will denote a family of 2^m *mutually independent* random vectors defined on $(\Omega, \mathcal{A}, \mathbb{P})$, each of dimension $s_{d,\ell,m} := d^2 \mathrm{Card}(K_{m,j}) = d^2 2^{\ell-m} n_\ell$ and such that

$$\mathbf{V}_j^{(m)} =^{\mathcal{D}} (\mathbf{X}_i \,, i \in K_{m,j}) \,. \tag{5.14}$$

The existence of such a family is ensured by the Skorohod lemma [66]. Indeed, since $(\Omega, \mathcal{A}, \mathbb{P})$ is assumed to be large enough to contain a sequence $(\delta_i)_{i \in \mathbb{Z}}$ of i.i.d. random variables uniformly distributed on $[0, 1]$ and independent of the sequence $(\mathbf{X}_i)_{i \geq 0}$, there exist measurable functions f_j such that the vectors $\mathbf{V}_j^{(m)} = f_j\big((\mathbf{X}_i \,, i \in K_{m,k})_{k=1,\dots,j}, \delta_j\big)$, $j = 1, \dots, 2^m$, are independent and satisfy (5.14).

Let $\pi_i^{(m)}$ be the i-th canonical projection from $\mathbb{K}^{s_{d,\ell,m}}$ onto $\mathbb{K}^{d^2}$, namely: for any vector $\mathbf{x} = (\mathbf{x}_i \,, i \in K_{m,j})$ of $\mathbb{K}^{s_{d,\ell,m}}$, $\pi_i^{(m)}(\mathbf{x}) = \mathbf{x}_i$. For any $i \in K_{m,j}$, let

$$\mathbf{X}_j^{(m)}(i) = \pi_i^{(m)}(\mathbf{V}_j^{(m)}) \quad \text{and} \quad \mathbb{S}_j^{(m)} = \sum_{i \in K_{m,j}} \mathbb{X}_j^{(m)}(i) \,, \tag{5.15}$$

where $\mathbb{X}_j^{(m)}(i)$ is the $d \times d$ random matrix associated with $\mathbf{X}_j^{(m)}(i)$. Recall that this means that the (k, ℓ)-th entry of $\mathbb{X}_j^{(m)}(i)$ is the $((\ell-1)d+k)$-th coordinate of the vector $\mathbf{X}_j^{(m)}(i)$.

With the above notations, we have

$$\mathbb{E}\mathrm{Tr} \exp\Big(t \sum_{i \in K_A} \mathbb{X}_i\Big) = \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\,\mathbb{S}_1^{(0)}}\Big) \,. \tag{5.16}$$

We are now in position to state the following lemma which will be a key step in the proof of Proposition 5.6 and allows decoupling when we deal with the Laplace transform of a sum of self adjoint random matrices.

Lemma 5.7. *Assume that* (5.5) *holds. Then for any* $t > 0$ *and any* $k \in \{0, \dots, \ell - 1\}$*,*

$$\mathbb{E}\mathrm{Tr}\exp\left(t \sum_{j=1}^{2^k} \mathbb{S}_j^{(k)}\right) \le \left(1 + \beta_{d_k+1} \mathrm{e}^{tMn_\ell 2^{\ell-k}}\right)^{2^k} \mathbb{E}\mathrm{Tr}\exp\left(t \sum_{j=1}^{2^{k+1}} \mathbb{S}_j^{(k+1)}\right),$$

where $(\mathbb{S}_j^{(k)})_{j=1,\dots,2^k}$ *is the family of mutually independent random matrices defined in* (5.15)*.*

This is a decoupling lemma in the sense that, for a given $k \in \{0, \dots, \ell - 1\}$, we can compare the Laplace transform of the sum of the 2^k independent matrices $\mathbb{S}_j^{(k)}$'s associated with the blocks $K_{k,j}$'s, by that of the 2^{k+1} independent matrices $\mathbb{S}_j^{(k+1)}$'s associated with the *smaller* blocks $K_{k+1,j}$'s.

For a better understanding of this zooming process, we recall that for any $j \in \{1, \dots, 2^k\}$, $K_{k,j} = K_{k+1,2j-1} \cup K_{k+1,2j}$ and we write

$$\mathbb{S}_j^{(k)} = \sum_{i \in K_{k,j}} \mathbb{X}_j^{(k)}(i) = \sum_{i \in K_{k+1,2j-1}} \mathbb{X}_j^{(k)}(i) + \sum_{i \in K_{k+1,2j}} \mathbb{X}_j^{(k)}(i)\,.$$

The lemma replaces $\mathbb{S}_j^{(k)}$ by $\mathbb{S}_{2j-1}^{(k+1)} + \mathbb{S}_{2j}^{(k+1)}$, where we recall that $\mathbb{S}_{2j-1}^{(k+1)}$ and $\mathbb{S}_{2j}^{(k+1)}$ are independent and have respectively, by (5.14), the same distribution as

$$\sum_{i \in K_{k+1,2j-1}} \mathbb{X}_j^{(k)}(i) \quad \text{and} \quad \sum_{i \in K_{k+1,2j}} \mathbb{X}_j^{(k)}(i)\,.$$

However, the replacement of each $\mathbb{S}_j^{(k)}$ is done up to an error depending on β_{d_k+1} and $\mathrm{Card}(K_{k,j}) = \mathrm{Card}(K_{k,1}) = n_\ell 2^{\ell-k}$. We recall that $\beta_{d_k+1} \le e^{-cd_k}$; so the larger the gap d_k is, the smaller β_{d_k+1} and the error eventually are. The Cantor-like set K_A is constructed, in Subsection 5.3.2, in such a way that for each $k \in \{0, \dots, \ell - 1\}$, β_{d_k+1} compensates $\mathrm{e}^{tMn_\ell 2^{\ell-k}}$ in order to get at the end satisfying error bounds.

We give now the proof, which explains how this process is done and how, for the replacement of each $\mathbb{S}_j^{(k)}$, the factor $\left(1 + \beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\right)$ appears.

Proof of Lemma 5.7. As mentioned before: for any $k \in \{0, \dots, \ell - 1\}$ and any $j \in \{1, \dots, 2^k\}$,

$$K_{k,j} = K_{k+1,2j-1} \cup K_{k+1,2j}$$

where the union is disjoint. Therefore, we set the following notation

$$\mathbb{S}_j^{(k)} = \mathbb{S}_{j,1}^{(k)} + \mathbb{S}_{j,2}^{(k)} \quad \text{and} \quad \mathbf{V}_j^{(k)} = (\mathbf{V}_{j,1}^{(k)}, \mathbf{V}_{j,2}^{(k)})\,,$$

where

$$\mathbb{S}_{j,1}^{(k)} := \sum_{i \in K_{k+1,2j-1}} \mathbb{X}_j^{(k)}(i) \quad , \quad \mathbb{S}_{j,2}^{(k)} := \sum_{i \in K_{k+1,2j}} \mathbb{X}_j^{(k)}(i),$$

$$\mathbf{V}_{j,1}^{(k)} := \Big(\mathbf{X}_j^{(k)}(i)\,, i \in K_{k+1,2j-1}\Big) \quad \text{and} \quad \mathbf{V}_{j,2}^{(k)} := \Big(\mathbf{X}_j^{(k)}(i)\,, i \in K_{k+1,2j}\Big)\,.$$

Note that there are exactly d_k integers separating $K_{k+1,2j-1}$ and $K_{k+1,2j}$ and that for any $i \in \{1, \dots, 2^{k+1}\}$,

$$\mathrm{Card}(K_{k+1,i}) = \mathrm{Card}(K_{k+1,1}) = 2^{\ell-(k+1)} n_\ell\,.$$

Recall that the probability space is assumed to be large enough to contain a sequence $(\delta_i, \eta_i)_{i\in\mathbb{Z}}$ of i.i.d. random variables uniformly distributed on $[0,1]^2$ independent of the sequence $(\mathbf{X}_i)_{i\geq 0}$. Therefore according to the remark on the existence of the family $(\mathbf{V}_j^{(m)})_{1\leq j\leq 2^m}$ made at the beginning of Section 5.3.3, the sequence $(\eta_i)_{i\in\mathbb{Z}}$ is independent of $(\mathbf{V}_j^{(m)})_{1\leq j\leq 2^m}$.

Lemma A.17 allows the following *coupling*: there exists a random vector $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ of size $d^2\mathrm{Card}(K_{k+1,2})$ measurable with respect to $\sigma(\eta_1) \vee \sigma(\mathbf{V}_{1,1}^{(k)}) \vee \sigma(\mathbf{V}_{1,2}^{(k)})$ such that

1. $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ has the same distribution as $\mathbf{V}_{1,2}^{(k)}$,

2. $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ is independent of $\sigma(\mathbf{V}_{1,1}^{(k)})$,

3. $\mathbb{P}(\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}) \leq \beta_{d_k+1}$.

We note that by Lemma A.17, $\mathbb{P}(\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}) = \beta\Big(\sigma(\mathbf{V}_{1,1}^{(k)}), \sigma(\mathbf{V}_{1,2}^{(k)})\Big)$ which, by relation (A.26), depends only on the joint distribution of $(\mathbf{V}_{1,1}^{(k)}, \mathbf{V}_{1,2}^{(k)})$. Therefore, by (5.14),

$$\beta\Big(\sigma(\mathbf{V}_{1,1}^{(k)}), \sigma(\mathbf{V}_{1,2}^{(k)})\Big) = \beta\Big(\sigma(\mathbf{X}_i\,, i \in K_{k+1,1}),\ \sigma(\mathbf{X}_i\,, i \in K_{k+1,2})\Big) \leq \beta_{d_k+1}\,.$$

Note that by construction, $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ is independent of $\sigma\Big(\mathbf{V}_{1,1}^{(k)}, (\mathbf{V}_j^{(k)})_{j=2,\dots,2^k}\Big)$.

For any $i \in K_{k+1,2}$, let

$$\widetilde{\mathbf{X}}_{1,2}^{(k)}(i) = \pi_i^{(k+1)}(\widetilde{\mathbf{V}}_{1,2}^{(k)}) \text{ and } \widetilde{\mathbb{S}}_{1,2}^{(k)} = \sum_{i \in K_{k+1,2}} \widetilde{\mathbb{X}}_{1,2}^{(k)}(i)\,,$$

where $\widetilde{\mathbb{X}}_{1,2}^{(k)}(i)$ is the $d \times d$ random matrix associated with the random vector $\widetilde{\mathbf{X}}_{1,2}^{(k)}(i)$.

With the notations above, we have

$$\begin{aligned}
&\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}=\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) + \mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) \\
&\qquad\le \mathbb{E}\mathrm{Tr}\exp\Big(t\mathbb{S}_{1,1}^{(k)} + t\widetilde{\mathbb{S}}_{1,2}^{(k)} + t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big) + \mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big). \quad (5.17)
\end{aligned}$$

(With usual convention, $\sum_{j=\ell}^{2^k}\mathbb{S}_j^{(k)}$ is the null vector if $\ell > 2^k$). By Golden-Thompson inequality, we have

$$\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big) \le \mathrm{Tr}\Big(\mathrm{e}^{t\mathbb{S}_1^{(k)}}\cdot\mathrm{e}^{t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}}\Big).$$

Since $\sigma\Big(\mathbf{V}_j^{(k)}\,,\, j=2,\ldots,2^k\Big)$ is independent of $\sigma\Big(\mathbf{V}_{1,1}^{(k)},\mathbf{V}_{1,2}^{(k)},\widetilde{\mathbf{V}}_{1,2}^{(k)}\Big)$, we get

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) \le \mathrm{Tr}\Big(\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{e}^{t\mathbb{S}_1^{(k)}}\Big)\cdot\mathbb{E}\exp\Big(t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big).$$

Note now the following fact: *if $\mathbb{U}$ is a $d \times d$ self-adjoint random matrix with entries defined on $(\Omega,\mathcal{A},\mathbb{P})$ and such that $\lambda_{\max}(\mathbb{U}) \le b$ a.s., then for any $\Gamma \in \mathcal{A}$,*

$$\mathbf{1}_\Gamma\mathrm{e}^{\mathbb{U}} \preceq \mathrm{e}^b\mathbb{I}_d\mathbf{1}_\Gamma \quad \textit{a.s. and so } \lambda_{max}\mathbb{E}\Big(\mathbf{1}_\Gamma\mathrm{e}^{\mathbb{U}}\Big) \le \mathrm{e}^b\mathbb{P}(\Gamma)\,.$$

Therefore if we consider $\mathbb{V}$ a $d \times d$ self-adjoint random matrix with entries defined on

$(\Omega, \mathcal{A}, \mathbb{P})$, *the following inequality is valid:*

$$\mathrm{Tr}\Big(\mathbb{E}(\mathbf{1}_\Gamma \mathrm{e}^{\mathbb{U}})\mathbb{E}(\mathrm{e}^{\mathbb{V}})\Big) \leq \mathrm{e}^b \mathbb{P}(\Gamma) \cdot \mathbb{E}\mathrm{Tr}(\mathrm{e}^{\mathbb{V}}) . \tag{5.18}$$

Notice now that $(\mathbf{X}_1^{(k)}(i)\,, i \in K_{k,1})$ has the same distribution as $(\mathbf{X}_i\,, i \in K_{k,1})$. Therefore $\lambda_{\max}(\mathbf{X}_1^{(k)}(i)) \leq M$ a.s. for any i, implying by Weyl's inequality that

$$\lambda_{\max}(t\,\mathbb{S}_1^{(k)}) \leq tM\mathrm{Card}(K_{k,1}) = tM2^{\ell-k}n_\ell \ \text{ a.s.}$$

Hence, applying (5.18) with $b = tM2^{\ell-k}n_\ell$, $\Gamma = \{\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}\}$ and $\mathbb{V} = t\sum_{j=2}^{2^k} \mathbb{S}_j^{(k)}$, and taking into account that $\mathbb{P}(\Gamma) \leq \beta_{d_k+1}$, we obtain

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}} \mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) \leq \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big) . \tag{5.19}$$

Note now that if $\mathbb{V}$ *and* $\mathbb{W}$ *are two independent random matrices with entries defined on* $(\Omega, \mathcal{A}, \mathbb{P})$ *and such that* $\mathbb{E}(\mathbb{W}) = \mathbf{0}$ *then*

$$\mathbb{E}\mathrm{Tr}\,\exp(\mathbb{V}) = \mathbb{E}\mathrm{Tr}\,\exp\Big(\mathbb{E}\Big(\mathbb{V} + \mathbb{W}|\sigma(\mathbb{V})\Big)\Big) .$$

Since $\mathrm{Tr} \circ \exp$ *is convex, it follows from Jensen's inequality applied to the conditional expectation that*

$$\mathbb{E}\mathrm{Tr}\,\exp(\mathbb{V}) \leq \mathbb{E}\Big(\mathbb{E}\Big(\mathrm{Tr}\,\mathrm{e}^{\mathbb{V}+\mathbb{W}}|\sigma(\mathbb{V})\Big)\Big) = \mathbb{E}\Big(Tr\,\mathrm{e}^{\mathbb{V}+\mathbb{W}}\Big) . \tag{5.20}$$

Since $\mathbb{E}(\mathbb{X}_1^{(k)}(i)) = \mathbb{E}(\mathbb{X}_i) = \mathbf{0}$ for any $i \in K_{k,1}$ and since $\sigma(\mathbb{S}_{1,1}^{(k)}, \widetilde{\mathbb{S}}_{1,2}^{(k)})$ and $\sigma(\mathbb{S}_j^{(k)}, j = 2, \ldots, 2^k)$ are independent, we can apply the inequality above with $\mathbb{W} = t(\mathbb{S}_{1,1}^{(k)} + \widetilde{\mathbb{S}}_{1,2}^{(k)})$ and $\mathbb{V} = t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}$. Therefore, starting from (5.19) and using (5.20), we get

$$\begin{aligned} &\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}} \mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) \\ &\qquad \leq \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\,\mathbb{E}\mathrm{Tr}\exp\Big(t\Big(\mathbb{S}_{1,1}^{(k)} + \widetilde{\mathbb{S}}_{1,2}^{(k)} + \sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big)\Big) . \end{aligned} \tag{5.21}$$

Starting from (5.17) and considering (5.21), it follows that

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\right) \le (1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M})\,\mathbb{E}\mathrm{Tr}\exp\left(t\Big(\mathbb{S}_{1,1}^{(k)}+\widetilde{\mathbb{S}}_{1,2}^{(k)}+\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big)\right). \quad (5.22)$$

Having replaced $\mathbb{S}_1^{(k)}$ by $\mathbb{S}_{1,1}^{(k)}+\widetilde{\mathbb{S}}_{1,2}^{(k)}$, the proof of Lemma 5.7 will then be achieved after having iterated this procedure 2^k-1 times more. For the sake of clarity, let us explain the way to go from the j-th step to the $(j+1)$-th step.

At the end of the j-th step, assume that we have constructed with the help of the coupling Lemma A.17, j random vectors $\widetilde{\mathbf{V}}_{i,2}^{(k)}$, $i=1,\dots,j$, each of dimension $d^2\mathrm{Card}(K_{k+1,1})$ and satisfying the following properties: for any i in $\{1,\dots,j\}$, $\widetilde{\mathbf{V}}_{i,2}^{(k)}$ is a measurable function of $(\mathbf{V}_{i,1}^{(k)},\mathbf{V}_{i,2}^{(k)},\eta_i)$, it has the same distribution as $\mathbf{V}_{i,2}^{(k)}$, is such that $\mathbb{P}(\widetilde{\mathbf{V}}_{i,2}^{(k)}\neq\mathbf{V}_{i,2}^{(k)})\le\beta_{d_k+1}$, is independent of $\mathbf{V}_{i,1}^{(k)}$ and it satisfies

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\exp&\left(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\right)\\ &\le\left(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\right)^j\cdot\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+t\sum_{i=j+1}^{2^k}\mathbb{S}_i^{(k)}\right),\end{aligned} \quad (5.23)$$

where we have implemented the following notation:

$$\widetilde{\mathbb{S}}_{i,2}^{(k)}=\sum_{r\in K_{k+1,2i}}\widetilde{\mathbb{X}}_{i,2}^{(k)}(r)\,. \quad (5.24)$$

In the notation above, $\widetilde{\mathbb{X}}_{i,2}^{(k)}(r)$ is the $d\times d$ random matrix associated with the random vector $\widetilde{\mathbf{X}}_{i,2}^{(k)}(r)$ of $\mathbb{K}^{d^2}$ defined by

$$\widetilde{\mathbf{X}}_{i,2}^{(k)}(r)=\pi_r^{(k+1)}(\widetilde{\mathbf{V}}_{i,2}^{(k)})\text{ for any }r\in K_{k+1,2i}\,.$$

Note that the induction assumption above has been proved at the beginning of the proof for $j=1$. Since, for any $m\in\{1,\dots,\ell\}$, $(\mathbf{V}_j^{(m)})_{1\le j\le 2^m}$ is a family of independent random vectors, then for any $i\in\{1,\dots,j\}$, $\widetilde{\mathbf{V}}_{i,2}^{(k)}$ is also independent of

$$\sigma\Big((\mathbf{V}_{\ell,1}^{(k)})_{\ell=1,\dots,i},(\widetilde{\mathbf{V}}_{\ell,2}^{(k)})_{\ell=1,\dots,i-1},(\mathbf{V}_\ell^{(k)})_{\ell=i+1,\dots,2^k}\Big)\,.$$

Now to show that the induction hypothesis also holds at step $j+1$, we proceed as follows. By Lemma A.17, there exists a random vector $\widetilde{\mathbf{V}}^{(k)}_{j+1,2}$ of size $d^2\mathrm{Card}(K_{k+1,1})$, measurable with respect to $\sigma(\eta_{j+1}) \vee \sigma(\mathbf{V}^{(k)}_{j+1,1}) \vee \sigma(\mathbf{V}^{(k)}_{j+1,2})$ such that:

1. $\widetilde{\mathbf{V}}^{(k)}_{j+1,2}$ has the same distribution as $\mathbf{V}^{(k)}_{j+1,2}$,
2. $\widetilde{\mathbf{V}}^{(k)}_{j+1,2}$ is independent of $\sigma(\mathbf{V}^{(k)}_{j+1,1})$,
3. $\mathbb{P}(\widetilde{\mathbf{V}}^{(k)}_{j+1,2} \neq \mathbf{V}^{(k)}_{j+1,2}) \leq \beta_{d_k+1}$.

The inequality in the last item above comes again from (5.14) and the equivalent definition (A.26) of the β-coefficients. Note that

$$\sigma\Big((\mathbf{V}^{(k)}_{i,1})_{i=1,\dots,j+1}, (\widetilde{\mathbf{V}}^{(k)}_{i,2})_{i=1,\dots,j}, (\mathbf{V}^{(k)}_{i})_{i=j+2,\dots,2^k}\Big)$$

and $\sigma(\widetilde{\mathbf{V}}^{(k)}_{j+1,2})$ are independent by construction. With the notation (5.24), we have the following decomposition:

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}\exp&\left(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\right)\\
&\leq \mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j+1}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\right)\\
&\quad+\mathbb{E}\Bigg(\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\right)\Bigg). \quad (5.25)
\end{aligned}$$

Using Golden-Thompson inequality, we have

$$\begin{aligned}
\mathrm{Tr}\exp&\left(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\right)\\
&\leq \mathrm{Tr}\Bigg(\exp\left(t\,\mathbb{S}^{(k)}_{j+1}\right)\cdot\exp\left(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\right)\Bigg).
\end{aligned}$$

Since the sigma algebra generated by $\Big((\mathbf{V}^{(k)}_{i,1})_{i=1,\dots,j}, (\widetilde{\mathbf{V}}^{(k)}_{i,2})_{i=1,\dots,j}, (\mathbf{V}^{(k)}_{i})_{i=j+2,\dots,2^k}\Big)$ is independent of that generated by $\Big(\mathbf{V}^{(k)}_{j+1,1}, \mathbf{V}^{(k)}_{j+1,2}, \widetilde{\mathbf{V}}^{(k)}_{j+1,2}\Big)$, we get

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq \mathbf{V}^{(k)}_{j+1,2}} \operatorname{Tr}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\Big)\Big)$$
$$\leq \operatorname{Tr}\Big(\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq \mathbf{V}^{(k)}_{j+1,2}}\mathrm{e}^{t\,\mathbb{S}^{(k)}_{j+1}}\Big)\cdot \mathbb{E}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\Big)\Big). \quad (5.26)$$

By Weyl's inequality,

$$\lambda_{\max}\Big(t\,\mathbb{S}^{(k)}_{j+1}\Big)\leq t\sum_{r\in K_{k,j+1}}\lambda_{\max}(\mathbb{X}^{(k)}_{j+1}(r)) \text{ a.s.}$$

As $\mathbf{V}^{(k)}_{j+1} =^{\mathcal{D}} (\mathbf{X}_i\,, i\in K_{k,j+1})$ and $\lambda_{\max}(\mathbb{X}_i)\leq M$ a.s. for any i, it follows that

$$\lambda_{\max}\Big(t\,\mathbb{S}^{(k)}_{j+1}\Big)\leq tM\mathrm{Card}(K_{k,j+1}) = tM2^{\ell-k}n_k \text{ a.s.}$$

Moreover, as $\widetilde{\mathbf{V}}^{(k)}_{j+1,2} =^{\mathcal{D}} \mathbf{V}^{(k)}_{j+1,2}$ and $\mathbf{V}^{(k)}_{j+1} =^{\mathcal{D}} (\mathbf{X}_i\,, i\in K_{k,j+1})$, then

$$\mathbb{E}(\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2}) = \mathbf{0}\,.$$

We also notice that $\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2}$ is independent of $\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}$. Therefore, starting from (5.26), an application of inequality (5.18) with $b=tM2^{\ell-k}n_\ell$, $\Gamma=\{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq \mathbf{V}^{(k)}_{j+1,2}\}$ and

$$\mathbb{V} = t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\,,$$

followed by an application of inequality (5.20) with $\mathbb{W}=t(\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2})$, gives

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq \mathbf{V}^{(k)}_{j+1,2}} \operatorname{Tr}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\Big)\Big)$$
$$\leq \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\times \mathbb{E}\operatorname{Tr}\exp\Big(t\sum_{i=1}^{j+1}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\Big). \quad (5.27)$$

Therefore, starting from (5.25) and using (5.27), we get

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+t\sum_{i=j+1}^{2^k}\mathbb{S}_i^{(k)}\right)$$
$$\leq\left(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\right)\times\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j+1}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+t\sum_{i=j+2}^{2^k}\mathbb{S}_i^{(k)}\right),$$

which combined with (5.23) implies that

$$\mathbb{E}\mathrm{Tr}\left(\mathrm{e}^{t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}}\right)$$
$$\leq\left(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\right)^{j+1}\times\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{j+1}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+t\sum_{i=j+2}^{2^k}\mathbb{S}_i^{(k)}\right),$$

proving the induction hypothesis for the step $j+1$. Finally 2^k steps of the procedure lead to

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\right)\leq\left(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\right)^{2^k}\times\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{2^k}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})\right). \qquad (5.28)$$

To end the proof of the lemma it suffices to notice the following facts: the random vectors $\mathbf{V}_{i,1}^{(k)},\widetilde{\mathbf{V}}_{i,2}^{(k)}$, $i=1,\ldots,2^k$, are mutually independent and such that $\mathbf{V}_{i,1}^{(k)}=^{\mathcal{D}}\mathbf{V}_{2i-1}^{(k+1)}$ and $\mathbf{V}_{i,2}^{(k)}=^{\mathcal{D}}\mathbf{V}_{2i}^{(k+1)}$. In addition, the random vectors $\mathbf{V}_i^{(k+1)}$, $i=1,\ldots,2^{k+1}$, are mutually independent. This obviously implies that

$$\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{2^k}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})\right)=\mathbb{E}\mathrm{Tr}\exp\left(t\sum_{i=1}^{2^{k+1}}\mathbb{S}_i^{(k+1)}\right),$$

which ends the proof of the lemma.

□

5.3.4 Proof of Proposition 5.6

Having proved the decoupling Lemma 5.7, we are now ready to prove Inequality (5.8) with K_A defined in Section 5.3.2.

Let us prove it first for the case where $0 < tM \leq 4/A$. We recall that $\mathbb{E}(\mathbb{X}_i) = \mathbf{0}$ for any $i \in K_A$ and we note that by Weyl's inequality,

$$\lambda_{\max}\Big(\sum_{i\in K_A} \mathbb{X}_i\Big) \leq \sum_{i\in K_A} \lambda_{\max}\big(\mathbb{X}_i\big) \leq AM \quad \text{a.s.}$$

Hence, applying Lemma A.13 applied with $K = \{1\}$ and $\mathbb{U}_1 = \sum_{i\in K_A} \mathbb{X}_i$, we get for any $t > 0$,

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A} \mathbb{X}_i\Big) \leq d\exp\Big(t^2 g(tAM)\lambda_{\max}\Big(\mathbb{E}\Big(\sum_{i\in K_A} \mathbb{X}_i\Big)^2\Big)\Big).$$

Therefore, we get by the definition of v^2, the fact that g is increasing, $tAM < 4$ and $g(4) \leq 3.1$, that

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A} \mathbb{X}_i\Big) \leq d\exp(3.1 \times At^2v^2)\,,$$

proving then (5.8).

We prove now Inequality (5.8) for the case where

$$4/A < tM \leq \min\Big(\frac{1}{2}, \frac{c\log 2}{32\log A}\Big) = \min\Big(\frac{1}{2}, \frac{c\delta}{16}\Big)\,.$$

Let

$$\kappa = \frac{c}{8} \quad \text{and} \quad k(t) = \inf\Big\{k \in \mathbb{Z} : \frac{A(1-\delta)^k}{2^k} \leq \min\Big(\frac{\kappa}{(tM)^2}, A\Big)\Big\}. \tag{5.29}$$

Note that if $t^2M^2 \leq \kappa/A$ then $k(t) = 0$ whereas $k(t) \geq 1$ if $t^2M^2 > \kappa/A$. In addition by the selection of ℓ_A, $A((1-\delta)/2)^\ell < 4/\delta$. Therefore $k(t) \leq \ell_A$ since $(tM)^2 \leq c\delta/32$.

Taking into account (5.16) and that $K_A = K_{0,1} = K_{1,1} \cup K_{1,2}$, we apply Lemma 5.7 for $k = 0$,

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A} \mathbb{X}_i\Big) = \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\mathbb{S}_j^{(0)}}\Big) \leq \Big(1 + \beta_{d_0+1}\mathrm{e}^{tMn_\ell 2^\ell}\Big)\,\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\mathbb{S}_1^{(1)} + t\mathbb{S}_2^{(1)}}\Big),$$

where we note that $\mathbb{S}_1^{(1)}$ and $\mathbb{S}_2^{(1)}$ are independent and have respectively the same distribution as $\sum_{i\in K_{1,1}} \mathbb{X}_i$ and $\sum_{i\in K_{1,2}} \mathbb{X}_i$ as explained in (5.14) and (5.15).

Considering the selection of $k(t)$ in (5.29) and applying Lemma 5.7 up to $k(t) - 1$, we get

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A}\mathbb{X}_i\Big) \le \prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}\,\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^{k(t)}}\mathbb{S}_j^{(k(t))}\Big)\,, \tag{5.30}$$

with the usual convention that $\prod_{k=0}^{-1} a_k = 1$.

Note that in the inequality above, $(\mathbb{S}_j^{(k(t))})_{j=1,\dots,2^{k(t)}}$ is a family of mutually independent random matrices defined in (5.15). They are then constructed from a family $(\mathbf{V}_j^{(k(t))})_{1\le j\le 2^{k(t)}}$ of $2^{k(t)}$ mutually independent random vectors that satisfy (5.14). Therefore we have that, for any $j \in 1,\dots,2^{k(t)}$,

$$\mathbb{S}_j^{(k(t))} =^{\mathcal{D}} \sum_{i\in K_{k(t),j}} \mathbb{X}_i\,.$$

Moreover, according to the remark on the existence of the family $(\mathbf{V}_j^{(k(t))})_{1\le j\le 2^{k(t)}}$ made at the beginning of Section 5.3.3, the entries of each random matrix $\mathbb{S}_j^{(k(t))}$ are measurable functions of $(\mathbf{X}_i,\delta_i)_{i\in\mathbb{Z}}$.

The rest of the proof consists of giving a suitable upper bound for

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^{k(t)}}\mathbb{S}_j^{(k(t))}\Big)\,.$$

With this aim, let p be a positive integer to be chosen later such that

$$2p \le \mathrm{Card}(K_{k(t),j}) := q\,. \tag{5.31}$$

Note that $q = 2^{\ell-k(t)}n_\ell$ and by (5.12)

$$q \ge \frac{A}{2^{k(t)+1}}\,.$$

Therefore if $k(t)=0$ then $q \ge A/2$ implying that $q\ge 2$ (since we have $4/A < tM \le 1$). Now if $k(t)\ge 1$ and therefore if $t^2M^2 > \kappa/A$, by the definition of $k(t)$, we have $q \ge \frac{\kappa}{(tM)^2}$ and then $q\ge 2$ since $(tM)^2 \le \kappa/2$. Hence in all cases, $q\ge 2$ implying that the selection of a positive integer p satisfying (5.31) is always possible.

Let $m_{q,p} = [q/(2p)]$. For any $j \in \{1, \dots, 2^{k(t)}\}$, we divide $K_{k(t),j}$ into $2m_{q,p}$ consecutive intervals $(J_{j,i}^{(k(t))}, 1 \le i \le 2m_{q,p})$ each containing p consecutive integers plus a remainder interval $J_{j,2m_{q,p}+1}^{(k(t))}$ containing $r = q - 2pm_{q,p}$ consecutive integers. Note that this last interval contains at most $2p-1$ integers. Let $\mathbb{X}_j^{(k(t))}(k)$ be the $d \times d$ random matrix associated with the random vector $\mathbf{X}_j^{(k(t))}(k)$ defined in (5.15) and define

$$\mathbb{Z}_{j,i} := \mathbb{Z}_{j,i}^{(k(t))} = \sum_{k \in K_{k(t),j} \cap J_{j,i}^{(k(t))}} \mathbb{X}_j^{(k(t))}(k) \, . \tag{5.32}$$

With this notation

$$\mathbb{S}_j^{(k(t))} = \sum_{i=1}^{m_{q,p}+1} \mathbb{Z}_{j,2i-1} + \sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i} \, .$$

Since $\mathrm{Tr} \circ \exp$ is a convex function, we get

$$\begin{aligned} &\mathbb{E}\mathrm{Tr} \exp \Big(t \sum_{j=1}^{2^{k(t)}} \mathbb{S}_j^{(k(t))} \Big) \\ &\qquad \le \frac{1}{2} \mathbb{E}\mathrm{Tr} \exp \Big(2t \sum_{j=1}^{2^{k(t)}} \sum_{i=1}^{m_{q,p}+1} \mathbb{Z}_{j,2i-1} \Big) + \frac{1}{2} \mathbb{E}\mathrm{Tr} \exp \Big(2t \sum_{j=1}^{2^{k(t)}} \sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i} \Big) \, . \end{aligned} \tag{5.33}$$

We start by giving an upper bound for $\mathbb{E}\mathrm{Tr} \exp \Big(2t \sum_{j=1}^{2^{k(t)}} \sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i} \Big)$. With this aim, let us define the following vectors

$$\mathbf{U}_{j,i} := \mathbf{U}_{j,i}^{(k(t))} = \Big(\mathbf{X}_j^{(k(t))}(k) \, , \, k \in K_{k(t),j} \cap J_{j,i}^{(k(t))} \Big) \tag{5.34}$$

and

$$\mathbf{W}_j := \mathbf{W}_j^{(k(t))} = \Big(\mathbf{U}_{j,i} \, , \, i \in \{1, \dots, 2m_{q,p}+1\} \Big) \, .$$

Proceeding by induction and using the coupling lemma A.17, one can construct random vectors $\mathbf{U}_{j,2i}^*$, $j = 1, \dots, 2^{k(t)}$, $i = 1, \dots, m_{q,p}$, that satisfy the following properties:

(i) $(\mathbf{U}_{j,2i}^*, (j,i) \in \{1, \dots, 2^{k(t)}\} \times \{1, \dots, m_{q,p}\})$ is a family of mutually independent random vectors,

(ii) $\mathbf{U}_{j,2i}^*$ has the same distribution as $\mathbf{U}_{j,2i}$,

(iii) $\mathbb{P}(\mathbf{U}_{j,2i}^* \neq \mathbf{U}_{j,2i}) \le \beta_{p+1}$.

Let us explain the construction. Recall first that $(\Omega, \mathcal{A}, \mathbb{P})$ is assumed to be rich enough to contain a sequence $(\eta_i)_{i\in\mathbb{Z}}$ of i.i.d. random variables with uniform distribution over $[0,1]$ independent of $(\mathbf{X}_i, \delta_i)_{i\in\mathbb{Z}}$ where we recall that the sequence $(\delta_i)_{i\in\mathbb{Z}}$ has been used to construct the independent random matrices $\mathbb{S}_j^{(k(t))}$, $j = 1, \dots, k(t)$, involved in inequality (5.30).

For any $j \in \{1, \dots, 2^{k(t)}\}$, let $\mathbf{U}^*_{j,2} = \mathbf{U}_{j,2}$, and construct the random vectors $\mathbf{U}^*_{j,2i}$, $i = 2, \dots, m_{q,p}$, recursively from $(\mathbf{U}^*_{j,2\ell}, 1 \le \ell \le i-1)$ as follows. According to Lemma A.17, there exists a random vector $\mathbf{U}^*_{j,2i}$ such that

$$\mathbf{U}^*_{j,2i} = f_{i,j}\Big((\mathbf{U}^*_{j,2\ell})_{1\le\ell\le i-1}, \mathbf{U}_{j,2i}, \eta_{i+(j-1)2^{k(t)}}\Big) \tag{5.35}$$

where $f_{i,j}$ is a measurable function, $\mathbf{U}^*_{j,2i}$ has the same law as $\mathbf{U}_{j,2i}$, is independent of $\sigma\Big(\mathbf{U}^*_{j,2\ell}, 1 \le \ell \le i-1\Big)$ and

$$\mathbb{P}(\mathbf{U}^*_{j,2i} \neq \mathbf{U}_{j,2i}) = \beta\Big(\sigma\Big(\mathbf{U}^*_{j,2\ell}, 1 \le \ell \le i-1\Big), \sigma(\mathbf{U}_{j,2i})\Big) \le \beta_{p+1}\,.$$

By construction, for any fixed $j \in \{1, \dots, 2^{k(t)}\}$, the random vectors $\mathbf{U}^*_{j,2i}$, $i = 1, \dots, m_{q,p}$, are mutually independent. In addition, by (5.35) and the fact that $(\mathbf{W}_j, j = 1, \dots, 2^{k(t)})$ is a family of mutually independent random vectors, we note that $(\mathbf{U}^*_{j,2i}, (i,j) \in \{1, \dots, m_{q,p}\} \times \{1, \dots, 2^{k(t)}\})$ is also so. Therefore, Items (i) and (ii) above are satisfied by the constructed random vectors $\mathbf{U}^*_{j,2i}$ $i = 1, \dots, m_{q,p}$, $j = 1, \dots, 2^{k(t)}$. Moreover, by (5.35), we have

$$\sigma\Big(\mathbf{U}^*_{j,2\ell}, 1 \le \ell \le i-1\Big) \subseteq \sigma\Big(\mathbf{U}_{j,2\ell}, 1 \le \ell \le i-1\Big) \vee \sigma\Big(\eta_{\ell+(j-1)2^{k(t)}}, 1 \le \ell \le i-1\Big)\,.$$

Since $(\eta_i)_{i\in\mathbb{Z}}$ is independent of $(\mathbf{X}_i, \delta_i)_{i\in\mathbb{Z}}$, we have

$$\beta\Big(\sigma\Big(\mathbf{U}^*_{j,2\ell}, 1 \le \ell \le i-1\Big), \sigma(\mathbf{U}_{j,2i})\Big) \le \beta\Big(\sigma\Big(\mathbf{U}_{j,2\ell}, 1 \le \ell \le i-1\Big), \sigma(\mathbf{U}_{j,2i})\Big)\,.$$

By relation (A.26), the quantity $\beta(\sigma(\mathbf{U}_{j,2\ell}, 1 \le \ell \le i-1), \sigma(\mathbf{U}_{j,2i}))$ depends only on the joint distribution of $((\mathbf{U}_{j,2\ell})_{1\le\ell\le i-1}, \mathbf{U}_{j,2i})$. By the definition (5.34) of the $\mathbf{U}_{j,\ell}$'s, the definition (5.15) of the $\mathbf{X}_j^{(k(t))}(k)$'s and (5.14), we infer that

$$\beta\Big(\sigma\Big(\mathbf{U}_{j,2\ell}, 1 \le \ell \le i-1\Big), \sigma(\mathbf{U}_{j,2i})\Big) \\ = \beta\Big(\sigma\Big(\mathbf{X}_k\,,\, k \in \cup_{\ell=1}^{i-1} K_{k(t),j} \cap J_{j,2\ell}^{(k(t))}\Big), \sigma\Big(\mathbf{X}_k\,,\, k \in K_{k(t),j} \cap J_{j,2i}^{(k(t))}\Big)\Big) \le \beta_{p+1}\,.$$

So, overall, the constructed random vectors $\mathbf{U}^*_{j,2i}$ $i = 1, \dots, m_{q,p}$, $j = 1, \dots, 2^{k(t)}$, satisfy also Item (iii) above.

Denote now

$$\mathbf{X}^*_{j,2i}(\ell) = \pi_\ell(\mathbf{U}^*_{j,2i})$$

where $\pi_i^{(m)}$ is the ℓ-th canonical projection from $\mathbb{K}^{pd^2}$ onto $\mathbb{K}^{d^2}$, namely: for any vector $\mathbf{x} = (\mathbf{x}_i\,, i \in \{1, \dots, p\})$ of $\mathbb{K}^{pd^2}$, $\pi_\ell(\mathbf{x}) = \mathbf{x}_\ell$. Let $\mathbb{X}^*_{j,2i}(\ell)$ be the $d \times d$ random matrix associated with $\mathbf{X}^*_{j,2i}(\ell)$ and define, for any $i = 1 \dots, m_{q,p}$,

$$\mathbb{Z}^*_{j,2i} = \sum_{\ell \in K_{k(t),j} \cap J_{j,2i}^{(k(t))}} \mathbb{X}^*_{j,2i}(\ell)\,.$$

Recalling the definition of $\mathbb{Z}_{j,2i}$ in (5.32), we observe that by Item (ii), $\mathbb{Z}^*_{j,2i} =^{\mathcal{D}} \mathbb{Z}_{j,2i}$, and that by Item (i), the random matrices $\mathbb{Z}^*_{j,2i}$, $i = 1, \dots, m_{q,p}$, $j = 1, \dots, 2^{k(t)}$, are mutually independent.

The aim now is to prove that the following inequality is valid:

$$\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}_{j,2i}\right) \\ \le \Big(1 + (m_{q,p} - 1)\mathrm{e}^{qtM}\beta_{p+1}\Big)^{2^{k(t)}} \mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i}\right). \quad (5.36)$$

Obviously, this can be done by induction if we can show that, for any ℓ in $\{1, \dots, 2^{k(t)}\}$,

$$\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}_{j,2i}\right) \\ \le \Big(1 + (m_{q,p} - 1)\mathrm{e}^{qtM}\ \beta_{p+1}\Big)\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{\ell}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}_{j,2i}\right). \quad (5.37)$$

To prove the inequality above, we set

$$\mathbb{C}_{\ell-1,\ell}(t) = 2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}} \mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}$$

and we write

$$\begin{aligned}
&\mathbb{E}\mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t)) \\
&= \mathbb{E}\Big(\prod_{i=2}^{m_{q,p}} \mathbf{1}_{\mathbf{U}_{\ell,2i}=\mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t))\Big) + \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq\mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t))\Big) \\
&\qquad\qquad \leq \mathbb{E}\mathrm{Tr}\exp(\mathbb{C}_{\ell,\ell+1}(t)) + \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq\mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t))\Big)\,. \qquad (5.38)
\end{aligned}$$

Note that the sigma algebra generated by the random vectors

$$(\mathbf{U}^{*(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},j\in\{1,\dots,\ell-1\}} \quad \text{and} \quad (\mathbf{U}^{(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},\,j\in\{\ell+1,\dots,2^{k(t)}\}}$$

is independent of $\sigma\Big((\mathbf{U}^{(k)}_{\ell,2i}, \mathbf{U}^{*(k)}_{\ell,2i})_{i\in\{1,\dots,m_{q,p}\}}\Big)$. This fact together with the Golden-Thomson inequality give

$$\begin{aligned}
&\mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq\mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big) \\
&\quad \leq \mathrm{Tr}\Big(\mathbb{E}\Big(\exp\Big(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}} \mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}\Big)\Big) \\
&\qquad\qquad\qquad \times\, \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq\mathbf{U}^*_{\ell,2i}} \exp\Big(2t\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{\ell,2i}\Big)\Big)\Big)\,.
\end{aligned}$$

By Weyl's inequality and (5.14), we infer that, almost surely,

$$\lambda_{\max}\Big(2t\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{\ell,2i}\Big) \leq 2t\sum_{i=1}^{m_{q,p}} \sum_{k\in K_{k(t),\ell}\cap J^{(k(t))}_{\ell,2i}} \lambda_{\max}\Big(\mathbb{X}_k\Big) \leq 2t m_{q,p} p M \leq tqM\,. \qquad (5.39)$$

Therefore, applying (5.18) with $b = tqM$, $\Gamma = \{\exists i \in \{2,\dots,m_{q,p}\} \,:\, \mathbf{U}_{\ell,2i} \neq \mathbf{U}^*_{\ell,2i}\}$ and

$$\mathbb{V} = 2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}} \mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}$$

and taking into account that

$$\mathbb{P}(\Gamma) \leq \sum_{i=2}^{m_{q,p}} \mathbb{P}(\mathbf{U}_{\ell,2i} \neq \mathbf{U}^*_{\ell,2i}) \leq (m_{q,p}-1)\beta_{p+1}\,,$$

we get

$$\begin{aligned}&\mathbb{E}\Big(\mathbf{1}_{\exists i \in \{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq \mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t))\Big)\\ &\qquad\leq (m_{q,p}-1)\beta_{p+1}\mathrm{e}^{qtM}\,\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}_{j,2i}\right).\end{aligned}$$

Noting that the sigma algebra generated by the random vectors

$$(\mathbf{U}^*_{j,2i})_{i\in\{1,\dots,m_{q,p}\},\,j\in\{1,\dots,\ell-1\}} \quad \text{and} \quad (\mathbf{U}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},j\in\{\ell+1,\dots,2^{k(t)}\}}$$

is independent of $\sigma\Big((\mathbf{U}^*_{\ell,2i})_{i\in\{1,\dots,m_{q,p}\}}\Big)$, and that, by construction,

$$\mathbb{E}(\mathbb{Z}^*_{\ell,2i}) = \mathbb{E}(\mathbb{Z}_{\ell,2i}) = \mathbf{0}\,,$$

then an application of inequality (5.20) yields

$$\begin{aligned}&\mathbb{E}\Big(\mathbf{1}_{\exists i \in \{2,\dots,m_{q,p}\}\,:\,\mathbf{U}_{\ell,2i}\neq \mathbf{U}^*_{\ell,2i}} \mathrm{Tr}\exp(\mathbb{C}_{\ell-1,\ell}(t)\Big)\\ &\qquad\leq \beta_{p+1}(m_{q,p}-1)\mathrm{e}^{qtM}\,\mathbb{E}\mathrm{Tr}\exp(\mathbb{C}_{\ell,\ell+1}(t))\,. \end{aligned}\tag{5.40}$$

Starting from (5.38) and taking into account (5.40), inequality (5.37) follows and so does inequality (5.36).

With the same arguments as above and with obvious notations, we infer that

$$\begin{aligned}&\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}_{j,2i-1}\right)\\ &\qquad\leq \left(1+m_{q,p}\mathrm{e}^{2qtM}\beta_{p+1}\right)^{2^{k(t)}}\mathbb{E}\mathrm{Tr}\exp\left(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}^*_{j,2i-1}\right).\end{aligned}\tag{5.41}$$

Note that to get the above inequality, we used instead of (5.39) that, almost surely,

$$\lambda_{\max}\Big(2t\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}_{\ell,2i-1}\Big)\le 2t\sum_{i=1}^{m_{q,p+1}}\sum_{k\in K_{k(t),\ell}\cap J^{(k(t))}_{\ell,2i-1}}\lambda_{\max}\big(\mathbb{X}_k\big)$$
$$\le 2Mt(m_{q,p}p+q-2pm_{q,p})=2Mt(q-pm_{q,p})$$
$$\le Mt(q+2p)\le 2tqM\,.$$

Starting from (5.30) and taking into account (5.33), (5.36) and (5.41), we then derive

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A}\mathbb{X}_i\Big)\le\Big(1+m_{q,p}\mathrm{e}^{2qtM}\beta_{p+1}\Big)^{2^{k(t)}}\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}$$
$$\times\Big(\frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i}\Big)+\frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}+1}\mathbb{Z}^*_{j,2i-1}\Big)\Big)\,. \qquad (5.42)$$

Now we choose

$$p=\Big[\frac{2}{tM}\Big]\wedge\Big[\frac{q}{2}\Big]\,.$$

Note that the random vectors $(\mathbb{Z}^*_{j,2i-1})_{i,j}$ are mutually independent and centered. Moreover,

$$2\lambda_{\max}(\mathbb{Z}^*_{j,2i-1})\le 2Mp\le\frac{4}{t}\quad\text{a.s.}$$

Therefore by using Lemma A.13 together with the definition of v^2 and the fact that

$$2^{k(t)}(m_{q,p}+1)p\le 2^{k(t)}q\le A\,,$$

we get

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}+1}\mathbb{Z}^*_{j,2i-1}\Big)\le d\exp(4\times 3.1\times At^2v^2)\,. \qquad (5.43)$$

Similarly, we obtain that

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^*_{j,2i}\Big)\le d\exp(4\times 3.1\times At^2v^2)\,. \qquad (5.44)$$

Next, by using Condition (5.2) and (5.12), we get

$$\log\Big(1+m_{q,p}\mathrm{e}^{2tqM}\beta_{p+1}\Big)^{2^{k(t)}} \le 2^{k(t)}m_{q,p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \le \frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp}\,. \tag{5.45}$$

Several situations can occur. Either $(tM)^2 \le \kappa/A$ and in this case $k(t)=0$ implying that $A/2 \le q \le A \le \kappa/(tM)^2$. If in addition $q \ge 4/(tM)$ then $p=[2/(tM)] \ge 1/tM$ (since $tM \le 1$) and

$$\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \le \frac{AtM}{2}\mathrm{e}^{2\kappa/(tM)}\mathrm{e}^{-c/(tM)} = \frac{AtM}{2}\mathrm{e}^{-3c/(4tM)}\,.$$

Since $\log_2 A \le c/(32tM)$, $A \ge 2$, and $\mathrm{e}^{-3c/(8tM)} \le 8tM/(3c)$, we get

$$\begin{aligned}\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} &\le \frac{AtM}{2}\exp\Big(-\frac{3c}{4tM}\Big) \le \frac{4A(tM)^2}{3c}\exp\Big(-\frac{3c}{8tM}\Big)\\ &\le \frac{4(tM)^2}{3cA^7}\exp\Big(-\frac{3c}{16tM}\Big) \le \frac{(tM)^2}{c}\exp\Big(-\frac{3c}{16tM}\Big)\,.\end{aligned}$$

If otherwise $q < 4/(tM)$ then $p=[q/2] \ge q/4$. Hence, since $2tM \le c/16$ (since $\log A \ge \log 2$) and $tM > 4/A$,

$$\begin{aligned}\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} &\le \frac{2A}{q}\mathrm{e}^{-3cq/16} \le 4\mathrm{e}^{-3cA/32}\\ &\le AtM\exp\Big(-\frac{3c}{8tM}\Big) \le \frac{(tM)^2}{c}\exp\Big(-\frac{3c}{32tM}\Big)\,,\end{aligned}$$

where we have used that $A/2 \le q$ for the second inequality, and that $\log_2 A \le c/(32tM)$, $A \ge 2$ and $\mathrm{e}^{-3c/(16tM)} \le 16tM/(3c)$ for the last one.

Either $(tM)^2 > \kappa/A$ and in this case $k(t) \ge 1$ and by using (5.12) and the definition of $k(t)$, we have

$$q \ge \frac{A}{2^{k(t)+1}} \ge \frac{\kappa}{4(tM)^2}\,. \tag{5.46}$$

If in addition $q \ge 4/(tM)$ then $p=[2/(tM)] \ge 1/tM$, and by (5.11) and the definition of $k(t)$,

$$q \le 2A\frac{(1-\delta)^\ell}{2^{k(t)}} \le \frac{2\kappa}{(tM)^2}\,.$$

It follows that

$$\begin{aligned}\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} &\leq \frac{AtM}{2}\exp\Big(\frac{4\kappa}{tM}\Big)\exp\Big(-\frac{c}{tM}\Big)\\ &\leq \frac{AtM}{2}\exp\Big(-\frac{c}{2tM}\Big) \leq \frac{(tM)^2}{c}\exp\Big(-\frac{c}{8tM}\Big),\end{aligned}$$

where we have used that $\log_2 A \leq c/(32tM)$, $A \geq 2$ and $\mathrm{e}^{-c/(4tM)} \leq 4tM/c$ for the last inequality. Now if $q < 4/(tM)$ then $p = [q/2] \geq q/4$. Hence, using again the fact that $2tM \leq c/16$ combined with (5.46), we get

$$\begin{aligned}\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} &\leq \frac{8A(tM)^2}{\kappa}\mathrm{e}^{-3cq/16} \leq \frac{8^2A(tM)^2}{c}\exp\Big(-\frac{3c^2}{16\times 4\times 8(tM)^2}\Big)\\ &\leq \frac{8(tM)^2}{c}\mathrm{e}^{-3c/(32tM)},\end{aligned}$$

where we have used that $\log_2 A \leq c^2/(32tM)^2$ and $A \geq 2$ for the last inequality.

So, overall, starting from (5.45), we get

$$\log\Big(1 + m_{q,p}\mathrm{e}^{2tqM}\beta_{p+1}\Big)^{2^{k(t)}} \leq \frac{8(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}. \tag{5.47}$$

We handle now the term $\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}$ only in the case where $(\kappa/A)^{1/2} < tM$, otherwise this term is equal to one. By taking into account (5.2), (5.10), (5.11) and the fact that $tM \leq c\delta/8$, we have

$$\begin{aligned}\log\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k} &\leq \sum_{k=0}^{k(t)-1} 2^k\exp\Big(-c\frac{A\delta(1-\delta)^k}{2^{k+1}} + 2tM\frac{A(1-\delta)^\ell}{2^k}\Big)\\ &\leq \sum_{k=0}^{k(t)-1} 2^k\exp\Big(-c\frac{A\delta(1-\delta)^k}{2^{k+2}}\Big)\\ &\leq 2^{k(t)}\exp\Big(-\frac{Ac\delta(1-\delta)^{k(t)-1}}{2^{k(t)+1}}\Big).\end{aligned}$$

By the definition of $k(t)$, we have $A\dfrac{(1-\delta)^{k(t)-1}}{2^{k(t)-1}} > \dfrac{\kappa}{(tM)^2}$. Therefore $2^{k(t)} \leq 2A\frac{(tM)^2}{\kappa}$.

Moreover

$$Ac\delta\frac{(1-\delta)^{k(t)-1}}{2^{k(t)+1}} > \frac{c\kappa\delta}{4(tM)^2} \geq \frac{2\kappa}{tM},$$

since $tM \leq c\delta/8$. It follows that

$$\begin{aligned}\log\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k} &\leq 2A\frac{(tM)^2}{\kappa}\exp\Big(-\frac{2\kappa}{tM}\Big)\\ &\leq \frac{(tM)^2}{c}\exp\Big(-\frac{3c}{32tM}\Big),\end{aligned} \tag{5.48}$$

where we have used the fact that $\log_2 A \leq c/(32tM)$.

So, overall, starting from (5.42) and considering the upper bounds (5.43), (5.44), (5.47) and (5.48), we get

$$\log\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A}\mathbb{X}_i\Big) \leq \log d + 4\times 3.1At^2v^2 + \frac{9(tM)^2}{c}\exp\Big(-\frac{3c}{32tM}\Big).$$

Therefore Inequality (5.8) also holds in the case where $4/A < tM \leq \min\big(\frac{1}{2}, \frac{c\log 2}{32\log A}\big)$. This ends the proof of the proposition. □

5.3.5 Proof of the Bernstein Inequality

Let $A_0 = A = n$ and $\mathbb{Y}^{(0)}(k) = \mathbb{X}_k$ for any $k = 1,\dots,A_0$. Let K_{A_0} be the discrete Cantor type set as defined from $\{1,\dots,A_0\}$ in Section 5.3.2. Let $A_1 = A_0 - \mathrm{Card}(K_{A_0})$ and define for any $k = 1,\dots,A_1$,

$$\mathbb{Y}^{(1)}(k) = \mathbb{X}_{i_k} \text{ where } \{i_1,\dots,i_{A_1}\} = \{1,\dots,A\}\setminus K_A.$$

Now for $i \geq 1$, let K_{A_i} be defined from $\{1,\dots,A_i\}$ exactly as K_A is defined from $\{1,\dots,A\}$. Set $A_{i+1} = A_i - \mathrm{Card}(K_{A_i})$ and $\{j_1,\dots,j_{A_{i+1}}\} = \{1,\dots,A_i\}\setminus K_{A_i}$. Define now

$$\mathbb{Y}^{(i+1)}(k) = \mathbb{Y}^{(i)}(j_k) \text{ for } k = 1,\dots,A_{i+1},$$

and set

$$L = L_n = \inf\{j \in \mathbb{N}^*\,,\ A_j \le 2\}\,.$$

Note that, for any $i \in \{0, \dots, L-1\}$, $A_i > 2$ and $\mathrm{Card}(K_{A_i}) \ge A_i/2$. Moreover $A_i \le n2^{-i}$. With this choice of L, we have

$$L \le \left[\frac{\log n - \log 2}{\log 2}\right] + 1\,.$$

By the above notations, we have

$$\begin{aligned}\sum_{k=1}^{n} \mathbb{X}_k := \sum_{k=1}^{A_0} \mathbb{Y}_k^{(0)} &= \sum_{k \in K_{A_0}} \mathbb{Y}_k^{(0)} + \sum_{k \in \{1,\dots,A\} \setminus K_{A_0}} \mathbb{Y}_k^{(0)} \\ &= \sum_{k \in K_{A_0}} \mathbb{Y}_k^{(0)} + \sum_{k=1}^{A_1} \mathbb{Y}_k^{(1)}\end{aligned}$$

Iterating the above process clearly gives the following decomposition

$$\sum_{k=1}^{n} \mathbb{X}_k = \sum_{i=0}^{L-1} \sum_{k \in K_{A_i}} \mathbb{Y}^{(i)}(k) + \sum_{k=1}^{A_L} \mathbb{Y}^{(L)}(k)\,. \tag{5.49}$$

Let

$$\mathbb{U}_i = \sum_{k \in K_{A_i}} \mathbb{Y}^{(i)}(k) \text{ for } 0 \le i \le L-1 \text{ and } \mathbb{U}_L = \sum_{k=1}^{A_L} \mathbb{Y}^{(L)}(k)\,,$$

For any positive x, let

$$h(c,x) = \min\left(\frac{1}{2}, \frac{c \log 2}{32 \log x}\right).$$

Noting that, for any $i \in \{0, \dots, L-1\}$, the self-adjoint random matrices $(Y^{(i)}(k))_k$ satisfy the condition (5.2) with the same constant c, we can apply Proposition 5.6 and get for any positive t satisfying $tM < h(c, n/2^i)$,

$$\log \mathbb{E}\mathrm{Tr}\Big(\exp(t\mathbb{U}_i)\Big) \le \log d + 4 \times 3.1 A_i t^2 v^2 + \frac{9(tM)^2}{c} \exp\Big(-\frac{3c}{32tM}\Big)\,. \tag{5.50}$$

Recalling that $A_i \le n2^{-i}$ and noting that $\exp(-\frac{3c}{32tM}) < (2^i/n)^4$ for $tM < h(c, n/2^i)$, we get

$$\begin{aligned}\log \mathbb{E}\mathrm{Tr}\Big(\exp(t\mathbb{U}_i)\Big) &\le \log d + 16t^2v^2\frac{n}{2^i} + \frac{9(tM)^2 2^{4i}}{cn^4} \\ &\le \log d + \frac{4t^2n}{2^i}\Big(2v + \frac{\sqrt{3}M2^{\frac{5i}{2}}}{\sqrt{c}n^{\frac{5}{2}}}\Big)^2 \\ &\le \log d + \frac{4t^2n2^{-i}(2v + \sqrt{3}\times 2^{5i/2}M/(n^{5/2}\sqrt{c}))^2}{1 - tM/h(c, n2^{-i})}\,. \end{aligned} \tag{5.51}$$

On the other hand, by Weyl's inequality,

$$\lambda_{\max}\Big(\mathbb{U}_L\Big) \le MA_L \le 2M\,.$$

Therefore by using Lemma A.13, for any positive t,

$$\mathbb{E}\mathrm{Tr}\Big(\exp(t\mathbb{U}_L)\Big) \le d\exp\Big(t^2 g(2tM)\lambda_{\max}\Big(\mathbb{E}(\mathbb{U}_L^2)\Big)\Big)\,.$$

Hence by the definition of v^2, for any positive t such that $tM < 1$, we get

$$\log \mathbb{E}\mathrm{Tr}\Big(\exp(t\mathbb{U}_L)\Big) \le \log d + 2t^2v^2 \le \log d + \frac{2t^2v^2}{1-tM}\,. \tag{5.52}$$

Let

$$\kappa_i = \frac{M}{h(c, n/2^i)} \text{ for } 0 \le i \le L-1 \text{ and } \kappa_L = M$$

and

$$\sigma_i = 2\frac{\sqrt{n}}{2^{i/2}}\Big(2v + \sqrt{3}\times\frac{2^iM}{n\sqrt{c}}\Big) \text{ for } 0 \le i \le L-1 \text{ and } \sigma_L = v\sqrt{2}\,.$$

By the above choice of L, we get

$$\sum_{i=0}^{L}\kappa_i \le M\Big(\sum_{i=0}^{L-1}\frac{1}{h(c, n/2^i)} + 1\Big) \le M\frac{\log n}{\log 2}\max\Big(2, \frac{32\log n}{c\log 2}\Big) = M\gamma(c,n)\,.$$

Moreover,

$$\begin{aligned}\sum_{i=0}^{L}\sigma_i &= 2\sqrt{n}\sum_{i=0}^{L-1}2^{-i/2}\Big(2v+\sqrt{3}\times\frac{2^{5i/2}M}{n^{5/2}\sqrt{c}}\Big) \; + v\sqrt{2}\\ &\le 14\sqrt{n}v + 2c^{-1/2}n^{-2}M2^{2L} + v\sqrt{2}\\ &\le 15\sqrt{n}v + 2c^{-1/2}M\,.\end{aligned}$$

Taking into account (5.51) and (5.52), we get overall by Lemma A.15, that for any positive t such that $tM < 1/\gamma(c,n)$,

$$\log \mathbb{E}\mathrm{Tr}\Big(\exp\Big(t\sum_{i=1}^{n}\mathbb{X}_i\Big)\Big) \le \log d + \frac{t^2 n\Big(15v + 2M/(cn)^{1/2}\Big)^2}{1 - tM\gamma(c,n)} := \gamma_n(t)\,. \tag{5.53}$$

To end the proof of the theorem, it suffices to notice that for any positive x

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^{n}\mathbb{X}_i\Big) \ge x\Big) \le \inf_{t>0\,:\,tM\le 1/\gamma(c,n)} \exp\Big(-tx + \gamma_n(t)\Big)\,,$$

where $\gamma_n(t)$ is defined in (5.53). □

Appendix A

Technical Lemmas

In this Appendix, we collect some technical lemmas in addition to some preliminary materials and tools.

A.1 On the Stieltjes transform of Gram matrices

Proposition A.1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $N \times n$ *random matrices, we have for any* $z = u + iv \in \mathbb{C}_+$,

$$\left|S_{\mathbf{A}\mathbf{A}^T}(z) - S_{\mathbf{B}\mathbf{B}^T}(z)\right| \leq \frac{\sqrt{2}}{Nv^2}\left(\mathrm{Tr}(\mathbf{A}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T)\right)^{1/2}\left(\mathrm{Tr}(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^T\right)^{1/2}.$$

Proof. For any $z = u + iv \in \mathbb{C}^+$, we get by integrating by parts

$$\begin{aligned}\left|S_{\mathbf{A}\mathbf{A}^T}(z) - S_{\mathbf{B}\mathbf{B}^T}(z)\right| &\leq \left|\int \frac{1}{x-z} dF^{\mathbf{A}\mathbf{A}^T}(x) - \int \frac{1}{x-z} dF^{\mathbf{B}\mathbf{B}^T}(x)\right| \\ &= \left|\int \frac{F^{\mathbf{A}\mathbf{A}^T}(x) - F^{\mathbf{B}\mathbf{B}^T}(x)}{(x-z)^2} dx\right| \\ &\leq \frac{1}{v^2}\int \left|F^{\mathbf{A}\mathbf{A}^T}(x) - F^{\mathbf{B}\mathbf{B}^T}(x)\right| dx\,. \end{aligned} \tag{A.1}$$

Now, $\int |F^{\mathbf{A}\mathbf{A}^T}(x) - F^{\mathbf{B}\mathbf{B}^T}(x)| dx$ is nothing but the Wasserstein distance of first order between the empirical measures of $\mathbf{AA}^T$ and $\mathbf{BB}^T$. To be more precise, if $\lambda_1 \geq \ldots \geq \lambda_N$ denote the eigenvalues of $\mathbf{AA}^T$ and $\bar{\lambda}_1 \geq \ldots \geq \bar{\lambda}_N$ denote those of $\mathbf{BB}^T$, then setting

$$\eta_n = \frac{1}{N} \sum_{k=1}^{N} \delta_{\lambda_k} \quad \text{and} \quad \bar{\eta}_n = \frac{1}{N} \sum_{k=1}^{N} \delta_{\bar{\lambda}_k} \, ,$$

we have that

$$\int \left| F^{\mathbf{AA}^T}(x) - F^{\mathbf{BB}^T}(x) \right| dx = W_1(\eta_n, \bar{\eta}_n) = \inf \mathbb{E}|X - Y| \, ,$$

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. [17], we have that

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

where π is a permutation belonging to the symmetric group $\mathcal{S}_N$ of $\{1, \ldots, N\}$. By standard arguments, involving the fact that if x, y, u, v are real numbers such that $x \leq y$ and $u > v$, then

$$|x - u| + |y - v| \geq |x - v| + |y - u|$$

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,

$$W_1(\eta_n, \bar{\eta}_n) = \int \left| F^{\mathbf{AA}^T}(x) - F^{\mathbf{BB}^T}(x) \right| dx \tag{A.2}$$

$$= \frac{1}{N} \sum_{k=1}^{N \wedge n} |\lambda_k - \bar{\lambda}_k| \, . \tag{A.3}$$

Since $\lambda_k = s_k^2$ and $\bar{\lambda}_k = \bar{s}_k^2$ where the s_k's and the $\bar{s}_k$'s are respectively the singular values of A and B, we get by Cauchy-Schwarz's inequality,

$$
\begin{aligned}
\sum_{k=1}^{N\wedge n}|\lambda_k-\bar{\lambda}_k| &\le \Big(\sum_{k=1}^{N\wedge n}|s_k+\bar{s}_k|^2\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}|s_k-\bar{s}_k|^2\Big)^{1/2}\\
&\le 2^{1/2}\Big(\sum_{k=1}^{N\wedge n}(s_k^2+\bar{s}_k^2)\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}|s_k-\bar{s}_k|^2\Big)^{1/2}\\
&\le 2^{1/2}\Big(\mathrm{Tr}(\mathbf{A}\mathbf{A}^T)+\mathrm{Tr}(\mathbf{B}\mathbf{B}^T)\Big)^{1/2}\Big(\sum_{k=1}^{N\wedge n}|s_k-\bar{s}_k|^2\Big)^{1/2}. \qquad (A.4)
\end{aligned}
$$

Next, by Hoffman-Wielandt's inequality, Corollary 7.3.8 in [37],

$$
\sum_{k=1}^{N\wedge n}|s_k-\bar{s}_k|^2 \le \mathrm{Tr}(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^T .
$$

Therefore,

$$
\sum_{k=1}^{N\wedge n}|\lambda_k-\bar{\lambda}_k| \le 2^{1/2}\Big(\mathrm{Tr}(\mathbf{A}\mathbf{A}^T)+\mathrm{Tr}(\mathbf{B}\mathbf{B}^T)\Big)^{1/2}\Big(\mathrm{Tr}(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^T\Big)^{1/2}. \qquad (A.5)
$$

The proof follows by considering (A.1) and (A.2) together with inequalities (A.4) and (A.5).

□

Control of the partial derivatives of the Stieltjes transform of Gram matrices

We recall from Definition 2.9 that the Stieltjes transform of a Gram matrix can be written as a function f of the matrix entries:

$$
f(\mathbf{x}) = \frac{1}{N}\mathrm{Tr}\Big(A(\mathbf{x})-z\mathbf{I}\Big)^{-1},
$$

where

$$
A(\mathbf{x}) = A(\mathbf{x}_1^T,\dots,\mathbf{x}_n^T) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^T .
$$

In this section, we give some upper bounds for the partial derivatives of f and we show how they can be computed. We start by introducing some notations:

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 $A(\mathbf{x})$, f instead of $f(\mathbf{x})$, and define $G = (A - z\mathbf{I})^{-1}$.

Lemma A.2. *Let $\mathbf{x}$ be a vector of $\mathbb{R}^{nN}$ with coordinates*

$$\mathbf{x} = \left(\mathbf{x}_1^T, \dots, \mathbf{x}_n^T\right) \text{ where for any } i \in \{1,\dots,n\},\ \mathbf{x}_i = \left(x_k^{(i)},\ k \in \{1,\dots,N\}\right)^T.$$

Let $z = u + \sqrt{-1}v \in \mathbb{C}^+$ and $f := f_z$ be the function defined in (2.37). Then, for any $i \in \{1,\dots,n\}$ and any $j,k,\ell,m \in \{1,\dots,N\}$, the following inequalities hold true:

$$|\partial^2_{mj} f(\mathbf{x})| \leq \frac{8}{v^3 n^2 N} \sum_{r=1}^{N} |x_r^{(i)}|^2 + \frac{2}{v^2 n N}, \tag{A.6}$$

$$|\partial^3_{\ell mj} f(\mathbf{x})| \leq \frac{48}{v^4 n^3 N} \left(\sum_{r=1}^{N} |x_r^{(i)}|^2\right)^{3/2} + \frac{24}{v^3 n^2 N} \left(\sum_{r=1}^{N} |x_r^{(i)}|^2\right)^{1/2}. \tag{A.7}$$

Proof. To prove the lemma, we shall proceed as in Theorem 1.3 of Chatterjee [20] but with some modifications since his computations are made in case where $A(\mathbf{x})$ is a Wigner matrix of order N.

Note that $\partial_j A$ is the matrix with

$$n^{-1}\left(x_1^{(i)}, \dots, x_{j-1}^{(i)}, 2x_j^{(i)}, x_{j+1}^{(i)}, \dots, x_N^{(i)}\right)$$

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:

$$\|\partial_j A\|_2 = \frac{1}{n}\left(2 \sum_{k=1\,,k\neq j}^{N} |x_k^{(i)}|^2 + 4|x_j^{(i)}|^2\right)^{1/2} \leq \frac{2}{n}\left(\sum_{k=1}^{N} |x_k^{(i)}|^2\right)^{1/2}. \tag{A.8}$$

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,

$$\|\partial^2_{mj} A\|_2 \leq \frac{2}{n}. \tag{A.9}$$

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

Now, by using (2.38), it follows that, for any $j \in \{1, \ldots, N\}$,

$$\partial_j f = -\frac{1}{N}\mathrm{Tr}(G(\partial_j A)G)\,. \tag{A.10}$$

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 $\{j,m\}$, $\{j,m,\ell\}$ and $\{j,m,\ell,k\}$. Therefore the first sum consists of 2 terms, the second one of 6 terms and the last one of 24 terms. Starting from (A.10) and applying repeatedly (2.38), we then derive the following cumbersome formulas for the partial derivatives up to the order four: for any $j,m,\ell,k \in \{1,\ldots,N\}$,

$$\begin{aligned}
\partial^2_{mj} f &= \frac{1}{N}\sum_{\{j',m'\}=\{j,m\}} \mathrm{Tr}\Big(G(\partial_{j'}A)G(\partial_{m'}A)G\Big) - \frac{1}{N}\mathrm{Tr}\Big(G(\partial^2_{mj}A)G\Big)\,, && \text{(A.11)}\\
\partial^3_{\ell mj} f &= -\frac{1}{N}\sum_{\{j',m',\ell'\}=\{j,m,\ell\}} \mathrm{Tr}\Big(G(\partial_{j'}A)G(\partial_{m'}A)G(\partial_{\ell'}A)G\Big) \\
&\quad +\frac{1}{N}\sum_{\{j',m'\}=\{j,m\}} \mathrm{Tr}\Big(G(\partial^2_{\ell j'}A)G(\partial_{m'}A)G + G(\partial_{j'}A)G(\partial^2_{\ell m'}A)G\Big) \\
&\quad +\frac{1}{N}\mathrm{Tr}\Big(G(\partial_\ell A)G(\partial^2_{mj}A)G\Big) + \frac{1}{N}\mathrm{Tr}\Big(G(\partial^2_{mj}A)G(\partial_\ell A)G\Big)\,. && \text{(A.12)}
\end{aligned}$$

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

$$|\mathrm{Tr}(G(\partial^2_{mj}A)G)| = |\mathrm{Tr}((\partial^2_{mj}A)G^2)| \le 2v^{-2}n^{-1}\,. \tag{A.13}$$

Next, to give an upper bound for $|\mathrm{Tr}\Big(G(\partial_j A)G(\partial_m A)G\Big)|$, we shall recall some properties of the Hilbert-Schmidt norm: Let $\mathbf{B}$ and $\mathbf{C}$ be two $N \times N$ complex matrices in $\mathcal{L}_2$, the set of Hilbert-Schmidt operators. Then

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

b. If $\mathbf{B}$ admits a spectral decomposition with eigenvalues $\lambda_1, \ldots, \lambda_N$, then

$$\max\{\|\mathbf{BC}\|_2, \|\mathbf{CB}\|_2\} \le \max_{1\le i\le N} |\lambda_i|.\|\mathbf{C}\|_2\,.$$

For a proof of these facts, one can check pages 55-58 in [76].

Using the properties of the Hilbert-Schmidt norm recalled above, the fact that the eigenvalues of G are all bounded by v^{-1}, and (A.8), we then derive that

$$\begin{aligned}
|\mathrm{Tr}(G(\partial_j A)G(\partial_m A)G)| &\leq \|G(\partial_j A)G\|_2.\|(\partial_m A)G\|_2 \\
&\leq \|G\|.\|(\partial_j A)G\|_2.\|\partial_m A\|_2.\|G\| \\
&\leq \|G\|^3.\|\partial_j A\|_2.\|\partial_m A\|_2 \\
&\leq \frac{4}{v^3 n^2}\sum_{k=1}^{N}\left|x_k^{(i)}\right|^2 . \qquad \text{(A.14)}
\end{aligned}$$

Starting from (A.11) and considering (A.13) and (A.14), the first inequality of Lemma A.2 follows.

Next, using again the above properties a. and b., the fact that the eigenvalues of G are all bounded by v^{-1}, (A.8) and (A.9), we get that

$$\begin{aligned}
|\mathrm{Tr}(G(\partial_j A)G(\partial_m A)G(\partial_\ell A)G)| &\leq \|G(\partial_j A)G(\partial_m A)G\|_2.\|(\partial_\ell A)G\|_2 \\
&\leq \|G(\partial_j A)G(\partial_m A)\|_2.\|G\|^2.\|\partial_\ell A\|_2 \\
&\leq \|G(\partial_j A)\|_2.\|G(\partial_m A)\|_2.\|G\|^2.\|\partial_\ell A\|_2 \\
&\leq \|G\|^4.\|\partial_j A\|_2.\|\partial_m A\|_2.\|\partial_\ell A\|_2 \\
&\leq \frac{8}{v^4 n^3}\left(\sum_{k=1}^{N}\left|x_k^{(i)}\right|^2\right)^{3/2} , \qquad \text{(A.15)}
\end{aligned}$$

and

$$\begin{aligned}
|\mathrm{Tr}(G(\partial^2_{\ell j} A)G(\partial_m A)G)| &\leq \|G(\partial^2_{\ell j} A)G\|_2.\|(\partial_m A)G\|_2 \\
&\leq \|G\|^2\|G(\partial^2_{\ell j} A)\|_2.\|\partial_m A\|_2 \\
&\leq \|G\|^3.\|\partial^2_{\ell j} A\|_2.\|\partial_m A\|_2 \\
&\leq \frac{4}{v^3 n^2}\left(\sum_{k=1}^{N}|x_k^{(i)}|^2\right)^{1/2} . \qquad \text{(A.16)}
\end{aligned}$$

The same last bound is obviously valid for $|\mathrm{Tr}(G(\partial_m A)G(\partial^2_{\ell j} A)G)|$. Hence, starting from (A.12) and considering (A.15) and (A.16), the second inequality of Lemma A.2 follows.

A.2 On the Stieltjes transform of symmetric matrices

The following lemma is proved by Götze *et al.* allowing us to compare the Stieltjes transforms of two symmetric matrices.

Lemma A.3. *(Lemma 2.1, [30]) Let $\mathbf{A}_n$ and $\mathbf{B}_n$ be two symmetric $n \times n$ matrices. Then, for any $z \in \mathbb{C}\backslash\mathbb{R}$,*

$$\Big|S_{\mathbb{A}_n}(z) - S_{\mathbb{B}_n}(z)\Big|^2 \leq \frac{1}{n^2|\operatorname{Im}(z)|^4}\operatorname{Tr}\Big((\mathbf{A}_n - \mathbf{B}_n)^2\Big)\,,$$

where $\mathbb{A}_n = n^{-1/2}\mathbf{A}_n$ and $\mathbb{B}_n = n^{-1/2}\mathbf{B}_n$.

On the partial derivatives of the Stieltjes transform of symmetric matrices

All along this section, we shall also use the fact that the Stieltjes transform of symmetric matrices is a smooth function of its entries. Indeed, let $N = n(n+1)/2$ and write elements of $\mathbb{R}^N$ as $\mathbf{x} = (x_{ij})_{1 \leq j \leq i \leq n}$. For any $z \in \mathbb{C}^+$, let $f := f_z$ be the function defined from $\mathbb{R}^N$ to $\mathbb{C}$ by

$$f(\mathbf{x}) = \frac{1}{n}\operatorname{Tr}(A(\mathbf{x}) - z\mathbf{I}_n)^{-1} \quad \text{for any } \mathbf{x} \in \mathbb{R}^N\,, \tag{A.17}$$

where $\mathbf{I}_n$ is the identity matrix of order n and $A(\mathbf{x})$ is the matrix defined

$$(A(\mathbf{x}))_{ij} = \frac{1}{\sqrt{n}}\begin{cases} x_{i,j} & \text{if } i \geq j \\ x_{j,i} & \text{if } i < j \end{cases}$$

The function f admits partial derivatives of all orders. In particular, denoting for any $\mathbf{u} \in \{(i,j)\,, 1 \leq j \leq i \leq n\}$, $\partial_{\mathbf{u}} f$ for $\partial f_n/\partial x_{\mathbf{u}}$, the following upper bounds hold: for any $z = x + iy \in \mathbb{C}^+$ and any $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\{(i,j)\,, 1 \leq j \leq i \leq n\}$,

$$|\partial_{\mathbf{u}} f| \leq \frac{2}{y^2 n^{3/2}}\,,\ |\partial_{\mathbf{u}}\partial_{\mathbf{v}} f| \leq \frac{4}{y^3 n^2} \ \text{ and } \ |\partial_{\mathbf{u}}\partial_{\mathbf{v}}\partial_{\mathbf{w}} f| \leq \frac{3 \times 2^{5/2}}{y^4 n^{5/2}}\,. \tag{A.18}$$

One can check the equalities (20) and (21) in [20] together with the computations on pages 2074-2075 of the same paper.

We state now the following lemma by Merlevède and Peligrad [45] allowing us to control the second order partial derivative of f:

Lemma A.4. *(Lemma 13, [45]) Let $z = x + iy \in \mathbb{C}^+$ and $f := f_z$ be defined by (A.17). Let $\{a_{ij}\,,\,1 \le j \le i \le n\}$ and $\{b_{ij}\,,\,1 \le j \le i \le n\}$ be two families of real numbers. Then, for any subset $\mathcal{I}_n$ of $\{(i,j)\,,1 \le j \le i \le n\}$ and any element $\mathbf{x}$ of $\mathbb{R}^N$,*

$$\Big|\sum_{\mathbf{u}\in\mathcal{I}_n}\sum_{\mathbf{v}\in\mathcal{I}_n} a_{\mathbf{u}} b_{\mathbf{v}}\,\partial_{\mathbf{u}}\partial_{\mathbf{v}} f(\mathbf{x})\Big| \le \frac{2}{y^3 n^2}\Big(\sum_{\mathbf{u}\in\mathcal{I}_n} a_{\mathbf{u}}^2 \sum_{\mathbf{v}\in\mathcal{I}_n} b_{\mathbf{v}}^2\Big)^{1/2}.$$

A.2.1 On the Gaussian interpolation technique

Next lemma is a consequence of the well-known Gaussian interpolation trick.

Lemma A.5. *Let $(Y_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ and $(Z_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ be two centered real-valued Gaussian processes. Let $\mathbf{Y}_n$ be the symmetric random matrix of order n defined by*

$$(\mathbf{Y}_n)_{ij} = \begin{cases} Y_{i,j} & \text{if } 1 \le j \le i \le n \\ Y_{j,i} & \text{if } 1 \le i < j \le n\,. \end{cases}$$

Denote $\mathbb{Y}_n = \frac{1}{\sqrt{n}}\mathbf{Y}_n$ and define similarly the symmetric matrix $\mathbb{Z}_n$ associated with $(Z_{k,\ell})_{k,\ell}$. Then, for any $z = x + \mathrm{i}y \in \mathbb{C}^+$,

$$\begin{aligned}&\mathbb{E}(S_{\mathbb{Y}_n}(z)) - \mathbb{E}(S_{\mathbb{Z}_n}(z)) \\ &\qquad = \frac{1}{2}\sum_{1\le\ell\le k\le n}\sum_{1\le j\le i\le n}\int_0^1 \Big(\mathbb{E}(Y_{k,\ell}Y_{i,j}) - \mathbb{E}(Z_{k,\ell}Z_{i,j})\Big)\mathbb{E}\Big(\partial_{k\ell}\partial_{ij} f(\mathbf{u}(t))\Big)dt \end{aligned} \tag{A.19}$$

where, for $t \in [0,1]$,

$$\mathbf{u}(t) = (\sqrt{t}\,Y_{k,\ell} + \sqrt{1-t}\,Z_{k,\ell}\,, 1 \le \ell \le k \le n)\,.$$

Moreover, we have for any $z = x + \mathrm{i}y$,

$$\Big|\mathbb{E}(S_{\mathbb{Y}_n}(z)) - \mathbb{E}(S_{\mathbb{Z}_n}(z))\Big| \le \frac{2}{n^2 y^3}\sum_{1\le\ell\le k\le n}\sum_{1\le j\le i\le n}|\mathbb{E}(Y_{k,\ell}Y_{i,j}) - \mathbb{E}(Z_{k,\ell}Z_{i,j})|\,.$$

Proof. Using the definition of f, we first write

$$\mathbb{E}(S_{\mathbb{Y}_n}(z)) = \mathbb{E} f\Big((Y_{k,\ell})_{1\le\ell\le k\le n}\Big) \text{ and } \mathbb{E}(S_{\mathbb{Z}_n}(z)) = \mathbb{E} f\Big((Z_{k,\ell})_{1\le\ell\le k\le n}\Big).$$

Equality (A.19) then follows from the usual interpolation trick (for an easy reference we cite Talagrand [67] Section 1.3, Lemma 1.3.1.). To obtain the above upper bound, it suffices then to take into account (A.18).

□

A.3 Other useful lemmas

We give now the following well-known lemma:

Lemma A.6. *If* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ *are three random vectors defined on a probability space* $(\Omega, \mathcal{K}, \mathbb{P})$, *such that* $\mathbf{X}$ *is independent of* $\sigma(\mathbf{Z})$ *and* $\sigma(\mathbf{Y}) \subset \sigma(\mathbf{Z})$. *Then, for any measurable function* g *such that* $\|g(\mathbf{X}, \mathbf{Y})\|_1 < \infty$,

$$\mathbb{E}(g(\mathbf{X}, \mathbf{Y})|\mathbf{Z}) = \mathbb{E}(g(\mathbf{X}, \mathbf{Y})|\mathbf{Y}) \ \textit{a.s.} \tag{A.20}$$

A.3.1 On Taylor expansions for functions of random variables

We give now the Taylor expansion for functions of random variables convenient for the Lindeberg method.

Lemma A.7. *Let* f *be a three times differentiable function from* $\mathbb{R}^{d+m}$ *to* $\mathbb{C}$, *with continuous and bounded third order partial derivatives, i.e. there exists a constant* L_3 *such that*

$$|\partial_i \partial_j \partial_k f(\mathbf{x})| \le L_3 \quad \textit{for all } i, j, k \textit{ and } \mathbf{x}.$$

Let $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors defined on a probability space* $(\Omega, \mathcal{K}, \mathbb{P})$ *taking values in* $\mathbb{R}^d$. *Assume that both vectors belong to* $\mathbb{L}^3(\mathbb{R}^d)$, *are centered at expectation and have the same covariance structure. Let* $\mathbf{Z}$ *be a random vector defined on the same probability* *space but taking values in* $\mathbb{R}^m$*. Assume in addition that* $\mathbf{Z}$ *is independent of* $\mathbf{X}$ *and* $\mathbf{Y}$*. Then, for any permutation* $\pi : \mathbb{R}^{d+m} \to \mathbb{R}^{d+m}$*, we have*

$$|\mathbb{E}f(\pi(\mathbf{X},\mathbf{Z})) - \mathbb{E}f(\pi(\mathbf{Y},\mathbf{Z}))| \leq \frac{L_3 d^2}{3}\Big(\sum_{j=1}^{d}\mathbb{E}(|X_j|^3) + \sum_{j=1}^{d}\mathbb{E}(|Y_j|^3)\Big).$$

The proof of this lemma is based on the following Taylor expansion for functions of several variables:

Lemma A.8. *Let* g *be a three times differentiable function from* $\mathbb{R}^p$ *to* $\mathbb{R}$ *with continuous third partial order derivatives and such that*

$$|\partial_i\partial_j\partial_k g(\mathbf{x})| \leq L_3 \text{ for all } i,j,k \text{ and } \mathbf{x}\,.$$

Then, for any $\mathbf{a} = (a_1,\ldots,a_p)$ *and* $\mathbf{b} = (b_1,\ldots,b_p)$ *in* $\mathbb{R}^p$*,*

$$g(\mathbf{b}) - g(\mathbf{a}) = \sum_{k=1}^{p}(b_j - a_j)\partial_j g(\mathbf{0}) + \frac{1}{2}\sum_{j,k=1}^{p}(b_jb_k - a_ja_k)\partial_j\partial_k g(\mathbf{0}) + R_3(\mathbf{a},\mathbf{b}).$$

with

$$|R_3(\mathbf{a},\mathbf{b})| \leq \frac{L_3}{6}\Big(\Big(\sum_{j=1}^{p}|a_j|\Big)^3 + \Big(\sum_{j=1}^{p}|b_j|\Big)^3\Big) \leq \frac{L_3p^2}{6}\Big(\sum_{j=1}^{p}|a_j|^3 + |b_j|^3\Big).$$

Proof of Lemma A.8. We use Taylor expansion of second order for functions with bounded partial derivatives of order three. It is well-known that

$$g(\mathbf{a}) - g(\mathbf{0}_p) = \sum_{j=1}^{p}a_j\partial_j g(\mathbf{0}_p) + \frac{1}{2}\sum_{j,k=1}^{p}a_ja_k\partial_j\partial_k g(\mathbf{0}_p) + R_3(\mathbf{a})\,,$$

where

$$|R_3(\mathbf{a})| \leq \frac{L_3}{6}\Big(\sum_{j=1}^{p}|a_j|\Big)^3 \leq \frac{L_3p^2}{6}\sum_{j=1}^{p}|a_j|^3\,.$$

By writing a similar expression for $g(\mathbf{b}) - g(\mathbf{0}_p)$ and subtracting them the result follows. □

Proof of Lemma A.7. For simplicity of the notation we shall prove it first for $f((\mathbf{X},\mathbf{Z})) -$

$f((\mathbf{Y},\mathbf{Z}))$. We start by applying Lemma A.8 to real and imaginary part of f and obtain

$$\begin{aligned}
&f(\mathbf{X},\mathbf{Z}) - f(\mathbf{Y},\mathbf{Z}) \\
&\qquad = \sum_{j=1}^{d}(X_j - Y_j)\partial_j f(\mathbf{0}_d,\mathbf{Z}) + \frac{1}{2}\sum_{j,k=1}^{d}(X_kX_j - Y_kY_j)\,\partial_j\partial_k f(\mathbf{0}_d,\mathbf{Z}) + R_3\,,
\end{aligned}$$

with

$$|R_3| \le \frac{L_3 d^2}{3}\Big(\sum_{j=1}^{d}|X_j|^3 + \sum_{j=1}^{d}|Y_j|^3\Big)\,.$$

By taking the expected value and taking into account the hypothesis of independence and the fact that $\mathbf{X}$ and $\mathbf{Y}$ are centered at expectations and have the same covariance structure, we obtain, for all $1 \le j \le d$

$$\mathbb{E}((X_j - Y_j)\,\partial_j f(\mathbf{0}_d,\mathbf{Z})) = (\mathbb{E}X_j - \mathbb{E}Y_j)\,\mathbb{E}\partial_j f(\mathbf{0}_d,\mathbf{Z}) = 0$$

and, for all $1 \le k,j \le d$,

$$\mathbb{E}(X_kX_j - Y_kY_j)\partial_j\partial_k f(\mathbf{0}_d,\mathbf{Z}) = (\mathbb{E}(X_kX_j) - \mathbb{E}(Y_kY_j))\mathbb{E}\partial_j\partial_k f(\mathbf{0}_d,\mathbf{Z}) = 0\,.$$

It follows that

$$\mathbb{E}f(\mathbf{X},\mathbf{Z}) - \mathbb{E}f(\mathbf{Y},\mathbf{Z}) = \mathbb{E}(R_3)\,,$$

with

$$|\mathbb{E}(R_3)| \le \frac{L_3 d^2}{3}\Big(\sum_{j=1}^{d}\mathbb{E}(|X_j|^3) + \sum_{j=1}^{d}\mathbb{E}(|Y_j|^3)\Big)\,.$$

It remains to note that the result remains valid for any permutation of variables $(\mathbf{X},\mathbf{Z})$. The variables in $\mathbf{X},\mathbf{Z}$ can hold any positions among the variables in function f since we just need all the derivatives of order three to be uniformly bounded. The difference in the proof consists only in re-denoting the partial derivatives; for instance instead of ∂_j we shall use ∂_{k_j} where k_j, $1 \le k_j \le d+m$ denotes the index of the variable X_j in $f(x_1, x_2, ..., x_{d+m})$.

$\square$

A.3.2 On the behavior of the Stieltjes transform of some Gaussian matrices

We provide next a technical lemma on the behavior of the expected value of Stieltjes transform of symmetric matrices with Gaussian entries. In Lemma A.9 and Proposition A.10 below, we consider a stationary real-valued centered Gaussian random field $(G_{k,\ell})_{(k,\ell)\in\mathbb{Z}^2}$ with covariance function given by: for any $(k,\ell)\in\mathbb{Z}^2$ and any $(i,j)\in\mathbb{Z}^2$,

$$\mathbb{E}(G_{k,\ell}G_{i,j}) = \gamma_{k-i,\ell-j}\,,$$

satisfying (3.7) and (3.8). We define then two symmetric matrices of order n, $\mathbb{G}_n = n^{-1/2}[g_{k,\ell}]_{k,\ell=1}^n$ and $\mathbb{W}_n = n^{-1/2}[W_{k,\ell}]_{k,\ell=1}^n$ where the entries $g_{k,\ell}$ and $W_{k,\ell}$ are defined respectively by

$$g_{k,\ell} = G_{\max(k,\ell),\min(k,\ell)} \quad \text{and} \quad W_{k,\ell} = \frac{1}{\sqrt{2}}\Big(G_{k,\ell}+G_{\ell,k}\Big)\,.$$

Lemma A.9. *For any $z\in\mathbb{C}\backslash\mathbb{R}$ the following convergence holds:*

$$\lim_{n\to\infty}\Big|\mathbb{E}(S_{\mathbb{G}_n}(z)) - \mathbb{E}(S_{\mathbb{W}_n}(z))\Big| = 0\,.$$

As a consequence of this lemma and Theorem 2 in [40], we obtain the following result concerning the limiting spectral distribution of both $\mathbb{G}_n$ and $\mathbb{W}_n$.

Proposition A.10. *For any $z\in\mathbb{C}\backslash\mathbb{R}$, $S_{\mathbb{G}_n}(z)$ and $S_{\mathbb{W}_n}(z)$ have almost surely the same limit, $S(z)$, defined by the relations* (3.9) *and* (3.10)*.*

Proof of Lemma A.9. According to Lemma A.5, for any $z\in\mathbb{C}\backslash\mathbb{R}$,

$$\begin{aligned}&\Big|\mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{W}_n}(z)\Big)\Big| \\ &\qquad\qquad \leq \frac{2}{n^2|\operatorname{Im}(z)|^3}\sum_{1\leq\ell\leq k\leq n}\ \sum_{1\leq j\leq i\leq n}|\mathrm{Cov}(G_{k,\ell},G_{i,j}) - \mathrm{Cov}(W_{k,\ell},W_{i,j})|\,.\end{aligned}$$

Taking into account (3.8), we get

$$\mathbb{E}(W_{k,\ell}W_{i,j}) = \gamma_{k-i,\ell-j} + \gamma_{k-j,\ell-i}\,. \tag{A.21}$$

Hence,

$$\left|\mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{W}_n}(z)\Big)\right| \le \frac{2}{n^2|\operatorname{Im}(z)|^3} \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} |\gamma_{k-j,\ell-i}| \,.$$

Using (3.8) and noticing that by stationarity $\gamma_{u,v} = \gamma_{-u,-v}$ for any $(u,v)\in\mathbb{Z}^2$, we get

$$\sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le n} |\gamma_{k-j,\ell-i}| \le 2 \sum_{1\le \ell\le k\le n} \sum_{1\le j\le i\le k} |\gamma_{k-j,\ell-i}| \,.$$

Let m_n be any positive integer less than n and write the following decomposition

$$\begin{aligned}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\sum_{i=1}^{k}\sum_{j=1}^{i}|\gamma_{k-j,\ell-i}| \le{}& \sum_{k=1}^{m_n}\sum_{\ell=1}^{k}\sum_{i=1}^{k}\sum_{j=1}^{k}|\gamma_{k-j,\ell-i}| \\ &+ \sum_{k=m_n+1}^{n}\sum_{\ell=1}^{k}\sum_{i=1}^{k}\sum_{j=1}^{i-m_n}|\gamma_{k-j,\ell-i}| + \sum_{k=m_n+1}^{n}\sum_{\ell=1}^{k}\sum_{i=1}^{k}\sum_{j=i-m_n+1}^{i}|\gamma_{k-j,\ell-i}| \,,\end{aligned}$$

with the convention that $\sum_{u=k}^{\ell} = 0$ if $k>\ell$. Straightforward computations lead to

$$\begin{aligned}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\sum_{i=1}^{k}\sum_{j=1}^{i}|\gamma_{k-j,\ell-i}| \le{}& m_n^2\sum_{p=0}^{m_n}\sum_{q=-m_n}^{m_n}|\gamma_{k-j,\ell-i}| \\ &+ n^2\sum_{p=m_n}^{n}\sum_{q=-n}^{n}|\gamma_{k-j,\ell-i}| + n\times m_n\sum_{p=0}^{n}\sum_{q=-m_n}^{m_n}|\gamma_{k-j,\ell-i}| \,.\end{aligned}$$

So overall, for any $z\in\mathbb{C}\backslash\mathbb{R}$,

$$\left|\mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big) - \mathbb{E}\Big(S_{\mathbb{W}_n}(z)\Big)\right| \le \frac{4}{|\operatorname{Im}(z)|^3}\bigg(\frac{2m_n}{n}\sum_{p=0}^{n}\sum_{q=-n}^{n}|\gamma_{p,q}| + \sum_{p\ge m_n}\sum_{q\in\mathbb{Z}}|\gamma_{p,q}|\bigg) \,,$$

for any $z\in\mathbb{C}\backslash\mathbb{R}$. The lemma then follows by taking into account (3.7) and by selecting m_n such that $m_n\to\infty$ and $m_n/n\to 0$.

□

Proof of Proposition A.10. The Borel-Cantelli lemma together with Theorem 17.1.1 in [56] imply that, for any $z\in\mathbb{C}\backslash\mathbb{R}$,

$$\lim_{n\to\infty}|S_{\mathbb{G}_n}(z) - \mathbb{E}\Big(S_{\mathbb{G}_n}(z)\Big)| = 0 \text{ and } \lim_{n\to\infty}|S_{\mathbb{W}_n}(z) - \mathbb{E}\Big(S_{\mathbb{W}_n}(z)\Big)| = 0 \text{ a.s.}$$

Therefore, the proposition follows by Lemma A.9 combined with Theorem 2 in [40] applied to $\mathbb{E}\Big(S_{\mathbb{W}_n}(z)\Big)$. Indeed the entries $(W_{k,\ell})_{1\le k,\ell\le n}$ of the matrix $n^{1/2}\mathbb{W}_n$ form a *symmetric* real-valued centered Gaussian random field whose covariance function satisfies (A.21). Hence relation (2.8) in [40] holds. In addition, by (3.7) , condition (2.9) in [40] is also satisfied. At this step, the reader should notice that Theorem 2 in [40] also requires additional conditions on the covariance function $\gamma_{k,\ell}$ (this function is denoted by $B(k,\ell)$ in this latter paper), namely $\gamma_{k,\ell} = \gamma_{\ell,k} = \gamma_{\ell,-k}$. In our case, the first holds (this is (3.8)) but not necessarily $\gamma_{\ell,k} = \gamma_{\ell,-k}$ since by stationarity we only have $\gamma_{\ell,k} = \gamma_{-\ell,-k}$. However a careful analysis of the proof of Theorem 2 in [40] (and in particular of their auxiliary lemmas) or of the proof of Theorem 17.2.1 in [56], shows that the only condition required on the covariance function to derive the limiting equation of the Stieljes transform is the absolute summability condition (2.9) in [40]. It is noteworthy to indicate that, in Theorem 2 of [40], the symmetry conditions on the covariance function $\gamma_{k,\ell}$ must only translate the fact that the entries of the matrix form a stationary *symmetric* real-valued centered Gaussian random field, so $\gamma_{k,\ell}$ has only to satisfy $\gamma_{k,\ell} = \gamma_{\ell,k} = \gamma_{-\ell,-k}$ for any $(k,\ell) \in \mathbb{Z}^2$.

$\square$

A.4 On operator functions

As matrices do not commute, certain tools available in the scalar setting cannot be straightforward extended. We shall recall in the first part of this section some preliminaries on the matrix exponential and logarithm.

A.4.1 On the matrix exponential and logarithm

We denote by $\mathbf{0}$ the zero matrix and by $\mathbb{I}_d$ the $d \times d$ identity matrix. For a symmetric matrix $\mathbf{A}$, we use the curly inequalities to denote the semi-definite ordering and we write $\mathbf{0} \preceq \mathbf{A}$ if $\mathbf{A}$ is positive semi-definite and $\mathbf{0} \prec \mathbf{A}$ if $\mathbf{A}$ is positive definite.

We recall that the matrix exponential is the operator mapping defined for any $d \times d$ self-adjoint matrix $\mathbf{A}$ by

$$\mathbf{A} \mapsto e^{\mathbf{A}} := \mathbb{I}_d + \sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k!}$$

and the matrix logarithm is defined as its inverse; i.e. $\log(e^{\mathbf{A}}) := \mathbf{A}$. The matrix logarithm is operator monotone. Namely, for any two self-adjoint matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{0} \prec \mathbf{A} \preceq \mathbf{B} \quad \Longrightarrow \quad \log(\mathbf{A}) \preceq \log(\mathbf{B})\,.$$

Moreover, we note that the matrix logarithm is operator concave: for any $t \in [0,1]$ and any self-adjoint matrices $\mathbf{A}$ and $\mathbf{B}$

$$t\log(\mathbf{A}) + (1-t)\log(\mathbf{B}) \preceq \log(t\mathbf{A} + (1-t)\mathbf{B})\,.$$

We also note that the matrix exponential is positive but it is neither operator monotone nor convex. However, the Trace exponential function $\mathbf{A} \mapsto \operatorname{Tr}\exp(\mathbf{A})$ is so. More precisely, for any self-adjoint matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A} \preceq \mathbf{B} \quad \Longrightarrow \quad \operatorname{Tr}\exp(\mathbf{A}) \leq \operatorname{Tr}\exp(\mathbf{B})$$

and for any $t \in [0,1]$,

$$\operatorname{Tr}\exp(t\mathbf{A} + (1-t)\mathbf{B}) \leq t\operatorname{Tr}\exp(\mathbf{A}) + (1-t)\operatorname{Tr}\exp(\mathbf{B})\,.$$

For more details on this issue, we refer to [12] and [35]. Having recalled the above functions, we can introduce now the following theorem, due to Lieb, on the concavity of the trace.

Theorem A.11. *(Theorem 6, [42]) Let* $\mathbf{B}$ *be a self-adjoint matrix. The operator mapping*

$$\mathbf{A} \mapsto \operatorname{Tr}\exp(\mathbf{B} + \log(\mathbf{A}))$$

is concave on the positive definite cone.

The fact that the matrix exponential does not convert sums into products unless the matrices commute makes many properties of the exponential function inextensible to the matrix setting. However, the following Golden-Thompson inequality allows us to compare the Trace exponential function of the sum of *two* matrices by the trace of the product. More precisely,

Lemma A.12. *(Golden-Thompson inequality, [29], [69]) Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two self-adjoint matrices, then*

$$\mathrm{Tr}\Big(\mathrm{e}^{\mathbf{A}+\mathbf{B}}\Big) \leq \mathrm{Tr}\Big(\mathrm{e}^{\mathbf{A}}\mathrm{e}^{\mathbf{B}}\Big)\,.$$

We note that this lemma *fails* for more than two matrices (see [12]).

We recall now some properties on the expectation of matrices. The trace and the expectation commute and for any random matrices $\mathbf{A}$ and $\mathbf{B}$ that are independent,

$$\mathbb{E}(\mathbf{AB}) = \mathbb{E}(\mathbf{A})\mathbb{E}(\mathbf{B})\,.$$

Moreover, the expectation preserves the semi-definite ordering:

$$\mathbf{A} \preceq \mathbf{B} \quad \Longrightarrow \quad \mathbb{E}(\mathbf{A}) \preceq \mathbb{E}(\mathbf{B})\,,$$

and every operator convex function admits an operator Jensen's inequality (see [35]). In particular, applying Jensen's inequality for the Trace exponential function yields

$$\mathrm{Tr}\exp(\mathbb{E}\mathbf{A}) \leq \mathbb{E}\mathrm{Tr}\exp(\mathbf{A})\,.$$

A.4.2 On the Matrix Laplace Transform

In this subsection, we collect some technical preliminary lemmas that are necessary for the proof in Section 5.3.

We start by the following lemma which is due to Tropp [70] and controls the matrix Laplace transform of the sum of *independent* centered self-adjoint matrices having uniformly bounded largest eigenvalues.

Lemma A.13. *([70]) Let* K *be a finite subset of positive integers. Consider a family* $(\mathbb{U}_k)_{k\in K}$ *of* $d \times d$ *self-adjoint random matrices that are mutually independent. Assume that for any* $k \in K$*,*

$$\mathbb{E}(\mathbb{U}_k) = \mathbf{0} \quad \textit{and} \quad \lambda_{\max}(\mathbb{U}_k) \leq B \textit{ a.s.}$$

where B *is a positive constant. Then for any* $t > 0$*,*

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k\in K}\mathbb{U}_k\Big) \leq d\exp\Big(t^2 g(tB)\lambda_{\max}\Big(\sum_{k\in K}\mathbb{E}(\mathbb{U}_k^2)\Big)\Big)\,, \tag{A.22}$$

where $g(x) = x^{-2}(\mathrm{e}^x - x - 1)$.

Under the form stated above, it is a combination of (1.13) and the following lemma that shows how the bound on the largest eigenvalue serves to control the Laplace transform of a zero-mean random matrix.

Lemma A.14. *(Lemma 6.7, [70]) Let* $\mathbb{U}$ *be a self-adjoint matrix satisfying*

$$\mathbb{E}\mathbb{U} = \mathbf{0} \quad \text{and} \quad \lambda_{\max}(\mathbb{U}) \leq 1 \text{ a.s.}$$

Then

$$\mathbb{E}\mathrm{e}^{t\mathbb{U}} \preceq \exp\Big((\mathrm{e}^t - t - 1) \cdot \mathbb{E}(\mathbb{U}^2)\Big).$$

Proof of Lemma A.13. By the above lemma, we have for any $k \in K$,

$$\mathbb{E}\mathrm{e}^{t\mathbb{U}_k} \preceq \exp\Big(\frac{1}{B^2}(\mathrm{e}^{tB} - tB - 1) \cdot \mathbb{E}(\mathbb{U}_k^2)\Big) = \exp\Big(t^2 g(tB) \cdot \mathbb{E}(\mathbb{U}_k^2)\Big).$$

Noting that the Trace exponential and the matrix logarithm are operator increasing, we plug the above bound in (1.13) and get

$$\begin{aligned}
\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k\in K}\mathbb{U}_k\Big) &\leq \mathrm{Tr}\exp\Big(\sum_{k\in K}\log\mathbb{E}\mathrm{e}^{t\mathbb{U}_k}\Big)\\
&\leq \mathrm{Tr}\exp\Big(t^2 g(tB)\sum_{k\in K}\mathbb{E}(\mathbb{U}_k^2)\Big)\\
&\leq d\exp\Big(t^2 g(tB)\lambda_{\max}\Big(\sum_{k\in K}\mathbb{E}(\mathbb{U}_k^2)\Big)\Big),
\end{aligned}$$

where the last inequality follows from the fact the trace is bounded by d times the largest eigenvalue. □

We restate the following lemma and give its proof:

Lemma A.15. *Let* $\mathbb{U}_0, \mathbb{U}_1, \dots$ *be a sequence of* $d \times d$ *self-adjoint random matrices. Assume that there exist positive constants* $\sigma_0, \sigma_1, \dots$ *and* $\kappa_0, \kappa_1, \dots$ *such that, for any* $i \geq 0$ *and any* t *in* $[0, 1/\kappa_i[$,

$$\log\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\,\mathbb{U}_i}\Big) \leq C_d + (\sigma_i t)^2/(1 - \kappa_i t),$$

where C_d is a positive constant depending only on d. Then, for any positive m and any t in $[0, 1/(\kappa_0 + \kappa_1 + \cdots + \kappa_m)[$,

$$\log \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{k=0}^{m}\mathbb{U}_k\Big) \leq C_d + (\sigma t)^2/(1-\kappa t),$$

where $\sigma = \sigma_0 + \sigma_1 + \cdots + \sigma_m$ and $\kappa = \kappa_0 + \kappa_1 + \cdots + \kappa_m$.

Proof. For any $i = 1, \ldots, m$, we define the functions γ_i by

$$\gamma_i(t) = (\sigma_i t)^2/(1-\kappa_i t) \ \text{ for } t \in [0, 1/\kappa_i[\ \text{ and } \gamma_i(t) = +\infty \ \text{ for } t \geq 1/\gamma_i\,.$$

The proof follows by induction on m. Considering the case $m = 1$, we let for any $t \geq 0$,

$$L(t) = \log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t(\mathbb{U}_0+\mathbb{U}_1)}\Big)\,,$$

and notice that by the Golden-Thompson inequality and the monotony of the matrix logarithm,

$$L(t) \leq \log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\,\mathbb{U}_0}\mathrm{e}^{t\,\mathbb{U}_1}\Big)\,. \tag{A.23}$$

We recall now the non-commutative Hölder inequality: *if $\mathbf{A}$ and $\mathbf{B}$ are $d\times d$ self-adjoint random matrices then, for any $1 \leq p, q \leq \infty$ with $p^{-1} + q^{-1} = 1$,*

$$|\mathrm{Tr}(\mathbf{AB})| \leq \|\mathbf{A}\|_{\mathbb{S}^p}\|\mathbf{B}\|_{\mathbb{S}^q}\,, \tag{A.24}$$

where

$$\|\mathbf{A}\|_{\mathbb{S}^p} = \Big(\sum_{i=1}^{d}|\lambda_i(\mathbf{A})|^p\Big)^{1/p} \quad \textit{and} \quad \|\mathbf{B}\|_{\mathbb{S}^q} = \Big(\sum_{i=1}^{d}|\lambda_i(\mathbf{B})|^q\Big)^{1/q}$$

are respectively the p-Schatten and q-Schatten norms of $\mathbf{A}$ and $\mathbf{B}$. We refer on this issue to exercise 1.3.9 in [68].

Starting from (A.23) and applying (A.24) with $A = \mathrm{e}^{t\,\mathbb{U}_0}$ and $B = \mathrm{e}^{t\,\mathbb{U}_0}$, we derive that for any $t > 0$ and any $p \in]1, \infty[$

$$L(t) \leq \log \mathbb{E}\Big(\|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathbb{S}^p}\|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathbb{S}^q}\Big)\,,$$

which gives by applying Hölder's inequality

$$L(t) \leq p^{-1} \log \mathbb{E} \|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathbb{S}^p}^p + q^{-1} \log \mathbb{E} \|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathbb{S}^q}^q \,.$$

Observe now that since $\mathbb{U}_0$ is self-adjoint

$$\|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathbb{S}^p}^p = \sum_{i=1}^d |\lambda_i(\mathrm{e}^{t\,\mathbb{U}_0})|^p = \sum_{i=1}^d \lambda_i(\mathrm{e}^{tp\,\mathbb{U}_0}) = \mathrm{Tr}\Big(\mathrm{e}^{tp\,\mathbb{U}_0}\Big)\,,$$

and similarly $\|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathbb{S}^q}^q = \mathrm{Tr}\Big(\mathrm{e}^{tq\,\mathbb{U}_1}\Big)$. So, overall,

$$L(t) \leq p^{-1} \log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{tp\,\mathbb{U}_0}\Big) + q^{-1} \log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{tq\,\mathbb{U}_1}\Big)\,. \tag{A.25}$$

For any t in $[0, 1/\kappa[$, take

$$u_t = \frac{\sigma_0}{\sigma}(1-\kappa t) + \kappa_0 t\,,$$

where we note that in this case, $\kappa = \kappa_0 + \kappa_1$ and $\sigma = \sigma_0 + \sigma_1$. With this choice

$$1 - u_t = \frac{\sigma_1}{\sigma}(1-\kappa t) + \kappa_1 t\,,$$

so that u_t belongs to $]0, 1[$. Applying Inequality (A.25) with $p = 1/u_t$, we get for any t in $[0, 1/\kappa[$,

$$\begin{aligned} L(t) &\leq C_d + u_t \gamma_0\Big(\frac{t}{u_t}\Big) + (1-u_t)\gamma_1\Big(\frac{t}{1-u_t}\Big) \\ &= C_d + \frac{\sigma_0^2 t^2}{u_t - \kappa_0 t} + \frac{\sigma_1^2 t^2}{1 - u_t - \kappa_1 t} \\ &= C_d + \frac{(\sigma t)^2}{1-\kappa t}\,, \end{aligned}$$

which completes the proof of Lemma A.15. □

A.4.3 Berbee's Coupling Lemmas

In this Section, we recall Berbee's maximal and classical coupling lemmas [11]. We start by the former one for random variables satisfying a β-mixing condition:

Lemma A.16. *(Berbee's maximal coupling lemma, [11]) Let $\mathcal{A}$ be a σ-field in $(\Omega, \mathcal{A}, \mathbb{P})$ and X be a random variable with values in some Polish space. Let δ be a random variable with uniform distribution over $[0,1]$, independent of the σ-field generated by X and $\mathcal{A}$. Then there exists a random variable X^* , with the same law as X, independent of X, such that*

$$\mathbb{P}(X \neq X^*) = \beta(\mathcal{A}, \sigma(X)) .$$

Furthermore, X^* is measurable with respect to the σ-field generated by $\mathcal{A}$ and (X, δ).

We state now Berbee's classical coupling lemma [11]:

Lemma A.17. *(Berbee's classical coupling lemma, [11]) Let X and Y be two random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in Borel spaces B_1 and B_2 respectively. Assume that $(\Omega, \mathcal{A}, \mathbb{P})$ is rich enough to contain a random variable δ with uniform distribution over $[0,1]$ independent of (X, Y). Then there exists a random variable $Y^* = f(X, Y, \delta)$, where f is a measurable function from $B_1 \times B_2 \times [0,1]$ into B_2, such that Y^* is independent of X, has the same distribution as Y and*

$$\mathbb{P}(Y \neq Y^*) = \beta(\sigma(X), \sigma(Y)) .$$

We finally note that the β-mixing coefficient $\beta(\sigma(X), \sigma(Y))$ has the following equivalent definition:

$$\beta(\sigma(X), \sigma(Y)) = \frac{1}{2} \| P_{X,Y} - P_X \otimes P_Y \| , \tag{A.26}$$

where $P_{X,Y}$ is the joint distribution of (X, Y) and P_X and P_Y are respectively the distributions of X and Y and, for two positive measures μ and ν, the notation $\|\mu - \nu\|$ denotes the total variation of $\mu - \nu$.

Bibliography

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:1000–1034, 2008. 18

[2] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on*, 48(3):569–579, 2002. 15, 18, 20

[3] G. W. Anderson and O. Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008. 2, 53, 66

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

[5] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition, 2010. 2, 37, 45, 56, 75, 82, 89, 91, 98

[6] Z. Bai and W. Zhou. Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, 18(2):425–442, 2008. 6, 7, 8, 26, 30

[7] M. Banna. Limiting spectral distribution of gram matrices associated with functionals of β-mixing processes. *Journal of Mathematical Analysis and Applications*, 433(1):416 – 433, 2016.

[8] M. Banna and F. Merlevède. Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *Journal of Theoretical Probability*, pages 1–39, 2013. 108

[9] M. Banna, F. Merlevède, and M. Peligrad. On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Processes and their Applications*, 125(7):2700 – 2726, 2015.

[10] M. Banna, F. Merlevède, and P. Youssef. Bernstein type inequality for a class of dependent random matrices. *arXiv preprint arXiv:1504.05834*, 2015.

[11] H. C. P. Berbee. *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1979. 13, 99, 104, 163, 164

[12] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. 159, 160

[13] S. Borovkova, R. Burton, and H. Dehling. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.*, 353(11):4261–4318, 2001. 93

[14] A. Boutet de Monvel and A. Khorunzhy. On the norm and eigenvalue distribution of large random matrices. *Ann. Probab.*, 27(2):913–944, 1999. 53, 57

[15] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Process. Related Fields*, 2(4):607–636, 1996. 59, 60

[16] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144, 2005. Update of, and a supplement to, the 1986 original. 92

[17] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012. 146

[18] A. Chakrabarty, R. S. Hazra, and D. Sarkar. From random matrices to long range dependence. *arXiv preprint arXiv:1401.0780*, 2014. 53, 60

[19] S. Chatterjee. Concentration inequalities with exchangeable pairs. *Ph. D. thesis, arXiv preprint math/0507526*, 2005. 16

[20] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34(6):2061–2076, 2006. 10, 39, 148, 152

[21] S. Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Relat. Fields*, 138(1-2):305–321, 2007. 16

[22] J. Dedecker, F. Merlevède, and M. Peligrad. A quenched weak invariance principle. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(3):872–898, 2014. 69

[23] J. Dedecker, F. Merlevède, and D. Volný. On the weak invariance principle for non-adapted sequences under projective criteria. *J. Theor. Probab.*, 20(4):971–1004, 2007. 28

[24] J. L. Doob. Stochastic processes. New York: Wiley. 654 S., 1953. 17

[25] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples. 12

[26] P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 30(1):63–82, 1994. 92, 93

[27] L. Giraitis, P. Kokoszka, and R. Leipus. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*, 16(1):3–22, 2000.

[28] V. L. Girko. *Theory of random determinants*, volume 45 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990. Translated from the Russian. 65

[29] S. Golden. Lower bounds for the Helmholtz function. *Phys. Rev. (2)*, 137:B1127–B1128, 1965. 15, 20, 160

[30] F. Götze, A. Naumov, and A. Tikhomirov. Semicircle law for a class of random matrices with dependent entries. *arXiv preprint arXiv:1211.0389*, 2012. 68, 77, 151

[31] U. Grenander and G. Szegö. *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley-Los Angeles, 1958. 51

[32] A. Guntuboyina and H. Leeb. Concentration of the spectral measure of large Wishart matrices with dependent entries. *Electron. Commun. Probab.*, 14:334–342, 2009. 9, 30, 73, 75, 89, 91

[33] W. Hachem, P. Loubaton, and J. Najim. The empirical eigenvalue distribution of a gram matrix: from independence to stationarity. *Markov Process. Related Fields*, 11(4):629–648, 2005. 6

[34] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics. 36

[35] F. Hansen and G. K. Pedersen. Jensen's inequality for operators and Löwner's theorem. *Math. Ann.*, 258(3):229–241, 1981/82. 159, 160

[36] F. Hofbauer and G. Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.*, 180(1):119–140, 1982. 93, 94

[37] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985. 147

[38] I. A. Ibragimov. Some limit theorems for stationary processes. *Teor. Verojatnost. i Primenen.*, 7:361–392, 1962. 117

[39] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[40] A. M. Khorunzhy and L. A. Pastur. On the eigenvalue distribution of the deformed Wigner ensemble of random matrices. In *Spectral operator theory and related topics*, volume 19 of *Adv. Soviet Math.*, pages 97–127. Amer. Math. Soc., Providence, RI, 1994. 53, 56, 72, 156, 158

[41] A. N. Kolmogorov and Y. A. Rozanov. On strong mixing conditions for stationary gaussian processes. *Theory of Probability & Its Applications*, 5(2):204–208, 1960.

[42] E. H. Lieb. Convex trace functions and the wigner-yanase-dyson conjecture. *Advances in Mathematics*, 11(3):267 – 288, 1973. 15, 159

[43] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.*, 42(3):906–945, 05 2014. 16, 17, 114

[44] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967. 3, 5

[45] F. Merlevède and M. Peligrad. On the empirical spectral distribution for matrices with long memory and independent rows. *arXiv preprint arXiv:1406.1216*, 2014. 25, 47, 60, 152

[46] F. Merlevède, M. Peligrad, and E. Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability. V: The Luminy volume.*, pages 273–292. Beachwood, OH: IMS, Institute of Mathematical Statistics, 2009. 18, 20, 111, 118

[47] F. Merlevède, M. Peligrad, and E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields*, 151(3-4):435–474, 2011. 18, 20, 118

[48] F. Merlevède, M. Peligrad, and S. Utev. Recent advances in invariance principles for stationary sequences. *Probab. Surv.*, 3:1–36, 2006.

[49] A. Mokkadem. Mixing properties of ARMA processes. *Stochastic Process. Appl.*, 29(2):309–315, 1988. 12

[50] A. Mokkadem. Propriétés de mélange des processus autorégressifs polynomiaux. (Mixing properties of polynomial autoregressive processes). *Ann. Inst. Henri Poincaré, Probab. Stat.*, 26(2):219–260, 1990. 17

[51] M. H. Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM Probab. Stat.*, 17:120–134, 2013.

[52] S. Orey. Recurrent Markov chains. *Pacific J. Math.*, 9:805–827, 1959. 92

[53] A. Pajor and L. Pastur. On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Stud. Math.*, 195(1):11–29, 2009. 8

[54] G. Pan. Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. *J. Multivariate Anal.*, 101(6):1330–1338, 2010. 6

[55] G. Pan, J. Gao, and Y. Yang. Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.*, 109(506):600–612, 2014. 8, 26

[56] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011. 2, 8, 56, 72, 157, 158

[57] D. Paulin, L. Mackey, and J. A. Tropp. Efron-stein inequalities for random matrices. *arXiv preprint arXiv:1408.3470*, 2014. 16, 17

[58] M. Peligrad and S. Utev. Central limit theorem for stationary linear processes. *Ann. Probab.*, 34(4):1608–1622, 2006.

[59] O. Pfaffel and E. Schlemm. Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.*, 31(2):313–329, 2011.

[60] O. Pfaffel and E. Schlemm. Limiting spectral distribution of a new random matrix model with dependence across rows and columns. *Linear Algebra Appl.*, 436(9):2966–2979, 2012. 11, 87

[61] M. B. Priestley. *Nonlinear and nonstationary time series analysis.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1988. 9

[62] R. Rashidi Far, T. Oraby, W. Bryc, and R. Speicher. On slow-fading MIMO systems with nonseparable correlation. *IEEE Trans. Inform. Theory*, 54(2):544–553, 2008. 46, 59

[63] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, 42:43–47, 1956.

[64] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995. 5, 50, 51

[65] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995. 6

[66] A. V. Skorohod. On a representation of random variables. *Teor. Verojatnost. i Primenen.*, 21(3):645–648, 1976. 121

[67] M. Talagrand. *Mean field models for spin glasses. Volume I*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics].* Springer-Verlag, Berlin, 2011. Basic examples. 153

[68] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2012. 2, 162

[69] C. J. Thompson. Inequality with applications in statistical mechanics. *J. Mathematical Phys.*, 6:1812–1813, 1965. 15, 20, 160

[70] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. 15, 16, 18, 20, 111, 160, 161

[71] G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492, 1997. 105

[72] V. A. Volkonskiĭ and Y. A. Rozanov. Some limit theorems for random functions. I. *Theor. Probability Appl.*, 4:178–197, 1959. 12

[73] C. Wang, B. Jin, and B. Miao. On limiting spectral distribution of large sample covariance matrices by varma (p, q). *Journal of Time Series Analysis*, 32(5):539–546, 2011. 49

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

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

[76] J. H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon Press, Oxford, 1965. 150

[77] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. Biometrika 20, A, 32-52 (1928)., 1928. 1

[78] W. B. Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005. 9

[79] W. B. Wu. Asymptotic theory for stationary processes. *Stat. Interface*, 4(2):207–226, 2011. 9, 28

[80] J. Yao. A note on a Marčenko-Pastur type theorem for time series. *Statist. Probab. Lett.*, 82(1):22–28, 2012. 7, 8, 25, 26, 49

[81] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986. 3, 5