Behaviour of non parametric estimators of second order statistics of high dimensional time series: a large random matrix approach

Alexis Rosuel

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

Thèse de doctorat

Spécialité : Signal, Image, Automatique

Alexis ROSUEL

Comportement d'estimateurs non paramétriques des statistiques du second ordre de séries temporelles de grande dimension via les grandes matrices aléatoires.

Behaviour of non-parametric estimators of second-order statistics of high dimensional time series: a large random matrix approach.

Thèse dirigée par Philippe LOUBATON et Cristina BUTUCEA. Soutenue le 14 décembre 2021.

Jury

Rapporteurs:	Arup BOSE	Professeur, Indian Statistical Institute
	Romain COUILLET	Professeur, Université Grenoble-Alpes
	Jeff Jianfeng YAO	Professeur, University of Hong Kong
Examinateurs:	Florence MERLEVÈDE	Professeur, Université Gustave Eiffel
Invité:	Jamal NAJIM	Directeur de recherche, Université Gustave Eiffel
Directeurs de thèse:	Cristina BUTUCEA	Professeur, Université Gustave Eiffel
	Philippe LOUBATON	Professeur, Université Gustave Eiffel

Résumé Les grandes matrices aléatoires se sont révélées depuis quelques temps être fondamentales en mathématiques (statistiques en grande dimension, algèbre d'opérateur, combinatoire, théorie des nombres, ...) et en physique (physique nucléaire, théorie quantique des champs, chaos quantique, ...). Leur utilisation en traitement statistique du signal et analyse des séries temporelles est en revanche plus récente. Elles s'avèrent être utiles quand les observations sont issues de séries temporelles multivariées de grande dimension (notée M) et que la taille de l'échantillon (noté N) n'est pas beaucoup plus grande que M, une situation qui devient très courante en raison du développement spectaculaire des dispositifs d'acquisition de données et des réseaux de capteurs. Ce contexte soulève un certain nombre de questions qui sont beaucoup étudiées par les chercheurs de la communauté des statistiques en grande dimension. L'exemple le plus parlant est lié au problème fondamental de l'estimation de la matrice de covariance car il est connu que son estimateur empirique classique se comporte mal quand N n'est pas significativement plus grand que M. Par conséquent, les schémas d'inférence statistique conventionnels basés sur des fonctionnelles de la matrice de covariance empirique peuvent mal se comporter. Afin de résoudre ce type de problème, les approches les plus populaires proposées sont basées sur une hypothèse de parcimonie des paramètres sous jacents. Malheureusement cette hypothèse n'est pas toujours valide. L'application de la théorie des grandes matrices aléatoires est une alternative prometteuse car, sous certaines hypothèses, il est possible de comprendre le comportement de fonctionnelles de la matrice de covariance empirique quand M et N sont simultanément grand, et d'utiliser ces résultats afin de proposer de nouvelles techniques d'inférence plus performantes. Bien qu'un certain nombre de travaux récents aillent dans cette direction, il reste un travail considérable à accomplir pour étudier des schémas d'inférence statistique basés sur des estimateurs empiriques des statistiques du second ordre de séries temporelles de grande dimensions. Le travail présenté est donc à l'interface entre grandes matrices aléatoire et statistiques des séries temporelles multivariées.

Mots clés: matrice aléatoires, statistique en grande dimension, traitement du signal, cohérence spectrale

Cette thèse a été préparée au Laboratoire IGM (CNRS, Univ. Paris-Est/MLV), 5 Boulevard Descartes, 77454 Marne-la-Vallée, France.

Abstract Large random matrices have been proved to be of fundamental importance in mathematics (high dimensional probability, operator algebras, combinatorics, number theory,...) and in physics (nuclear physics, quantum fields theory, quantum chaos,..) for a long time. The use of large random matrices is more recent in statistical signal processing and time series analysis. The corresponding tools turn out to be useful when the observation is a large dimension (say M) multivariate time series and the sample size N is not much larger than M, a situation that becomes very common due to the spectacular development of data acquisition devices and sensor networks. This context poses several new difficult statistical problems that are intensively studied by the high-dimensional statistics community. The most significant example is related to the fundamental problem of estimating the covariance matrix of the observation because the standard empirical covariance matrix is known to perform poorly if N is not significantly larger than M. As a result, the conventional statistical inference schemes that are based on functionals of the empirical covariance matrix may perform poorly. To mitigate this conceptual difficulty, the most popular approaches were based on the design of inference schemes using some possible degree of sparsity of the underlying parameters. However, sparsity is a property that does not necessarily hold. The use of large random matrix theory is an appealing alternative because, under some assumptions on the observations, it is possible to identify the behaviour of certain functionals of the empirical covariance matrix when M and N are both large, and to use the corresponding results to design new improved performance inference schemes. While some papers produced several valuable results, considerable work remains to be done to exploit the potential of large random matrix techniques in the context of statistics of high-dimensional Gaussian time series. The proposed works in this thesis are thus at the interface between large random matrices and the statistics of multivariate time series.

Keywords: random matrix, high-dimensional statistics, signal processing, spectral coherency

This thesis has been prepared at Laboratoire IGM (CNRS, Univ. Paris-Est/MLV), 5 Boulevard Descartes, 77454 Marne-la-Vallée, France.

Contents

List of Figures

List of Tables

Introduction

This thesis is devoted to the asymptotic study of estimates of the spectral coherency matrix of large dimensional Gaussian time series. If one denotes by M the dimension of the time series and by N the number of observations collected, classical time series analysis usually deals with the case where M is fixed and $N \to +\infty$. This context enables us to derive several important asymptotic results. These low-dimensional results in general break down when M is considered as a function of N, such that $M(N) \xrightarrow{N \to +\infty} +\infty$. Modern statistical problems are commonly faced with this context, either because the dimensionality M of the problem is large, or because it is not possible to collect a sufficiently large number of observations N so that N is in practice not much greater than M (this can happen for instance because the observed time series is not stationary and its parameters are rapidly evolving, or because the duration of the observation period is small). In these situations, it becomes tricky to apply low-dimensional results from statistical analysis. This relatively new asymptotic regime motivated a certain number of research papers, and an important proportion of them considered using results from random matrix theory. Random matrix theory is concerned with asymptotic properties of the eigenstructure of high-dimensional random matrices when both M and N are diverging such that $M/N \to c \in (0, +\infty)$. This approach turned out to be insightful, and in particular shed light on the behaviour of the eigenvalues and eigenvectors of sample covariance and correlation matrices, both objects which are of prime importance for statistical analysis applications. This random matrix-based approach also allowed to design new improved performance approaches. However, other properties related to the behaviour of time series requires more information than those provided by the covariance matrix, and some applications may require detailed knowledge of the spectral density and spectral coherency matrices in the high-dimensional regime. The goal of this manuscript is to present some results obtained on the behaviour of a well-known estimator of the spectral coherency matrix of a high-dimensional Gaussian time series, and how these results can be used to address some important statistical signal processing problems in particular whether the components of the M–dimensional time series are independent or not.

0.1 Motivation

We consider an M–dimensional stationary complex Gaussian[1] time series $\mathbf{y}$. A time series $(\mathbf{y}_n)_{n \in \mathbb{Z}}$ is said to be Gaussian if and only if for any finite set of time points $N_1, N_2, \ldots, N_k$, $\mathbf{y}_{N_1}, \mathbf{y}_{N_2}, \ldots, \mathbf{y}_{N_k}$ are jointly normally distributed. If we denote $\mathbf{R}_u = \mathbb{E}[\mathbf{y}_{n+u}\mathbf{y}_n^*]$ its autocovariance at lag u, one can define the *spectral density matrix* at frequency ν by

$$\mathbf{S}(\nu) = (s_{ij}(\nu))_{i,j=1}^M = \sum_{u \in \mathbb{Z}} \mathbf{R}_u e^{-\mathrm{i}2\pi u \nu} \tag{1}$$

[1] The complex circular Gaussian distribution with variance σ^2 is denoted as $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ and defined as $X_1 + iX_2$ where $\mathbf{X}_1, \mathbf{X}_2 \sim \mathcal{N}(0, \sigma^2/2)$ i.i.d. A random vector $\mathbf{x}$ of $\mathbb{C}^n$ follows the $\mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{R})$ distribution if $\mathbf{b}^*\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{b}^*\mathbf{R}\mathbf{b})$ for all deterministic (column) vector $\mathbf{b}$ and a fixed $n \times n$ positive definite matrix $\mathbf{R}$.

and the *spectral coherency matrix* (see for instance [13, Chapter 7-6], [52, Chapter 5.5]) at frequency ν, an analogue of correlation in the frequency domain, by

$$\mathbf{C}(\nu) = (c_{ij}(\nu))_{i,j=1}^{M} = \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\mathbf{S}(\nu)\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}} \tag{2}$$

where $\mathrm{dg}(\mathbf{S}(\nu)) = \mathbf{S}(\nu) \odot \mathbf{I}_M$, with $\odot$ denoting the Hadamard product (entrywise product) and $\mathbf{I}_M$ is the M–dimensional identity matrix.

These quantities, $\mathbf{S}(\nu)$ and $\mathbf{C}(\nu)$, provide useful information about the second-order dynamics properties of $\mathbf{y}$, which from a statistical point of view could be important to understand and extract from a finite sample of $(\mathbf{y}_n)_{n=1,\dots,N}$. In other words, spectral density aggregates information of autocovariance of different lag orders l at a specific frequency $\nu \in [-\frac{1}{2}, \frac{1}{2}]$. For instance, it is shown in [68] that $c_{ij}(\nu)$, the coherence between the time series i and j at frequency ν, is equivalent to the squared correlation between ν-oscillatory components of the two time series. Note also that the autocovariance function can be recovered from the spectral density by

$$\mathbf{R}_l = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\mathrm{i}\pi l\nu}\mathbf{S}(\nu)\,\mathrm{d}\nu$$

In this manuscript, we will detail some applications in statistical analysis and signal processing, such as the test of independence between the M components of $\mathbf{y}$, and the detection of a low-rank spatially correlated signal within a high-dimensional noise. To achieve these goals, we will rely on the smoothed periodogram estimator of $\mathbf{S}(\nu)$ (also called averaged periodogram estimator in the literature) of $\mathbf{y}$ at frequency ν defined by

$$\hat{\mathbf{S}}_N(\nu) = (\hat{s}_{N,ij}(\nu))_{i,j=1}^{M} = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_N\left(\nu + \frac{b}{N}\right)\boldsymbol{\xi}_N\left(\nu + \frac{b}{N}\right)^* \tag{3}$$

where B is an even integer, which represents the smoothing span, and

$$\boldsymbol{\xi}_N(\nu) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{y}_n e^{-2\mathrm{i}\pi(n-1)\nu}$$

is the renormalized Fourier transform of $(\mathbf{y}_n)_{n=1\dots,N}$. The corresponding estimated spectral coherency matrix is defined as:

$$\hat{\mathbf{C}}_N(\nu) = (\hat{c}_{N,ij})_{i,j=1}^{M}(\nu) = \mathrm{dg}\left(\hat{\mathbf{S}}_N(\nu)\right)^{-\frac{1}{2}}\hat{\mathbf{S}}_N(\nu)\mathrm{dg}\left(\hat{\mathbf{S}}_N(\nu)\right)^{-\frac{1}{2}} \tag{4}$$

The following chapters of this manuscript are motivated by the understanding of some properties of $\hat{\mathbf{C}}_N(\nu)$ when the dimension M of $\mathbf{y}$ is large and the components of $\mathbf{y}$ are independent time series. Before stating the obtained results, we recall that the properties of quantities derived from $\hat{\mathbf{S}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$ have first been studied in the regime where $N \to \infty$ and M is fixed, a regime that we will call *low-dimensional* in the rest of this manuscript. It is well known (see the paragraph on the literature review paragraph below for more details) that $\hat{\mathbf{S}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$ are consistent estimates of $\mathbf{S}(\nu)$ and $\mathbf{C}(\nu)$ in the low-dimensional regime if $B \to +\infty$ and $\frac{B}{N} \to 0$. Under mild assumptions on the memory of the time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$, $\hat{\mathbf{S}}_N(\nu)$ is moreover an asymptotically normal estimate of $\mathbf{S}(\nu)$, which can, in turn, be used to study the asymptotic performance of various tests based on $\hat{\mathbf{C}}_N(\nu)$.

In the case where $M \to +\infty$, to ensure that $\hat{\mathbf{S}}_N(\nu)$ is still a consistent estimator of $\mathbf{S}(\nu)$, it is necessary that $\frac{M}{B} \to 0$. A simple example which shows this is the following: consider $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ a sequence of i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables. In this context, for each ν, the renormalized Fourier transform vectors $(\boldsymbol{\xi}_N(\nu + b/N))_{b=-B/2,\dots,B/2}$ are mutually independent $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$ random vectors.

The spectral density estimate $\hat{\mathbf{S}}_N(\nu)$ thus coincides with the sample covariance matrix of these $(B+1)$ M–dimensional vectors. If one denotes by c_N the ratio $\frac{M}{B+1}$ and suppose that $c_N \to c$, it is known since [63] that $\|\hat{\mathbf{S}}_N(\nu) - \mathbf{S}(\nu)\| \xrightarrow{N\to+\infty} (1+\sqrt{c})^2 - 1$ almost surely. This quantity is null if and only if $c = 0$, which means that $\frac{M}{B} \to 0$.

Nowadays, in many practical applications involving high-dimensional signals and/or moderate sample size, the ratio $\frac{M}{N}$ may not be small enough to be able to choose B so that $\frac{M}{B}$ is much smaller than 1 and $\frac{B}{N}$ small enough. In this context, the behaviour of $\hat{\mathbf{C}}_N(\nu)$ is not accurately predicted by the results obtained in the low-dimensional regime ($\frac{B}{N} \to 0$ with either M fixed or $\frac{M}{B} \to 0$). In particular, $\hat{\mathbf{C}}_N(\nu)$ has no reason to be close from $\mathbf{C}(\nu)$. In this situation, it appears more relevant to choose B of the same order of magnitude as M, and to rely on the *high-dimensional regime* in which $M := M(N), B := B(N)$ and N converge to infinity such that $\frac{M}{B}$ converges to a positive constant $c \in (0, +\infty)$ while $\frac{B}{N}$ converges to zero. As seen below, large random matrix methods then allows to have a good understanding of the behaviour of $\hat{\mathbf{C}}_N(\nu)$. In the following, we moreover assume that

$$B = \mathcal{O}(N^\alpha) \text{ for some } \alpha \in \left(\frac{1}{2}, 1\right)$$

Of course, since $\frac{M}{B} \to c \in (0, +\infty)$, it is clear that M is also $\mathcal{O}(N^\alpha)$ for $\alpha \in \left(\frac{1}{2}, 1\right)$, and $\frac{M}{N} \to 0$. Note that α is taken greater than $\frac{1}{2}$ so that M is not too small compared to N. Otherwise, it becomes again reasonable in applications to find a B such that simultaneously $M \ll B$ and $B \ll N$. This means that, in the case $\alpha < \frac{1}{2}$, choosing B of the same order of magnitude as M has, in practice, little interest and would only make our results more complex to state. This is why we discard this case. Lastly, in the case where $\frac{M}{N} \to c \in (0, +\infty)$ (so B must also be chosen of the same order of magnitude as N and M) the entries of $\hat{\mathbf{S}}_N(\nu)$ are not even consistent. Analysis based on the spectral density is not recommended anymore, and one has to rely on other tools, such as the behaviour of the eigenvalues of empirical covariance matrix from random matrix theory. The conclusion of this paragraph is that in this manuscript, we will rely on the following *high-dimensional regime* defined by:

$$M = M(N), B = B(N), B = \mathcal{O}(N^\alpha) \text{ for } \alpha \in \left(\frac{1}{2}, 1\right), \text{ while } \frac{M}{B+1} := c_N \xrightarrow{N\to+\infty} c \in (0, +\infty) \tag{5}$$

which will be abbreviated by $\xrightarrow{N\to+\infty}$. This regime is relevant to study $\hat{\mathbf{S}}_N(\nu)$ defined in (3) using large random matrix methods, since it can be written as $\hat{\mathbf{S}}_N(\nu) = \mathbf{\Sigma}_N(\nu)\mathbf{\Sigma}_N(\nu)^*$ where $\mathbf{\Sigma}_N(\nu)$ is the $M \times (B+1)$ matrix defined by

$$\mathbf{\Sigma}_N(\nu) = \frac{1}{\sqrt{B+1}} \left(\boldsymbol{\xi}_N \left(\nu - \frac{B}{2N}\right), \ldots, \boldsymbol{\xi}_N \left(\nu + \frac{B}{2N}\right) \right).$$

This can be seen as the empirical covariance matrix applied to the Fourier transforms vectors $\boldsymbol{\xi}_N(\nu - \frac{B}{2N}), \ldots, \boldsymbol{\xi}_N(\nu + \frac{B}{2N})$, It can thus be expected that random matrix theory provides tools to understand the eigenvalues of $\hat{\mathbf{S}}_N(\nu)$ under our high-dimensional regime.

We finish this paragraph by making more precise the two quantities derived from $\hat{\mathbf{C}}_N(\nu)$ that will be studied in this manuscript when the components of $\mathbf{y}$ are assumed independent. First, if one denotes $\hat{\lambda}_1(\nu), \ldots, \hat{\lambda}_M(\nu)$ the eigenvalues of $\hat{\mathbf{C}}_N(\nu)$, a quantity of interest to study is the Linear Spectral Statistic (LSS) (see [5]) $\sum_{m=1}^M f(\hat{\lambda}_m(\nu))$ for some function f defined on $\mathbb{R}_+$ satisfying some regularity conditions. This quantity is related to the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N(\nu)$ since writing F_ν the empirical eigenvalue distribution function defined by:

$$\hat{F}_\nu(t) = \frac{1}{M}\text{card}\{\hat{\lambda}_m(\nu) : \hat{\lambda}_m(\nu) \leq t\}$$

we see that the LSS corresponds to test the empirical probability distribution against the function f. Moreover, this quantity is of importance for testing the independence of the M–time series, since it can be used to compare the eigenvalue distribution of $\hat{\mathbf{C}}_N(\nu)$ with the distribution expected under the null hypothesis $\mathcal{H}_0$ where the time series are independent. Still motivated by the independence test, the second object of interest will be the off-diagonal entries of $\hat{\mathbf{C}}_N(\nu)$, and more precisely $\sup_\nu \sup_{i \neq j} |\hat{c}_{N,ij}(\nu)|^2$. This manuscript compiles results obtained on the behaviour of these two objects (LSS and maximum off-diagonal entry of $\hat{\mathbf{C}}_N(\nu)$) under various model assumptions for the time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ in the high-dimensional asymptotic regime described above.

0.2 Literature review

In this section, we present some important results related to the questions addressed in this manuscript. We will divide this section into two paragraphs: the first one presents some of the typical topics of interest around the spectral density and coherency of time series in the low-dimensional regime, while the second one presents some of the main useful results in random matrix theory.

0.2.1 On the spectral density/coherency matrices

Multivariate spectral density estimation is an important problem in time series and signal processing. The spectral density of a stationary multivariate time series is the frequency domain analogue of covariance and is based on the Fourier transform of the autocovariance function. It aggregates information across different lags and among the components of a multivariate time series. Compared to the Pearson correlation matrix, it provides a richer description of a potential cross-sectional dependence. While most of the classical results on time series analysis through frequency domain tools are available in [13] and [14], we recall in this paragraph some of them to help the reader be familiar with the objects used in this manuscript.

Periodogram. As this manuscript is interested in the behaviour of the smoothed periodogram estimator, it could be of importance to collect some results concerning the periodogram estimator, which is defined as $\boldsymbol{\xi}_N(\nu)\boldsymbol{\xi}_N(\nu)^*$, or simply $|\xi_N(\nu)|^2$ in the univariate case. The periodogram is a fundamental tool in spectral analysis. In the case where $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is a complex Gaussian time series, it can be shown that the Fourier transform at frequencies $\nu_1, \ldots, \nu_L$ written $\boldsymbol{\xi}_N(\nu_1), \ldots, \boldsymbol{\xi}_N(\nu_L)$ are asymptotically independent $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{S}(\nu_l))$ random variables. Moreover, in the univariate case, it is known that $|\xi_N(\nu)|^2$ is asymptotically unbiased for $s(\nu)$ but not consistent due to its non-diminishing variance. For instance, for i.i.d Gaussian white noise $y_n \sim \mathcal{N}(0, \sigma^2)$, the variance of $|\xi_N(\nu)|^2$ is of the order σ^4 [14, Proposition 10.3.2]. Moreover, it is shown in [30] that for any sequence of independent, normally distributed random variables with mean zero and finite variance σ^2, the random variables $\{|\xi_N(\frac{k}{N})|^2, k \in \{0, \ldots, N-1\}\}$ are independent and have exponential distributions with σ^2 as their common mean. It is also shown in [1] that for a stationary ergodic process satisfying $\inf_{\nu\in[0,1]} s(\nu) > 0$ and additional moment assumptions on the innovations, then almost surely:

$$\lim_{N\to+\infty} \sup_{\nu\in[0,1]} \frac{|\xi_N(\nu)|^2}{s(\nu) \log N} = 1$$

Smoothed periodogram. While these results for the periodogram $\boldsymbol{\xi}_N(\nu)\boldsymbol{\xi}_N(\nu)^*$ are useful, they suffer as shown above from the non-consistency of this estimator in the regime where M is fixed and $N \to +\infty$. To achieve consistency, it is common to resort to smooth periodograms over nearby frequencies. An example of estimator in this more general class of estimator is called the smoothed periodogram and is defined as follows:

$$\hat{\mathbf{S}}_N(\nu) = \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_N\left(\nu + \frac{B}{N}\right) \boldsymbol{\xi}_N\left(\nu + \frac{B}{N}\right)^*$$

where B is a parameter called the spectral window. Under mild assumptions on the spectral window, some distributional results are also available for the windowed periodogram estimator, see for instance [13], [10], [66],[20],[57] and [97].

Before moving to the next section, we note that it is possible to define more general estimator of smoothed periodogram using arbitrary functions, but this manuscript only considers the smoothed periodogram estimator defined by (3) (also called the averaged periodogram). This allows to keep computations tractable and as simple as possible while providing clues on the behaviour of high-dimensional phenomena. We believe that considering the more general class of estimator would not change much our understanding of the asymptotic properties of the spectral coherency estimator in high dimension (up to some constants independent of N, but the speed of convergence would remain unchanged), but would greatly increase the derivation complexity and decrease the readability of our results.

0.2.2 On large dimensional general sample covariance matrices

Even though this manuscript is interested in understanding the behaviour of an estimator of the spectral density matrix of high-dimensional stationary processes, it has been seen that our smoothed periodogram estimator can be seen as the empirical covariance matrix of the Fourier vectors associated with the observations $(\mathbf{y}_n)_{n=1,\ldots,N}$, so it is insightful to review some of the known results from random matrix theory on the sample covariance matrix.

While the behaviour of the eigenvalues of sample covariance matrices is well-known in the low-dimensional regime (where the dimension M is fixed while the sample size $N \to +\infty$) since the foundation work in [2], the high-dimensional regime requires more work. The aim of this subsection is of course not to provide a general overview of the results developed regarding the behaviour of the eigenvalues of large dimensional sample covariance matrices, but instead to introduce the reader to the main ideas used later in this manuscript. This subsection will cover several subjects, from the pure white noise model to several extensions incorporating dependence between the entries, either within the rows (which will sometimes be called *time dependence*) or within the columns (which will sometimes be called *spatial dependence*).

Marcenko-Pastur distribution. For the rest of this section we consider the sample covariance matrix $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ where $\mathbf{Y}_N$ is a matrix of size $M \times N$, and the asymptotic regime considered will be $M, N \to +\infty$ such that $\frac{M}{N} \to c \in (0, +\infty)$. We stress that this notation is not the one that will be used in this manuscript (we will use $M \times (B+1)$ matrices such that $\frac{M}{B+1} \to c \in (0, +\infty)$), but is a standard one used in the random matrix literature. [63] is one of the first papers which realized that there might be new problems in estimating the spectral properties of large-dimensional covariance matrices when $\frac{M}{N}$ is not small. In particular, they showed the surprising result that in the case of i.i.d. data with variance 1, the eigenvalues of the sample covariance matrix $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ do not concentrate around 1 (the value of all population eigenvalues), but rather were spread out on the interval $[(1-\sqrt{\frac{M}{N}})^2, (1+\sqrt{\frac{M}{N}})^2]$ when $M \leq N$ (otherwise, the matrix $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ is rank deficient and there are almost surely exactly $M - N$ zero eigenvalues). This implies that when the ratio $\frac{M}{N}$ is not small, the sample covariance matrix is not a good estimator of the population covariance. Fortunately, they also have shown that the empirical eigenvalue distribution of the sample covariance matrix is asymptotically non-random, which enables the design of new improved estimators, a better understanding of the quantities related to sample covariance matrices, and more generally the foundations of random matrix theory applied to statistical data analysis. More precisely, consider $\mathbf{Y}_N$ an $M \times N$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries and consider the random measure $\hat{\mu}_N$ defined by

$$\mathrm{d}\hat{\mu}_N = \frac{1}{M}\sum_{m=1}^{M} \delta_{\hat{\lambda}_{m,N}}$$

where $\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{M,N}$ are the M eigenvalues of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ and δ_x the Dirac measure at point x. It is shown in [63] that there exist a non-random measure $\mu_{mp,c}$, depending on $c := \lim \frac{M}{N} \in (0, +\infty)$, such that almost surely

$$\hat{\mu}_N - \mu_{mp,c} \xrightarrow[M\to+\infty]{w} 0$$

where the convergence is the weak convergence of measures. The measure $\mu_{mp,c}$ is defined by:

$$\mathrm{d}\mu_{mp,c}(\lambda) = \left(1 - \frac{1}{c}\right)_+ \delta_0(\lambda) + \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2c\pi\lambda}\mathbb{1}_{[\lambda_-,\lambda_+]}(\lambda) \tag{6}$$

where $(\cdot)_+ = \max(\cdot, 0)$ and $\lambda_\pm = (1 \pm \sqrt{c})^2$. When $c > 1$, the $M - N$ zero eigenvalues are represented by the term $\left(1 - \frac{1}{c}\right)_+ \delta_0(\lambda)$. When $c < 1$, the Marcenko-Pastur distribution is supported on the interval $[\lambda_-, \lambda_+]$, also called the bulk of the distribution. Their method relies on the Stieltjes transform of the empirical eigenvalue distribution, defined as:

$$\hat{m}_N(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\hat{\mu}_N(\lambda)}{\lambda - z}, \quad \text{for all } z \in \mathbb{C} \setminus \mathbb{R}_+$$

or equivalently by the resolvent $(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* - z\mathbf{I}_M)^{-1}$ since $\hat{m}_N(z)$ can also be expressed as follows:

$$\hat{m}_N(z) = \frac{1}{M}\mathrm{tr}\,\left(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* - z\mathbf{I}_M\right)^{-1}.$$

The authors of [63] showed that the function $\hat{m}_N$ satisfies asymptotically the equation $zcX^2 + (z + c - 1)X + 1 = 0$, which admits the solution $z \mapsto t(z)$ in the set of Stieltjes transforms equal to:

$$t(z) = \frac{-(z + c - 1) + \sqrt{(z - \lambda_+)(z - \lambda_-)}}{2zc}$$

where for $z = \rho e^{i\theta}$, $\sqrt{z} = \sqrt{\rho}e^{i\theta/2}$ for $\theta \in (0, 2\pi)$. For a sequence $(c_N)_{N\geq 1}$, we denote

$$t_N(z) = \frac{-(z + c_N - 1) + \sqrt{(z - \lambda_+)(z - \lambda_-)}}{2zc_N}$$

where it is understood that $\lambda_\pm = (1 \pm \sqrt{c_N})^2$. It will be convenient in the next chapters to introduce now the Stieltjes transform of $\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N$:

$$\underline{\hat{m}}_N(z) = \frac{1}{N}\mathrm{tr}\,\left(\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N - z\mathbf{I}_N\right)^{-1}$$

($(\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N - z\mathbf{I}_N)^{-1}$ is sometimes called the co-resolvent) and it can be shown equivalently that $\underline{\hat{m}}_N$ converges asymptotically to $\tilde{t}$, satisfying:

$$\tilde{t}(z) = ct(z) - \frac{1 - c}{z}.$$

Lastly, $t(z)$ and $\tilde{t}(z)$ can be expressed as the unique solution in the set of Stieltjes transforms carried by $\mathbb{R}_+$ of the following coupled equations:

$$t(z) = \frac{-1}{z(1+\tilde{t}(z))}, \quad \tilde{t}(z) = \frac{-1}{z(1+ct(z))}.$$

Similarly, for a sequence $(c_N)_{N\geq 1}$, we denote:

$$t_N(z) = \frac{-1}{z(1+\tilde{t}_N(z))}, \quad \tilde{t}_N(z) = \frac{-1}{z(1+c_N t_N(z))}. \tag{7}$$

It is also of interest to consider the limiting spectral distribution of the sample correlation matrix defined by $\mathrm{dg}(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*)^{-\frac{1}{2}}\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*\mathrm{dg}(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*)^{-\frac{1}{2}}$. It was first studied by [46] under second-moment conditions, where the author showed that in the white noise case, the limiting spectral distribution of the sample correlation also almost surely converges weakly towards the Marcenko-Pastur distribution.

Extensions with spatial dependence. While the limiting spectral distribution of the white model (with no correlation between the rows and the columns of $\mathbf{Y}_N$) has been studied in detail in [85],[84] and [4], there has also been a large amount of effort made to relax the independence structure imposed between the entries Y_{ij}, and understand how dependence between the entries of $\mathbf{Y}_N$ have an impact on the limiting spectral distribution of the sample covariance matrix. We start with the model where the samples $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent, but the components of $\mathbf{y}_i$ are not independent. The authors in [85] considered the case where $\mathbf{Y}_N = \mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N$, and $\mathbf{X}_N$ is a matrix of size $M \times N$ with i.i.d. entries of finite second moment while $\mathbf{\Sigma}_N$ is a non-negative definite matrix independent of $\mathbf{X}_N$. The Marcenko-Pastur equation corresponds to the case where $\mathbf{\Sigma}_N = \mathbf{I}_M$. In the other case, where $\mathbf{\Sigma}_N \neq \mathbf{I}_M$, $\mathbf{Y}_N$ contains spatial correlation. Under the assumption that the spectral distribution of the matrices $(\mathbf{\Sigma}_N)_{N\geq 0}$ almost surely converges weakly to a probability density function H, then with probability one as $N \to +\infty$ while $\frac{M}{N} \to c$, the empirical spectral distribution of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ almost surely converges weakly to a non-random p.d.f. This limiting spectral distribution is given implicitly by an equation whose Stieltjes transform $z \mapsto m(z)$ is the unique solution in the set $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}_+\}$ of:

$$m(z) = \int_{\mathbb{R}} \frac{\mathrm{d}H(t)}{t(1-c-czm(z))-z}$$

Note that in the white noise case, $H = \delta_1$, and we recover the equation defining the Marcenko-Pastur distribution. Moreover, [86] derived analytical properties concerning the limiting spectral distribution from the study of this equation.

Another model of importance for applications and specifically for Chapter 3 and Chapter 4 is the well-known spiked model. This corresponds to the case where $\mathbf{\Sigma}_N$ is a finite K-rank perturbation of the identity:

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{I}_M, \quad \text{with } \mathbf{B}_N \text{ positive and } \mathrm{rank}(\mathbf{B}_N) = K$$

so that the $M-K$ smallest eigenvalues of $\mathbf{\Sigma}_N$ are equal to 1. It is then of interest to study the eigenvalues of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* = \frac{1}{N}\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^*\mathbf{\Sigma}_N^{\frac{1}{2}}$ where $\mathbf{X}_N$ is an $M \times N$ matrix of i.i.d. zero-mean variance one entries. This model is also called the multiplicative spiked model and has been studied in [6] and [7]. Denoting $\hat{\mu}_N$ the empirical eigenvalue distribution of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$, it is shown that almost surely,

$$\hat{\mu}_N \xrightarrow[N\to+\infty]{w} \mu_{mp,c}$$

However, all the eigenvalues of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ do not concentrate around the bulk $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$.

It is shown in [7] that the K largest eigenvalues $\hat{\lambda}_{1,N}, \dots, \hat{\lambda}_{K,N}$ of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ can converge to limits outside of the bulk if the K non-zero eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ converges towards the sufficiently large values $\lambda_1, \dots, \lambda_K$. More precisely, for any $k = 1, \dots, K$,

$$\hat{\lambda}_{k,N} \xrightarrow[N\to+\infty]{a.s.} \begin{cases} \frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k} & \text{if } \lambda_k > 1+\sqrt{c} \\ (1+\sqrt{c})^2 & \text{else.} \end{cases}$$

Note that $\lambda_k > 1+\sqrt{c}$ implies that $\frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k} > (1+\sqrt{c})^2$, which means that asymptotically the k–th largest eigenvalues will effectively leave the bulk of the eigenvalues, and create a spike localised around the values $\frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k}$.

Extensions with time dependence. We now focus on the case where the samples $\mathbf{y}_1, \dots, \mathbf{y}_N$ are correlated through time instead of correlated through space. There are several ways of considering such a model. For instance, one could consider $\mathbf{Y}_N = \mathbf{X}_N\mathbf{\Sigma}_N^{\frac{1}{2}}$, (instead of $\mathbf{Y}_N = \mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N$ in the previous paragraph) and study the sample covariance matrix $\hat{\mathbf{\Sigma}}_N = \frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* = \frac{1}{N}\mathbf{X}_N\mathbf{\Sigma}_N\mathbf{X}_N^*$. However, it can be seen that this model is not so different than the one developed from [85], as one could consider instead $\underline{\hat{\mathbf{\Sigma}}_N} = \frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N = \frac{1}{M}\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^*\mathbf{\Sigma}_N^{\frac{1}{2}}$, whose limiting eigenvalue distribution is known. It remains to observe that $\hat{\mathbf{\Sigma}}_N$ and $\underline{\hat{\mathbf{\Sigma}}_N}$ have exactly the same eigenvalues, except $N-M$ zeros if $N > M$ ($M-N$ zeros otherwise).

Another model, which will be of great importance in Chapter 2 is the following: suppose that there exists an $N \times N$ positive Hermitian matrices $\mathbf{\Theta}_1, \dots, \mathbf{\Theta}_M$ such that

$$\begin{pmatrix} y_{m,1} \\ \vdots \\ y_{m,N} \end{pmatrix} = \mathbf{\Theta}_m^{\frac{1}{2}} \begin{pmatrix} x_{m,1} \\ \vdots \\ x_{m,N} \end{pmatrix}$$

where $(x_{m,1}, \dots, x_{m,N}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ and $(x_{m_1,1}, \dots, x_{m_1,N})$ is independent of $(x_{m_2,1}, \dots, x_{m_2,N})$ for $m_1 \neq m_2$. This model, studied in [92] and [48], is useful to represent M–dimensional observations, where the dimensions are independent, but each one has its own covariance structure described by $\mathbf{\Theta}_m$. This model will be treated in detail in Chapter 2.

Lastly, a model related to those developed in this manuscript is to suppose that each dimension is generated as a time series. This model is of importance since an important problem in applications is to test whether a collection of time series are uncorrelated or not. In this case, the Marcenko–Pastur distribution is not anymore a good approximation of the eigenvalue distribution of the sample covariance matrix, and moreover this model cannot be expressed as Silverstein's sample covariance matrices from the previous paragraph. For instance, the author in [98] considered that each dimension of $\mathbf{y}$ is generated as an independent univariate linear process:

$$y_{m,n} = \sum_{k\geq 0} a_k \epsilon_{m,n-k}$$

where for each m, $(\epsilon_{m,n})_{n\in\mathbb{Z}}$ is a real-valued and weakly stationary white noise with mean zero and variance 1. Under the assumption that the filtering sequence is summable ($\sum_{k\geq 0} |a_k| < +\infty$) and moment condition on the innovations ($\mathbb{E}\epsilon_n^4 < +\infty$), it is shown that the spectral distribution of $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ converges towards a non-random probability distribution F whose Stieltjes transform $z \mapsto m(z)$ satisfies:

$$z = \frac{-1}{m(z)} + \int_0^1 \frac{1}{cm(z) + cs^{-1}(\nu)} \,\mathrm{d}\nu$$

where s is the spectral density of the linear process $(y_{m,n})_{n\geq 0}$ (by assumption, the spectral densities for all the dimensions are identical) equal to:

$$s(\nu) = \left| \sum_{k \geq 0} a_k e^{2i\pi k\nu} \right|^2.$$

The assumption that each time series is generated by the same linear filter is quite strong, and in this manuscript, we will not suppose this, which means that we suppose that there exists M sequences of filtering coefficients $(a_{1,k})_{k\geq 0}, \ldots, (a_{M,k})_{k\geq 0}$ (with additional assumptions that will be made precise in the following chapters) such that

$$y_{m,n} = \sum_{k \geq 0} a_{m,k} \epsilon_{m,n-k}$$

where the noises are still independent weakly stationary white noise.

Extensions with spatial and time dependence. Naturally, one may ask whether it is possible to simultaneously handle dependence within rows and columns since multivariate populations with spatio-temporal dependence can be found in many practical situations and are of considerable interest in statistical methods and applied areas. For instance, it is of interest to consider a general sample covariance matrix model of the form $\frac{1}{N}\mathbf{T}_{2N}^{\frac{1}{2}}\mathbf{X}_N^*\mathbf{T}_{1N}\mathbf{X}_N\mathbf{T}_{2N}^{\frac{1}{2}}$, where $\mathbf{T}_{1N}$ and $\mathbf{T}_{2N}$ and non-negative sequence of matrices, and $\mathbf{T}_{1N}$ are Hermitian, whose size coincides with $\mathbf{X}_N \in \mathbb{C}^{M\times N}$. In the case where $\mathbf{T}_{1,N}$ is positive semidefinite, this model is known as the separable covariance model, since the covariance of the data matrix is the Kronecker product of $\mathbf{T}_{1,N}$ and $\mathbf{T}_{2,N}$. Among the works which studied this model, we can refer to [71], [49], [18], [37] and the reference therein.

0.2.3 Previous works using the asymptotic regime (5)

Relatively few papers studied the behaviour of the smoothed periodograms $\hat{\mathbf{S}}_N(\nu)$ under the regime defined by (5). For instance, it was observed in [11] that as the dimension of the time series increases, the estimation risk of the smoothed periodogram also increases. This observation led to new directions of research to improve estimation and testing performance, using methods relying on shrinkage, thresholding (see e.g. [87]), and random matrix theory. The authors of [11], inspired by the well-known spreading phenomena of the eigenvalues of high-dimensional sample covariance matrices around the population eigenvalues, defined a shrinkage estimator by:

$$\hat{\mathbf{S}}_N^{(shrink)}(\nu) = p(\nu)\hat{\mathbf{S}}_N(\nu) + r(\nu)\mathbf{I}_M$$

for some explicit functions p and r controlling the shrinkage level for each frequency, and proved that this method can reduce the risk associated with the estimation of the spectral density matrix. Note that the asymptotic regime they considered is similar to the one used in this manuscript, as they required:

$$0 < \inf_{N\geq 1} \frac{M}{B} \leq \sup_{N\geq 1} \frac{M}{B} < +\infty \text{ and } \frac{\sqrt{M}B}{N} \to 0$$

which is equivalent to $M = o(N^{2/3})$ (and we recall that in this manuscript, we will assume $M = \mathcal{O}(N^\alpha)$ for some $\frac{1}{2} < \alpha < 1$). This approach has been improved later in [29] and the reference therein.

Another related approach to test the independence of high-dimensional time series is found in [61] and [60]. The authors consider L consecutive observations of the m–th time series starting at time n, namely

$$\mathbf{y}_{m,n}^L = [y_{m,n}, \ldots, y_{m,n+L-1}]^T$$

and from this built an ML–dimensional column vector

$$\mathbf{y}_n^L = [(\mathbf{y}_{1,n}^L)^T, \ldots, (\mathbf{y}_{M,n}^L)^T]$$

They denote by $\mathcal{R}_L$ the $ML \times ML$ covariance matrix of this random vector, i.e. $\mathcal{R}_L = \mathbb{E}[\mathbf{y}_n^L(\mathbf{y}_n^L)^*]$ where $(\cdot)^*$ stands for transpose conjugate. This matrix is sometimes referred as the spatio-temporal covariance matrix. Clearly, the M series $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$ are uncorrelated if and only if for each integer L, matrix $\mathcal{R}_L$ is block-diagonal, namely

$$\mathcal{R}_L = \mathrm{Bdiag}(\mathcal{R}_L)$$

where, for an $ML \times ML$ matrix $\mathbf{A}$, $\mathrm{Bdiag}(\mathbf{A})$ is the block-diagonal matrix of the same dimension whose $L \times L$ blocks are those of $\mathbf{A}$. We notice that the $L \times L$ diagonal blocks of $\mathrm{Bdiag}(\mathcal{R}_L)$ are the $L \times L$ Toeplitz matrices $\mathbf{R}_{m,L}$, $m = 1, \ldots, M$, defined by

$$(\mathbf{R}_{m,L})_{kk'} = r_m(k - k')$$

where $(r_m(k))_{k\in\mathbb{Z}}$ is the covariance sequence of the m–th time series. [61] and [60] denote by $\mathcal{R}_{corr,L}$ the block correlation matrix defined by

$$\mathcal{R}_{corr,L} = \mathrm{Bdiag}(\mathcal{R}_L)^{-\frac{1}{2}}\mathcal{R}_L\mathrm{Bdiag}(\mathcal{R}_L)^{-\frac{1}{2}}.$$

Consequently, $\mathcal{R}_L$ is block diagonal for each L if and only if $\mathcal{R}_{corr,L} = \mathbf{I}_{ML}$ for each L. A possible way to test whether the M components of the time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ are uncorrelated thus consists in estimating $\mathcal{R}_{corr,L}$ for a suitable value of L, and subsequently comparing the corresponding estimate with $\mathbf{I}_{ML}$. For this, the authors consider the standard sample estimate $\hat{\mathcal{R}}_{corr,L}$ defined by

$$\hat{\mathcal{R}}_{corr,L} = \mathrm{Bdiag}(\hat{\mathcal{R}}_L)^{-\frac{1}{2}}\hat{\mathcal{R}}_L\mathrm{Bdiag}(\hat{\mathcal{R}}_L)^{-\frac{1}{2}}$$

where $\hat{\mathcal{R}}_L$ is the empirical spatio-temporal covariance matrix given by

$$\hat{\mathcal{R}}_L = \frac{1}{ML}\sum_{n=1}^{N} \mathbf{y}_n^L(\mathbf{y}_n^L)^*$$

and where $\mathrm{Bdiag}(\hat{\mathcal{R}}_L)$ is the corresponding block diagonal matrix:

$$\mathrm{Bdiag}(\hat{\mathcal{R}}_L) = \begin{pmatrix} \hat{\mathbf{R}}_{1,L} & & \\ & \ddots & \\ & & \hat{\mathbf{R}}_{M,L} \end{pmatrix}$$

with $\hat{\mathbf{R}}_{m,L}$ for $m = 1, \ldots, M$, denoting the corresponding $L \times L$ diagonal blocks. Under the assumption of independence of the M components of the time series and under the asymptotic regime where $\frac{ML}{N} \xrightarrow{N\to+\infty} c$, they showed that

$$\int f(\lambda)\,\mathrm{d}\hat{\mu}_N(\lambda) - \int f(\lambda)\,\mathrm{d}\mu_{mp,c} \to 0$$

where $\mathrm{d}\hat{\mu}_N = \frac{1}{ML}\sum_{k=1}^{N}\delta_{\hat{\lambda}_k}$, and $(\hat{\lambda}_k)_{k=1,\ldots,ML}$ are the eigenvalues of $\hat{\mathcal{R}}_L$.

0.3 Contributions of the thesis

The general topic of this thesis is to understand the properties of the smoothed periodogram estimator of the spectral coherency matrix of an M–dimensional time series $\mathbf{y}$ when M is large and the components of $\mathbf{y}$ are independent. For all chapters, we will use the high-dimensional regime defined above in (5) that we recall here: M and B are integer functions of N such that $\frac{M}{B} \xrightarrow{N\to+\infty} c \in (0,+\infty)$ and $M = \mathcal{O}(N^{\alpha})$ for $\frac{1}{2} < \alpha < 1$. In this section, we will sometimes omit the explicit dependencies on the frequency ν and on the complex number z to reduce notations.

0.3.1 Contribution of Chapter 1

The motivation of this chapter is to evaluate the behaviour of certain Linear Spectral Statistics (LSS) of $\hat{\mathbf{C}}_N(\nu)$ defined in (4) under the hypothesis $\mathcal{H}_0$ that the M components $(y_{1,n})_{n\in\mathbb{Z}}, \dots, (y_{M,n})_{n\in\mathbb{Z}}$ of $\mathbf{y}$ are mutually uncorrelated and in our high-dimensional asymptotic regimes. Let the empirical spectral measure of $\hat{\mathbf{C}}_N(\nu)$ be defined by

$$\mathrm{d}\hat{\mu}_{N,\nu} = \frac{1}{M}\sum_{m=1}^{M} \delta_{\lambda_m(\hat{\mathbf{C}}_N(\nu))}$$

where $(\lambda_m(\mathbf{A}))_{m=1,\dots,M}$ represents the M eigenvalues of the $M \times M$ matrix $\mathbf{A}$, and δ_x is the Dirac mass at point x. An LSS of $\hat{\mathbf{C}}_N(\nu)$ can be written $\int_{\mathbb{R}} f(\lambda)\,\mathrm{d}\hat{\mu}_N(\nu,\lambda)$ or equivalently $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$ for some function f defined on $\mathbb{R}_+$ satisfying some regularity assumptions if needed. In this chapter, we study the quantity

$$\sup_{\nu\in[0,1]} \left| \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{mp,c_N} \right|$$

to show that the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N(\nu)$ converges towards the Marcenko-Pastur distribution. In order to study this quantity, we first work with a fixed ν, and later extend our results over $\nu \in [0,1]$ using a Lipschitz argument. For each fixed ν, we consider the intermediate quantity $\tilde{\mathbf{C}}_N(\nu)$ defined by

$$\tilde{\mathbf{C}}_N(\nu) = \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}}\hat{\mathbf{S}}_N(\nu)\mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}}$$

and show that $\tilde{\mathbf{C}}_N(\nu)$ can be written

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu)$$

where $\mathbf{X}_N(\nu)$ is an $M \times (B+1)$ matrix with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries and $\tilde{\mathbf{\Delta}}_N(\nu)$ is an error term converging towards 0 in spectral norm at rate $\frac{B}{N}$ (see below for more details on matrix $\tilde{\mathbf{\Delta}}_N(\nu)$). We now state the global decomposition used to approach the quantity of interest $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$:

$$\begin{aligned}\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) &= \left(\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))\right) + \left(\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))\right)\\ &+ \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right)\right)\\ &+ \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right) - \int f\,\mathrm{d}\mu_{mp,c_N}\right) + \int f\,\mathrm{d}\mu_{mp,c_N}\end{aligned}$$

Concerning the first step of the decomposition, we use the Helffer-Sjöstrand formula, which makes possible to transfer the study of the LSS to the resolvents $\tilde{\mathbf{Q}}_N(z) := (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$ and $\hat{\mathbf{Q}}_N(z) := (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$. Under assumptions on f (regularity and compactness of its support), the Helffer-

Sjöstrand formula states that

$$\int f \,\mathrm{d}\mu = \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(x,y)s_\mu(x+iy)\,\mathrm{d}x\,\mathrm{d}y$$

where s_μ is the Stieltjes transform of μ, $\bar{\partial} = \partial_x + i\partial_y$ and $\Phi_k(f) : \mathbb{C} \to \mathbb{C}$ the function defined on $\mathbb{C}$ by

$$\Phi_k(f)(x,y) = \sum_{l=0}^{k} \frac{(iy)^l}{l!} f^{(l)}(x)\rho(y)$$

where $\rho : \mathbb{R} \to \mathbb{R}^+$ is smooth, compactly supported, with value 1 in a neighbourhood of 0. Function $\Phi_k(f)$ coincides with f on the real line and extends it to the complex plane. Applying this in our context, we write that

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) = \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(x,y)\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{Q}}_N(x+iy) - \tilde{\mathbf{Q}}_N(x+iy))\,\mathrm{d}x\,\mathrm{d}y.$$

It is then possible to study the difference

$$\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} = \frac{1}{M}\mathrm{tr}\ \left(\hat{\mathbf{Q}}_N(z)(\tilde{\mathbf{C}}_N - \hat{\mathbf{C}}_N)\tilde{\mathbf{Q}}_N(z)\right)$$

To state our quantitative result about the behaviour of $\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$, we first need to define the concept of *stochastic domination*, where we say that the family of non-negative random variables $X = (X_N)_{N\geq 1}$ is stochastically dominated by the deterministic sequence $(u_N)_{N\geq 1}$ if for all $\epsilon > 0$, there exists $N_0(\epsilon)$ large enough and some $\gamma(\epsilon) > 0$ such that for any $N > N_0(\epsilon)$:

$$\mathbb{P}\left[X_N > N^\epsilon u_N\right] \leq \exp -N^\gamma$$

If $(X_N)_{N\geq 1}$ is stochastically dominated by $(u_N)_{N\geq 1}$ we use the notation $X_N \prec u_N$, and if for some complex valued family X we have $|X| \prec u_N$, we also write $X = \mathcal{O}_\prec(u_N)$. With this tool at hand, we show that the following stochastic domination holds uniformly in $\nu \in [0,1]$:

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) = -\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{2s_m(\nu)}\right) v_N < \tilde{\mathcal{D}}_N, f > + \mathcal{O}_\prec\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right)$$

where $v_N = \left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{b}{N}\right)^2\right)$ and $< \tilde{\mathcal{D}}_N, f >$ corresponds to the action of the distribution $\tilde{\mathcal{D}}_N$, whose Stieltjes transform is explicitly known, on f (this quantity will be made precise later). This allows us to focus on the behaviour of $\tilde{\mathbf{C}}_N(\nu)$. We then show by standard Gaussian concentration inequalities, that

$$\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) = \mathcal{O}_\prec\left(\frac{1}{B}\right)$$

so it remains to study $\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))$. Under the assumption that the M components of the time series are independent, we show that $\tilde{\mathbf{C}}_N(\nu)$ is well approximated by a white noise Wishart matrix for each ν:

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu)$$

where $\mathbf{X}_N(\nu)$ is an $M \times (B+1)$ matrix with i.i.d. $\mathcal{N}_\mathbb{C}(0,1)$ entries, $\sup_{\nu\in\mathcal{V}_N}\|\tilde{\mathbf{\Delta}}_N(\nu)\| = \mathcal{O}_\prec(\frac{B}{N})$ and $\mathcal{V}_N = \{\frac{k}{N}, k = 0,\ldots,N-1\}$. After Gaussian computations (using especially the Gaussian integration by part formula), it is possible to obtain the following estimation:

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))\right] - \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right)\right] = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$
$$- \frac{1}{M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{2s_m(\nu)} < \tilde{\mathcal{D}}_N, f > v_N + \mathcal{O}\left(\frac{1}{N} + \left(\frac{B}{N}\right)^3\right)$$

where $\mathcal{D}_N$ is another distribution whose Stieltjes transform is explicitly known. The strategy to obtain this result consists in considering $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{Q}}_N(z)\right] - \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \mathbf{Q}_N(z)\right]$, where $\mathbf{Q}_N(z) = (\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} - z\mathbf{I}_M)^{-1}$ is the resolvent associated with $\mathbf{X}_N\mathbf{X}_N^*$. Using long and very tedious Gaussian calculations, it holds that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z))\right] = -\left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}\right)\ \tilde{p}_N(z)\ v_N +$$
$$\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2\ p_N(z)\ v_N + \mathcal{O}_z\left(\frac{B}{N}\right)^3$$

where $\tilde{p}_N$ is the Stieltjes transform of the compactly supported distribution $\tilde{D}_N$ introduced previously, p_N is also the Stieltjes transform of another compactly supported distributions D_N. Their close-form expression is known:

$$\tilde{p}_N(z) = (zt_N(z))', \quad p_N(z) = \frac{-c_N(zt_N(z)\tilde{t}_N(z))^3}{1 - c_N(zt_N(z)\tilde{t}_N(z))^2}$$

where we recall that t_N and $\tilde{t}_N$ have been defined in (7). It remains to state a classical result from random matrix theory which ensures that for complex Gaussian white noise matrix $\mathbf{X}_N(\nu)$,

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] - \int f\ \mathrm{d}\mu_{mp,c_N} = \mathcal{O}\left(\frac{1}{B^2}\right)$$

where μ_{mp,c_N} is the Marcenko-Pastur distribution with parameter $c_N = \frac{M}{B+1}$. Combining all the previous estimations, we obtain that for each ν,

$$T_N(\nu) = \mathcal{O}_\prec\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right).$$

where

$$T_N(\nu) = \frac{1}{M}\mathrm{tr}\ \hat{\mathbf{C}}_N(\nu) - \int f\ \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$

This result informally states that under the assumption of independence of the components of the multidimensional time series, the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N(\nu)$ is close to the Marcenko-Pastur distribution, up to a correction of order $\mathcal{O}(\frac{B}{N})^2$ and an error rate of order $\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$. By a union bound over the grid $\mathcal{V}_N$ of cardinality N, it is clear by the definition of the stochastic domination that $\sup_{\nu\in\mathcal{V}_N}|T_N(\nu)| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$. It remains to show that $\nu \mapsto T_N(\nu)$ is Lipschitz with a Lipschitz constant of order N^p for some finite p to obtain the desired result:

$$\sup_{\nu\in[0,1]} |T_N(\nu)| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$$

However, we remark that this result does not allow to propose a statistical test since nothing is known about the asymptotic distribution of $\sup_{\nu\in[0,1]} |T_N(\nu)|$, so it impossible to design a test with a controlled type-I error. Nevertheless, the results obtained in this chapter represent a first step toward obtaining the asymptotic distribution of $\sup_{\nu\in[0,1]} |T_N(\nu)|$. This will be discussed in more details in the Perspectives section 0.4.

0.3.2 Contribution of Chapter 2

To evaluate explicitly the $\mathcal{O}(\frac{B}{N})^2$ correction term in the expression of $\frac{1}{M}\mathrm{tr}\ \hat{\mathbf{C}}_N(\nu) - \int f\,\mathrm{d}\mu_{mp,c_N}$, the approach used in Chapter 1 turns out to require extremely long and tedious calculations. In this chapter, we propose an alternative approach that allows us to recover this quantity much more easily. Instead of using the decomposition:

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{C}}_N(\nu)\right] - \int f\,\mathrm{d}\mu_{mp,c_N} = \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{C}}_N(\nu)\right] - \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] + \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] - \int f\,\mathrm{d}\mu_{mp,c_N}$$

we propose to directly study the behaviour of the resolvent $\mathbb{E}(\tilde{\mathbf{C}}_N(\nu) - z\mathbf{I}_M)^{-1}$ by using the covariance structure of the Fourier transforms $(\boldsymbol{\xi}_N(\nu + \frac{b}{N}))_{b=-B/2,\ldots,B/2}$ and Gaussian calculus. Note that $\tilde{\mathbf{C}}_N(\nu) = \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)\boldsymbol{\Sigma}_N(\nu)^*\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ where

$$\boldsymbol{\Sigma}_N(\nu) = \frac{1}{\sqrt{B+1}}\left(\boldsymbol{\xi}_N\left(\nu - \frac{B}{2N}\right),\ldots,\boldsymbol{\xi}_N\left(\nu + \frac{B}{2N}\right)\right)$$

so $\tilde{\mathbf{C}}_N(\nu)$ can be interpreted as the empirical covariance matrix of $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)$. Under the assumption of spatial independence, the rows of $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)$ are independent. We define $\boldsymbol{\Theta}_{N,m}(\nu)$ the covariance matrix of the normalized Fourier transforms at frequency ν of the m–th component of the time series:

$$\boldsymbol{\Theta}_{N,m}(\nu) = \mathbb{E}\left(\frac{\xi_{N,m}\left(\nu - \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}},\ldots,\frac{\xi_{N,m}\left(\nu + \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}\right)^*\left(\frac{\xi_{N,m}\left(\nu - \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}},\ldots,\frac{\xi_{N,m}\left(\nu + \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}\right)$$

By classical results on Fourier transforms (see [13]), for any $m \geq 1$,

$$\mathrm{Cov}\left(\xi_{N,m}\left(\nu + \frac{b}{N}\right), \xi_{N,m}\left(\nu + \frac{b'}{N}\right)\right) = s_m\left(\nu + \frac{b}{N}\right)\mathbb{1}(b = b') + \mathcal{O}\left(\frac{1}{N}\right)$$

This implies that for each m and ν, $\boldsymbol{\Theta}_{N,m}(\nu)$ is approximately the identity matrix, and means that the columns of $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)$ are almost independent. This model allows explicit calculations for the behaviour of $\tilde{\mathbf{Q}}_N(z) := (\tilde{\mathbf{C}}_N(\nu) - z\mathbf{I}_M)^{-1}$. We show that for each $z \in \mathbb{C}_+$, there exists deterministic matrices $\tilde{\mathbf{T}}_N(z,\nu)$ such that

$$\frac{1}{M}\mathrm{tr}\ \mathbb{E}\tilde{\mathbf{Q}}_N(\nu, z) - \frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}_N(\nu, z) = \mathcal{O}_z\left(\frac{1}{B^2}\right)$$

where $\mathcal{O}_z(\frac{1}{B^2})$ means that there exists polynomials whose constants and degrees are independent of N and z such that this quantity is less than $P_1(|z|)P_2(\frac{1}{\mathrm{Im} z})\frac{1}{B^2}$. Moreover, using proper assumptions on the time series, it is possible to show that $\mathcal{O}_z(\frac{1}{B^2})$ is uniform in ν. Finally, the matrix $\tilde{\mathbf{T}}_N(\nu,z)$ is deterministic and is defined implicitly by the unique solution of equation:

$$\tilde{\mathbf{T}}_N(\nu,z) = \mathrm{dg}\left(\frac{-1}{z(1+\overline{\delta}_{N,m}(\nu,z))}, m=1,\ldots,M\right)$$

where the $(\overline{\delta}_{N,m})_{m\geq 1}$ are defined as the solution of the system of M equations:

$$\overline{\delta}_{N,m}(\nu,z) = \frac{1}{B+1}\mathrm{tr}\left(-z\mathbf{I}_{B+1} + \frac{c}{M}\sum_{k=1}^{M}\frac{\boldsymbol{\Theta}_{N,k}(\nu)}{1+\overline{\delta}_{N,k}(\nu,z)}\right)^{-1}\boldsymbol{\Theta}_{N,m}(\nu), \quad m\in[M]$$

We now use the fact that the columns are almost independent, so that it is reasonable to recover a Marcenko-Pastur type behaviour for $\tilde{\mathbf{T}}_N$. We show that the following estimates hold:

$$\sup_{m\geq 1}\sup_{\nu\in[0,1]}\|\boldsymbol{\Theta}_{N,m}(\nu) - \mathbf{I}_{B+1}\| = \mathcal{O}\left(\frac{B}{N}\right)$$

$$\frac{1}{B+1}\sup_{m\geq 1}\sup_{\nu\in[0,1]}\mathrm{tr}\left(\boldsymbol{\Theta}_{N,m}(\nu) - \mathbf{I}_{B+1}\right) - \Upsilon_{N,m}(\nu) = \mathcal{O}\left(\frac{B}{N}\right)^3$$

where $\Upsilon_{N,m}(\nu)$ is the $\mathcal{O}(\frac{B}{N})^2$ term equal to $\Upsilon_{N,m}(\nu) = \frac{1}{2}\frac{s''_m(\nu)}{\hat{s}_m(\nu)}v_N$. This allows to prove that $\tilde{\mathbf{T}}_N(z) - t_N(z)\mathbf{I}_M \to 0$ for any $z\in\mathbb{C}_+$, and obtain that for any $f\in C^p$ with p large enough,

$$\mathbb{E}\frac{1}{M}\mathrm{tr}\; f(\tilde{\mathbf{C}}_N(\nu)) - \int f\,\mathrm{d}\mu_{mp,c_N} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$
$$- \left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}\right)^2 < \tilde{\mathcal{D}}_N, f > v_N + \mathcal{O}\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right)$$

where the $\mathcal{O}(\cdot)$ is uniform in $\nu\in[0,1]$, and where $\tilde{\mathcal{D}}_N$ and $\mathcal{D}_N$ are distributions whose Stieltjes transforms at point z are equal respectively to $(zt_N(z))'$ and $\frac{-c_N(zt_N(z)\tilde{t}_N(z))^3}{1-c_N(zt_N(z)\tilde{t}_N(z))^2}$.

0.3.3 Contribution of Chapter 3

In this chapter, we consider an M–dimensional time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ defined as

$$\mathbf{y}_n = \mathbf{u}_n + \mathbf{v}_n$$

where $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ represents a useful signal and where $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ represents an additive noise. The useful signal is modeled as the output of an unknown stable $M\times K$ MIMO filter $(\mathbf{H}_k)_{k\in\mathbb{Z}}$ driven by a non-observable K–dimensional complex Gaussian white noise $(\boldsymbol{\epsilon}_n)_{n\in\mathbb{Z}}$ with $\mathbb{E}[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^*] = \mathbf{I}_K$, i.e.

$$\mathbf{u}_n = \sum_{k\in\mathbb{Z}}\mathbf{H}_k\boldsymbol{\epsilon}_{n-k}$$

with probability one. We notice that K represents the number of sources in the context of array processing. For each $\nu\in[0,1]$, we denote by $\mathbf{H}(\nu)$ the Fourier transform of $(\mathbf{H}_k)_{k\in\mathbb{Z}}$:

$$\mathbf{H}(\nu) = \sum_{k\in\mathbb{Z}}\mathbf{H}_k\mathrm{e}^{-\mathrm{i}2\pi\nu k}$$

We suppose that the signal satisfies the following short-memory assumption:

$$\sup_{m\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)\|\mathbf{H}_k\| < +\infty$$

while its magnitude is scaled such that

$$\lim_{M\to\infty}\max_{m=1,\ldots,M}\max_{\nu\in[0,1]}\|\mathbf{h}_m(\nu)\|_2 = 0.$$

where $\|\mathbf{h}_m(\nu)\|_2^2 = \sum_{k=1}^{K}|(\mathbf{H})_{m,k}(\nu)|^2$ represents the total power received by the sensor m from the K–dimensional signal at frequency ν. This implies that the total power of the signal is equal to

$$\mathbb{E}\|\mathbf{u}_{m,n}\|^2 = \int_0^1 \|\mathbf{h}_m(\nu)\|^2\,\mathrm{d}\nu = o(1)$$

Concerning the noise, $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is modeled as an M–dimensional stationary complex Gaussian time series such that its component time series $(v_{1,n})_{n\in\mathbb{Z}},\ldots,(v_{M,n})_{n\in\mathbb{Z}}$ are mutually independent. We assume that the noise satisfies the short-memory assumption:

$$\sup_{m\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)^2|r_m(k)| < +\infty$$

where $r_m(k) = \mathbb{E}[v_{m,n}\overline{v}_{m,n-k}]$, and that the spectral density is uniformly bounded from below:

$$\inf_{m\geq 1}\inf_{\nu\in[0,1]} s_m(\nu) > 0$$

where $s_m(\nu) = \sum_{k\in\mathbb{Z}} r_m(k)e^{-\mathrm{i}2\pi\nu k}$. This assumption implies that the total noise power satisfies the following: there exists constants $0 < C_1 \leq C_2 < +\infty$ such that

$$C_1 < \inf_{N\geq 1}\frac{1}{M}\mathbb{E}\|\mathbf{v}_N\|^2 \leq \sup_{N\geq 1}\frac{1}{M}\mathbb{E}\|\mathbf{v}_N\|^2 < C_2$$

Therefore, the signal to noise ratio $\frac{\mathbb{E}\|\mathbf{u}_N\|^2}{\mathbb{E}\|\mathbf{v}_N\|^2}$ satisfies $\frac{\mathbb{E}\|\mathbf{u}_N\|^2}{\mathbb{E}\|\mathbf{v}_N\|^2} = \mathcal{O}(\frac{1}{M})$. The objective of Chapter 3 is to detect the presence of the signal in this context. We start by noticing that the spectral density matrix $\mathbf{S}_\mathbf{y}$ of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is given by

$$\mathbf{S}_\mathbf{y}(\nu) = \mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)$$

where $\mathbf{S}_\mathbf{v}(\nu) = \mathrm{dg}\left(s_1(\nu),\ldots,s_M(\nu)\right)$. Similarly, by the definition of the spectral coherency matrix, it holds that

$$\mathbf{C}_\mathbf{y}(\nu) = \left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)\right)\left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}$$

Under the assumptions on the total power of the signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$, it is reasonable to approximate $\mathrm{dg}(\mathbf{S}_\mathbf{y}(\nu))^{-\frac{1}{2}} = \left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}$ by $\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}$ and $\mathbf{C}_\mathbf{y}(\nu)$ by

$$\mathbf{C}_\mathbf{y}(\nu) \approx \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)\right)\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}} = \mathbf{I}_M + \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}$$

To estimate $\mathbf{C}_\mathbf{y}$, we consider the frequency-smoothed periodogram $\hat{\mathbf{C}}_{N,\mathbf{y}}$ defined by (4). We first prove that $\hat{\mathbf{C}}_\mathbf{y}(\nu)$ behaves asymptotically as a colored Wishart matrix with a population covariance equal to $\mathbf{I}_M + \mathbf{S_v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S_v}(\nu)^{-\frac{1}{2}}$ in spectral norm, and uniformly in ν:

$$\max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{C}}_{N,\mathbf{y}}(\nu) - \mathbf{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} \mathbf{\Xi}(\nu)^{\frac{1}{2}} \right\| \xrightarrow[M\to\infty]{a.s.} 0$$

where $\mathcal{V}_N$ is the set of Fourier frequencies $\left\{\frac{k}{N}, k = 0, \dots, N-1\right\}$,

$$\mathbf{\Xi}(\nu) = \mathbf{I}_M + \mathbf{S_v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S_v}(\nu)^{-\frac{1}{2}}.$$

and $\mathbf{X}_N(\nu)$ is an $M \times (B+1)$ matrix with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries. This result, used in conjunction with Weyl's inequalities [43, Th. 4.3.1], implies in particular that each eigenvalue of the spectral coherency matrix $\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)$ behaves as its counterpart of the Wishart matrix

$$\mathbf{W}_N(\nu) = \mathbf{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} \mathbf{\Xi}(\nu)^{\frac{1}{2}}$$

that is

$$\max_{m=1,\dots,M} \max_{\nu \in \mathcal{V}_N} \left| \lambda_m\left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)\right) - \lambda_m\left(\mathbf{W}_N(\nu)\right) \right| \xrightarrow[M\to\infty]{a.s.} 0.$$

We remark now that $\mathbf{\Xi}(\nu)$ is a finite rank K perturbation of the identity matrix $\mathbf{I}_M$. This suggests to use the results of the spike model covariance matrix from [7] to describe precisely the behaviour of the largest eigenvalues of $\mathbf{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} \mathbf{\Xi}(\nu)^{\frac{1}{2}}$. This enables to design a statistics to test for $\mathcal{H}_0$ against $\mathcal{H}_1$ defined by:

$$\begin{aligned} \mathcal{H}_0 &: (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{v}_n)_{n\in\mathbb{Z}} \\ \mathcal{H}_1 &: (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{u}_n)_{n\in\mathbb{Z}} + (\mathbf{v}_n)_{n\in\mathbb{Z}} \end{aligned}$$

We consider the largest eigenvalue of $\hat{\mathbf{C}}_N(\nu)$ over the frequencies of $\mathcal{V}_N$, and compare it with λ_+:

$$T_{N,\epsilon} = \mathbb{1}_{[\lambda^+ + \epsilon, \infty)}\left(\max_{\nu \in \mathcal{V}_N} \lambda_1\left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)\right)\right).$$

where we recall that $\lambda_+ = (1+\sqrt{c})^2$. This test is proven to be consistent in our high-dimensional regime as follows: if under Hypothesis $\mathcal{H}_1$, there exists $\gamma_\infty \geq 0$ such that

$$\max_{\nu \in \mathcal{V}_N} \lambda_1\left(\mathbf{S_v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S_v}(\nu)^{-\frac{1}{2}}\right) \xrightarrow[M\to\infty]{} \gamma_\infty > \sqrt{c}$$

then defining

$$\phi(x) = \begin{cases} \frac{(x+1)(x+c)}{x} & \text{if } x > \sqrt{c} \\ \lambda^+ & \text{if } x \leq \sqrt{c} \end{cases}$$

we prove that for all $0 < \epsilon < \phi(\gamma_\infty) - \lambda^+$ and $i \in \{0,1\}$,

$$\mathbb{P}_i\left(\lim_{M\to\infty} T_{N,\epsilon} = i\right) = 1$$

where $\mathbb{P}_i$ is the underlying probability measure under Hypothesis $\mathcal{H}_i$. Note that γ_∞, the maximum eigenvalue of the finite rank perturbation, may be interpreted as a certain SNR metric in the frequency domain. Finally, we prove that a test based on the linear spectral statistics of $\hat{\mathbf{C}}_N(\nu)$

does not allow discrimination between these two hypotheses. Consider the linear spectral statistics

$$L_{N,\varphi}(\nu) = \frac{1}{M}\sum_{m=1}^{M} \varphi\left(\lambda_m\left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)\right)\right)$$

where φ belongs to a certain class of functions, and here we require that $\varphi \in \mathcal{C}^1((0,+\infty))$. It will be proved that

$$\max_{\nu\in\mathcal{V}_N}\left|L_{N,\varphi}(\nu) - \int_{\mathbb{R}}\varphi(\lambda)f(\lambda)\mathrm{d}\lambda\right| \xrightarrow[M\to\infty]{a.s.} 0$$

where f is the density of the Marcenko-Pastur distribution for $c<1$ given by

$$f(\lambda) = \frac{\sqrt{(\lambda-\lambda^-)(\lambda^+-\lambda)}}{2\pi c\lambda}\mathbb{1}_{[\lambda^-,\lambda^+]}(\lambda)$$

with $\lambda^\pm = (1\pm\sqrt{c})^2$. This shows that linear spectral statistics of the sample spectral coherency matrix converge to the same limit regardless of whether the observations contain only pure noise or signal-plus-noise contributions. This shows that any test statistic solely relying on linear spectral statistics of the sample spectral coherency matrix is unable to distinguish between absence or presence of useful signal, and cannot be consistent in the high-dimensional regime.

0.3.4 Contribution of Chapter 4

In this chapter we consider the same settings and context as that of Chapter 3, except that we do not assume that $\lim_{M\to\infty}\max_{m=1,\dots,M}\max_{\nu\in[0,1]}\|\mathbf{h}_m(\nu)\|_2 = 0$. This assumption implied that the total power received by each sensor m from the K–dimensional signal at frequency ν converges toward zero. In some context, it can be of interest to not rely on this assumption, and it will be seen that computations can still be done explicitly. The main consequence of this is that we cannot anymore rely on the approximation $\mathrm{dg}(\mathbf{S}_\mathbf{y}(\nu)) \approx \mathrm{dg}(\mathbf{S}_\mathbf{v}(\nu))$ and we instead have to write $\mathrm{dg}(\mathbf{S}_\mathbf{y}(\nu)) = \mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)$. We denote $\mathbf{D}_\mathbf{u}(\nu) = \mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*)$ and suppose that there exists for each $\nu\in[0,1]$ a sequence of $M\times M$ diagonal matrices $(\mathbf{D}_{\mathbf{u},\infty}(\nu))_{N\geq 1}$ with a bounded number of non-zero entries such that

$$\|\mathbf{D}_\mathbf{u}(\nu) - \mathbf{D}_{\mathbf{u},\infty}(\nu)\| \xrightarrow{N\to+\infty} 0.$$

The condition on the number of non-zero entries of $\mathbf{D}_{\mathbf{u},\infty}(\nu)$ ensures that this matrix rank remains asymptotically finite, which will be useful to apply results on spike models from [7]. Using the same strategy as in Chapter 3, we show that for each $\nu\in\mathcal{V}_N$, there exist a random matrix $\mathbf{X}_N(\nu)$ with i.i.d. $\mathcal{N}_\mathbb{C}(0,1)$ entries such that

$$\max_{\nu\in\mathcal{V}_N}\left\|\hat{\mathbf{C}}_N(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\right\| \xrightarrow{N\to+\infty} 0$$

where

$$\boldsymbol{\Xi}(\nu) = \mathbf{I}_M + (\mathbf{D}_{\mathbf{u},\infty}(\nu)+\mathbf{S}_\mathbf{v}(\nu))^{-\frac{1}{2}}(\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu))(\mathbf{D}_{\mathbf{u},\infty}(\nu)+\mathbf{S}_\mathbf{v}(\nu))^{-\frac{1}{2}}$$

This is in contrast to the result from Chapter 3, where the population covariance matrix was equal to $\mathbf{I}_M + \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*)(\mathbf{S}_\mathbf{v}(\nu))^{-\frac{1}{2}}$. The new expression of $\boldsymbol{\Xi}(\nu)$ implies that the perturbation term may contains negative spike since the matrix $\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)$ is not necessarily positive definite. Moreover, as the rank of $\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)$ is finite, results on spike models from [7] prove that the largest and smallest eigenvalues of $\hat{\mathbf{C}}_N(\nu)$ may leave the bulk of the Marcenko-Pastur distribution if a certain SNR condition is respected. A new statistical test to detect the presence or absence of signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ is proposed:

$$T_\epsilon = \mathbf{1}\left(\sup_{\nu\in\mathcal{V}_N} \lambda_1(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) > \lambda^+ + \epsilon \text{ or } \inf_{\nu\in\mathcal{V}_N} \lambda_M(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) < \lambda^- - \epsilon\right)$$

and it is shown that this test is consistent to test between the null hypothesis $\mathcal{H}_0$ and the alternative $\mathcal{H}_1$ if

$$\gamma_\infty^+ > \sqrt{c} \text{ or } \gamma_\infty^- < -\sqrt{c}.$$

where γ_∞^+ and γ_∞^- are defined as:

$$\max_{\nu\in\mathcal{V}_N} \lambda_1\left(\mathbf{\Omega}(\nu)\right) \xrightarrow[M\to\infty]{} \gamma_\infty^+ > \sqrt{c}$$
$$\min_{\nu\in\mathcal{V}_N} \lambda_M\left(\mathbf{\Omega}(\nu)\right) \xrightarrow[M\to\infty]{} \gamma_\infty^- < -\sqrt{c}$$

with $\mathbf{\Omega}(\nu) = \left(\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu)\right)^{-\frac{1}{2}}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)\right)\left(\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu)\right)^{-\frac{1}{2}}$.

0.3.5 Contribution of Chapter 5

This chapter stands apart from all the other chapters, as it is not interested in the eigenvalue behaviour of the spectral coherency matrix, but instead is interested in the behaviour of its individual entries. More precisely, we consider M jointly stationary complex Gaussian time series $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ and for all $i,j \in \{1,\ldots,M\}$, we denote by s_{ij} and c_{ij} the spectral density and spectral coherence between $(y_{i,n})_{n\in\mathbb{Z}}$ and $(y_{j,n})_{n\in\mathbb{Z}}$ given respectively by $s_{ij}(\nu) = (\mathbf{S}(\nu))_{ij}$ and $c_{ij}(\nu) = (\mathbf{C}(\nu))_{ij}$ where $\mathbf{S}(\nu)$ and $\mathbf{C}(\nu)$ are respectively the spectral density/coherency matrices defined in (1) and (2). This chapter is motivated by the problem of testing the independence of a large number of Gaussian time series:

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |s_{ij}(\nu)|^2 = 0,$$

or equivalently

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |c_{ij}(\nu)|^2 = 0,$$

This suggests to compute consistent estimators of these quantities, and test their closeness to zero. Assuming that N observations $(y_{1,n})_{n=1,\ldots,N}, \ldots, (y_{M,n})_{n=1,\ldots,N}$ are available for each time series, we consider the frequency smoothed estimate $\hat{c}_{N,ij}$ of c_{ij} given by (4). Under the hypothesis

$$\mathcal{H}_0 : (y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}} \text{ are mutually uncorrelated},$$

and some technical assumptions on the spectral densities of the M time series, it is shown that in the high-dimensional regime described above in (5),

$$\mathbb{P}\left((B+1)\max_{(i,j,\nu)\in\mathcal{I}_N} |\hat{c}_{N,ij}(\nu)|^2 \le t + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2}\right) \xrightarrow{N\to+\infty} e^{-e^{-t}} \tag{8}$$

where

$$\mathcal{I}_N := \{(i,j,\nu) : i,j \in [M] \text{ such that } i<j,\ \nu\in\mathcal{G}_N\}$$

with $[M] = \{1, \dots, M\}$ and where

$$\mathcal{G}_N := \left\{ k\frac{B+1}{N} : k \in \mathbb{N}, 0 \le k \le \frac{N}{B+1} \right\}$$

is the subset of the Fourier frequencies

$$\mathcal{V}_N := \left\{ \frac{k}{N} : k \in \mathbb{N}, 0 \le k \le N-1 \right\}$$

with elements spaced by a distance $(B+1)/N$. Thus, (8) states that $\max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{N,ij}(\nu)|^2$, after proper normalization and centering, converges in distribution to a type I extreme value distribution, also known as Gumbel distribution. As it will be clear in the proof, the term $\log \frac{M(M-1)}{2}$ is related to the maximum over (i,j) while the term $\log \frac{N}{B+1}$ is related to the maximum over $\nu \in \mathcal{G}_N$. The main steps of the proof are as follows. We start by writing the linear causal representation

$$y_{m,n} = \sum_{k\ge 0} a_{m,k}\epsilon_{m,n-k}$$

where $(\epsilon_{1,k})_{k\in\mathbb{Z}}, \dots, (\epsilon_{M,k})_{k\in\mathbb{Z}}$ are mutually independent sequences of $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. random variables, and $(a_{1,k})_{k\in\mathbb{N}}, \dots, (a_{M,k})_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Denote

$$h_m(\nu) = \sum_{k=0}^{+\infty} a_{m,k} e^{-2\mathrm{i}\pi k\nu}$$

so that $|h_m(\nu)|^2 = s_m(\nu)$. It is shown in [94] (and extended under weaker assumptions in [67] and [19]) that the Fourier transforms $\xi_{N,y_m}(\nu)$ behaves like $h_m(\nu)\xi_{N,\epsilon_m}(\nu)$ in the following way: for any $\kappa > 0$:

$$\mathbb{P}\left[\lim_{N\to+\infty} |\xi_{N,y_m}(\nu) - h_m(\nu)\xi_{N,\epsilon_m}(\nu)| > \kappa \right] = 0 \tag{9}$$

This suggests to define $\tilde{s}_{N,ij}(\nu)$, an approximation of $\hat{s}_{N,ij}(\nu) := (\hat{\mathbf{S}}_N(\nu))_{ij}$ where $\hat{\mathbf{S}}_N(\nu)$ is defined by (3), as:

$$\tilde{s}_{N,ij}(\nu) = \frac{1}{B+1} \sum_{b=-B/2}^{B/2} h_i\left(\nu + \frac{b}{N}\right) \overline{h_j\left(\nu + \frac{b}{N}\right)} \xi_{N,\epsilon_i}\left(\nu + \frac{b}{N}\right) \overline{\xi_{N,\epsilon_j}\left(\nu + \frac{b}{N}\right)}$$

Equation (9) is then key to prove that $(B+1)\sup_{i,j,\nu} |\hat{c}_{N,ij}(\nu)|^2$ has the same limiting distribution as $(B+1)\sup_{i,j,\nu} |\tilde{c}_{N,ij}(\nu)|^2$ where $\tilde{c}_{N,ij}(\nu) = \frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)}$ with

$$\sigma^2_{N,ij}(\nu) = \frac{1}{B+1} \sum_{b=-B/2}^{B/2} s_i\left(\nu + \frac{b}{N}\right) s_j\left(\nu + \frac{b}{N}\right)$$

This decomposition is helpful since we decoupled the action of the filter from the innovations, so that it remains now to deal with the Fourier transforms of the i.i.d. random variables $\epsilon_{m,n}$. In order to show the limiting Gumbel distribution for $\sup_{i,j,\nu} |\tilde{c}_{N,ij}(\nu)|^2$, we rely on a Poisson approximation from [3] based on the Chen-Stein method. This method has been used by [45] to study the behaviour of the largest entry of the Pearson correlation matrix of i.i.d. data. Moreover, this method requires to compute the two following moderate deviations: there exists a constant $\eta > 0$ such that for any

$C > 0$, we have

$$\max_{t \in [0, CB^\eta]} \max_{(i,j,\nu) \in \mathcal{I}} \left| \mathbb{P}\left((B+1) \frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)} > t^2 \right) e^{t^2} - 1 \right| \xrightarrow[N \to \infty]{} 0 \tag{10}$$

and

$$\max_{t,s \in [0, CB^\eta]} \max_{\substack{(i,j,\nu) \in \mathcal{I} \\ (i',j',\nu) \in \mathcal{I}_{(i,j,\nu)}}} \left| \mathbb{P}\left((B+1) \frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)} > t^2, \, (B+1) \frac{|\tilde{s}_{N,i'j'}(\nu)|^2}{\sigma^2_{N,i'j'}(\nu)} > s^2 \right) \times e^{t^2+s^2} - 1 \right| \xrightarrow[N \to \infty]{} 0.$$

Using these estimates ends the proof of (8).

0.4 Perspectives

We end this Introduction by stating some possible directions of future works which could improve the results presented in this manuscript.

- First, as mentioned in the contributions of Chapter 1, more information about the asymptotic distribution of

$$T_N = \sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\, f(\hat{\mathbf{C}}_N(\nu)) - \int f \, \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M} \sum_{m=1}^{M} \frac{s'_m(\nu)}{s_m(\nu)} \right)^2 < \mathcal{D}_N, f > v_N \right|$$

is needed to control the type-I error of a statistical test based on it (for instance $\mathbb{1}(T_N > \kappa)$ for some threshold κ). However, we feel that it is very difficult to obtain such a result. A more reasonable approach would be the following one: write, as in Chapter 5, the linear causal representation $y_{m,n} = \sum_{k \in \mathbb{Z}} a_{m,k} \epsilon_{m,n-k}$ with $(\epsilon_{1,n})_{n \in \mathbb{Z}}, \ldots, (\epsilon_{M,n})_{n \in \mathbb{Z}}$ mutually independent sequences of $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. random variables, and $(a_{1,k})_{k \in \mathbb{Z}}, \ldots, (a_{M,k})_{k \in \mathbb{Z}} \in \ell^2(\mathbb{N})$. Define $h_m(\nu) = \sum_{k \in \mathbb{Z}} a_{m,k} e^{-2\mathrm{i}\pi k \nu}$, $\mathbf{H}(\nu) = \mathrm{dg}(h_1(\nu), \ldots, h_M(\nu))$, and $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu) = (\boldsymbol{\xi}_{N,\epsilon_1}(\nu), \ldots, \boldsymbol{\xi}_{N,\epsilon_M}(\nu))^T$. Using the well-known approximation $\boldsymbol{\xi}_{N,\mathbf{y}}(\nu) \approx \mathbf{H}(\nu) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu)$, which has already been used in Chapter 5, it makes sense to consider $\check{\mathbf{S}}_N(\nu)$, an approximation of $\hat{\mathbf{S}}_N(\nu)$, defined by

$$\check{\mathbf{S}}_N(\nu) = \frac{1}{B+1} \sum_{-B/2}^{B/2} \mathbf{H}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}\left(\nu + \frac{b}{N}\right)^* \mathbf{H}\left(\nu + \frac{b}{N}\right)^*$$

and $\check{\mathbf{C}}_N(\nu)$, an approximation of $\hat{\mathbf{C}}_N(\nu)$ defined by

$$\check{\mathbf{C}}_N(\nu) = \mathrm{dg}(\check{\mathbf{C}}_N(\nu))^{-\frac{1}{2}} \check{\mathbf{C}}_N(\nu) \mathrm{dg}(\check{\mathbf{C}}_N(\nu))^{-\frac{1}{2}}.$$

For $\nu_1 \neq \nu_2 \in \mathcal{G}_N := \left\{ k \frac{B+1}{N} : k \in \mathbb{N}, 0 \leq k \leq \frac{N}{B+1} \right\}$, $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu_1)$ is independent of $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu_2)$, so it is clear that $\frac{1}{M} \mathrm{tr}\, f(\check{\mathbf{C}}_N(\nu_1))$ is also independent of $\frac{1}{M} \mathrm{tr}\, f(\check{\mathbf{C}}_N(\nu_2))$. This should enable to obtain the asymptotic distribution of $\check{T}_N$ defined by

$$\check{T}_N = \sup_{\nu \in \mathcal{G}_N} \left| \frac{1}{M} \mathrm{tr}\, f(\check{\mathbf{C}}(\nu)) - \int f \, \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M} \sum_{m=1}^{M} \frac{s'_m(\nu)}{s_m(\nu)} \right)^2 < \mathcal{D}_N, f > v_N \right|$$

It would remain then to control the error rate between $\check{\mathbf{C}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$ to ensure that $\frac{1}{M}\mathrm{tr}\ f(\check{\mathbf{C}}_N(\nu))$ and $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$ follow asymptotically the same distribution. This would represent an improvement over our results from Chapter 1 and Chapter 2, since it would allow us to control the type-I error of the test statistics $\mathbb{1}(\sup_{\nu \in \mathcal{G}_N} |T_N(\nu)| > \kappa)$. However, it is still unclear how to control the supremum over the continuous range of frequencies $[0, 1]$ instead of the discrete grid $\mathcal{G}_N$.

- In chapter 1 we focused on $\sup_\nu |T_N(\nu)|$, but it could be useful for the applications to study some kind of integrated statistics as was done in [25], such as $\int_{\nu \in [0,1]} |T_N(\nu)|^2 \, \mathrm{d}\nu$.

- Once the asymptotic distribution of T_N is known under $\mathcal{H}_0$, it would be natural to study the power of such a test under a defined class of alternatives where the spectral density matrix $\mathbf{S}(\nu)$ is not diagonal for every ν, using similar tools as those used in [15] for instance.

- The Gaussian assumption on the time series is crucial in all the chapters in this manuscript, and it could be of interest to extend our results to non-Gaussian time series. For instance, in Chapter 1, the Gaussian assumption on the observations was necessary to approximate $\tilde{\mathbf{C}}_N(\nu)$ by $\frac{1}{B+1}\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*$ where $\mathbf{X}_N(\nu)$ is an $M \times (B+1)$ matrix with i.i.d. entries. Without the Gaussian assumption, the entries of $\mathbf{X}_N(\nu)$ are not anymore i.i.d., thus preventing us to follow the strategy used in the proofs of Chapter 1. However, the approach used in Chapter 2 seems more adapted to handle non-Gaussian cases, as the corresponding model was first studied in [92] under condition of finite moment of order eight. However, the majority of the remaining work would be about controlling the new error rates implied by the weaker moment conditions.

- It should be possible to obtain results about the local behaviour of the eigenvalues of $\hat{\mathbf{S}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$. For the applications in signal detection, it would be valuable to prove a Tracy-Widom behaviour for the largest eigenvalues of the spectral density/coherency matrices.

- Lastly, we stress the fact that we systematically approximated $\hat{\mathbf{C}}_N(\nu)$ by $\tilde{\mathbf{C}}_N(\nu) = \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\hat{\mathbf{S}}_N(\nu)\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$, and then studied $\tilde{\mathbf{C}}_N(\nu)$ instead of $\hat{\mathbf{C}}_N(\nu)$. The advantage of $\tilde{\mathbf{C}}_N(\nu)$ over $\hat{\mathbf{C}}_N(\nu)$ is that the renormalization by $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ is not random anymore. Nevertheless, it also requires that $\mathrm{dg}(\mathbf{S}(\nu))$ does not vanish, therefore imposing the technical condition $\inf_{\nu\in[0,1]} \inf_{m\geq 1} s_m(\nu) > 0$. It could be of interest to find another approach to avoid having to rely on this assumption.

0.5 Journal/Conference papers associated with the manuscript

- Chapter 1 has been accepted by the Electronic Journal of Statistics, and will soon be published.

- Chapter 3 has first been published at the ICASSP conference: Alexis Rosuel et al. "On the frequency domain detection of high dimensional time series". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 8782–8786, and later extended and published by IEEE Transactions on Signal Processing: A. Rosuel et al. "On the Detection of Low-Rank Signal in the Presence of Spatially Uncorrelated Noise: A Frequency Domain Approach". In: *IEEE Transactions on Signal Processing* 69 (2021), pp. 4458–4473. DOI: 10.1109/TSP.2021.3099644. We only present in this manuscript the IEEE TSP paper in Chapter 3.

- Chapter 5 is currently under review by the Journal of Multivariate Analysis.

0.6 Notations

- We assume that $M := M(N)$ and $B = B(N)$ are integer functions of N, and that $M(N), B(N) \xrightarrow{N\to+\infty} +\infty$ in such a way that $c_N := \frac{M}{B+1}$ satisfies $0 < \inf_{N\geq 1} c_N \leq \sup_{N\geq 1} c_N < +\infty$ and $c_N \xrightarrow{N\to+\infty} c \in (0,+\infty)$. Moreover, there exists $\frac{1}{2} < \alpha < 1$ such that $B = \mathcal{O}(N^\alpha)$. This implies in particular that $\frac{B}{N} \xrightarrow{N\to+\infty} 0$ and $\frac{M}{N} \xrightarrow{N\to+\infty} 0$. This regime will be referred as $N \to +\infty$ in the following.

- The various quantities introduced above and studied in the manuscript will usually depend on N and ν, but we choose to omit the explicit notation on the dependence of these parameters when the context is clear in order to simplify the notations. For instance, $\hat{\mathbf{S}}$ can be used to represents $\hat{\mathbf{S}}_N(\nu)$, which is the estimated spectral density matrix at frequency ν obtained from the sample $\mathbf{y}_1, \ldots, \mathbf{y}_N$. Similarly, $\hat{\mathbf{C}}_N(\nu), \hat{c}_{N,ij}(\nu), \boldsymbol{\xi}_N(\nu), \ldots$ can be represented by $\hat{\mathbf{C}}, \hat{c}_{ij}, \boldsymbol{\xi}, \ldots$.

- $\mathbb{R}_+$ and $\mathbb{R}_-$ represent respectively the set of all non-negative numbers and non-positive numbers, and we denote $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$ and $\mathbb{R}_-^* = \mathbb{R}_- \setminus \{0\}$. We also define $\mathbb{C}_+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$.

- By a nice constant, we mean a positive deterministic constant which does not depend on the parameters M, B and N, or on the complex variable z. In the following, C will represent a generic nice constant whose value may change from one line to the other. A nice polynomial P is a polynomial whose degree and coefficients are nice constants.

- For a matrix $\mathbf{A}$, the notations $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ represent its spectral norm and Frobenius norm respectively. The transpose and the conjugate transpose of $\mathbf{A}$ are respectively denoted by $\mathbf{A}^T$ and $\mathbf{A}^*$. For a matrix $\mathbf{B}$ of the same size, $\mathbf{A} \geq \mathbf{B}$ stands for $\mathbf{A} - \mathbf{B}$ is non-negative definite. If moreover A is a square matrix, $\mathrm{Im}(\mathbf{A})$ is the Hermitian matrix defined by $\mathrm{Im}(\mathbf{A}) = \frac{\mathbf{A}-\mathbf{A}^*}{2\mathrm{i}}$.

- $C_c(\mathbb{R}, \mathbb{R})$ represents the set of all real valued compactly supported functions defined on $\mathbb{R}$. $C^{(k)}(\mathbb{R}, \mathbb{R})$ represents the set of all real valued k–differentiable functions defined on $\mathbb{R}$. $\mathcal{C}^1(I)$ (resp. $\mathcal{C}_c^1(I)$) represents the set of continuously differentiable functions (resp. continuously differentiable functions with compact support) on an open set I.

- If ζ is a random variable, we denote by ζ° the zero mean random variable defined by $\zeta^\circ = \zeta - \mathbb{E}\zeta$.

- The complex circular Gaussian distribution with variance σ^2 is denoted as $\mathcal{N}_\mathbb{C}(0, \sigma^2)$ and a random vector $\mathbf{x}$ of $\mathbb{C}^n$ follows the $\mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{R})$ distribution if $\mathbf{b}^*\mathbf{x} \sim \mathcal{N}_\mathbb{C}(0, \mathbf{b}^*\mathbf{R}\mathbf{b})$ for all deterministic (column) vector $\mathbf{b}$ and a fixed $n \times n$ positive definite matrix $\mathbf{R}$.

Introduction (Français)

Ce manuscrit de thèse est consacré à l'étude asymptotique de la matrice de cohérence spectrale estimée de séries temporelles gaussiennes de grande dimension. Si l'on désigne par M la dimension de la série temporelle et N le nombre d'observations, la théorie des séries temporelles classique traite le cas où M est fixé et $N \to +\infty$. Ce contexte permet de déduire plusieurs résultats importants. Ces résultats étudiés en petite dimension échouent en général quand M est considéré comme une fonction de N, de façon que $M(N) \xrightarrow{N\to+\infty} +\infty$. Les problèmes statistiques actuels sont souvent confrontés à ce contexte, soit parce que la dimension M du problème est grande, soit parce que ce n'est pas possible de collecter un nombre suffisamment grand d'observations N de façon que N soit en pratique pas beaucoup plus grand que M (ce peut être le cas par exemple parce que la série temporelle observée n'est pas stationnaire et que ses paramètres évoluent rapidement, ou parce que la durée de l'observation est petite). Dans ces situations, il devient délicat d'appliquer les résultats théorie statistique en petite dimension. Ce nouveau régime asymptotique a suscité une certaine quantité de travaux de recherche, et une importante proportion d'entre eux a utilisé les résultats de la théorie des matrices aléatoires. La théorie des matrices aléatoires étudie les propriétés asymptotiques des valeurs propres de grandes matrices aléatoires quand M et N divergent de façon que $M/N \to c \in (0, +\infty)$. Cette approche s'est révélée intéressante, et en particulier a permis de mettre en évidence les propriétés des valeurs propres et des vecteurs propres des matrices de covariance et de corrélation, qui sont des objets de première importance pour les applications statistiques. Cette approche basée sur les matrices aléatoires a également permis de concevoir de nouvelles approches plus performantes. Cependant, d'autres propriétés liées au comportement de la série temporelle nécessitent plus d'informations que celles fournies par la matrice de covariance, et certaines applications peuvent nécessiter des informations détaillées sur la matrice de densité spectrale et la matrice de cohérence spectrale dans le régime asymptotique des grandes dimensions. Le but de ce manuscrit est de présenter quelques résultats obtenus sur le comportement d'un estimateur bien connu de la matrice de cohérence spectrale d'une série temporelle gaussienne de grande dimension, et de montrer comment ces résultats peuvent être utilisés pour étudier quelques problèmes importants de traitement du signal, notamment pour tester si les composantes d'une série temporelle M-dimensionnelle sont indépendantes ou non.

0.7 Motivation

Considérons une série temporelle stationnaire complexe Gaussienne[2] de dimension M. Une série temporelle $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ est dite Gaussienne si et seulement si pour tout ensemble fini d'instant $N_1, N_2, \ldots, N_k$, $\mathbf{y}_{N_1}, \mathbf{y}_{N_2}, \ldots, \mathbf{y}_{N_k}$ sont normalement distribués. Si on note $\mathbf{R}_u = \mathbb{E}[\mathbf{y}_{n+u}\mathbf{y}_n^*]$ son autocovariance au lag u, on peut définir la matrice spectrale de densité $\mathbf{S}(\nu)$ à la fréquence ν par

$$\mathbf{S}(\nu) = (s_{ij}(\nu))_{i,j=1}^M = \sum_{u\in\mathbb{Z}} \mathbf{R}_u e^{-\mathrm{i}2\pi u\nu} \tag{11}$$

[2]La distribution circulaire Gaussienne de variance σ^2 est notée $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ et définie par $X_1 + iX_2$ où $X_1, X_2 \sim \mathcal{N}(0, \sigma^2/2)$. i.i.d. Un vecteur aléatoire $\mathbf{x}$ dans $\mathbb{C}^n$ suit la distribution $\mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{R})$ si $\mathbf{b}^*\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{b}^*\mathbf{R}\mathbf{b})$ pour tous les vecteurs (colonne) $\mathbf{b}$ déterministes et un matrice $n \times n$ positive définie $\mathbf{R}$.

et la matrice de cohérence spectrale (voir par exemple [13, chapitre 7-6], [52, chapitre 5.5]) à la fréquence ν, un analogue de la corrélation dans le domaine fréquenciel, par

$$\mathbf{C}(\nu) = (c_{ij}(\nu))_{i,j=1}^{M} = \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\mathbf{S}(\nu)\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}} \tag{12}$$

où $\mathrm{dg}(\mathbf{S}(\nu)) = \mathbf{S}(\nu) \odot \mathbf{I}_M$, avec $\odot$ désignant le produit Hadamard (produit entrée par entrée) et $\mathbf{I}_M$ désignant la matrice identité de taille M.

Ces quantités, $\mathbf{S}(\nu)$ et $\mathbf{C}(\nu)$, fournissent des informations utiles sur les propriétés dynamiques de second ordre de $\mathbf{y}$, ce qui, du point de vue statistique, pourrait être important à comprendre et à extraire d'un échantillon fini $(\mathbf{y}_n)_{n=1,\ldots,N}$. En effet, la densité spectrale combine l'information de l'autocovariance de différents lags l à une fréquence $\nu \in [-\frac{1}{2}, \frac{1}{2}]$. Par exemple, il est montré dans [68] que $c_{ij}(\nu)$, la cohérence entre les séries i et j à la fréquence ν, est équivalente au carré de la corrélation entre les composantes ν-oscillatoires des deux séries temporelles. On note aussi que l'autocovariance peut être retrouvée à partir de la densité spectrale par

$$\mathbf{R}_l = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\mathrm{i}\pi l\nu}\mathbf{S}(\nu)\,\mathrm{d}\nu$$

Dans ce manuscrit, nous détaillerons quelques applications en analyse statistique et en traitement du signal, telles que le test d'indépendance entre les M composantes de $\mathbf{y}$, et la détection d'une signal de petit rang corrélé spatialement dans un espace de grande dimension. Pour atteindre ces objectifs, nous nous appuierons sur le périodogramme moyenné, un estimateur de $\mathbf{S}(\nu)$ à la fréquence ν défini par

$$\hat{\mathbf{S}}_N(\nu) = (\hat{s}_{N,ij}(\nu))_{i,j=1}^{M} = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_N\left(\nu + \frac{b}{N}\right)\boldsymbol{\xi}_N\left(\nu + \frac{b}{N}\right)^* \tag{13}$$

où B est un entier pair, qui représente la taille du fenêtrage, et

$$\boldsymbol{\xi}_N(\nu) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{y}_n e^{-2\mathrm{i}\pi(n-1)\nu}$$

est la transformée de Fourier renormalisée de $(\mathbf{y}_n)_{n=1\ldots,N}$. Le matrice de cohérence spectrale estimée est alors définie comme ceci:

$$\hat{\mathbf{C}}_N(\nu) = (\hat{c}_{N,ij})_{i,j=1}^{M}(\nu) = \mathrm{dg}\left(\hat{\mathbf{S}}_N(\nu)\right)^{-\frac{1}{2}}\hat{\mathbf{S}}_N(\nu)\mathrm{dg}\left(\hat{\mathbf{S}}_N(\nu)\right)^{-\frac{1}{2}} \tag{14}$$

Les chapitres suivants sont motivés par la compréhension de certaines propriétés de $\hat{\mathbf{C}}_N(\nu)$ quand la dimension M de $\mathbf{y}$ est grande et les composantes de $\mathbf{y}$ sont indépendantes. Avant de mentionner les résultats obtenus, nous rappelons que les propriétés de quantités dérivées de $\hat{\mathbf{S}}_N(\nu)$ et $\hat{\mathbf{C}}_N(\nu)$ ont d'abord été étudiées dans le régime où $N \to \infty$ et M est fixe, un régime que nous appellerons des *petites dimensions* dans le reste de ce manuscrit. Il est bien connu (voir le paragraphe sur la revue de la littérature ci-dessous pour plus de détails) que $\hat{\mathbf{S}}_N(\nu)$ et $\hat{\mathbf{C}}_N(\nu)$ sont des estimateurs consistants de $\mathbf{S}(\nu)$ et $\mathbf{C}(\nu)$ dans le régime des petites dimensions si $B \to +\infty$ et $\frac{B}{N} \to 0$. Sous des hypothèses supplémentaires sur la mémoire des séries temporelles $(\mathbf{y}_n)_{n\in\mathbb{Z}}$, $\hat{\mathbf{S}}_N(\nu)$ est en outre un estimateur asymptotiquement normal de $\mathbf{S}(\nu)$, qui peut, à son tour, être utilisée pour étudier la performance asymptotique de différents tests basés sur $\hat{\mathbf{C}}_N(\nu)$.

Dans le cas où $M \to +\infty$, pour que $\hat{\mathbf{S}}_N(\nu)$ soit toujours un estimateur consistante de $\mathbf{S}(\nu)$, il est nécessaire que $\frac{M}{B} \to 0$. Un exemple simple qui illustre cela est le suivant. Considérons $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ une suite de variables aléatoires indépendantes $\mathcal{N}_{\mathbb{C}}(0,1)$. Dans ce contexte, pour chaque ν, les vecteurs de Fourier $(\boldsymbol{\xi}_N(\nu + b/N))_{b=-B/2,\ldots,B/2}$ sont des vecteurs de variables aléatoires indépendantes de loi

$\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$, et la densité spectrale estimée $\hat{\mathbf{S}}_N(\nu)$ est égale à la matrice de covariances de ces $(B+1)$ vecteurs M–dimensionnel. Si on désigne par c_N le ratio $\frac{M}{B+1}$ et on suppose que $c_N \to c$, il est connu depuis [63] que $\|\hat{\mathbf{S}}_N(\nu) - \mathbf{S}(\nu)\| \xrightarrow{N \to +\infty} (1+\sqrt{c})^2 - 1$ presque sûrement. Cette quantité est nulle si et seulement si $c = 0$, ce qui signifie que $\frac{M}{B} \to 0$.

De nos jours, dans la plupart des applications impliquant des signaux de grande dimension et/ou un petit échantillon, le rapport $\frac{M}{N}$ peut ne pas être suffisamment petit pour choisir B de manière à satisfaire $\frac{M}{B}$ et $\frac{B}{N}$ nettement inférieur à 1. Dans ce contexte, le comportement de $\hat{\mathbf{C}}_N(\nu)$ n'est pas bien prédit par les résultats obtenus dans le domaine des petites dimensions ($\frac{B}{N} \to 0$ avec M fixé ou $\frac{M}{B} \to 0$). En particulier, $\hat{\mathbf{C}}_N(\nu)$ n'a pas de raison d'être proche de $\mathbf{C}(\nu)$. Dans ce cas, il semble plus intéressant de choisir B du même ordre de grandeur que M, et de se baser sur le régime des grandes dimensions dans lequel $M := M(N), B := B(N)$ et N convergent vers l'infini tel que $\frac{M}{B}$ converge vers une constante positive $c \in (0, +\infty)$ tandis que $\frac{B}{N}$ converge vers zéro. Comme on le voit ci-dessous, les méthodes de grandes matrices aléatoires rendent alors possible une bonne compréhension du comportement de $\hat{\mathbf{C}}_N(\nu)$. Dans ce qui suit, on suppose en outre que

$$B = \mathcal{O}(N^{\alpha}) \text{ pour un certain } \alpha \in \left(\frac{1}{2}, 1\right)$$

Bien entendu, puisque $\frac{M}{B} \to c \in (0, +\infty)$, il est clair que M est également $\mathcal{O}(N^{\alpha})$ pour $\alpha \in \left(\frac{1}{2}, 1\right)$, et $\frac{M}{N} \to 0$. On note que α est pris supérieur à $\frac{1}{2}$ de manière à ce que M ne soit pas trop petit par rapport à N. Dans le cas contraire, il apparaîtrait de nouveau intéressant de choisir B de manière à ce que simultanément $M \ll B$ et $B \ll N$. Ceci signifie que, dans le cas $\alpha < \frac{1}{2}$, choisir B du même ordre de grandeur que M ne présente pas d'intérêt et ne ferait que rendre nos résultats plus complexes à exposer. C'est pourquoi nous écartons ce cas. Enfin, dans le cas où $\frac{M}{N} \to c \in (0, +\infty)$ (de sorte que B doit aussi être choisi du même ordre de grandeur que N et M), les entrées de $\hat{\mathbf{S}}_N(\nu)$ ne sont même pas consistantes. Une analyse basée sur la densité spectrale n'est plus recommendable, et il faut alors se fier à d'autres approches, telles que celles basées sur le comportement des valeurs propres de la matrice empirique de covariance issue de la théorie des matrices aléatoires. Pour conclure ce paragraphe, nous allons dans la suite de ce manuscrit nous appuyer sur le régime des grandes dimensions défini par $M = M(N), B = B(N)$ avec:

$$B = \mathcal{O}(N^{\alpha}) \text{ pour } \alpha \in \left(\frac{1}{2}, 1\right), \text{ tandis que } \frac{M}{B+1} := c_N \xrightarrow{N \to +\infty} c \in (0, +\infty) \tag{15}$$

qui sera abrégé par $\xrightarrow{N \to +\infty}$. Ce régime est pertinent pour étudier $\hat{\mathbf{S}}_N(\nu)$ défini dans (13) en utilisant des méthodes de matrices aléatoires en grande dimension, puisque l'on peut écrire $\hat{\mathbf{S}}_N(\nu) = \mathbf{\Sigma}_N(\nu)\mathbf{\Sigma}_N(\nu)^*$ où $\mathbf{\Sigma}_N(\nu)$ est la $M \times (B+1)$ matrice définie par

$$\mathbf{\Sigma}_N(\nu) = \frac{1}{\sqrt{B+1}} \left(\boldsymbol{\xi}_N \left(\nu - \frac{B}{2N}\right), \ldots, \boldsymbol{\xi}_N \left(\nu + \frac{B}{2N}\right) \right).$$

Cela représente la matrice de covariance empirique appliquée aux vecteurs de Fourier $\boldsymbol{\xi}_N(\nu - \frac{B}{2N}), \ldots, \boldsymbol{\xi}_N(\nu + \frac{B}{2N})$, et il est donc attendu que la théorie des matrices aléatoires fournisse des outils pour comprendre les valeurs propres de $\hat{\mathbf{S}}_N(\nu)$ sous notre régime des grandes dimensions.

Nous terminons ce paragraphe en précisant les deux quantités dérivées de $\hat{\mathbf{C}}_N(\nu)$ que nous étudierons dans ce manuscrit lorsque les composantes de $\mathbf{y}$ sont supposées indépendantes. Premièrement, si l'on note $\hat{\lambda}_1(\nu), \ldots, \hat{\lambda}_M(\nu)$ les valeurs propres de $\hat{\mathbf{C}}_N(\nu)$, on peut considérer la Statistique Linéaire Spectrale (LSS en anglais) défini par $\sum_{m=1}^{M} f(\hat{\lambda}_m(\nu))$ pour une fonction f définie sur $\mathbb{R}_+$ satisfaisant des conditions de régularité. Cette quantité est liée à la distribution empirique des valeurs propres de $\hat{\mathbf{C}}_N(\nu)$ puisque, en notant F_ν la distribution empirique des valeurs propres définie par:

$$\hat{F}_\nu(t) = \frac{1}{M}\text{card}\{\hat{\lambda}_m(\nu) : \hat{\lambda}_m(\nu) \leq t\}$$

On voit que la LSS correspond à tester la mesure de probabilité empirique contre la fonction f. De plus, cette quantité est importante pour tester l'indépendance des M séries temporelles, puisqu'elle peut être utilisée pour comparer la distribution des valeurs propres de $\hat{\mathbf{C}}_N(\nu)$ avec la distribution attendue sous l'hypothèse nulle $\mathcal{H}_0$, où les séries temporelles sont indépendantes. Toujours motivé par le test d'indépendance, le second objet d'intérêt sera les entrées hors-diagonales de $\hat{\mathbf{C}}_N(\nu)$, et plus précisément $\sup_\nu \sup_{i\neq j} |\hat{c}_{N,ij}(\nu)|^2$. Ce manuscrit compile les résultats obtenus sur le comportement de ces deux objets (LSS et maximum des entrées hors-diagonales de $\hat{\mathbf{C}}_N(\nu)$) sous différentes hypothèses de modèle pour les séries temporelles $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ dans le régime asymptotique des grandes dimensions décrit ci-dessus.

0.8 État de l'art

Dans cette section, nous présentons quelques résultats importants liés aux questions traitées dans ce manuscrit. Nous diviserons cette section en deux paragraphes: le premier présente les principaux sujets d'intérêt autour de la densité et cohérence spectrale de séries temporelles dans le régime des petites dimensions, tandis que le second présente quelques résultats importants en théorie des matrices aléatoires.

0.8.1 Autour de la matrice de densité/cohérence spectrale

L'estimation de la matrice de densité spectrale est un problème important en traitement du signal. La densité spectrale d'une série temporelle stationnaire est l'analogue en domaine fréquentiel de la covariance et est basée sur la transformée de Fourier de la fonction d'autocovariance. Elle agrège l'information au travers de différents lags et parmi les composantes d'une série temporelle multivariée. Comparée à la matrice de corrélation, elle offre une description plus riche d'une potentielle dépendance croisée. Alors que la plupart des résultats classiques sur l'analyse des séries temporelles utilisant des outils du domaine fréquentiel sont disponibles dans [13] et [14], nous en rappelons ici quelques-uns pour aider le lecteur à se familiariser avec les objets utilisés dans ce manuscrit.

Periodogramme. Ce manuscrit étant intéressé par le comportement du périodogramme moyenné, il peut être important de réunir d'abord quelques résultats concernant le périodogramme, qui est défini comme $\boldsymbol{\xi}_N(\nu)\boldsymbol{\xi}_N(\nu)^*$, ou simplement $|\xi_N(\nu)|^2$ dans le cas univarié. Le périodogramme est un outil fondamental en analyse spectrale. Dans le cas où $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ est une série temporelle complexe gaussienne, il est possible de montrer que la transformée de Fourier à des fréquences $\nu_1,\ldots,\nu_L$ écrite $\boldsymbol{\xi}_N(\nu_1),\ldots,\boldsymbol{\xi}_N(\nu_L)$ sont des variables aléatoires asymptotiquement indépendantes de loi $\mathcal{N}_\mathbb{C}(\mathbf{0},\mathbf{S}(\nu_l))$. De plus, dans le cas univarié, il est connu que $|\xi_N(\nu)|^2$ est un estimateur asymptotiquement non-biaisé de $s(\nu)$ mais pas consistant à cause de sa variance qui ne tend pas vers zéro. Par exemple, pour un bruit blanc gaussien i.i.d. $y_n \sim \mathcal{N}(0,\sigma^2)$, la variance de $|\xi_N(\nu)|^2$ est de l'ordre σ^4 [14, Proposition 10.3.2]. De plus, il est montré dans [30] que pour n'importe quelle suite de variables aléatoires indépendantes, gaussiennes, centrées, et de variance finie σ^2, les variables aléatoires $\{|\xi_N(\frac{k}{N})|^2, k\in\{0,\ldots,N-1\}\}$ sont indépendantes et ont une distribution exponentielle avec σ^2 comme valeur pour leur espérance commune. Il est également montré dans [1] que pour un processus stationnaire ergodique satisfaisant $\inf_{\nu\in[0,1]} s(\nu) > 0$ et des hypothèses supplémentaires sur les innovations, alors presque sûrement :

$$\lim_{N\to+\infty} \sup_{\nu\in[0,1]} \frac{|\xi_N(\nu)|^2}{s(\nu)\log N} = 1$$

Périodogramme moyenné. Bien que ces résultats sur le périodogramme $\xi_N(\nu)\xi_N(\nu)^*$ soient utiles, ils souffrent toutefois comme montré ci-dessus de la non-consistance de cet estimateur dans le régime où M est fixe et N tend vers l'infini. Pour atteindre la consistance, il est commun de recourir à une moyennisation des périodogrammes sur des fréquences proches. Un exemple d'estimateur dans cette classe plus générale d'estimateurs est appelé le périodogramme moyenné et est défini comme suit :

$$\hat{\mathbf{S}}_N(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_N\left(\nu + \frac{B}{N}\right)\boldsymbol{\xi}_N\left(\nu + \frac{B}{N}\right)^*$$

où B est un paramètre appelé fenêtre spectrale. Sous des hypothèses supplémentaires sur la fenêtre spectrale, certains résultats de distribution limite sont également disponibles pour le périodogramme moyenné, voir par exemple [13], [10], [66],[20],[57] et [97].

Avant de passer à la section suivante, nous précisons qu'il est possible de définir plus généralement un estimateur de la densité spectrale à l'aide d'une fonction de fenêtrage arbitraire, mais ce manuscrit ne traitera que de l'estimateur de la densité spectrale à l'aide d'une fonction de fenêtrage définie par (13) (aussi appelé l'estimateur de la densité spectrale moyennée). Cela permet de garder les calculs tractables et aussi simples que possible tout en fournissant des indices sur le comportement de cet estimateur en grande dimension. Nous pensons que considérer la classe plus générale d'estimateur ne changerait pas substantiellement notre compréhension des propriétés asymptotiques de l'estimateur de la cohérence spectrale en grande dimension (à des constantes près indépendantes de N et avec une vitesse de convergence inchangée), mais cela augmenterait considérablement la complexité des preuves et réduirait la lisibilité de nos résultats.

0.8.2 Autour des grandes matrices de covariance empirique

Bien que ce manuscrit a pour objectif de comprendre le comportement du spectre de la matrice de cohérence spectrale d'un processus stationnaire en grande dimension, il a été observé que le périodogramme moyenné peut être vu comme la matrice de covariance empirique des vecteurs de Fourier associés aux observations $(\mathbf{y}_n)_{n=1,\ldots,N}$, il est donc intéressant de revoir certains résultats connus de la théorie des matrices aléatoires sur la matrice de covariance empirique.

Bien que le comportement des valeurs propres de la matrice de covariance empirique en petite dimension est bien connu depuis les travaux fondateurs de [2], le grande dimension nécessite une plus grande attention. L'objectif de cette section n'est bien entendu pas de fournir une vue exhaustive des résultats développés concernant le comportement des valeurs propres d'une matrice de covariance empirique de grande dimension, mais plutôt d'introduire le lecteur aux idées principales utilisées plus tard dans ce manuscrit. Cette section couvrira plusieurs sujets, allant du modèle blanc jusqu'à plusieurs extensions incorporant une dépendance au sein des lignes (qui sera parfois appelée *dépendance temporelle*) ou au sein des colonnes (qui sera parfois appelée *dépendance spatiale*).

Distribution de Marcenko-Pastur. Pour la suite de cette section, nous considérons la matrice de covariance empirique $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$, où $\mathbf{Y}_N$ est une matrice de taille $M \times N$, et le régime considéré sera $M, N \to +\infty$ tel que $\frac{M}{N} \to c \in (0, +\infty)$. Nous soulignons le fait que cette notation n'est pas celle que nous utiliserons dans ce manuscrit (nous utiliserons des matrices de taille $M \times (B+1)$ tel que $\frac{M}{B+1} \to c \in (0, +\infty)$), mais est une notation utilisée classiquement dans la littérature sur les matrices aléatoires. Le travail de [63] est un des premiers qui a réalisé que l'estimation des propriétés spectrales de matrices de covariance empirique de grande dimension pose de nouveaux problèmes quand $\frac{M}{N}$ n'est pas petit. Ils ont montré que dans le cas d'entrées i.i.d. avec variance 1, les valeurs propres de la matrice de covariance empirique $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ ne se concentrent pas autour de 1 (l'unique valeur propre population), mais se répartissent sur l'intervalle $[(1-\sqrt{\frac{M}{N}})^2, (1+\sqrt{\frac{M}{N}})^2]$ quand $M \leq N$

(sinon, la matrice $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ est dégénérée et il y a presque sûrement exactement $M-N$ valeurs propres nulles). Ceci implique que quand le rapport $\frac{M}{N}$ n'est pas petit, la matrice de covariance empirique n'est pas un bon estimateur de la matrice de covariance. Néanmoins, ils ont également montré que la distribution empirique des valeurs propres de la matrice de covariance empirique est asymptotiquement non-aléatoire, ce qui permet d'une part la construction de nouveaux estimateurs améliorés, d'autre part une meilleure compréhension des quantités spectrales liées à la matrice de covariance estimée, et à plus généralement fondé les bases de la théorie des matrices aléatoires appliquée à l'analyse statistique des données. Plus précisément, considérons $\mathbf{Y}_N$ une matrice de taille $M \times N$ avec des entrées i.i.d. $\mathcal{N}(0,1)$ et considérons la mesure aléatoire $\hat{\mu}_N$ définie par

$$\mathrm{d}\hat{\mu}_N = \frac{1}{M}\sum_{m=1}^{M} \delta_{\hat{\lambda}_{m,N}}$$

où $\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{M,N}$ sont les M valeurs propres de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ et δ_x la mesure de Dirac au point x. Il est montré dans [63] l'existence d'une mesure non-aléatoire $\mu_{mp,c}$, dépendante de $c := \lim \frac{M}{N} \in (0,+\infty)$, telle que presque sûrement

$$\hat{\mu}_N - \mu_{mp,c} \xrightarrow[M\to+\infty]{w} 0$$

où la convergence est la convergence faible des mesures. La mesure $\mu_{mp,c}$ est définie par:

$$\mathrm{d}\mu_{mp,c}(\lambda) = \left(1-\frac{1}{c}\right)_+ \delta_0(\lambda) + \frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{2c\pi\lambda}\mathbb{1}_{[\lambda_-,\lambda_+]}(\lambda) \tag{16}$$

où $(\cdot)_+ = \max(\cdot, 0)$ et $\lambda_\pm = (1\pm\sqrt{c})^2$. Quand $c > 1$, les $M-N$ valeurs propres nulles sont représentées par le terme $\left(1-\frac{1}{c}\right)_+\delta_0(\lambda)$. Quand $c < 1$, la distribution de Marcenko-Pastur est supportée sur l'intervalle $[\lambda_-,\lambda_+]$, qui est appelé le bulk de la distribution. Leur méthode de preuve repose sur la transformée de Stieltjes de la distribution empirique des valeurs propres, définie comme:

$$\hat{m}_N(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\hat{\mu}_N(\lambda)}{\lambda - z}, \quad \text{pour tout } z \in \mathbb{C}\setminus\mathbb{R}_+ \tag{17}$$

ou de manière équivalente par la résolvante $(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* - z\mathbf{I}_M)^{-1}$ puisque $\hat{m}_N(z)$ peut aussi s'exprimer comme suit:

$$\hat{m}_N(z) = \frac{1}{M}\mathrm{tr}\,\left(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* - z\mathbf{I}_M\right)^{-1}.$$

Les auteurs de [63] ont montré que la fonction $\hat{m}_N$ satisfait asymptotiquement l'équation $zcX^2 + (z+c-1)X + 1 = 0$, qui admet pour solution $z \mapsto t(z)$ dans l'ensemble des transformées de Stieltjes égale à:

$$t(z) = \frac{-(z+c-1)+\sqrt{(z-\lambda_+)(z-\lambda_-)}}{2zc} \tag{18}$$

où pour $z = \rho e^{i\theta}$, $\sqrt{z} = \sqrt{\rho}e^{i\theta/2}$ pour $\theta \in (0, 2\pi)$. Étant donné une suite $(c_N)_{N\geq 1}$, nous noterons

$$t_N(z) = \frac{-(z+c_N-1)+\sqrt{(z-\lambda_+)(z-\lambda_-)}}{2zc_N}$$

où il est entendu que $\lambda_\pm = (1\pm\sqrt{c_N})^2$. On va maintenant introduire la transformée de Stieltjes de $\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N$:

$$\underline{m}_N(z) = \frac{1}{N}\mathrm{tr}\,\left(\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N - z\mathbf{I}_N\right)^{-1}$$

$((\frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N - z\mathbf{I}_N)^{-1}$ est parfois appelée la co-résolvante) et on peut montrer de la même manière que $\underline{\hat{m}}_N$ converge asymptotiquement vers $\tilde{t}$, satisfaisant:

$$\tilde{t}(z) = ct(z) - \frac{1-c}{z}.$$

Enfin, $t(z)$ et $\tilde{t}(z)$ peuvent être exprimés comme l'unique solution dans le domaine des transformées de Stieltjes portées par $\mathbb{R}_+$ des équations couplées suivantes :

$$t(z) = \frac{-1}{z(1+\tilde{t}(z))}, \quad \tilde{t}(z) = \frac{-1}{z(1+ct(z))}.$$

De la même manière, pour une suite $(c_N)_{N\geq 1}$, nous noterons :

$$t_N(z) = \frac{-1}{z(1+\tilde{t}_N(z))}, \quad \tilde{t}_N(z) = \frac{-1}{z(1+c_N t_N(z))}. \tag{19}$$

Il est aussi intéressant de considérer la distribution spectrale limite de la matrice de corrélation empirique définie par $\mathrm{dg}(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*)^{-\frac{1}{2}}\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*\mathrm{dg}(\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*)^{-\frac{1}{2}}$. Cette distribution a été étudiée pour la première fois par [46] sous des conditions de second moment fini, où l'auteur a montré que, dans le cas blanc, la distribution spectrale limite de la matrice de corrélation empirique converge faiblement vers la distribution de Marcenko-Pastur.

Extensions avec dépendence spatiale. Bien que la distribution spectrale limite dans le cas blanc (indépendance entre les lignes et les colonnes de $\mathbf{Y}_N$) ait été étudiée en détail dans [85],[84] et [4], il y a eu une grande quantité de travaux écrit afin de relaxer la structure d'indépendance entre les entrées Y_{ij}, et comprendre comment les dépendances entre les entrées de $\mathbf{Y}_N$ ont un impact sur la distribution spectrale limite de la matrice de covariance empirique. Nous commençons avec le modèle où les échantillons $\mathbf{y}_1, \dots, \mathbf{y}_N$ sont indépendants, mais les composantes de $\mathbf{y}_i$ ne sont pas indépendantes. Les auteurs dans [85] ont considéré le cas où $\mathbf{Y}_N = \mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N$, et $\mathbf{X}_N$ est une matrice de taille $M \times N$ avec des valeurs i.i.d. de deuxième moment fini tandis que $\mathbf{\Sigma}_N$ est une matrice définie positive indépendante de $\mathbf{X}_N$. Le modèle de Marcenko-Pastur correspond au cas où $\mathbf{\Sigma}_N = \mathbf{I}_M$. Dans le cas contraire, où $\mathbf{\Sigma}_N \neq \mathbf{I}_M$, $\mathbf{Y}_N$ contient de la corrélation spatiale. Sous l'hypothèse que la distribution spectrale des matrices $(\mathbf{\Sigma}_N)_{N\geq 0}$ converge presque sûrement faiblement vers une densité de probabilité H, alors avec une probabilité égale à 1, étant donné que $N \to +\infty$ tandis que $\frac{M}{N} \to c$, la distribution spectrale empirique $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ converge presque sûrement faiblement vers une densité de probabilité non-aléatoire. Cette distribution spectrale limite est donnée implicitement par une équation dont la transformée de Stieltjes $z \mapsto m(z)$ est l'unique solution dans l'ensemble $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}_+\}$ de:

$$m(z) = \int_{\mathbb{R}} \frac{\mathrm{d}H(t)}{t(1-c-czm(z)) - z}$$

On note que dans le cas du bruit blanc, $H = \delta_1$, on retrouve l'équation définissant la distribution de Marcenko-Pastur. De plus, [86] déduit des propriétés analytiques concernant la distribution limite des valeurs propres à partir de l'étude de cette équation.

Un autre modèle important pour les applications et, plus particulièrement, pour le Chapitre 3 et le Chapitre 4, est le modèle connu sous le nom de modèle spike. Il correspond au cas où $\mathbf{\Sigma}_N$ est une perturbation de rang fini K de l'identité :

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \mathbf{I}_M, \quad \text{avec } \mathbf{B}_N \text{ positif et } \mathrm{rank}(\mathbf{B}_N) = K$$

de sorte que les $M - K$ premières valeurs propres de $\mathbf{\Sigma}_N$ sont égales à 1. Il est alors intéressant d'étudier les valeurs propres de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* = \frac{1}{N}\mathbf{\Sigma}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^*\mathbf{\Sigma}_N^{\frac{1}{2}}$ où $\mathbf{X}_N$ est une matrice de taille $M \times N$ à éléments i.i.d. de moyenne nulle et variance un. Ce modèle est aussi appelé modèle multiplicatif spike et a été étudié dans [6] et [7]. En notant $\hat{\mu}_N$ la distribution empirique des valeurs propres de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$, il est montré que presque sûrement,

$$\hat{\mu}_N \xrightarrow[N\to+\infty]{w} \mu_{mp,c}$$

Toutefois, les valeurs propres de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ ne se concentrent pas toutes autour de l'intervalle $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$. Il est montré dans [7] que les plus grandes valeurs propres $\hat{\lambda}_{1,N}, \ldots, \hat{\lambda}_{K,N}$ de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ peuvent converger vers des valeurs hors de l'intervalle $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ si les K valeurs propres positives de $\mathbf{B}_N\mathbf{B}_N^*$ convergent vers des valeurs suffisamment grandes $\lambda_1, \ldots, \lambda_K$. Plus précisément, pour tout $k = 1, \ldots, K$,

$$\hat{\lambda}_{k,N} \xrightarrow[N\to+\infty]{p.s.} \begin{cases} \frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k} & \text{si } \lambda_k > 1+\sqrt{c} \\ (1+\sqrt{c})^2 & \text{sinon.} \end{cases}$$

On note que $\lambda_k > 1+\sqrt{c}$ implique que $\frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k} > (1+\sqrt{c})^2$, ce qui signifie que, asymptotiquement, les valeurs propres les plus grandes vont effectivement quitter l'intervalle $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$, et créer un spike localisé autour de la valeur $\frac{(\lambda_k+1)(\lambda_k+c)}{\lambda_k}$.

Extensions avec dépendance temporelle. Nous étudions maintenant le cas où les échantillons $\mathbf{y}_1, \ldots, \mathbf{y}_N$ sont corrélés temporellement plutôt que corrélés spatialement. Il existe plusieurs manières de considérer un tel modèle. Par exemple, on pourrait considérer $\mathbf{Y}_N = \mathbf{X}_N\mathbf{\Sigma}_N^{1/2}$, (au lieu de $\mathbf{Y}_N = \mathbf{\Sigma}_N^{1/2}\mathbf{X}_N$ dans le paragraphe précédent) et étudier la matrice de covariance empirique $\hat{\mathbf{\Sigma}}_N = \frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^* = \frac{1}{N}\mathbf{X}_N\mathbf{\Sigma}_N\mathbf{X}_N^*$. Cependant, il apparaît que ce modèle n'est pas tellement différent que celui développé dans [85]: il suffit de considérer $\underline{\hat{\mathbf{\Sigma}}_N} = \frac{1}{M}\mathbf{Y}_N^*\mathbf{Y}_N = \frac{1}{M}\mathbf{\Sigma}_N^{1/2}\mathbf{X}_N\mathbf{X}_N^*\mathbf{\Sigma}_N^{1/2}$, dont la distribution limite des valeurs propres est connue. Il reste à observer que $\hat{\mathbf{\Sigma}}_N$ et $\underline{\hat{\mathbf{\Sigma}}_N}$ ont exactement les mêmes valeurs propres, à l'exception de $N-M$ zéros si $N > M$ ($M-N$ zéros sinon), pour décrire complètement les valeurs propres de ce modèle.

Un autre modèle qui aura une importance cruciale dans le Chapitre 2 est le suivant : supposons qu'il existe des matrices positives Hermitiennes $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_M$ de taille $N \times N$ telles que

$$\begin{pmatrix} y_{m,1} \\ \vdots \\ y_{m,N} \end{pmatrix} = \mathbf{\Theta}_m^{\frac{1}{2}} \begin{pmatrix} x_{m,1} \\ \vdots \\ x_{m,N} \end{pmatrix}$$

où $(x_{m,1}, \ldots, x_{m,N}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_N)$ et $(x_{m_1,1}, \ldots, x_{m_1,N})$ est indépendant de $(x_{m_2,1}, \ldots, x_{m_2,N})$ pour $m_1 \neq m_2$. Ce modèle, étudié dans [92] et [48], est utile pour représenter des observations M-dimensionnelles, où les dimensions sont indépendantes, mais chaque dimension a sa propre structure de covariance décrite par $\mathbf{\Theta}_m$. Ce modèle sera traité en détail dans le Chapitre 2.

Pour finir, un modèle lié à ceux développés dans ce manuscrit est de supposer que chaque dimension est générée comme une série temporelle. Ce modèle est d'importance puisque l'un des problèmes importants dans les applications est de tester si une collection de séries temporelles sont corrélées ou non. Dans ce cas, la distribution de Marcenko-Pastur n'est plus une bonne approximation de la distribution des valeurs propres de la matrice de covariance empirique, et d'autre part, ce modèle ne peut être exprimé sous la forme du modèle développé dans [85]. Pour prendre un exemple, [98] considère que chaque dimension de $\mathbf{y}$ est générée par un processus linéaire scalaire indépendant des autres dimensions :

$$y_{m,n} = \sum_{k \geq 0} a_k \epsilon_{m,n-k}$$

où pour chaque m, $(\epsilon_{m,n})_{n\in\mathbb{Z}}$ est une série de valeurs réelles et stationnaires à moyenne nulle et de variance 1. Sous l'hypothèse que le filtre est sommable ($\sum_{k\geq 0} |a_k| < +\infty$) et une condition de moment fini sur les innovations ($\mathbb{E}\epsilon_n^4 < +\infty$), il est montré que la distribution limite des valeurs propres de $\frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^*$ converge vers une distribution de probabilité déterministe F dont la transformée de Stieltjes $z \mapsto m(z)$ satisfait :

$$z = \frac{-1}{m(z)} + \int_0^1 \frac{1}{cm(z) + cs^{-1}(\nu)} \, \mathrm{d}\nu$$

où s est la densité spectrale du processus linéaire $(y_{m,n})_{n\geq 0}$ (toutes les dimensions $m = 1, \ldots, M$ ont la même densité spectrale par hypothèse) égales à :

$$s(\nu) = \left| \sum_{k\geq 0} a_k e^{2\mathrm{i}\pi k\nu} \right|^2 .$$

L'hypothèse que chaque série temporelle est générée par le même processus est assez forte, et dans ce manuscrit, nous ne supposerons pas cela, ce qui veut dire que nous supposerons qu'il existe M filtres $(a_{1,k})_{k\geq 0}, \ldots, (a_{M,k})_{k\geq 0}$ (avec des hypothèses supplémentaires qui seront précisées dans les chapitres suivants) tels que

$$Y_{m,n} = \sum_{k\geq 0} a_{m,k} \epsilon_{m,n-k}$$

où les innovations sont comme précedement du bruit blanc.

Extensions avec de la dépendance temporelle et spatiale. Naturellement, on peut se demander s'il est possible de traiter simultanément la dépendance sur les lignes et les colonnes, puisque des échantillons avec dépendance spatio-temporelles peuvent être l'objet de nombreuses situations pratiques et sont d'intérêt dans le domaine du traitement du signal par exemple. Par exemple, il est intéressant de considérer un modèle de covariance plus général de la forme $\frac{1}{N}\mathbf{T}_{2N}^{\frac{1}{2}}\mathbf{X}_N^*\mathbf{T}_{1N}\mathbf{X}_N\mathbf{T}_{2N}^{\frac{1}{2}}$, où $\mathbf{T}_{1N}$ et $\mathbf{T}_{2N}$ sont des suites de matrices positives, et les $\mathbf{T}_{1N}$ sont hermitiennes, dont les taille coïncident avec celles des $\mathbf{X}_N \in \mathbb{C}^{M\times N}$. Dans le cas où $\mathbf{T}_{1,N}$ est semi-définie positive, ce modèle est connu sous le nom de modèle à covariance séparable, puisque la covariance des données s'écrit comme le produit de Kronecker des matrices $\mathbf{T}_{1,N}$ et $\mathbf{T}_{2,N}$. Parmi les travaux qui ont étudié ce modèle, on peut citer [71], [49], [18], [37] et les références citées dans ces papiers.

0.8.3 Travaux précédents utilisant le régime asymptotique (15)

Relativement peu de travaux ont étudié le comportement du périodogramme moyenné $\hat{\mathbf{S}}_N(\nu)$ dans le régime défini par (15). Par exemple, il a été observé dans [11] que, quand la dimension de la série temporelle diverge, le risque d'estimation du périodogramme moyenné diverge également. Cette observation a conduit à de nouvelles directions de recherche pour améliorer les performances d'estimation et de test, en utilisant des méthodes reposant sur le shrinkage, le thresholding (voir par exemple [87]), et la théorie des matrices aléatoires. Les auteurs de [11], inspirés par le phénomène bien connu de répartition des valeurs propres de la matrice de covariance empirique de grande dimension autour des valeurs propres population, ont défini un estimateur de shrinkage par:

$$\hat{\mathbf{S}}_N^{(shrink)}(\nu) = p(\nu)\hat{\mathbf{S}}_N(\nu) + r(\nu)\mathbf{I}_M$$

pour des fonctions explicites p et r contrôlant le niveau de shrinkage pour chaque fréquence, et ont démontré que cette méthode peut réduire le risque associé à l'estimation de la matrice de densité spectrale. On note que le régime asymptotique qu'ils ont considéré est similaire à celui utilisé dans ce manuscrit, puisqu'ils utilisent :

$$0 < \inf_{N \geq 1} \frac{M}{B} \leq \sup_{N \geq 1} \frac{M}{B} < +\infty \text{ et } \frac{\sqrt{M}B}{N} \to 0$$

ce qui est équivalent à $M = o(N^{2/3})$ (et on rappelle que dans ce manuscrit, on suppose $M = \mathcal{O}(N^{\alpha})$ pour $\frac{1}{2} < \alpha < 1$). Cette approche a été améliorée plus tard dans [29] et les références citées dedans.

Une autre approche liée aux séries temporelles de grande dimension est présente dans [61] et [60]. Les auteurs considèrent L observations consécutives de la m-ème série temporelle à partir du temps n :

$$\mathbf{y}_{m,n}^{L} = [y_{m,n}, \ldots, y_{m,n+L-1}]^{T}$$

et à partir de ceci construisent un vecteur ML-dimensionnel

$$\mathbf{y}_{n}^{L} = [(\mathbf{y}_{1,n}^{L})^{T}, \ldots, (\mathbf{y}_{M,n}^{L})^{T}].$$

Ils désignent par $\mathcal{R}_L$ la matrice de covariance $ML \times ML$ de ce vecteur aléatoire, c'est-à-dire $\mathcal{R}_L = \mathbb{E}[\mathbf{y}_n^L(\mathbf{y}_n^L)^*]$ où $(\cdot)^*$ signifie le transposé conjugué. Cette matrice est parfois appelée matrice de covariance spatio-temporelle. Il est clair que les M séries $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$ sont indépendantes si et seulement si pour chaque entier L, la matrice $\mathcal{R}_L$ est diagonale, c'est-à-dire

$$\mathcal{R}_L = \mathrm{Bdiag}(\mathcal{R}_L)$$

où, pour une matrice $ML \times ML$ $\mathbf{A}$, $\mathrm{Bdiag}(\mathbf{A})$ est la matrice diagonale de la même dimension dont les $L \times L$ blocs sont ceux de $\mathbf{A}$. On remarque que les $L \times L$ blocs diagonaux de $\mathrm{Bdiag}(\mathcal{R}_L)$ sont les $L \times L$ matrices Toeplitz $\mathbf{R}_{m,L}$, $m = 1, \ldots, M$, définies par

$$(\mathbf{R}_{m,L})_{kk'} = r_m(k - k')$$

où $(r_m(k))_{k\in\mathbb{Z}}$ est la séquence de covariance de la m–ème série temporelle. [61] et [60] définissent par $\mathcal{R}_{corr,L}$ la matrice de corrélation bloc-diagonale par:

$$\mathcal{R}_{corr,L} = \mathrm{Bdiag}(\mathcal{R}_L)^{-\frac{1}{2}} \mathcal{R}_L \mathrm{Bdiag}(\mathcal{R}_L)^{-\frac{1}{2}}.$$

De cette manière, $\mathcal{R}_L$ est bloc-diagonale pour tout L si et seulement si $\mathcal{R}_{corr,L} = \mathbf{I}_{ML}$ pour tout L. Une manière de tester si les M composantes de la série temporelle $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ sont décorrélées consiste à estimer $\mathcal{R}_{corr,L}$ pour des valeurs de L adéquates, puis comparer ces estimées avec $\mathbf{I}_{ML}$. Pour faire cela, les auteurs considèrent l'estimateur classique $\hat{\mathcal{R}}_{corr,L}$ défini par:

$$\hat{\mathcal{R}}_{corr,L} = \mathrm{Bdiag}(\hat{\mathcal{R}}_L)^{-\frac{1}{2}} \hat{\mathcal{R}}_L \mathrm{Bdiag}(\hat{\mathcal{R}}_L)^{-\frac{1}{2}}$$

où $\hat{\mathcal{R}}_L$ est la matrice de covariance spatio-temporelle empirique définie par

$$\hat{\mathcal{R}}_L = \frac{1}{ML} \sum_{n=1}^{N} \mathbf{y}_n^L (\mathbf{y}_n^L)^*$$

et où $\mathrm{Bdiag}(\hat{\mathcal{R}}_L)$ est la matrice bloc-diagonale correspondante:

$$\mathrm{Bdiag}(\hat{\mathcal{R}}_L) = \begin{pmatrix} \hat{\mathbf{R}}_{1,L} & & \\ & \ddots & \\ & & \hat{\mathbf{R}}_{M,L} \end{pmatrix}$$

avec $\hat{\mathbf{R}}_{m,L}$ pour $m = 1, \ldots, M$, représentant les $L \times L$ blocs-diagonaux correspondants. Sous l'hypothèse d'indépendance des M composants de la série temporelle et sous le régime asymptotique où $\frac{ML}{N} \xrightarrow{N \to +\infty} c$, ils montrent que

$$\int f(\lambda)\, \mathrm{d}\hat{\mu}_N(\lambda) - \int f(\lambda)\, \mathrm{d}\mu_{mp,c} \to 0$$

où $\mathrm{d}\hat{\mu}_N = \frac{1}{ML} \sum_{k=1}^{N} \delta_{\hat{\lambda}_k}$, et $(\hat{\lambda}_k)_{k=1,\ldots,ML}$ sont les valeurs propres de $\hat{\mathcal{R}}_L$.

0.9 Contributions de cette thèse

Le sujet général de cette thèse est de comprendre certaines propriétés de quantités issues du périodogramme moyenné de séries temporelles M–dimensionnelles $\mathbf{y}$ quand M est grand et les composantes de $\mathbf{y}$ sont indépendantes. Dans tous les chapitres suivant, nous utiliserons le régime des grandes dimensions défini plus haut dans (15) que nous rappelons ici : M et B sont des fonctions de N à valeur entières telles que $\frac{M}{B} \xrightarrow{N \to +\infty} c \in (0, +\infty)$ et $M = \mathcal{O}(N^\alpha)$ pour $\frac{1}{2} < \alpha < 1$. Dans cette section, nous omettrons parfois la dépendance explicite en la fréquence ν et en le complexe z afin de simplifier les notations.

0.9.1 Contribution du Chapitre 1

La motivation de ce chapitre est d'évaluer le comportement de certaines statistiques linéaire des valeurs propres (LSS) de $\hat{\mathbf{C}}_N(\nu)$ définies en (14) sous l'hypothèse $\mathcal{H}_0$ que les M composantes $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ de $\mathbf{y}$ sont mutuellement décorrélées et dans notre régime asymptotique des grandes dimensions. Soit la mesure spectrale empirique de $\hat{\mathbf{C}}_N(\nu)$ définie par

$$\mathrm{d}\hat{\mu}_{N,\nu} = \frac{1}{M} \sum_{m=1}^{M} \delta_{\lambda_m(\hat{\mathbf{C}}_N(\nu))}$$

où $(\lambda_m(\mathbf{A}))_{m=1,\ldots,M}$ représentent les M valeurs propres de la matrice $\mathbf{A}$ de taille $M \times M$, et δ_x est la mesure de Dirac au point x. Une LSS de $\hat{\mathbf{C}}_N(\nu)$ peut alors s'écrire $\int_{\mathbb{R}} f(\lambda)\, \mathrm{d}\hat{\mu}_N(\nu, \lambda)$ ou de manière équivalente $\frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$ pour une certaine fonction f definie sur $\mathbb{R}_+$ satisfaisant certaines propriétés de régularité si nécessaire. Dans ce chapitre, nous étudions la quantité

$$\sup_{\nu\in[0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f\, \mathrm{d}\mu_{mp,c_N} \right|$$

et montrons que la mesure empirique des valeurs propres de $\hat{\mathbf{C}}_N(\nu)$ converge vers la distribution de Marcenko-Pastur. Afin d'étudier cette quantité, nous travaillons d'abord à ν fixé, et étendons ensuite notre résultat uniformément sur $\nu \in [0, 1]$ en utilisant un argument de Lipschitzité. Pour chaque ν fixé, nous considérons l'intermédiaire de calcul $\tilde{\mathbf{C}}_N(\nu)$ défini par

$$\tilde{\mathbf{C}}_N(\nu) = \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} \hat{\mathbf{S}}_N(\nu) \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}}$$

et montrons que $\tilde{\mathbf{C}}_N(\nu)$ peut s'écrire sous la forme

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu)$$

où $\mathbf{X}_N(\nu)$ est une matrice $M \times (B+1)$ à entrées i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ et $\tilde{\mathbf{\Delta}}_N(\nu)$ est un terme d'erreur convergeant vers 0 en norme spectrale à la vitesse $\frac{B}{N}$ (voir plus bas pour plus de détails sur la matrice $\tilde{\mathbf{\Delta}}_N(\nu)$). Nous énonçons maintenant la décomposition utilisées pour étudier la quantité $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$:

$$\begin{aligned}\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) = &\left(\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))\right) + \left(\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu))\right) \\ &+ \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right)\right) \\ &+ \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right) - \int f\,\mathrm{d}\mu_{mp,c_N}\right) + \int f\,\mathrm{d}\mu_{mp,c_N}\end{aligned}$$

Concernant la première étape de cette décomposition, nous utilisons la formule de Helffer-Sjöstrand, qui permet de transférer l'étude des LSS vers celle des résolvantes $\tilde{\mathbf{Q}}_N(z) := (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$ et $\hat{\mathbf{Q}}_N(z) := (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$. Sous hypothèses sur f (régularité et support compact), la formule de Helffer-Sjöstrand fournit que

$$\int f\,\mathrm{d}\mu = \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(x,y)s_\mu(x+iy)\,\mathrm{d}x\,\mathrm{d}y$$

où s_μ est la transformée de Stieltjes de μ, $\bar{\partial} = \partial_x + i\partial_y$ et $\Phi_k(f) : \mathbb{C} \to \mathbb{C}$ est la fonction définie sur $\mathbb{C}$ par

$$\Phi_k(f)(x,y) = \sum_{l=0}^{k} \frac{(iy)^l}{l!} f^{(l)}(x)\rho(y)$$

où $\rho : \mathbb{R} \to \mathbb{R}^+$ est smooth, à support compact, de valeur 1 dans un voisinage de 0. La fonction $\Phi_k(f)$ coincide avec f sur l'axe réel et l'étend sur le plan complexe. Dans notre contexte, nous pouvons écrire que

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) = \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(x,y)\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{Q}}_N(x+iy) - \tilde{\mathbf{Q}}_N(x+iy))\,\mathrm{d}x\,\mathrm{d}y.$$

Il est ensuite possible d'étudier la différence

$$\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} = \frac{1}{M}\mathrm{tr}\ \left(\hat{\mathbf{Q}}_N(z)(\tilde{\mathbf{C}}_N - \hat{\mathbf{C}}_N)\tilde{\mathbf{Q}}_N(z)\right)$$

Pour énoncer notre résultat sur le comportement de $\frac{1}{M}\mathrm{tr}\ (\hat{\mathbf{C}}_N - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{C}}_N - z\mathbf{I}_M)^{-1}$, nous avons préalablement besoin de définir le concept de *domination stochastique*, où l'on dit que qu'une famille de variable aléatoire positive $X = (X_N)_{N\geq 1}$ est stochastiquement dominée par une suite $(u_N)_{N\geq 1}$ si pour tout $\epsilon > 0$, il existe $N_0(\epsilon)$ suffisamment grand et un certain $\gamma(\epsilon) > 0$ tel que pour tout $N > N_0(\epsilon)$:

$$\mathbb{P}\left[X_N > N^\epsilon u_N\right] \leq \exp -N^\gamma$$

Si $(X_N)_{N\geq 1}$ est stochastiquement dominé par $(u_N)_{N\geq 1}$, nous utilisons la notation $X_N \prec u_N$, et si pour tout famille de variable aléatoire complexe X nous avons $|X| \prec u_N$, nous écrirons aussi $X = \mathcal{O}_\prec(u_N)$. En utilisant cette notation, nous montrons que le résultat suivant est valide uniformément en $\nu \in [0,1]$:

$$\frac{1}{M}\text{tr } f(\hat{\mathbf{C}}_N(\nu)) - \frac{1}{M}\text{tr } f(\tilde{\mathbf{C}}_N(\nu)) = -\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s_m''(\nu)}{2s_m(\nu)}\right) v_N < \tilde{\mathcal{D}}_N, f > + \mathcal{O}_{\prec}\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right)$$

où $v_N = \left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{b}{N}\right)^2\right)$ et $< \tilde{\mathcal{D}}_N, f >$ correspond à l'action de la distribution $\tilde{\mathcal{D}}_N$, dont la transformée de Stieltjes est explicitement connue, sur f (cette quantité sera précisée plus loin). Cela nous permet de nous concentrer sur l'étude du comportement de $\tilde{\mathbf{C}}_N(\nu)$ au lieu de celui de $\hat{\mathbf{C}}_N(\nu)$. Nous montrons ensuite, via des arguments de concentration gaussienne, que

$$\frac{1}{M}\text{tr } f(\tilde{\mathbf{C}}_N(\nu)) - \mathbb{E}\frac{1}{M}\text{tr } f(\tilde{\mathbf{C}}_N(\nu)) = \mathcal{O}_{\prec}\left(\frac{1}{B}\right)$$

afin que de cette manière il ne reste qu'à étudier $\mathbb{E}\frac{1}{M}\text{tr } f(\tilde{\mathbf{C}}_N(\nu))$. Sous l'hypothèse d'indépendance des M composantes de la série temporelle, nous montrons que $\tilde{\mathbf{C}}_N(\nu)$ peut s'approximer pour chaque ν par une matrice de Wishart de bruit blanc:

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu)$$

où $\mathbf{X}_N(\nu)$ est une matrice de taille $M \times (B+1)$ à éléments i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, $\sup_{\nu \in \mathcal{V}_N} \|\tilde{\mathbf{\Delta}}_N(\nu)\| = \mathcal{O}_{\prec}(\frac{B}{N})$ et $\mathcal{V}_N = \{\frac{k}{N}, k = 0, \ldots, N-1\}$. Après quelques étapes de calcul gaussien (en particulier la formule d'intégration par partie gaussienne), il est possible d'obtenir l'estimation suivante:

$$\mathbb{E}\left[\frac{1}{M}\text{tr } f(\tilde{\mathbf{C}}_N(\nu))\right] - \mathbb{E}\left[\frac{1}{M}\text{tr } f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right)\right] = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s_m'(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$
$$- \frac{1}{M}\sum_{m=1}^{M}\frac{s_m''(\nu)}{2s_m(\nu)} < \tilde{\mathcal{D}}_N, f > v_N + \mathcal{O}\left(\frac{1}{N} + \left(\frac{B}{N}\right)^3\right)$$

où $\mathcal{D}_N$ est une autre distribution dont la transformée de Stieltjes est aussi connue explicitement. La stratégie pour obtenir ce résultat consiste à considérer $\mathbb{E}\left[\frac{1}{M}\text{tr } \tilde{\mathbf{Q}}_N(z)\right] - \mathbb{E}\left[\frac{1}{M}\text{tr } \mathbf{Q}_N(z)\right]$, où $\mathbf{Q}_N(z) = (\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} - z\mathbf{I}_M)^{-1}$ est la résolvante associée à $\mathbf{X}_N\mathbf{X}_N^*$. Après de long et fastidieux calcul gaussiens, on montre que

$$\mathbb{E}\left[\frac{1}{M}\text{Tr}(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z))\right] = -\left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s_m''(\nu)}{s_m(\nu)}\right) \tilde{p}_N(z)\, v_N +$$
$$\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s_m'(\nu)}{s_m(\nu)}\right)^2 p_N(z)\, v_N + \mathcal{O}_z\left(\frac{B}{N}\right)^3$$

où $\tilde{p}_N$ est la transformée de Stieltjes de la distribution à support compact $\tilde{D}_N$ introduite précedemment, et p_N est aussi la transformée de Stieltjes d'une autre distribution à support compact D_N. Leur expression exacte est connue:

$$\tilde{p}_N(z) = (zt_N(z))', \quad p_N(z) = \frac{-c_N(zt_N(z)\tilde{t}_N(z))^3}{1 - c_N(zt_N(z)\tilde{t}_N(z))^2}$$

où nous rappelons que t_N et $\tilde{t}_N$ ont étées définies dans (19). Il reste alors à énnoncer un résultat bien connu en matrice aléatoire qui assure que pour une matrice de bruit blanc gaussien complexe

$\mathbf{X}_N(\nu)$,

$$\mathbb{E}\left[\frac{1}{M}\text{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] - \int f\ \mathrm{d}\mu_{mp,c_N} = \mathcal{O}\left(\frac{1}{B^2}\right)$$

où μ_{mp,c_N} est la distribution de Marcenko-Pastur de paramètre $c_N = \frac{M}{B+1}$. En rassemblant toutes les estimations précédentes, nous obtenons que pour chaque ν,

$$T_N(\nu) = \mathcal{O}_{\prec}\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right).$$

où

$$T_N(\nu) = \frac{1}{M}\text{tr}\ \hat{\mathbf{C}}_N(\nu) - \int f\ \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$

De manière informelle, ce résultat indique que sous l'hypothèse d'indépendance des séries temporelles, la distribution empirique des valeurs propres de $\hat{\mathbf{C}}_N(\nu)$ est proche de la distribution de Marcenko-Pastur, à une correction près d'ordre $\mathcal{O}(\frac{B}{N})^2$ et à une erreur près d'ordre $\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$. Par une borne d'union sur la grille $\mathcal{V}_N$ de cardinal N, il est clair, par la définition de la domination stochastique, que $\sup_{\nu\in\mathcal{V}_N}|T_N(\nu)| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$. Il reste à montrer pour obtenir le résultat final que $\nu \mapsto T_N(\nu)$ est Lipschitz avec une constante de Lipschitz d'ordre N^p pour un certain p fini:

$$\sup_{\nu\in[0,1]}|T_N(\nu)| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3$$

Cependant, nous remarquons que ce résultat ne permet pas de proposer un test statistique car la distribution asymptotique de $\sup_{\nu\in[0,1]}|T_N(\nu)|$ reste inconnue. Il est donc impossible de contrôler l'erreur de première espèce de cette statistique de test. Néanmoins, le résultat obtenu dans ce chapitre représente un premier pas vers l'obtention de la distribution asymptotique de $\sup_{\nu\in[0,1]}|T_N(\nu)|$. Ce sujet sera décrit plus en détail dans la section Perspectives 0.10.

0.9.2 Contribution du Chapitre 2

Pour évaluer explicitement le terme de correction d'ordre $\mathcal{O}(\frac{B}{N})^2$ dans l'expression de $\frac{1}{M}\text{tr}\ \hat{\mathbf{C}}_N(\nu) - \int f\ \mathrm{d}\mu_{mp,c_N}$, l'approche utilisée dans le Chapitre 1 se révèle être extrêmement longue et fastidieuse en terme de calculs. Dans ce chapitre, nous proposons une approche alternative qui permet de retrouver cette quantité bien plus simplement. Au lieu d'utiliser la décomposition :

$$\begin{aligned}\mathbb{E}\left[\frac{1}{M}\text{tr}\ \tilde{\mathbf{C}}_N(\nu)\right] - \int f\ \mathrm{d}\mu_{mp,c_N} &= \mathbb{E}\left[\frac{1}{M}\text{tr}\ \tilde{\mathbf{C}}_N(\nu)\right] - \mathbb{E}\left[\frac{1}{M}\text{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] \\ &\quad + \mathbb{E}\left[\frac{1}{M}\text{tr}\ \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\right] - \int f\ \mathrm{d}\mu_{mp,c_N}\end{aligned}$$

nous proposons d'étudier directement le comportement de la résolvante $\mathbb{E}(\tilde{\mathbf{C}}_N(\nu) - z\mathbf{I}_M)^{-1}$ en utilisant la structure de covariance des transformée de Fourier $(\boldsymbol{\xi}_N(\nu + \frac{b}{N}))_{b=-B/2,\ldots,B/2}$ et du calcul gaussien. Notons que $\tilde{\mathbf{C}}_N(\nu) = \text{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)\boldsymbol{\Sigma}_N(\nu)^*\text{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ où

$$\boldsymbol{\Sigma}_N(\nu) = \frac{1}{\sqrt{B+1}}\left(\boldsymbol{\xi}_N\left(\nu - \frac{B}{2N}\right), \ldots, \boldsymbol{\xi}_N\left(\nu + \frac{B}{2N}\right)\right)$$

donc $\tilde{\mathbf{C}}_N(\nu)$ peut s'interpréter comme la matrice de covariance empirique de $\text{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)$. Sous l'hypothèse d'indépendance spatiale, les lignes de $\text{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}_N(\nu)$ sont indépendantes. Nous définissons $\boldsymbol{\Theta}_{N,m}(\nu)$ la matrice de covariance des transformée de Fourier renormalisées à la fréquence

ν de la m–ème composante de la série temporelle:

$$\mathbf{\Theta}_{N,m}(\nu) = \mathbb{E}\left(\frac{\xi_{N,m}\left(\nu - \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}, \ldots, \frac{\xi_{N,m}\left(\nu + \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}\right)^* \left(\frac{\xi_{N,m}\left(\nu - \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}, \ldots, \frac{\xi_{N,m}\left(\nu + \frac{B}{2N}\right)}{\sqrt{s_m(\nu)}}\right)$$

D'après des résultats classique sur les transformée de Fourier (voir [13]), pour tout $m \geq 1$,

$$\mathrm{Cov}\left(\xi_{N,m}\left(\nu + \frac{b}{N}\right), \xi_{N,m}\left(\nu + \frac{b'}{N}\right)\right) = s_m\left(\nu + \frac{b}{N}\right)\mathbb{1}(b = b') + \mathcal{O}\left(\frac{1}{N}\right)$$

Ceci implique que pour tout m et ν, $\mathbf{\Theta}_{N,m}(\nu)$ peut s'approximer par l'identité $\mathbf{I}_{B+1}$, ce qui signifie que les colonnes de $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\mathbf{\Sigma}_N(\nu)$ sont presque indépendantes. Ce modèle permet d'expliciter un équivalent déterministe de $\tilde{\mathbf{Q}}_N(z) := (\tilde{\mathbf{C}}_N(\nu) - z\mathbf{I}_M)^{-1}$. On montrera que pour tout $z \in \mathbb{C}_+$, il existe une matrice déterministe $\tilde{\mathbf{T}}_N(z, \nu)$ telle que

$$\frac{1}{M}\mathrm{tr}\ \mathbb{E}\tilde{\mathbf{Q}}_N(\nu, z) - \frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}_N(\nu, z) = \mathcal{O}_z\left(\frac{1}{B^2}\right)$$

où $\mathcal{O}_z(\frac{1}{B^2})$ indique qu'il existe des polynômes P_1 et P_2 dont les constantes et degrés sont indépendants de N et z tel que la quantité en question soit bornée par $P_1(|z|)P_2(\frac{1}{\mathrm{Im}z})\frac{1}{B^2}$. De plus, sous certaines hypothèses sur les séries temporelles, il est possible de montrer que la quantité $\mathcal{O}_z(\frac{1}{B^2})$ est en fait indépendante de ν. Finalement, la matrice $\tilde{\mathbf{T}}_N(\nu, z)$ est déterministe et est définie implicitement par l'unique solution de l'équation:

$$\tilde{\mathbf{T}}_N(\nu, z) = \mathrm{dg}\left(\frac{-1}{z(1 + \overline{\delta}_{N,m}(\nu, z))}, m = 1, \ldots, M\right)$$

où les $(\overline{\delta}_{N,m})_{m\geq 1}$ sont définis par la solution dans l'ensemble des transformée de Stieltjes du système de M équations suivantes:

$$\overline{\delta}_{N,m}(\nu, z) = \frac{1}{B+1}\mathrm{tr}\ \left(-z\mathbf{I}_{B+1} + \frac{c}{M}\sum_{k=1}^{M}\frac{\mathbf{\Theta}_{N,k}(\nu)}{1 + \overline{\delta}_{N,k}(\nu, z)}\right)^{-1}\mathbf{\Theta}_{N,m}(\nu), \quad m \in [M]$$

Nous utilisons maintenant le fait que les colonnes sont presque indépendantes, et il est donc raisonnable de s'attendre à retrouver un comportement de type Marcenko-Pastur pour $\tilde{\mathbf{T}}_N$. Nous montrerons que l'estimation suivante est valide:

$$\sup_{m\geq 1}\sup_{\nu\in[0,1]}\|\mathbf{\Theta}_{N,m}(\nu) - \mathbf{I}_{B+1}\| = \mathcal{O}\left(\frac{B}{N}\right)$$

$$\frac{1}{B+1}\sup_{m\geq 1}\sup_{\nu\in[0,1]}\mathrm{tr}\ (\mathbf{\Theta}_{N,m}(\nu) - \mathbf{I}_{B+1}) - \Upsilon_{N,m}(\nu) = \mathcal{O}\left(\frac{B}{N}\right)^3$$

où $\Upsilon_{N,m}(\nu)$ est le terme d'ordre $\mathcal{O}(\frac{B}{N})^2$ égal à $\Upsilon_{N,m}(\nu) = \frac{1}{2}\frac{s''_m(\nu)}{\hat{s}_m(\nu)}v_N$. Ceci permet de prouver que $\tilde{\mathbf{T}}_N(z) - t_N(z)\mathbf{I}_M \to 0$ pour tout $z \in \mathbb{C}_+$, et d'obtenir que pour tout $f \in C^p$ avec p suffisamment grand,

$$\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \int f\,\mathrm{d}\mu_{mp,c_N} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 < \mathcal{D}_N, f > v_N$$
$$- \left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}\right)^2 < \tilde{\mathcal{D}}_N, f > v_N + \mathcal{O}\left(\frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3\right)$$

où le $\mathcal{O}(\cdot)$ est uniforme en $\nu \in [0,1]$, et où $\tilde{\mathcal{D}}_N$ et $\mathcal{D}_N$ sont des distributions dont la transformée de Stieltjes au point z sont égales respectivement à $(zt_N(z))'$ et $\frac{-c_N(zt_N(z)\tilde{t}_N(z))^3}{1-c_N(zt_N(z)\tilde{t}_N(z))^2}$.

0.9.3 Contribution du Chapitre 3

Dans ce chapitre, nous considérons une série temporelle M–dimensionnelle $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ définie par

$$\mathbf{y}_n = \mathbf{u}_n + \mathbf{v}_n$$

où $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ représente le signal utile et où $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ représente un bruit additif. Le signal utile est modélisé comme la sortie d'un filtre MIMO $M \times K$ inconnu et stable $(\mathbf{H}_k)_{k\in\mathbb{Z}}$ dont l'entrée est un bruit blanc gaussien K–dimensionnel non-observable $(\boldsymbol{\epsilon}_n)_{n\in\mathbb{Z}}$ avec $\mathbb{E}[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^*] = \mathbf{I}_K$, i.e.

$$\mathbf{u}_n = \sum_{k\in\mathbb{Z}}\mathbf{H}_k\boldsymbol{\epsilon}_{n-k}$$

avec probabilité 1. Nous remarquons que K représente le nombre de sources dans le contexte du traitement d'antenne. Pour tout $\nu \in [0,1]$, nous notons $\mathbf{H}(\nu)$ la transformée de Fourier de la séquence $(\mathbf{H}_k)_{k\in\mathbb{Z}}$:

$$\mathbf{H}(\nu) = \sum_{k\in\mathbb{Z}}\mathbf{H}_k\mathrm{e}^{-\mathrm{i}2\pi\nu k}$$

Nous supposons que le signal satisfait l'hypothèse suivante dite de mémoire courte:

$$\sup_{m\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)\|\mathbf{H}_k\| < +\infty$$

où la magnitude du signal satisfait

$$\lim_{M\to\infty}\max_{m=1,\ldots,M}\max_{\nu\in[0,1]}\|\mathbf{h}_m(\nu)\|_2 = 0.$$

où $\|\mathbf{h}_m(\nu)\|_2^2 = \sum_{k=1}^{K}|(\mathbf{H})_{m,k}(\nu)|^2$ représente la puissance totale reçue sur le capteur m par le signal K–dimensionnel à la fréquence ν. Ceci implique que la puissance totale du signal est d'ordre

$$\mathbb{E}\|\mathbf{u}_{m,n}\|^2 = \int_0^1 \|\mathbf{h}_m(\nu)\|^2\,\mathrm{d}\nu = o(1)$$

Concernant le bruit, $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ est modélisé par une série temporelle gaussienne complexe M–dimensionnelle telle que ses composantes $(v_{1,n})_{n\in\mathbb{Z}}, \ldots, (v_{M,n})_{n\in\mathbb{Z}}$ sont mutuellement indépendantes. Nous supposons que le bruit satisfait aussi une hypothèse de mémoire courte:

$$\sup_{m\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)^2|r_m(k)| < +\infty$$

où $r_m(k) = \mathbb{E}[v_{m,n}\overline{v}_{m,n-k}]$. Nous supposons aussi que les densités spectrales associées à ces séries temporelles sont uniformément bornées inférieurement par zéro:

$$\inf_{m\geq 1}\inf_{\nu\in[0,1]} s_m(\nu) > 0$$

où $s_m(\nu) = \sum_{k\in\mathbb{Z}} r_m(k)e^{-\mathrm{i}2\pi\nu k}$. Cette hypothèse implique que la puissance totale du bruit satisfait la condition suivante: il existe des constantes $0 < C_1 \leq C_2 < +\infty$ telles que

$$C_1 < \inf_{N\geq 1}\frac{1}{M}\mathbb{E}\|\mathbf{v}_N\|^2 \leq \sup_{N\geq 1}\frac{1}{M}\mathbb{E}\|\mathbf{v}_N\|^2 < C_2$$

Ainsi, le rapport signal-sur-bruit (SNR) $\frac{\mathbb{E}\|\mathbf{u}_N\|^2}{\mathbb{E}\|\mathbf{v}_N\|^2}$ vérifie $\frac{\mathbb{E}\|\mathbf{u}_N\|^2}{\mathbb{E}\|\mathbf{v}_N\|^2} = \mathcal{O}(\frac{1}{M})$. L'objectif du Chapitre 3 est de détecter la présence de signal dans ce contexte. Pour commencer, nous remarquons que la matrice de densité spectrale $\mathbf{S}_\mathbf{y}$ de $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ est donnée par

$$\mathbf{S}_\mathbf{y}(\nu) = \mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)$$

où $\mathbf{S}_\mathbf{v}(\nu) = \mathrm{dg}\left(s_1(\nu),\ldots,s_M(\nu)\right)$. De manière similaire, par la définition de la matrice de cohérence spectrale,

$$\mathbf{C}_\mathbf{y}(\nu) = \left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)\right)\left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}$$

Sous les hypothèses sur la puissance du signal et du bruit, il est raisonnable d'approximer $\mathrm{dg}(\mathbf{S}_\mathbf{y}(\nu))^{-\frac{1}{2}} = \left(\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_\mathbf{v}(\nu)\right)^{-\frac{1}{2}}$ par $\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}$ et $\mathbf{C}_\mathbf{y}(\nu)$ par

$$\mathbf{C}_\mathbf{y}(\nu) \approx \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_\mathbf{v}(\nu)\right)\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}} = \mathbf{I}_M + \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}$$

Afin d'estimer $\mathbf{C}_\mathbf{y}$, nous considérons l'estimateur de la cohérence spectrale construit à partir du périodogramme moyenné, noté $\hat{\mathbf{C}}_{N,\mathbf{y}}$ et défini en (14). Nous montrons en premier que $\hat{\mathbf{C}}_\mathbf{y}(\nu)$ se comporte asymptotiquement en norme spectrale et uniformément en ν comme une matrice de Wishart "coloré" dont la matrice de covariance population est égale à $\mathbf{I}_M + \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}$:

$$\max_{\nu\in\mathcal{V}_N}\left\|\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\right\| \xrightarrow[M\to\infty]{p.s.} 0$$

où $\mathcal{V}_N$ est l'ensemble des fréquences de Fourier $\left\{\frac{k}{N}, k=0,\ldots,N-1\right\}$,

$$\boldsymbol{\Xi}(\nu) = \mathbf{I}_M + \mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_\mathbf{v}(\nu)^{-\frac{1}{2}}.$$

et $\mathbf{X}_N(\nu)$ est une matrice de taille $M\times(B+1)$ à entrées i.i.d. $\mathcal{N}_\mathbb{C}(0,1)$. Ce résultat, utilisé en conjonction avec les inégalités de Weyl [43, Th. 4.3.1], implique en particulier que chaque valeur propre de la matrice de cohérence spectrale estimée $\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)$ se comporte comme sa contrepartie Wishart:

$$\mathbf{W}_N(\nu) = \boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\boldsymbol{\Xi}(\nu)^{\frac{1}{2}}$$

c'est à dire

$$\max_{m=1,\ldots,M}\max_{\nu\in\mathcal{V}_N}\left|\lambda_m\left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu)\right) - \lambda_m\left(\mathbf{W}_N(\nu)\right)\right| \xrightarrow[M\to\infty]{p.s.} 0.$$

Nous remarquons maintenant que $\boldsymbol{\Xi}(\nu)$ est une perturbation de rang fini K de l'identité $\mathbf{I}_M$. Ceci suggère d'utiliser les résultats concernant les modèles spike multiplicatif développés dans [7] afin de décrire précisément le comportement des plus grandes valeurs propres de $\boldsymbol{\Xi}(\nu)^{\frac{1}{2}}\frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1}\boldsymbol{\Xi}(\nu)^{\frac{1}{2}}$. Ceci permet de proposer une statistique permettant de tester l'hypothèse $\mathcal{H}_0$ contre l'hypothèse $\mathcal{H}_1$

définies par :

$$\mathcal{H}_0 : (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{v}_n)_{n\in\mathbb{Z}}$$
$$\mathcal{H}_1 : (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{u}_n)_{n\in\mathbb{Z}} + (\mathbf{v}_n)_{n\in\mathbb{Z}}$$

Nous considérons la plus grande valeur propre de $\hat{\mathbf{C}}_N(\nu)$ à travers les fréquences de Fourier $\mathcal{V}_N$, et comparons cette valeur à λ_+ :

$$T_{N,\epsilon} = \mathbb{1}_{[\lambda^+ + \epsilon, \infty)} \left(\max_{\nu \in \mathcal{V}_N} \lambda_1 \left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu) \right) \right)$$

où nous rappelons que $\lambda_+ = (1+\sqrt{c})^2$. Nous montrons que ce test est consistant sous le régime asymptotique des grandes dimensions : si sous l'hypothèse $\mathcal{H}_1$, il existe $\gamma_\infty \geq 0$ tel que

$$\max_{\nu \in \mathcal{V}_N} \lambda_1 \left(\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \mathbf{H}(\nu)\mathbf{H}(\nu)^* \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \right) \xrightarrow[M\to\infty]{} \gamma_\infty > \sqrt{c}$$

alors, en définissant

$$\phi(x) = \begin{cases} \frac{(x+1)(x+c)}{x} & \text{if } x > \sqrt{c} \\ \lambda^+ & \text{if } x \leq \sqrt{c} \end{cases}$$

nous montrons que pour tout $0 < \epsilon < \phi(\gamma_\infty) - \lambda^+$ et $i \in \{0,1\}$,

$$\mathbb{P}_i \left(\lim_{M\to\infty} T_{N,\epsilon} = i \right) = 1$$

où $\mathbb{P}_i$ est la mesure de probabilité sous-jacente à l'hypothèse $\mathcal{H}_i$. On note que γ_∞, la limite de la plus grande valeur propre de la perturbation de rang K, peut s'interpreter comme un certain SNR dans le domaine fréquentiel. Finalement, nous montrons qu'un test basé sur les statistiques linéaires des valeurs propres de $\hat{\mathbf{C}}_N(\nu)$ ne permet pas de séparer les deux hypothèses. On considère la LSS suivante :

$$L_{N,\varphi}(\nu) = \frac{1}{M} \sum_{m=1}^{M} \varphi \left(\lambda_m \left(\hat{\mathbf{C}}_{N,\mathbf{y}}(\nu) \right) \right)$$

où φ appartient à une certaine classe de fonction, ici cette classe étant $\varphi \in \mathcal{C}^1((0,+\infty))$. Nous montrerons que

$$\max_{\nu \in \mathcal{V}_N} \left| L_{N,\varphi}(\nu) - \int_{\mathbb{R}} \varphi(\lambda) f(\lambda) \mathrm{d}\lambda \right| \xrightarrow[M\to\infty]{p.s.} 0$$

où f est la densité de la distribution de Marcenko-Pastur pour $c < 1$ donnée par

$$f(\lambda) = \frac{\sqrt{(\lambda - \lambda^-)(\lambda^+ - \lambda)}}{2\pi c \lambda} \mathbb{1}_{[\lambda^-, \lambda^+]}(\lambda)$$

avec $\lambda^\pm = (1 \pm \sqrt{c})^2$. Ceci montre que les LSS de la matrice de cohérence spectrale estimée convergent vers la même limite, indépendamment du fait que les observations contiennent uniquement du bruit ou du bruit et signal. Ceci montre que toute statistique de test se reposant uniquement sur une LSS de la matrice de cohérence spectrale estimée est incapable de distinguer de manière consistante la présence ou l'absence de signal dans notre régime des grandes dimensions.

0.9.4 Contribution du Chapitre 4

Dans ce chapitre nous considérons le même contexte que dans le Chapitre 3, à l'exception que nous ne supposons plus que $\lim_{M\to\infty} \max_{m=1,\ldots,M} \max_{\nu\in[0,1]} \|\mathbf{h}_m(\nu)\|_2 = 0$. Cette hypothèse indiquait que la puissance totale émise par le signal K−dimensionnel reçue à la fréquence ν sur chaque capteur m converge vers zéro. Dans certain cas, il peut être d'intérêt de ne pas supposer que c'est le cas, et nous montrons qu'il est encore possible d'explicitement écrire le comportement au premier ordre des valeurs propres spike de la matrice de cohérence spectrale estimée dans ce cas plus général. La conséquence principale de cet affaiblissement d'hypothèse est que l'approximation $\mathrm{dg}(\mathbf{S}_{\mathbf{y}}(\nu)) \approx \mathrm{dg}(\mathbf{S}_{\mathbf{v}}(\nu))$ n'est plus valide, et nous devons à la place écrire $\mathrm{dg}(\mathbf{S}_{\mathbf{y}}(\nu)) = \mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) + \mathbf{S}_{\mathbf{v}}(\nu)$. Nous notons $\mathbf{D}_{\mathbf{u}}(\nu) = \mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*)$ et supposons qu'il existe pour chaque fréquence $\nu \in [0,1]$ une séquence de matrices diagonale $(\mathbf{D}_{\mathbf{u},\infty}(\nu))_{N\geq 1}$ de taille $M \times M$ avec un nombre fini d'entrées non-nulles telle que

$$\|\mathbf{D}_{\mathbf{u}}(\nu) - \mathbf{D}_{\mathbf{u},\infty}(\nu)\| \xrightarrow{N\to+\infty} 0.$$

La condition sur le nombre d'entrées non-nulles de $\mathbf{D}_{\mathbf{u},\infty}(\nu)$ assure que cette matrice reste de rang fini asymptotiquement, ce qui sera utile pour appliquer les résultats sur les modèles spike issus de [7]. En utilisant la même stratégie qu'en Chapitre 3, nous montrons que pour tout $\nu \in \mathcal{V}_N$, il existe une matrice aléatoire $\mathbf{X}_N(\nu)$ à entrées i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ telle que

$$\max_{\nu\in\mathcal{V}_N} \left\| \hat{\mathbf{C}}_N(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \right\| \xrightarrow{N\to+\infty} 0$$

où

$$\boldsymbol{\Xi}(\nu) = \mathbf{I}_M + (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}} (\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)) (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}}$$

Ceci est en contraste avec le résultat du Chapitre 3, où la matrice de covariance population s'écrivait $\mathbf{I}_M + \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} (\mathbf{H}(\nu)\mathbf{H}(\nu)^*) (\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}$. Cette nouvelle expression de $\boldsymbol{\Xi}(\nu)$ nous indique que le terme de perturbation peut induire des spikes négatifs, car la matrice $\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)$ n'est pas nécessairement définie positive. De plus, comme le rang de $\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)$ est fini, les résultats sur les modèles spike de [7] prouvent que la plus grande et plus petite valeur propre de $\hat{\mathbf{C}}_N(\nu)$ peut s'échapper du bulk de la distribution de Marcenko-Pastur si une certaine condition sur le SNR est respectée. Une nouvelle statistique pour tester la présence ou l'absence de signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ est en conséquence proposée:

$$T_\epsilon = \mathbf{1}\left(\sup_{\nu\in\mathcal{V}_N} \lambda_1(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) > \lambda^+ + \epsilon \text{ ou } \inf_{\nu\in\mathcal{V}_N} \lambda_M(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) < \lambda^- - \epsilon \right)$$

et il est montré que cette statistique est consistante pour séparer l'hypothèse nulle $\mathcal{H}_0$ de l'alternative $\mathcal{H}_1$ si

$$\gamma_\infty^+ > \sqrt{c} \text{ ou } \gamma_\infty^- < -\sqrt{c}.$$

où γ_∞^+ et γ_∞^- sont définis par:

$$\max_{\nu\in\mathcal{V}_N} \lambda_1(\boldsymbol{\Omega}(\nu)) \xrightarrow[M\to\infty]{} \gamma_\infty^+ > \sqrt{c}$$
$$\min_{\nu\in\mathcal{V}_N} \lambda_M(\boldsymbol{\Omega}(\nu)) \xrightarrow[M\to\infty]{} \gamma_\infty^- < -\sqrt{c}$$

avec $\boldsymbol{\Omega}(\nu) = (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}} (\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)) (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}}$.

0.9.5 Contribution du Chapitre 5

Ce chapitre se démarque des autres, car il s'interesse non pas aux valeurs propres de la matrice de cohérence spectrale estimée, mais plutôt au comportement de ses entrées individuelles. Plus précisement, nous considérons M séries temporelles gaussiennes complexes stationnaires $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ et pour tout $i,j \in \{1,\ldots,M\}$, nous notons s_{ij} et c_{ij} la densité spectrale et la cohérence spectrale entre $(y_{i,n})_{n\in\mathbb{Z}}$ et $(y_{j,n})_{n\in\mathbb{Z}}$ données respectivement par $s_{ij}(\nu) = (\mathbf{S}(\nu))_{ij}$ et $c_{ij}(\nu) = (\mathbf{C}(\nu))_{ij}$ où $\mathbf{S}(\nu)$ et $\mathbf{C}(\nu)$ sont respectivement les matrices de densité/cohérence spectrales définies en (11) et (12). Ce chapitre est lui aussi motivé par le problème du test d'indépendance d'un grand nombre de séries temporelles gaussiennes:

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |s_{ij}(\nu)|^2 = 0,$$

ou de manière équivalente

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |c_{ij}(\nu)|^2 = 0,$$

Ceci suggère de calculer un estimateur consistant de ces quantités, et de tester s'ils sont suffisamment proches de zéro. Étant donné N observations pour chaque série temporelle $(y_{1,n})_{n=1,\ldots,N}, \ldots, (y_{M,n})_{n=1,\ldots,N}$, nous considérons l'estimateur issu du périodogramme moyenné $\hat{c}_{N,ij}$ de c_{ij} donné par (14). Sous l'hypothèse

$$\mathcal{H}_0 : (y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}} \text{ sont mutuellement décorrélés,}$$

et quelques hypothèses techniques supplémentaires sur les densités spectrales des M séries temporelles, nous montrons que dans le régime des grandes dimensions décrit en (15),

$$\mathbb{P}\left((B+1) \max_{(i,j,\nu)\in\mathcal{I}_N} |\hat{c}_{N,ij}(\nu)|^2 \le t + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2} \right) \xrightarrow{N\to+\infty} e^{-e^{-t}} \tag{20}$$

où

$$\mathcal{I}_N := \{(i,j,\nu) : i,j \in [M] \text{ tel que } i<j,\ \nu \in \mathcal{G}_N\}$$

avec $[M] = \{1,\ldots,M\}$ et où

$$\mathcal{G}_N := \left\{ k\frac{B+1}{N} : k \in \mathbb{N}, 0 \le k \le \frac{N}{B+1} \right\}$$

est l'ensemble inclut dans celui des fréquences de Fourier

$$\mathcal{V}_N := \left\{ \frac{k}{N} : k \in \mathbb{N}, 0 \le k \le N-1 \right\}$$

dont les éléments sont séparés d'une distance $(B+1)/N$ (au lieu de $1/N$ dans le cas de $\mathcal{F}_N$). Ainsi, (20) annonce que $\max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{N,ij}(\nu)|^2$, après un recentrage et une renormalisation adaptée, converge en distribution vers une distribution des valeurs extrêmes de type I, aussi connue sous le nom de distribution de Gumbel. Il apparaîtra alors clair au cours de la preuve que le terme en $\log\frac{M(M-1)}{2}$ est lié au fait de prendre le maximum sur $(i,j) \in [M]^2$ avec $i \neq j$, tandis que le terme en $\log\frac{N}{B+1}$ est une conséquence du maximum sur $\nu \in \mathcal{G}_N$. Les étapes essentielles de la preuve sont les suivantes. Nous écrivons la représentation linéaire causale

$$y_{m,n} = \sum_{k\ge 0} a_{m,k}\epsilon_{m,n-k}$$

où $(\epsilon_{1,k})_{k\in\mathbb{Z}}, \ldots, (\epsilon_{M,k})_{k\in\mathbb{Z}}$ sont des séquences mutuellement indépendantes de variables aléatoires $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d, et $(a_{1,k})_{k\in\mathbb{N}}, \ldots, (a_{M,k})_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$. On note aussi

$$h_m(\nu) = \sum_{k=0}^{+\infty} a_{m,k} e^{-2i\pi k\nu}$$

tel que $|h_m(\nu)|^2 = s_m(\nu)$. Il est montré dans [94] (et étendu sous des hypothèses plus faibles dans [67] et [19]) que les transformées de Fourier $\xi_{N,y_m}(\nu)$ se comportent comme $h_m(\nu)\xi_{N,\epsilon_m}(\nu)$ au sens suivant: pour tout $\kappa > 0$:

$$\mathbb{P}\left[\lim_{N\to+\infty} |\xi_{N,y_m}(\nu) - h_m(\nu)\xi_{N,\epsilon_m}(\nu)| > \kappa\right] = 0 \tag{21}$$

Ceci suggère de définir $\tilde{s}_{N,ij}(\nu)$, une approximation de $\hat{s}_{N,ij}(\nu) := (\hat{\mathbf{S}}_N(\nu))_{ij}$ où $\hat{\mathbf{S}}_N(\nu)$ est défini par (13), comme:

$$\tilde{s}_{N,ij}(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} h_i\left(\nu + \frac{b}{N}\right)\overline{h_j\left(\nu + \frac{b}{N}\right)}\xi_{N,\epsilon_i}\left(\nu + \frac{b}{N}\right)\overline{\xi_{N,\epsilon_j}\left(\nu + \frac{b}{N}\right)}$$

L'équation (21) est ensuite clé pour prouver que $(B+1)\sup_{i,j,\nu} |\hat{c}_{N,ij}(\nu)|^2$ a la même distribution limite que $(B+1)\sup_{i,j,\nu} |\tilde{c}_{N,ij}(\nu)|^2$ où $\tilde{c}_{N,ij}(\nu) = \frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)}$ avec

$$\sigma^2_{N,ij}(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} s_i\left(\nu + \frac{b}{N}\right) s_j\left(\nu + \frac{b}{N}\right)$$

Cette décomposition permet de découpler l'action du filtre des innovations, afin qu'il ne reste qu'à gérer des transformées de Fourier de séquences i.i.d. de variables aléatoires notées $\epsilon_{m,n}$. Afin de montrer que $\sup_{i,j,\nu} |\tilde{c}_{N,ij}(\nu)|^2$ suit asymptotiquement une loi Gumbel, nous utilisons un théorème d'approximation Poisson issu de [3] basé sur la méthode de Chen-Stein. Cette méthode a été utilisée dans [45] afin d'étudier le comportement de la plus grande entrée hors-diagonale de matrice de corrélation construite à partir d'entrées i.i.d. Cette méthode nécessite de prouver le résultat de déviation modérée suivant : il existe une constante $\eta > 0$ telle que pour tout $C > 0$,

$$\max_{t\in[0,CB^\eta]} \max_{(i,j,\nu)\in\mathcal{I}} \left|\mathbb{P}\left((B+1)\frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)} > t^2\right)e^{t^2} - 1\right| \xrightarrow[N\to\infty]{} 0 \tag{22}$$

et

$$\max_{t,s\in[0,CB^\eta]} \max_{\substack{(i,j,\nu)\in\mathcal{I} \\ (i',j',\nu)\in\mathcal{I}_{(i,j,\nu)}}} \left|\mathbb{P}\left((B+1)\frac{|\tilde{s}_{N,ij}(\nu)|^2}{\sigma^2_{N,ij}(\nu)} > t^2,\ (B+1)\frac{|\tilde{s}_{N,i'j'}(\nu)|^2}{\sigma^2_{N,i'j'}(\nu)} > s^2\right)\right.$$
$$\left.\times\, e^{t^2+s^2} - 1\right| \xrightarrow[N\to\infty]{} 0.$$

Ceci permet de terminer la preuve de (20).

0.10 Perspectives

Nous clôturons cette Introduction en énonçant quelques directions possibles qui permettraient de prolonger les travaux présentés dans ce manuscrit.

- Tout d'abord, tel que mentionné dans le paragraphe exposant les contributions du Chapitre 1, plus d'information sur la distribution limite de

$$T_N = \sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int f\, \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M} \sum_{m=1}^{M} \frac{s_m'(\nu)}{s_m(\nu)} \right)^2 < \mathcal{D}_N, f > v_N \right|$$

  est necessaire pour controler l'erreur de type-I d'une statistique de test basée dessus (tel que par exemple $\mathbb{1}(T_N > \kappa)$ pour un certain seuil κ). Cependant, nous pensons qu'il sera difficile d'obtenir ce résultat. Une approche plus raisonnable serait la suivante : à partir de la représentation linéaire causale, tel que fait dans le Chapitre 5, $y_{m,n} = \sum_{k \in \mathbb{Z}} a_{m,k} \epsilon_{m,n-k}$ où $(\epsilon_{1,n})_{n \in \mathbb{Z}}, \ldots, (\epsilon_{M,n})_{n \in \mathbb{Z}}$ sont des sequences mutuellement indépendantes de variables $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d, et $(a_{1,k})_{k \in \mathbb{Z}}, \ldots, (a_{M,k})_{k \in \mathbb{Z}} \in \ell^2(\mathbb{N})$. Definissons $h_m(\nu) = \sum_{k \in \mathbb{Z}} a_{m,k} e^{-2\mathrm{i}\pi k \nu}$, $\mathbf{H}(\nu) = \mathrm{dg}(h_1(\nu), \ldots, h_M(\nu))$, et $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu) = (\boldsymbol{\xi}_{N,\epsilon_1}(\nu), \ldots, \boldsymbol{\xi}_{N,\epsilon_M}(\nu))^T$. En utilisant l'approximation $\boldsymbol{\xi}_{N,\mathbf{y}}(\nu) \approx \mathbf{H}(\nu) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu)$, qui a déjà été utilisé dans le Chapitre 5, il fait sens de considérer $\check{\mathbf{S}}_N(\nu)$, une approximation de $\hat{\mathbf{S}}_N(\nu)$, définie par

$$\check{\mathbf{S}}_N(\nu) = \frac{1}{B+1} \sum_{-B/2}^{B/2} \mathbf{H}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}\left(\nu + \frac{b}{N}\right)^* \mathbf{H}\left(\nu + \frac{b}{N}\right)^*$$

  et $\check{\mathbf{C}}_N(\nu)$, une approximation de $\hat{\mathbf{C}}_N(\nu)$ définie par

$$\check{\mathbf{C}}_N(\nu) = \mathrm{dg}(\check{\mathbf{C}}_N(\nu))^{-\frac{1}{2}} \check{\mathbf{C}}_N(\nu) \mathrm{dg}(\check{\mathbf{C}}_N(\nu))^{-\frac{1}{2}}.$$

  Pour $\nu_1 \neq \nu_2 \in \mathcal{G}_N := \left\{ k \frac{B+1}{N} : k \in \mathbb{N}, 0 \leq k \leq \frac{N}{B+1} \right\}$, $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu_1)$ est indépendant de $\boldsymbol{\xi}_{N,\boldsymbol{\epsilon}}(\nu_2)$, et il est donc clair que $\frac{1}{M} \mathrm{tr}\ f(\check{\mathbf{C}}_N(\nu_1))$ est aussi independent de $\frac{1}{M} \mathrm{tr}\ f(\check{\mathbf{C}}_N(\nu_2))$. Ceci devrait permettre d'obtenir la distribution asymptotique de $\check{T}_N$ define par

$$\check{T}_N = \sup_{\nu \in \mathcal{G}_N} \left| \frac{1}{M} \mathrm{tr}\ f(\check{\mathbf{C}}(\nu)) - \int f\, \mathrm{d}\mu_{mp,c_N} - \left(\frac{1}{M} \sum_{m=1}^{M} \frac{s_m'(\nu)}{s_m(\nu)} \right)^2 < \mathcal{D}_N, f > v_N \right|$$

  Il resterait à controler le terme d'erreur entre $\check{\mathbf{C}}_N(\nu)$ et $\hat{\mathbf{C}}_N(\nu)$ pour s'assurer que $\frac{1}{M} \mathrm{tr}\ f(\check{\mathbf{C}}_N(\nu))$ et $\frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu))$ suivent asymptotiquement la même distribution. Ceci représenterait une amélioration de notre résultat présenté en Chapitre 1 et Chapitre 2, car cela permettrait de controler l'erreur de type-I de la statistique de test $\mathbb{1}(\sup_{\nu \in \mathcal{G}_N} |T_N(\nu)| > \kappa)$. Cependent, la manière de gérer le supremum sur l'interval continu de fréquences $[0,1]$ à la place de la grille discrete $\mathcal{G}_N$ reste peu claire.

- Dans le Chapitre 1 nous nous sommes concentrés sur $\sup_\nu |T_N(\nu)|$, mais il aurait aussi pu être intéressant d'étudier des statistiques intégrées sur les fréquences, tel que cela a été fait dans [25], où l'auteur s'est intéressé à des statistiques de la forme $\int_{\nu \in [0,1]} |T_N(\nu)|^2 \,\mathrm{d}\nu$.

- Une fois la distribution asymptotique de T_N sous $\mathcal{H}_0$ obtenu, il serait naturel d'étudier la puissance de test sous une certaine classe d'alternative où la densité spectrale $\mathbf{S}(\nu)$ n'est pas diagonale pour tout ν, en utilisant des outils similaire à ceux utilisés dans [15] par exemple.

- L'hypothèse de gaussianité des séries temporelles est cruciale dans tous les chapitres de ce manuscrit, et il pourrait être intéressant d'étendre nos résultats dans le cas de séries temporelles non-gaussiennes. Par exemple, dans le Chapitre 1, l'hypothèse de gaussianité était nécessaire pour approximer $\tilde{\mathbf{C}}_N(\nu)$ par $\frac{1}{B+1}\mathbf{X}_N(\nu)\mathbf{X}_N(\nu)^*$ où $\mathbf{X}_N(\nu)$ est une matrice de taille $M \times (B+1)$ à entrées i.i.d. Sans l'hypothèse gaussienne, les entrées de notre matrice $\mathbf{X}_N(\nu)$ ne serait plus i.i.d. (mais seulement de covariance nulle), ce qui empêcherait de poursuivre la stratégie de preuve utilisée dans le Chapitre 1. Cependant, nous remarquons aussi que l'approche utilisée dans le Chapitre 2 semble plus adaptée à manipuler le cas non-gaussien, car le modèle correspondant a été historiquement étudié dans [92] sous une hypothèse de moment-fini d'ordre huit. Cependant, la majorité du travail restant serait de contrôler les nouveaux termes d'erreurs qui apparaîtraient du à cette condition plus faible.

- Il devrait être possible d'obtenir des résultats concernant le comportement local du spectre de $\hat{\mathbf{S}}_N(\nu)$ et $\hat{\mathbf{C}}_N(\nu)$. Concernant les applications en détection de signal, il serait utile de prouver un comportement de type Tracy-Widom pour la plus grande valeur propre de la matrice de densité/cohérence spectrale.

- Enfin, nous soulevons le fait que nous avons systematiquement approximé $\hat{\mathbf{C}}_N(\nu)$ par $\tilde{\mathbf{C}}_N(\nu) = \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\hat{\mathbf{S}}_N(\nu)\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$, puis étudié $\tilde{\mathbf{C}}_N(\nu)$ au lieu de $\hat{\mathbf{C}}_N(\nu)$. L'avantage de $\tilde{\mathbf{C}}_N(\nu)$ sur $\hat{\mathbf{C}}_N(\nu)$ est que la renormalisation en $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ n'est plus aléatoire. Néanmoins, ceci necessite maintenant que $\mathrm{dg}(\mathbf{S}(\nu))$ ne s'annule plus, ce qui impose donc la condition technique suivante $\inf_{\nu\in[0,1]}\inf_{m\geq 1} s_m(\nu) > 0$. Il pourrait être interesant de trouver une approche alternative pour ne pas avoir besoin de cette hypothèse.

0.11 Papiers de Journal/Conférence associés à ce manuscrit

- Le Chapitre 1 a été accepté à Electronic Journal of Statistics, et sera prochainement publié.

- Le Chapitre 3 a été d'abord publié à la conférence ICASSP : Alexis Rosuel et al. "On the frequency domain detection of high dimensional time series". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 8782–8786, et plus tard republié sous une version étendue dans IEEE Transactions on Signal Processing: A. Rosuel et al. "On the Detection of Low-Rank Signal in the Presence of Spatially Uncorrelated Noise: A Frequency Domain Approach". In: *IEEE Transactions on Signal Processing* 69 (2021), pp. 4458–4473. DOI: `10.1109/TSP.2021.3099644`. Nous présentons dans ce manuscrit au Chapitre 3 uniquement la version acceptée à IEEE TSP.

- Le Chapitre 5 est en ce moment en review à Journal of Multivariate Analysis.

0.12 Notations

- Nous supposons que $M := M(N)$ et $B = B(N)$ sont des fonctions à valeur entières de N, et que $M(N), B(N) \xrightarrow{N\to+\infty} +\infty$ tel que $c_N := \frac{M}{B+1}$ satisfait $0 < \inf_{N\geq 1} c_N \leq \sup_{N\geq 1} c_N < +\infty$ et $c_N \xrightarrow{N\to+\infty} c \in (0, +\infty)$. De plus, nous supposons qu'il existe $\frac{1}{2} < \alpha < 1$ tel que $B = \mathcal{O}(N^\alpha)$. Ceci implique en particulier que $\frac{B}{N} \xrightarrow{N\to+\infty} 0$ et $\frac{M}{N} \xrightarrow{N\to+\infty} 0$. La notation $N \to +\infty$ fera reférence à ce regime dans toute la suite.

- Les quantités introduites et étudiées au cours de ce manuscrit vont généralement dépendre de N et ν, mais nous choisirons, quand le contexte le permettra, de supprimer la dépendance explicite afin de simplifier les notations. Par exemple, $\hat{\mathbf{S}}$ pourra être utilisé pour représenter $\hat{\mathbf{S}}_N(\nu)$, l'estimateur de la matrice de densité spectrale à la fréquence ν obtenue à partir des observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$. De manière similaire, $\hat{\mathbf{C}}_N(\nu), \hat{c}_{N,ij}(\nu), \boldsymbol{\xi}_N(\nu), \ldots$ pourront être représentés par $\hat{\mathbf{C}}, \hat{c}_{ij}, \boldsymbol{\xi}, \ldots$.

- $\mathbb{R}_+$ et $\mathbb{R}_-$ représentent respectivement l'ensemble des réels positifs et négatifs, et nous notons $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$ et $\mathbb{R}_-^* = \mathbb{R}_- \setminus \{0\}$. Nous définissons $\mathbb{C}_+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$.

- Par "nice constant", nous désignerons une constante déterministe positive qui ne dépend ni des paramètres M, B et N, ni de la variable complexe z. Dans la suite, C représentera une constante générique donc la valeur exacte n'a pas d'importance et qui pourra changer d'une ligne à l'autre. Un "nice polynomial" P est un polynôme dont le degré et les coefficients sont des "nice constants".

- Pour une matrice $\mathbf{A}$, les notations $\|\mathbf{A}\|$ et $\|\mathbf{A}\|_F$ représentent respectivement la norme spectrale et la norme de Frobenius. La transposée et la transconjugué de $\mathbf{A}$ sont respectivement noté $\mathbf{A}^T$ et $\mathbf{A}^*$. Pour une matrice $\mathbf{B}$ de la même taille que $\mathbf{A}$, $\mathbf{A} \geq \mathbf{B}$ indique que $\mathbf{A} - \mathbf{B}$ est définie positive. Si de plus A est une matrice carré, $\mathrm{Im}(\mathbf{A})$ désigne la matrice Hermitienne définie par $\mathrm{Im}(\mathbf{A}) = \frac{\mathbf{A}-\mathbf{A}^*}{2\mathrm{i}}$.

- $C_c(\mathbb{R}, \mathbb{R})$ représente l'ensemble des fonctions à valeur réelle à support compact définie sur $\mathbb{R}$. $C^{(k)}(\mathbb{R}, \mathbb{R})$ représente l'ensemble des fonctions k–differentiables à valeur réelle définies sur $\mathbb{R}$.

- Si ζ est une variable aléatoire, nous noterons ζ° la variable aléatoire centrée définie par $\zeta^\circ = \zeta - \mathbb{E}\zeta$.

- La distribution gaussienne complexe circulaire de variance σ^2 sera notée $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ et un vecteur aléatoire $\mathbf{x}$ dans $\mathbb{C}^n$ suivra la distribution $\mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{R})$ si $\mathbf{b}^*\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{b}^*\mathbf{R}\mathbf{b})$ pour tout vecteur colonne déterministe $\mathbf{b}$ et une matrice $\mathbf{R}$ de taille $n \times n$ définie positive.

Part I

Linear spectral statistics of the estimated spectral coherency

Chapter 1

Approach based on the approximation of the estimated coherency matrix by a Wishart matrix

1.1 Introduction

1.1.1 The addressed problem and the results

We consider an M–variate zero-mean complex Gaussian stationary time series [1] $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ and assume that the samples $\mathbf{y}_1,\ldots,\mathbf{y}_N$ are available. We introduce the traditional frequency smoothed periodogram estimate $\hat{\mathbf{S}}(\nu)$ of the spectral density of $\mathbf{y}$ at frequency ν defined by

$$\hat{\mathbf{S}}(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_{\mathbf{y}}\left(\nu+\frac{b}{N}\right)\boldsymbol{\xi}_{\mathbf{y}}\left(\nu+\frac{b}{N}\right)^* \tag{1.1}$$

where B is an even integer, which represents the smoothing span, and

$$\boldsymbol{\xi}_{\mathbf{y}}(\nu) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\mathbf{y}_n e^{-2i\pi(n-1)\nu} \tag{1.2}$$

is the renormalized Fourier transform of $(\mathbf{y}_n)_{n=1\ldots,N}$. The corresponding estimated spectral coherency matrix is defined as:

$$\hat{\mathbf{C}}(\nu) = \mathrm{dg}\left(\hat{\mathbf{S}}(\nu)\right)^{-\frac{1}{2}}\hat{\mathbf{S}}(\nu)\mathrm{dg}\left(\hat{\mathbf{S}}(\nu)\right)^{-\frac{1}{2}} \tag{1.3}$$

where $\mathrm{dg}(\hat{\mathbf{S}}(\nu)) = \hat{\mathbf{S}}(\nu)\odot\mathbf{I}_M$, with $\odot$ denoting the Hadamard product (ie. entrywise product) and $\mathbf{I}_M$ is the M–dimensional identity matrix. Under the hypothesis $\mathcal{H}_0$ that the M components $(y_{1,n})_{n\in\mathbb{Z}},\ldots,(y_{M,n})_{n\in\mathbb{Z}}$ of $\mathbf{y}$ are mutually uncorrelated, we evaluate the behaviour of certain Linear Spectral Statistics (LSS) of the eigenvalues of $\hat{\mathbf{C}}(\nu)$ in asymptotic regimes where $N\to+\infty$ and both $M=M(N)$ and $B=B(N)$ converge towards $+\infty$ in such a way that $M(N)=\mathcal{O}(N^\alpha)$ for $\alpha\in(1/2,1)$ and $c_N=\frac{M(N)}{B(N)}\to c$ where $c\in(0,1)$. We denote by $\mu_{MP}^{(c)}$ the Marcenko-Pastur distribution with parameter $c<1$ defined by

$$d\mu_{MP}^{(c)}(\lambda) = \frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{2\pi c\lambda}\mathbb{1}_{\lambda\in[\lambda_-;\lambda_+]}(\lambda)\,\mathrm{d}\lambda, \quad \lambda_\pm = (1\pm\sqrt{c})^2$$

[1] any finite linear combination x of the components of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is a complex Gaussian random variable, i.e. $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ are independent zero-mean Gaussian random variables having the same variance

and define the sequences $(u_N)_{N\geq 1}$ and $(v_N)_{N\geq 1}$ by

$$u_N = \frac{1}{B} + \frac{\sqrt{B}}{N} + \left(\frac{B}{N}\right)^3 \tag{1.4}$$

and

$$v_N = \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{b}{N}\right)^2. \tag{1.5}$$

We notice that

$$u_N = \mathcal{O}\left(\frac{1}{B}\right)\mathbf{1}_{\frac{1}{2}\leq\alpha\leq\frac{2}{3}} + \mathcal{O}\left(\frac{\sqrt{B}}{N}\right)\mathbf{1}_{\frac{2}{3}\leq\alpha\leq\frac{4}{5}} + \mathcal{O}\left(\frac{B}{N}\right)^3\mathbf{1}_{\alpha\geq\frac{4}{5}} \tag{1.6}$$

and $v_N = \mathcal{O}\left((\frac{B}{N})^2\right)$, as well as $\frac{u_N}{v_N} \to 0$ if $\alpha > 2/3$ and $\frac{u_N}{v_N} \to +\infty$ if $\alpha < 2/3$. Then, if $(s_m)_{m=1,\ldots,M}$ represent the spectral densities of the scalar time series $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$, for each function f defined on $\mathbb{R}^+$ and $\mathcal{C}^\infty$ in a neighbourhood of the support $[\lambda_-;\lambda_+]$ of $\mu_{MP}^{(c)}$, it holds that for each $\epsilon > 0$, there exist a $\gamma(\epsilon) := \gamma > 0$ such that for each N large enough:

$$\mathbb{P}\left[\sup_{\nu\in[0,1]}\left|\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)} - r_N(\nu)\,\phi_N(f)\,v_N\,\mathbf{1}_{\alpha>2/3}\right| > N^\epsilon u_N\right] \leq \exp -N^\gamma \tag{1.7}$$

where $r_N(\nu)$ is defined by

$$r_N(\nu) = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s_m'(\nu)}{s_m(\nu)}\right)^2 \tag{1.8}$$

and where $\phi_N(f)$ is a deterministic $\mathcal{O}(1)$ term which coincides with the action of function f on a certain compactly supported distribution D_N (to be made precised later) depending on the Marcenko-Pastur distribution $\mu_{MP}^{(c_N)}$. In other words, under $\mathcal{H}_0$, uniformly w.r.t. the frequency ν, $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)$ behaves as $\int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$. If $\alpha \leq 2/3$, with high probability, the order of magnitude of the corresponding error is not larger than $u_N = \frac{1}{B} = \mathcal{O}(\frac{1}{N^\alpha})$. If $\alpha > 2/3$, $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$ behaves as the deterministic $\mathcal{O}(\frac{B}{N})^2$ term $r_N(\nu)\,\phi_N(f)\,v_N$, and the rate of convergence towards 0 of the corrected statistics $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)} - r_N(\nu)\,\phi_N(f)\,v_N\,\mathbf{1}_{\alpha>2/3}$ appears to be u_N which satisfies $\frac{u_N}{v_N} \to 0$.

Our approach is based on the observation that in the above asymptotic regime, $\hat{\mathbf{S}}(\nu)$ can be interpreted as the sample covariance matrix of the large vectors $(\boldsymbol{\xi}_{\mathbf{y}}(\nu+\frac{b}{N}))_{b=-B/2,\ldots,B/2}$. Classical time series analysis results suggest that the vectors $(\boldsymbol{\xi}_{\mathbf{y}}(\nu+\frac{b}{N}))_{b=-B/2,\ldots,B/2}$ appear as "nearly" i.i.d. zero mean complex random vectors with covariance matrix $\mathbf{S}(\nu)$ where $\mathbf{S}(\nu) = \mathrm{dg}\,(s_1(\nu),\ldots,s_M(\nu))$. $\hat{\mathbf{C}}(\nu)$ can be interpreted as the sample autocorrelation matrix of the above vectors. As it is well-known that the empirical eigenvalue distribution of the sample autocorrelation matrix of i.i.d. large random vectors converges towards the Marcenko-Pastur distribution (see e.g. [46]), it is not surprising that $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)$ behaves as $\int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)}$. Our main results are thus obtained using tools borrowed from large random matrix theory (see e.g. [70], [5]) and from frequency domain time series analysis techniques (see e.g. [13]).

1.1.2 Motivation

This chapter is motivated by the problem of testing whether the components of $\mathbf{y}$ are uncorrelated or not when the dimension M of $\mathbf{y}$ is large and the number of observations N is significantly larger than M. For this, a possible way would be to estimate the spectral coherency matrix, equal to $\mathbf{I}_M$ at each frequency ν under $\mathcal{H}_0$, by the standard estimate $\hat{\mathbf{C}}(\nu)$ defined by (1.3) for a relevant choice of B, and to compare, for example, the supremum over ν of the spectral norm $\|\hat{\mathbf{C}}(\nu) - \mathbf{I}_M\|$ to a threshold. To understand the conditions under which such an approach should provide satisfying results, we mention that under some mild extra assumptions, it can be shown that

$$\sup_{\nu} \|\hat{\mathbf{S}}(\nu) - \mathbf{S}(\nu)\| \xrightarrow[N\to+\infty]{a.s.} 0$$

as well as

$$\sup_{\nu} \|\hat{\mathbf{C}}(\nu) - \mathbf{I}_M\| \xrightarrow[N\to+\infty]{a.s.} 0$$

in asymptotic regimes where N, B, M converge towards $+\infty$ in such a way that $\frac{B}{N} \to 0$ and $\frac{M}{B} \to 0$. Therefore, $\hat{\mathbf{C}}(\nu)$ is likely to be close to $\mathbf{I}_M$ for each ν if both $\frac{B}{N}$ and $\frac{M}{B}$ are small enough. However, if M is large and the number of available samples N is not arbitrarily large w.r.t. M, it may be impossible to choose the smoothing span B in such a way that $\frac{B}{N} \ll 1$ and $\frac{M}{B} \ll 1$. In such a context, the predictions provided by the asymptotic regime $\frac{B}{N} \to 0$ and $\frac{M}{B} \to 0$ will not be accurate, and any test comparing $\hat{\mathbf{C}}(\nu)$ to $\mathbf{I}_M$ for each ν will provide poor results. To solve this issue, we propose to choose B of the same order of magnitude as M. In this case, $\hat{\mathbf{C}}(\nu)$ has of course no reason to be close to $\mathbf{I}_M$ for each ν. If $\frac{M}{N}$, or equivalently if $\frac{B}{N}$ is small enough, the asymptotic regime where both M and B converge towards $+\infty$ at the same rate appears relevant to understand the behaviour of $\hat{\mathbf{C}}(\nu)$. We mention in particular that the condition $\alpha > 1/2$ implies that the rate of convergence of $\frac{M}{N}$ towards 0 is moderate, which is in accordance with practical situations in which the sample size is not arbitrarily large. Our asymptotic results thus suggest that if $\frac{M}{N}$ is small enough and if B is chosen of the same order of magnitude as M, then it seems reasonable to test that the components of $\mathbf{y}$ are uncorrelated by comparing

$$\frac{1}{u_N} \sup_{\nu\in[0,1]} \left| \frac{1}{M} \mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu) \right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)} - \hat{r}_N(\nu)\, \phi_N(f)\, v_N\, \mathbf{1}_{\alpha>2/3} \right|$$

to a well chosen threshold, where $\hat{r}_N(\nu)$ represents an estimate of $r_N(\nu)$ accurate enough to keep equal to u_N the convergence rate towards 0 of the modified statistics. We notice that our results just characterize the order of magnitude of the above statistics under $\mathcal{H}_0$, and that we do not provide asymptotic approximation of its distribution. While the derivation of such an approximation would be quite useful to design a well defined statistical test and to study and compare its performance with existing approaches, our results represent a first necessary step that has its own interest. We notice that we consider the supremum on the whole frequency interval $[0, 1]$ because, compared to a solution where the maximum is over a low number of fixed frequencies, this allows to increase the power of the test in contexts of alternatives for which, under $\mathcal{H}_1$,

$$\nu \to \left| \frac{1}{M} \mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu)) \right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)} - \hat{r}_N(\nu)\, \phi_N(f)\, v_N\, \mathbf{1}_{\alpha>2/3} \right| \tag{1.9}$$

exhibits narrow peaks that would not be visible on a low density frequency grid. We also mention that other statistics could also be considered, e.g. the integral on the frequency domain of function (1.9) or of the square of this function.

We finally remark that the most usual asymptotic regime considered in the context of large random matrices is $M \to +\infty, N \to +\infty$ in such a way that $\frac{M}{N}$ converges towards a non zero constant. In this regime, it is still possible to develop large random matrix-based approaches testing that the components of $\mathbf{y}$ are uncorrelated or not, see e.g. the contribution [69] to be presented below which, under the extra assumption that the components of $\mathbf{y}$ share the same spectral density, is based on a Gaussian approximation of linear spectral statistics of the empirical covariance matrix $\hat{\mathbf{R}}_N$ defined by

$$\hat{\mathbf{R}}_N = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^* \tag{1.10}$$

under $\mathcal{H}_0$. However, when the ratio $\frac{M}{N}$ is small enough, the asymptotic regime considered in the present chapter seems more relevant than the standard large random matrix regime $M \to +\infty, N \to +\infty$, and test statistics that depend on the estimated spectral coherency matrix $\hat{\mathbf{C}}(\nu)$ should provide better performance than functionals of the matrix $\hat{\mathbf{R}}_N$.

1.1.3 On the literature

The problem of testing whether various jointly stationary and jointly Gaussian time series are uncorrelated is an important problem that was extensively addressed in the past. Apart from a few works that will be discussed later, almost all the previous contributions addressed the case where the number M of available time series remains finite as the sample size increases. Two classes of methods were mainly studied. The first class uses lag domain approaches based on the observation that M jointly stationary time series $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ are mutually uncorrelated if and only if for each integer L, the covariance matrix of the ML dimensional vector $\mathbf{y}_n^{(L)}$ defined by

$$\mathbf{y}_n^{(L)} = (y_{1,n}, \ldots, y_{1,n+L-1}, \ldots, y_{M,n}, \ldots, y_{M,n+L-1})^T$$

is block diagonal. The lag domain approach was in particular used in [38] for $M = 2$, and extended and developed in [51], [55], [40], [42], [22] and [41].

The second approach is based on the observation that the M jointly stationary time series $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ are uncorrelated if and only the spectral density matrix $\mathbf{S}(\nu)$ of $\mathbf{y}_n = (y_{1,n}, \ldots, y_{M,n})^T$ is diagonal for each frequency ν, or equivalently, if its spectral coherence matrix $\mathbf{C}(\nu)$ is reduced to $\mathbf{I}_M$ for each ν. [93] is one of the first related contribution. This work was followed by [24], [88], as well as [25].

We now review the existing works devoted to the case where the number M of time series converges towards $+\infty$. The particular context where the observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are i.i.d. and where the ratio $\frac{M}{N}$ converges towards a constant $d \in (0, 1)$ is the most popular. In contrast to the asymptotic regime considered in the present chapter, M and N are of the same order of magnitude. This is because, in this context, the time series are mutually uncorrelated if and only the covariance matrix $\mathbb{E}[\mathbf{y}_n \mathbf{y}_n^*]$ is diagonal. Therefore, it is reasonable to consider test statistics that are functionals of the sample covariance matrix $\hat{\mathbf{R}}_N$ defined by (1.10). In particular, when the observations are Gaussian random vectors, the generalized likelihood ratio test (GLRT) consists in comparing the test statistics $\log\det(\hat{\mathbf{C}}_N)$ to a threshold, where $\hat{\mathbf{C}}_N$ represents the sample autocorrelation matrix. [46] proved that under $\mathcal{H}_0$, the empirical eigenvalue distribution of $\hat{\mathbf{C}}_N$ converges almost surely towards the Marcenko-Pastur distribution $\mu_{MP}^{(d)}$ and therefore, that $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}_N)\right)$ converges towards $\int f d\mu_{MP}^{(d)}$ for each bounded continuous function f. In the Gaussian case, [47] also established a central limit theorem (CLT) for $\log\det(\hat{\mathbf{C}}_N)$ under $\mathcal{H}_0$ using the moment method. In the real Gaussian case, [21] remarked that $\left(\det\hat{\mathbf{C}}_N\right)^{N/2}$ is the product of independent beta distributed random variables. Therefore, $\log\det(\hat{\mathbf{C}}_N)$ appears as the sum of independent random variables, thus deducing the CLT. More recently, in [64] is established a CLT on LSS of $\hat{\mathbf{C}}_N$ in the Gaussian case using large random matrix techniques when the covariance matrix $\mathbb{E}[\mathbf{y}_n \mathbf{y}_n^*]$ is not necessarily diagonal. This allows studying the asymptotic performance of the GLRT under a certain class of alternatives. We also mention that [45] studied the behaviour of $\max_{i,j} |(\hat{\mathbf{C}}_N)_{i,j}|$ under $\mathcal{H}_0$, and established that $\max_{i,j} |(\hat{\mathbf{C}}_N)_{i,j}|$, after recentering and appropriate normalization, converges in distribution towards a Gumbel distribution, which, of course, allows to test the hypothesis $\mathcal{H}_0$. This first contribution was extended later in several works, in particular in [17] who considered the case where the samples $\mathbf{y}_1, \ldots, \mathbf{y}_N$ have some specific correlation pattern. Still, in the asymptotic regime $\frac{M}{N} \to d$, [69] proposed to test hypothesis $\mathcal{H}_0$ when the components of $\mathbf{y}$ share the same spectral density. In this case, the rows of the $M \times N$ matrix $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ are independent and identically distributed under $\mathcal{H}_0$. [69] established a central limit theorem for linear spectral statistics of the empirical covariance matrix $\hat{\mathbf{R}}_N$ defined by (1.10), and used this test statistics to check whether $\mathcal{H}_0$ holds or not. We notice that the results of [69] are valid in the non-Gaussian case.

In our knowledge, no existing work studied the behaviour of linear spectral statistics of the matrix $\hat{\mathbf{C}}(\nu)$ in the asymptotic regime defined in the present chapter. However, we mention that this regime was considered in [11] to solve a completely different problem, i.e. the use of shrinkage in the frequency domain in order to enhance the performance of the spectral density estimate (1.1) when the components of $\mathbf{y}$ are not uncorrelated. We notice that $\frac{B^{3/2}}{N}$ is supposed to converge towards 0 in [11]. When $B = \mathcal{O}(N^\alpha)$, this condition is equivalent to $\alpha < 2/3$, while we rather study situations where $\alpha > 1/2$. We finally mention that our works [77] and [76] also consider the present asymptotic regime and study respectively the behaviour of $\sup_{i<j, \nu \in \mathcal{G}_N} |\hat{\mathbf{C}}_{i,j}(\nu)|$ ($\mathcal{G}_N$ is the set $\{k\frac{B+1}{N}, k = 0, \ldots, \frac{N}{B+1}\}$) and the largest eigenvalues of $\hat{\mathbf{C}}(\nu)$ in the presence of an extra signal, independent from $\mathbf{y}$, and having a low-rank spectral density matrix.

1.1.4 General approach

To simplify the notations, we denote by $\psi_N(f, \nu)$ the statistics defined by

$$\psi_N(f, \nu) = \frac{1}{M} \mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \, \mathrm{d}\mu_{MP}^{(c_N)} - r_N(\nu) \, \phi_N(f) \, v_N \, \mathbf{1}_{\alpha > 2/3}. \tag{1.11}$$

To be able to study the behaviour of $\sup_\nu |\psi_N(f, \nu)|$, we establish exponential concentration inequalities that allow to evaluate $\mathbb{P}(|\psi_N(f, \nu)| > N^\epsilon u_N)$ for each ν as well as $\mathbb{P}(\sup_{\nu \in \mathcal{V}_N} |\psi_N(f, \nu)| > N^\epsilon u_N)$ for some relevant finite discrete grid $\mathcal{V}_N$ of the interval $[0, 1]$. (1.7) is then obtained by using Lipschitz properties of function $\nu \to \psi_N(f, \nu)$.

To evaluate $\mathbb{P}(|\psi_N(f, \nu)| > N^\epsilon u_N)$ for each ν, we use the following approach:

- We first study the behaviour of the modified sample spectral coherency matrix $\tilde{\mathbf{C}}(\nu)$ defined by
$$\tilde{\mathbf{C}}(\nu) = \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} \hat{\mathbf{S}}(\nu) \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}}. \tag{1.12}$$
We notice that $\tilde{\mathbf{C}}(\nu)$ is obtained from $\hat{\mathbf{C}}(\nu)$ by replacing the estimated diagonal matrix $\mathrm{dg}\left(\hat{\mathbf{S}}(\nu)\right)$ by its true value $\mathrm{dg}\left(\mathbf{S}(\nu)\right)$. Using classical results of [13], we establish that for each ν, $\tilde{\mathbf{C}}(\nu)$ can be represented as
$$\tilde{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}(\nu) \tag{1.13}$$
where $\mathbf{X}(\nu)$ is an $M \times (B+1)$ random matrix with $\mathcal{N}_{\mathbb{C}}(0, 1)$ i.i.d. entries, and $\tilde{\mathbf{\Delta}}(\nu)$ is another matrix such that, for any $\epsilon > 0$, there exists $\gamma > 0$, independent from ν, such that for each

large enough $N \in \mathbb{N}$:

$$\mathbb{P}\left[\|\tilde{\boldsymbol{\Delta}}(\nu)\| > N^{\epsilon}\frac{B}{N}\right] \leq \exp -N^{\gamma}.$$

We deduce from (1.13) that $\hat{\mathbf{C}}(\nu)$ can be written as

$$\hat{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \boldsymbol{\Delta}(\nu) \tag{1.14}$$

where $\boldsymbol{\Delta}(\nu)$ satisfies the concentration inequality

$$\mathbb{P}\left[\|\boldsymbol{\Delta}(\nu)\| > N^{\epsilon}\left(\frac{1}{\sqrt{B}} + \frac{B}{N}\right)\right] \leq \exp -N^{\gamma}$$

for each $\epsilon > 0$, where γ does not depend on ν. Using (1.13) and (1.14), we establish that the eigenvalues of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$ are localized with high probability in a neighbourhood of the support of the Marcenko-Pastur distribution $\mu_{MP}^{(c)}$.

$\tilde{\mathbf{C}}(\nu)$ appears as a useful intermediate matrix because the study of $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)}$ is based on the evaluation of each term of the following decomposition:

$$\begin{aligned}
\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)} &= \frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) + \\
&\quad \frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right] + \\
&\quad \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] + \\
&\quad \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)}.
\end{aligned} \tag{1.15}$$

Using the above-mentioned results related to the localization of the eigenvalues of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$, we also argue that it is sufficient to do so when f is compactly supported.

- The term $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)$ is studied using the Helffer-Sjöstrand formula which allows, in a certain sense, to be back to the study of $\frac{1}{M}\mathrm{Tr}\left(\hat{\mathbf{Q}}(z) - \tilde{\mathbf{Q}}(z)\right)$ for $z \in \mathbb{C}^+$, where $\hat{\mathbf{Q}}(z)$ and $\tilde{\mathbf{Q}}(z)$ represent the resolvents of matrices $\hat{\mathbf{C}}(\nu)$ and $\tilde{\mathbf{C}}(\nu)$ (see below for a formal definition). Using (1.13) and (1.14), we express $\frac{1}{M}\mathrm{Tr}\left(\hat{\mathbf{Q}}(z) - \tilde{\mathbf{Q}}(z)\right)$ in terms of the resolvent $\mathbf{Q}(z)$ of the matrix $\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}$. As the matrix $\mathbf{X}(\nu)$ is Gaussian, it is possible to use standard Gaussian tools (Poincaré-Nash inequality and the integration by parts formula) to have a good understanding of the behaviour of $\mathbf{Q}(z)$, and to prove that for each $\epsilon > 0$, there exists γ independent from ν such that

$$\begin{aligned}
\mathbb{P}\Bigg(\Bigg|\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \\
\left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s_m''(\nu)}{s_m(\nu)}\right)\tilde{\phi}_N(f)\, v_N\, \mathbf{1}_{\alpha>2/3}\Bigg| > N^{\epsilon}u_N\Bigg) \leq \exp -N^{\gamma}
\end{aligned}$$

where $\tilde{\phi}_N(f)$ is a deterministic term defined as the action of f on a compactly supported distribution $\tilde{D}_N$ depending on $\mu_{MP}^{(c_N)}$.

- Using a standard Gaussian concentration inequality as well as the structure of the matrix

$\tilde{\mathbf{C}}(\nu)$, we obtain that for each $\epsilon > 0$, there exists γ independent from ν such that

$$\mathbb{P}\left[\left|\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right]\right| > N^{\epsilon}\frac{1}{B}\right] \leq \exp -N^{\gamma} \tag{1.16}$$

for each N large enough.

- We then analyse the deterministic term $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right]$ using the Helffer-Sjöstrand formula. We first show that for each $z \in \mathbb{C}^+$, $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}(\tilde{\mathbf{Q}}(z) - \mathbf{Q}(z))\right]$ is a $\mathcal{O}(\frac{B}{N})^2$ term, a non obvious result because the relation (1.13) just leads to the conclusion that the above term is $\mathcal{O}(\frac{B}{N})$. Moreover, using long and very tedious Gaussian calculations, we obtain that if $\alpha > \frac{2}{3}$, it holds that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}(\tilde{\mathbf{Q}}(z) - \mathbf{Q}(z))\right] = -\left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}\right)\tilde{p}_N(z)\, v_N + \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 p_N(z)\, v_N + \mathcal{O}\left(\frac{B}{N}\right)^3$$

where p_N and $\tilde{p}_N$ are the Stieltjes transforms of the compactly supported distributions D_N and $\tilde{D}_N$ introduced previously. This immediately implies that if $\alpha \leq \frac{2}{3}$, then

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] = \mathcal{O}\left(\frac{B}{N}\right)^2 = o\left(\frac{1}{B}\right) = o(u_N)$$

while if $\alpha > \frac{2}{3}$, then,

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] = -\left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}\right)\tilde{\phi}_N(f)\, v_N + \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 \phi_N(f)\, v_N + \mathcal{O}(u_N)$$

because $(\frac{B}{N})^3 \ll u_N$ if $2/3 < \alpha \leq 4/5$ and $(\frac{B}{N})^3$ is equivalent to u_N if $\alpha > 4/5$.

- Finally, classical results imply that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\, d\mu_{MP}^{(c_N)} = \mathcal{O}\left(\frac{1}{B^2}\right) = o(u_N).$$

Gathering the above approximations and using the Lipschitz properties of function $\nu \to \psi(f,\nu)$, we finally obtain (1.7).

We also indicate how the use of lag window estimators of the spectral densities $(s_m)_{m=1,\ldots,M}$ allows to design an estimator $\hat{r}_N(\nu)$ of $r_N(\nu)$ defined by (1.8) for which the rate of convergence towards 0 of the statistics $\hat{\psi}_N(f,\nu)$ obtained by replacing $r_N(\nu)$ by $\hat{r}_N(\nu)$ in Eq. (1.11) is still u_N. In particular, we establish that for each $\epsilon > 0$, $\mathbb{P}\left(\sup_\nu |\hat{\psi}_N(f,\nu)| > N^{\epsilon} u_N\right)$ converges towards 0 exponentially.

1.1.5 Assumptions and general notations

Assumption 1.1. *For each $m \geq 1$, $(y_{m,n})_{n\in\mathbb{Z}}$ is a zero mean stationary complex Gaussian time series, ie.*

1. $\mathbb{E}[y_{m,n}] = 0$ *for any* $m \geq 1$ *and any* $n \in \mathbb{Z}$
2. *every finite linear combination x of the random variables $(y_{m,n})_{n\in\mathbb{Z}}$ is a $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ distributed random variable for some σ^2, i.e.* $\mathrm{Re}(x)$ *and* $\mathrm{Im}(x)$ *are independent and $\mathcal{N}(0,\sigma^2/2)$ distributed.*

Assumption 1.2. *If $m_1 \neq m_2$, then the scalar time series $(y_{m_1,n})_{n\in\mathbb{Z}}$ and $(y_{m_2,n})_{n\in\mathbb{Z}}$ are independent.*

We now formulate the following assumptions on the growth rate of the quantities N, M, B:

Assumption 1.3.

$$B, M = \mathcal{O}(N^\alpha) \text{ where } \frac{1}{2} < \alpha < 1, \quad \frac{M}{B+1} = c_N, \quad c_N \xrightarrow[N\to+\infty]{} c \in (0,1).$$

As $M = M(N)$ converges towards $+\infty$, we assume that an infinite sequence $(y_{1,n})_{n\in\mathbb{Z}}, (y_{2,n})_{n\in\mathbb{Z}}, \ldots, (y_{k,n})_{n\in\mathbb{Z}}, \ldots$ of mutually independent zero mean complex Gaussian time series is given.

We denote by $(s_m)_{m\geq 1}$ the corresponding sequence of spectral densities (i.e. s_m coincides with the spectral density of the times series $(y_{m,n})_{n\in\mathbb{Z}}$). For each $m \geq 1$, we denote by $r_m = (r_{m,u})_{u\in\mathbb{Z}}$ the autocovariance sequence of $(y_{m,n})_{n\in\mathbb{Z}}$, i.e. $r_{m,u} = \mathbb{E}[y_{m,n+u} y^*_{m,n}]$. We formulate the following assumptions on $(s_m)_{m\geq 1}$ and $(r_m)_{m\geq 1}$:

Assumption 1.4. *The time series $((y_{m,n})_{n\in\mathbb{Z}})_{m\geq 1}$ are such that:*

$$\inf_{m\geq 1} \inf_{\nu\in[0,1]} |s_m(\nu)| > 0 \tag{1.17}$$

and

$$\sup_{m\geq 1} \sum_{u\in\mathbb{Z}} (1+|u|)^{\gamma_0} |r_{m,u}| < +\infty \tag{1.18}$$

where $\gamma_0 \geq 3$. Assumption (1.18) of course implies that the spectral densities $(s_m)_{m\geq 1}$ are $\mathcal{C}^3$ and that

$$\sup_{m\geq 1} \sup_{\nu\in[0,1]} |s_m^{(i)}(\nu)| < +\infty$$

for $i = 0,1,2,3$ ($s_m^{(i)}$ represents the derivative of order i of s_m). We notice that (1.18) holds as soon as we have

$$\sup_{m\geq 1} |r_{m,u}| \leq \frac{C}{|u|^{1+\gamma_0+\delta}}$$

for each $u \neq 0$ as well as $\sup_{m\geq 1} |r_{m,0}| < \infty$ ($C > 0$ and $\delta > 0$ represent constants). If z represents the backward shift operator, a simple example of time series satisfying Assumption 1.4 is to consider an ARMA time series generated as

$$y_{m,n} = [h_m(z)]\epsilon_{m,n}$$

where $((\epsilon_{m,n})_{n\in\mathbb{Z}})_{m\geq 1}$ are mutually independent i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ sequences, and where $h_m(z) = \frac{b_m(z)}{a_m(z)}$, a_m and b_m being 2 polynomials having no pole or zero in the closed unit disk $\overline{\mathbb{D}}$. Moreover, $\sup_{m\geq 1} \max(\deg(a_m), \deg(b_m)) < +\infty$, and if $(z_{k,m})_{k=1,\ldots,\deg(b_m)}$ and $(p_{k,m})_{k=1,\ldots,\deg(a_m)}$ are the zeros of b_m and a_m, then we should have

$$\inf_{m\geq 1}\inf_k \mathrm{dist}(z_{k,m},\overline{\mathbb{D}}) > 0, \ \inf_{m\geq 1}\inf_k \mathrm{dist}(p_{k,m},\overline{\mathbb{D}}) > 0$$
$$\sup_{m\geq 1}\sup_k |z_{k,m}| < +\infty, \ \sup_{m\geq 1}\sup_k |p_{k,m}| < +\infty.$$

It is easy to check that (1.18) holds for each $\gamma_0 > 0$, and that (1.17) is verified as well.

Notations. A zero mean complex valued random vector $\mathbf{y}$ is said to be $\mathcal{N}_{\mathbb{C}}(0,\mathbf{\Sigma})$ distributed if $\mathbb{E}(\mathbf{y}\mathbf{y}^*) = \mathbf{\Sigma}$ and if each linear combination x of the entries of $\mathbf{y}$ is a complex Gaussian random variable, i.e. $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ are independent Gaussian random variables sharing the same variance. If x is a random variable, we denote by x° the random variable defined by

$$x^\circ = x - \mathbb{E}[x]. \tag{1.19}$$

If $\mathbf{A}$ is a $P \times Q$ matrix, $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ denote its spectral norm and Frobenius norm respectively. If $P = Q$ and $\mathbf{A}$ is Hermitian, $\lambda_1(\mathbf{A}) \geq \ldots \geq \lambda_P(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$. The spectrum of $\mathbf{A}$, which is here the set of its eigenvalues $(\lambda_k(\mathbf{A}))_{k=1,\ldots,P}$, is denoted by $\sigma(\mathbf{A})$. For $\mathbf{A}$ and $\mathbf{B}$ square Hermitian matrices, if all the eigenvalues of $\mathbf{A}-\mathbf{B}$ are non negative, we write $\mathbf{A} \geq \mathbf{B}$. We define $\mathrm{Re}\,\mathbf{A} = (\mathbf{A}+\mathbf{A}^*)/2$ and $\mathrm{Im}\,\mathbf{A} = (\mathbf{A}-\mathbf{A}^*)/2$ where $\mathbf{A}^*$ is the conjugate transpose of the matrix $\mathbf{A}$.

$\mathcal{C}^p$ represents the set of all real-valued functions defined on $\mathbb{R}$ whose first p derivatives exist and are continuous, and $\mathcal{C}^p_c$ is the set of all compactly supported functions of $\mathcal{C}^p$.

We recall that $\mathbf{S}(\nu)$ represents the $M \times M$ diagonal matrix $\mathbf{S}(\nu) = \mathrm{dg}(s_1(\nu),\ldots,s_M(\nu))$. We notice that $\mathbf{S}$ depends on M, thus on N (through $M := M(N)$), but we often omit to mention the corresponding dependency in order to simplify the notations. In the following, we will denote by $\mathbf{y}_m$ the N–dimensional vector $\mathbf{y}_m = (y_{m,1},\ldots,y_{m,N})^T$.

A nice constant is a positive a constant that does not depend on the frequency ν, the time series index m, the complex variable z of the various resolvents and Stieltjes transforms used throughout the chapter, as well as on the dimensions B, M and N. A nice polynomial is a polynomial whose degree and coefficients are nice constants. If $z \in \mathbb{C}^+$ and if P_1 and P_2 are two nice polynomials, terms such as $P_1(z)P_2(\frac{1}{\mathrm{Im}z})$ play an important role in the following. C and $C(z)$ will represent a generic notation for respectively a nice constant and a term $P_1(z)P_2(\frac{1}{\mathrm{Im}z})$, and the values of C and $C(z)$ may change from one line to the other.

If $(a_N)_{N\geq 1}$ and $(b_N)_{N\geq 1}$ are two sequences of positive real numbers, we write $a_N << b_N$ if $\frac{a_N}{b_N} \to 0$ when $N \to +\infty$.

We also recall how a function can be applied to Hermitian matrices. For an $M \times M$ Hermitian matrix $\mathbf{A}$ with spectral decomposition $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ where $\mathbf{\Lambda} = \mathrm{dg}(\lambda_m, m = 1,\ldots,M)$ and the $(\lambda_m)_{m=1,\ldots,M}$ are the real eigenvalues of $\mathbf{A}$, then for any function f defined on $\mathbb{R}$, we define $f(\mathbf{A})$ as:

$$f(\mathbf{A}) = \mathbf{U}\begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_M)\end{pmatrix}\mathbf{U}^*$$

$\mathbb{C}^+$ is the upper half-plane of $\mathbb{C}$, i.e. the set of all complex numbers z for which $\mathrm{Im}\,z > 0$.

For μ a probability measure, its Stieltjes transform s_μ is the function defined on $\mathbb{C} \setminus \mathrm{Supp}\mu$ as

$$s_\mu(z) = \int \frac{\mathrm{d}\mu(\lambda)}{\lambda - z}. \tag{1.20}$$

We recall that

$$|s_\mu(z)| \leq \frac{1}{\mathrm{Im}\, z} \tag{1.21}$$

for each $z \in \mathbb{C}^+$. Moreover, if μ is carried by $\mathbb{R}^+$, then for any $a > 0$, the function $-\frac{1}{z(1+as_\mu(z))}$ is also the Stieltjes transform of a probability distribution carried by $\mathbb{R}^+$, a property which implies that

$$\left|\frac{1}{1 + as_\mu(z)}\right| \leq \frac{|z|}{\mathrm{Im} z} \tag{1.22}$$

for each $z \in \mathbb{C}^+$ (see [35], Proposition 5-1, item 4).

If $\lambda_1, \ldots, \lambda_M$ denote the eigenvalues of a Hermitian matrix $\mathbf{A}$ and if $\mu := \frac{1}{M}\sum_{i=1}^M \delta_{\lambda_i}$ denotes the empirical eigenvalue distribution of $\mathbf{A}$, then we have the following relation:

$$s_\mu(z) = \frac{1}{M}\mathrm{tr}\ \mathbf{Q}_\mathbf{A}(z)$$

where $\mathbf{Q}_\mathbf{A}(z)$ represents the resolvent of $\mathbf{A}$ defined by

$$\mathbf{Q}_\mathbf{A}(z) = (\mathbf{A} - z\mathbf{I}_M)^{-1}. \tag{1.23}$$

We finally mention the following useful control for the norm $\mathbf{Q}_\mathbf{A}$. For each $z \in \mathbb{C}^+$, we have

$$\|\mathbf{Q}_\mathbf{A}\| \leq \frac{1}{\mathrm{Im}\, z}. \tag{1.24}$$

1.1.6 Overview of the chapter

We first recall in Section 1.2 useful technical tools: in Paragraph 1.2.1, the concept of stochastic domination adapted from [27] which allows to considerably simplify the exposition of the following results, in Paragraph 1.2.2 some useful properties of the extreme eigenvalues and of the resolvent of large Wishart matrices, two well-known Gaussian concentration inequalities expressed using the stochastic domination framework in Paragraphs 1.2.3 and 1.2.4, and the Helffer-Sjöstrand formula in Paragraph 1.2.5. We establish in Section 1.3 the stochastic representations (1.13) and (1.14) of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$. In Section 1.4, we prove for each ν the concentration of $|\psi_N(f,\nu)|$ defined by (1.11), and indicate how it is possible to estimate the term $r_N(\nu)$ in order to keep equal to u_N the rate of convergence of the statistics $\hat{\psi}_N(f,\nu)$ obtained by replacing $r_N(\nu)$ by $\hat{r}_N(\nu)$ in (1.11). In Section 1.5, we establish Lipschitz properties for the functions $\nu \to \psi_N(f,\nu)$ and $\nu \to \hat{\psi}_N(f,\nu)$ that allow to establish the concentration of $\sup_\nu |\psi_N(f,\nu)|$ and $\sup_\nu |\hat{\psi}_N(f,\nu)|$. We finally provide in Section 1.6 some numerical simulations that support our results.

1.2 Useful technical tools

1.2.1 Stochastic domination

We now present the concept of stochastic domination introduced in [27]. A nice introduction to this tool can also be found in the lecture notes [8].

Definition 1.1. ***Stochastic Domination.*** *Let*

$$X = (X^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}), \quad Y = (Y^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)})$$

be two families of nonnegative random variables, where $U^{(N)}$ is a set that may possibly depend on N. We say that X is stochastically dominated by Y if for all (small) $\epsilon > 0$, there exists some $\gamma > 0$ (which of course depends on ϵ) such that:

$$\mathbb{P}\left[X^{(N)}(u) > N^{\epsilon} Y^{(N)}(u)\right] \leq \exp -N^{\gamma}$$

for each $u \in U^{(N)}$ and for each large enough $N > N_0(\epsilon)$, where $N_0(\epsilon)$ is independent of u, or equivalently

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^{\epsilon} Y^{(N)}(u)\right] \leq \exp -N^{\gamma}. \tag{1.25}$$

for each large enough $N > N_0(\epsilon)$. If X is stochastically dominated by Y we use the notation $X^{(N)}(u) \prec Y^{(N)}(u)$. To simplify the notations, we will very often denote $X^{(N)} \prec Y^{(N)}$ or $X \prec Y$ when the context will be clear enough. Moreover, if for some complex valued family X we have $|X| \prec Y$ we also write $X = \mathcal{O}_{\prec}(Y)$.

Finally, we say that a family of events $\Xi = \Xi^{(N)}(u)$ holds with exponentially high (small) probability if there exist N_0 and $\gamma > 0$ such that for $N \geq N_0$, $\mathbb{P}[\Xi_N(u)] > 1 - \exp -N^{\gamma}$ ($\mathbb{P}[\Xi_N(u)] < \exp -N^{\gamma}$) for each $u \in U^{(N)}$.

Remark 1.1. *Suppose $(X_N)_{N\in\mathbb{N}}$ is a sequence of positive random variables, satisfying $X_N \prec a_N N^{\epsilon}$ for any $\epsilon > 0$ for some positive real numbers sequence $(a_N)_{N\in\mathbb{N}}$. It turns out that this precisely means that $X_N \prec a_N$. Indeed, consider an arbitrary $\epsilon' > 0$. By the stochastic domination property of X_N, one can take ϵ such that $0 < \epsilon < \epsilon'$ and write*

$$\mathbb{P}\left[X_N > a_N \times N^{\epsilon'}\right] \leq \mathbb{P}\left[X_N > a_N \times N^{\epsilon} \times \underbrace{N^{\epsilon'-\epsilon}}_{\gg 1}\right] \leq \mathbb{P}\left[X_N > a_N \times N^{\epsilon}\right]$$

which goes to zero exponentially since $X_N \prec a_N N^{\epsilon}$ for the ϵ chosen. This argument will be used in the proof of Lemma 1.7.

Lemma 1.1. *Take four families of non negative random variables X_1, X_2, Y_1 and Y_2 defined as in Definition 1.1. Then the following holds:*

$$X_1 \prec Y_1 \text{ and } X_2 \prec Y_2 \implies X_1 + X_2 \prec Y_1 + Y_2 \text{ and } X_1 X_2 \prec Y_1 Y_2.$$

We omit the proof of this lemma.

Remark 1.2. *Note that Definition 1.1 is slightly different from the original one [27] which states that the left hand side of (1.25) should be bounded by a quantity of order N^{-D} for any finite $D > 0$. In the present chapter, all the random variables are Gaussian, and exponential concentration rates can be achieved.*

1.2.2 Properties of the eigenvalues and of the resolvent of large Wishart matrices

In this chapter we will at multiple occasion use properties of the eigenvalues of matrices $\frac{\mathbf{X}_N \mathbf{X}_N^*}{B+1}$ where $\mathbf{X}_N$ is an $M \times (B+1)$ complex Gaussian matrix with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries when $M = M(N)$ and $B = B(N)$ follow Assumption 1.3.

Concentration of the largest and the smallest eigenvalues

We first recall concentration results of the largest and smallest eigenvalue of $\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}$ due to [34]. We have for any $\epsilon > 0$

$$\mathbb{P}\left[\lambda_M\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) < (1-\sqrt{c})^2 - \epsilon\right] \leq (B+1)\exp -C(B+1)\epsilon^2 \tag{1.26}$$

$$\mathbb{P}\left[\lambda_1\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) > (1+\sqrt{c})^2 + \epsilon\right] \leq (B+1)\exp -C(B+1)\epsilon^2 \tag{1.27}$$

for some nice constant C.

Consider for $\epsilon > 0$, the ϵ–expansion of the support of the Marchenko-Pastur distribution $\mu_{MP}^{(c)}$:

$$\mathrm{Supp}\mu_{MP}^{(c)} + \epsilon := \left[(1-\sqrt{c})^2 - \epsilon, (1+\sqrt{c})^2 + \epsilon\right]$$

and the event:

$$\Lambda_{N,\epsilon} = \left\{\sigma\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) \subset \mathrm{Supp}\mu_{MP}^{(c)} + \epsilon\right\}. \tag{1.28}$$

It is clear that using (1.26) and (1.27), $\Lambda_{N,\epsilon}$ holds with exponentially high probability for any $\epsilon > 0$. This will be of high importance in the following since it will enable us to work on events of exponentially high probability where the norm of $\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}$ and the norm of its inverse are bounded.

Finally, the following (weaker) statement is a simple consequence of the equations (1.26) and (1.27), which will sometimes be enough in the following:

$$\lambda_1\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) + \frac{1}{\lambda_M\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right)} \prec 1. \tag{1.29}$$

We finally notice that if we consider a family $\mathbf{X}_N(u) \in \mathbb{C}^{M\times(B+1)}$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, $u \in U^{(N)}$, where $U^{(N)}$ is a certain set possibly depending on N, then (1.26) and (1.27) hold for each $u \in U^{(N)}$ because the constant C in (1.26) and (1.27) is universal. This implies that the stochastic domination (1.29) is still satisfied by the family $\mathbf{X}_N(u)$, $u \in U^{(N)}$. Moreover, the family of events $\Lambda_{N,\epsilon}(u)$ defined by (1.28) when $\mathbf{X}_N$ is replaced by $\mathbf{X}_N(u)$ still holds with exponentially high probability.

Asymptotic behaviour of the resolvent of $\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}$

We next review known results related to the asymptotic behaviour of the resolvent $\mathbf{Q}_N(z)$ of matrix $\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}$ that can be deduced from standard Gaussian tools. The Poincaré-Nash inequality (see e.g. [70, Proposition 2.1.6] in the Gaussian real case and Eq. (18) in [36] in the complex Gaussian case) implies immediately that the following Lemma holds.

Lemma 1.2. *Consider deterministic $M \times M$ and $(B+1)\times(B+1)$ matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}$. Then, it holds that*

$$\mathrm{Var}\frac{1}{M}\mathrm{tr}\ \mathbf{A}\mathbf{Q}_N^i(z) \leq \frac{C(z)}{M^2}\frac{1}{M}\mathrm{tr}\ \mathbf{A}\mathbf{A}^* \tag{1.30}$$

$$\mathrm{Var}\frac{1}{M}\mathrm{tr}\ \left(\frac{\mathbf{X}\tilde{\mathbf{A}}\mathbf{X}^*}{B+1}\mathbf{Q}_N^i(z)\right) \leq \frac{C(z)}{M^2}\frac{1}{B+1}\mathrm{tr}\ \tilde{\mathbf{A}}\tilde{\mathbf{A}}^* \tag{1.31}$$

for $i = 1, 2$

We recall that $C(z)$ represents a generic notation for $P_1(z)P_2(\frac{1}{\mathrm{Im}z})$ where P_1 and P_2 are nice polynomials.

The integration by parts formula states that if $h(\mathbf{X}, \mathbf{X}^*)$ is a $\mathcal{C}^1$ function of the entries of $\mathbf{X}$ and $\mathbf{X}^*$ with polynomially bounded first derivatives, then, it holds that

$$\mathbb{E}(X_{ij} h(\mathbf{X}, \mathbf{X}^*)) = \mathbb{E}|X_{ij}|^2 \mathbb{E}\left[\frac{\partial h}{\partial \overline{X}_{ij}}(\mathbf{X}, \mathbf{X}^*)\right]. \tag{1.32}$$

(1.32), in conjunction with the Poincaré-Nash inequality, allows to evaluate easily the asymptotic behaviour of the entries of $\mathbb{E}(\mathbf{Q}_N(z))$ (see e.g. [70]). We first notice that properties of the distribution of the matrix $\mathbf{X}_N$ immediately imply that $\mathbb{E}(\mathbf{Q}_N(z))$ is reduced to $\beta_N(z)\mathbf{I}_M$ where $\beta_N(z)$ coincides with $\mathbb{E}(\mathbf{Q}_{m,m}(z))$ for each m. Then, it holds that

$$\beta_N(z) = t_N(z) + \epsilon_N(z) \tag{1.33}$$

where the error term $\epsilon_N(z)$ satifies $|\epsilon_N(z)| \leq \frac{C(z)}{M^2}$ and where $t_N(z)$ is the Stieltjes transform of the Marcenko-Pastur distribution $\mu_{MP}^{(c_N)}$. In other words, $t_N(z)$ is the unique Stieltjes transform satisfying the equation

$$t_N(z) = \frac{1}{-z + \frac{1}{1+c_N t_N(z)}}. \tag{1.34}$$

It is also convenient to define $\tilde{t}_N(z)$ by

$$\tilde{t}_N(z) = -\frac{1}{z(1 + c_N t_N(z))} \tag{1.35}$$

so that $t_N(z)$ is also given by

$$t_N(z) = -\frac{1}{z(1 + \tilde{t}_N(z))}. \tag{1.36}$$

It is well-known that $\tilde{t}_N(z)$ is the Stieltjes transform of the probability distribution $c_N \mu_{MP}^{(c_N)} + (1 - c_N)\delta_0$.

We finally mention that $\mathbb{E}(\mathbf{Q}_N'(z)) = \mathbb{E}(\mathbf{Q}_N^2(z)) = \beta_N'(z)\mathbf{I}_M$ (where $'$ stands for the derivative w.r.t. z), and that $\epsilon_N'(z) = \beta_N'(z) - t_N'(z)$ still satisfies

$$|\epsilon_N'(z)| \leq \frac{C(z)}{M^2}. \tag{1.37}$$

1.2.3 Concentration of functionals of Gaussian entries

It is well-known (see e.g. [89, Th. 2.1.12]) that for any 1-Lipschitz real valued function f defined on $\mathbb{R}^N$ and any N–dimensional random variable $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I}_N)$, there exists a universal constant C such that:

$$\mathbb{P}\left[|f(\mathbf{X}) - \mathbb{E}f(\mathbf{X})| > t\right] \leq C \exp -Ct^2. \tag{1.38}$$

This inequality is still valid when $\mathbf{X} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$: in this context, $f(\mathbf{X})$ is replaced by a real-valued function $f(\mathbf{X}, \mathbf{X}^*)$ depending on the entries of $\mathbf{X}$ and $\mathbf{X}^*$. $f(\mathbf{X}, \mathbf{X}^*)$ can of course be written as $f(\mathbf{X}, \mathbf{X}^*) = \tilde{f}(\sqrt{2}\mathrm{Re}(\mathbf{X}), \sqrt{2}\mathrm{Im}(\mathbf{X}))$ for some function $\tilde{f}$ defined on $\mathbb{R}^{2N}$. As $(\sqrt{2}\mathrm{Re}(\mathbf{X}), \sqrt{2}\mathrm{Im}(\mathbf{X}))$ is $\mathcal{N}(0, \mathbf{I}_{2N})$ distributed, the concentration inequality is still valid for $f(\mathbf{X}, \mathbf{X}^*) = \tilde{f}(\sqrt{2}\mathrm{Re}(\mathbf{X}), \sqrt{2}\mathrm{Im}(\mathbf{X}))$. We just finally mention that f, considered as a function of $(\mathbf{X}, \mathbf{X}^*)$, and $\tilde{f}$ have Lipschitz constants that are of the same order of magnitude. More precisely, if we define the differential operators $\frac{\partial}{\partial z}$ and $\frac{\partial}{\partial \bar{z}}$ by

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}, \quad \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}$$

we can verify immediately that

$$\sum_{i=1}^{N}\left(\left|\frac{\partial f}{\partial X_i}\right|^2 + \left|\frac{\partial f}{\partial X_i^*}\right|^2\right) = \| \left(\nabla f\right)_{(\mathbf{X},\mathbf{X}^*)} \|^2 = 4\| \left(\nabla \tilde{f}\right)_{(\sqrt{2}\mathrm{Re}(\mathbf{X}),\sqrt{2}\mathrm{Im}(\mathbf{X}))} \|^2.$$

Within the stochastic domination framework, the concentration inequality (1.38) implies that for a family $\mathbf{X}_N(u) \sim \mathcal{N}(0, \mathbf{I}_N)$ for $u \in U^{(N)}$:

$$|f(\mathbf{X}_N(u)) - \mathbb{E}f(\mathbf{X}_N(u))| \prec 1$$

The proof is immediate: consider $\epsilon > 0$ and obtain that

$$\mathbb{P}[|f(\mathbf{X}_N(u)) - \mathbb{E}f(\mathbf{X}_N(u))| > N^\epsilon] \leq C \exp -CN^{2\epsilon}$$

for each u as expected. This result can easily be extended in the complex case, ie. when $\mathbf{X}_N(u) \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$.

1.2.4 Hanson-Wright inequality

The Hanson-Wright inequality [79] is useful to control deviations of a quadratic form from its expectation. While it is proved in the real case in [79], it can easily be understood that it can be extended in the complex case as follows: let $\mathbf{X} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$ and $\mathbf{A} \in \mathbb{C}^{N\times N}$. Then

$$\mathbb{P}[|\mathbf{X}^*\mathbf{A}\mathbf{X} - \mathbb{E}\mathbf{X}^*\mathbf{A}\mathbf{X}| > t] \leq 2\exp -C\min\left(\frac{t^2}{\|\mathbf{A}\|_F^2}, \frac{t}{\|\mathbf{A}\|}\right). \tag{1.39}$$

We now write (1.39) in the stochastic domination framework. Consider a family of independent $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables $(X_n(u))_{n=1,\ldots,N}$ where $u \in U^{(N)}$ and a sequence of $N\times N$ matrices $\mathbf{A}_N(u)$ that possibly depend on u. Take $\epsilon > 0$ and $t = N^\epsilon \|\mathbf{A}_N(u)\|_F$. Since $\|\mathbf{A}_N(u)\| > 0$, $\|\mathbf{A}_N(u)\|_F > 0$, and $\|\mathbf{A}_N(u)\| \leq \|\mathbf{A}_N(u)\|_F$:

$$\begin{aligned}\min\left(\frac{t}{\|\mathbf{A}_N(u)\|}, \frac{t^2}{\|\mathbf{A}_N(u)\|_F^2}\right) &= \min\left(N^\epsilon\frac{\|\mathbf{A}_N(u)\|_F}{\|\mathbf{A}_N(u)\|}, N^{2\epsilon}\frac{\|\mathbf{A}_N(u)\|_F^2}{\|\mathbf{A}_N(u)\|_F^2}\right)\\ &\geq \min(N^\epsilon, N^{2\epsilon}) = N^\epsilon.\end{aligned}$$

Denote $\mathbf{X}_N(u) = (X_1(u), \ldots, X_N(u))^T$. For any $u \in U^{(N)}$, it holds that:

$$\mathbb{P}\left[|\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u) - \mathbb{E}\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u)| > N^\epsilon\|\mathbf{A}_N(u)\|_F\right] \leq 2\exp -CN^\epsilon. \tag{1.40}$$

We can therefore rewrite (1.40) as the following stochastic domination:

$$|\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u) - \mathbb{E}\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u)| \prec \|\mathbf{A}_N(u)\|_F. \tag{1.41}$$

1.2.5 Helffer-Sjöstrand formula

If μ is a probability measure, the Helffer-Sjöstrand formula can be seen as an alternative to the Stieltjes inversion formula that allows to express $\int f d\mu$ in terms of the Stieltjes transform $s_\mu(z)$ of μ (see (1.20)) when f is a regular enough compactly supported function. In order to introduce this tool, we consider a class $\mathcal{C}^{k+1}$ compactly supported function f for a certain integer k, and denote by $\Phi_k(f) : \mathbb{C} \to \mathbb{C}$ the function defined on $\mathbb{C}$ by

$$\Phi_k(f)(x+iy) = \sum_{l=0}^{k} \frac{(iy)^l}{l!} f^{(l)}(x)\rho(y)$$

where $\rho : \mathbb{R} \to \mathbb{R}^+$ is smooth, compactly supported, with value 1 in a neighbourhood of 0. Function $\Phi_k(f)$ coincides with f on the real line and extends it to the complex plane. Let $\bar{\partial} = \partial_x + i\partial_y$. It is well-known that

$$\bar{\partial}\Phi_k(f)(x+iy) = \frac{(iy)^k}{k!} f^{(k+1)}(x) \tag{1.42}$$

(a proof of this result can be found in [23] or [39]) if y belongs to the neighbourhood of 0 in which ρ is equal to 1. The Helffer-Sjöstrand formula can be written as

$$\int f \,\mathrm{d}\mu = \frac{1}{\pi}\mathrm{Re} \int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(z) s_\mu(z) \,\mathrm{d}x \,\mathrm{d}y. \tag{1.43}$$

In order to understand why the integral at the right hand side of (1.43) is well defined, we take, to fix the ideas, $\rho \in \mathcal{C}^\infty$ such that $\rho(y) = 1$ for $|y| \leq 1$ and $\rho(y) = 0$ for $|y| > 2$, and denote by $[a_1, a_2]$ an interval containing the support of f. Then, it appears that the integral on $\mathbb{C}^+$ is in fact over the compact set $\mathcal{D} = \{x + iy : x \in [a_1, a_2], y \in [0, 2]\}$. Moreover, as $|\mathbf{s}_\mu(z)| \leq \frac{1}{y}$ if $z \in \mathcal{D}$ (see (1.21)), (1.42) for $k = 1$ leads to the conclusion that

$$|\bar{\partial}\Phi_k(f)(z)s_\mu(z)| \leq C$$

for $z \in \{x + iy \in \mathcal{D}, y \leq 1\}$. Therefore, the right hand side of (1.43) is well defined.

We finally mention that the Helffer-Sjöstrand formula continues to remain valid for any compactly supported distribution D (see e.g. [59], section 9). The Stieltjes transform of D, denoted by $s_D(z)$, is defined for each $z \in \mathbb{C}^+$ as the action of the function $\lambda \to \frac{1}{\lambda - z}$ on D, i.e. $s_D(z) =< D, \frac{1}{\lambda - z} >$, and satisfies

$$|s_D(z)| \leq C\left(1 + \frac{1}{(\mathrm{Im} z)^{n_0}}\right)$$

for each $z \in \mathbb{C}^+$ where n_0 is related to the order of the distribution. We refer the reader to [16] (Theorem 4.3) and the references therein for more details on Stieltjes transforms of distributions. Then, if f is a $\mathcal{C}^\infty$ function supported by $[a_1, a_2]$, $< D, f >$ is given by

$$< D, f >= \frac{1}{\pi}\mathrm{Re} \int_{\mathcal{D}} \bar{\partial}\Phi_k(f)(z) s_D(z) \,\mathrm{d}x \,\mathrm{d}y \tag{1.44}$$

for $k \geq n_0$. We also recall that an alternative expression for $< D, f >$ is given by the Stieltjes inversion formula, also valid for distributions, i.e.

$$< D, f >= \frac{1}{\pi} \lim_{y \to 0} \int_{a_1}^{a_2} f(\lambda) \,\mathrm{Im} s_D(\lambda + iy) \, d\lambda. \tag{1.45}$$

1.3 Stochastic representations of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$

The first step is to show that $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$ can be approximated by the sample covariance matrix of a sequence of i.i.d. Gaussian random vectors, and to control the order of magnitude of the corresponding errors. This is the objective of the following result.

Theorem 1.1. *Under Assumptions 1.1, 1.2, 1.3 and 1.4, for any $\nu \in [0, 1]$, there exists an $M \times$*

$(B+1)$ *random matrix* $\mathbf{X}_N(\nu)$ *with* $\mathcal{N}_{\mathbb{C}}(0,1)$ *i.i.d. entries, and two matrices* $(\tilde{\mathbf{\Delta}}_N(\nu), \mathbf{\Delta}_N(\nu))$ *such that:*

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu), \quad \|\tilde{\mathbf{\Delta}}_N(\nu)\| \prec \frac{B}{N} \tag{1.46}$$

$$\hat{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1} + \mathbf{\Delta}_N(\nu), \quad \|\mathbf{\Delta}_N(\nu)\| \prec \frac{1}{\sqrt{B}} + \frac{B}{N}. \tag{1.47}$$

Remark 1.3. *Therefore, up to small additive perturbations,* $\tilde{\mathbf{C}}_N(\nu)$ *and* $\hat{\mathbf{C}}_N(\nu)$ *appear as empirical covariance matrices of i.i.d.* $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ *random vectors. We thus expect that* $\tilde{\mathbf{C}}_N(\nu)$ *and* $\hat{\mathbf{C}}_N(\nu)$ *will satisfy a number of useful properties of empirical covariance matrices of i.i.d.* $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ *random vectors.*

In particular, Theorem 1.1 allows to make precise the location of the eigenvalues of $\tilde{\mathbf{C}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$. In order to formulate the corresponding result, we define some notations. We introduce the events $\Lambda^{\tilde{\mathbf{C}}}_{N,\epsilon}(\nu)$ and $\Lambda^{\hat{\mathbf{C}}}_{N,\epsilon}(\nu)$ defined by

$$\Lambda^{\tilde{\mathbf{C}}}_{N,\epsilon}(\nu) = \{\sigma(\tilde{\mathbf{C}}_N(\nu)) \subset \text{Supp}\mu^{(c)}_{MP} + \epsilon\} \tag{1.48}$$

$$\Lambda^{\hat{\mathbf{C}}}_{N,\epsilon}(\nu) = \{\sigma(\hat{\mathbf{C}}_N(\nu)) \subset \text{Supp}\mu^{(c)}_{MP} + \epsilon\}. \tag{1.49}$$

Then, we establish in the following the Corollary:

Corollary 1.1. *For each* $\epsilon > 0$*, the family of events* $\Lambda^{\tilde{\mathbf{C}}}_{N,\epsilon}(\nu), N \geq 1, \nu \in [0,1]$ *and* $\Lambda^{\hat{\mathbf{C}}}_{N,\epsilon}(\nu), N \geq 1, \nu \in [0,1]$ *hold with exponential high probability.*

Remark 1.4. *In the following, we will often omit to mention that the various matrices under consideration depend on* N *and* ν*. Matrices* $\hat{\mathbf{C}}_N(\nu), \tilde{\mathbf{C}}_N(\nu), \mathbf{X}_N(\nu), \mathbf{\Delta}_N(\nu), \ldots$ *will therefore be denoted by* $\hat{\mathbf{C}}(\nu), \tilde{\mathbf{C}}(\nu), \mathbf{X}(\nu), \mathbf{\Delta}(\nu), \ldots$ *or* $\hat{\mathbf{C}}, \tilde{\mathbf{C}}, \mathbf{X}, \mathbf{\Delta}, \ldots$*. We will also denote* $\Lambda^{\tilde{\mathbf{C}}}_{N,\epsilon}(\nu)$ *and* $\Lambda^{\hat{\mathbf{C}}}_{N,\epsilon}(\nu)$ *by* $\Lambda^{\tilde{\mathbf{C}}}_{\epsilon}(\nu)$ *or* $\Lambda^{\tilde{\mathbf{C}}}_{\epsilon}$ *and* $\Lambda^{\hat{\mathbf{C}}}_{\epsilon}(\nu)$ *or* $\Lambda^{\hat{\mathbf{C}}}_{\epsilon}$*.*

The proof of Theorem 1.1 will proceed in three steps: first we provide the result for matrix $\tilde{\mathbf{C}}(\nu)$, then control the deviations between $\text{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ and $\text{dg}(\hat{\mathbf{S}}(\nu))^{-\frac{1}{2}}$, and finally extend the stochastic representation of $\tilde{\mathbf{C}}(\nu)$ to $\hat{\mathbf{C}}(\nu)$.

1.3.1 Step 1: Stochastic representation of $\tilde{\mathbf{C}}$

In order to establish (1.46), we prove the following Proposition.

Proposition 1.1. *Under Assumptions 1.1, 1.2, 1.3 and 1.4, for any* $\nu \in [0,1]$*, there exist an* $M \times (B+1)$ *random matrix* $\mathbf{X}_N(\nu)$ *with* $\mathcal{N}_{\mathbb{C}}(0,1)$ *i.i.d. entries, and another matrix* $\mathbf{\Gamma}_N(\nu)$ *such that:*

$$\tilde{\mathbf{C}}_N(\nu) = \frac{(\mathbf{X}_N(\nu) + \mathbf{\Gamma}_N(\nu))(\mathbf{X}_N(\nu) + \mathbf{\Gamma}_N(\nu))^*}{B+1} \tag{1.50}$$

where the family of random variables $\frac{\|\mathbf{\Gamma}_N(\nu)\|^2}{B+1}, \nu \in [0,1]$ *satisfies*

$$\frac{\|\mathbf{\Gamma}_N(\nu)\|^2}{B+1} \prec \frac{B^2}{N^2}. \tag{1.51}$$

Proof. Denote by $\mathbf{\Sigma}$ the $M \times (B+1)$ random matrix defined by

$$\mathbf{\Sigma} = \left(\boldsymbol{\xi}_{\mathbf{y}}(\nu - \frac{B}{2N}), \ldots, \boldsymbol{\xi}_{\mathbf{y}}(\nu + \frac{B}{2N})\right) \tag{1.52}$$

where we recall that the normalized Fourier transform $\boldsymbol{\xi}_{\mathbf{y}}$ is defined in (1.2), so that $\hat{\mathbf{S}}$ defined in (1.1) is equal to $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*/(B+1)$. Denote by $\boldsymbol{\omega}_m$ the m–th row of $\boldsymbol{\Sigma}$. In other words, $\boldsymbol{\omega}_m$ coincides with the $(B+1)$–dimensional Gaussian complex row vector defined by:

$$\boldsymbol{\omega}_m = \left(\xi_{y_m}(\nu - \frac{B}{2N}), \ldots, \xi_{y_m}(\nu + \frac{B}{2N})\right).$$

The covariance matrix $\mathbb{E}[\boldsymbol{\omega}_m^* \boldsymbol{\omega}_m]$ of $\boldsymbol{\omega}$ is given by:

$$\mathbb{E}[\boldsymbol{\omega}_m^* \boldsymbol{\omega}_m] = \mathbb{E}\left[\left\{\xi_{y_m}(\nu + \frac{b_1}{N})^* \xi_{y_m}(\nu + \frac{b_2}{N})\right\}_{b_1, b_2 = -B/2}^{B/2}\right]$$

By Lemma 1.12 in Appendix, we have for b and $b_1 \neq b_2$:

$$\mathbb{E}\left[\left|\xi_{y_m}\left(\nu + \frac{b}{N}\right)\right|^2\right] = s_m\left(\nu + \frac{b}{N}\right) + \mathcal{O}\left(\frac{1}{N}\right)$$
$$\mathbb{E}\left[\xi_{y_m}(\nu + \frac{b_1}{N})^* \xi_{y_m}(\nu + \frac{b_2}{N})\right] = \mathcal{O}\left(\frac{1}{N}\right)$$

where the error is uniform over $m \geq 1$ and $\nu \in [0,1]$. Therefore one can claim that there exists some Hermitian matrix $\boldsymbol{\Upsilon}_m(\nu)$ and some nice constant C such that:

$$\mathbb{E}[\boldsymbol{\omega}_m^* \boldsymbol{\omega}_m] = \mathrm{dg}\left(s_m\left(\nu + \frac{b}{N}\right) : b = -B/2, \ldots, B/2\right) + \boldsymbol{\Upsilon}_m$$

where $\boldsymbol{\Upsilon}_m$ satisfies

$$\sup_{m \geq 1, b_1, b_2} \left|(\boldsymbol{\Upsilon}_m)_{b_1, b_2}\right| \leq \frac{C}{N}.$$

Moreover, the regularity of the mapping $\nu \mapsto s_m(\nu)$ specified in Assumption 1.4 implies that there exists quantities ϵ_m such that:

$$s_m(\nu + \frac{b}{N}) = s_m(\nu) + s_m'(\nu)\,\frac{b}{N} + \frac{1}{2}\,s_m''(\nu)\,(\frac{b}{N})^2 + \epsilon_m(\nu + \frac{b}{N})$$

where:

$$\sup_{m \geq 1} \sup_{-B/2 \leq b \leq B/2} |\epsilon_m(\nu + \frac{b}{N})| \leq C\left(\frac{B}{N}\right)^3$$

for some nice constant C. Therefore, it holds that

$$\begin{aligned}
&\mathrm{dg}\left(s_m(\nu + \frac{b}{N}) : b = -B/2, \ldots, B/2\right) \\
&\quad = s_m(\nu)\mathbf{I}_{B+1} + s_m'(\nu)\,\mathrm{dg}\left(\frac{b}{N} : b = -B/2, \ldots, B/2\right) + \\
&\quad\quad \frac{1}{2}\,s_m''(\nu)\,\mathrm{dg}\left((\frac{b}{N})^2 : b = -B/2, \ldots, B/2\right) + \\
&\quad\quad\quad \mathrm{dg}\left(\epsilon_m(\nu + \frac{b}{N}) : b = -B/2, \ldots, B/2\right).
\end{aligned}$$

If we define matrix $\boldsymbol{\Phi}_m$ as:

$$\boldsymbol{\Phi}_m = \frac{1}{s_m}\left[\boldsymbol{\Upsilon}_m + \mathrm{dg}\left(\mathbf{s}_m(\nu + \frac{b}{N}) - s_m(\nu) : b = -B/2, \ldots, B/2\right)\right]$$

then the following relations hold:

$$\mathbb{E}[\boldsymbol{\omega}_m^* \boldsymbol{\omega}_m] = s_m \left(\mathbf{I}_{B+1} + \boldsymbol{\Phi}_m\right), \qquad \sup_{m \geq 1, b_1 \neq b_2} |(\boldsymbol{\Phi}_m)_{b_1,b_2}| \leq \frac{C}{N}, \quad \sup_{m \geq 1, b} |(\boldsymbol{\Phi}_m)_{b,b}| \leq \frac{CB}{N} \tag{1.53}$$

as well as

$$\frac{1}{B+1} \mathrm{tr}\ \boldsymbol{\Phi}_m = \frac{1}{2} \frac{s''_m(\nu)}{s_m(\nu)} v_N + \mathcal{O}\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right) \tag{1.54}$$

where we recall that v_N is defined by (1.5). The spectral norm of $\boldsymbol{\Phi}_m$ can be roughly bounded by the following inequality:

$$\sup_{m \geq 1} \|\boldsymbol{\Phi}_m\| \leq \sup_{m \geq 1} \sup_{-B/2 \leq b_1 \leq B/2} \sum_{b_2 = -B/2}^{B/2} |(\boldsymbol{\Phi}_m)_{b_1,b_2}| \leq C \frac{B}{N}.$$

Moreover, it is easily checked that the Frobenius norm of $\frac{\boldsymbol{\Phi}_m}{B+1}$ satisfies

$$\left\| \frac{\boldsymbol{\Phi}_m}{B+1} \right\|_F \leq C \, \frac{\sqrt{B}}{N} = \mathcal{O}(u_N). \tag{1.55}$$

Using the Gaussianity of the vector $\boldsymbol{\omega}_m$ and the expression (1.53), we obtain that $\boldsymbol{\omega}_m$ can be represented as

$$\boldsymbol{\omega}_m = \sqrt{s_m} \, \mathbf{x}_m \left(\mathbf{I} + \boldsymbol{\Phi}_m\right)^{1/2}, \quad \mathbf{x}_m \sim \mathcal{N}_{\mathbb{C}}(0, I_{B+1}) \tag{1.56}$$

where $\mathbf{x}_{m_1}$ and $\mathbf{x}_{m_2}$ are independent for $m_1 \neq m_2$. This comes from the mutual independence of the time series $((y_{m,n})_{n \in \mathbb{Z}})_{m=1,\ldots,M}$. It is clear that $\left(\mathbf{I} + \boldsymbol{\Phi}_m\right)^{1/2}$ can be written as

$$\left(\mathbf{I} + \boldsymbol{\Phi}_m\right)^{1/2} = \mathbf{I} + \boldsymbol{\Psi}_m \tag{1.57}$$

where the matrix $\boldsymbol{\Psi}_m$ satisfies

$$\sup_m \|\boldsymbol{\Psi}_m\| \leq C \, \frac{B}{N} \tag{1.58}$$

Therefore, it holds that:

$$\boldsymbol{\omega}_m = \sqrt{s_m} \mathbf{x}_m \left(\mathbf{I} + \boldsymbol{\Psi}_m\right) = \sqrt{s_m} \left(\mathbf{x}_m + \mathbf{x}_m \boldsymbol{\Psi}_m\right)$$

We denote by $\mathbf{X}$ and $\boldsymbol{\Gamma}$ the $M \times (B+1)$ matrices with rows $(\mathbf{x}_m)_{m=1,\ldots,M}$, and $(\mathbf{x}_m \boldsymbol{\Psi}_m)_{m=1,\ldots,M}$ respectively. Then, it holds that

$$\boldsymbol{\Sigma} = \mathrm{dg}\left(\sqrt{s_m}, m = 1, \ldots, M\right) \left(\mathbf{X} + \boldsymbol{\Gamma}\right) \tag{1.59}$$

where we recall that $\boldsymbol{\Sigma}$ is defined by (1.52). We recall the definition of the matrix $\tilde{\mathbf{C}}$ given by

$$\begin{aligned} \tilde{\mathbf{C}} &= \mathrm{dg}(\sqrt{s_m}, m = 1, \ldots, M)^{-1/2} \, \hat{\mathbf{S}} \, \mathrm{dg}(\sqrt{s_m}, m = 1, \ldots, M)^{-1/2} \\ &= \mathrm{dg}(\sqrt{s_m}, m = 1, \ldots, M)^{-1/2} \frac{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*}{B+1} \mathrm{dg}(\sqrt{s_m}, m = 1, \ldots, M)^{-1/2}. \end{aligned} \tag{1.60}$$

The representation (1.59) implies that $\tilde{\mathbf{C}}$ can also be written as

$$\tilde{\mathbf{C}} = \frac{(\mathbf{X} + \boldsymbol{\Gamma})(\mathbf{X} + \boldsymbol{\Gamma})^*}{B+1}.$$

Equivalently, for each m_1, m_2, the entry $(\tilde{\mathbf{C}})_{m_1,m_2}$ is given by

$$(\tilde{\mathbf{C}})_{m_1,m_2} = \frac{1}{B+1}\mathbf{x}_{m_1}(\mathbf{I}+\mathbf{\Phi}_{m_1})^{1/2}(\mathbf{I}+\mathbf{\Phi}_{m_2})^{1/2}\mathbf{x}_{m_2}^*. \tag{1.61}$$

This completes the proof of (1.50). It remains to show (1.51). We denote by $\mathbf{Z}$ the $M \times M$ matrix $\mathbf{Z} = \frac{1}{B+1}\mathbf{\Gamma}\mathbf{\Gamma}^*$. As $\|\mathbf{Z}\|$ satisfies

$$\|\mathbf{Z}\| \leq \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| + \|\mathbb{E}\mathbf{Z}\|$$

it is enough to prove the two following facts:

$$\|\mathbb{E}\mathbf{Z}\| \leq C\frac{B^2}{N^2} \tag{1.62}$$

$$\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| \prec \frac{B^2}{N^2}. \tag{1.63}$$

We start with (1.62). The definition of $\mathbf{\Gamma}$ leads to

$$\mathbb{E}[\mathbf{Z}_{i,j}] = \frac{1}{B+1}\mathbb{E}[\mathbf{\Gamma}\mathbf{\Gamma}^*]_{i,j} = \frac{1}{B+1}\mathbb{E}[\mathbf{x}_i\mathbf{\Psi}_i\mathbf{\Psi}_j^*\mathbf{x}_j^*] = \delta_{ij}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Psi}_i\mathbf{\Psi}_j^*$$

so that it is clear that $\mathbb{E}[\mathbf{Z}]$ is the diagonal matrix with diagonal entries $(\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Psi}_m\mathbf{\Psi}_m^*)_{m=1,\ldots,M}$. By the estimation in equation (1.58), we easily have (1.62).

It remains to prove (1.63). We use the observation that $\|\mathbf{Z}-\mathbb{E}[\mathbf{Z}]\| = \max_{\|\mathbf{h}\|=1}|\mathbf{h}^*(\mathbf{Z}-\mathbb{E}[\mathbf{Z}])\mathbf{h}|$, and use a classical ϵ–net argument that allows to deduce the behaviour of $\|\mathbf{Z}-\mathbb{E}[\mathbf{Z}]\|$ from the behaviour of any recentered quadratic form $\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}$ where $\mathbf{g} \in \mathbb{C}^M$ is a deterministic unit norm vector. We thus first concentrate $\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}$ using the Hanson-Wright inequality (1.41). For this, we need to express $\mathbf{g}^*\mathbf{Z}\mathbf{g}$ as a quadratic form of a certain complex Gaussian random vector with i.i.d. entries. We denote by $\mathbf{z}$ the M–dimensional random vector $\mathbf{z} = \frac{\mathbf{\Gamma}^*(\nu)\mathbf{g}}{\sqrt{B+1}}$. Its covariance matrix $\mathbf{G} = \mathbf{G}(\nu)$ is equal to

$$\mathbf{G}(\nu) = \mathbb{E}[\mathbf{z}\mathbf{z}^*] = \frac{1}{B+1}\sum_{m=1}^{M}|\mathbf{g}_m|^2(\mathbf{\Psi}_m(\nu))^*\mathbf{\Psi}_m(\nu).$$

Therefore, $\mathbf{z}$ can be written as $\mathbf{z} = \mathbf{G}^{1/2}\mathbf{w}$ for some $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ random vector. As a consequence, the quadratic form $\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}$ can be written as

$$\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g} = \mathbf{w}^*\mathbf{G}\mathbf{w} - \mathbb{E}\mathbf{w}^*\mathbf{G}\mathbf{w}.$$

The Hanson-Wright inequality (1.41) can now be applied:

$$|\mathbf{w}^*\mathbf{G}\mathbf{w} - \mathbb{E}\mathbf{w}^*\mathbf{G}\mathbf{w}| \prec \|\mathbf{G}\|_F. \tag{1.64}$$

Since $\sum_{m=1}^M |\mathbf{g}_m|^2 = 1$, it is clear that $\|\mathbf{G}\| \leq \frac{1}{B+1}\sup_{m=1,\ldots,M}\|\mathbf{\Psi}_m(\nu)\|^2$. Therefore, (1.58) and the rough bound $\|\mathbf{G}\|_F^2 \leq (B+1)\|\mathbf{G}\|^2$ leads to

$$\|\mathbf{G}\| \leq C\frac{1}{B+1}\left(\frac{B}{N}\right)^2, \quad \|\mathbf{G}\|_F^2 \leq C\frac{1}{B+1}\left(\frac{B}{N}\right)^4 \tag{1.65}$$

The substitution of (1.65) in equation (1.64) gives the following control of $\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}$:

$$|\mathbf{g}^*\mathbf{Z}\mathbf{g} - \mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}| \prec \frac{1}{\sqrt{B}}\left(\frac{B}{N}\right)^2 \tag{1.66}$$

Consider $\epsilon > 0$, and an ϵ–net N_ϵ of $\mathbb{C}^M$, that is a set of $\mathbb{C}^M$ unit norm vectors $\{\mathbf{h}_k : k = 1, \ldots, \mathcal{K}\}$ such that for each unit norm vector $\mathbf{u} \in \mathbb{C}^M$, there exists a vector $\mathbf{h} \in N_\epsilon$ for which $\|\mathbf{u} - \mathbf{h}\| \leq \epsilon$. It is well known that the cardinality of N_ϵ is bounded by $C_0 \left(\frac{1}{\epsilon}\right)^{2M}$ where C_0 is a universal constant. Then, denote $\mathbf{g}_s$ a (random) unit norm vector such that $|\mathbf{g}_s^* \mathbf{Z} \mathbf{g}_s - \mathbb{E}\mathbf{g}_s^* \mathbf{Z} \mathbf{g}_s| = \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\|$, and define $\mathbf{h}_s \in \mathbf{N}_\epsilon$ as the closest vector from $\mathbf{g}_s$. Therefore, we have

$$\begin{aligned}
\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| &= |\mathbf{g}_s^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{g}_s| \\
&= |(\mathbf{g}_s^* - \mathbf{h}_s^* + \mathbf{h}_s^*)(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{g}_s - \mathbf{h}_s + \mathbf{h}_s)| \\
&\leq |(\mathbf{g}_s^* - \mathbf{h}_s^*)(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{g}_s - \mathbf{h}_s)| + |(\mathbf{g}_s^* - \mathbf{h}_s^*)(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}_s| \\
&\qquad + |\mathbf{h}_s^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{g}_s - \mathbf{h}_s)| + |\mathbf{h}_s^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}_s|.
\end{aligned}$$

It is clear that:

$$|(\mathbf{g}_s^* - \mathbf{h}_s^*)(\mathbf{Z} - \mathbb{E}\mathbf{Z})(\mathbf{g}_s - \mathbf{h}_s)| \leq \epsilon^2 \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\|, \quad |(\mathbf{g}_s^* - \mathbf{h}_s^*)(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}_s| \leq \epsilon \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\|$$

and

$$\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| \leq |\mathbf{h}_s^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}_s| + \epsilon^2 \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| + 2\epsilon \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\|$$

which leads to

$$(1 - 2\epsilon - \epsilon^2)\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| \leq |\mathbf{h}_s^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}_s|.$$

This implies that for each $t > 0$,

$$\{\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| > t\} \subset \cup_{h \in N_\epsilon} \{|\mathbf{h}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}| > C_1 t\}$$

where $C_1 = (1 - 2\epsilon - \epsilon^2)$. Using the union bound, we obtain that

$$\mathbb{P}\left[\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| > t\right] \leq \sum_{h \in N_\epsilon} \mathbb{P}\left[|\mathbf{h}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}| > C_1 t\right]. \tag{1.67}$$

Here, we would like to use equation (1.66). By the definition of $\prec$, (1.66) is valid uniformly on any set of vector with cardinality polynomial in N. Here, the cardinality of the set N_ϵ is a $\mathcal{O}(\epsilon^{-2M})$ term and therefore exponential in M. As a consequence, we have to accept to lose some speed when going from the stochastic domination of $|\mathbf{g}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{g}|$ for a fixed $\mathbf{g}$ to the same stochastic domination but uniformly over N_ϵ.

More specifically, write again (1.66) but here without the notation $\prec$ in order to understand precisely how a change in speed affects the probability. Take t_N a sequence of positive numbers such that $t_N \geq B^2/N^2$. Using the estimates (1.65) of $\|\mathbf{G}\|$ and $\|\mathbf{G}\|_F^2$, and the fact that $\min(a_1, a_2) > \min(b_1, b_2)$ when $a_1 > b_1$ and $a_2 > b_2$, we obtain that there exist some nice constant $C > 0$ such that:

$$\min\left(\frac{t_N}{\|\mathbf{G}\|}, \frac{t_N^2}{\|\mathbf{G}\|_F^2}\right) \geq C\, B \min\left(t_N \left(\frac{N}{B}\right)^2, \left(t_N \left(\frac{N}{B}\right)^2\right)^2\right) = C\, B\, t_N \left(\frac{N}{B}\right)^2.$$

The Hanson-Wright inequality (1.39) provides:

$$\mathbb{P}\left[|\mathbf{g}^*\left[\mathbf{Z} - \mathbb{E}\mathbf{Z})\right]\mathbf{g}| > C_1 t_N\right] \leq 2\,\exp\left\{-CB\frac{t_N}{(B/N)^2}\right\}$$

for some nice constant C that depends on C_1. Finally, the union bound on N_ϵ gives:

$$\mathbb{P}\left[\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\| > t_N\right] \leq \sum_{h \in N_\epsilon} \mathbb{P}\left[|\mathbf{h}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}| > C_1 t_N\right]$$
$$\leq 2\, C_0 \exp\left\{-CB\frac{t_N}{(B/N)^2} + 2M \log\frac{1}{\epsilon}\right\}. \tag{1.68}$$

If we take $t_N = N^{\epsilon'}(B^2/N^2)$, then, there exists $\gamma > 0$ such that

$$\exp\left\{-CB\frac{t_N}{(B/N)^2} + 2CM\log\frac{1}{\epsilon}\right\} \leq \exp -N^\gamma$$

holds for each N large enough. (1.67) thus implies (1.63). This completes the proof of (1.50). □

Corollary 1.2 is a rewriting of Proposition 1.1 in a more concise way. Define:

$$\tilde{\mathbf{\Delta}} = \frac{\mathbf{X}\mathbf{\Gamma}^* + \mathbf{\Gamma}\mathbf{X}^* + \mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1}. \tag{1.69}$$

Corollary 1.2. *For any $\nu \in [0,1]$, $\tilde{\mathbf{C}}(\nu)$ can be written as*

$$\tilde{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}(\nu) \tag{1.70}$$

where the family of random variable $\|\tilde{\mathbf{\Delta}}(\nu)\|$, $\nu \in [0,1]$ satisfies

$$\|\tilde{\mathbf{\Delta}}\| \prec \frac{B}{N}. \tag{1.71}$$

Proof. Let $\nu \in [0,1]$. By equation (1.51) from Theorem 1.1 and equation (1.29) from Paragraph 1.2.2, we have the two following estimates:

$$\frac{\|\mathbf{\Gamma}\|}{\sqrt{B+1}} \prec \frac{B}{N}, \quad \frac{\|\mathbf{X}\|}{\sqrt{B+1}} \prec 1$$

The result is immediate using decomposition $\tilde{\mathbf{\Delta}}$ from (1.69): □

We now take benefit of Corollary 1.2 to establish the first part of Corollary 1.1 and to analyse the location of the eigenvalues of matrices $\hat{\mathbf{S}}$. We denote by $\mathbf{D}$ and $\hat{\mathbf{D}}$ the matrices $\mathbf{D} = \mathbf{D}(\nu) := \mathrm{dg}(\mathbf{S}(\nu))^{\frac{1}{2}}$ and $\hat{\mathbf{D}} = \hat{\mathbf{D}}(\nu) := \mathrm{dg}(\hat{\mathbf{S}}(\nu))^{\frac{1}{2}}$. Denote by $\bar{s}$ and $\underline{s}$ the quantities such that:

$$\underline{s} := \inf_{m \geq 1} \inf_{\nu \in [0,1]} s_m(\nu), \quad \bar{s} := \sup_{m \geq 1} \sup_{\nu \in [0,1]} s_m(\nu)$$

which are by Assumption 1.4 in $(0, +\infty)$. We consider the event:

$$\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu) = \left\{\sigma(\hat{\mathbf{S}}(\nu)) \subset \mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon\right\} \tag{1.72}$$

where the notation $\mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}]$ stands for $[(1-\sqrt{c})^2 \underline{s}, (1+\sqrt{c})^2 \bar{s}]$. Note that in our settings, $c \in (0,1)$ so $\mathrm{Supp}\mu_{MP}^{(c)}$ is bounded and away from zero. In conjunction with Assumption 1.4, the same holds for $\mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}]$. We also note that $\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu)$ of course depends on N.

Corollary 1.3. *For any $\epsilon > 0$, the families of events $\Lambda_\epsilon^{\tilde{\mathbf{C}}}(\nu)$, $\nu \in [0,1]$ and $\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu)$, $\nu \in [0,1]$ hold with exponentially high probability.*

Proof. Equation (1.70) implies that

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1} - \|\tilde{\mathbf{\Delta}}\|\mathbf{I}_M \leq \tilde{\mathbf{C}} \leq \frac{\mathbf{X}\mathbf{X}^*}{B+1} + \|\tilde{\mathbf{\Delta}}\|\mathbf{I}_M.$$

Therefore, the event $\{\lambda_1(\tilde{\mathbf{C}}) > (1+\sqrt{c})^2 + \epsilon\}$ is included in $\{\lambda_1(\frac{\mathbf{X}\mathbf{X}^*}{B+1}) + \|\tilde{\mathbf{\Delta}}\| > (1+\sqrt{c})^2 + \epsilon\}$, which is itself included in

$$\left\{\lambda_1(\frac{\mathbf{X}\mathbf{X}^*}{B+1}) > (1+\sqrt{c})^2 + \epsilon/2\right\} \cup \left\{\|\tilde{\mathbf{\Delta}}\| > \epsilon/2\right\}.$$

Therefore,

$$\mathbb{P}\left[\lambda_1(\tilde{\mathbf{C}}) > (1+\sqrt{c})^2 + \epsilon\right] \leq \mathbb{P}\left[\lambda_1(\frac{\mathbf{X}\mathbf{X}^*}{B+1}) > (1+\sqrt{c})^2 + \epsilon/2\right] + \mathbb{P}\left[\|\tilde{\mathbf{\Delta}}\| > \epsilon/2\right].$$

Equations (1.27) and (1.71) imply that $\mathbb{P}\left[\lambda_1(\tilde{\mathbf{C}}) > (1+\sqrt{c})^2 + \epsilon\right]$ converges towards 0 exponentially. A similar evaluation of $\mathbb{P}\left[\lambda_M(\tilde{\mathbf{C}}) < (1-\sqrt{c})^2 - \epsilon\right]$ leads to the same conclusion. This, in turn, establishes that $\Lambda_\epsilon^{\tilde{\mathbf{C}}}(\nu), \nu \in [0,1]$ holds with exponential high probability.

In order to establish that the same property holds for $\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu), \nu \in [0,1]$, we just need to write (1.12) as $\hat{\mathbf{S}} = \mathbf{D}^{1/2}\tilde{\mathbf{C}}\mathbf{D}^{1/2}$. Therefore, for each $k = 1, \ldots, M$, the eigenvalues of $\hat{\mathbf{S}}$ satisfy

$$\underline{s}\,\lambda_M(\tilde{\mathbf{C}}) \leq \lambda_k(\hat{\mathbf{S}}) \leq \bar{s}\,\lambda_1(\tilde{\mathbf{C}}).$$

This, of course, implies that $\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu), \nu \in [0,1]$ holds with exponential high probability (indeed, one can change ϵ to $\tilde{\epsilon}$ such that $(\mathrm{Supp}\mu_{MP}^{(c)} + \tilde{\epsilon}) \times [\underline{s}, \bar{s}] \subset \mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon$.

□

Remark 1.5. *Corollary 1.3 implies the following weaker property, which will be useful:*

$$\|\hat{\mathbf{S}}(\nu)\| \prec 1. \tag{1.73}$$

Before ending the section and proving Theorem 1.1, we need some stochastic control on the diagonal elements of $\hat{\mathbf{S}}$ in order to evaluate $\mathbf{\Theta}$ defined by

$$\mathbf{\Theta} := \hat{\mathbf{C}} - \tilde{\mathbf{C}}. \tag{1.74}$$

Using the definition of $\hat{\mathbf{C}}$ from (1.3) and $\tilde{\mathbf{C}}$ from (1.12), $\mathbf{\Theta}$ can be written as

$$\mathbf{\Theta} = (\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2})\hat{\mathbf{S}}\hat{\mathbf{D}}^{-1/2} + \mathbf{D}^{-1/2}\hat{\mathbf{S}}(\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}). \tag{1.75}$$

Since we proved that $\|\hat{\mathbf{S}}\| \prec 1$, it remains to show that $\|\hat{\mathbf{D}}^{-1/2}\|$ and $\|\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}\|$ can also be stochastically dominated by some relevant quantity in order to control $\|\mathbf{\Theta}\|$. Define

$$\hat{s}_m(\nu) := \hat{\mathbf{S}}_{m,m}(\nu) \tag{1.76}$$

the diagonal elements of $\hat{\mathbf{S}}(\nu)$ spectral density estimator (note that they coincide with the traditional smoothed periodogram estimator of the spectral density s_m). The aim of the following Paragraph 1.3.2 is to establish stochastic domination results for $\hat{s}_m$, $\|\hat{\mathbf{D}}^{-1/2}\|$ and $\|\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}\|$.

1.3.2 Step 2: Estimates for $\hat{s}_m(\nu)$

We write $s_m(\nu) := s_m$, $\mathbf{D}(\nu) := \mathbf{D}$, in order to simplify the notations. Define as in (1.72) the following quantity

$$\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu) = \{\sigma(\hat{\mathbf{D}}(\nu)) \subset [\underline{s}, \bar{s}] + \epsilon\}. \tag{1.77}$$

Lemma 1.3. *Let $\epsilon > 0$. The family of events $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu), \nu \in [0,1]$ holds with exponentially high probability.*

Proof. See Appendix 1.B. □

Roughly speaking, this ensures that with exponentially high probability, $\hat{s}_m$ stays bounded and away from zero. This result implies the following (weaker) statement, but will still be enough for some proofs and reduces the complexity of the arguments.

Lemma 1.4. *The family of random variables $(|\hat{s}_m(\nu)| + \frac{1}{|\hat{s}_m(\nu)|})_{m=1,\ldots,M}$, $\nu \in [0,1]$, satisfies*

$$\left(|\hat{s}_m| + \frac{1}{|\hat{s}_m|}\right) \prec 1.$$

Proof. Immediate from Lemma 1.3. □

Lemma 1.5. *The set of random variable $(|\hat{s}_m(\nu)^{-1/2} - s_m(\nu)^{-1/2}|)_{m=1,\ldots,M}$ and $(|\sqrt{\frac{s_m(\nu)}{\hat{s}_m(\nu)}} - 1|)_{m=1,\ldots,M}$, $\nu \in [0,1]$, satisfies*

$$|\hat{s}_m^{-1/2} - s_m^{-1/2}| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2}, \quad \left|\sqrt{\frac{s_m}{\hat{s}_m}} - 1\right| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2}. \tag{1.78}$$

Proof. See Appendix 1.C □

1.3.3 Step 3: Stochastic representation of $\hat{\mathbf{C}}$

We are now in a position to prove the result concerning $\hat{\mathbf{C}}$ of Theorem 1.1 and of Corollary 1.1.

Proof. We have first to control the operator norm of:

$$\boldsymbol{\Delta} = \hat{\mathbf{C}} - \frac{\mathbf{X}\mathbf{X}^*}{B+1} = \hat{\mathbf{C}} - \tilde{\mathbf{C}} + \tilde{\mathbf{C}} - \frac{\mathbf{X}\mathbf{X}^*}{B+1} = \boldsymbol{\Theta} + \tilde{\boldsymbol{\Delta}}. \tag{1.79}$$

The operator norm of $\|\tilde{\boldsymbol{\Delta}}\|$ has already been proved in Corollary 1.2 to satisfy $\|\tilde{\boldsymbol{\Delta}}\| \prec (\frac{B}{N})$. Moreover, recall that $\boldsymbol{\Theta}$ can be written as a function of $\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}$ in (1.75), so that one can use Lemma 1.4 and Lemma 1.5 to dominate each term and get:

$$\|\boldsymbol{\Theta}\| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2}. \tag{1.80}$$

Summing the estimate of $\boldsymbol{\Theta}$ and the one of $\tilde{\boldsymbol{\Delta}}$, one gets:

$$\|\boldsymbol{\Delta}\| \prec \frac{1}{\sqrt{B}} + \frac{B}{N}$$

which is the desired result. □

As a consequence, we state here Corollary 1.4 about the localization of the eigenvalues of $\hat{\mathbf{C}}(\nu)$.

Corollary 1.4. *For each $\epsilon > 0$, we define $\Lambda_{\epsilon}^{\hat{\mathbf{C}}}(\nu)$ as the event*

$$\Lambda_{\epsilon}^{\hat{\mathbf{C}}}(\nu) = \left\{ \sigma(\hat{\mathbf{C}}(\nu)) \subset \mathrm{Supp}\mu_{MP}^{(c)} + \epsilon \right\}. \tag{1.81}$$

Then, the family of events $\Lambda_{\epsilon}^{\hat{\mathbf{C}}}(\nu), \nu \in [0,1]$ holds with exponentially high probability.

Proof. We simply write:

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1} - \|\mathbf{\Delta}\|\mathbf{I}_M \leq \hat{\mathbf{C}} \leq \frac{\mathbf{X}\mathbf{X}^*}{B+1} + \|\mathbf{\Delta}\|\mathbf{I}_M$$

and use the same arguments as in the proof of Corollary 1.3. □

1.4 Stochastic domination of the family $\psi_N(f,\nu), N \geq 1, \nu \in [0,1]$

We have first to define the distribution D_N introduced in the definition (1.11) of $\psi_N(f,\nu)$. For this, we consider the function $p_N(z)$ defined by

$$p_N(z) = -\frac{c_N\,(z\,t_N(z)\,\tilde{t}_N(z))^3}{1 - c(z\,t_N(z)\,\tilde{t}_N(z))^2} \tag{1.82}$$

where we recall that t_N and $\tilde{t}_N$ are defined by (1.34) and (1.35). Then (see Lemma 9.2 in [59]), p_N is the Stieltjes transform of a distribution whose support is contained in the support $\mathrm{Supp}\mu_{MP}^{(c_N)} = [(1-\sqrt{c}_N)^2, (1+\sqrt{c}_N)^2]$ of the Marcenko-Pastur distribution $\mu_{MP}^{(c_N)}$. This distribution is D_N introduced in (1.11). In the following, we consider LSS for function f satisfying the following assumptions.

Assumption 1.5. *f is defined on $\mathbb{R}_+$ and there exists some $\epsilon > 0$ such that its restriction on $\mathrm{Supp}_{MP}^{(c)} + \epsilon$ is $\mathcal{C}^{\infty}$.*

We now state the main result of this section.

Theorem 1.2. *Let f be a function satisfying the conditions of Assumption 1.5. Then, under Assumptions 1.1, 1.2, 1.3 and 1.4, the family $|\psi_N(f,\nu)|, N \geq 1, \nu \in [0,1]$ satisfies*

$$|\psi_N(f,\nu)| \prec u_N. \tag{1.83}$$

Before starting the proof of Theorem 1.2, we first mention that it is sufficient to establish (1.83) when f is compactly supported by a neighbourhood of $\mathrm{Supp}\mu_{MP}^{(c)}$. To justify this claim, we consider $\kappa > 0$ and define $\chi : \mathbb{R} \to \mathbb{R}$ as a $\mathbb{C}^{\infty}$ function such that:

$$\chi(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \mathrm{Supp}\mu_{MP}^{(c)} + \kappa \\ 0 & \text{if } \lambda \notin \mathrm{Supp}\mu_{MP}^{(c)} + 2\kappa. \end{cases} \tag{1.84}$$

We consider the function $\bar{f}$ given by $\bar{f} = f \times \chi$. Then, as $c_N \to c$, for N large enough, $\mathrm{Supp}\mu_{MP}^{(c_N)}$ is contained in $\mathrm{Supp}\mu_{MP}^{(c)} + \kappa$. Therefore, $f = \bar{f}$ on $\mathrm{Supp}\mu_{MP}^{(c_N)}$ for N large enough, and it holds that $< D_N, f > = < D_N, \bar{f} >$ and $\int f d\mu_{MP}^{(c_N)} = \int \bar{f} d\mu_{MP}^{(c_N)}$. For each $\epsilon > 0$, we express $\mathbb{P}(|\psi_N(f,\nu)| > N^{\epsilon} u_N)$ as

$$\begin{aligned}
\mathbb{P}(|\psi_N(f,\nu)| > N^{\epsilon} u_N) &= \mathbb{P}(|\psi_N(f,\nu)| > N^{\epsilon} u_N, \Lambda_{\kappa}^{\hat{\mathbf{C}}}(\nu)) + \mathbb{P}(|\psi_N(f,\nu)| > N^{\epsilon} u_N, (\Lambda_{\kappa}^{\hat{\mathbf{C}}}(\nu))^c) \\
&\leq \mathbb{P}(|\psi_N(f,\nu)| > N^{\epsilon} u_N, \Lambda_{\kappa}^{\hat{\mathbf{C}}}(\nu)) + \mathbb{P}\left((\Lambda_{\kappa}^{\hat{\mathbf{C}}}(\nu))^c\right) \\
&\leq \mathbb{P}(|\psi_N(\bar{f},\nu)| > N^{\epsilon} u_N) + \mathbb{P}\left((\Lambda_{\kappa}^{\hat{\mathbf{C}}}(\nu))^c\right)
\end{aligned}$$

where the last inequality follows from the observation that $\frac{1}{M}\text{tr}\ f(\hat{\mathbf{C}}) = \frac{1}{M}\text{tr}\ \bar{f}(\hat{\mathbf{C}})$ on $\Lambda_\kappa^{\hat{\mathbf{C}}}(\nu)$. Moreover, the family of events $\Lambda_\kappa^{\hat{\mathbf{C}}}(\nu)$ holds with exponential high probability, which implies that $\mathbb{P}\left((\Lambda_\kappa^{\hat{\mathbf{C}}}(\nu))^c\right)$ converges towards 0 exponentially fast. Therefore, $|\psi_N(\bar{f},\nu)| \prec u_N$ implies (1.83) as expected. From now on, we thus assume that the function f is supported by $\text{Supp}\mu_{MP}^{(c)} + 2\kappa$

In order to establish (1.83), we evaluate the four terms of the righhandside of (1.15).

1.4.1 Step 1: Evaluation of $\mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)}$

We evaluate this term using the Helffer-Sjöstrand formula. We keep the notations of paragraphs 1.2.5 and 1.2.2: we assume that the support of f is included in $[a_1, a_2]$ with $a_1 = (1-\sqrt{c_N})^2 - 2\kappa$ and $a_2 = (1+\sqrt{c_N})^2 + 2\kappa$. Moreover, the resolvent of the matrix $\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}$ is denoted $\mathbf{Q}_N(z)$ (we omit to mention that the matrices depend on ν), and $\beta_N(z)$ represents $\mathbb{E}((\mathbf{Q}_N(z))_{mm})$ for each m. We also denote by $\epsilon_N(z)$ the error term defined by (1.33) which satisfies $|\epsilon_N(z)| \leq \frac{1}{M^2}P_1(|z|)P_2(\frac{1}{\text{Im}z})$ on $\mathbb{C}^+$ for some nice polynomials P_1 and P_2. Then, for $k \geq \deg(P_2)$, it holds that

$$\mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)} = \frac{1}{\pi}\text{Re}\int_{\mathcal{D}} \bar{\partial}\Phi_k(f)(z)(\beta_N(z) - t_N(z))\,\mathrm{d}x\,\mathrm{d}y$$

where $\mathcal{D}$ is defined as in paragraph 1.2.5. $\int_{\mathcal{D}} \left|\bar{\partial}\Phi_k(f)(z)\right| P_1(|z|)P_2(\frac{1}{\text{Im}z})\,\mathrm{d}x\,\mathrm{d}y$ is finite, and by (1.33), the following bound holds:

$$\left|\mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| \leq \frac{1}{M^2}\int_{\mathcal{D}} \left|\bar{\partial}\Phi_k(f)(z)\right| P_1(|z|)P_2(\frac{1}{\text{Im}z})\,\mathrm{d}x\,\mathrm{d}y \leq \frac{C}{B^2}$$

for some nice constant C. We have therefore established the following result.

Lemma 1.6. *There exists a nice constant C such that, for each ν,*

$$\left|\mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| \leq \frac{C}{B^2}. \tag{1.85}$$

1.4.2 Step 2: Evaluation of $\frac{1}{M}\text{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right]$

In order to evaluate the above term, we use the Gaussian concentration inequality introduced in Paragraph 1.2.3. We recall that $\tilde{\mathbf{C}}$ can be interpreted as a function of $(\mathbf{X}, \mathbf{X}^*)$ (see (1.61))). Therefore, $\frac{1}{M}\text{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)$ can be written as $g(\mathbf{X}, \mathbf{X}^*)$ for some real valued function g. We establish in the following that g is $\mathcal{O}(\frac{1}{B})$–Lipschitz, which in turn, will imply that

$$\left|\frac{1}{M}\text{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\text{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right]\right| \prec \frac{1}{B}. \tag{1.86}$$

For this, we evaluate

$$\|\nabla g(\mathbf{X}, \mathbf{X}^*)\|^2 = \sum_{i,j}\left|\frac{\partial g}{\partial X_{i,j}}\right|^2 + \left|\frac{\partial g}{\partial \overline{X_{i,j}}}\right|^2 = 2\sum_{i,j}\left|\frac{\partial g}{\partial X_{i,j}}\right|^2. \tag{1.87}$$

Using classic identities for the derivation of Hermitian matrices, we obtain that

$$\frac{1}{M}\frac{\partial \operatorname{tr}\, f(\tilde{\mathbf{C}})}{\partial X_{ij}} = \frac{1}{M}\operatorname{tr}\, \left(f'(\tilde{\mathbf{C}})\frac{\partial \tilde{\mathbf{C}}}{\partial X_{ij}} \right)$$

Straightforward calculations lead to

$$\sum_{i,j} \left| \frac{\partial g}{\partial X_{i,j}} \right|^2 = \frac{1}{M^2(B+1)^2} \sum_{i=1}^{M} \left(f'(\tilde{\mathbf{C}})(\mathbf{X}+\mathbf{\Gamma})(\mathbf{I}+\mathbf{\Phi}_i)(\mathbf{X}+\mathbf{\Gamma})^* f'(\tilde{\mathbf{C}}) \right)_{ii}.$$

Using $\sup_i \|\mathbf{I}+\mathbf{\Phi}_i\| \leq C$ for some nice constant C as well as $\tilde{\mathbf{C}} = \frac{1}{B+1}(\mathbf{X}+\mathbf{\Gamma})(\mathbf{X}+\mathbf{\Gamma})^*$, we obtain immediately that

$$\sum_{i,j} \left| \frac{\partial g}{\partial X_{i,j}} \right|^2 \leq \frac{C}{B^2}\,\frac{1}{M}\operatorname{tr}\, \left(f'^2(\tilde{\mathbf{C}})\tilde{\mathbf{C}} \right).$$

As $f \in C^\infty$ and is compactly supported, the function $\lambda \to \lambda\, f'^2(\lambda)$ is bounded by some constant, and there exists a nice constant C such that

$$\|\nabla g(\mathbf{X}, \mathbf{X}^*)\|^2 \leq \frac{C}{B^2}.$$

This proves that g is $\mathcal{O}(\frac{1}{B})$–Lipschitz. Paragraph 1.2.3 thus leads to (1.86).

1.4.3 Step 3: Evaluation of $\frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\operatorname{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)$

The goal of this paragraph is to establish the following Proposition.

Proposition 1.2. *Let $\tilde{D}_N$ the distribution supported by $\mathrm{Supp}(\mu_{MP}^{(c_N)})$ with Stieltjes transform*

$$\tilde{p}_N(z) = (z\, t_N(z))' = \frac{(z\, t_N(z)\, \tilde{t}_N(z))^2}{1 - c(z\, t_N(z)\, \tilde{t}_N(z))^2}. \tag{1.88}$$

Then, if we denote $< \tilde{D}_N, f >$ by $\tilde{\phi}_N(f)$, we have

$$\left| \frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\operatorname{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \left(\frac{1}{2M}\sum_{m=1}^{M} \frac{s_m''(\nu)}{s_m(\nu)} \right) \tilde{\phi}_N(f)\, v_N\, \mathbf{1}_{\alpha>2/3} \right| \prec u_N. \tag{1.89}$$

Remark 1.6. *(1.89) implies that $\left| \frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\operatorname{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) \right| \prec \frac{1}{B}$ if $\alpha \leq 2/3$. If $\alpha > 2/3$, the dominant term of $\frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\operatorname{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)$ is the deterministic $\mathcal{O}\left(\frac{B}{N}\right)^2$ term $\left(\frac{1}{M}\sum_{m=1}^{M} \frac{s_m''(\nu)}{s_m(\nu)}\right) \tilde{\phi}_N(f)\, v_N$, and its substraction from $\frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \frac{1}{M}\operatorname{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)$ allows to retrieve a term stochastically dominated by u_N.*

Remark 1.7. *We notice that (1.80) leads immediately to*

$$\left| \frac{1}{M}\operatorname{tr}\, f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\operatorname{tr}\, f(\tilde{\mathbf{C}})(\nu) \right| \prec \frac{B}{N} + \frac{1}{\sqrt{B}} \tag{1.90}$$

an approximation which is considerably more pessimistic than (1.89). As seen below, the derivation of (1.89) is rather demanding, and is based on subtle effects. In order to understand why (1.90) can *be improved, we consider the simple case $f(\lambda) = \log \lambda$. We thus have*

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})(\nu) = \frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right)$$

which depends only on the estimators $(\hat{s}_m(\nu))_{m=1,\ldots,M}$. We just provide a brief analysis of the above term. For this, we first remark that it is possible to study $\frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right)$ on the event $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ defined by (1.77). For each m, we expand around s_m the logarithm up to the second order, and obtain that

$$\frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right) = -\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)\frac{1}{s_m} + \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2}\left(\frac{\hat{s}_m - s_m}{\theta_m}\right)^2 \tag{1.91}$$

where for each m, θ_m is located between s_m and $\hat{s}_m$. Lemma 1.16 allows to conclude that the second term of the right hand side of (1.91) is dominated by $\frac{1}{B} + \left(\frac{B}{N}\right)^4 = \mathcal{O}(u_N)$ term. In order to evaluate the first term of the r.h.s. of (1.91), we note that (1.159) leads to

$$\frac{1}{M}\sum_{m=1}^{M}\left(\mathbb{E}(\hat{s}_m - s_m)\right)\frac{1}{s_m} = \frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m}{s_m}v_N + \mathcal{O}\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right).$$

As $\frac{\hat{s}_m}{s_m} = \frac{\mathbf{x}_m(\mathbf{I}+\mathbf{\Phi}_m)\mathbf{x}_m^*}{B+1}$ we finally remark that

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{s}_m - \mathbb{E}(\hat{s}_m)}{s_m}$$

can be interpreted as a recentered quadratic form of the $M(B+1)$–dimensional vector $\mathbf{x} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_M^T)^T$. The stochastic domination relation

$$\left|\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{s}_m - \mathbb{E}(\hat{s}_m)}{s_m}\right| \prec \frac{1}{B}$$

then follows from the Hanson-Wright inequality. Putting all the pieces together, and using that $\frac{1}{B} + \mathcal{O}\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right) = \mathcal{O}(u_N)$ and that $v_N = o(u_N)$ if $\alpha < 2/3$, we obtain that

$$\left|\frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right) + \frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m(\nu)}{s_m(\nu)}v_N\,\mathbf{1}_{\alpha>2/3}\right| \prec u_N.$$

Comparing this result with (1.89), we deduce that $< \tilde{D}_N, f >= -1$. We just check this formula directly. For this, we notice that function $z \to \log z$ is holomorphic inside a neighbourhood of the interval $[a_1, a_2]$. We consider the expression (1.45) of $< \tilde{D}_N, f >$ and remark that if $(\partial\mathcal{R}_\epsilon)_-$ denotes the negatively oriented contour

$$(\partial\mathcal{R}_\epsilon)_- = \{\lambda \pm i\epsilon, \lambda \in [a_1, a_2]\} \cup \{a_1 + iy, y \in [-\epsilon, \epsilon]\} \cup \{a_2 + iy, y \in [\epsilon, -\epsilon]\}$$

then, by (1.45), $< \tilde{D}_N, f >$ can also be written as the contour integral

$$< \tilde{D}_N, f >= \lim_{\epsilon\to 0}\frac{1}{2i\pi}\int_{(\partial\mathcal{R}_\epsilon)_-}\log z\,\tilde{p}_N(z)\,dz.$$

But, the above contour integral does not depend on ϵ, so that for each ϵ, we have

$$< \tilde{D}_N, f >= \frac{1}{2i\pi} \int_{(\partial \mathcal{R}_\epsilon)_-} \log z \, \tilde{p}_N(z) \, dz.$$

Using the expression of $\tilde{p}_N(z)$ and the integration by parts trick, we get that

$$< \tilde{D}_N, f >= -\frac{1}{2i\pi} \int_{(\partial \mathcal{R}_\epsilon)_-} t_N(z) \, dz.$$

Taking the limit $\epsilon \to 0$, and using the Stieltjes inversion formula for the Marcenko-Pastur distribution $\mu_{MP}^{(c_N)}$, we finally obtain that

$$< \tilde{D}_N, f >= -\frac{1}{\pi} \lim_{\epsilon \to 0} \int_{a_1}^{a_2} \mathrm{Im}(t_N(\lambda + i\epsilon)) \, d\lambda = -\mu_{MP}^{(c_N)}([a_1, a_2]) = -1$$

which is the expected result.

Proof. We now establish (1.89). In order to simplify the notations, we put

$$\tilde{r}_N(\nu) = \frac{1}{2M} \sum_{m=1}^{M} \frac{s''_m(\nu)}{s_m(\nu)}. \tag{1.92}$$

The Helffer-Sjöstrand formula implies that

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}) - \tilde{r}_N(\nu)\, \tilde{\phi}_N(f)\, v_N\, \mathbf{1}_{\alpha > 2/3} = \frac{1}{\pi}\mathrm{Re} \int_{\mathcal{D}} \mathrm{d}x \, \mathrm{d}y \, \bar{\partial}\Phi_k(f)(z) \left[\frac{1}{M} (\mathrm{tr}\ \hat{\mathbf{Q}}(z) - \mathrm{tr}\ \tilde{\mathbf{Q}}(z)) - \tilde{r}_N(\nu)\, \tilde{p}_N(z)\, v_N\, \mathbf{1}_{\alpha > 2/3} \right].$$

Reduction to the study of $\int_{\mathcal{D}} \mathrm{d}x \, \mathrm{d}y \, \bar{\partial}\Phi_k(f)(z) \frac{1}{M} \sum_{m=1}^{M} (z\mathbf{Q})'_{mm} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1 \right)$

We define

$$\zeta = \int_{\mathcal{D}} \mathrm{d}x \, \mathrm{d}y \, \bar{\partial}\Phi_k(f)(z) \frac{1}{M} \sum_{m=1}^{M} (z\mathbf{Q})'_{mm} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1 \right) \tag{1.93}$$

where we recall that the row vectors $(\mathbf{x}_m)_{m=1,\ldots,M}$ are the rows of the i.i.d. matrix $\mathbf{X}$. We establish in this paragraph that

$$\left| \int_{\mathcal{D}} \mathrm{d}x \, \mathrm{d}y \, \bar{\partial}\Phi_k(f)(z) \left(\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \tilde{r}_N(\nu)\, \tilde{p}_N(z)\, v_N\, \mathbf{1}_{\alpha > 2/3} \right) - \zeta \right| \prec u_N. \tag{1.94}$$

It turns out that by Lemma 1.7 and Lemma 1.9 in Paragraph 1.4.3 below, ζ satisfies the key properties:

$$|\zeta| \leq |\zeta - \mathbb{E}\zeta| + |\mathbb{E}\zeta| \prec \frac{1}{B}.$$

(1.89) will then follow directly from (1.94).

Plugging in the integral expression of ζ, and using the expression (1.88), we get:

$$\left|\int_{\mathcal{D}} \bar{\partial}\Phi_k(f)(z)\left(\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\} - \tilde{r}_N(\nu)\ \tilde{p}_N(z)\ v_N\ \mathbf{1}_{\alpha>2/3}\right)\mathrm{d}x\,\mathrm{d}y - \zeta\right|$$
$$= \left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\bar{\partial}\Phi_k(f)(z)\times\right.$$
$$\left.\left(\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\} - \tilde{r}_N\ (zt_N(z))'\ v_N\ \mathbf{1}_{\alpha>2/3} - \frac{1}{M}\sum_{m=1}^{M}(z\mathbf{Q})'_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right)\right)\right|.$$

We recall the definition of $\boldsymbol{\Theta} := \hat{\mathbf{C}} - \tilde{\mathbf{C}}$ from (1.74). We will proceed in three steps, which, in turn, will imply (1.94):

1.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\left(\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\} + \frac{1}{M}\mathrm{tr}\ \{\mathbf{Q}^2\boldsymbol{\Theta}\}\right)\right| \prec u_N \tag{1.95}$$

2.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\left(\frac{1}{M}\ \mathrm{tr}\ \{\mathbf{Q}^2\boldsymbol{\Theta}\} - 2\,\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2}-\mathbf{I})\right)\right| \prec u_N \tag{1.96}$$

3.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\times\right.$$
$$\left(2\,\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\mathbf{I}-\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2}) - \tilde{r}_N\ (zt_N(z))'\ v_N\ \mathbf{1}_{\alpha>2/3} -\right.$$
$$\left.\left.\frac{1}{M}\sum_{m=1}^{M}(z\mathbf{Q})'_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right)\right)\right| \prec u_N \tag{1.97}$$

Step 1. Using the well-known identity $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{B}-\mathbf{A})\mathbf{A}^{-1}$, we express $\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}$ as:

$$\hat{\mathbf{Q}} - \tilde{\mathbf{Q}} = -\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}. \tag{1.98}$$

We claim that it is possible to approximate tr $\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}$ by tr $\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}$. Indeed, we have

$$\begin{aligned}
&|\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}} - \mathrm{tr}\ \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}| \\
&\quad = |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}} - \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}} + \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}} - \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q} + \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q} - \mathrm{tr}\ \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}| \\
&\quad \le |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}} - \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}| + |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}} - \mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}| + |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q} - \mathrm{tr}\ \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}| \\
&\quad := T_1 + T_2 + T_3.
\end{aligned}$$

The following rough bounds are enough to control T_1 (we used (1.24) to control the norm of the resolvents):

$$T_1 = |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}(\hat{\mathbf{Q}}-\tilde{\mathbf{Q}})| = |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}| \le M\|\tilde{\mathbf{Q}}\|^2\|\hat{\mathbf{Q}}\|\|\|\boldsymbol{\Theta}\|^2 \le \frac{1}{\mathrm{Im}^3 z}M\|\boldsymbol{\Theta}\|^2.$$

Concerning T_2 and T_3, we write similarly that $\tilde{\mathbf{Q}} - \mathbf{Q} = -\tilde{\mathbf{Q}}\tilde{\mathbf{\Delta}}\mathbf{Q}$, and obtain that

$$T_2 = |\mathrm{tr}\ \tilde{\mathbf{Q}}\mathbf{\Theta}\tilde{\mathbf{Q}} - \mathrm{tr}\ \tilde{\mathbf{Q}}\mathbf{\Theta}\mathbf{Q}| \leq M\|\tilde{\mathbf{Q}}\|^2\|\mathbf{Q}\|\|\tilde{\mathbf{\Delta}}\|\mathbf{\Theta}\| \leq \frac{1}{\mathrm{Im}^3 z} M\|\tilde{\mathbf{\Delta}}\|\|\mathbf{\Theta}\|$$

$$T_3 = |\mathrm{tr}\ \tilde{\mathbf{Q}}\mathbf{\Theta}\mathbf{Q} - \mathrm{tr}\ \mathbf{Q}\mathbf{\Theta}\mathbf{Q}| \leq M\|\tilde{\mathbf{Q}}\|\|\mathbf{Q}\|^2\|\tilde{\mathbf{\Delta}}\|\mathbf{\Theta}\| \leq \frac{1}{\mathrm{Im}^3 z} M\|\tilde{\mathbf{\Delta}}\|\|\mathbf{\Theta}\|.$$

Plugging these estimations into the left hand side of (1.95), we obtain that

$$\begin{aligned}
\left| \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \overline{\partial}\Phi_k(f)(z) \left(\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \frac{1}{M}\mathrm{tr}\ \{\mathbf{Q}^2\mathbf{\Theta}\} \right) \right| & \\
\leq \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial}\Phi_k(f)(z)| \frac{1}{M}\ (T_1 + T_2 + T_3) & \\
\leq C(\|\mathbf{\Theta}\|^2 + 2\|\tilde{\mathbf{\Delta}}\|\mathbf{\Theta}\|). &
\end{aligned}$$

Moreover, the concentration results (1.80) for $\|\mathbf{\Theta}\|$ and (1.71) for $\|\tilde{\mathbf{\Delta}}\|$ from Proposition 1.1, imply that

$$\|\mathbf{\Theta}\|^2 + 2\|\mathbf{\Theta}\|\|\tilde{\mathbf{\Delta}}\| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3 = u_N.$$

This finally establishes (1.95).

Step 2. We claim that:

$$\left\| \mathbf{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \right) \right\| \prec u_N. \tag{1.99}$$

We recall that $\hat{\mathbf{S}}$ can be written using the definition (1.12) of $\tilde{\mathbf{C}}$, and use the decomposition (1.70) of $\tilde{\mathbf{C}}$ from Corollary 1.2. Using these results, we get that

$$\hat{\mathbf{S}} = \mathbf{D}^{1/2}\, \tilde{\mathbf{C}}\, \mathbf{D}^{1/2} = \mathbf{D}^{1/2} \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}} \right) \mathbf{D}^{1/2}.$$

Plugging this expression of $\hat{\mathbf{S}}$ into (1.75), we obtain easily that

$$\begin{aligned}
\mathbf{\Theta} =& (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}) \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}} \right) \mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} \\
& + \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}} \right) (\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \\
:=& \mathbf{\Theta}_1 + \mathbf{\Theta}_2.
\end{aligned}$$

As $\tilde{\mathbf{\Delta}}$ is a negligible quantity, one should expect that the leading quantity in $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ is respectively $(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}$ and $\frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})$. To prove it, write:

$$\begin{aligned}
\left\| \mathbf{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} \right\| & \\
= \left\| (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\tilde{\mathbf{\Delta}}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} \right\| & \\
\leq \|\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\|\|\tilde{\mathbf{\Delta}}\|\|\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\|. &
\end{aligned} \tag{1.100}$$

$\tilde{\mathbf{\Delta}}$ is controlled by (1.71) from Corollary 1.2, and $\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}$ is controlled by (1.78) from Lemma 1.5 (it is a diagonal matrix which elements are stochastically dominated by Lemma 1.5). Moreover, from Lemma 1.5, it holds that $\|\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\| \prec 1$. Combining these estimates into (1.100), one gets:

$$\left\|\boldsymbol{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\right\| \prec \left(\frac{1}{\sqrt{B}} + \frac{B^2}{N^2}\right)\frac{B}{N} = \mathcal{O}(u_N). \tag{1.101}$$

Using that $\|\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2}$ as well as (1.29) from Paragraph 1.2.2 to control the norm of $\mathbf{X}\mathbf{X}^*/(B+1)$, one can further approximate $(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}$ by $(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}$. In particular, it is easy to check that

$$\left\|\boldsymbol{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right\| \prec u_N. \tag{1.102}$$

Similarly for $\boldsymbol{\Theta}_2$, one would obtain:

$$\left\|\boldsymbol{\Theta}_2 - \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right\| \prec u_N. \tag{1.103}$$

Combining (1.102) and (1.103), we obtain (1.99). To finish the proof of Step 2, it remains to consider tr $\mathbf{Q}^2\boldsymbol{\Theta}$ and prove (1.96). Remark that $\mathbf{X}\mathbf{X}^*/(B+1)$ and its resolvent $\mathbf{Q}$ commutes.

$$\begin{aligned}\operatorname{tr}\,\mathbf{Q}^2&\left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\\ &= 2\operatorname{tr}\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\end{aligned} \tag{1.104}$$

Therefore, using (1.104):

$$\begin{aligned}&\left|\frac{1}{M}\operatorname{tr}\,\mathbf{Q}^2\boldsymbol{\Theta} - 2\,\frac{1}{M}\operatorname{tr}\,\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\right|\\ &\quad= \left|\frac{1}{M}\,\operatorname{tr}\,\mathbf{Q}^2\boldsymbol{\Theta} - \frac{1}{M}\,\operatorname{tr}\,\mathbf{Q}^2\left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right|\\ &\quad\leq \|\mathbf{Q}\|^2\left\|\boldsymbol{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\|\end{aligned} \tag{1.105}$$

so that the left hand side of (1.99) is recognised in the right hand side of (1.105). We can finally prove (1.96) by following the same idea as in Step 1:

$$\begin{aligned}&\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\,\frac{1}{M}\left(\operatorname{tr}\,\{\mathbf{Q}^2\boldsymbol{\Theta}\} - 2\operatorname{tr}\,\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\right)\right|\\ &\quad\leq \left\|\boldsymbol{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\|\\ &\qquad\times \underbrace{\int_{\mathcal{D}} |\overline{\partial}\Phi_k(f)(z)|\frac{1}{\operatorname{Im}^2 z}\,\mathrm{d}x\,\mathrm{d}y}_{<+\infty}.\end{aligned}$$

This proves (1.96) and ends Step 2.

Step 3. By definition of the resolvent, the following identity holds $\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} - z\mathbf{I}_M\right)\mathbf{Q}(z) = \mathbf{I}_M$, which leads to the so-called resolvent identity:

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q} = \mathbf{I}_M + z\mathbf{Q}. \tag{1.106}$$

Using (1.106) as well the identity $\mathbf{Q}'(z) = \mathbf{Q}^2(z)$ one can write:

$$\frac{1}{M}\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\mathbf{I}-\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}) = \frac{1}{M}\mathrm{tr}\ (\mathbf{I}+z\mathbf{Q})\mathbf{Q}(\mathbf{I}-\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2})$$
$$= \frac{1}{M}\sum_{m=1}^{M}(z\mathbf{Q})'_{mm}\left(1-\sqrt{\frac{s_m}{\hat{s}_m}}\right). \tag{1.107}$$

To handle $1-\sqrt{\frac{s_m}{\hat{s}_m}}$ we use the following Taylor expansion: define the function h by $h(u) = 1-\frac{1}{\sqrt{u}}$, with $h'(u) = \frac{1}{2}\frac{1}{u^{3/2}}$ and $h''(u) = -\frac{3}{4}\frac{1}{u^{5/2}}$. A Taylor expansion to the second order of h around 1 provides:

$$h\left(\frac{\hat{s}_m}{s_m}\right) = h(1) + \left(\frac{\hat{s}_m}{s_m}-1\right)h'(1) + \frac{1}{2}\left(\frac{\hat{s}_m}{s_m}-1\right)^2 h''(\theta_m)$$
$$= \frac{1}{2s_m}(\hat{s}_m - s_m) + \frac{1}{2}\frac{h''(\theta_m)}{s_m^2}(\hat{s}_m - s_m)^2$$

where θ_m is some random quantity between $\hat{s}_m$ and s_m. Therefore (1.107) becomes

$$\frac{1}{M}\ \mathrm{tr}\ \left((z\mathbf{Q})'(\mathbf{I}-\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2})\right)$$
$$= \frac{1}{M}\ \mathrm{tr}\ \left((z\mathbf{Q})'\mathrm{dg}\left(\frac{\hat{s}_m - s_m}{2s_m} + \frac{1}{2}\frac{h''(\theta_m)(\hat{s}_m - s_m)^2}{s_m^2} : m \in \{1,\ldots,M\}\right)\right).$$

Lemma 1.3 implies that the set $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ defined by (1.77) holds with exponentially high probability. Therefore, it is sufficient to study the term $\frac{1}{M}$ tr $(z\mathbf{Q})'(\mathbf{I}-\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2})$ on the event $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$. If $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ holds, θ_m belongs to $[\underline{s}, \bar{s}] + \epsilon$ for each $m \in \{1,\ldots,M\}$, and $\sup_{m\geq 1}|h''(\theta_m)|$ is bounded by a nice constant. Moreover, as $\inf_\nu \inf_{m\geq 1} s_m(\nu)$ is bounded away from zero, there exists a nice constant C for which the inequality

$$\left|\frac{1}{M}\ \mathrm{tr}\ \left((z\mathbf{Q})'\mathrm{dg}\left(\frac{1}{2}\frac{h''(\theta_m)(\hat{s}_m - s_m)^2}{s_m^2} : m \in \{1,\ldots,M\}\right)\right)\right|$$
$$\leq C(\|\mathbf{Q}\| + z\|\mathbf{Q}\|^2)\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)^2 \leq C(z)\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)^2$$

holds on $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$, where we recall that $C(z)$ can be written as $P_1(|z|)P_2(\frac{1}{\mathrm{Im} z})$ for some nice polynomials P_1 and P_2. Following again the same argument as in Step 1, we obtain that

$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\bar{\partial}\Phi_k(f)(z)\left\{\frac{1}{M}\ \mathrm{tr}\ (z\mathbf{Q})'(\mathbf{I}-\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2})\right.\right.$$
$$\left.\left. -\frac{1}{M}\mathrm{tr}\ (z\mathbf{Q})'\mathrm{dg}\left(\frac{\hat{s}_m - s_m}{2s_m} : m \in \{1,\ldots,M\}\right)\right\}\right|$$
$$\leq C\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)^2$$

on $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ provided $k \geq \mathrm{Deg}(P_2)$. Lemma 1.16 in Appendix implies that

$$\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)^2 \prec \frac{1}{B} + \frac{B^4}{N^4} = \mathcal{O}(u_N).$$

We have thus shown that

$$\Big| \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y \overline{\partial} \Phi_k(f)(z) \left\{ \frac{1}{M} \mathrm{tr}\ (z\mathbf{Q})'(\mathbf{I} - \hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2}) \right. \\ \left. - \frac{1}{M}\ \mathrm{tr}\ (z\mathbf{Q})' \mathrm{dg} \left(\frac{\hat{s}_m - s_m}{2s_m} : m \in \{1, \ldots, M\} \right) \right\} \Big| \\ \prec u_N.$$

We denote by $\eta_N(z)$ the term defined by

$$\eta_N(z) = \frac{1}{M} \sum_{m=1}^{M} (z\mathbf{Q})'_{mm} \left(\frac{\hat{s}_m - s_m}{s_m} - \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1 \right) \right) - \tilde{r}_N(z)(zt_N(z))' v_N \mathbf{1}_{\alpha > 2/3} \tag{1.108}$$

and define δ_N as

$$\delta_N = \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \overline{\partial} \Phi_k(f)(z)\, \eta_N(z).$$

In order to establish (1.97), it is sufficient to prove that $|\delta_N| \prec u_N$. For this, we first remark that $\hat{s}_m = s_m \frac{\mathbf{x}_m(\mathbf{I}+\mathbf{\Phi}_m)\mathbf{x}_m^*}{B+1}$, so that $\eta_N(z)$ can also be written as

$$\eta_N(z) = \frac{1}{M} \sum_{m=1}^{M} (z\mathbf{Q})'_{mm} \frac{\mathbf{x}_m \mathbf{\Phi}_m \mathbf{x}_m^*}{B+1} - \tilde{r}_N(zt_N(z))' v_N \mathbf{1}_{\alpha > 2/3}. \tag{1.109}$$

We express $\eta_N(z)$ as $\eta_N(z) = \eta_{1,N}(z) + \eta_{2,N}(z) + \eta_{3,N}(z)$ where $(\eta_{i,N})_{i=1,2,3}$ are defined by

$$\eta_{1,N}(z) = \frac{1}{M} \sum_{m=1}^{M} (z\mathbf{Q})'_{mm} \left(\frac{\mathbf{x}_m \mathbf{\Phi}_m \mathbf{x}_m^*}{B+1} - \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Phi}_m \right)$$

$$\eta_{2,N}(z) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}[(z\mathbf{Q})'_{mm}] \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Phi}_m - \tilde{r}_N(zt_N(z))' v_N \mathbf{1}_{\alpha > 2/3}$$

$$\eta_{3,N}(z) = \frac{1}{M} \sum_{m=1}^{M} ((z\mathbf{Q})'_{mm})^{\circ} \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Phi}_m$$

and denote by $(\delta_{i,N})_{i=1,2,3}$ the contributions of $(\eta_{i,N})_{i=1,2,3}$ to δ_N. We recall the definition (1.19) of $((z\mathbf{Q})'_{mm})^{\circ}$. In order to evaluate $\delta_{1,N}$, we note that $|(z\mathbf{Q})'_{mm}| = |Q_{mm} + z\mathbf{Q}^2_{mm}| \leq C(z)$ and that

$$|\eta_{1,N}(z)| \leq C(z) \sup_{m=1,\ldots,M} \left| \frac{\mathbf{x}_m \mathbf{\Phi}_m \mathbf{x}_m^*}{B+1} - \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Phi}_m \right|.$$

Therefore, for k large enough, $\delta_{1,N}$ satisfies $|\delta_{1,N}| \leq C \sup_{m=1,\ldots,M} \left| \frac{\mathbf{x}_m \mathbf{\Phi}_m \mathbf{x}_m^*}{B+1} - \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Phi}_m \right|$. The Hanson-Wright inequality as well as the bound (1.55) of the Frobenius norm of $\mathbf{\Phi}_m$ imply that $|\delta_{1,N}| \prec u_N$. We now evaluate $\delta_{2,N}$. For this, we notice that the results reviewed in Paragraph 1.2.2 imply that $\mathbb{E}(z\mathbf{Q})'_{mm} = (z\beta_N(z))' = (zt_N(z))' + (z\epsilon_N(z))'$ where $|(z\epsilon_N(z))'| \leq \frac{C(z)}{M^2}$. Therefore, using (1.54), we obtain that

$$\begin{aligned} \eta_{2,N}(z)) = & \left(\frac{1}{M} \sum_{m=1}^{2M} \frac{s''_m}{s_m} \right) (zt_N(z))'\, v_N + \epsilon_{1,N}(z) - \tilde{r}_N\, (zt_N(z))'\, v_N\, \mathbf{1}_{\alpha > 2/3} \\ = & \tilde{r}_N\, (zt_N(z))'\, v_N\, \mathbf{1}_{\alpha \leq 2/3} + \epsilon_{1,N}(z) \end{aligned}$$

where $\epsilon_{1,N}(z)$ satisfies $|\epsilon_{1,N}(z)| \leq C(z)\, u_N$. We then deduce that $|\eta_{2,N}(z)| \leq C(z) u_N$ because if

$\alpha \leq 2/3$, $v_N \leq u_N$. This implies that $|\delta_{2,N}| = \mathcal{O}(u_N)$. In order to address $\delta_{3,N}$, we interpret $\delta_{3,N}$ as a function g of $(\mathbf{X}, \mathbf{X}^*)$, and use the Gaussian concentration inequality presented in Paragraph 1.2.3. In particular, we satisfy that

$$\|\nabla g\| \leq C \frac{1}{\sqrt{B}} \left(\frac{B}{N}\right)^2 = o(u_N).$$

As $\mathbb{E}(\delta_{3,N}) = 0$, this leads immediately to $|\delta_{3,N}| \prec u_N$. We just check that

$$\sum_{i,j} \left|\frac{\partial g}{\partial X_{ij}}\right|^2 \leq C \frac{1}{B} \left(\frac{B}{N}\right)^4. \tag{1.110}$$

For this, we express $(z\mathbf{Q})'_{mm}$ as $(z\mathbf{Q})'_{mm} = Q_{mm} + z\mathbf{Q}^2_{mm}$ and notice that

$$\begin{aligned}
\frac{\partial Q_{mm}}{\partial X_{ij}} &= - Q_{mi} \left(\frac{\mathbf{X}^*}{B+1}\mathbf{Q}\right)_{jm} \\
\frac{\partial \mathbf{Q}^2_{mm}}{\partial X_{ij}} &= - (\mathbf{Q}^2)_{mi} \left(\frac{\mathbf{X}^*}{B+1}\mathbf{Q}\right)_{jm} - Q_{mi} \left(\frac{\mathbf{X}^*}{B+1}\mathbf{Q}^2\right)_{jm}.
\end{aligned}$$

Using the Jensen inequality, we obtain that

$$\left|\frac{\partial g}{\partial X_{ij}}\right|^2 \leq \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\bar{\partial}\Phi_k(f)(z)|^2 \frac{1}{M} \sum_{m=1}^{M} \left|\frac{\partial (z\mathbf{Q})'_{mm}}{\partial X_{ij}}\right|^2 \left(\frac{1}{B+1} \mathrm{tr}\, \mathbf{\Phi}_m\right)^2.$$

Summing over i, j leads to the expected evaluation of (1.110) and to $|\delta_{3,N}| \prec u_N$. This, in turn, completes the proof of (1.97) and of (1.94).

Up to the Lemma 1.7 and Lemma 1.9, Theorem 1.2 is proved. □

Proof of Lemma 1.7 and Lemma 1.9

We now establish Lemma 1.7 and Lemma 1.9.

Lemma 1.7. *The family of random variables $\zeta(\nu) - \mathbb{E}\zeta(\nu)$, $\nu \in [0,1]$ satisfies the following property:*

$$|\zeta(\nu) - \mathbb{E}\zeta(\nu)| \prec \frac{1}{B}. \tag{1.111}$$

Proof. ζ defined by (1.93) can be written as

$$\begin{aligned}
\zeta = &\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \bar{\partial}\Phi_k(f)(z) \frac{1}{M} \sum_{m=1}^{M} Q_{mm} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right) + \\
&\qquad \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \bar{\partial}\Phi_k(f)(z) \frac{1}{M} \sum_{m=1}^{M} z(\mathbf{Q}^2)_{mm} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right) \\
:= &\ \zeta_1 + \zeta_2.
\end{aligned}$$

In the following, we omit to evaluate $|\zeta_1(\nu) - \mathbb{E}(\zeta_1(\nu))|$, and just establish that $|\zeta_2(\nu) - \mathbb{E}(\zeta_2(\nu))| \prec \frac{1}{B}$ using the Gaussian concentration inequality from Paragraph 1.2.3.

Recall that $\|\mathbf{x}_m\|_2^2$ is a $\chi^2_{2(B+1)}$ random variable. Therefore it is clear that:

$$\left|\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right| \prec \frac{1}{\sqrt{B}}.$$

Knowing this, the idea is to show that, conditioned on the event where the random variables $\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right)_{m=1,\dots,M}$ are localized, which holds with exponentially high probability, ζ_2 is a $\mathcal{O}(\frac{1}{B^{1-\epsilon}})$–Lipschitz function of the entries of the matrix $\mathbf{X}$ for any $\epsilon > 0$. Let $0 < \epsilon < \frac{1}{2}$, and define the family of events $A_{m,\epsilon}(\nu)$, $m = 1, \dots, M$, $\nu \in [0,1]$ given by

$$A_{m,\epsilon}(\nu) = \left\{ \frac{\|\mathbf{x}_m(\nu)\|_2^2}{B+1} \in \left[1 - \frac{B^\epsilon}{\sqrt{B}}, 1 + \frac{B^\epsilon}{\sqrt{B}}\right] \right\} \tag{1.112}$$

as well as $A_\epsilon(\nu) = \cap_{m=1}^M A_{m,\epsilon}(\nu)$. It is clear that the family of events $A_{m,\epsilon}(\nu)$, $m = 1, \dots, M$, $\nu \in [0,1]$ holds with exponentially high probability, and that the same property holds for the family $A_\epsilon(\nu), \nu \in [0,1]$. We claim that there exists a family of $\mathcal{C}^\infty$ functions $(g_{B,\epsilon})_{B\geq 1}$ satisfying

$$g_{B,\epsilon}(t) = \begin{cases} t - 1 & \text{if } t \in [1 - \frac{B^\epsilon}{\sqrt{B}}, 1 + \frac{B^\epsilon}{\sqrt{B}}] \\ 0 & \text{if } t \notin [1 - 2\frac{B^\epsilon}{\sqrt{B}}, 1 + 2\frac{B^\epsilon}{\sqrt{B}}] \end{cases}$$

and

$$\sup_t |g_{B,\epsilon}(t)| \leq C\, \frac{B^\epsilon}{\sqrt{B}}, \quad \sup_t |g'_{B,\epsilon}(t)| \leq C \tag{1.113}$$

for each B, where C is a nice constant. Indeed consider $h \in C^\infty$ such that it satisfies $|h(t)| \leq 2|t|$ for each t and

$$h(t) = \begin{cases} t & \text{if } t \in [-1, 1] \\ 0 & \text{if } t \notin [-2, 2]. \end{cases}$$

Then, it is easy to check that the family $(g_{B,\epsilon})_{B\geq 1}$ defined by

$$g_{B,\epsilon}(t) = \frac{B^\epsilon}{\sqrt{B}}\, h\left(\frac{\sqrt{B}}{B^\epsilon}\,(t-1)\right)$$

satisfies the requirements (1.113).

We define $\tilde{\zeta}_{2,\epsilon}$ by

$$\tilde{\zeta}_{2,\epsilon} = \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y \bar{\partial}\Phi_k(f)(z) \frac{1}{M} \sum_{m=1}^M (z\mathbf{Q}^2)_{mm}\, g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)$$

and notice that ζ_2 and $\tilde{\zeta}_{2,\epsilon}$ coincide on the exponentially high probability event $A_\epsilon(\nu)$. We claim that if $|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| \prec \frac{1}{B^{1-\epsilon}}$, then $|\zeta_2 - \mathbb{E}(\zeta_2)| \prec \frac{1}{B^{1-\epsilon}}$. Since ϵ is arbitrary and $B^\epsilon = \mathcal{O}(N^{\alpha\epsilon})$, Remark 1.1 will imply that $|\zeta_2 - \mathbb{E}(\zeta_2)| \prec \frac{1}{B}$. To justify this, we evaluate $\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > \frac{1}{B^{1-\epsilon}} N^\delta\right)$ for each $\delta > 0$. It holds that

$$\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > \frac{N^{\alpha\epsilon+\delta}}{B}\right) \leq \mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > \frac{N^{\alpha\epsilon+\delta}}{B}, A_\epsilon\right) + \mathbb{P}(A_\epsilon^c).$$

As $\mathbb{P}(A_\epsilon^c)$ converges towards zero exponentially, we have just to consider

$$\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > \frac{N^{\alpha\epsilon+\delta}}{B}, A_\epsilon\right)$$

and write, since ζ_2 and $\tilde{\zeta}_{2,\epsilon}$ coincide on A_ϵ,

$$\mathbb{P}\left(|\zeta_2-\mathbb{E}(\zeta_2)|>\frac{N^{\alpha\epsilon+\delta}}{B},A_\epsilon\right)=\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\zeta_2)|>\frac{N^{\alpha\epsilon+\delta}}{B},A_\epsilon\right)$$
$$\leq \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|>\frac{N^{\alpha\epsilon+\delta}}{B}-|E(\zeta_2-\tilde{\zeta}_{2,\epsilon})|,A_\epsilon\right).$$

We now prove that $|\mathbb{E}(\zeta_2-\tilde{\zeta}_{2,\epsilon})|$ converges towards 0 exponentially. For this, we notice that as ζ_2 and $\tilde{\zeta}_{2,\epsilon}$ coincide on A_ϵ, then

$$|\mathbb{E}(\zeta_2-\tilde{\zeta}_{2,\epsilon})|=\left|\mathbb{E}((\zeta_2-\tilde{\zeta}_{2,\epsilon})\mathbb{1}_{A_\epsilon^c})\right|\leq\left(\mathbb{E}\left|\zeta_2-\tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2}(\mathbb{P}(A_\epsilon^c))^{1/2}.$$

A rough evaluation of $\left(\mathbb{E}\left|\zeta_2-\tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2}$ leads to $\left(\mathbb{E}\left|\zeta_2-\tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2}\leq C$ for some nice constant C. Therefore, $\left(\mathbb{E}\left|\zeta_2-\tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2}(\mathbb{P}(A_\epsilon^c))^{1/2}$, and thus $|\mathbb{E}(\zeta_2-\tilde{\zeta}_{2,\epsilon})|$, converge towards 0 exponentially. For each N large enough, we thus have

$$\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|>\frac{N^{\alpha\epsilon+\delta}}{B}-|\mathbb{E}(\zeta_2-\tilde{\zeta}_{2,\epsilon})|,A_\epsilon\right)\leq\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|>\frac{N^{\alpha\epsilon+\delta/2}}{B},A_\epsilon\right)$$
$$\leq\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|>\frac{N^{\alpha\epsilon+\delta/2}}{B}\right).$$

We have therefore established that

$$\mathbb{P}\left(|\zeta_2-\mathbb{E}(\zeta_2)|>\frac{N^{\alpha\epsilon+\delta}}{B},A_\epsilon\right)\leq\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|>\frac{N^{\alpha\epsilon+\delta/2}}{B}\right)$$

which finally justifies that if $|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|\prec\frac{B^\epsilon}{B}$, then $|\zeta_2-\mathbb{E}(\zeta_2)|\prec\frac{B^\epsilon}{B}$.

Therefore, it remains to prove that $|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|\prec\frac{B^\epsilon}{B}$. This is true by Lemma 1.8 below. The stochastic domination relation $|\zeta_1-\mathbb{E}\zeta_1|\prec\frac{B^\epsilon}{B}$ is proved similarly. This completes the proof of Lemma 1.7. □

Lemma 1.8.

$$|\tilde{\zeta}_{2,\epsilon}-\mathbb{E}(\tilde{\zeta}_{2,\epsilon})|\prec\frac{B^\epsilon}{B}.$$

Proof. In the following, we evaluate the norm square of the gradient of $\tilde{\zeta}_{2,\epsilon}$ w.r.t. the variables $X_{i,j},X_{i,j}^*$ and just compute $\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2$ because $\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}^*}\right|^2$ is of the same order of magnitude.

We recall that

$$\frac{\partial(\mathbf{Q}^2)_{mm}}{\partial X_{ij}}=-\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right). \tag{1.114}$$

Moreover it is clear that

$$\frac{\partial}{\partial X_{ij}}\left(g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right)=\delta_{im}\frac{\overline{X_{m,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right). \tag{1.115}$$

Collecting the derivatives (1.114) and (1.115) we get after some algebra that

$$\frac{\partial}{\partial X_{ij}}\left(\sum_{m=1}^{M}(\mathbf{Q}^2)_{mm}g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right)=\frac{\overline{X_{i,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_i\|_2^2}{B+1}\right)(\mathbf{Q}^2)_{ii}$$
$$-\sum_{m=1}^{M}g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right). \quad (1.116)$$

It remains to control $\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2$. From the integral representation of $\tilde{\zeta}_{2,\epsilon}$, the derivative with respect to X_{ij} is applied only on the integrand as follows:

$$\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}=\frac{1}{M}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\frac{\partial}{\partial X_{ij}}\left(\sum_{m=1}^{M}z(\mathbf{Q}^2)_{mm}g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right).$$

Plugging in the derivative computed in (1.116) we get:

$$\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}=\frac{1}{M}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\,z\left\{\frac{\overline{X_{i,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_i\|_2^2}{B+1}\right)(\mathbf{Q}^2)_{ii}\right.$$
$$\left.-\sum_{m=1}^{M}g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right)\right\}.$$

Using the bounds of $g_{B,\epsilon}$ and $g'_{B,\epsilon}$ from inequalities (1.113), the observation that $g'_{B,\epsilon}(t)=0$ if $|t-1|\geq\frac{2B^\epsilon}{\sqrt{B}}$, and that $|z|$ is bounded on $\mathcal{D}$, one can write:

$$\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2\leq\frac{C}{M^2}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2\left|\frac{\overline{X_{i,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_i\|_2^2}{B+1}\right)(\mathbf{Q}^2)_{ii}\right.$$
$$\left.-\sum_{m=1}^{M}g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right)\right|^2$$
$$\leq\frac{C}{M^2}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2\left|\frac{\overline{X_{i,j}}}{B+1}(\mathbf{Q}^2)_{ii}\right|^2\mathbb{1}\left(\left|\frac{\|\mathbf{x}_i\|_2^2}{B+1}-1\right|\leq\frac{2B^\epsilon}{\sqrt{B}}\right)$$
$$+\frac{C}{M^2}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2\left(\frac{B^\epsilon}{\sqrt{B}}\right)^2\left|\sum_{m=1}^{M}\frac{|(\mathbf{Q}^2)_{mi}|\,|(\mathbf{X}^*\mathbf{Q})_{jm}|}{B+1}\right|^2$$
$$+\frac{C}{M^2}\int_{\mathcal{D}}\mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2\left(\frac{B^\epsilon}{\sqrt{B}}\right)^2\left|\sum_{m=1}^{M}\frac{|Q_{mi}|\,|(\mathbf{X}^*\mathbf{Q}^2)_{jm}|}{B+1}\right|^2$$
$$:=\frac{C}{M^2}(T_{ij}^{(1)}+T_{ij}^{(2)}+T_{ij}^{(3)}).$$

It remains to sum over i, j.

$$\begin{aligned}
&\sum_{i,j=1}^{M} T_{ij}^{(1)} \\
&= \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i,j=1}^{M} \left|\frac{\overline{X_{i,j}}}{B+1}(\mathbf{Q}^2)_{ii}\right|^2 \mathbb{1}\left(\left|\frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1\right| \leq \frac{2B^\epsilon}{\sqrt{B}}\right) \\
&\leq \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2\, \mathbb{1}\left(\left|\frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1\right| \leq \frac{2B^\epsilon}{\sqrt{B}}\right) \sum_{j=1}^{M} \left|\frac{\overline{X_{i,j}}}{B+1}\right|^2 \\
&= \frac{C}{B+1} \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2 \frac{\|\mathbf{x}_i\|_2^2}{B+1}\, \mathbb{1}\left(\left|\frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1\right| \leq \frac{2B^\epsilon}{\sqrt{B}}\right) \\
&\leq \frac{C}{B+1}(1 + \frac{2B^\epsilon}{\sqrt{B}}) \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2.
\end{aligned}$$

Since

$$\sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2 \leq M\|\mathbf{Q}\|^4$$

it can be written that:

$$\sum_{i,j=1}^{M} T_{ij}^{(1)} \leq C\frac{M}{B+1} \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \|\mathbf{Q}\|^4.$$

Inspecting $T_{ij}^{(2)}$, one can see that by Jensen's inequality

$$\left|\sum_{m=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right| |(\mathbf{X}^*\mathbf{Q})_{jm}|\right|^2 \leq M \sum_{m=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right|^2 |(\mathbf{X}^*\mathbf{Q})_{jm}|^2$$

so summing over i and j provides:

$$\sum_{i,j=1}^{M} T_{ij}^{(2)} \leq \frac{B^{2\epsilon}M}{(B+1)^3} \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{m=1}^{M} \left(\sum_{i=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right|^2\right) \left(\sum_{j=1}^{M} |(\mathbf{X}^*\mathbf{Q})_{jm}|^2\right).$$

Notice that since $\sum_{i=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right|^2$ is the square euclidean norm of line m of $\mathbf{Q}^2$:

$$\sum_{i=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right|^2 \leq \|\mathbf{Q}^2\|^2 \leq \|\mathbf{Q}\|^4.$$

Moreover,

$$\begin{aligned}
\sum_{m=1}^{M} \left(\sum_{j=1}^{M} |(\mathbf{X}^*\mathbf{Q})_{jm}|^2\right) &= \mathrm{tr}\ \mathbf{X}^*\mathbf{Q}\mathbf{Q}^*\mathbf{X} \\
&= (B+1)\mathrm{tr}\ \left((\mathbf{I} + z\mathbf{Q})\,\mathbf{Q}^*\right) \leq M(B+1)(\|\mathbf{Q}\| + |z|\|\mathbf{Q}\|^2)
\end{aligned}$$

therefore
$$\sum_{i,j=1}^{M} T_{ij}^{(2)} \leq B^{2\epsilon}\left(\frac{M}{B+1}\right)^2 \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2 \|\mathbf{Q}^4\|(\|\mathbf{Q}\| + |z|\|\mathbf{Q}\|^2)$$

and similarly for $T_{ij}^{(3)}$ one gets:

$$\sum_{i,j=1}^{M} T_{ij}^{(3)} \leq B^{2\epsilon}\left(\frac{M}{B+1}\right)^2 \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2 \|\mathbf{Q}^2\|(\|\mathbf{Q}\|^3 + |z|\|\mathbf{Q}\|^4).$$

Collecting the terms in $T_{ij}^{(1)}, T_{ij}^{(2)}$ and $T_{ij}^{(3)}$, and since $M/(B+1) = \mathcal{O}(1)$ by Assumption 1.3, we can write:

$$\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 \leq \frac{C}{M^2}B^{2\epsilon}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\bar{\partial}\Phi_k(f)(z)\right|^2 (\|\mathbf{Q}\|^4 + \|\mathbf{Q}^5\| + |z|\|\mathbf{Q}^6\|)$$

As $\|\mathbf{Q}\|^4 + \|\mathbf{Q}^5\| + |z|\|\mathbf{Q}^6\| \leq C(z)$, we obtain that for k large enough,

$$\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 = \mathcal{O}\left(\frac{B^{2\epsilon}}{B^2}\right)$$

as expected. □

It remains to study $\mathbb{E}[\zeta]$, and establish the following Lemma.

Lemma 1.9.
$$|\mathbb{E}\zeta| = \mathcal{O}\left(\frac{1}{B}\right).$$

Proof. As in the proof of Lemma 1.7, we only consider

$$\mathbb{E}[\zeta_2] = \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\frac{1}{M}\sum_{m=1}^{M} z\mathbb{E}\left[(\mathbf{Q}^2)_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right)\right]$$

as $\mathbb{E}[\zeta_1]$ is shown to be also $\mathcal{O}(\frac{1}{B})$ with the same argument. As $\mathbb{E}\left[\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right] = 0$, we have

$$\mathbb{E}\left[(\mathbf{Q}^2)_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right)\right] = \mathbb{E}\left[\left((\mathbf{Q}^2)_{mm} - \mathbb{E}[(\mathbf{Q}^2)_{mm}]\right)\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right)\right].$$

Apply now the Cauchy-Schwartz inequality:

$$|\mathbb{E}[\zeta_2]| \leq \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\bar{\partial}\Phi_k(f)(z)|\frac{1}{M}\sum_{m=1}^{M}|z|\sqrt{\mathrm{Var}(\mathbf{Q}^2)_{mm}}\sqrt{\mathbb{E}\left|\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right|^2}. \tag{1.117}$$

As it is clear that $\mathbb{E}\left|\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right|^2 = \mathcal{O}\left(\frac{1}{B}\right)$, it remains to control $\mathrm{Var}(\mathbf{Q}^2)_{mm} = \mathrm{Var}(\mathrm{tr}\ \mathbf{Q}^2\mathbf{e}_m\mathbf{e}_m^T)$ where $(\mathbf{e}_m)_{m=1,\ldots,M}$ is the canonical basis of $\mathbb{C}^M$. A direct application of (1.30) for $i = 2$ leads immediately to

$$\mathrm{Var}(\mathbf{Q}^2)_{mm} \leq \frac{C(z)}{B} \tag{1.118}$$

for some nice constant C. Using (1.118) in (1.117), we get that for k large enough:

$$\begin{aligned}|\mathbb{E}\zeta_2| &\leq \frac{1}{B}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\bar{\partial}\Phi_k(f)(z)|\sqrt{C(z)} \\ &\leq \frac{1}{B}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\bar{\partial}\Phi_k(f)(z)|(1+C(z)) \leq C\frac{1}{B}.\end{aligned}$$

This completes the proof of Lemma 1.9. $\square$

Remark 1.8. *We notice that, instead of using (1.15), an alternative approach to study* $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f\, d\mu_{MP}^{(c_N)}$ *could have been based on the decomposition*

$$\begin{aligned}\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)} = \frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)\right] + \\ \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)\right] - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right] + \\ \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] + \\ \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] - \int_{\mathbb{R}^+} f\,\mathrm{d}\mu_{MP}^{(c_N)}.\end{aligned} \tag{1.119}$$

The first term of the r.h.s. of (1.119) can be addressed using the Gaussian concentration inequality. However, the calculations are more complicated than the evaluation of $\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right)\right]$ *because, considered as a function of* $(\mathbf{X}, \mathbf{X}^*)$, $\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)$ *is not a Lipschitz function. Using techniques similar to those developed to evaluate* $\zeta - \mathbb{E}(\zeta)$ *(see Lemma 1.7), it could however be shown that*

$$\left|\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)\right]\right| \prec \frac{1}{B}. \tag{1.120}$$

In order to evaluate the second term of the r.h.s. of (1.119), one should prove that

$$\mathbb{E}\left[\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\left(\bar{\partial}\Phi_k(f)(z)\frac{1}{M}\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \tilde{r}_N(\nu)\ \tilde{p}_N(z)\ v_N\ \mathbf{1}_{\alpha>2/3}\right) - \zeta\right] = \mathcal{O}(u_N) \tag{1.121}$$

and $\mathbb{E}(\zeta) = \mathcal{O}(\frac{1}{B})$. *The proof of (1.121) does not appear simpler than the proof of (1.94): the 3 steps that allowed to establish (1.94) should still be used, except that the stochastic domination properties should be replaced by properties of the mathematical expectation of the various terms. However, proving stochastic domination appears simpler than showing the desired properties of the above mathematical expectations. In sum, while the use of decomposition (1.119) allows to avoid Lemma 1.7, the justification of (1.120) needs to develop tools that are similar to those of Lemma 1.7, and the proof of (1.121) tends to be more complicated than the proof of (1.94). This explains why we have chosen to use decomposition (1.15) rather than (1.119).*

1.4.4 Step 4: evaluation of $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right]$

The Helffer-Sjöstrand formula implies that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] = \frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\,\left(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z)\right)\right].$$

Therefore, we are back to evaluate $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\,\left(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z)\right)\right]$.

In order to simplify the exposition of the results of this paragraph, we introduce the following notation. If $(h_N(z))_{N\geq 1}$ is a sequence of complex-valued functions defined on $\mathbb{C}^+$ and if $(w_N)_{N\geq 1}$ is a sequence of positive real numbers, the notation $h_N(z) = \mathcal{O}_z(w_N)$ means that there exists two nice polynomials P_1 and P_2 such that $|h_N(z)| \leq w_N P_1(|z|)P_2(\frac{1}{\mathrm{Im}z})$ for each $z \in \mathbb{C}^+$.

In this paragraph, we establish the following Proposition.

Proposition 1.3. $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\,\left(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z)\right)\right]$ *can be written as*

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\,\left(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z)\right)\right] = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m}{s_m}\right)^2 p_N(z)\, v_N - \left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m}{s_m}\right)\tilde{p}_N(z)\, v_N + \mathcal{O}_z\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right). \quad (1.122)$$

The Helffer-Sjöstrand formula thus leads to the following Corollary:

Corollary 1.5. $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right]$ *is given by*

$$\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right] = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m}{s_m}\right)^2 \phi_N(f)\, v_N - \left(\frac{1}{2M}\sum_{m=1}^{M}\frac{s''_m}{s_m}\right)\tilde{\phi}_N(f)\, v_N + \mathcal{O}\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right). \quad (1.123)$$

Corollary 1.5 first implies that $\mathbb{E}\left[\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \frac{1}{M}\mathrm{Tr}\left(f(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1})\right)\right]$ is $\mathcal{O}\left(\frac{B}{N}\right)^2$, a result which is not a priori obvious. In particular, the stochastic representation (1.46) of the matrix $\tilde{\mathbf{C}}$ can be shown to provide the more pessimistic $\mathcal{O}(\frac{B}{N})$ rate of convergence. The comparison of (1.123) with (1.89) also leads to the conclusion that if $\alpha > 2/3$, the dominant $\mathcal{O}\left(\frac{B}{N}\right)^2$ deterministic term of $\frac{1}{M}\mathrm{tr}\,\left(f(\hat{\mathbf{C}}(\nu)) - f(\tilde{\mathbf{C}}(\nu)\right)$ is cancelled by the second term of the righthandside of (1.123), thus explaining the structure of the $\mathcal{O}\left(\frac{B}{N}\right)^2$ deterministic correction of $\frac{1}{M}\mathrm{tr}\,\left(f(\hat{\mathbf{C}}(\nu)) - \int f\, d\mu_{MP}^{(c_N)}\right.$. In particular, establishing (1.122) (and thus (1.123)) will complete the proof of Theorem 1.2.

Proof. The proof of (1.122) is based on the Gaussian tools reviewed in Paragraph 1.2.2, and needs long and very tedious calculations. Therefore, we just provide a sketch of proof. In particular, we justify that $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\,\left(\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z)\right)\right]$ is a $\mathcal{O}_z\left(\frac{B}{N}\right)^2$ term, but do not establish its expression (1.122).

The starting point of the proof is to express $\tilde{\mathbf{Q}} - \mathbf{Q}$ as

$$\tilde{\mathbf{Q}} - \mathbf{Q} = -\tilde{\mathbf{Q}}\tilde{\mathbf{\Delta}}\mathbf{Q} = -\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q} + \mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q} - \tilde{\mathbf{Q}}\ \tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}.$$

Therefore, $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z))\right]$ can be written as

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{Q}} - \mathbf{Q})\right] = -\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\mathbf{Q}^2\tilde{\mathbf{\Delta}})\right] + \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\mathbf{Q}^2\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}})\right] - \mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{Q}}\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}})\right]. \quad (1.124)$$

It is clear that the moduli of the second and third terms of the right hand side of (1.124) are controlled by $C(z)\mathbb{E}(\|\tilde{\mathbf{\Delta}}\|^2)$ and $C(z)\mathbb{E}(\|\tilde{\mathbf{\Delta}}\|^3)$ respectively. We now state the following useful Lemma, proved in the Appendix, which implies that these terms are $\mathcal{O}_z\left(\frac{B}{N}\right)^2$ and $\mathcal{O}_z\left(\frac{B}{N}\right)^3$ respectively.

Lemma 1.10. *For each $k \geq 1$, there exist a nice constant C depending on k such that* $\mathbb{E}\left(\|\tilde{\mathbf{\Delta}}\|^k\right) \leq C\ \left(\frac{B}{N}\right)^k$

In order to prove that $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ (\tilde{\mathbf{Q}}_N(z) - \mathbf{Q}_N(z))\right] = \mathcal{O}_z\left(\frac{B}{N}\right)^2$, we thus have to check that

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\tilde{\mathbf{\Delta}}\right] = \mathcal{O}_z\left(\frac{B}{N}\right)^2. \quad (1.125)$$

For this, we first express $\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\tilde{\mathbf{\Delta}}\right]$ as

$$\mathbb{E}\left[\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\tilde{\mathbf{\Delta}}\right] = \mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right) + \mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{X}\mathbf{\Gamma}^*}{B+1}\right) + \mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1}\right).$$

The third term of the right hand side is clearly $\mathcal{O}_z((\frac{B}{N})^2)$. We thus need to check that the first two terms are also $\mathcal{O}_z((\frac{B}{N})^2)$. We just verify this property for the first term. For this, we evaluate $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right)$ using the Gaussian tools, and take the derivative w.r.t. z to obtain the expression of $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right)$.

In order to simplify the notations, we denote by $\mathbf{W}$ the matrix $\mathbf{W} = \frac{\mathbf{X}}{\sqrt{B+1}}$, and denote by $\mathbf{w}_1 = \frac{\mathbf{x}_1}{\sqrt{B+1}}, \ldots, \mathbf{w}_M = \frac{\mathbf{x}_M}{\sqrt{B+1}}$ its M rows. In particular, the row m of the matrix $\frac{\mathbf{\Gamma}}{\sqrt{B+1}}$ coincides with $\mathbf{w}_m\mathbf{\Psi}_m$ where we recall that matrix $\mathbf{\Psi}_m$ is defined by (1.57). If $(\mathbf{e}_1, \ldots, \mathbf{e}_m)$ represents the canonical basis of $\mathbb{C}^M$, $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right)$ can be written as

$$\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right) = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\left(\mathbf{w}_m\mathbf{\Psi}_m\mathbf{W}^*\mathbf{Q}\mathbf{e}_m\right).$$

We now state the following Lemma whose proof is given in Appendix. We recall that $\beta_N(z) = \mathbb{E}((\mathbf{Q}_N(z))_{mm}$ for each m.

Lemma 1.11. *If* $\mathbf{A}$ *represents a* $(B+1)\times(B+1)$ *matrix, the following equality holds*

$$\mathbb{E}\left(\mathbf{w}_m \mathbf{A}\mathbf{W}^*\mathbf{Q}\mathbf{e}_m\right) = \frac{\beta}{1+\beta c}\frac{1}{B+1}\mathrm{tr}\ \mathbf{A} - \mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}\mathbf{A}\mathbf{W}^*\mathbf{Q}\right)^{\circ}\mathbf{Q}^{\circ}_{m,m}\right] + \frac{\beta c}{1+\beta c}\mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}\mathbf{A}\mathbf{W}^*\mathbf{Q}\right)^{\circ}\frac{1}{B+1}\mathrm{tr}\ \mathbf{Q}^{\circ}\right]. \quad (1.126)$$

Using (1.22) in the case $\mathbf{s}_\mu(z) = \beta(z)$ as well as (1.33), we easily obtain that $\frac{\beta}{1+\beta c} = \frac{t}{1+ct} + \epsilon_1(z) = -zt(z)\tilde{t}(z) + \epsilon_1(z)$ where $\epsilon_1(z) = \mathcal{O}_z(\frac{1}{B^2})$. Moreover, it follows from (1.37) that $\epsilon_1'(z)$ is also a $\mathcal{O}_z(\frac{1}{B^2})$. We now use (1.126) for $\mathbf{A} = \mathbf{\Psi}_m$, and differentiate (1.126) for $\mathbf{A} = \mathbf{\Psi}_m$ w.r.t. z. Using the Schwartz inequality the inequalities (1.30) and (1.31), and (1.58), we obtain immediately that

$$\mathbb{E}\left(\mathbf{w}_m \Psi_m \mathbf{W}^*\mathbf{Q}\mathbf{e}_m\right) = -\left(zt(z)\tilde{t}(z)\right)'\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Psi}_m + \mathcal{O}_z\left(\frac{B}{N}\right)^3 + \mathcal{O}_z\left(\frac{1}{\sqrt{BN}}\right)$$

and that

$$\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right) = -\left(zt(z)\tilde{t}(z)\right)'\frac{1}{B+1}\mathrm{tr}\ \left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m\right) + \mathcal{O}_z\left(\frac{B}{N}\right)^3 + \mathcal{O}_z\left(\frac{1}{\sqrt{BN}}\right).$$

It is easily checked that

$$\mathbf{\Psi}_m = (\mathbf{I} + \mathbf{\Phi}_m)^{1/2} - \mathbf{I} = \frac{1}{2}\mathbf{\Phi}_m - \frac{1}{8}\mathbf{\Phi}_m^2 + \mathbf{\Xi}_m$$

where $\|\mathbf{\Xi}_m\| \le C\left(\frac{B}{N}\right)^3$. It is easily seen that $\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m^2 = \left(\frac{s_m'}{s_m}\right)^2 v_N + \mathcal{O}((\frac{B}{N})^3 + \frac{1}{N})$. Using (1.54), we thus obtain that

$$\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right) = -(zt_N(z)\tilde{t}_N(z))'\left(\frac{1}{M}\sum_{m=1}^{M}(\frac{s_m''}{2s_m} - \frac{(s_m')^2}{8(s_m)^2})\right)v_N + \mathcal{O}_z(u_N)$$

because

$$\mathcal{O}_z\left(\frac{B}{N}\right)^3 + \mathcal{O}_z\left(\frac{1}{\sqrt{BN}}\right) + \mathcal{O}_z\left(\frac{1}{N}\right) = \mathcal{O}_z(u_N).$$

We have thus established that $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1}\right)$ is a $\mathcal{O}_z\left(\frac{B}{N}\right)^2$ term, and have evaluated the corresponding principal term. Using similar calculations, we can obtain easily the expression of the $\mathcal{O}_z\left(\frac{B}{N}\right)^2$ term of $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^2\tilde{\mathbf{\Delta}}\right)$. In order to establish (1.123), it is necessary to evaluate the $\mathcal{O}_z\left(\frac{B}{N}\right)^2$ term of $\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}\tilde{\mathbf{\Delta}}\mathbf{Q}\right)$. This step needs very long calculations that are omitted. □

1.4.5 Estimation of $r_N(\nu)$

The term $\sup_\nu |\psi_N(f,\nu)|$ depends on the unknown true spectral densities $(s_m)_{m=1,\dots,M}$ through the term $r_N(\nu)$ defined by (1.8). In order to be able to use Theorem 1.2 in practice, it appears necessary to estimate $r_N(\nu)$ by an accurate enough estimate $\hat{r}_N(\nu)$, and to replace $\psi_N(f,\nu)$ by

$\hat{\psi}_N(f,\nu)$ defined by

$$\hat{\psi}_N(f,\nu) = \frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \, \mathrm{d}\mu_{MP}^{(c_N)} - \hat{r}_N(\nu)\, \phi_N(f)\, v_N \, \mathbf{1}_{\alpha>2/3} \tag{1.127}$$

$\hat{r}_N(\nu)$ has to be chosen in such a way that $|\hat{\psi}_N(f,\nu)| \prec u_N$, a condition that will be verified if $|\hat{r}_N(\nu) - r_N(\nu)| \prec \frac{u_N}{v_N}$ if $\alpha > \frac{2}{3}$. A natural choice for $\hat{r}_N(\nu)$ would be to replace the true spectral densities $(s_m)_{m=1,\ldots,M}$ by their frequency smoothed estimates $(\hat{s}_m)_{m=1,\ldots,M}$ defined by (1.76), and the derivatives $(s'_m)_{m=1,\ldots,M}$ by $(\hat{s}'_m)_{m=1,\ldots,M}$. However, $\hat{s}'_m$ is not an accurate estimate of s'_m so that the corresponding estimate of $r_N(\nu)$ does not satisfy $|\hat{r}_N(\nu) - r_N(\nu)| \prec \frac{u_N}{v_N}$ if $\alpha > \frac{2}{3}$. If $L < N$ is an integer, we introduce the lag window estimator $\hat{s}_{m,L}$ of s_m defined by

$$\hat{s}_{m,L}(\nu) = \int_0^1 |\xi_{\mathbf{y}_m}(\mu)|^2 w_L(\nu-\mu) d\mu = \sum_{l=-L}^{L} \hat{r}_{m,l}\, e^{-2i\pi l\nu} \tag{1.128}$$

where $w_L(\nu) = \sum_{l=-L}^{L} e^{-2i\pi l\nu}$ is the Fourier transform of the rectangular window and $\hat{r}_{m,l}$ represents the biased estimate of the autocovariance coefficient $r_{m,l}$ of y_m at lag l defined by

$$\hat{r}_{m,l} = \frac{1}{N}\sum_{n=1}^{N-l} y_{m,n+l} y^*_{m,n} \tag{1.129}$$

and $\hat{r}_{m,-l} = \hat{r}^*_{m,l}$ for $l \geq 0$. Then, the following result holds.

Proposition 1.4. *Assume that $L = L(N) = \mathcal{O}(N^{\frac{1}{2\gamma_0+1}})$, where $\gamma_0 \geq 3$ is defined by (1.18). Then, the estimate $\hat{r}_N(\nu)$ defined by*

$$\hat{r}_N(\nu) = \left(\frac{1}{M}\sum_{m=1}^{M} \frac{\hat{s}'_{m,L}(\nu)}{\hat{s}_{m,L}(\nu)}\right)^2 \tag{1.130}$$

satisfies

$$|\hat{r}_N(\nu) - r_N(\nu)| \prec \frac{1}{N^{(\gamma_0-1)/(2\gamma_0+1)}} \tag{1.131}$$

$$|\hat{r}_N(\nu) - r_N(\nu)| \prec \frac{u_N}{v_N} \text{ if } \alpha > \tfrac{2}{3} \tag{1.132}$$

as well as

$$|\hat{\psi}_N(f,\nu)| \prec u_N. \tag{1.133}$$

Proof. We denote by $\mathbf{d}_N(\nu)$ the N–dimensional vector defined by $\mathbf{d}_N(\nu) = (1, e^{-2i\pi\nu}, \ldots, e^{-2i\pi(N-1)\nu})^T$. We recall that $\mathbf{y}_m$ is the N–dimensional vector $\mathbf{y}_m = (y_{m,1}, \ldots, y_{m,N})^T$ which can be written as $\mathbf{y}_m = \mathbf{R}_m^{1/2}\mathbf{z}_m$ where $\mathbf{R}_m = \mathbb{E}(\mathbf{y}_m\mathbf{y}_m^*)$ and $\mathbf{z}_m$ is $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$ distributed. It is clear that $\hat{s}_{m,L}(\nu)$ can be written as

$$\hat{s}_{m,L}(\nu) = \mathbf{z}_m^*\mathbf{R}_m^{1/2}\boldsymbol{\Omega}(\nu)\mathbf{R}_m^{1/2}\mathbf{z}_m$$

with

$$\boldsymbol{\Omega}(\nu) = \frac{1}{N}\int \mathbf{d}_N(\mu)\mathbf{d}_N(\mu)^* w_L(\nu-\mu) d\mu$$

while $\hat{s}'_{m,L}(\nu)$ is equal to

$$\hat{s}'_{m,L}(\nu) = \mathbf{z}_m^*\mathbf{R}_m^{1/2}\boldsymbol{\Omega}'(\nu)\mathbf{R}_m^{1/2}\mathbf{z}_m$$

with

$$\mathbf{\Omega}'(\nu) = \frac{-2i\pi}{N} \int \mathbf{d}_N(\mu)\mathbf{d}_N(\mu)^* \left(\sum_{l=-L}^{L} l\, e^{-2i\pi l(\nu-\mu)} \right) d\mu$$

It is easy to check that $\|\mathbf{\Omega}'(\nu)\|_F = \mathcal{O}(\frac{L^{3/2}}{N^{1/2}})$ and therefore that $\|\mathbf{R}_m^{1/2}\mathbf{\Omega}'(\nu)\mathbf{R}_m^{1/2}\|_F = \mathcal{O}(\frac{L^{3/2}}{N^{1/2}})$. The Hanson-Wright inequality leads immediately to $|\hat{s}'_{m,L}(\nu) - \mathbb{E}(\hat{s}'_{m,L}(\nu))| \prec \frac{L^{3/2}}{N^{1/2}}$. Moreover, it is easy to check that (1.18) implies that

$$|\mathbb{E}(\hat{s}'_{m,L}(\nu)) - s'_m(\nu)| \leq \frac{C}{L^{\gamma_0 - 1}}$$

where C is a nice constant. For $L = L(N)$ in such a way that $\mathcal{O}(\frac{L^{3/2}}{N^{1/2}}) = \frac{1}{L^{\gamma_0-1}}$, i.e. $L = \mathcal{O}(N^{\frac{1}{2\gamma_0+1}})$, we obtain that

$$|\hat{s}'_{m,L}(\nu) - s'_m(\nu)| \prec \frac{1}{N^{(\gamma_0-1)/(2\gamma_0+1)}}.$$

Moreover, a similar analysis leads to

$$|\hat{s}_{m,L}(\nu) - s_m(\nu)| \prec \frac{1}{N^{\gamma_0/(2\gamma_0+1)}}$$

from which we deduce that the estimate $\hat{r}_N(\nu)$ defined by (1.130) satisfies (1.131). It is then easily checked that if $\gamma_0 \geq 3$, then (1.132) holds, which implies that $|\hat{\psi}_N(f,\nu)| \prec u_N$ holds. □

1.5 Use of Lipschitz properties of the functions $\nu \to \psi_N(f,\nu)$ and $\nu \to \hat{\psi}_N(f,\nu)$

In this section, we establish Lipschitz properties of $\nu \to \psi_N(f,\nu)$ and $\nu \to \hat{\psi}_N(f,\nu)$, and deduce that the stochastic domination properties (1.83) and (1.133) are still valid for $\sup_{\nu\in[0,1]} |\psi_N(f,\nu)|$ and $\sup_{\nu\in[0,1]} |\hat{\psi}_N(f,\nu)|$ where $\hat{\psi}_N(f,\nu)$ is defined by (1.127, 1.130).

1.5.1 Lipschitz properties

The goal of this paragraph is to prove the following Proposition.

Proposition 1.5. *Functions $\nu \to \psi_N(f,\nu)$ and $\nu \to \hat{\psi}_N(f,\nu)$ satisfy*

$$\sup_{\delta\neq 0} \sup_{\nu\in[0,1]} \frac{\|\psi_N(f,\nu) - \psi_N(f,\nu+\delta)\|}{|\delta|} \prec MN^{3/2} \tag{1.134}$$

$$\sup_{\delta\neq 0} \sup_{\nu\in[0,1]} \frac{\|\hat{\psi}_N(f,\nu) - \hat{\psi}_N(f,\nu+\delta)\|}{|\delta|} \prec MN^{3/2} \tag{1.135}$$

In the following, we just establish (1.135). For this, we evaluate separately the Lipschitz constants of $\nu \to \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ and of $\nu \to \hat{r}_N(\nu)$.

Lipschitz constant of $\nu \to \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$

To show that $\nu \to \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ is $MN^{3/2}$-Lipschitz with overwhelming probability, we need to establish a number of intermediate properties.

Proposition 1.6. *It holds that*

$$\sup_{\delta\neq 0} \sup_{\nu\in[0,1]} \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu+\delta)\|}{|\delta|} \prec MN^{3/2}. \tag{1.136}$$

Proof. Let $\delta \in \mathbb{R}$ and $\nu \in [0,1]$. As the random variables $(y_{m,n})_{m=1,\dots,M,n=1,\dots,N}$ are complex Gaussian and that $\sup_{m\geq 1} \mathbb{E}|y_{m,n}|^2 < +\infty$, the family $(y_{m,n})_{m=1,\dots,M,n=1,\dots,N}$ satisfies $|y_{m,n}| \prec 1$. Therefore, it holds that

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} |y_{m,n}| \prec \sqrt{N}. \tag{1.137}$$

For the same reasons, the family $\xi_{y_m}(\nu), m = 1,\dots,M$, $\nu \in [0,1]$ satisfies

$$|\xi_{y_m}(\nu)| \prec 1. \tag{1.138}$$

We also claim that

$$\sup_{\nu\in[0,1]} |\xi_{y_m}(\nu)| \prec 1. \tag{1.139}$$

In order to verify (1.139), we first observe that for any $n \geq 1$, we have the following control:

$$|e^{-2i\pi n\nu} - e^{-2i\pi n(\nu+\delta)}| \leq 2|\sin \pi n\delta| \leq 2\pi n|\delta|.$$

(1.137) implies that

$$\begin{aligned}
\sup_{\delta\neq 0} \sup_{\nu\in[0,1]} & \left| \frac{\xi_{y_m}(\nu) - \xi_{y_m}(\nu+\delta)}{\delta} \right| \\
&= \sup_{\delta\neq 0} \sup_{\nu\in[0,1]} \frac{1}{\sqrt{N}} \left| \sum_{n=1}^{N} y_{m,n} \frac{e^{-2i\pi n\nu} - e^{-2i\pi n(\nu+\delta)}}{\delta} \right| \\
&\leq 2\pi N \frac{1}{\sqrt{N}} \sum_{n=1}^{N} |y_{m,n}| \\
&\prec N^{3/2}.
\end{aligned} \tag{1.140}$$

We consider a frequency $\nu_* \in [0,1]$ (depending on m) where $|\xi_{y_m}(\nu)|$ is maximum, and have thus to establish that for each $\epsilon > 0$, there exists $\gamma > 0$ depending only on ϵ such that

$$\mathbb{P}(|\xi_{y_m}(\nu_*)| > N^\epsilon) \leq \exp -N^\gamma$$

for each N larger than a certain integer $N_0(\epsilon)$. We introduce the discrete set

$$\mathcal{V}_N^p = \left\{ \frac{k}{N^p} : k \in \{0,\dots,N^p - 1\} \right\} \tag{1.141}$$

whose cardinality is $|\mathcal{V}_N^p| = N^p$. We notice that (1.138) in conjunction with the union bound implies that $\sup_{\nu_p\in\mathcal{V}_N^p} |\xi_{y_m}(\nu_p)| \prec 1$. We denote by $\nu_{*,p}$ the element of $\mathcal{V}_N^p$ for which $|\nu_* - \nu_p|$ is minimum, and notice that $|\nu_* - \nu_{*,p}| \leq \frac{1}{N^p}$. Then, we have the following inequality

$$\begin{aligned}
&\mathbb{P}(|\xi_{y_m}(\nu_*)| > N^\epsilon) \\
&\quad \leq \mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2} \right) + \mathbb{P}\left(|\xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2} \right) \\
&\quad \leq \mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2} \right) + \mathbb{P}\left(\sup_{\nu_p\in\mathcal{V}_N^p} |\xi_{y_m}(\nu_p)| > \frac{N^\epsilon}{2} \right).
\end{aligned} \tag{1.142}$$

As $\sup_{\nu_p\in\mathcal{V}_N^p} |\xi_{y_m}(\nu_p)| \prec 1$, the second term of the right hand side of (1.142) converges exponentially towards 0. In order to evaluate the first term of the r.h.s. of (1.142), we use (1.140), and obtain that

$$\mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2}\right) \leq \mathbb{P}\left(N\frac{1}{\sqrt{N}}\sum_{n=1}^N |y_{m,n}| \geq \frac{\pi}{2|\nu_* - \nu_{*,p}|}N^\epsilon\right)$$
$$\leq \mathbb{P}\left(\frac{1}{\sqrt{N}}\sum_{n=1}^N |y_{m,n}| \geq \frac{\pi}{2}N^{p+\epsilon-1}\right).$$

We choose p so that $p-1 > 3/2$, and use (1.137) to conclude that $\mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2}\right)$ converges towards 0 exponentially. This establishes (1.139).

In order to complete the proof of Proposition 1.6, we consider an individual entry $\hat{s}_{ij}(\nu)$ of $\hat{\mathbf{S}}(\nu)$ for $i, j \leq M$, and write

$$\begin{aligned}
&|\hat{s}_{ij}(\nu) - \hat{s}_{ij}(\nu+\delta)| \\
&= \frac{1}{B+1}\left|\sum_{b=-B/2}^{B/2} \xi_i\left(\nu + \frac{b}{N}\right)\xi_j\left(\nu + \frac{b}{N}\right)^* \right.\\
&\qquad \left. - \xi_i\left(\nu + \delta + \frac{b}{N}\right)\xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right| \\
&\leq \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_i\left(\nu + \frac{b}{N}\right)\left(\xi_j\left(\nu + \frac{b}{N}\right)^* - \xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right)\right| \\
&\qquad + \left|\left(\xi_i\left(\nu + \frac{b}{N}\right) - \xi_i\left(\nu + \delta + \frac{b}{N}\right)\right)\xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right|.
\end{aligned}$$

Using the estimations (1.139) and (1.140), we get:

$$\sup_{i,j}\sup_{\delta\neq 0}\sup_{\nu\in[0,1]}\left|\frac{\hat{s}_{ij}(\nu) - \hat{s}_{ij}(\nu+\delta)}{\delta}\right| \prec N^{3/2} \tag{1.143}$$

and deduce (1.136) from the rough bound

$$\begin{aligned}
\sup_{\nu\in[0,1]}\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu+\delta)\| &\leq \sup_{\nu\in[0,1]}\sup_i \sum_j |\hat{s}_{ij}(\nu) - \hat{s}_{ij}(\nu+\delta)| \\
&\leq M\sup_{\nu\in[0,1]}\sup_{i,j}|\hat{s}_{ij}(\nu) - \hat{s}_{ij}(\nu+\delta)|.
\end{aligned}$$

□

Combining the eigenvalue localisation result from Corollary 1.3 and the Lipschitz behaviour of $\hat{\mathbf{S}}$ from Proposition 1.6, the following statement holds.

Corollary 1.6. *(ν uniform version of Corollary 1.3.) Denote for $\epsilon > 0$:*

$$\begin{aligned}
\Lambda_\epsilon^{\hat{\mathbf{S}}} &= \left\{\forall\nu\in[0,1] : \sigma(\hat{\mathbf{S}}(\nu)) \subset \mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon\right\} \\
\Lambda_\epsilon^{\hat{\mathbf{D}}} &= \left\{\forall\nu\in[0,1] : \sigma(\hat{\mathbf{D}}(\nu)) \subset [\underline{s}, \bar{s}] + \epsilon\right\}.
\end{aligned}$$

Then, $\Lambda_\epsilon^{\hat{\mathbf{S}}}$ and $\Lambda_\epsilon^{\hat{\mathbf{D}}}$ hold with exponentially high probability.

Proof. As the proof for $\Lambda_\epsilon^{\hat{\mathbf{D}}}$ is strictly similar to the one of $\Lambda_\epsilon^{\hat{\mathbf{S}}}$, we will only write the arguments for $\Lambda_\epsilon^{\hat{\mathbf{S}}}$. For any fixed $\nu\in[0,1]$, Corollary 1.3 ensures that $\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu)$ holds with exponentially high probability. For $p \geq 1$, we still consider the set $\mathcal{V}_N^p$ defined by (1.141) and denote by $\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}}$ the event defined by

$$\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}} = \left\{\forall \nu_p \in \mathcal{V}_N^p : \sigma(\hat{\mathbf{S}}(\nu_p)) \subset \mathrm{Supp}\mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon\right\}$$

which is $\Lambda_\epsilon^{\hat{\mathbf{S}}}$ but where ν runs only on the finite grid $\mathcal{V}_N^p$. It is immediate (by the union bound) that $\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}}$ holds with exponentially high probability for any fixed $p \in \mathbb{N}$. Moreover, it is clear from the definitions of $\Lambda_\epsilon^{\hat{\mathbf{S}}}$ and $\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}}$ that $\Lambda_\epsilon^{\hat{\mathbf{S}}} \subset \Lambda_{\epsilon,p}^{\hat{\mathbf{S}}}$. We now show the following inclusion:

$$\begin{aligned}\left(\Lambda_\epsilon^{\hat{\mathbf{S}}}\right)^c &\subset \left(\Lambda_{\epsilon/2,p}^{\hat{\mathbf{S}}}\right)^c \\ &\cup \left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > \epsilon/2 \text{ where } \nu_p^* \in \underset{\nu_p \in \mathcal{V}_N^p}{\mathrm{argmin}}\, |\nu - \nu_p|\right\}. \end{aligned} \tag{1.144}$$

Suppose that $(\Lambda_\epsilon^{\hat{\mathbf{S}}})^c$ is realized, and denote by $\nu^* \in [0,1]$ a frequency such that $\sigma(\hat{\mathbf{S}})(\nu^*) \not\subset \mathrm{Supp}_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon$. Denote also $\nu_p^* \in \mathrm{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu_p - \nu^*|$. We just consider the case where $\lambda_1(\hat{\mathbf{S}}(\nu^*)) > \bar{s}(1+\sqrt{c})^2 + \epsilon$, since in the case where $\lambda_M(\hat{\mathbf{S}}(\nu^*)) < \underline{s}(1-\sqrt{c})^2 - \epsilon$, the proof is similar. Then, either:

1. $\|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu^*)\| \leq \epsilon/2$, which implies the following estimation for the location of $\lambda_1(\hat{\mathbf{S}}(\nu_p^*))$:

$$\lambda_1(\hat{\mathbf{S}}(\nu^*)) - \frac{\epsilon}{2} \leq \lambda_1(\hat{\mathbf{S}}(\nu_p^*)) \leq \lambda_1(\hat{\mathbf{S}}(\nu^*)) + \frac{\epsilon}{2}$$

   and in particular, $\lambda_1(\hat{\mathbf{S}}(\nu_p^*)) \geq \bar{s}(1+\sqrt{c})^2 + \epsilon/2$. This means that $\left(\Lambda_{\epsilon/2,p}^{\hat{\mathbf{S}}}\right)^c$ holds.

2. $\|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu^*)\| > \epsilon/2$, which exactly means that $\left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > \epsilon/2 \text{ where } \nu_p^* \in \mathrm{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\}$ is realized

(1.144) is now proved.

We already showed that $\left(\Lambda_{\epsilon/2,p}^{\hat{\mathbf{S}}}\right)^c$ holds with exponentially small probability, and establish now that the set

$$\left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu)\| > \epsilon/2 \text{ where } \nu_p^* \in \underset{\nu_p \in \mathcal{V}_N^p}{\mathrm{argmin}}\, |\nu - \nu_p|\right\}$$

has the same property. To justify this claim, we note that Proposition 1.6 implies that for each $\kappa > 0$, the probability

$$\mathbb{P}\left[\left\{\exists \nu, \nu' \in [0,1],\ \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu')\|}{|\nu - \nu'|} > N^\kappa M N^{3/2}\right\}\right]$$

converges to 0 exponentially fast. As the following inclusion

$$\begin{aligned}&\left\{\exists \nu \in [0,1],\ \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\|}{|\nu - \nu_p^*|}, > N^\kappa M N^{3/2}, \text{ where } \nu_p^* \in \underset{\nu_p \in \mathcal{V}_N^p}{\mathrm{argmin}}\, |\nu - \nu_p|\right\} \\ &\qquad\subset \left\{\exists \nu, \nu' \in [0,1],\ \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu')\|}{|\nu - \nu'|} > N^\kappa M N^{3/2}\right\}\end{aligned}$$

holds, we get that

$$\mathbb{P}\left[\left\{\exists \nu \in [0,1],\ \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > |\nu - \nu_p^*| N^\kappa M N^{3/2}\right\}\right] \to 0$$

exponentially fast. Moreover, as for each ν, $|\nu - \nu_p^*| \leq \frac{1}{N^p}$, we obtain that

$$\mathbb{P}\left[\left\{\exists \nu \in [0,1],\ \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > \frac{1}{N^p} N^{\kappa} M N^{3/2}\right\}\right] \to 0$$

exponentially fast as well. For p large enough, $N^{\kappa}\frac{1}{N^p} M N^{3/2}$ will finally become smaller than $\epsilon/2$. This proves that

$$\left\{\exists \nu \in [0,1],\ \|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu)\| > \epsilon/2 \text{ where } \nu_p^* \in \underset{\nu_p \in \mathcal{V}_N^p}{\operatorname{argmin}} |\nu - \nu_p|\right\}$$

holds with exponentially small probability.

The same argument can be used to control $\Lambda_{\epsilon}^{\hat{\mathbf{D}}}$. This completes the proof of Corollary 1.6. □

We deduce immediately from Corollary 1.6 the following result that can be seen as a refinement of (1.73) and of Lemma 1.3.

Corollary 1.7. *It holds that*

$$\sup_{\nu \in [0,1]} \|\hat{\mathbf{D}}(\nu)^{-1/2}\| \prec 1, \qquad \sup_{\nu \in [0,1]} \|\hat{\mathbf{S}}(\nu)\| \prec 1.$$

A useful consequence of this is the following Corollary, which states that the Lipschitz result holds for $\hat{\mathbf{C}}(\nu)$.

Corollary 1.8. *It holds that*

$$\sup_{\delta \neq 0} \sup_{\nu \in [0,1]} \left\| \frac{\hat{\mathbf{C}}(\nu) - \hat{\mathbf{C}}(\nu+\delta)}{\delta} \right\| \prec M N^{3/2} \tag{1.145}$$

Proof. For more clarity in the following argument, denote $\nu_1 = \nu$ and $\nu_2 = \nu + \delta$. Recall that $\hat{\mathbf{D}} = \mathrm{dg}\hat{\mathbf{S}}$. Using the definition of $\hat{\mathbf{C}}$ from equation (1.3), we write:

$$\begin{aligned}
\hat{\mathbf{C}}(\nu_2) - \hat{\mathbf{C}}(\nu_1) &= \hat{\mathbf{D}}^{-1/2}(\nu_2)\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1)\hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1) \\
&= (\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1))\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2) \\
&\qquad + \hat{\mathbf{D}}^{-1/2}(\nu_1)(\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1)).
\end{aligned}$$

Moreover, we write

$$\begin{aligned}
&\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1) \\
&\qquad = (\hat{\mathbf{S}}(\nu_2) - \hat{\mathbf{S}}(\nu_1))\hat{\mathbf{D}}^{-1/2}(\nu_2) + \hat{\mathbf{S}}(\nu_1)(\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1)).
\end{aligned}$$

Therefore, applying the operator norm, we get by the triangle inequality:

$$\begin{aligned}
\|\hat{\mathbf{C}}(\nu_2) - \hat{\mathbf{C}}(\nu_1)\| \leq\ & \|\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_2)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2)\| \\
&+ \|\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_2) - \hat{\mathbf{S}}(\nu_1)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2)\| \\
&+ \|\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_1)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1)\|
\end{aligned}$$

It is easy to check that

$$\sup_{\delta \neq 0} \sup_{|\nu_2 - \nu_1| = \delta} \left\| \frac{\hat{\mathbf{D}}^{-1/2}(\nu_2) - \hat{\mathbf{D}}^{-1/2}(\nu_1)}{\delta} \right\| \prec N^{3/2}$$

holds. Therefore, Proposition 1.6 and Corollary 1.7 immediately imply (1.145). $\square$

Finally, we can write for the spectrum of $\hat{\mathbf{C}}$ the same kind of result as in Corollary 1.6.

Corollary 1.9. *For each $\epsilon > 0$, we define $\Lambda_\epsilon^{\hat{\mathbf{C}}}$ as the event*

$$\Lambda_\epsilon^{\hat{\mathbf{C}}} = \left\{\forall \nu \in [0,1] : \sigma(\hat{\mathbf{C}}(\nu)) \subset \mathrm{Supp}\mu_{MP}^{(c)} + \epsilon\right\}.$$

Then, $\Lambda_\epsilon^{\hat{\mathbf{C}}}$ holds with exponentially high probability.

Proof. The proof is similar to the proof of Corollary 1.6 and is thus omitted. $\square$

We finally use the above results to prove that $\nu \to \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f\, d\mu_{MP}^{(c_N)}$ is $MN^{3/2}$-Lipschitz with overwhelming probability. For this, we establish the following Proposition.

Proposition 1.7. *It holds that*

$$\sup_{\delta \neq 0} \sup_{\nu \in [0,1]} \frac{1}{|\delta|} \left| \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu+\delta)) \right| \prec MN^{3/2}. \tag{1.146}$$

Proof. By Corollary 1.9, the event $\Lambda_\epsilon^{\hat{\mathbf{C}}}$ holds with exponentially high probability. Therefore, it is sufficient to establish that

$$\mathbf{1}_{\Lambda_\epsilon^{\hat{\mathbf{C}}}} \sup_{\delta \neq 0} \sup_{\nu \in [0,1]} \frac{1}{|\delta|} \left| \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu+\delta)) - \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) \right| \prec MN^{3/2}$$

We express $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu+\delta)) - \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ as

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu+\delta)) - \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) = \frac{1}{M}\sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu+\delta))) - f(\lambda_m(\hat{\mathbf{C}}(\nu)).$$

As f is $\mathcal{C}^\infty$ on a neighborhood of $\mathrm{Supp}_{MP}^{(c)}$, on the set $\Lambda_\epsilon^{\hat{\mathbf{C}}}$, there exist some random quantities $(\tilde{\lambda}_m)_{1\leq m\leq M}$ between $\lambda_m(\hat{\mathbf{C}}(\nu))$ and $\lambda_m(\hat{\mathbf{C}}(\nu+\delta))$ such that

$$\frac{1}{M}\sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu+\delta))) - f(\lambda_m(\hat{\mathbf{C}}(\nu))$$
$$= \frac{1}{M}\sum_{m=1}^{M} \Big(\lambda_m(\hat{\mathbf{C}}(\nu+\delta)) - \lambda_m(\hat{\mathbf{C}}(\nu))\Big)\ f'(\tilde{\lambda}_m).$$

Using the following eigenvalue inequality for Hermitian matrices:

$$\Big|\lambda_m(\hat{\mathbf{C}}(\nu+\delta)) - \lambda_m(\hat{\mathbf{C}}(\nu))\Big| \leq \|\hat{\mathbf{C}}(\nu+\delta) - \hat{\mathbf{C}}(\nu)\|$$

in conjunction with the fact that $\sup_{1\leq m\leq M} |f'(\tilde{\lambda}_m|$ is bounded by some nice constant C on the event $\Lambda_\epsilon^{\hat{\mathbf{C}}}$, we obtain that

$$\mathbb{P}\left[\sup_{\delta\neq 0}\sup_{\nu\in[0,1]} \left|\frac{1}{M}\sum_{m=1}^{M} f'(\tilde{\lambda}_m)(\lambda_m(\hat{\mathbf{C}}(\nu+\delta)) - \lambda_m(\hat{\mathbf{C}}(\nu))\right| > |\delta| N^\kappa MN^{3/2},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$
$$\leq \mathbb{P}\left[\sup_{\delta\neq 0}\sup_{\nu\in[0,1]} C\|\hat{\mathbf{C}}(\nu+\delta) - \hat{\mathbf{C}}(\nu)\| > |\delta| N^\kappa MN^{3/2},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$

(1.145) finally leads to (1.146).

□

Lipschitz constant of $\nu \to \hat{r}_N(\nu)$.

The function $\nu \to \hat{r}_N(\nu)$ satisfies the following property:

Proposition 1.8.

$$\sup_{\delta \neq 0} \sup_{\nu \in [0,1]} \frac{1}{|\delta|} \left| \hat{r}_N(\nu+\delta) - \hat{r}_N(\nu) \right| \prec N^{3/(2\gamma_0+1)}. \tag{1.147}$$

We just provide the main steps the proof, and leave the details to the reader. We first prove that $\sup_{\nu \in [0,1]} \sum_{m=1}^{M} \frac{1}{\hat{s}_{m,L}(\nu)} \prec 1$ by verifying that the event $\{\forall \nu \in [0,1], \forall m = 1, \ldots, M, \hat{s}_{m,L}(\nu) \in [\underline{s}, \bar{s}] + \epsilon\}$ holds with exponentially high probability. Then, we establish that $\nu \to \hat{s}_{m,L}(\nu)$ and $\nu \to \hat{s}'_{m,L}(\nu)$ are $N^{2/(2\gamma_0+1)}$ Lipschitz and $N^{3/(2\gamma_0+1)}$ Lipschitz with overwhelming probability. This leads immediately to (1.147).

As $v_N \, N^{3/(2\gamma_0+1)} \ll M N^{3/2}$, Propositions 1.7 and 1.8 lead to (1.135). This completes the proof of Proposition 1.5.

1.5.2 Stochastic domination of $\sup_{\nu \in [0,1]} |\psi_N(f,\nu)|$ and $\sup_{\nu \in [0,1]} |\hat{\psi}_N(f,\nu)|$

We are now in a position to establish the main result of this chapter.

Theorem 1.3. $\sup_{\nu \in [0,1]} |\psi_N(f,\nu)|$ *and* $\sup_{\nu \in [0,1]} |\hat{\psi}_N(f,\nu)|$ *satisfy the following stochastic domination property:*

$$\sup_{\nu \in [0,1]} |\psi_N(f,\nu)| \quad \prec \quad u_N \tag{1.148}$$

$$\sup_{\nu \in [0,1]} |\hat{\psi}_N(f,\nu)| \quad \prec \quad u_N. \tag{1.149}$$

Proof. We just establish (1.149). We consider $\epsilon > 0$ and evaluate

$$\mathbb{P}\left[\sup_{\nu \in [0,1]} \left|\hat{\psi}_N(f,\nu)\right| > N^{\epsilon} u_N\right].$$

We denote by $\nu^* \in [0,1]$ an element where the supremum is achieved, and consider ν_p^* the closest element of $\mathcal{V}_N^p$ to ν^*, where we recall that $\mathcal{V}_N^p$ is defined by (1.141). Therefore, one can write:

$$\begin{aligned} \mathbb{P}\left[\sup_{\nu \in [0,1]} \left|\hat{\psi}_N(f,\nu)\right| > N^{\epsilon} u_N\right] \leq & \mathbb{P}\left[\left|\hat{\psi}_N(f,\nu^*) - \hat{\psi}_N(f,\nu_p^*))\right| > \frac{1}{2} N^{\epsilon} u_N\right] + \\ & \mathbb{P}\left[\left|\hat{\psi}_N(f,\nu_p^*)\right| > \frac{1}{2} N^{\epsilon} u_N\right]. \end{aligned}$$

(1.133) implies that $\mathbb{P}\left[\left|\hat{\psi}_N(f,\nu_p^*)\right| > \frac{1}{2} N^{\epsilon} u_N\right]$ converges exponentially towards 0. It thus remains to study $\mathbb{P}\left[\left|\hat{\psi}_N(f,\nu^*) - \hat{\psi}_N(f,\nu_p^*))\right| > \frac{1}{2} N^{\epsilon} u_N\right]$. For this, we of course use (1.135), Corollary 1.8, and write

$$\mathbb{P}\left[\left|\hat{\psi}_N(f,\nu^*) - \hat{\psi}_N(f,\nu_p^*))\right| > \frac{1}{2}N^{\epsilon}u_N\right]$$
$$= \mathbb{P}\left[\left|\frac{\hat{\psi}_N(f,\nu^*) - \hat{\psi}_N(f,\nu_p^*))}{\nu^* - \nu_p^*}\right| > \frac{1}{2|\nu^* - \nu_p^*|}N^{\epsilon}u_N\right]$$
$$\leq \mathbb{P}\left[\left|\frac{\hat{\psi}_N(f,\nu^*) - \hat{\psi}_N(f,\nu_p^*))}{\nu^* - \nu_p^*}\right| > \frac{1}{2}N^p N^{\epsilon}u_N\right].$$

If we choose p large enough, $MN^{3/2}$ satisfies $MN^{3/2} \ll N^p u_N$, and $\mathbb{P}\left[\left|\frac{\hat{\psi}_N(f,\nu^*)-\hat{\psi}_N(f,\nu_p^*))}{\nu^*-\nu_p^*}\right| > \frac{1}{2}N^p N^{\epsilon}u_N\right]$ converges towards 0 exponentially as expected. This completes the proof of (1.149). $\square$

1.6 Numerical simulations

In this section we examine the impact of the correction quantity $r_N(\nu)\phi_N(f)v_N$ when $\alpha > \frac{2}{3}$ and see how it improves the estimation of the LSS $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$. More precisely, we start by examining the behaviour of the LSS

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|$$

that we abbreviate by $LSS(f,\nu)$, and the impact of the correction term

$$\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s_m'(\nu)}{s_m(\nu)}\right)^2 \phi_N(f)\left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{b}{N}\right)^2\right) = r_N(\nu)\phi_N(f)v_N$$
$$\left(\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{s}_m'(\nu)}{\hat{s}_m(\nu)}\right)^2 \phi_N(f)\left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{b}{N}\right)^2\right) = \hat{r}_N(\nu)\phi_N(f)v_N.$$

under $\mathcal{H}_0$. We recall that $\phi_N(f)$ is the deterministic term defined as the action of f on the compactly supported distribution D_N, whose Stieltjes transform is:

$$p_N(z) = -\frac{c_N(zt_N(z)\tilde{t}_N(z))^3}{1 - c_N(zt_N(z)\tilde{t}_N(z))^2}.$$

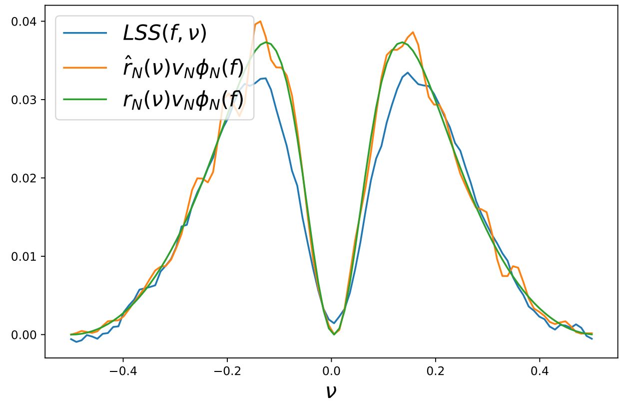
Motivated by [64], we consider $f(\lambda) = (\lambda - 1)^2$ where it can be verified with a bit of algebra and residue calculus that

$$\int_{\mathbb{R}} f(\lambda)\,\mathrm{d}\mu_{MP}^{(c_N)}(\lambda) = c_N.$$

and $\phi_N(f) = c_N$. Take $\mathbf{y}_n$ generated by the following simple model:

$$\mathbf{y}_{n+1} = \mathbf{A}\mathbf{y}_n + \boldsymbol{\epsilon}_n \tag{1.150}$$

where $(\boldsymbol{\epsilon}_n)_{n\in\mathbb{Z}}$ is an independent sequence of $\mathcal{N}_{\mathbb{C}}(0,\mathbf{I}_M)$ distributed random vectors, and where $\mathbf{A}$ is the diagonal matrix defined by $\mathbf{A} = \theta\,\mathbf{I}_M$ for $\theta \in \mathbb{C}$ such that $|\theta| < 1$. Under (1.150), each time series is independent AR(1) processes. In Figure 1.1 is represented on the left the values of the LSS associated to $f(\lambda) = (\lambda-1)^2$ for each $\nu \in (0,1)$ when $(N,B,M,L) = (10119, 1600, 800, 21)$ (so $\alpha = 0.8$ and $c = 1/2$) and $\theta = 0.4$, where we recall that L represent the lag window size in



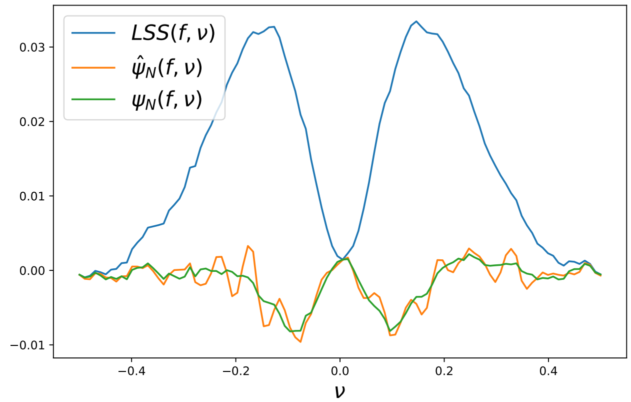


Figure 1.1: Linear Spectral Statistics vs the correction term. $f(\lambda) = (\lambda - 1)^2$, $(N, B, M, L) = (10119, 1600, 800, 21)$, and $\theta = 0.4$.

the estimation of $\hat{r}_N(\nu)$. We see that the correction term captures the majority of the deviation of the LSS from zero. Moreover, the correction where the spectral densities s_m and s'_m are estimated still provide a good approximation of the $\mathcal{O}(\frac{B}{N})^2$ term. On the right side is represented the LSS against $\psi_N(f,\nu)$ and $\hat{\psi}_N(f,\nu)$. We again observe that the majority of the deviation from zero of the LSS is corrected by the $\mathcal{O}(\frac{B}{N})^2$ terms. Around $\nu = \pm 0.1$, the corrections' precision seems to have degraded. This can be understood since $\nu = \pm 0.1$ corresponds to peaks in s'_m, which leads to greater estimation errors for $\hat{s}'_m$ at this frequency than for the other ones.

We now recall that we proved in Theorem 1.3 that

$$\sup_{\nu \in [0,1]} \left| \frac{1}{M} \operatorname{tr} f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right| = \mathcal{O}\left(\frac{1}{B}\right) \mathbb{1}_{1/2 \leq \alpha \leq 2/3} + \mathcal{O}\left(\frac{B}{N}\right)^2 \mathbb{1}_{\alpha \geq 2/3}$$

and that the right hand-side is minimal when the rates $\frac{1}{B}$ and $\left(\frac{B}{N}\right)^2$ coincides, ie. when $\alpha = \frac{2}{3}$. In order to support these statements, we represent in Figure 1.2 the value of

$$\sup_{\nu \in [0,1]} \left| \frac{\frac{1}{M} \operatorname{tr} f(\hat{\mathbf{C}}(\nu)) - \int f \, \mathrm{d}\mu_{mp,c}}{\int f \, \mathrm{d}\mu_{mp,c}} \right|$$

which is the relative error between $\frac{1}{M}\operatorname{tr} f(\hat{\mathbf{C}}(\nu))$ and its corresponding asymptotic limit as a function of $\alpha \in (0.5, 0.9)$ for different values of N, and averaged (using the square root of the mean squared error) over 10^2 realizations. The corresponding values of M and B are computed as $B = N^\alpha$, $M = B \times c$ and taking the floor function. These numerical results tend to confirm the fact that the error between the considered statistic and its asymptotic deterministic approximation tends to be dominated by two different phenomena depending on whether $M \ll \mathcal{O}(N^{2/3})$ or $M \gg \mathcal{O}(N^{2/3})$. In the first case, the main contribution to the error corresponds to the $\frac{1}{B}$ terms arising from multiple terms in the decomposition of $\frac{1}{M}\operatorname{tr} f(\hat{\mathbf{C}}(\nu))$ while in the second case, the deterministic error term

$$\left(\frac{1}{M}\sum_{m=1}^{M} \frac{s'_m(\nu)}{s_m(\nu)}\right)^2 \phi_N(f) \left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2} \left(\frac{b}{N}\right)^2\right)$$

of order $\mathcal{O}(\frac{B}{N})^2$ is dominant. We also observe that the optimum choice of α appears to be close to $\frac{2}{3}$ (which corresponds to the case where the two error rates coincide) even for relatively small values of N, M and B.

In Figure 1.3, we represent the same quantities for fixed (N, M, B) but for various values of ν.

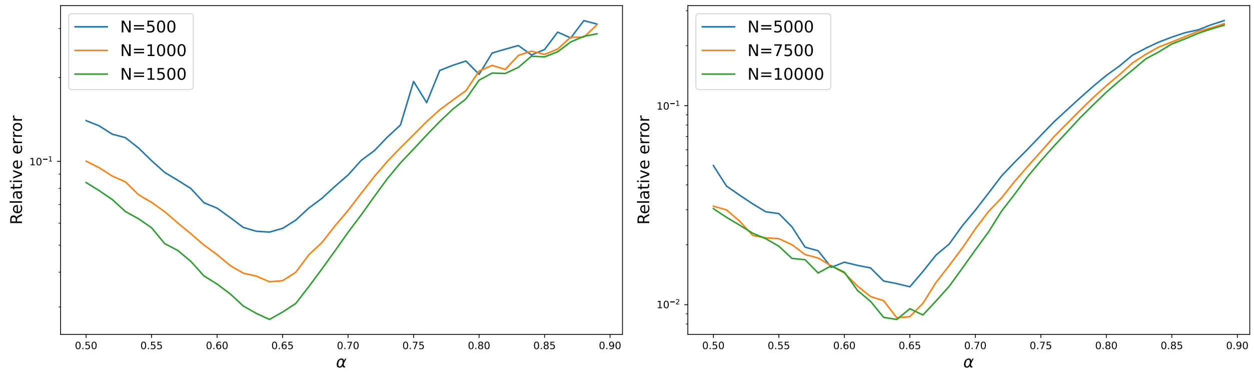


Figure 1.2: Evolution of the error of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ with respect to the Marcenko-Pastur limit $(c = 0.5)$ as a function of α for $\nu = 0.1$, averaged over 100 realizations of the statistic. Small values of N on the left and larger values of N on the right.

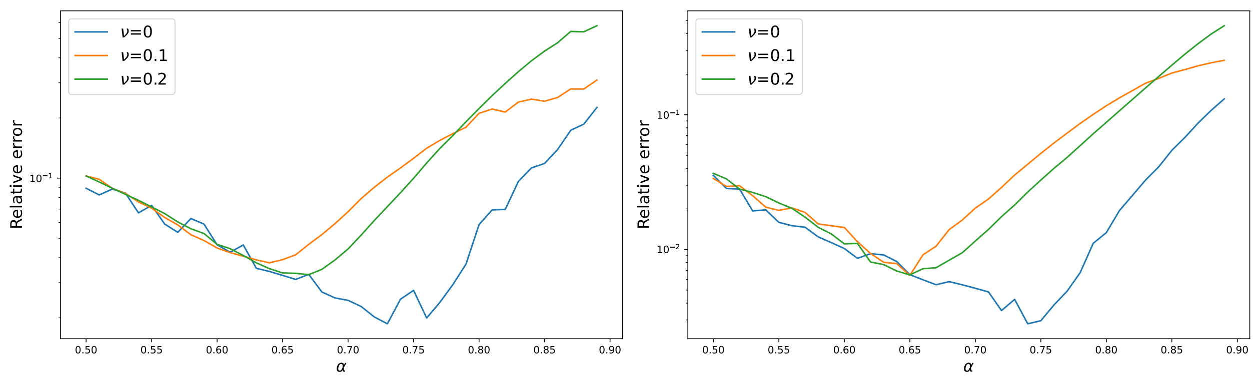


Figure 1.3: Evolution of the error of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ with respect to the Marcenko-Pastur limit $(c = 0.5)$ as a function of α averaged over 100 realizations of the statistic. $N = 1000$ on the left and $N = 10000$ on the right.

We notice that the behaviour for $\nu = 0$ is different than for $\nu = 0.1, 0.2$, as the transition seems to appear near $\alpha = \frac{3}{4}$ instead of $\alpha = \frac{2}{3}$. This can be understood by recalling that

- in our context, s'_m is null for $\nu = 0$ but not for $\nu = 0.1$ and $\nu = 0.2$
- the $\mathcal{O}(\frac{B}{N})^2$ correction term is proportional to $s'_m(\nu)$,

Therefore, in the case $\nu = 0$, the $\mathcal{O}(\frac{B}{N})^2$ correction term is exactly equal to 0, and as a consequence, we should directly obtain the convergence of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ towards $\int f\, \mathrm{d}\mu_{mp,c_N}$ at the stronger rate $\frac{\sqrt{B}}{N}$ if $\alpha \in (\frac{2}{3}, \frac{4}{5})$ and $(\frac{B}{N})^3$ if $\alpha \in (\frac{4}{5}, 1)$. However, our theoretical results would also implies that the transition would happen for $\alpha = \frac{4}{5}$ instead of the observed $\frac{3}{4}$ in Figure 1.3. Since $\alpha = \frac{3}{4}$ is the minimum rate achieved by $\frac{1}{B} + (\frac{B}{N})^3$, we conjecture that for $\nu = 0$, the error term proportional to $\frac{\sqrt{B}}{N}$ in fact vanishes so that it remains only:

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f\, \mathrm{d}\mu_{mp,c_N} = \mathcal{O}_{\prec}\left(\frac{1}{B} + \left(\frac{B}{N}\right)^3\right)$$

for $\nu = 0$.

We now present another way to check that the derived speed of convergence towards zero in

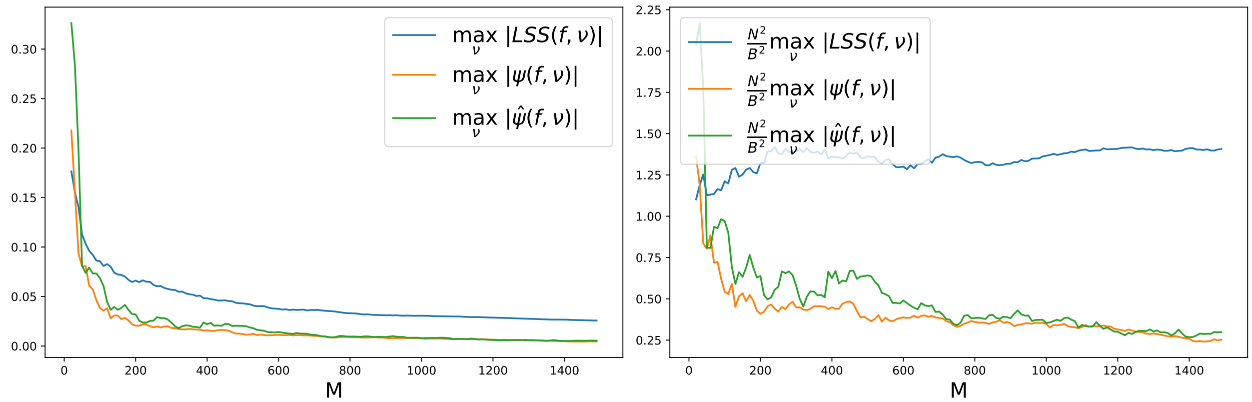


Figure 1.4: $\sup_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|$ against $\sup_{\nu\in\mathcal{V}_N}\psi_N(f,\nu)$ and $\sup_{\nu\in\mathcal{V}_N}\hat{\psi}_N(f,\nu)$ as functions of M. On the right the quantities are rescaled by $(\frac{N}{B})^2$. $\alpha = 0.8$, $c = 1/2$, $\theta = 0.4$

Theorem 1.3 hold true:

$$\sup_{\nu\in[0,1]}\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| = \mathcal{O}\left(\frac{1}{B}\right)\mathbb{1}_{1/2\leq\alpha\leq 2/3} + \mathcal{O}\left(\frac{B}{N}\right)^2\mathbb{1}_{\alpha\geq 2/3}$$
$$\sup_{\nu\in[0,1]}|\psi(f,\nu)| = \mathcal{O}(u_N).$$

In the following we take $c = \frac{M}{B+1} = \frac{1}{2}$ and $\alpha = 4/5$. In this case we recall that $u_N = \mathcal{O}(\frac{B}{N})^3$. On the left of Figure 1.4 is represented for $M \in \{20, 30, \ldots, 1500\}$ the value of $\sup_{\nu\in[0,1]}|LSS(f,\nu)|$ against $\sup_{\nu\in[0,1]}|\psi(f,\nu)|$ and $\sup_{\nu\in[0,1]}|\hat{\psi}(f,\nu)|$. On the right of Figure 1.4 we rescale all quantities by $(\frac{N}{B})^2$ and observe, in accordance with Theorem 1.2 that $LSS(f,\nu)$ remains $\mathcal{O}(1)$ while the corrected quantities are $o(1)$. Finally, in Figure 1.5 are represented $\sup_{\nu\in[0,1]}|\psi(f,\nu)|$ and $\sup_{\nu\in[0,1]}|\hat{\psi}(f,\nu)|$ rescaled by $(\frac{N}{B})^3$, and observe that these quantities are now $\mathcal{O}(1)$, again in accordance with Theorem 1.2.

In Figure 1.6 is represented 20000 realisations of the LSS $\sup_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|$ against its improved estimations $\sup_{\nu\in\mathcal{V}_N}|\psi_N(f,\nu)|$ and $\sup_{\nu\in\mathcal{V}_N}|\hat{\psi}_N(f,\nu)|$. We see that the oracle corrected statistics $\psi(f,\nu)$ is more concentrated around 0, and that its estimated counterpart $\hat{\psi}(f,\nu)$ is close to $\psi(f,\nu)$ but exhibits more spread due to the additional estimation step of $\hat{s}_m(\nu)$.

Overall, we have seen in this section of numerical simulation that the behaviour of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ is appropriately described by the statement of Theorem 1.3. The approximation of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ by its Marcenko-Pastur equivalent holds for moderate values of N, M and B since the approximation error is of order $\mathcal{O}(\frac{1}{B}) = \mathcal{O}(\frac{1}{N^\alpha})$ for $\alpha \in (\frac{1}{2}, \frac{2}{3})$ which is a quite fast rate, and otherwise of order $\mathcal{O}(\frac{B}{N})^2 = \mathcal{O}(\frac{1}{N^{2(1-\alpha)}})$ which is also a fast rate while α is not close to 1. Typically, we saw in Figure 1.2 and Figure 1.3 that for $(N, \alpha, c) = (500, \frac{1}{2}, \frac{1}{2})$ (so $(B, M) = (22, 11)$) is enough to observe a good fit between $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ and $\int f\,\mathrm{d}\mu_{mp,c_N}$ as the relative error is about 10%. However, in order to observe in practice the improvement expected by considering the $\mathcal{O}(\frac{B}{N})^2$ correction term, N, M and B have to be chosen much larger. This can be explained by considering the ratio of the rates in the case of the corrected statistics versus the non-corrected statistics. It is at best equal to

$$\sup_{\alpha\in(\frac{2}{3},1)}\frac{\frac{\sqrt{B}}{N} + \left(\frac{B}{N}\right)^3}{\left(\frac{B}{N}\right)^2} = \frac{1}{N^{\frac{1}{5}}}$$

which grows very slowly with N. This is why in the simulation in Figure 1.4 it was necessary to

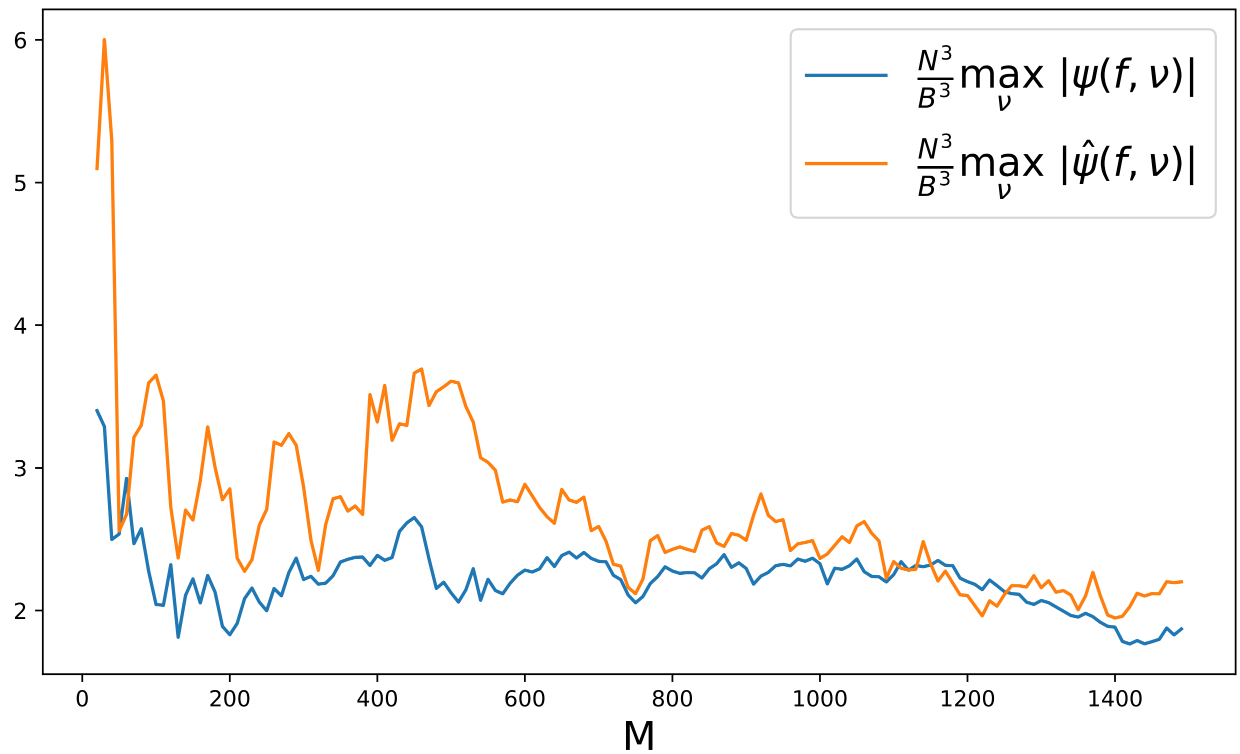


Figure 1.5: $\sup_{\nu\in\mathcal{V}_N}\psi_N(f,\nu)$ and $\sup_{\nu\in\mathcal{V}_N}\hat{\psi}_N(f,\nu)$ rescaled by $(\frac{N}{B})^3$ as functions of M. $\alpha=0.8$, $c=1/2$, $\theta=0.4$.

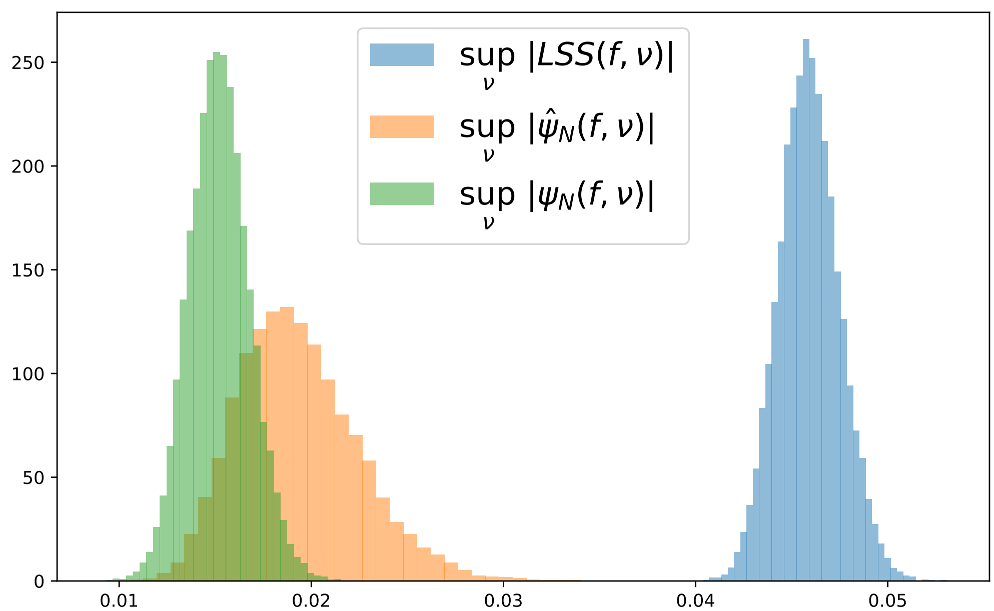


Figure 1.6: $\sup_{\nu\in\mathcal{V}_N}|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))-\int f\,\mathrm{d}\mu_{MP}^{(c_N)}|$ against $\sup_{\nu\in\mathcal{V}_N}|\psi_N(f,\nu)|$ and $\sup_{\nu\in\mathcal{V}_N}|\hat{\psi}_N(f,\nu)|$. $(N,B,M,L)=(4254,800,400,16)$, $\theta=0.4$.

increase M to approximately $M = 200$ when $(\alpha, c) = (0.8, 0.5)$ (so $(N, B) = (1788, 400)$) to observe a noticeable improvement in the estimation error of $\frac{1}{M}\text{tr } f(\hat{\mathbf{C}}(\nu))$.

Appendix

1.A Proof of Lemma 1.12

Lemma 1.12 is a slight variation of Theorem 4.3.2 [13].

Lemma 1.12. *For any ν_1 and ν_2 in $[0,1]$, such that there exists $k \in \{0,1,\ldots,N-1\}$ satisfying $\nu_2 - \nu_1 = k/N$, the following bound holds:*

$$\sup_{m\geq 1} \left|\mathbb{E}\left[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*\right] - s_m(\nu_1)\delta_{\nu_1=\nu_2}\right| = \mathcal{O}\left(\frac{1}{N}\right). \tag{1.151}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*\right] \\
&\qquad = \frac{1}{N}\sum_{n_1,n_2=1}^{N} \mathbb{E}[y_{m,n_1}y^*_{m,n_2}]e^{-2i\pi(n_1-1)\nu_1}e^{2i\pi(n_2-1)\nu_2} \\
&\qquad = \frac{1}{N}\sum_{n_1,n_2=1}^{N} r_{m,n_1-n_2}e^{-2i\pi(n_1-1)\nu_1+2i\pi(n_2-1)\nu_2} \\
&\qquad = \frac{1}{N}\sum_{u=-(N-1),n_1,n_2\in 0,\ldots,N-1}^{(N-1)} r_{m,u}\sum_{n_1-n_2=u} e^{-2i\pi n_1\nu_1+2i\pi n_2\nu_2}
\end{aligned}$$

Splitting this expression for $u=0, u>0$ and $u<0$ provides

$$\begin{aligned}
\mathbb{E}[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*] = \frac{1}{N}r_{m,0}\sum_{n_1=0}^{N-1} e^{-2i\pi n_1(\nu_2-\nu_1)} \\
+ \frac{1}{N}\sum_{u=1}^{(N-1)} r_{m,u}\sum_{n_2=0}^{N-1-u} e^{-2i\pi(u+n_2)\nu_1}e^{2i\pi n_2\nu_2} \\
+ \frac{1}{N}\sum_{u=-(N-1)}^{-1} r_{m,u}\sum_{n_2=-u}^{N-1} e^{-2i\pi(u+n_2)\nu_1}e^{2i\pi n_2\nu_2}.
\end{aligned} \tag{1.152}$$

The first term of the right hand side of (1.152) can be computed in the case $\nu_1=\nu_2$:

$$\frac{1}{N}r_{m,0}\sum_{n_1=0}^{N-1} e^{-2i\pi n_1(\nu_2-\nu_1)} = r_{m,0}$$

and in the case $\nu_1 \neq \nu_2$,

$$\frac{1}{N} r_{m,0} \sum_{n_1=0}^{N-1} e^{-2i\pi n_1 \frac{k}{N}} = 0.$$

Therefore, the first term of the right hand side of (1.152) is equal to $r_{m,0}\delta_{\nu_1=\nu_2}$.

Consider now the second term of (1.152) (where $u > 0$):

$$\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} \sum_{n_2=0}^{N-1-u} e^{-2i\pi(u+n_2)\nu_1} e^{2i\pi n_2 \nu_2}$$
$$= \frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)}. \tag{1.153}$$

The right hand side of (1.153) can also be explicitly written in the case $\nu_1 = \nu_2$:

$$\begin{aligned}\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)} &= \frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1}(N-u) \\ &= \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1}\frac{N-u}{N} \\ &= \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} - \frac{1}{N}\sum_{u=1}^{N-1} u\, r_{m,u} e^{2i\pi u \nu_1}.\end{aligned}$$

By Assumption 1.4, $\sup_{m\geq 1}\sum_{u\in\mathbb{Z}} |u||r_{m,u}| < +\infty$, so we have:

$$\sup_{m\geq 1} \frac{1}{N}\left|\sum_{u=1}^{N-1} u\, r_{m,u} e^{2i\pi u \nu_1}\right| = \mathcal{O}\left(\frac{1}{N}\right).$$

Therefore:

$$\sup_{m\geq 1}\left|\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)} - \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1}\right| = \mathcal{O}\left(\frac{1}{N}\right). \tag{1.154}$$

In the case where $\nu_1 \neq \nu_2$, note that $\nu_1 - \nu_2 = k/N$ with $k \neq 0$, therefore:

$$\sum_{n_2=0}^{N-1} e^{-2i\pi n_2(\nu_2-\nu_1)} = \sum_{n_2=0}^{N-1} e^{-2i\pi n_2 \frac{k}{N}} = 0. \tag{1.155}$$

Using (1.155), one can rewrite the right hand side of (1.153) as

$$\begin{aligned}\left|\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u}e^{-2i\pi u\nu_1}\sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)}\right| \\ &= \left|-\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u}e^{-2i\pi u\nu_1}\sum_{n_2=N-u}^{N} e^{-2i\pi n_2(\nu_2-\nu_1)}\right| \\ &\leq \frac{1}{N}\sum_{u=1}^{N-1}|u||r_{m,u}|\end{aligned}$$

which, again by Assumption 1.4, provides the bound:

$$\sup_{m\geq 1}\left|\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u}e^{-2i\pi u\nu_1}\sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)}\right| = \mathcal{O}\left(\frac{1}{N}\right). \tag{1.156}$$

Combining (1.154) and (1.156), the second term of the right hand side of (1.152) can be estimated as follow:

$$\sup_{m\geq 1}\left|\frac{1}{N}\sum_{u=1}^{(N-1)} r_{m,u}e^{-2i\pi u\nu_1}\sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)} - \delta_{\nu_1=\nu_2}\sum_{u=1}^{N-1} r_{m,u}e^{-2i\pi u\nu_1}\right| = \mathcal{O}\left(\frac{1}{N}\right).$$

The term for $u<0$ in equation (1.152) is similar. Gathering the three terms of equation (1.152) leads to

$$\sup_{m\geq 1}\left|\mathbb{E}[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*] - \delta_{\nu_1=\nu_2}\left(\sum_{u=-(N-1)}^{N-1} r_{m,u}e^{-2i\pi u\nu_1}\right)\right| = \mathcal{O}\left(\frac{1}{N}\right). \tag{1.157}$$

Finally, using again Assumption 1.4 we have:

$$\left|\sum_{|u|>N} r_m(u)e^{-2i\pi u\nu_1}\right| \leq \frac{1}{N}\sum_{|u|>N}|u||r_m(u)| = \mathcal{O}\left(\frac{1}{N}\right).$$

Inserting this into equation (1.157), we obtain equation (1.151) □

1.B Proof of Lemma 1.3

Proof. Consider the complement of the event $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ and notice that:

$$\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)^c \subset \{\exists m\in\{1,\ldots,M\} : \hat{s}_m > \bar{s}+\epsilon\} \cup \{\exists m\in\{1,\ldots,M\} : \hat{s}_m < \underline{s}-\epsilon\}. \tag{1.158}$$

We start by proving that the first set of the right hand side of (1.158) holds with exponentially small probability, ie. for any $\epsilon>0$, there exist $\gamma>0$ such that:

$$\mathbb{P}\left[\exists m\in\{1,\ldots,M\} : \hat{s}_m > \bar{s}+\epsilon\right] \leq \exp -N^\gamma.$$

By Lemma 1.13 (see below), $|\mathbb{E}\hat{s}_m - s_m| = \mathcal{O}(B^2/N^2)$ so for N large enough, this bias term will be smaller than $\epsilon/2$. Moreover, for any $m\in\{1,\ldots,M\}$, $s_m-\bar{s}\leq 0$. Therefore, one can write for large enough N:

$$\begin{aligned}
&\mathbb{P}\left[\exists m \in \{1,\ldots,M\} : \hat{s}_m > \bar{s} + \epsilon\right] \\
&\qquad = \mathbb{P}\left[\sup_{m\in\{1,\ldots,M\}} (\hat{s}_m - \mathbb{E}\hat{s}_m + \mathbb{E}\hat{s}_m - s_m + s_m - \bar{s}) > \epsilon\right] \\
&\qquad \leq \mathbb{P}\left[\sup_{m\in\{1,\ldots,M\}} |\hat{s}_m - \mathbb{E}\hat{s}_m| > \epsilon/2\right]
\end{aligned}$$

which holds with exponentially high probability by Lemma 1.14 (see below). The proof for the lower bound is similar. □

It remains to prove Lemma 1.13 and Lemma 1.14. Concerning the proof of Lemma 1.13, we follow the same approach as the one used in Theorem 5.4.2 in [13].

Lemma 1.13. *For any $\nu \in [0,1]$, the following results hold:*

$$\mathbb{E}(\hat{s}_m(\nu)) - s_m(\nu) = \frac{s''_m(\nu)}{2} v_N + \mathcal{O}\left(\left(\frac{B}{N}\right)^3 + \frac{1}{N}\right) \tag{1.159}$$

and

$$\sup_{m=1,\ldots,M} |\mathbb{E}\hat{s}_m(\nu) - s_m(\nu)| = \mathcal{O}\left(\frac{B}{N}\right)^2. \tag{1.160}$$

Proof. It is clear that $\hat{s}_m(\nu) = \hat{\mathbf{S}}_{m,m}(\nu) = s_m(\nu)\tilde{\mathbf{C}}_{m,m}(\nu)$ can be written as

$$\hat{s}_m(\nu) = s_m(\nu)\frac{x_m(\mathbf{I} + \mathbf{\Phi}_m)x_m^*}{B+1}. \tag{1.161}$$

Therefore, $\mathbb{E}(\hat{s}_m(\nu)) = s_m(\nu)(1 + \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m)$. (1.159) thus follows immediately from (1.54). (1.160) is an immediate consequence of (1.159). □

Lemma 1.14. *The family of random variables* $\sup_{m=1,\ldots,M} |\hat{s}_m(\nu) - \mathbb{E}[\hat{s}_m(\nu)]|, \nu \in [0,1]$ *satisfies*

$$\sup_{m=1,\ldots,M} |\hat{s}_m - \mathbb{E}[\hat{s}_m]| \prec \frac{1}{\sqrt{B}}. \tag{1.162}$$

Proof. (1.161) implies that $\hat{s}_m - \mathbb{E}[\hat{s}_m]$ can be written as $\hat{s}_m - \mathbb{E}[\hat{s}_m] = s_m\left(\frac{x_m(\mathbf{I}+\mathbf{\Phi}_m)x_m^*}{B+1} - \frac{1}{B+1}\mathrm{Tr}(\mathbf{I} + \mathbf{\Phi}_m)\right)$. It is clear that $\sup_m \|\frac{(\mathbf{I}+\mathbf{\Phi}_m)}{B+1}\|_F \leq \frac{C}{B}$ for some nice constant C. Therefore, (1.162) leads immediately to (1.41) . □

1.C Proof of Lemma 1.5

Proof. These estimates can be proved in a compact way by using the calculus rules available in the stochastic domination framework introduced in Definition 1.1 and proved in Lemma 1.1. Using

Lemma 1.4 and Lemma 1.15 (see below):

$$\begin{aligned}\left|\frac{1}{\sqrt{\hat{s}_m}}-\frac{1}{\sqrt{s_m}}\right| &= \left|\frac{\sqrt{s_m}-\sqrt{\hat{s}_m}}{\sqrt{s_m}\sqrt{\hat{s}_m}}\right| \\ &\leq \underbrace{\left|\sqrt{s_m}-\sqrt{\hat{s}_m}\right|}_{\mathcal{O}_\prec(\frac{1}{\sqrt{B}}+\frac{B^2}{N^2})} \times \underbrace{\left|\sqrt{\frac{1}{s_m}}\right|}_{\mathcal{O}_\prec(1)} \times \underbrace{\left|\sqrt{\frac{1}{\hat{s}_m}}\right|}_{\mathcal{O}_\prec(1)} \\ &\prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}.\end{aligned}$$

The second inequality is similar to prove:

$$\begin{aligned}\left|\sqrt{\frac{s_m}{\hat{s}_m}}-1\right| &= \left|\frac{\sqrt{s_m}-\sqrt{\hat{s}_m}}{\sqrt{\hat{s}_m}}\right| \\ &\leq \underbrace{|s_m-\hat{s}_m|}_{\mathcal{O}_\prec(\frac{1}{\sqrt{B}}+\frac{B^2}{N^2})} \times \underbrace{\left|\frac{1}{\hat{s}_m(\sqrt{s_m+\hat{s}_m})}\right|}_{\mathcal{O}_\prec(1)} \\ &\prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}.\end{aligned}$$

□

Lemma 1.15. *The family of random variables* $(\sup_{m=1,\ldots,M}|\hat{s}_m(\nu)-s_m(\nu)|)$, $\nu\in[0,1]$ *satisfies*

$$\sup_{m=1,\ldots,M}|\hat{s}_m-s_m| \prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}.$$

Proof. It is sufficient to check that the family of random variables $(|\hat{s}_m-s_m|)_{m=1,\ldots,M}, \nu\in[0,1]$ satisfies $|\hat{s}_m-s_m|\prec\frac{1}{\sqrt{B}}+\frac{B^2}{N^2}$. Using Lemma 1.13 and Lemma 1.14, we obtain as expected that

$$|\hat{s}_m-s_m| = |s_m-\mathbb{E}\hat{s}_m+\mathbb{E}\hat{s}_m-\hat{s}_m| \leq \underbrace{|s_m-\mathbb{E}\hat{s}_m|}_{\mathcal{O}(\frac{B^2}{N^2})}+\underbrace{|\mathbb{E}\hat{s}_m-\hat{s}_m|}_{\mathcal{O}_\prec(\frac{1}{\sqrt{B}})} \prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}.$$

□

1.D Proof of Lemma 1.16

Lemma 1.16. *The set of random variable* $(\sum_{m=1}^M|\hat{s}_m(\nu)-s_m(\nu)|^2)$, $\nu\in[0,1]$ *satisfies*

$$\sum_{m=1}^M|\hat{s}_m-s_m|^2 \prec 1+\frac{B^5}{N^4}.$$

Proof. Using Lemma 1.15, we have

$$|\hat{s}_m-s_m|^2 \prec \frac{1}{B}+\frac{B^4}{N^4}$$

and summing over $m = 1 \dots M$, one immediately get:

$$\sum_{m=1}^{M} |\hat{s}_m - s_m|^2 \prec 1 + \frac{B^5}{N^4}.$$

$\square$

1.E Proof of Lemma 1.10

We express $\tilde{\mathbf{\Delta}}$ as $\tilde{\mathbf{\Delta}} = \frac{\mathbf{X}\mathbf{\Gamma}^*}{B+1} + \frac{\mathbf{\Gamma}\mathbf{X}^*}{B+1} + \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1}$. Therefore, we have

$$\|\tilde{\mathbf{\Delta}}\|^k \leq C \left(\left\| \frac{\mathbf{X}}{\sqrt{B+1}} \right\|^k \left\| \frac{\mathbf{\Gamma}}{\sqrt{B+1}} \right\|^k + \left\| \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1} \right\|^k \right).$$

Using the Schwartz inequality, we obtain that

$$\mathbb{E}\|\tilde{\mathbf{\Delta}}\|^k \leq C \left(\left(\mathbb{E} \left\| \frac{\mathbf{X}\mathbf{X}^*}{B+1} \right\|^k \right)^{1/2} \left(\mathbb{E} \left\| \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1} \right\|^k \right)^{1/2} + \mathbb{E} \left\| \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1} \right\|^k \right).$$

It is well-known that $\mathbb{E}\left(\left\| \frac{\mathbf{X}\mathbf{X}^*}{B+1} \right\|^k \right) \leq C$ for some nice constant depending on k. Therefore, we establish that

$$\mathbb{E} \left\| \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1} \right\|^k \leq C \left(\frac{B}{N} \right)^{2k}$$

a property which will imply that $\mathbb{E}\|\tilde{\mathbf{\Delta}}\|^k \leq C \left(\frac{B}{N}\right)^k$. For this, we put $\mathbf{Z} = \frac{\mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1}$. As (1.62) holds, it remains to verify that $\mathbb{E}(\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\|^k) = \mathcal{O}\left(\frac{B}{N}\right)^{2k}$. For this, we use the concentration inequality (1.68). We choose $t_N = w^{1/k}(\frac{B}{N})^2$, and obtain that

$$\mathbb{P}\left[\frac{\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\|}{(\frac{B}{N})^2} > w^{1/k} \right] \leq 2\, C_0 \exp -CB(w^{1/k} - w_0^{1/k}) \tag{1.163}$$

for some $w_0 > 0$. If we denote by z_N the random variable $z_N = \left(\frac{\|\mathbf{Z} - \mathbb{E}(\mathbf{Z})\|}{(\frac{B}{N})^2} \right)^k$, we have to establish that $\mathbb{E}(z_N) = \mathcal{O}(1)$. For this, we express $\mathbb{E}(z_N)$ as

$$\mathbb{E}(z_N) = \int_0^{+\infty} \mathbb{P}(z_N > w)\, dw = \int_0^{w_0} \mathbb{P}(z_N > w) dw + \int_{w_0}^{+\infty} \mathbb{P}(z_N > w)\, dw.$$

As $\mathbb{P}(z_N > w) = \mathbb{P}(z_N^{1/k} > w^{1/k})$, (1.163) immediately implies that $\mathbb{E}(z_N) = \mathcal{O}(1)$.

1.F Proof of Lemma 1.11

We denote by η_m the term of interest, i.e. $\eta_m = \mathbb{E}(\mathbf{w}_m \mathbf{A}\mathbf{W}^* \mathbf{Q}\mathbf{e}_m)$. It can be written as

$$\eta_m = \sum_{n_1} \left(\sum_{n_2, m'} \mathbb{E}(\mathbf{W}_{m,n_2} \overline{\mathbf{W}}_{m',n_1} \mathbf{Q}_{m',m}) \mathbf{A}_{n_2,n_1} \right).$$

The integration by parts formula (1.32) leads to

$$\mathbb{E}(\mathbf{W}_{m,n_2}\overline{\mathbf{W}}_{m',n_1}\mathbf{Q}_{m',m}) = \delta_{m-m'}\delta_{n_1-n_2}\frac{1}{B+1}\mathbb{E}(\mathbf{Q}_{m,m}) - \frac{1}{B+1}\mathbb{E}\left[\overline{\mathbf{W}}_{m',n_1}(\mathbf{Q}\mathbf{W})_{m',n_2}\mathbf{Q}_{m,m}\right].$$

Therefore, we obtain that

$$\sum_{n_2,m'}\mathbb{E}(\mathbf{W}_{m,n_2}\overline{\mathbf{W}}_{m',n_1}\mathbf{Q}_{m',m})\mathbf{A}_{n_2,n_1} = \frac{1}{B+1}\mathbb{E}(\mathbf{Q}_{m,m})\mathbf{A}_{n_1,n_1} - \frac{1}{B+1}\mathbb{E}\left[(\mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A})_{n_1,n_1}\mathbf{Q}_{m,m}\right]$$

and that

$$\begin{aligned}\eta_m &= \mathbb{E}[\mathbf{Q}_{m,m}]\frac{1}{B+1}\mathrm{tr}\ \mathbf{A} - \mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right)\mathbf{Q}_{m,m}\right] \\ &= \beta\frac{1}{B+1}\mathrm{tr}\ \mathbf{A} - \beta\,\mathbb{E}\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right) - \mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right)^\circ\mathbf{Q}^\circ_{m,m}\right]. \end{aligned} \tag{1.164}$$

In order to evaluate $\mathbb{E}\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right)$, we note that

$$\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right) = \frac{1}{M}\sum_{m=1}^{M}\eta_m = \beta\frac{1}{B+1}\mathrm{tr}\ \mathbf{A} - \beta\, c\,\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right) - \mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right)^\circ\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^\circ\right]$$

from which we deduce that

$$\mathbb{E}\left(\frac{1}{M}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right) = \frac{\beta}{1+\beta c}\,\frac{1}{B+1}\mathrm{tr}\ \mathbf{A} - \frac{1}{1+\beta c}\mathbb{E}\left[\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{W}^*\mathbf{Q}\mathbf{W}\mathbf{A}\right)^\circ\frac{1}{M}\mathrm{tr}\ \mathbf{Q}^\circ\right].$$

Plugging this relation into (1.164) leads immediately to (1.126).

Chapter 2

Approach based on large random matrix models with independent rows

2.1 Introduction

In Chapter 1, to estimate $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int f \,\mathrm{d}\mu_{mp,c_N}$, we used the following decomposition:

$$
\begin{aligned}
&\left(\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))\right) + \left(\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))\right) \\
&\qquad + \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right) \\
&\qquad\qquad + \left(\mathbb{E}\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right) - \int f \,\mathrm{d}\mu_{mp,c_N}\right)
\end{aligned}
$$

where for each ν, $\mathbf{X}(\nu)$ is an $M \times (B+1)$ matrix with iid $\mathcal{N}_{\mathbb{C}}(0,1)$ entries. The goal of this chapter is to provide an alternative approach for the study of $\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \int f \,\mathrm{d}\mu_{mp,c_N}$, which will turn out to provide more easily the $\mathcal{O}(\frac{B}{N})^2$ correction terms than with the method developed in Chapter 1. More precisely, we will show that the quantity of interest $\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))$ can be approximated by a deterministic quantity $\int f \,\mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)}$ which satisfies, under the assumption of independence of the M time series,

$$
\begin{aligned}
\int f \,\mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)} - \int f \,\mathrm{d}\mu_{mp,c_N} = &- < \tilde{\mathcal{D}}_N, f > \tilde{r}_N(\nu) v_N \\
&+ < \mathcal{D}_N, f > r_N(\nu) v_N + \mathcal{O}\left(\frac{B}{N}\right)^3
\end{aligned}
\tag{2.1}
$$

where $\mathcal{O}(\cdot)$ is uniform in $\nu \in [0,1]$, and where $< \tilde{\mathcal{D}}_N, f >$, $< \mathcal{D}_N, f >$, $\tilde{r}_N(\nu)$, $r_N(\nu)$, v_N and c_N are defined as in Chapter 1. Moreover, using Proposition 3 from [48], it will be possible to prove that $\mu^{\tilde{\mathbf{T}}(\nu)}$ satisfies

$$
\frac{1}{M}\mathrm{tr}\ \mathbb{E} f(\tilde{\mathbf{C}}(\nu)) - \int f \,\mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)} = \mathcal{O}\left(\frac{1}{B^2}\right) \tag{2.2}
$$

uniformly over $\nu \in [0,1]$. To obtain (2.2), we show that $\mu^{\tilde{\mathbf{T}}(\nu)}$ is defined as the spectral measure associated with the Stieltjes transform $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}(\nu, z)$, with $\tilde{\mathbf{T}}(\nu, z)$ defined as the solution of the following coupled system of equations

$$\begin{cases} \tilde{\mathbf{T}}(\nu, z) = \mathrm{dg}\left(\frac{-1}{z(1+\overline{\delta}_m(\nu,z))}, m = 1, \ldots, M\right) \\ \\ \overline{\mathbf{T}}(\nu, z) = \left(-z\mathbf{I}_{B+1} - \frac{1}{M}\sum_{m=1}^{M} cz\tilde{T}_{mm}(\nu, z)\mathbf{\Theta}_{N,m}(\nu)\right)^{-1} \end{cases}$$

where

$$\overline{\delta}_m(\nu, z) = \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_{N,m}(\nu)\overline{\mathbf{T}}(\nu, z)$$

and $\mathbf{\Theta}_{N,m}(\nu)$ are the deterministic sequence of matrices defined by

$$(\mathbf{\Theta}_{N,m})_{bb'}(\nu) = \mathbb{E}\left[\frac{\xi_m\left(\nu - \frac{b}{2N}\right)}{\sqrt{s_m(\nu)}} \overline{\frac{\xi_m\left(\nu - \frac{b'}{2N}\right)}{\sqrt{s_m(\nu)}}}\right]. \tag{2.3}$$

To show (2.1), we use the fact that under the joint independence assumption of the M time series, the matrices $\mathbf{\Theta}_{N,m}(\nu)$ are all close to $\mathbf{I}_{B+1}$ as $N \to +\infty$. This allows to extract the dominant terms from the expression of $\int f\, \mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)} - \int f\, \mathrm{d}\mu_{mp,c_N}$.

2.1.1 Comparison with [60]

This work can be compared with [60] since they share similarities in their goals and the tools used. Concerning the goals, both are motivated by the test of independence of multidimensional time series. The approach proposed by the authors in [60] relies on the lag domain. They consider the column vector gathering L consecutive observations of the m–th time series, namely

$$\mathbf{y}_{m,n}^L = [y_{m,n}, \ldots, y_{m,n+L-1}]^T$$

and from this built an ML–dimensional column vector

$$\mathbf{y}_n^L = \left[(\mathbf{y}_{1,n}^L)^T, \ldots, \left(\mathbf{y}_{M,n}^L\right)^T\right]^T.$$

Denoting by $\mathcal{R}_L$ the $ML \times ML$ spatio-temporal covariance matrix of this random vector, i.e. $\mathcal{R}_L = \mathbb{E}[\mathbf{y}_N^L(\mathbf{y}_N^L)^*]$, the M Gaussian time series are independent if and only if $\mathcal{R}_L$ is block-diagnoal, or equivalently if $\mathcal{R}_{corr,L} = \mathbf{I}_{ML}$ where

$$\mathcal{R}_{corr,L} = \mathcal{B}_L^{-1/2}\mathcal{R}_L\mathcal{B}_L^{-1/2}$$

with

$$\mathcal{B}_L = \mathrm{Bdiag}(\mathcal{R}_L).$$

To test if $\mathcal{R}_{corr,L}$ is equal to $\mathbf{I}_{ML}$, the authors from [60] studied the behaviour of LSS of $\hat{\mathcal{R}}_{corr,L}$, an estimator of $\mathcal{R}_{corr,L}$ defined by

$$\hat{\mathcal{R}}_{corr,L} = \hat{\mathcal{B}}_L^{-1/2}\hat{\mathcal{R}}_L\hat{\mathcal{B}}_L^{-1/2}, \quad \hat{\mathcal{R}}_L = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n^L(\mathbf{y}_n^L)^*.$$

The authors considered the setting where $\frac{ML}{N} := c_N \xrightarrow{N\to+\infty} c \in (0, +\infty)$, the M–time series are jointly independent, and they satisfy the following short memory assumption:

$$\text{there exists a constant } \gamma_0 > 0 \text{ such that } \sup_{m\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)^{\gamma_0}|r_m(k)| < +\infty$$

where $(r_m(k))_{k\in\mathbb{Z}}$ is the autocovariance sequence associated with the m–th component of the time series $\mathbf{y}$. Denote $(\hat{\lambda}_m)_{m\in[ML]}$ the eigenvalues of $\hat{\mathcal{R}}_L$ and

$$\hat{\mu}_N(I) = \frac{1}{ML}\sum_{m=1}^{ML} \mathbb{1}(\hat{\lambda}_m \in I).$$

The authors then followed the same approach as in this chapter, and showed that the spectral measure $\hat{\mu}_N$ can be approximated by a deterministic measure μ_N, which allows to prove the following approximation:

$$\int \phi(\lambda)\,\mathrm{d}\hat{\mu}_N(\lambda) - \int \phi(\lambda)\,\mathrm{d}\mu_{mp,c_N}(\lambda) \prec \frac{1}{L^{\gamma_0}} + \frac{1}{M} + \frac{1}{M\sqrt{L}} + \frac{1}{M^2} + \frac{1}{L^2}$$

where ϕ is a suitable function and $\mathrm{d}\hat{\mu}_N$ is the empirical eigenvalue distribution associated with $\hat{\mathcal{R}}_{corr,L}$.

We now compare this result (concerning the behaviour of the eigenvalues of the lag autocovariance matrix) with the one developed in this chapter (about the behaviour of the eigenvalues of the spectral coherency matrix). First, we consider only the case where $\gamma_0 \geq 2$ since this is the assumption made in Chapter 1. In this case, the result from [60] can be simplified:

$$\int \phi(\lambda)\,\mathrm{d}\hat{\mu}_N(\lambda) - \int \phi(\lambda)\,\mathrm{d}\mu_{mp,c_N}(\lambda) \prec \frac{1}{M} + \frac{1}{L^2}.$$

Moreover, we see that the parameter $\frac{1}{L}$ plays the role of the quantity $\frac{B}{N}$ in this work. Indeed, if we denote $\alpha < 1$ such that $M = \mathcal{O}(N^\alpha)$, we see that the condition $\frac{ML}{N} \to c$ implies that $\frac{1}{L} = \mathcal{O}(N^{1-\alpha})$. In this chapter, M and B are both $\mathcal{O}(N^\alpha)$, so $\frac{B}{N} = \mathcal{O}(N^{1-\alpha})$. Since we show that

$$\int \phi(\lambda)\,\mathrm{d}\hat{\mu}_{N,\nu}(\lambda) - \int \phi(\lambda)\,\mathrm{d}\mu_{mp,c_N}(\lambda) \prec \frac{1}{M} + \left(\frac{B}{N}\right)^2$$

one can see that up to the identification $\frac{1}{L} \leftrightarrow \frac{B}{N}$, the eigenvalues of $\hat{\mathcal{R}}_L$ and $\hat{\mathbf{C}}(\nu)$ converges towards the Marcenko-Pastur distribution at the same rate. However, in this chapter we could obtain the explicit expression of the second order term (which is $\mathcal{O}\left(\frac{B}{N}\right)^2$), while it has not been done in [60].

2.1.2 Notations

In the rest of the chapter we will often use the simpler notation $\boldsymbol{\Theta}_m$ instead of $\boldsymbol{\Theta}_{m,N}(\nu)$ (it will be clear from the context that the values of N and ν have been fixed). Moreover, note that $\tilde{\mathbf{C}}(\nu)$ can be written as the empirical covariance matrix of the normalized Fourier transform matrix $\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}(\nu)$:

$$\tilde{\mathbf{C}}(\nu) = \left(\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}(\nu)\right)\left(\mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}(\nu)\right)^*$$

where

$$\mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}}\boldsymbol{\Sigma}(\nu) = \frac{1}{\sqrt{B+1}}\begin{bmatrix} \frac{\zeta_1\left(\nu-\frac{B}{2N}\right)}{\sqrt{s_1(\nu)}} & \dots & \frac{\zeta_1\left(\nu+\frac{B}{2N}\right)}{\sqrt{s_1(\nu)}} \\ \vdots & & \vdots \\ \frac{\zeta_M\left(\nu-\frac{B}{2N}\right)}{\sqrt{s_M(\nu)}} & \dots & \frac{\zeta_M\left(\nu+\frac{B}{2N}\right)}{\sqrt{s_M(\nu)}} \end{bmatrix}.$$

We define $\tilde{\mathbf{Q}}(\nu,z)$ (respectively $\overline{\mathbf{Q}}(\nu,z)$) as the resolvent (respectively co-resolvent) of $\tilde{\mathbf{C}}(\nu)$:

$$\tilde{\mathbf{Q}}(\nu,z) = \left(\mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} \boldsymbol{\Sigma}(\nu)\boldsymbol{\Sigma}(\nu)^* \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} - z\mathbf{I}_M\right)^{-1}$$
$$\overline{\mathbf{Q}}(\nu,z) = \left(\boldsymbol{\Sigma}(\nu)^* \mathrm{dg}\left(\mathbf{S}(\nu)\right)^{-1} \boldsymbol{\Sigma}(\nu) - z\mathbf{I}_{B+1}\right)^{-1}.$$

The spectral radius of a matrix $\mathbf{M}$ is denoted $\rho(\mathbf{M})$, and is defined by:

$$\rho(\mathbf{M}) = \max\{|\lambda_i|, \lambda_i \text{ eigenvalue of } \mathbf{M}\}.$$

while the l_1 norm of $\mathbf{M}$ is defined by $\|\mathbf{M}\|_1 = \sup_{k\in[M]} \sum_{l\in[M]} |M_{kl}|$. The spectral norm is denoted $\|\mathbf{M}\| := \sqrt{\lambda_1(\mathbf{M}\mathbf{M}^*)}$ where $\lambda_1(\mathbf{M}\mathbf{M}^*)$ is the largest eigenvalue of $\mathbf{M}\mathbf{M}^*$. We also recall the resolvent identity: if $\mathbf{Q} = (\mathbf{X}\mathbf{X}^* - z\mathbf{I})$, then

$$\mathbf{I} + z\mathbf{Q} = \mathbf{Q}\mathbf{X}\mathbf{X}^* = \mathbf{X}\mathbf{X}^*\mathbf{Q}.$$

Finally we denote $[M] = \{1, \ldots, M\}$.

2.1.3 Some results on the Stieltjes transform

As the Stieltjes transform will be a central tool in this chapter, we recall some basic useful facts. We begin with

Definition 2.1. *Let μ be a finite positive measure carried by $\mathbb{R}_+$. The Stieltjes transform of μ is the function s defined for $z \in \mathbb{C} \setminus \mathbb{R}_+$ by*

$$s(z) = \int_{\mathbb{R}_+} \frac{\mathrm{d}\mu(\lambda)}{\lambda - z}.$$

In the following, the class of Stieltjes transform of finite positive measures carried by $\mathbb{R}_+$ is denoted $\mathcal{S}(\mathbb{R}_+)$.

We now state some properties of the elements of $\mathcal{S}(\mathbb{R}_+)$.

Lemma 2.1. *Let $s \in \mathcal{S}(\mathbb{R}_+)$ and μ its associated measure. Then we have the following results:*

1. *s is analytic on $\mathbb{C}_+$,*
2. *$\mathrm{Im} s(z) > 0$ if $\mathrm{Im} z > 0$, and $\mathrm{Im} s(z) < 0$ if $\mathrm{Im} z < 0$,*
3. *$\mathrm{Im}(zs(z)) > 0$ if $\mathrm{Im} z > 0$, and $\mathrm{Im}(zs(z)) < 0$ if $\mathrm{Im} z < 0$,*
4. *$|s(z)| \leq \frac{\mu(\mathbb{R}_+)}{d(z,\mathbb{R}_+)}$, where $d(z, \mathbb{R}_+)$ is the euclidean distance from z to $\mathbb{R}_+$,*
5. *$\mu(\mathbb{R}_+) = \lim_{y\to+\infty} -iys(iy)$.*

Proof. All the stated properties are standard materials, see e.g. Appendix of [53]. □

Conversely, we have the following lemma.

Lemma 2.2. *Let f be an analytic function over $\mathbb{C}_+$ such that $f(z) \in \mathbb{C}_+$ and $zf(z) \in \mathbb{C}_+$ if $z \in \mathbb{C}_+$. If $\lim_{y\to+\infty} -iyf(iy) = 1$, then f is the Stieltjes transform of a probability measure μ carried by $\mathbb{R}_+$, and f is given by*

$$f(z) = \int_{\mathbb{R}_+} \frac{\mu(\mathrm{d}\lambda)}{\lambda - z}.$$

Moreover, f has an analytic continuation on $\mathbb{C} \setminus \mathbb{R}_+$.

Proof. See Appendix of [53]. □

In this chapter, we will need to extend the concept of Stietljes transform to matrix-valued functions, as it was done in [35] for instance. Therefore, we recall here Proposition 2.2 from [35].

Lemma 2.3. *Let F be an $N \times N$ matrix-valued function over $\mathbb{C}_+$ such that*

1. *the function F is analytic,*
2. *the matrices* $\mathrm{Im}F(z)$ *and* $\mathrm{Im}zF(z)$ *satisfy*

$$\mathrm{Im}F(z) \geq 0 \quad \text{and} \quad \mathrm{Im}(zF(z)) \geq 0 \quad \text{for } z \in \mathbb{C}_+,$$

3. *For any $\epsilon > 0$, $\sup_{y \geq \epsilon} \|iyF(iy)\| < +\infty$.*

Then there exists a positive $N \times N$ matrix-valued measure μ carried over $\mathbb{R}_+$ such that

$$F(z) = \int \frac{\mu(\mathrm{d}\lambda)}{\lambda - z}.$$

Moreover, the total mass of the measure μ is equal to

$$\mu(\mathbb{R}_+) = -\lim_{y \to +\infty} iyF(iy).$$

Proof. See Appendix A of [35]. □

The most important Stieltjes transform in this chapter will be denoted $\tilde{t}_N$ and represents the Stieltjes transform of the Marcenko-Pastur distribution. For a parameter $c_N > 0$, $\tilde{t}_N$ is defined as the solution in the set of Stieltjes transforms of

$$zc_N\tilde{t}_N^2(z) + (z + c_N - 1)\tilde{t}_N(z) + 1 = 0.$$

From this equation, it is clear that $\tilde{t}_N$ satisfies the following useful relation:

$$\tilde{t}_N(z) = \frac{1}{-z + \frac{1}{1+c_N\tilde{t}_N(z)}}. \tag{2.4}$$

Finally, we define $\bar{t}_N(z)$ as:

$$\bar{t}_N(z) = c_N\tilde{t}_N(z) - \frac{1-c_N}{z}$$

so that $(\tilde{t}_N(z), \bar{t}_N(z))$ can be written as the solution in the set of the Stieltjes transform of the following coupled equations:

$$\tilde{t}_N(z) = \frac{-1}{z(1+\bar{t}_N(z))}, \quad \bar{t}_N(z) = \frac{-1}{z(1+c_N\tilde{t}_N(z))}.$$

Note that we do not use the more traditional notation $z \mapsto \tilde{t}_N(z)$ since we will show later that $\tilde{\mathbf{T}} \approx \tilde{t}_N\mathbf{I}_M$. In the rest of the chapter, we will also use the notation $\tilde{t}_N$ and $\bar{t}_N$ instead of $\tilde{t}_N$ and $\bar{t}_N$. We end this section with a useful preliminary result.

2.2 Statement of the results

Throughout this chapter we rely on the same assumptions as those from Chapter 1. Proposition 2.1 states that the expectation of the resolvent of $\tilde{\mathbf{C}}(\nu)$ written $\mathbb{E}\tilde{\mathbf{Q}}(\nu, z)$ and the expectation of the co-resolvent of $\tilde{\mathbf{C}}(\nu)$ denoted $\mathbb{E}\overline{\mathbf{Q}}(\nu, z)$ are asymptotically close to deterministic quantities. Before that, we need to introduce them and show that they are well-defined. For this, we first need Appendix I-B from [92] that we rewrite here.

Lemma 2.4. *Let $(\mathbf{\Theta}_m)_{m\in[M]}$ be a sequence of $(B+1)\times(B+1)$ Hermitian positive definite matrices such that $\sup_{m\in[M]}\|\mathbf{\Theta}_m\| < +\infty$. There exists a unique solution $(\overline{\delta}_1,\ldots,\overline{\delta}_M)$ in $\mathcal{S}(\mathbb{R}_+)$, the set of Stieltjes transform carried by $\mathbb{R}_+$, satisfying the following equations:*

$$\overline{\delta}_m = \frac{1}{B+1}\mathrm{tr}\left(-z\mathbf{I}_{B+1} + \frac{c}{M}\sum_{i=1}^{M}\frac{\mathbf{\Theta}_i}{1+\overline{\delta}_i}\right)^{-1}\mathbf{\Theta}_m, \quad m\in[M]. \tag{2.5}$$

Moreover, if we denote $\overline{\mu}_m$ the measure associated with $\overline{\delta}_m$, we have

$$\overline{\mu}_m(\mathbb{R}_+) = \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m.$$

Proof. See Appendix I-B from [92] for the existence and unicity in the set $\mathcal{S}(\mathbb{R}_+)$ of the functions $\overline{\delta}_m$. Lastly, it can be checked that $\lim_{y\to+\infty} -iy\overline{\delta}_m(iy) = \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m$. □

Let the deterministic matrices $\tilde{\mathbf{T}}(z)$ and $\overline{\mathbf{T}}(z)$ be defined by

$$\begin{cases} \tilde{\mathbf{T}}(\nu,z) = \mathrm{dg}\left(\frac{-1}{z(1+\overline{\delta}_m(\nu,z))}, m=1,\ldots,M\right) \\ \overline{\mathbf{T}}(\nu,z) = \left(-z\mathbf{I}_{B+1} - \frac{1}{M}\sum_{m=1}^{M} cz\tilde{T}_{mm}(\nu,z)\mathbf{\Theta}_m(\nu)\right)^{-1} \end{cases} \tag{2.6}$$

where the matrices $\mathbf{\Theta}_m(\nu)$ are defined by (2.3). It can be checked that under Assumption 1.4, $\sup_{\nu\in[0,1]}\sup_{N\geq 1}\sup_{m\in[M]}\|\mathbf{\Theta}_m(\nu)\| < +\infty$. Note that the matrices $\tilde{\mathbf{T}}(\nu,z)$ and $\overline{\mathbf{T}}(\nu,z)$ are fully deterministic.

Remark 2.1. *Note that $\overline{\delta}_m(\nu,z)$ is equivalently defined by:*

$$\overline{\delta}_m(\nu,z) := \frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(\nu,z)\mathbf{\Theta}_m(\nu).$$

Moreover, the following important properties for $\tilde{\mathbf{T}}$ and $\overline{\mathbf{T}}$ hold. The statement of these properties and the proofs are very similar to Proposition 5.1 from [35], but we detail them for the sake of completeness.

Lemma 2.5. *Let $\tilde{\mathbf{T}}$ and $\overline{\mathbf{T}}$ defined by (2.6). The following properties hold:*

1. *For each $\nu\in[0,1]$, $z\mapsto\tilde{\mathbf{T}}(\nu,z)$ and $z\mapsto\overline{\mathbf{T}}(\nu,z)$ are holomorphic on $\mathbb{C}_+$,*
2. *For each $\nu\in[0,1]$, if $z\in\mathbb{C}_+$, then $\mathrm{Im}\tilde{\mathbf{T}}(\nu,z)\geq 0$, $\mathrm{Im}(z\tilde{\mathbf{T}}(\nu,z))\geq 0$,$\mathrm{Im}\overline{\mathbf{T}}(\nu,z)\geq 0$ and $\mathrm{Im}(z\overline{\mathbf{T}}(\nu,z))\geq 0$. Moreover, there exists a positive $M\times M$ matrix-valued measure $\mu^{\tilde{\mathbf{T}}(\nu)} = (\mu_{ij}^{\tilde{\mathbf{T}}(\nu)})_{ij}$ and a positive $(B+1)\times(B+1)$ matrix-valued measure $\mu^{\overline{\mathbf{T}}(\nu)} = (\mu_{ij}^{\overline{\mathbf{T}}(\nu)})_{ij}$ such that*

   $$\mu^{\tilde{\mathbf{T}}(\nu)}(\mathbb{R}_+) = \mathbf{I}_M, \quad \mu^{\overline{\mathbf{T}}(\nu)}(\mathbb{R}_+) = \mathbf{I}_{B+1}$$
   $$\tilde{\mathbf{T}}(\nu,z) = \int_{\mathbb{R}}\frac{\mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)}(\lambda)}{\lambda - z}, \quad \overline{\mathbf{T}}(\nu,z) = \int_{\mathbb{R}}\frac{\mathrm{d}\mu^{\overline{\mathbf{T}}(\nu)}(\lambda)}{\lambda - z}$$

   for $z\in\mathbb{C}_+$,
3. *For each $\nu\in[0,1]$, the following inequalities hold true:*

   $$\forall z\in\mathbb{C}_+,\quad \tilde{\mathbf{T}}(\nu,z)\tilde{\mathbf{T}}(\nu,z)^* \leq \frac{\mathbf{I}_M}{\mathrm{Im}^2 z}, \quad \text{and} \quad \overline{\mathbf{T}}(\nu,z)\overline{\mathbf{T}}(\nu,z)^* \leq \frac{\mathbf{I}_{B+1}}{\mathrm{Im}^2 z},$$

4. The sequence of measures $(\frac{1}{M}\mathrm{tr}\ \mu^{\tilde{\mathbf{T}}(\nu)})_{N\geq 1}$ *and* $(\frac{1}{B+1}\mu^{\overline{\mathbf{T}}(\nu)})_{N\geq 1}$ *are tight.*

In particular, $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}$ *and* $\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}$ *are Stieltjes transforms of probability measures.*

Proof. We fix a $\nu \in [0,1]$ and omit the dependence in ν for $\tilde{\mathbf{T}}$ and $\overline{\mathbf{T}}$ in the rest of the proof. We start by proving items 1 to 3 for the matrix $\tilde{\mathbf{T}}$. Since it is shown in Lemma 2.4 that the mappings $z \mapsto \overline{\delta}_m(z)$ are Stieltjes transform carried by $\mathbb{R}_+$, by [35, Proposition 5-1, item 4] $z \mapsto (-z(1+\overline{\delta}_m))^{-1}$ for any $m \geq 1$ is also a Stieltjes transform carried by $\mathbb{R}_+$. Moreover, by Lemma 2.2, it remains to prove that $\lim_{y\to+\infty} iy\tilde{T}_{mm}(iy) = -1$. We write

$$iy\tilde{T}_{mm}(iy) = iy\frac{-1}{iy(1+\overline{\delta}_m(iy))} = \frac{-1}{1+\frac{iy\overline{\delta}_m(iy)}{iy}}$$

and since $\overline{\delta}_m$ are Stieltjes transforms of measures with total mass $\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m < +\infty$, we have that $\lim_{y\to+\infty} iy\overline{\delta}_m(iy)$ is finite, which proves that $\lim_{y\to+\infty} iy\tilde{T}_{mm}(iy) = -1$ for each m. This implies that $\lim_{y\to+\infty} iy\tilde{\mathbf{T}}(iy) = -\mathbf{I}_M$, and ends the proof of items 1 to 3 for $\tilde{\mathbf{T}}$.

We now prove item 1 for the matrix $\overline{\mathbf{T}}$. Since we proved before that $z \mapsto \tilde{T}_{mm}(z)$ is a Stieltjes transform, we only have to show that $\overline{\mathbf{T}}(z)^{-1}$ does not vanishes on $\mathbb{C}_+$. This is immediate since for any $z \in \mathbb{C}_+$, using the fact that $\tilde{T}_{mm}$ is a Stieltjes transform,

$$\mathrm{Im}\overline{\mathbf{T}}(z)^{-1} = -\mathrm{Im}z\mathbf{I}_{B+1} - \frac{c}{M}\mathrm{Im}(z\tilde{T}_{mm}(z))\mathbf{\Theta}_m < 0.$$

Concerning item 2, we have

$$\mathrm{Im}\overline{\mathbf{T}} = (\mathrm{Im}z)\overline{\mathbf{T}}\overline{\mathbf{T}}^* + \frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk})\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}^*$$

which is positive. Similarly,

$$\mathrm{Im}z\overline{\mathbf{T}} = |z|^2\overline{\mathbf{T}}\overline{\mathbf{T}}^*\left(\frac{c}{M}\sum_{m=1}^{M}\mathbf{\Theta}_m\mathrm{Im}\tilde{T}_{mm}\right)$$

which is also positive. Lastly,

$$\lim_{y\to+\infty} iy\overline{\mathbf{T}}(iy) = \lim_{y\to+\infty} iy\left(-iy\mathbf{I}_{B+1} - \frac{c}{M}\sum_{m=1}^{M}\tilde{T}_{mm}(iy)\mathbf{\Theta}_m\right)^{-1} = -\mathbf{I}_{B+1}$$

since $|\tilde{T}_{mm}(iy)| \xrightarrow{y\to+\infty} 0$ and $\sup_{N\geq 1}\sup_{m\in[M]}\|\mathbf{\Theta}_m\| < +\infty$. Therefore, it is clear that $\sup_{y>\epsilon}\|iy(\overline{\mathbf{T}}(iy))\| < +\infty$. By Lemma 2.3, there exists a matrix valued measure $\mu^{\overline{\mathbf{T}}(\nu)}$ carried over $\mathbb{R}_+$ such that

$$\overline{\mathbf{T}}(z) = \int\frac{\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda)}{\lambda - z}$$

with $\mu^{\overline{\mathbf{T}}(\nu)}(\mathbb{R}_+) = -\lim_{y\to+\infty} iy\overline{\mathbf{T}}(iy) = \mathbf{I}_{B+1}$ as announced. To prove item 3, we shall prove the two following inequalities

$$\forall z \in \mathbb{C}_+, \frac{\overline{\mathbf{T}} - \overline{\mathbf{T}}^*}{2i} \leq \frac{\mathbf{I}_M}{\mathrm{Im}z} \quad \text{and} \quad \frac{\overline{\mathbf{T}} - \overline{\mathbf{T}}^*}{2i} \geq \mathrm{Im}(z)\overline{\mathbf{T}}\overline{\mathbf{T}}^*.$$

The first one is shown by writing:

$$\frac{\overline{\mathbf{T}}-\overline{\mathbf{T}}^*}{2i} = \mathrm{Im}\overline{\mathbf{T}} = \mathrm{Im}\int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}(\nu)}(\mathrm{d}\lambda)}{\lambda - z} = \mathrm{Im}z \int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}(\nu)}(\mathrm{d}\lambda)}{|\lambda - z|^2} \leq \frac{\mathrm{Im}z}{\mathrm{Im}^2 z}\int_{\mathbb{R}} \mathrm{d}\mu^{\overline{\mathbf{T}}(\nu)} = \frac{\mathbf{I}_{B+1}}{\mathrm{Im}z}.$$

Now,

$$\frac{\overline{\mathbf{T}}-\overline{\mathbf{T}}^*}{2i} = (\mathrm{Im}z)\overline{\mathbf{T}}\overline{\mathbf{T}}^* + \frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk})\overline{\mathbf{T}}\boldsymbol{\Theta}_k\overline{\mathbf{T}}^* \geq (\mathrm{Im}z)\overline{\mathbf{T}}\overline{\mathbf{T}}^*$$

since $\frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk})\overline{\mathbf{T}}\boldsymbol{\Theta}_k\overline{\mathbf{T}}^* \geq 0$.

We now prove item 4 for $\mu^{\overline{\mathbf{T}}}$. By Markov's inequality, it is sufficient to check that

$$\sup_{N\geq 1}\int_{\mathbb{R}} \lambda\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda) < +\infty.$$

By [35, Lemma C.1],

$$\int_{\mathbb{R}} \lambda\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda) = \lim_{y\to+\infty}\mathrm{Re}\left\{-iy\left(iy\frac{1}{M}\mathrm{tr}\ \overline{\mathbf{T}}(iy)+1\right)\right\}. \tag{2.7}$$

By the definition of $\overline{\mathbf{T}}$, the following equality holds

$$\overline{\mathbf{T}}(iy)\left(-iy\mathbf{I}_{B+1} + \frac{c}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\Theta}_m}{1+\overline{\delta}_m(iy)}\right) = \mathbf{I}_{B+1}$$

which by taking the normalized trace implies that

$$1 + \frac{1}{B+1}\mathrm{tr}\ iy\overline{\mathbf{T}}(iy) = \frac{c}{M}\sum_{m=1}^{M}\frac{\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(iy)\boldsymbol{\Theta}_m}{1+\overline{\delta}_m(iy)} = \frac{c}{M}\sum_{m=1}^{M}\frac{\overline{\delta}_m(iy)}{1+\overline{\delta}_m(iy)}. \tag{2.8}$$

Plugging (2.8) into (2.7), we get:

$$\int_{\mathbb{R}} \lambda\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda) = \lim_{y\to+\infty}\mathrm{Re}\left\{-iy\frac{c}{M}\sum_{m=1}^{M}\frac{\overline{\delta}_m(iy)}{1+\overline{\delta}_m(iy)}\right\} = \lim_{y\to+\infty}\frac{c}{M}\sum_{m=1}^{M}\frac{\mathrm{Re}\left\{-iy\overline{\delta}_m(iy)\right\}}{|1+\overline{\delta}_m(iy)|^2}.$$

Since it has been recalled in Lemma 2.4 that $\overline{\delta}_m$ are Stieltjes transforms of finite measures, it holds that

$$\sup_{N\geq 1}\sup_{m\in[M]}\lim_{y\to+\infty}|\overline{\delta}_m(iy)| = 0$$

and that for each m, $\lim_{y\to+\infty} -iy\overline{\delta}_m(iy) = \frac{1}{B+1}\mathrm{tr}\ \boldsymbol{\Theta}_m < +\infty$ uniformly in $m \in [M]$ and $N \geq 1$. This ends the proof of item 4 for $\mu^{\overline{\mathbf{T}}}$. The proof for $\mu^{\tilde{\mathbf{T}}}$ is similar. □

We can now state the following important result, which states that $\mathbb{E}\tilde{\mathbf{Q}}(\nu,z)$ behaves asymptotically as $\tilde{\mathbf{T}}(\nu,z)$.

Proposition 2.1. *Let $\tilde{\mathbf{T}}$ defined by (2.6). Under Assumption 1.1 to Assumption 1.4, for each $z \in \mathbb{C}_+$,*

$$\sup_{\nu\in[0,1]}\left|\frac{1}{M}\mathrm{tr}\ \mathbb{E}\tilde{\mathbf{Q}}(\nu,z) - \frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}(\nu,z)\right| = \mathcal{O}_z\left(\frac{1}{B^2}\right)$$

$$\sup_{\nu\in[0,1]}\left|\frac{1}{B+1}\mathrm{tr}\ \mathbb{E}\overline{\mathbf{Q}}(\nu,z) - \frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(\nu,z)\right| = \mathcal{O}_z\left(\frac{1}{B^2}\right).$$

Proof. Using Gaussian tools (Gaussian integration by part and Nash-Poincaré inequality), Proposition 3 from [48] proves that for each ν,

$$\frac{1}{B+1}\mathrm{tr}\ \mathbb{E}\overline{\mathbf{Q}}(\nu,z) - \frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(\nu,z) = \mathcal{O}_z\left(\frac{1}{B^2}\right)$$

and it is easy to check that the bound is uniform in $\nu \in [0,1]$, since it has been proved in Chapter 1 that

$$\sup_{N\geq 1}\sup_{\nu\in[0,1]}\sup_{m\in[M]}\|\boldsymbol{\Theta}_{m,N}(\nu)\| < +\infty.$$

Denote $\mathbf{X} := \mathrm{dg}(\mathbf{S}(\nu))^{-\frac{1}{2}}\boldsymbol{\Sigma}(\nu)$. Since $\mathbf{X}\overline{\mathbf{Q}}\mathbf{X}^* = z\tilde{\mathbf{Q}} + \mathbf{I}_M$, to obtain the result on $\mathbb{E}\tilde{\mathbf{Q}}$, it is sufficient to study $\mathbb{E}\mathbf{X}\overline{\mathbf{Q}}\mathbf{X}^*$ using the same Gaussian tools. We omit this part. $\square$

As a direct consequence of the Helffer-Sjöstrand formula (the proof is very similar to the proof of 1.3), the following corollary holds.

Corollary 2.1. *Under Assumption 1.1, to Assumption 1.5,*

$$\sup_{\nu\in[0,1]}\left|\mathbb{E}\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)}\right| = \mathcal{O}\left(\frac{1}{B^2}\right).$$

We now use the fact that for $\nu \neq \nu'$ such that $\nu - \nu' = \frac{k}{N}$ for some integer k and each m, the Fourier transforms of $y_{m,1},\dots,y_{m,N}$ at these frequencies have a covariance close to zero, which leads to the fact that the family of matrices $\boldsymbol{\Theta}_m(\nu)$ for $m \in [M]$ and $\nu \in [0,1]$ are all close to the identity matrix. We already proved in Chapter 1 that $\boldsymbol{\Theta}_{m,N}(\nu)$ satisfies

$$\boldsymbol{\Theta}_{m,N}(\nu) = \mathbf{I}_{B+1} + \boldsymbol{\Phi}_{m,N}(\nu), \qquad \sup_{m\in[M]}\sup_{\nu\in[0,1]}\|\boldsymbol{\Phi}_{m,N}(\nu)\| = \mathcal{O}\left(\frac{B}{N}\right).$$

and that the following lemma holds:

Lemma 2.6. *Under Assumption 1.2 to Assumption 1.4, the covariance matrices $\boldsymbol{\Theta}_m(\nu)$ satisfies the following estimates:*

$$\sup_{m\in[M]}\sup_{\nu\in[0,1]}\|\boldsymbol{\Theta}_{m,N}(\nu) - \mathbf{I}_{B+1}\| = \mathcal{O}\left(\frac{B}{N}\right) \tag{2.9}$$

$$\sup_{m\in[M]}\sup_{\nu\in[0,1]}\left|\frac{1}{B+1}\mathrm{tr}\ (\boldsymbol{\Theta}_{m,N}(\nu) - \mathbf{I}_{B+1}) - \Upsilon_{m,N}(\nu)\right| = \mathcal{O}\left(\frac{B}{N}\right)^3 \tag{2.10}$$

where $\Upsilon_{m,N}(\nu)$ is the $\mathcal{O}(\frac{B}{N})^2$ term equal to

$$\Upsilon_{m,N}(\nu) = \frac{1}{2}\frac{s''_m(\nu)}{s_m(\nu)}\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\frac{b^2}{N^2}.$$

Informally, the matrices $\boldsymbol{\Theta}_{m,N}(\nu)$ are uniformly close to the identity. Since we know that in the white noise case ($\boldsymbol{\Theta}_{m,N}(\nu) = \mathbf{I}_{B+1}$ for each $N \geq 1$, $m \geq 1$ and each $\nu \in [0,1]$), the limiting spectral distribution is the Marcenko-Pastur distribution, it is reasonable to expect that the limiting spectral distribution of $\tilde{\mathbf{T}}$ is close to the Marcenko-Pastur distribution. We now state the second important result of this chapter, which states that the deterministic matrices $\tilde{\mathbf{T}}(\nu,z)$ are close to $\tilde{t}_N(z)\mathbf{I}_M$.

Proposition 2.2. *Recall that we defined*

$$\tilde{r}_N(\nu) = \left(\frac{1}{2M} \sum_{m=1}^{M} \frac{s''_m(\nu)}{s_m(\nu)} \right), \quad r_N(\nu) = \left(\frac{1}{M} \sum_{m=1}^{M} \frac{s'_m(\nu)}{s_m(\nu)} \right)^2 \tag{2.11}$$

and

$$\tilde{p}_N(z) = (z\tilde{t}_N(z))', \quad p_N(z) = \frac{-c_N(z\tilde{t}_N(z)\bar{t}_N(z))^3}{1 - c_N(z\tilde{t}_N(z)\bar{t}_N(z))^2}. \tag{2.12}$$

Under Assumption 1.4, for each $z \in \mathbb{C}_+$, the matrix $\tilde{\mathbf{T}}$ is close to the Stieltjes transform of the Marcenko-Pastur distribution in the following way:

$$\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ \tilde{\mathbf{T}}(\nu, z) - \tilde{t}_N(z) + \tilde{p}_N(z)\tilde{r}_N(\nu)v_N - p_N(z)r_N(\nu)v_N \right| = \mathcal{O}_z \left(\frac{B}{N} \right)^3. \tag{2.13}$$

Remark 2.2. *One could also approximate $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}(\nu, z)$ by $\tilde{t}_N(z)$, but the estimation error rate would be worse, of order $\mathcal{O}_z(\frac{B}{N})^2$ instead of $\mathcal{O}_z(\frac{B}{N})^3$.*

The proof is deferred to Section 2.3. As for Proposition 2.1, the Helffer-Sjöstrand formula implies directly the following corollary.

Corollary 2.2. *Under Assumption 1.1 to Assumption 1.5,*

$$\sup_{\nu \in [0,1]} \left| \int_{\mathbb{R}} f \, \mathrm{d}\mu^{\tilde{\mathbf{T}}(\nu)} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{mp,c_N} + < \tilde{\mathcal{D}}_N, f > \tilde{r}_N(\nu)v_N - < \mathcal{D}_N, f > r_N(\nu)v_N \right| = \mathcal{O} \left(\frac{B}{N} \right)^3.$$

Proof. Immediate following the same steps as in the proof of Corollary 2.1. □

Combining Corollary 2.1 and Corollary 2.2, we obtain the following final result.

Theorem 2.1. *Under Assumption 1.1 to Assumption 1.5,*

$$\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathbb{E}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{mp,c_N} + < \tilde{\mathcal{D}}_N, f > \tilde{r}_N(\nu)v_N - < \mathcal{D}_N, f > r_N(\nu)v_N \right| = \mathcal{O} \left(\frac{B}{N} \right)^3.$$

This allows to recover exactly the statement of Theorem 1.2 from Chapter 1.

2.3 Proof of Proposition 2.2

Proof. By a succession of approximation of $\tilde{\mathbf{T}}$ and $\overline{\mathbf{T}}$ by $\tilde{t}_N \mathbf{I}_M$ and $\bar{t}_N \mathbf{I}_{B+1}$, we show that the following system of equations between $\tilde{\epsilon} := \frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}} - \tilde{t}_N$ and $\bar{\epsilon} := \frac{1}{B+1}\overline{\mathbf{T}} - \bar{t}_N$ holds:

$$
\begin{aligned}
\begin{pmatrix} \tilde{\epsilon}_N \\ \bar{\epsilon}_N \end{pmatrix} &= \begin{bmatrix} 0 & z\tilde{t}_N^2 \\ c_N z \bar{t}_N^2 & 0 \end{bmatrix} \begin{pmatrix} \tilde{\epsilon}_N \\ \bar{\epsilon}_N \end{pmatrix} \\
&+ \begin{pmatrix} z\tilde{t}_N^2 \bar{t}_N \frac{1}{M} \sum_{m=1}^M \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N z^2 \tilde{t}_N^3 \bar{t}_N^2 \frac{1}{B+1}\mathrm{tr}\ \left\{ \left(\frac{1}{M}\sum_{m=1}^M \mathbf{\Phi}_m \right)^2 \right\} \\ c_N z \tilde{t}_N \bar{t}_N^2 \frac{1}{M} \sum_{m=1}^M \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N^2 z^2 \tilde{t}_N^2 \bar{t}_N^3 \frac{1}{B+1}\mathrm{tr}\ \left\{ \left(\frac{1}{M}\sum_{m=1}^M \mathbf{\Phi}_m \right)^2 \right\} \end{pmatrix} \\
&+ \begin{pmatrix} z\tilde{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N \\ c_N z \bar{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N \end{pmatrix} \\
&+ \begin{pmatrix} \mathcal{O}_z\left(\frac{B}{N}\right)^3 \\ \mathcal{O}_z\left(\frac{B}{N}\right)^3 \end{pmatrix}.
\end{aligned}
\tag{2.14}
$$

Before proving the validity of this equation, we show that (2.14) implies that $\tilde{\epsilon}_N$ and $\bar{\epsilon}_N$ are both $\mathcal{O}_z(\frac{B}{N})^2$. the first component of (2.14) can be written

$$
\tilde{\epsilon}_N = z\tilde{t}_N^2 \bar{\epsilon}_N + z\tilde{t}_N^2 \bar{t}_N \frac{1}{M} \sum_{m=1}^M \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N z^2 \tilde{t}_N^3 \bar{t}_N^2 \frac{1}{B+1}\mathrm{tr}\ \left\{ \left(\frac{1}{M}\sum_{m=1}^M \mathbf{\Phi}_m \right)^2 \right\} + z\tilde{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N + \mathcal{O}_z\left(\frac{B}{N}\right)^3
$$

while the second component provides

$$
\bar{\epsilon}_N = c_N z \bar{t}_N^2 \tilde{\epsilon}_N + c_N z \tilde{t}_N \bar{t}_N^2 \frac{1}{M} \sum_{m=1}^M \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N^2 z^2 \tilde{t}_N^2 \bar{t}_N^3 \frac{1}{B+1}\mathrm{tr}\ \left\{ \left(\frac{1}{M}\sum_{m=1}^M \mathbf{\Phi}_m \right)^2 \right\} + c_N z \bar{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N + \mathcal{O}_z\left(\frac{B}{N}\right)^3.
$$

Combining both results, we can write:

$$
\tilde{\epsilon}_N (1 - c_N (z\tilde{t}_N \bar{t}_N)^2) = z\tilde{t}_N^2 \left(c_N z \tilde{t}_N \bar{t}_N^2 \frac{1}{M} \sum_{m=1}^M \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N^2 z^2 \tilde{t}_N^2 \bar{t}_N^3 \frac{1}{B+1}\mathrm{tr}\ \left\{ \left(\frac{1}{M}\sum_{m=1}^M \mathbf{\Phi}_m \right)^2 \right\} \right) + z\tilde{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N + \mathcal{O}_z\left(\frac{B}{N}\right)^3. \tag{2.15}
$$

The following Lemma 2.7 proves that $\tilde{\epsilon}_N$ and $\bar{\epsilon}_N$ are $\mathcal{O}_z(\frac{B}{N})$. Its proof is deferred to Appendix 2.A.

Lemma 2.7.

$$
\|\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M\| = \mathcal{O}_z\left(\frac{B}{N}\right) \tag{2.16}
$$

$$
\|\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}\| = \mathcal{O}_z\left(\frac{B}{N}\right). \tag{2.17}
$$

Lemma 2.7 implies that ϵ_N and $\tilde{\epsilon}_N$ are $\mathcal{O}_z(\frac{B}{N})$ so consequently $\bar{\epsilon}_N \tilde{\epsilon}_N = \mathcal{O}_z(\frac{B}{N})^2$. We recall that by Lemma 2.6, $\frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m=\mathcal{O}_z(\frac{B}{N})^2$ and $\frac{1}{B+1}\mathrm{tr}\ \left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2=\mathcal{O}_z(\frac{B}{N})^2$. Lastly, by Lemma 1.1 from [59], $1-c_N|z\tilde{t}_N\bar{t}_N|^2>\frac{\mathrm{Im}^4 z}{2(|z|^2+\lambda_-^2)}$, where $\lambda_-=(1-\sqrt{c_N})^2$. We conclude by (2.15) that $\tilde{\epsilon}_N=\mathcal{O}_z(\frac{B}{N})^2$. Showing that $\bar{\epsilon}_N=\mathcal{O}_z(\frac{B}{N})^2$ is proved similarly. Therefore, $\bar{\epsilon}_N\tilde{\epsilon}_N=\mathcal{O}_z(\frac{B}{N})^3$, and (2.14) becomes:

$$\begin{bmatrix}1 & -z\tilde{t}_N^2\\ -c_N z\bar{t}_N^2 & 1\end{bmatrix}\begin{pmatrix}\tilde{\epsilon}_N\\ \bar{\epsilon}_N\end{pmatrix}=\begin{pmatrix} z\tilde{t}_N^2\bar{t}_N\frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m+c_N z^2\tilde{t}_N^3\bar{t}_N^2\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2\right\}\\ c_N z\tilde{t}_N\bar{t}_N^2\frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m+c_N^2 z^2\tilde{t}_N^2\bar{t}_N^3\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2\right\}\end{pmatrix}$$
$$+\begin{pmatrix}\mathcal{O}_z\left(\frac{B}{N}\right)^3\\ \mathcal{O}_z\left(\frac{B}{N}\right)^3\end{pmatrix}.$$

By Lemma 1.1 from [59], the inverse of the matrix $\begin{bmatrix}1 & -z\tilde{t}_N^2\\ -c_N z\bar{t}_N^2 & 1\end{bmatrix}$ exists and is given by:

$$\frac{1}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}\begin{bmatrix}1 & z\tilde{t}_N^2\\ c_N z\bar{t}_N^2 & 1\end{bmatrix}$$

so $(\tilde{\epsilon}_N,\bar{\epsilon}_N)$ can be expressed as:

$$\begin{pmatrix}\tilde{\epsilon}_N\\ \bar{\epsilon}_N\end{pmatrix}=\frac{1}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}\times$$
$$\begin{bmatrix}1 & z\tilde{t}_N^2\\ c_N z\bar{t}_N^2 & 1\end{bmatrix}\begin{pmatrix}\frac{z\tilde{t}_N^2\bar{t}_N}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m+\frac{c_N z^2\tilde{t}_N^3\bar{t}_N^2}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2\right\}\\ \frac{c_N z\tilde{t}_N\bar{t}_N^2}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m+\frac{c_N^2 z^2\tilde{t}_N^2\bar{t}_N^3}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2\right\}\end{pmatrix}+\begin{pmatrix}\mathcal{O}_z\left(\frac{B}{N}\right)^3\\ \mathcal{O}_z\left(\frac{B}{N}\right)^3\end{pmatrix}.$$

This provides:

$$\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}-\tilde{t}_N=\frac{z\tilde{t}_N^2\bar{t}_N+c_N z^2\tilde{t}_N^3\bar{t}_N^2}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}\frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m$$
$$+\frac{c_N z^2\tilde{t}_N^3\bar{t}_N^2+c^2z^3\tilde{t}_N^4\bar{t}_N^3}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2\right\}+\mathcal{O}_z\left(\frac{B}{N}\right)^3.$$

Note that it is easily verified that:

$$\frac{z\tilde{t}_N^2\bar{t}_N+c_N z^2\tilde{t}_N^3\bar{t}_N^2}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}=(z\tilde{t}_N(z))'=\tilde{p}_N(z)$$
$$\frac{c_N z^2\tilde{t}_N^3\bar{t}_N^2+c^2z^3\tilde{t}_N^4\bar{t}_N^3}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}=\frac{-c_N(z\tilde{t}_N\bar{t}_N)^3}{1-c_N(z\tilde{t}_N\bar{t}_N)^2}=p_N(z).$$

Finally, it remains to show that

$$\frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \boldsymbol{\Phi}_m = \tilde{r}_N(\nu)v_N + \mathcal{O}\left(\frac{B}{N}\right)^3 \tag{2.18}$$

$$\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\Phi}_m\right)^2\right\} = r_N(\nu)v_N + \mathcal{O}\left(\frac{B}{N}\right)^3. \tag{2.19}$$

Recall that

$$\boldsymbol{\Phi}_m(\nu) = \mathrm{dg}\left(\frac{b}{N}\frac{s'_m(\nu)}{s_m(\nu)} + \frac{b^2}{N^2}\frac{s''_m(\nu)}{2s_m(\nu)}, b=-\frac{B}{2},\ldots,\frac{B}{2}\right) + \mathcal{O}\left(\frac{B}{N}\right)^3\mathbf{I}_{B+1} + \mathcal{O}\left(\frac{1}{N}\right)\mathbb{1}\mathbb{1}^T$$

so (2.18) is clear. Concerning (2.19), it can be checked as follows:

$$\begin{aligned}\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\Phi}_m\right)^2\right\} &= \frac{1}{B+1}\mathrm{tr}\ \left\{\frac{1}{M^2}\sum_{m_1,m_2=1}^{M}\boldsymbol{\Phi}_{m_1}\boldsymbol{\Phi}_{m_2}\right\} \\ &= \frac{1}{B+1}\frac{1}{M^2}\sum_{m_1,m_2=1}^{M}\mathrm{tr}\ \{\boldsymbol{\Phi}_{m_1}\boldsymbol{\Phi}_{m_2}\}\end{aligned}$$

so we now evaluate tr $\{\boldsymbol{\Phi}_{m_1}\boldsymbol{\Phi}_{m_2}\}$:

$$\begin{aligned}\boldsymbol{\Phi}_{m_1}(\nu)\boldsymbol{\Phi}_{m_2}(\nu) = \left(\frac{s'_{m_1}(\nu)}{s_{m_1}(\nu)}\frac{s'_{m_2}(\nu)}{s_{m_2}(\nu)}\right)\mathrm{dg}\left(\left(\frac{b}{N}\right)^2, b=-\frac{B}{2},\ldots,\frac{B}{2}\right) \\ + \mathcal{O}\left(\frac{B}{N}\right)^3\mathbf{I}_{B+1} + \mathcal{O}\left(\frac{1}{N}\right)\mathbb{1}\mathbb{1}^T\end{aligned}$$

which means that

$$\frac{1}{B+1}\mathrm{tr}\ (\boldsymbol{\Phi}_{m_1}(\nu)\boldsymbol{\Phi}_{m_2}(\nu)) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\frac{s'_{m_1}(\nu)}{s_{m_1}(\nu)}\frac{s'_{m_2}(\nu)}{s_{m_2}(\nu)}\right)\left(\frac{b}{N}\right)^2 + \mathcal{O}\left(\frac{B}{N}\right)^3.$$

Finally, summing over m_1 and m_2 provides:

$$\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\Phi}_m\right)^2\right\} = v_N\left(\frac{1}{M^2}\sum_{m_1,m_2=1}^{M}\left(\frac{s'_{m_1}(\nu)}{s_{m_1}(\nu)}\frac{s'_{m_2}(\nu)}{s_{m_2}(\nu)}\right)\right) + \mathcal{O}\left(\frac{B}{N}\right)^3$$

which can be written in the form

$$\frac{1}{B+1}\mathrm{tr}\ \left\{\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\Phi}_m\right)^2\right\} = v_N\left(\frac{1}{M}\sum_{m=1}^{M}\frac{s'_m(\nu)}{s_m(\nu)}\right)^2 + \mathcal{O}\left(\frac{B}{N}\right)^3.$$

Modulo equation (2.14), Proposition 2.2 is proved.

Proof of (2.14). We start by expressing $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}(\nu,z) - \tilde{t}_N(z)$ in terms of $\frac{1}{M}\mathrm{tr}\ \overline{\mathbf{T}}(\nu,z) - \bar{t}_N(z)$. Recall that

$$\tilde{t}_N(z) = \frac{-1}{z(1+\bar{t}_N(z))}, \quad \bar{t}_N(z) = \frac{-1}{z(1+c_N\tilde{t}_N(z))}$$

so we can write

$$\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M = z\tilde{t}_N \tilde{\mathbf{T}} \mathrm{dg}\left(\overline{\delta}_m - \overline{t}_N, m \in [M]\right)$$

where $\overline{\delta}_m$ can be decomposed as the following sum of two terms:

$$\overline{\delta}_m = \frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Theta}_m \overline{\mathbf{T}}) = \frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}} + \frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m \overline{\mathbf{T}})$$

so that

$$\begin{aligned}\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M &= z\tilde{t}_N \tilde{\mathbf{T}} \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}} + \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m \overline{\mathbf{T}} - \overline{t}_N, m \in [M]\right)\\ &= z\tilde{t}_N \overline{\epsilon}_N \tilde{\mathbf{T}} + z\tilde{t}_N \tilde{\mathbf{T}} \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m \overline{\mathbf{T}}), m \in [M]\right)\end{aligned}$$

Take the normalized trace, and we obtain the following expression for $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}} - \tilde{t}_N$:

$$\tilde{\epsilon}_N = z\tilde{t}_N \overline{\epsilon}_N \frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}} + z\tilde{t}_N \frac{1}{M}\mathrm{tr}\ \left(\tilde{\mathbf{T}} \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m \overline{\mathbf{T}}, m \in [M]\right)\right). \tag{2.20}$$

The first term of the rhs of (2.20) is clearly equal to

$$z\tilde{t}_N \overline{\epsilon}_N \left(\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}} - \tilde{t}_N + \tilde{t}_N\right) = z\tilde{t}_N^2 \overline{\epsilon}_N + z\tilde{t}_N \overline{\epsilon}_N \tilde{\epsilon}_N. \tag{2.21}$$

Concerning the second term of the rhs of (2.20), we write $\tilde{\mathbf{T}} = \tilde{t}_N \mathbf{I}_M + (\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M)$ and $\overline{\mathbf{T}} = \overline{t}_N \mathbf{I}_{B+1} + (\overline{\mathbf{T}} - \overline{t}_N \mathbf{I}_{B+1})$, so that:

$$\begin{aligned}&\frac{1}{M}\mathrm{tr}\ \left(\tilde{\mathbf{T}} \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m \overline{\mathbf{T}}), m \in [M]\right)\right)\\ &\quad= \frac{1}{M}\mathrm{tr}\ \left(\tilde{t}_N \mathbf{I}_M \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m \overline{t}_N \mathbf{I}_{B+1}), m \in [M]\right)\right)\\ &\quad+ \frac{1}{M}\mathrm{tr}\ \left((\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M) \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m \overline{t}_N \mathbf{I}_{B+1}), m \in [M]\right)\right)\\ &\quad+ \frac{1}{M}\mathrm{tr}\ \left(\tilde{t}_N \mathbf{I}_M \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m (\overline{\mathbf{T}} - \overline{t}_N \mathbf{I}_{B+1})), m \in [M]\right)\right)\\ &\quad+ \frac{1}{M}\mathrm{tr}\ \left((\tilde{\mathbf{T}} - \tilde{t}_N \mathbf{I}_M) \mathrm{dg}\left(\frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m (\overline{\mathbf{T}} - \overline{t}_N \mathbf{I}_{B+1}), m \in [M]\right)\right).\end{aligned}$$

By Lemma 2.6 and Lemma 2.7, it is clear that the second and fourth terms are of order $\mathcal{O}_z(\frac{B}{N})^3$. The first one is equal to

$$\tilde{t}_N \overline{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m \tag{2.22}$$

while the third one is equal to:

$$\tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\ (\mathbf{\Phi}_m (\overline{\mathbf{T}} - \overline{t}_N \mathbf{I}_{B+1})). \tag{2.23}$$

It remains to extract the dominant quantity from this term. Since $\sup_{m\in[M]} \|\mathbf{\Phi}_m\| = \mathcal{O}_z(\frac{B}{N})$, it is sufficient to expand $\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}$ up to terms of spectral norm $\mathcal{O}_z(\frac{B}{N})^2$. Developing $\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}$:

$$
\begin{aligned}
\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1} &= \overline{\mathbf{T}} \left(-z(1 + c_N \tilde{t}_N)\mathbf{I}_{B+1} + z\left(\mathbf{I}_{B+1} + \frac{c_N}{M}\sum_{m=1}^{M} \tilde{T}_{mm}\mathbf{\Theta}_m\right)\right) \bar{t}_N \mathbf{I}_{B+1} \\
&= c_N z \bar{t}_N \overline{\mathbf{T}} \frac{1}{M}\sum_{m=1}^{M} \left(\tilde{T}_{mm}\mathbf{\Theta}_m - \tilde{t}_N \mathbf{I}_{B+1}\right) \\
&= T_1 + T_2
\end{aligned} \tag{2.24}
$$

where

$$
\begin{aligned}
T_1 &= c_N z \bar{t}_N^2 \frac{1}{M}\sum_{m=1}^{M} \left((\tilde{T}_{mm} - \tilde{t}_N + \tilde{t}_N)(\mathbf{\Phi}_m + \mathbf{I}_{B+1}) - \tilde{t}_N \mathbf{I}_{B+1}\right) \\
T_2 &= c_N z \bar{t}_N (\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}) \frac{1}{M}\sum_{m=1}^{M} \left((\tilde{T}_{mm} - \tilde{t}_N + \tilde{t}_N)(\mathbf{\Phi}_m + \mathbf{I}_{B+1}) - \tilde{t}_N \mathbf{I}_{B+1}\right).
\end{aligned}
$$

Concerning T_1, it is clear that

$$
T_1 = c_N z \bar{t}_N^2 \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Phi}_m + c_N z \bar{t}_N^2 \tilde{\epsilon}_N \mathbf{I}_{B+1} + c_N z \bar{t}_N^2 \frac{1}{M}\sum_{m=1}^{M} ((\tilde{T}_{mm} - \tilde{t}_N)\mathbf{\Phi}_m) \tag{2.25}
$$

while for T_2,

$$
T_2 = c_N z \bar{t}_N (\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1})\tilde{\epsilon}_N + c_N z \tilde{t}_N \bar{t}_N (\overline{\mathbf{T}} - \bar{t}\mathbf{I}_{B+1}) \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Phi}_m + \mathcal{O}_z\left(\frac{B}{N}\right)^3 \tag{2.26}
$$

where the $\mathcal{O}_z(\cdot)$ means that the spectral norm of $\|T_2 - c_N z \bar{t}_N (\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1})\tilde{\epsilon}_N - c_N z \tilde{t}_N^2 (\overline{\mathbf{T}} - \bar{t}\mathbf{I}_{B+1}) \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Phi}_m\| = \mathcal{O}_z(\frac{B}{N})^3$. Combining (2.25) and (2.26) into (2.24) we get:

$$
\begin{aligned}
\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1} = {} & c_N z \bar{t}_N^2 \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Phi}_m + c_N z \bar{t}_N^2 \tilde{\epsilon}_N \mathbf{I}_{B+1} + c_N z \bar{t}_N^2 \frac{1}{M}\sum_{m=1}^{M} ((\tilde{T}_{mm} - \tilde{t}_N)\mathbf{\Phi}_m) \\
& + c_N z \bar{t}_N (\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1})\tilde{\epsilon}_N + c_N z \tilde{t}_N \bar{t}_N (\overline{\mathbf{T}} - \bar{t}\mathbf{I}_{B+1}) \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Phi}_m + \mathcal{O}_z\left(\frac{B}{N}\right)^3.
\end{aligned} \tag{2.27}
$$

Inserting (2.27) in (2.23), we get the following expression for the dominant quantity:

$$\begin{aligned}
\tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m(\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1})\right) &= \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m\left(c_N z \bar{t}_N^2 \tilde{t}_N \frac{1}{M}\sum_{m'=1}^{M}\mathbf{\Phi}_{m'}\right)\right) \\
&+ \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m\left(c_N z \bar{t}_N^2 \tilde{\epsilon}_N \mathbf{I}_{B+1}\right)\right) \\
&+ \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m\left(c_N z \bar{t}_N^2 \frac{1}{M}\sum_{m'=1}^{M}((T_{m'm'} - \tilde{t}_N)\mathbf{\Phi}_{m'})\right)\right) \\
&+ \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m\left(c_N z \bar{t}_N (\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1})\tilde{\epsilon}_N\right)\right) \\
&+ \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m\left(c_N z \tilde{t}_N \bar{t}_N (\overline{\mathbf{T}} - \bar{t}\mathbf{I}_{B+1})\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)\right) \\
&+ \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M} \frac{1}{B+1}\mathrm{tr}\left(\mathbf{\Phi}_m \mathcal{O}_z\left(\frac{B}{N}\right)^3\right) \\
&= c_N z \tilde{t}_N^2 \bar{t}_N^2 \frac{1}{B+1}\mathrm{tr}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2 + \mathcal{O}_z\left(\frac{B}{N}\right)^3 \qquad (2.28)
\end{aligned}$$

where only the first term is $\mathcal{O}_z\left(\frac{B}{N}\right)^2$. Combining (2.22) and (2.28), the second term of the rhs of (2.20) is equal to

$$z\tilde{t}_N^2 \bar{t}_N \frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N z^2 \tilde{t}_N^3 \bar{t}_N^2 \frac{1}{B+1}\mathrm{tr}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2 + \mathcal{O}_z\left(\frac{B}{N}\right)^3. \qquad (2.29)$$

Combining the estimates (2.21) and (2.29) for the first and second term of the rhs of (2.20) we finally get:

$$\begin{aligned}
\tilde{\epsilon}_N = z\tilde{t}_N^2 \bar{\epsilon}_N + z\tilde{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N + z\tilde{t}_N^2 \bar{t}_N \frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m \\
+ c_N z^2 \tilde{t}_N^3 \bar{t}_N^2 \frac{1}{B+1}\mathrm{tr}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right)^2 + \mathcal{O}_z\left(\frac{B}{N}\right)^3 \qquad (2.30)
\end{aligned}$$

which is exactly the first component in the equation (2.14).

Expansion of $\frac{1}{M}\mathrm{tr}\ \overline{\mathbf{T}}(\nu, z) - \bar{t}_N(z)$. We start with $\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}$. This quantity has already been computed up to order $\mathcal{O}_z(\frac{B}{N})^3$ in spectral norm in (2.27), so taking the normalized trace we obtain directly

$$\begin{aligned}
\bar{\epsilon}_N = c_N z \bar{t}_N^2 \tilde{t}_N \frac{1}{M}\sum_{m=1}^{M}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m + c_N z \bar{t}_N^2 \tilde{\epsilon}_N + c_N z \bar{t}_N^2 \frac{1}{M}\sum_{m=1}^{M}\left((\tilde{T}_{mm} - \tilde{t}_N)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Phi}_m\right) \\
+ c_N z \bar{t}_N \bar{\epsilon}_N \tilde{\epsilon}_N + c_N z \tilde{t}_N \bar{t}_N \frac{1}{B+1}\mathrm{tr}\left((\overline{\mathbf{T}} - \bar{t}\mathbf{I}_{B+1})\frac{1}{M}\sum_{m=1}^{M}\mathbf{\Phi}_m\right) + \mathcal{O}_z\left(\frac{B}{N}\right)^3. \qquad (2.31)
\end{aligned}$$

The last term of the rhs needs more attention, so we develop again $\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1}$:

$$\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1} = c_N z \bar{t}_N \overline{\mathbf{T}} \frac{1}{M} \sum_{m=1}^{M} \left(\tilde{T}_{mm} \mathbf{\Theta}_m - \tilde{t}_N \mathbf{I}_{B+1} \right) = c_N z \bar{t}_N \tilde{\epsilon}_N \overline{\mathbf{T}} + c_N z \bar{t}_N \overline{\mathbf{T}} \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \tilde{T}_{mm}$$

and get:

$$\begin{aligned}
& c_N \bar{t}_N z \tilde{t}_N \frac{1}{B+1} \operatorname{tr} \left(\left(\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1} \right) \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \right) \\
&\quad = c_N \bar{t}_N z \tilde{t}_N \frac{1}{B+1} \operatorname{tr} \left(c_N z \bar{t}_N \tilde{\epsilon}_N \overline{\mathbf{T}} \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \right) \\
&\qquad + c_N \bar{t}_N z \tilde{t}_N \frac{1}{B+1} \operatorname{tr} \left(\left(c_N z \bar{t}_N \overline{\mathbf{T}} \frac{1}{M} \sum_{m'=1}^{M} \mathbf{\Phi}_{m'} T_{m'm'} \right) \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \right) \\
&\quad = c_N \bar{t}_N z \tilde{t}_N \frac{1}{B+1} \operatorname{tr} \left(\left(c_N z \bar{t}_N \overline{\mathbf{T}} \frac{1}{M} \sum_{m'=1}^{M} \mathbf{\Phi}_{m'} T_{m'm'} \right) \frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \right) \\
&\qquad + \mathcal{O}_z \left(\frac{B}{N} \right)^3 \\
&\quad = c_N^2 z^2 \tilde{t}_N^2 \bar{t}_N^3 \frac{1}{B+1} \operatorname{tr} \left\{ \left(\frac{1}{M} \sum_{m=1}^{M} \mathbf{\Phi}_m \right)^2 \right\} + \mathcal{O}_z \left(\frac{B}{N} \right)^3 .
\end{aligned} \tag{2.32}$$

Plugging (2.32) into (2.31) we get (2.14) as expected.

□

Appendix

2.A Proof of Lemma 2.7

Proof. Recall that $\overline{\mathbf{T}} = (-z\mathbf{I}_{B+1} - \frac{c}{M}\sum_{m=1}^{M} z\tilde{T}_{mm}\mathbf{\Theta}_m)^{-1}$. By approximating $\tilde{\mathbf{T}}$ by $\tilde{t}_N \mathbf{I}_M$, we introduce the intermediate quantities

$$\overline{\mathbf{T}}_{mp} = \left(-z\mathbf{I}_{B+1} - \frac{c}{M}\sum_{m=1}^{M} z\tilde{t}_N\mathbf{\Theta}_m\right)^{-1} = \left(-z\mathbf{I}_{B+1} - cz\tilde{t}_N\mathbf{I}_{B+1} - \frac{c}{M}\sum_{m=1}^{M} z\tilde{t}_N\mathbf{\Phi}_m\right)^{-1}$$

and

$$\tilde{\mathbf{T}}_{mp} = \mathrm{dg}\left(\frac{-1}{z(1+\overline{\delta}_{m,mp})}, m \in [M]\right)$$

where $\overline{\delta}_{m,mp}$ is the Stieltjes transform defined by

$$\overline{\delta}_{m,mp} = \frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}_{mp}\mathbf{\Theta}_m.$$

Similar to the quantities $\tilde{\mathbf{T}}$ and $\overline{\mathbf{T}}$, the following properties hold for $\tilde{\mathbf{T}}_{mp}$ and $\overline{\mathbf{T}}_{mp}$.

Lemma 2.8. *The following properties hold:*

1. *For each $\nu \in [0,1]$, $z \mapsto \tilde{\mathbf{T}}_{mp}(\nu, z)$ and $z \mapsto \overline{\mathbf{T}}_{mp}(\nu, z)$ are holomorphic on $\mathbb{C}_+$,*

2. *For each $\nu \in [0,1]$, if $z \in \mathbb{C}_+$, then $\mathrm{Im}\tilde{\mathbf{T}}_{mp}(\nu, z) \geq 0$, $\mathrm{Im}(z\tilde{\mathbf{T}}_{mp}(\nu, z)) \geq 0$,$\mathrm{Im}\overline{\mathbf{T}}_{mp}(\nu, z) \geq 0$ and $\mathrm{Im}(z\overline{\mathbf{T}}_{mp}(\nu, z)) \geq 0$. Moreover, there exists a positive $M \times M$ matrix-valued measure $\mu^{\tilde{\mathbf{T}}_{mp}(\nu)} = (\mu_{ij}^{\tilde{\mathbf{T}}_{mp}(\nu)})_{ij}$ and a positive $(B+1)\times(B+1)$ matrix-valued measure $\mu^{\overline{\mathbf{T}}_{mp}(\nu)} = (\mu_{ij}^{\overline{\mathbf{T}}_{mp}(\nu)})_{ij}$ such that*

$$\mu^{\tilde{\mathbf{T}}_{mp}(\nu)}(\mathbb{R}_+) = \mathbf{I}_M, \quad \mu^{\overline{\mathbf{T}}_{mp}(\nu)}(\mathbb{R}_+) = \mathbf{I}_{B+1}$$
$$\tilde{\mathbf{T}}_{mp}(\nu, z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu^{\tilde{\mathbf{T}}_{mp}(\nu)}(\lambda)}{\lambda - z}, \quad \overline{\mathbf{T}}_{mp}(\nu, z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu^{\overline{\mathbf{T}}_{mp}(\nu)}(\lambda)}{\lambda - z}$$

   for $z \in \mathbb{C}_+$,

3. *For each $\nu \in [0,1]$, the following inequalities hold true:*

$$\forall z \in \mathbb{C}_+, \quad \tilde{\mathbf{T}}_{mp}(\nu, z)\tilde{\mathbf{T}}_{mp}(\nu, z)^* \leq \frac{\mathbf{I}_M}{\mathrm{Im}^2 z}, \quad \textit{and} \quad \overline{\mathbf{T}}_{mp}(\nu, z)\overline{\mathbf{T}}_{mp}(\nu, z)^* \leq \frac{\mathbf{I}_{B+1}}{\mathrm{Im}^2 z},$$

4. *The sequence of measures $(\frac{1}{M}\mathrm{tr}\ \mu^{\tilde{\mathbf{T}}_{mp}(\nu)})_{N\geq 1}$ and $(\frac{1}{B+1}\mu^{\overline{\mathbf{T}}_{mp}(\nu)})_{N\geq 1}$ are tight.*

In particular, $\frac{1}{M}\mathrm{tr}\ \tilde{\mathbf{T}}_{mp}$ and $\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}_{mp}$ are Stieltjes transforms of probability measures.

Proof. Identical to the proof of Lemma 2.5. □

Denote $\overline{\mathbf{\Delta}}_{mp} = \overline{\mathbf{T}}_{mp} - \bar{t}_N \mathbf{I}_{B+1}$ so that

$$\overline{\mathbf{T}} - \bar{t}_N \mathbf{I}_{B+1} = \overline{\mathbf{T}} - \overline{\mathbf{T}}_{mp} + \overline{\mathbf{\Delta}}_{mp}.$$

We first show that $\|\overline{\mathbf{\Delta}}_{mp}\| = \mathcal{O}_z(\frac{B}{N})$. Indeed,

$$\begin{aligned}
\overline{\mathbf{\Delta}}_{mp} &= \overline{\mathbf{T}}_{mp} - \bar{t}_N \mathbf{I}_{B+1} \\
&= \bar{t}_N \overline{\mathbf{T}}_{mp} (\bar{t}_N^{-1} \mathbf{I}_{B+1} - \overline{\mathbf{T}}_{mp}^{-1}) \\
&= \bar{t}_N \overline{\mathbf{T}}_{mp} \left(-z(1 + c\tilde{t}_N) + z(1 + c\tilde{t}_N + \frac{c}{M} \sum_{m=1}^{M} \tilde{t}_N \mathbf{\Phi}_m) \right) \\
&= \frac{1}{M} \sum_{m=1}^{M} cz\tilde{t}_N \bar{t}_N \overline{\mathbf{T}}_{mp} \mathbf{\Phi}_m.
\end{aligned}$$

Lemma 2.6 implies that $\|\overline{\mathbf{\Delta}}_{mp}\| = \mathcal{O}_z(\frac{B}{N})$. Similarly, we show that $\tilde{\mathbf{\Delta}}_{mp} = \tilde{\mathbf{T}}_{mp} - \tilde{t}_N \mathbf{I}_M$ satisfies $\|\tilde{\mathbf{\Delta}}_{mp}\| = \mathcal{O}_z(\frac{B}{N})$:

$$\begin{aligned}
\tilde{\mathbf{T}}_{mp} - \tilde{t}_N \mathbf{I}_M &= \tilde{t}_N \tilde{\mathbf{T}}_{mp} (\tilde{t}_N^{-1} - \tilde{\mathbf{T}}_{mp}^{-1}) \\
&= \tilde{t}_N \tilde{\mathbf{T}}_{mp} (-z(1 + \bar{t}_N) + z(1 + \bar{\delta}_{m,mp})) \\
&= \tilde{t}_N \tilde{\mathbf{T}}_{mp} z(\bar{\delta}_{m,mp} - \bar{t}_N).
\end{aligned}$$

Since $\bar{\delta}_{m,mp} - \bar{t}_N = \frac{1}{B+1} \mathrm{tr}\,(\overline{\mathbf{T}}_{mp} \mathbf{\Theta}_m - \bar{t}_N \mathbf{I}_{B+1})$, it is cleat that

$$\bar{\delta}_{m,mp} - \bar{t}_N = \frac{1}{B+1} \mathrm{tr}\,(\overline{\mathbf{T}}_{mp} - \bar{t}_N \mathbf{I}_{B+1} + \overline{\mathbf{T}}_{mp} \mathbf{\Phi}_m) \leq \|\overline{\mathbf{\Delta}}_{mp}\| + \|\overline{\mathbf{T}}_{mp}\| \sup_{m \in [M]} \|\mathbf{\Phi}_m\| = \mathcal{O}_z\left(\frac{B}{N}\right).$$

We recap the two approximation results that we obtained:

$$\|\tilde{\mathbf{\Delta}}_{mp}\| = \mathcal{O}_z\left(\frac{B}{N}\right), \quad \|\overline{\mathbf{\Delta}}_{mp}\| = \mathcal{O}_z\left(\frac{B}{N}\right). \tag{2.33}$$

We now show that there exists matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ such that

$$\begin{pmatrix} \bar{\delta}_1 - \bar{\delta}_{1,mp} \\ \vdots \\ \bar{\delta}_M - \bar{\delta}_{M,mp} \end{pmatrix} = \mathbf{M}_1 \mathbf{M}_2 \begin{pmatrix} \bar{\delta}_1 - \bar{\delta}_{1,mp} \\ \vdots \\ \bar{\delta}_M - \bar{\delta}_{M,mp} \end{pmatrix} + \begin{pmatrix} \tilde{\epsilon}_1 \\ \vdots \\ \tilde{\epsilon}_M \end{pmatrix} \tag{2.34}$$

where $\sup_{m \in [M]} |\tilde{\epsilon}_m| = \mathcal{O}_z(\frac{B}{N})$. Consider $\overline{\mathbf{T}} - \overline{\mathbf{T}}_{mp}$:

$$\begin{aligned}
\overline{\mathbf{T}} - \overline{\mathbf{T}}_{mp} &= \overline{\mathbf{T}} (\overline{\mathbf{T}}_{mp}^{-1} - \overline{\mathbf{T}}^{-1}) \overline{\mathbf{T}}_{mp} \\
&= \overline{\mathbf{T}} (\frac{c}{M} \sum_{k=1}^{M} z(\tilde{T}_{kk} - \tilde{t}_N) \mathbf{\Theta}_k) \overline{\mathbf{T}}_{mp} \\
&= \frac{1}{M} \sum_{k=1}^{M} cz(\tilde{T}_{kk} - \tilde{t}_N) \overline{\mathbf{T}} \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp}.
\end{aligned}$$

We now decompose $\tilde{T}_{kk} - \tilde{t}_N = \tilde{T}_{kk} - \tilde{T}_{kk,mp} + \tilde{T}_{kk,mp} - \tilde{t}_N$ and obtain

$$\overline{\mathbf{T}} - \overline{\mathbf{T}}_{mp} = \frac{1}{M} \sum_{k=1}^{M} cz(\tilde{T}_{kk} - \tilde{T}_{kk,mp}) \overline{\mathbf{T}} \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp} + \frac{1}{M} \sum_{k=1}^{M} cz(\tilde{T}_{kk,mp} - \tilde{t}_N) \overline{\mathbf{T}} \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp}. \tag{2.35}$$

Denote $\tilde{\epsilon}_k = cz(\tilde{T}_{kk,mp} - \tilde{t}_N)$, which by (2.33) satisfies $\sup_{m\in[M]} |\tilde{\epsilon}_m| = \mathcal{O}_z(\frac{B}{N})$. Moreover, developing $z\tilde{T}_{kk} - z\tilde{T}_{kk,mp}$, we get:

$$z(\tilde{T}_{kk} - \tilde{T}_{kk,mp}) = z^2\tilde{T}_{kk}\tilde{T}_{kk,mp}(\tilde{T}_{kk,mp}^{-1} - \tilde{T}_{kk}^{-1}) = z^2\tilde{T}_{kk}\tilde{T}_{kk,mp}(\overline{\delta}_k - \overline{\delta}_{k,mp}). \tag{2.36}$$

Plugging (2.36) into (2.35), we get

$$\overline{\mathbf{T}} - \overline{\mathbf{T}}_{mp} = \frac{1}{M}\sum_{k=1}^{M} cz^2\tilde{T}_{kk}\tilde{T}_{kk,mp}(\overline{\delta}_k - \overline{\delta}_{k,mp})\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp} + \frac{1}{M}\sum_{k=1}^{M}\tilde{\epsilon}_k\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}.$$

Multiplying by $\mathbf{\Theta}_m$ and taking the normalized trace, we obtain

$$\overline{\delta}_m - \overline{\delta}_{m,mp} = \frac{1}{M}\sum_{k=1}^{M} cz^2\tilde{T}_{kk}\tilde{T}_{kk,mp}(\overline{\delta}_k - \overline{\delta}_{k,mp})\frac{1}{B+1}\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_m + \frac{1}{M}\sum_{k=1}^{M}\tilde{\epsilon}_k\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_m.$$

The quantity $\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_m$ is $\mathcal{O}_z(1)$. Renaming $\tilde{\epsilon}_m$ in $\tilde{\epsilon}_m\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_m$, we proved that the system of equation (2.34) involving the $\overline{\delta}_m - \overline{\delta}_{m,mp}$ holds with

$$\mathbf{M}_1 = \frac{c}{M}\begin{bmatrix} \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_1\overline{\mathbf{T}} & \dots & \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_M\overline{\mathbf{T}} \\ \vdots & & \vdots \\ \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_1\overline{\mathbf{T}} & \dots & \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_M\overline{\mathbf{T}} \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} z^2\tilde{T}_{11}\tilde{T}_{11,mp} & & 0 \\ & \ddots & \\ 0 & & z^2\tilde{T}_{MM}\tilde{T}_{MM,mp} \end{bmatrix}.$$

The following Lemma 2.9 provides useful bounds for quantities related to $\mathbf{M}_1\mathbf{M}_2$.

Lemma 2.9. *For $N_0 \in \mathbb{N}$ large enough, there exist nice polynomials P_1 and P_2 such that for any z in the set $\mathcal{S}_N$ defined by*

$$\mathcal{S}_N = \left\{ z \in \mathbb{C}_+ : \frac{B}{N}P_1(|z|)P_2\left(\frac{1}{\mathrm{Im} z}\right) \leq 1 \right\}$$

the following inequalities hold:

$$\sup_{N\geq N_0} \rho(\mathbf{M}_1(z)\mathbf{M}_2(z))) \leq 1 - \frac{1}{2}\frac{C_1}{C_2}\frac{\mathrm{Im}^4 z}{16(\eta^2+|z|^2)^2} \tag{2.37}$$

$$\sup_{N\geq N_0} \left\|(\mathbf{I}_M - \mathbf{M}_1(z)\mathbf{M}_2(z))^{-1}\right\|_1 \leq 2\frac{C_2}{C_1}\frac{16(\eta^2+|z|^2)^2}{\mathrm{Im}^4 z} \tag{2.38}$$

where C_1, C_2 are nice constants.

The proof of Lemma 2.9 is deferred to Appendix 2.B. Denote $\mathcal{S}_N \subset \mathbb{C}_+$ such that for any $z \in \mathcal{S}_N$, the spectral radius of $\mathbf{M}_1\mathbf{M}_2$ satisfies $\rho(\mathbf{M}_1\mathbf{M}_2) \leq 1 - \frac{1}{2}\frac{C_1}{C_2}\frac{\mathrm{Im}^4 z}{16(\eta^2+|z|^2)^2}$ so the linear equation (2.34) can be inverted:

$$\begin{pmatrix} \overline{\delta}_1 - \overline{\delta}_{1,mp} \\ \vdots \\ \overline{\delta}_M - \overline{\delta}_{M,mp} \end{pmatrix} = (\mathbf{I}_M - \mathbf{M}_1\mathbf{M}_2)^{-1}\begin{pmatrix} \tilde{\epsilon}_1 \\ \vdots \\ \tilde{\epsilon}_M \end{pmatrix}.$$

Moreover, Lemma 2.9 proves that on the subset $\mathcal{S}_N$ defined by

$$\mathcal{S}_N = \left\{ z \in \mathbb{C}_+ : \frac{B}{N} P_1(|z|) P_2\left(\frac{1}{\mathrm{Im} z}\right) \leq 1 \right\}$$

where P_1 and P_2 are nice polynomials, $\sup_{N \geq N_0} \left\| (\mathbf{I}_M - \mathbf{M}_1(z)\mathbf{M}_2(z))^{-1} \right\|_1 = \mathcal{O}_z(1)$, so the following holds:

$$\max_{m \in [M]} \{ |\overline{\delta}_m - \overline{\delta}_{m,mp}| \} \leq \left\| (\mathbf{I}_M - \mathbf{M}_1 \mathbf{M}_2)^{-1} \right\|_1 \sup_{m \in [M]} |\tilde{\epsilon}_m| = \mathcal{O}_z\left(\frac{B}{N}\right)$$

For $z \in \mathbb{C}_+ \backslash \mathcal{S}_N$, we have $1 \leq \frac{B}{N} P_1(|z|) P_2\left(\frac{1}{\mathrm{Im} z}\right)$. Moreover, we recall that $\overline{\delta}_m$ and $\overline{\delta}_{m,mp}$ are Stieltjes transform of measures carried by $\mathbb{R}_+$ with total mass equal to $\frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_m$, so it is possible to write

$$\sup_{m \in [M]} \left| \overline{\delta}_m - \overline{\delta}_{m,mp} \right| \leq \frac{2}{\mathrm{Im} z} \frac{1}{B+1} \sup_{m \in [M]} |\mathrm{tr}\ \mathbf{\Theta}_m| \leq 2C \frac{B}{N} P_1(|z|) P_2\left(\frac{1}{\mathrm{Im} z}\right) = \mathcal{O}_z\left(\frac{B}{N}\right).$$

This ends the proof of Lemma 2.7. □

2.B Proof of Lemma 2.9

This proof follows closely the one from [48, Proposition 10].

Proof. $\mathbf{M}_1$ and $\mathbf{M}_2$ depend simultaneously on $(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}_{mp})$ and $(\overline{\mathbf{T}}, \overline{\mathbf{T}}_{mp})$, which make their study difficult. We first show how we can decouple these quantities. By the Cauchy-Schwartz inequality, we have:

$$\begin{aligned}
\left| \mathrm{tr}\ \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_l \overline{\mathbf{T}} \right| &\leq \sqrt{\mathrm{tr}\ \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_l \overline{\mathbf{T}}_{mp}^*} \sqrt{\mathrm{tr}\ \mathbf{\Theta}_l \overline{\mathbf{T}} \mathbf{\Theta}_k \overline{\mathbf{T}}^*} \\
\left| z^2 \tilde{T}_{kk} \tilde{T}_{kk,mp} \right| &\leq \sqrt{|z \tilde{T}_{kk}|^2} \sqrt{|z \tilde{T}_{kk,mp}|^2}.
\end{aligned}$$

The matrix with entries $((\frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_k \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_l \overline{\mathbf{T}}_{mp}^*)(\frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_l \overline{\mathbf{T}} \mathbf{\Theta}_k \overline{\mathbf{T}}^*))_{kl}$ can be written as the Hadamard product $\mathbf{N}_{11} \odot \mathbf{N}_{12}$ where

$$\mathbf{N}_{11} = \frac{c}{M} \begin{bmatrix} \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_1 \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_1 \overline{\mathbf{T}}_{mp}^* & \dots & \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_1 \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_M \overline{\mathbf{T}}_{mp}^* \\ \vdots & & \vdots \\ \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_M \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_1 \overline{\mathbf{T}}_{mp}^* & \dots & \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_M \overline{\mathbf{T}}_{mp} \mathbf{\Theta}_M \overline{\mathbf{T}}_{mp}^* \end{bmatrix} \tag{2.39}$$

$$\mathbf{N}_{12} = \frac{c}{M} \begin{bmatrix} \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_1 \overline{\mathbf{T}} \mathbf{\Theta}_1 \overline{\mathbf{T}}^* & \dots & \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_M \overline{\mathbf{T}} \mathbf{\Theta}_1 \overline{\mathbf{T}}^* \\ \vdots & & \vdots \\ \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_1 \overline{\mathbf{T}} \mathbf{\Theta}_M \overline{\mathbf{T}}^* & \dots & \frac{1}{B+1} \mathrm{tr}\ \mathbf{\Theta}_M \overline{\mathbf{T}} \mathbf{\Theta}_M \overline{\mathbf{T}}^* \end{bmatrix}. \tag{2.40}$$

Similarly the matrix with diagonal entries $(|z\tilde{T}_{kk}|^2 |z\tilde{T}_{kk,mp}|^2)_{k \in [M]}$ can be written $\mathbf{N}_{21} \odot \mathbf{N}_{22}$ where

$$\mathbf{N}_{21} = \begin{bmatrix} |z\tilde{T}_{11,mp}|^2 & & 0 \\ & \ddots & \\ 0 & & |z\tilde{T}_{MM,mp}|^2 \end{bmatrix} \tag{2.41}$$

$$\mathbf{N}_{22} = \begin{bmatrix} |z\tilde{T}_{11}|^2 & & 0 \\ & \ddots & \\ 0 & & |z\tilde{T}_{MM}|^2 \end{bmatrix}. \tag{2.42}$$

By [43, Theorem 8.1.18], denoting $|\mathbf{M}_1 \mathbf{M}_2|$ the matrix with entries $(|(\mathbf{M}_1 \mathbf{M}_2)_{ij}|)_{i,j=1}^M$, $\rho(\mathbf{M}_1 \mathbf{M}_2) \leq$

$\rho(|\mathbf{M}_1\mathbf{M}_2|)$, and using [44, Lemma 5.7.9], we have

$$\rho(|\mathbf{M}_1\mathbf{M}_2|) \leq \sqrt{\rho(\mathbf{N}_{11}\mathbf{N}_{21})}\sqrt{\rho(\mathbf{N}_{12}\mathbf{N}_{22})}. \tag{2.43}$$

Therefore, it remains to bound the two following spectral radii: $\rho(\mathbf{N}_{11}\mathbf{N}_{21})$ and $\rho(\mathbf{N}_{12}\mathbf{N}_{22})$.

Bound for $\rho(\mathbf{N}_{12}\mathbf{N}_{22})$. We will show that it is possible to obtain a linear system involving the positive quantities $(\mathrm{Im}(z\tilde{T}_{mm})$ and $\mathrm{Im}\overline{\delta}_m)$ with the matrix $\mathbf{N}_{12}\mathbf{N}_{22}$. Moreover,

$$\mathrm{Im}(z\tilde{T}_{mm}) = \frac{z\tilde{T}_{mm} - \overline{z\tilde{T}_{mm}}}{2i} = (z\tilde{T}_{mm})\overline{z\tilde{T}_{mm}}\frac{1+\overline{\delta}_m - 1 - \overline{\overline{\delta}_m}}{2i} = |z\tilde{T}_{mm}|^2\mathrm{Im}\overline{\delta}_m$$

so that

$$\begin{pmatrix} \mathrm{Im}(z\tilde{T}_{11}) \\ \vdots \\ \mathrm{Im}(z\tilde{T}_{MM}) \end{pmatrix} = \begin{bmatrix} |z\tilde{T}_{11}|^2 & & 0 \\ & \ddots & \\ 0 & & |z\tilde{T}_{MM}|^2 \end{bmatrix} \begin{pmatrix} \mathrm{Im}\overline{\delta}_1 \\ \vdots \\ \mathrm{Im}\overline{\delta}_M \end{pmatrix} = \mathbf{N}_{22} \begin{pmatrix} \mathrm{Im}\overline{\delta}_1 \\ \vdots \\ \mathrm{Im}\overline{\delta}_M \end{pmatrix} \tag{2.44}$$

where we recognized $\mathbf{N}_{22}$. Moreover, since $\mathbf{\Theta}_m$ is Hermitian, $\mathrm{Im}(\overline{\delta}_m) = \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\mathrm{Im}\overline{\mathbf{T}}$, and

$$\begin{aligned}
\mathrm{Im}\overline{\mathbf{T}} &= \frac{\overline{\mathbf{T}} - \overline{\mathbf{T}}^*}{2i} \\
&= \frac{1}{2i}\overline{\mathbf{T}}\left(\overline{-z\mathbf{I}_{B+1} - \frac{c}{M}\sum_{k=1}^{M} z\tilde{T}_{kk}\mathbf{\Theta}_k} + z\mathbf{I}_{B+1} + \frac{c}{M}\sum_{k=1}^{M} z\tilde{T}_{kk}\mathbf{\Theta}_k\right)\overline{\mathbf{T}}^* \\
&= \overline{\mathbf{T}}\left((\mathrm{Im}z)\mathbf{I}_{B+1} + \frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk}\mathbf{\Theta}_k)\right)\overline{\mathbf{T}}^* \\
&= (\mathrm{Im}z)\overline{\mathbf{T}\mathbf{T}}^* + \frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk})\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}^*
\end{aligned}$$

so that

$$\mathrm{Im}(\overline{\delta}_m) = \frac{\mathrm{Im}z}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\overline{\mathbf{T}\mathbf{T}}^* + \frac{c}{M}\sum_{k=1}^{M}\mathrm{Im}(z\tilde{T}_{kk})\frac{1}{B+1}\mathrm{tr}\ \left\{\mathbf{\Theta}_m\overline{\mathbf{T}}\mathbf{\Theta}_k\overline{\mathbf{T}}^*\right\}.$$

This leads to:

$$\begin{pmatrix} \mathrm{Im}\,\overline{\delta}_1 \\ \vdots \\ \mathrm{Im}\,\overline{\delta}_M \end{pmatrix} = \frac{c}{M}\begin{bmatrix} \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_1\overline{\mathbf{T}}\mathbf{\Theta}_1\overline{\mathbf{T}}^* & \ldots & \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_1\overline{\mathbf{T}}\mathbf{\Theta}_M\overline{\mathbf{T}}^* \\ \vdots & & \vdots \\ \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}}\mathbf{\Theta}_1\overline{\mathbf{T}}^* & \ldots & \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}}\mathbf{\Theta}_M\overline{\mathbf{T}}^* \end{bmatrix}\begin{pmatrix} \mathrm{Im}(z\tilde{T}_{11}) \\ \vdots \\ \mathrm{Im}(z\tilde{T}_{MM}) \end{pmatrix} + \begin{pmatrix} \mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\overline{\mathbf{T}\mathbf{T}}^* \\ \vdots \\ \mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\overline{\mathbf{T}\mathbf{T}}^* \end{pmatrix}$$

where we recognize $\mathbf{N}_{12}$.

$$\begin{pmatrix} \mathrm{Im}\,\overline{\delta}_1 \\ \vdots \\ \mathrm{Im}\,\overline{\delta}_M \end{pmatrix} = \mathbf{N}_{12}\begin{pmatrix} \mathrm{Im}(z\tilde{T}_{11}) \\ \vdots \\ \mathrm{Im}(z\tilde{T}_{MM}) \end{pmatrix} + \begin{pmatrix} \mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\overline{\mathbf{T}\mathbf{T}}^* \\ \vdots \\ \mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m\overline{\mathbf{T}\mathbf{T}}^* \end{pmatrix}. \tag{2.45}$$

Combining (2.44) and (2.45), we get:

$$\begin{pmatrix} \operatorname{Im}\overline{\delta}_1 \\ \vdots \\ \operatorname{Im}\overline{\delta}_M \end{pmatrix} = \mathbf{N}_{12}\mathbf{N}_{22} \begin{pmatrix} \operatorname{Im}\overline{\delta}_1 \\ \vdots \\ \operatorname{Im}\overline{\delta}_M \end{pmatrix} + \begin{pmatrix} \frac{\operatorname{Im} z}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_1 \overline{\mathbf{T}}\overline{\mathbf{T}}^* \\ \vdots \\ \frac{\operatorname{Im} z}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_M \overline{\mathbf{T}}\overline{\mathbf{T}}^* \end{pmatrix}.$$

Denote

$$\mathbf{u} = (\operatorname{Im}\overline{\delta}_1, \ldots, \operatorname{Im}\overline{\delta}_M)^T$$
$$\mathbf{v} = \left(\frac{\operatorname{Im} z}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_1 \overline{\mathbf{T}}\overline{\mathbf{T}}^*, \ldots, \frac{\operatorname{Im} z}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_M \overline{\mathbf{T}}\overline{\mathbf{T}}^*\right)^T.$$

We say that a vector or a matrix is positive if each of its entries is positive. By Lemma 2.5, for each $m \geq 1$ the mapping $z \mapsto \overline{\delta}_m(z)$ is a Stieltjes transform, so $\operatorname{Im}\overline{\delta}_m(z) > 0$ for $z \in \mathbb{C}_+$. This implies that $\mathbf{u}$ is a positive vector. It is also clear that the entries of $\mathbf{v}$ and $\mathbf{N}_{22}$ are positive. Concerning the matrix $\mathbf{N}_{12}$, its entries can be written as

$$\frac{1}{B+1}\operatorname{tr}\, \left\{ \left(\boldsymbol{\Theta}_m^{\frac{1}{2}}\overline{\mathbf{T}}\boldsymbol{\Theta}_k^{\frac{1}{2}}\right)\left(\boldsymbol{\Theta}_m^{\frac{1}{2}}\overline{\mathbf{T}}\boldsymbol{\Theta}_k^{\frac{1}{2}}\right)^* \right\}$$

and since $\operatorname{tr}\, \mathbf{A}\mathbf{A}^* \geq 0$ for any matrix $\mathbf{A}$, it is clear that $\mathbf{N}_{12}$ is a matrix with positive elements. Write $\mathbf{N}_{12}\mathbf{N}_{22}\mathbf{u} = \mathbf{u} - \mathbf{v}$. It is clear that, since $u_l > 0$ for each l, each entry of $\mathbf{u} - \mathbf{v}$ is bounded as follows:

$$(\mathbf{u} - \mathbf{v})_l = u_l\left(1 - \frac{v_l}{u_l}\right) \leq u_l\left(1 - \frac{\min_{l\in[M]} v_l}{\max_{l\in[M]} u_l}\right).$$

By [43, Corollary 8.1.30], this implies the following bound on the spectral radius of $\mathbf{N}_{12}\mathbf{N}_{22}$: for any $z \in \mathbb{C}_+$,

$$\rho(\mathbf{N}_{12}\mathbf{N}_{22}) \leq 1 - \frac{\min_{l\in[M]} v_l}{\max_{l\in[M]} u_l}. \tag{2.46}$$

It remains now to find lower bounds for $\min_{l\in[M]} v_l$ and $\frac{1}{\max_{l\in[M]} u_l}$ when $z \in \mathbb{C}_+$. Since $\sup_{m\in[M]} \|\boldsymbol{\Theta}_m - \mathbf{I}_{B+1}\| = \mathcal{O}(\frac{B}{N})$, it is clear that there exists $N_0 \in \mathbb{N}$ such that

$$0 < C_1 < \inf_{N\geq N_0} \inf_{m\in[M]} \inf_{\nu\in[0,1]} \frac{1}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_m \leq \sup_{N\geq N_0} \sup_{m\in[M]} \sup_{\nu\in[0,1]} \frac{1}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_m < C_2 < +\infty$$

for some nice constants C_1 and C_2. This implies the following bounds:

$$\sup_{N\geq N_0} \max_{m\in[M]} (\operatorname{Im}\overline{\delta}_m) \leq \sup_{N\geq N_0} \sup_{m\in[M]} \frac{1}{B+1}\left|\operatorname{tr}\, \boldsymbol{\Theta}_m \overline{\mathbf{T}}\right| \leq \frac{C_2}{\operatorname{Im} z}$$

so

$$\sup_{N\geq N_0} \max_{l\in[M]} u_l \leq \frac{C_2}{\operatorname{Im} z}. \tag{2.47}$$

It remains to lower bound $\inf_{N\geq N_0} \min_{l\in[M]} v_l$. It is clear that since $\boldsymbol{\Theta}$ is Hermitian,

$$v_l = \operatorname{Im} z \frac{1}{B+1}\operatorname{tr}\, \boldsymbol{\Theta}_l \overline{\mathbf{T}}\overline{\mathbf{T}}^* \geq C_1 \operatorname{Im}(z) \frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}\overline{\mathbf{T}}^* \tag{2.48}$$

so it remains to lower bound $\frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}\overline{\mathbf{T}}^*$. By Lemma 12 from [48],

$$\frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}(z)\overline{\mathbf{T}}(z)^* \geq \left|\frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}(z)\right|^2 \geq \operatorname{Im}\left(\frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}(z)\right)^2.$$

Rewrite $\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(z)$ as $\int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda)}{\lambda - z}$ so that

$$\mathrm{Im}\left(\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}(z)\right) = \mathrm{Im}\left(\int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda)}{\lambda - z}\right) = \mathrm{Im}(z)\left(\int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda)}{|\lambda - z|^2}\right).$$

Since by Lemma 2.5 the measure $\mu^{\overline{\mathbf{T}}}$ is tight, there exists $\eta > 0$ (not depending on N) such that $\mu^{\overline{\mathbf{T}}}([0,\eta]) \geq \frac{1}{2}$ for all $N \geq N_0$. This implies that

$$\mathrm{Im}(z)\left(\int_{\mathbb{R}} \frac{\mu^{\overline{\mathbf{T}}}(\mathrm{d}\lambda)}{|\lambda - z|^2}\right) \geq \int_0^{\eta} \frac{\mathrm{Im}(z)\,\mathrm{d}\mu^{\overline{\mathbf{T}}}(\lambda)}{2(\eta^2 + |z|^2)}\mu^{\overline{\mathbf{T}}}([0,\eta]) \geq \frac{\mathrm{Im} z}{4(\eta^2 + |z|^2)}. \tag{2.49}$$

Plugging (2.49) into (2.48), we get:

$$\inf_{N \geq N_0} \inf_{\nu \in [0,1]} \inf_{l \in [M]} v_l \geq C_1 \frac{\mathrm{Im}^3 z}{16(\eta^2 + |z|^2)^2}.$$

This gives the desired bound for $\rho(\mathbf{N}_{12}\mathbf{N}_{22})$.

$$\rho(\mathbf{N}_{12}\mathbf{N}_{22}) \leq 1 - \frac{C_1}{C_2}\frac{\mathrm{Im}^4 z}{16(\eta^2 + |z|^2)^2}. \tag{2.50}$$

Bound for $\rho(\mathbf{N}_{11}\mathbf{N}_{21})$. The proof for this bound is similar to the previous one, so we will omit some details. Write

$$\mathrm{Im}(z\tilde{T}_{mm,mp}) = |z\tilde{T}_{mm,mp}|^2 \mathrm{Im}\,\overline{\delta}_{m,mp}.$$

This provides the following equation:

$$\begin{pmatrix} \mathrm{Im}(z\tilde{T}_{11,mp}) \\ \vdots \\ \mathrm{Im}(z\tilde{T}_{MM,mp}) \end{pmatrix} = \begin{bmatrix} |z\tilde{T}_{11,mp}|^2 & & 0 \\ & \ddots & \\ 0 & & |z\tilde{T}_{MM,mp}|^2 \end{bmatrix} \begin{pmatrix} \mathrm{Im}\,\overline{\delta}_{1,mp} \\ \vdots \\ \mathrm{Im}\,\overline{\delta}_{M,mp} \end{pmatrix}$$

where we recognize $\mathbf{N}_{21}$:

$$\begin{pmatrix} \mathrm{Im}(z\tilde{T}_{11,mp}) \\ \vdots \\ \mathrm{Im}(z\tilde{T}_{MM,mp}) \end{pmatrix} = \mathbf{N}_{21} \begin{pmatrix} \mathrm{Im}\,\overline{\delta}_{1,mp} \\ \vdots \\ \mathrm{Im}\,\overline{\delta}_{M,mp} \end{pmatrix}. \tag{2.51}$$

Since $\mathbf{\Theta}_m$ is a Hermitian matrix, one can express $\mathrm{Im}\,\overline{\delta}_{m,mp}$ as:

$$\mathrm{Im}\,\overline{\delta}_{m,mp} = \frac{1}{B+1}\mathrm{tr}\ \mathrm{Im}(\mathbf{\Theta}_m \overline{\mathbf{T}}_{mp}) = \frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_m \mathrm{Im}\,\overline{\mathbf{T}}_{mp}$$

and write $\mathrm{Im}\,\overline{\mathbf{T}}_{mp}$ as:

$$\begin{aligned}
\mathrm{Im}\,\overline{\mathbf{T}}_{mp} &= \frac{\overline{\mathbf{T}}_{mp} - \overline{\mathbf{T}}_{mp}^*}{2i} \\
&= \overline{\mathbf{T}}_{mp}\left(\frac{z - \overline{z}}{2i}\mathbf{I}_{B+1} + \frac{c}{M}\sum_{k=1}^{M} \frac{z\tilde{t}_N - \overline{z\tilde{t}_N}}{2i}\mathbf{\Theta}_k\right)\overline{\mathbf{T}}_{mp}^* \\
&= \mathrm{Im}(z)\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \frac{c}{M}\sum_{k=1}^{M} \mathrm{Im}(z\tilde{t}_N)\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}^*.
\end{aligned}$$

Using $\tilde{t}_N = \tilde{T}_{kk,mp} + \tilde{t}_N - \tilde{T}_{kk,mp}$, we get:

$$\operatorname{Im} \overline{\mathbf{T}}_{mp} = \operatorname{Im}(z)\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \frac{c}{M}\sum_{k=1}^{M} \operatorname{Im}(z\tilde{T}_{kk,mp})\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}^* \\ + \frac{c}{M}\sum_{k=1}^{M} \operatorname{Im}(z(\tilde{t}_N - \tilde{T}_{kk,mp}))\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}^*$$

so left-multiplying by $\mathbf{\Theta}_m$ and taking the normalized trace we obtain

$$\operatorname{Im} \overline{\delta}_{m,mp} = \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_m\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \frac{c}{M}\sum_{k=1}^{M} \operatorname{Im}(z\tilde{T}_{kk,mp})\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_m\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}^* + \epsilon_m$$

where $\epsilon_m = \frac{c}{M}\sum_{k=1}^{M} \operatorname{Im}(z(\tilde{t}_N - \tilde{T}_{kk,mp}))\frac{1}{B+1}\operatorname{tr}\, \overline{\mathbf{T}}_{mp}\mathbf{\Theta}_k\overline{\mathbf{T}}_{mp}^*\mathbf{\Theta}_m$ is by (2.33) a $\mathcal{O}_z(\frac{B}{N})$ quantity. so we obtain the following equations:

$$\begin{pmatrix} \operatorname{Im} \overline{\delta}_{1,mp} \\ \vdots \\ \operatorname{Im} \overline{\delta}_{M,mp} \end{pmatrix} = \frac{c}{M}\begin{bmatrix} \frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}^* & \dots & \frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}^* \\ \vdots & & \vdots \\ \frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}^* & \dots & \frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}^* \end{bmatrix}\begin{pmatrix} \operatorname{Im}(z\tilde{T}_{11,mp}) \\ \vdots \\ \operatorname{Im}(z\tilde{T}_{MM,mp}) \end{pmatrix} \\ + \begin{pmatrix} \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_1 \\ \vdots \\ \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_M \end{pmatrix}$$

and we recover matrix $\mathbf{N}_{11}$:

$$\begin{pmatrix} \operatorname{Im} \overline{\delta}_{1,mp} \\ \vdots \\ \operatorname{Im} \overline{\delta}_{M,mp} \end{pmatrix} = \mathbf{N}_{11}\begin{pmatrix} \operatorname{Im}(z\tilde{T}_{11,mp}) \\ \vdots \\ \operatorname{Im}(z\tilde{T}_{MM,mp}) \end{pmatrix} + \begin{pmatrix} \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_1 \\ \vdots \\ \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_M \end{pmatrix}. \tag{2.52}$$

Combining (2.51) and (2.52), we get:

$$\begin{pmatrix} \operatorname{Im} \overline{\delta}_{1,mp} \\ \vdots \\ \operatorname{Im} \overline{\delta}_{M,mp} \end{pmatrix} = \mathbf{N}_{11}\mathbf{N}_{21}\begin{pmatrix} \operatorname{Im} \overline{\delta}_{1,mp} \\ \vdots \\ \operatorname{Im} \overline{\delta}_{M,mp} \end{pmatrix} + \begin{pmatrix} \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_1 \\ \vdots \\ \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_M \end{pmatrix}. \tag{2.53}$$

This system of equations is similar to the one obtained previously, but it involves the additional terms $\epsilon_1, \dots, \epsilon_M$. Denote $\mathbf{u} = (\operatorname{Im} \overline{\delta}_{1,mp}, \dots, \operatorname{Im} \overline{\delta}_{M,mp})^T$ and

$$\mathbf{v} = \begin{pmatrix} \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_1\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_1 \\ \vdots \\ \operatorname{Im}(z)\frac{1}{B+1}\operatorname{tr}\, \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}_{mp}^* + \epsilon_M \end{pmatrix}.$$

The mappings $z \mapsto \overline{\delta}_{m,mp}(z)$ for $m \geq 1$ are Stieltjes transforms, so their imaginary parts are positive. This implies that $\mathbf{u}$ is a positive vector. By the same argument as before, $\mathbf{N}_{11}\mathbf{N}_{21}$ is also a positive matrix. Moreover, the following bounds holds:

$$\sup_{N \geq N_0} \max_{l \in [M]} u_l \leq \frac{C_2}{\operatorname{Im} z}$$

Moreover, by (2.8) the measure associated with $\frac{1}{B+1}\mathrm{tr}\ \overline{\mathbf{T}}_{mp}$ is a tight measure, so it is possible to repeat the argument to bound $\mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}\mathbf{T}}^*$ and obtain that $\mathrm{Im}(z)\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_M\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}^*_{mp}$ is positive for each m, and satisfies:

$$\inf_{N\geq N_0}\min_{l\in[M]}\frac{\mathrm{Im}z}{B+1}\mathrm{tr}\ \mathbf{\Theta}_l\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}^*_{mp}\geq C_1\frac{\mathrm{Im}^3 z}{16(\eta^2+|z|^2)^2}.$$

However, to show that $\mathbf{v}$ is a positive vector, more care is needed to handle the perturbations ϵ_m. Since the ϵ_m are small in the sense that there exist nice polynomials (the coefficients and degrees do not depend on N and z) P_1 and P_2 such that $\sup_{m\in[M]}|\epsilon_m|\leq\frac{B}{N}P_1(|z|)P_2\left(\frac{1}{\mathrm{Im}z}\right)$, it is possible to write:

$$\inf_{m\in[M]}\epsilon_m\geq-\sup_{m\in[M]}|\epsilon_m|\geq-\frac{B}{N}P_1(|z|)P_2\left(\frac{1}{\mathrm{Im}z}\right).$$

Denote by $\mathcal{S}_N$ the following subset of $\mathbb{C}$

$$\mathcal{S}_N=\left\{z\in\mathbb{C}:C_1\frac{\mathrm{Im}^3 z}{16(\eta^2+|z|^2)^2}-\frac{B}{N}P_1(|z|)P_2\left(\frac{1}{\mathrm{Im}z}\right)>\frac{1}{2}C_1\frac{\mathrm{Im}^3 z}{16(\eta^2+|z|^2)^2}\right\}$$

so that on $\mathcal{S}_N$,

$$\min_{l\in[M]}v_l=\inf_{N\geq N_0}\min_{l\in[M]}\left(\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Theta}_l\overline{\mathbf{T}}_{mp}\overline{\mathbf{T}}^*_{mp}-\epsilon_m\right)\geq\frac{1}{2}C_1\frac{\mathrm{Im}^3 z}{16(\eta^2+|z|^2)^2}.$$

It is clear that $\mathcal{S}_N$ is of the form

$$\mathcal{S}_N=\left\{z\in\mathbb{C}:\frac{B}{N}\tilde{P}_1(|z|)\tilde{P}_2\left(\frac{1}{\mathrm{Im}z}\right)\leq 1\right\}\tag{2.54}$$

where $\tilde{P}_1$ and $\tilde{P}_2$ are nice polynomials, and $\mathbf{v}$ is a positive vector on $\mathcal{S}_N$. It is now possible to use [43, Corollary 8.1.30] and write the following bound on the spectral radius of $\mathbf{N}_{11}\mathbf{N}_{21}$:

$$\rho(\mathbf{N}_{11}\mathbf{N}_{21})\leq 1-\frac{\min_{l\in[M]}v_l}{\max_{l\in[M]}u_l}$$

which proves that for $z\in\mathcal{S}_N$:

$$\rho(\mathbf{N}_{11}(z)\mathbf{N}_{12}(z))\leq 1-\frac{1}{2}\frac{C_1}{C_2}\frac{\mathrm{Im}^4 z}{16(\eta^2+|z|^2)^2}.\tag{2.55}$$

Bound for $\rho(\mathbf{M}_1\mathbf{M}_2)$ Using (2.55) and (2.50) in (2.43), we obtain (2.37) as desired.

Bound for $\|(\mathbf{I}_M-\mathbf{M}_1\mathbf{M}_2)^{-1}\|_1$. We defined previously matrices $\mathbf{N}_{11},\mathbf{N}_{12},\mathbf{N}_{21}$ and $\mathbf{N}_{22}$ such that for each $i,j\in[M]$,

$$|(\mathbf{M}_1\mathbf{M}_2)_{i,j}|\leq\sqrt{|(\mathbf{N}_{11}\mathbf{N}_{21})_{i,j}|}\sqrt{|(\mathbf{N}_{12}\mathbf{N}_{22})_{i,j}|}.\tag{2.56}$$

By Lemma 13 from [48], this implies that

$$\|(\mathbf{I}_M-\mathbf{M}_1\mathbf{M}_2)^{-1}\|_1\leq\sqrt{\|(\mathbf{I}_M-\mathbf{N}_{11}\mathbf{N}_{21})^{-1}\|_1}\sqrt{\|(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}\|_1}$$

so it is enough to control and $\|(\mathbf{I}_M-\mathbf{N}_{11}\mathbf{N}_{21})^{-1}\|_1$ and $\|(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}\|_1$. Since $\rho(\mathbf{N}_{12}\mathbf{N}_{22})<1$, the inverse $(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}$ is well defined. Moreover, using $(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}=\sum_{k\geq 0}(\mathbf{N}_{12}\mathbf{N}_{22})^k$, the entries of $(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}$ are all positive. Finally, $(\mathbf{I}_M-\mathbf{N}_{12}\mathbf{N}_{22})^{-1}$ satisfies $(\mathbf{I}_M-$

$\mathbf{N}_{12}\mathbf{N}_{22})^{-1}\mathbf{u} = \mathbf{v}$ where we recall that

$$\mathbf{u} = (\operatorname{Im}\overline{\delta}_1, \ldots, \operatorname{Im}\overline{\delta}_M)^T$$
$$\mathbf{v} = \left(\frac{\operatorname{Im} z}{B+1}\operatorname{tr}\,\boldsymbol{\Theta}_1\overline{\mathbf{T}\mathbf{T}}^*, \ldots, \frac{\operatorname{Im} z}{B+1}\operatorname{tr}\,\boldsymbol{\Theta}_M\overline{\mathbf{T}\mathbf{T}}^*\right)^T.$$

This implies that

$$\|(\mathbf{I}_M - \mathbf{N}_{12}\mathbf{N}_{22})^{-1}\|_1 \leq \frac{\max_k u_k}{\min_k v_k} \leq \frac{C_2}{C_1}\frac{16(\eta^2 + |z|^2)^2}{\operatorname{Im}^4 z}. \tag{2.57}$$

Moreover, we similarly have for $\|(\mathbf{I}_M - \mathbf{N}_{11}\mathbf{N}_{21})^{-1}\|_1$ the following bound:

$$\|(\mathbf{I}_M - \mathbf{N}_{11}\mathbf{N}_{21})^{-1}\|_1 \leq \frac{\max_k u_k}{\min_k v_k}$$

where for $z \in \mathcal{S}_N$ defined in (2.54) gives

$$\|(\mathbf{I}_M - \mathbf{N}_{11}\mathbf{N}_{21})^{-1}\|_1 \leq 2\frac{C_2}{C_1}\frac{16(\eta^2 + |z|^2)^2}{\operatorname{Im}^4 z}. \tag{2.58}$$

Combining (2.57) and (2.58) in (2.56), we get (2.38) as expected.

□

Part II

Detection of low-rank signal

Chapter 3

Detection when all the signal to noise ratio per sensor vanishes

3.1 Introduction

Detecting the presence of an unknown multivariate signal corrupted by noise is one of the fundamental problems in signal processing, which is found in many applications including array and radar processing, wireless communications, radio-astronomy or seismology among others. In a statistical framework, this problem is usually formulated as the following binary hypothesis test, where the objective is to discriminate between the null hypothesis $\mathcal{H}_0$ and the alternative hypothesis $\mathcal{H}_1$ defined as

$$\begin{aligned} \mathcal{H}_0 &: (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{v}_n)_{n\in\mathbb{Z}} \\ \mathcal{H}_1 &: (\mathbf{y}_n)_{n\in\mathbb{Z}} = (\mathbf{u}_n)_{n\in\mathbb{Z}} + (\mathbf{v}_n)_{n\in\mathbb{Z}} \end{aligned} \tag{3.1}$$

where $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is the M-variate observed signal, and where $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ and $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ represent a non observable signal of interest and the noise respectively, both modeled in this chapter as mutually independent zero-mean complex Gaussian stationary time series.

Without further knowledge on the covariance function of $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ and/or $(\mathbf{u}_n)_{n\in\mathbb{Z}}$, or access to "noise only" samples, the test problem (3.1) is ill-posed, even for temporally white time series $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ and $(\mathbf{u}_n)_{n\in\mathbb{Z}}$, and one needs to exploit additional information on the covariance structure of the useful signal and noise. One common assumption, widely used in the context of array processing and multi-antenna communications, is to consider that the noise $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is spatially uncorrelated. Moreover, when the receiver antennas are not calibrated, it is reasonable to assume that the spectral densities of the components of the noise may not coincide, see e.g. [72], [82], [12], [54]. This will be the context of the present chapter.

A first class of tests is based on the observation that the noise is spatially uncorrelated if and only if the matrices $\mathbf{R}_{\mathbf{v}}(\ell) = \mathbb{E}[\mathbf{v}_n\mathbf{v}_{n-\ell}^*]$ are diagonal for all $\ell \in \mathbb{Z}$, whereas if the useful signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ is assumed spatially correlated, $\mathbf{R}_{\mathbf{u}}(\ell) = \mathbb{E}[\mathbf{u}_n\mathbf{u}_{n-\ell}^*]$ is non-diagonal for some $\ell \in \mathbb{Z}$. Under this assumption, the problem in (3.1) can be formulated as the following correlation test:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{R}_{\mathbf{y}}(\ell) = \mathrm{dg}\left(\mathbf{R}_{\mathbf{y}}(\ell)\right) \text{ for all } \ell \in \mathbb{Z} \\ \mathcal{H}_1 &: \mathbf{R}_{\mathbf{y}}(\ell) \neq \mathrm{dg}\left(\mathbf{R}_{\mathbf{y}}(\ell)\right) \text{ for some } \ell \in \mathbb{Z} \end{aligned} \tag{3.2}$$

where $\mathbf{R}_{\mathbf{y}}(\ell) = \mathbb{E}[\mathbf{y}_n\mathbf{y}_{n-\ell}^*]$ and $\mathrm{dg}\left(\mathbf{R}_{\mathbf{y}}(\ell)\right) = \mathbf{R}_{\mathbf{y}}(\ell) \odot \mathbf{I}_M$, where $\odot$ is the element-wise (Hadamard) product and $\mathbf{I}_M$ the $M \times M$ identity matrix. A number of previous works developed lag domains tests that specifically tackle the above problem, see e.g. [38], [42], [55], [41], [73], [50]. Also relevant are the approaches in [72] and [82], where the possible useful signal is supposed to be the output of a filter driven by a low-dimensional white noise sequence.

Our focus here is on another type of formulation, referred to as *frequency domain approach*, which consists in rewriting problem (3.2) as

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{S}_{\mathbf{y}}(\nu) = \mathrm{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right) \text{ for all } \nu \in [0,1] \\ \mathcal{H}_1 &: \mathbf{S}_{\mathbf{y}}(\nu) \neq \mathrm{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right) \text{ for some } \nu \in [0,1] \end{aligned} \tag{3.3}$$

where $\mathbf{S}_{\mathbf{y}}(\nu)$ is the $M \times M$ spectral density matrix of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ at frequency ν, defined by

$$\mathbf{S}_{\mathbf{y}}(\nu) = \sum_{k\in\mathbb{Z}} \mathbf{R}_{\mathbf{y}}(k) e^{-2\mathrm{i}\pi\nu k}.$$

This problem is equivalent to testing whether the *spectral coherence matrix* (see for instance [13, Chapter 7-6], [52, Chapter 5.5])

$$\mathbf{C}_{\mathbf{y}}(\nu) = \mathrm{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} \mathbf{S}_{\mathbf{y}}(\nu) \mathrm{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} \tag{3.4}$$

is equal to $\mathbf{I}_M$ for all frequencies $\nu \in [0,1]$. In this approach, usual test statistics are mostly based on consistent sample estimates of $\mathbf{S}_{\mathbf{y}}(\nu)$ or $\mathbf{C}_{\mathbf{y}}(\nu)$ that are compared to a diagonal matrix or to the identity $\mathbf{I}_M$ respectively. Previous works that developed this approach include [93], [88], [24], [25]. In particular, [93] considered the following frequency smoothed-periodogram estimator $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ defined by

$$\hat{\mathbf{S}}_{\mathbf{y}}(\nu) = \frac{1}{B+1} \sum_{b=-\frac{B}{2}}^{\frac{B}{2}} \boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right)^* \tag{3.5}$$

with $\boldsymbol{\xi}_{\mathbf{y}}(\nu) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{y}_n e^{-2i\pi n\nu}$ the renormalized finite Fourier transform of $(\mathbf{y}_n)_{n=0,\dots,N-1}$, B the *smoothing span*, assumed to be an even number, and where $\boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right)^*$ is the conjugate transpose of the vector $\boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right)$. [93] was devoted to the study of the limiting distribution of

$$\log\left\{\prod_{i=1}^{P} \det\,(\hat{\mathbf{S}}_{\mathbf{y}}(\nu_i)) / \prod_{m=1}^{M} \hat{s}_{m,m}(\nu_i)\right\}$$

for some properly defined subset of frequencies $(\nu_i)_{i=1,\dots,P}$, where $\hat{s}_{m,m}(\nu) = \left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)_{m,m}$. When $M = 2$, [24] considered a general kernel estimator of $\mathbf{S}_{\mathbf{y}}(\nu)$:

$$\tilde{\mathbf{S}}_{\mathbf{y}}(\nu) = \sum_{b=-\frac{N}{2}}^{\frac{N}{2}} w_N\left(\frac{b}{N}\right) \boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{\mathbf{y}}\left(\nu + \frac{b}{N}\right)^*$$

where w_N is a weight function satisfying some specific properties and a test statistic of the form

$$\frac{1}{N} \sum_{n=1}^{N} \frac{|(\tilde{\mathbf{S}}_{\mathbf{y}})_{12}(\nu)|^2}{(\tilde{\mathbf{S}}_{\mathbf{y}})_{11}(\nu)(\tilde{\mathbf{S}}_{\mathbf{y}})_{22}(\nu)}$$

which is proven to be, after proper recentring and renormalization, asymptotically normally distributed. Finally, [88] and [25] considered the more general class of test statistics, defined by:

$$\int_{-1/2}^{1/2} K\left((\tilde{\mathbf{S}}_{\mathbf{y}})_{12}(\nu)\right) d\nu \text{ and } \int_{-1/2}^{1/2} \left\|\psi\left((\tilde{\mathbf{S}}_{\mathbf{y}})_{12}(\nu), \nu\right)\right\|^2 d\nu$$

for some well-defined functions K and ψ, and where $\|\cdot\|$ is the Euclidian norm. They proved that these quantities asymptotically follow normal distributions. In the present chapter, we focus on the natural estimator (see e.g. [13, Chapter 7-6], [52, Chapter 8-4]) of $\mathbf{C}_{\mathbf{y}}$, defined by

$$\hat{\mathbf{C}}_{\mathbf{y}}(\nu) = \mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} \hat{\mathbf{S}}_{\mathbf{y}}(\nu)\mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} \tag{3.6}$$

where $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ is the frequency-smoothed periodogram estimate defined by (3.5). Note that adding a weight to the matrices $\boldsymbol{\xi}_{\mathbf{y}}(\nu + \frac{b}{N})\boldsymbol{\xi}_{\mathbf{y}}(\nu + \frac{b}{N})^*$ leads to a more general class of estimators of $\mathbf{S}_{\mathbf{y}}(\nu)$. The study of this more general class of estimators involves different techniques and random matrix models than the ones used here, and is therefore out of the scope of this chapter.

3.1.1 Low vs High-dimensional regime

The performance of the test statistics developed in the above mentioned previous works is usually studied in the *low-dimensional regime* where $N \to \infty$ and M is fixed. It is well known (see for instance [13]) that $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ and $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ are consistent estimates if $B \to +\infty$ and $\frac{B}{N} \to 0$. Under mild assumptions on the memory of the time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$, $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ is a consistent and asymptotically normal estimate of $\mathbf{C}_{\mathbf{y}}(\nu)$, which can in turn be used to study the asymptotic performance of the various tests based on $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$. In practice, the above asymptotic regime allows to predict the actual performance of the tests quite accurately, provided the ratio $\frac{M}{N}$ is small enough. If this condition is not met, test statistics based on $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ may be of delicate use, as the choice of the smoothing span B must meet the constraints $\frac{B}{M}$ much larger than 1 (because B is supposed to converge towards $+\infty$) as well as $\frac{B}{N}$ small enough (because $\frac{B}{N}$ is supposed to converge towards 0).

Nowadays, in many practical applications involving high-dimensional signals and/or moderate sample size, the ratio $\frac{M}{N}$ may not be small enough to be able to choose B to meet $\frac{B}{M}$ much larger than 1 and $\frac{B}{N}$ small enough. Therefore, the results obtained in the low-dimensional regime may fail to provide accurate predictions of the behaviour of the aforementioned test statistics. In this situation, one may rely on the more relevant *high-dimensional regime* in which M, B, N converge to infinity such that $\frac{M}{B}$ converges to a positive constant while $\frac{B}{N}$ converges to zero.

In comparison to the low-dimensional regime, the literature concerning correlation tests for the frequency domain in the high-dimensional regime is quite scarce. Recent results obtained in [62] show that under hypothesis $\mathcal{H}_0$, the empirical eigenvalue distribution of the spectral coherence estimate $\hat{\mathbf{C}}(\nu)$ behaves in the high-dimensional regime as the well-known Marcenko-Pastur distribution [63]. The result of [62] allows predicting the performance under $\mathcal{H}_0$ of a large class of test statistics based on

$$L_\varphi(\nu) = \frac{1}{M}\sum_{m=1}^{M} \varphi\left(\lambda_m(\hat{\mathbf{C}}_{\mathbf{y}}(\nu))\right)$$

where $\lambda_1(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)), \ldots, \lambda_M(\hat{\mathbf{C}}_{\mathbf{y}}(\nu))$ are the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, and φ belongs to a certain functional class. Such family of statistics L_φ, called *linear spectral statistics* (LSS) of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, include in particular the choice $\varphi(x) = \log x$, i.e. $L_\varphi(\nu) = \frac{1}{M}\log\det\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ and the choice $\varphi(x) = (x-1)^2$, i.e. $L_\varphi(\nu) = \frac{1}{M}\|\hat{\mathbf{C}}_{\mathbf{y}}(\nu) - \mathbf{I}_M\|_F^2$, where $\|\cdot\|_F$ represents the Frobenius norm.

In this chapter, we consider the study of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ in the high-dimensional regime under the special alternative $\mathcal{H}_1$ for which the useful signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ is modelled as the output of a stable MIMO filter driven by a K–dimensional white complex Gaussian noise. In the context where the intrinsic dimension K is fixed while $M, N, B \to \infty$, it is shown that the empirical eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ still converges to the Marcenko-Pastur distribution, showing that the test statistic based on $L_\varphi(\nu)$ is unable to discriminate between hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in the high-dimensional regime. Nevertheless, we also prove that, provided that the signal-to-noise ratio is large enough, the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ asymptotically splits from the support of the Marcenko-Pastur distribution. We can therefore exploit this result to design a new frequency domain test statistic, which is shown to be consistent in the high-dimensional regime. This result is connected to the widely studied *spiked models* in Random Matrix Theory, defined as low-rank perturbations of large random matrices. These models were extensively studied in the context of sample covariance matrices of independent identically distributed high-dimensional vectors, see e.g. [7]. We however notice that papers addressing the behaviour of the corresponding sample correlation matrices are quite scarce, see [65] when the low-rank perturbation affects only the first components of the observations.

3.1.2 Related works

Although the asymptotic framework differs from the high-dimensional regime considered here, we also mention the series of studies [32, 31] in the econometrics field, which consider a similar model under $\mathcal{H}_1$. In these works, it is assumed that $M, N \to \infty$ so the ratio $\frac{M}{N}$ remains bounded, while the K non-zero eigenvalues of the spectral density $\mathbf{S}_{\mathbf{u}}(\nu)$ of $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ are assumed to converge towards $+\infty$ at rate M. This last assumption, which ensures that the Signal-to-Noise Ratio (SNR) $\frac{\mathbb{E}\|\mathbf{u}_n\|^2}{\mathbb{E}\|\mathbf{v}_n\|^2}$ remains bounded away from 0 as $M \to \infty$, significantly facilitates the design of consistent detection methods. Nevertheless, while relevant in the domain of econometrics, this assumption may be unrealistic in several applications of array processing, where the challenge is to manage situations in which the SNR converges towards 0 at rate $\frac{1}{M}$. This situation is the one considered in this chapter and, in that case, the results of [32, 31] cannot be used. We discuss this point further in Section 3.2 below.

The rest of the chapter is organized as follows. In Section 3.2, we formally introduce the model of signals used, as well as the required technical assumptions. In section 3.3, we introduce informally the proposed test statistic and illustrate its behaviour to provide some intuition before a more rigorous presentation. In section 3.4, we study some approximation results for the spectral coherence $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ which are useful to study the linear spectral statistics considered here. This study is then used in Section 3.5 to introduce a new test statistic that is consistent in the high-dimensional regime. Finally, Section 3.6 provides some simulations illustrating its performance and comparisons against other relevant approaches.

Notations. For a complex matrix $\mathbf{A}$, we denote by $\mathbf{A}^*$ its conjugate transpose, and by $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$ its spectral and Frobenius norms respectively. If $\mathbf{A}$ is a $n \times n$ complex matrix, we denote by $\mathrm{tr}\,(\mathbf{A})$ its trace, and by $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ its eigenvalues; if moreover $\mathbf{A}$ is Hermitian, they are sorted in decreasing order $\lambda_1(\mathbf{A}) \geq \ldots \geq \lambda_n(\mathbf{A})$. The $n \times n$ identity matrix is denoted as $\mathbf{I}_n$. The expectation of a complex random variable Z is denoted by $\mathbb{E}[Z]$. The complex circular Gaussian distribution with variance σ^2 is denoted as $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ and a random vector $\mathbf{x}$ of $\mathbb{C}^n$ follows the $\mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{R})$ distribution if $\mathbf{b}^*\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{b}^*\mathbf{R}\mathbf{b})$ for all deterministic (column) vector $\mathbf{b}$ and a fixed $n \times n$ positive definite matrix $\mathbf{R}$. Finally, $\mathcal{C}^1(I)$ (resp. $\mathcal{C}^1_c(I)$) represents the set of continuously differentiable functions (resp. continuously differentiable functions with compact support) on an open set I.

3.2 Model and assumptions

Let us consider an M–dimensional observed time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ defined as

$$\mathbf{y}_n = \mathbf{u}_n + \mathbf{v}_n \tag{3.7}$$

where $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ represents a useful signal and where $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ represents an additive noise. The useful signal is modeled as the output of an unknown stable MIMO filter $(\mathbf{H}_k)_{k\in\mathbb{Z}}$ driven by a non-observable K–dimensional complex Gaussian white noise $(\boldsymbol{\epsilon}_n)_{n\in\mathbb{Z}}$ with $\mathbb{E}[\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^*] = \mathbf{I}_K$, i.e.

$$\mathbf{u}_n = \sum_{k\in\mathbb{Z}} \mathbf{H}_k \boldsymbol{\epsilon}_{n-k}$$

with probability one. We notice that K represents the number of sources in the context of array processing. $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is modeled as an M–dimensional stationary complex Gaussian time series such that its component time series $(v_{1,n})_{n\in\mathbb{Z}}, \ldots, (v_{M,n})_{n\in\mathbb{Z}}$ are mutually independent.

For each $m = 1, \ldots, M$, we denote by $(r_m(k))_{k\in\mathbb{Z}}$ the covariance function of $(v_{m,n})_{n\in\mathbb{Z}}$, i.e. $r_m(k) = \mathbb{E}[v_{m,n}\overline{v_{m,n-k}}]$, which satifies the following memory assumption.

Assumption 3.1. *The covariance coefficients decay sufficiently fast in the lag domain, in the sense that*

$$\sup_{m\geq 1} \sum_{k\in\mathbb{Z}} (1+|k|)^2 |r_m(k)| < \infty. \tag{3.8}$$

In particular, Assumption 3.1 implies that the spectral density s_m of $(v_{m,n})_{n\in\mathbb{Z}}$, given by

$$s_m(\nu) = \sum_{k\in\mathbb{Z}} r_m(k) \mathrm{e}^{-\mathrm{i}2\pi\nu k}$$

satisfies

$$\sup_{m\geq 1} \sup_{\nu\in[0,1]} s_m(\nu) < \infty.$$

Assumption 3.1 is in particular satisfied as soon as the condition

$$|r_m(k)| \leq \frac{C}{(1+|k|)^{3+\delta}} \tag{3.9}$$

holds for each $k \in \mathbb{Z}$ and each $m \geq 1$, where C and δ are positive constants. As the autocovariance function of ARMA signals decreases exponentially towards 0, Assumption 3.1 thus holds if the time series $(v_m)_{m\geq 1}$ are ARMA signals, provided some extra purely technical conditions that allow to manage the supremum over m in (3.8) are met. As the spectral coherence matrix of $(\mathbf{v}_n)_{n\in\mathbb{Z}}$, involves a renormalization by the inverse of the spectral densities s_m, we also need that s_m does not vanish for each m.

Assumption 3.2. *The spectral densities are uniformly bounded away from zero, that is*

$$\inf_{m\geq 1} \min_{\nu\in[0,1]} s_m(\nu) > 0.$$

Assumptions 3.1 and 3.2 also imply that the total noise power satisfies

$$0 < \inf_{M\geq 1} \frac{1}{M}\mathbb{E}\|\mathbf{v}_n\|_2^2 \leq \sup_{M\geq 1} \frac{1}{M}\mathbb{E}\|\mathbf{v}_n\|_2^2 < \infty. \tag{3.10}$$

The next assumption is related to the signal part $(\mathbf{u}_n)_{n\in\mathbb{Z}}$. For each $\nu \in [0,1]$, we denote by $\mathbf{H}(\nu)$ the Fourier transform of $(\mathbf{H}_k)_{k\in\mathbb{Z}}$, i.e.

$$\mathbf{H}(\nu) = \sum_{k\in\mathbb{Z}} \mathbf{H}_k \mathrm{e}^{-\mathrm{i}2\pi\nu k}$$

and by $\mathbf{h}^1(\nu), \ldots, \mathbf{h}^M(\nu)$ the rows of $\mathbf{H}(\nu)$.

Assumption 3.3. *The MIMO filter coefficient matrices are such that*

$$\sup_{M\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)\left\|\mathbf{H}_k\right\|_2 < \infty \tag{3.11}$$

and

$$\lim_{M\to\infty}\max_{m=1,\ldots,M}\max_{\nu\in[0,1]}\left\|\mathbf{h}^m(\nu)\right\|_2 = 0. \tag{3.12}$$

When K is fixed while $M\to\infty$, condition (3.11) in Assumption 3.3 implies that the total useful signal power remains bounded, i.e.

$$\mathbb{E}\left\|\mathbf{u}_n\right\|_2^2 = \sum_{k\in\mathbb{Z}}\left\|\mathbf{H}_k\right\|_F^2 = \mathcal{O}(1) \tag{3.13}$$

so that, using (3.10), the SNR vanishes at rate $\frac{1}{M}$, i.e.

$$\frac{\mathbb{E}\|\mathbf{u}_n\|_2^2}{\mathbb{E}\|\mathbf{v}_n\|_2^2} = \mathcal{O}\left(\frac{1}{M}\right). \tag{3.14}$$

Likewise, condition (3.12) in Assumption 3.3 implies that the SNR per time series vanishes, i.e.

$$\frac{\mathbb{E}|u_{m,n}|^2}{\mathbb{E}|v_{m,n}|^2} = \frac{\int_0^1\|\mathbf{h}^m(\nu)\|_2^2\mathrm{d}\nu}{\int_0^1 s_m(\nu)\mathrm{d}\nu} = o(1) \tag{3.15}$$

as $M\to\infty$. We finally notice that (3.11) is stronger than (3.13). While $\mathbb{E}\left\|\mathbf{u}_n\right\|_2^2 = \mathcal{O}(1)$ is a rather fundamental assumption that allows to control the behaviour of the signal to noise ratio, the extra condition $\sup_{m\geq 1}\sum_k |k|\left\|\mathbf{H}_k\right\|_2 < \infty$ is essentially motivated by technical reasons (it is needed to establish Theorem 3.2). However, it is clearly not restrictive in practice.

Remark 3.1. *Conditions (3.11) and (3.12) in Assumption 3.3 are especially relevant in the context of array processing, where M represents the number of sensors, which may be large [90, 91]. In this context, (3.14) represents the SNR before matched filtering, while (3.15) represents the SNR per sensor. The use of spatial filtering techniques, which combine the observations $y_{1,n},\ldots,y_{M,n}$ across the M sensors, allows to increase the SNR by a factor M when the second order statistics of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ are known, which leads to an SNR after matched filtering of the order of magnitude $\mathcal{O}(1)$. Thus, despite the apparent low SNR, reliable information on the useful signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ can potentially still be extracted from the observed signal $(\mathbf{y}_n)_{n\in\mathbb{Z}}$.*

Let $\mathbf{S}_{\mathbf{y}}$ denote the spectral density of $(\mathbf{y}_n)_{n\in\mathbb{Z}}$, given by

$$\mathbf{S}_{\mathbf{y}}(\nu) = \mathbf{H}(\nu)\mathbf{H}(\nu)^* + \mathbf{S}_{\mathbf{v}}(\nu)$$

where $\mathbf{S}_{\mathbf{v}}(\nu) = \mathrm{dg}\left(s_1(\nu),\ldots,s_M(\nu)\right)$. To estimate $\mathbf{S}_{\mathbf{y}}$, we consider in this chapter a frequency-smoothed periodogram $\hat{\mathbf{S}}_{\mathbf{y}}$, which we defined in (3.5). In the classical low-dimensional regime where $B,N\to\infty$ while M,K remain fixed, it is well-known [13] that

$$\mathbb{E}[\hat{\mathbf{S}}_{\mathbf{y}}(\nu)] = \mathbf{S}_{\mathbf{y}}(\nu) + \mathcal{O}\left(\frac{B^2}{N^2}\right)$$

and

$$\mathbb{E}\left\|\hat{\mathbf{S}}_{\mathbf{y}}(\nu) - \mathbb{E}[\hat{\mathbf{S}}_{\mathbf{y}}(\nu)]\right\|_2^2 = \mathcal{O}\left(\frac{1}{B}\right).$$

Thus, in this regime, $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ is a consistent estimator of $\mathbf{S}_{\mathbf{y}}(\nu)$ as long as $B \to \infty$ and $\frac{B}{N} \to 0$. Likewise, the sample Spectral Coherence Matrix (SCM, not to be confused with the sample covariance matrix, which will not be used in this chapter) defined in (3.6) is a consistent estimator of the true SCM $\mathbf{C}_{\mathbf{y}}(\nu)$ defined in (3.4). When $M \to +\infty$ and $\frac{M}{N} \to 0$, it can be shown that, under some additional mild extra assumptions, the consistency of $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ and $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ in the spectral norm sense still holds provided that B is chosen in such a way that $\frac{B}{N} \to 0$ and $\frac{M}{B} \to 0$. In practice, for finite values of M and N, the above asymptotic regime will allow to predict the performance of various inference schemes in situations where it is possible to choose B in such a way that $\frac{M}{B}$ and $\frac{B}{N}$ are both small enough. Nevertheless, when the dimension M is large and the sample size N is not unlimited, or equivalently if $\frac{M}{N}$ is not small enough, such a choice of B may be impossible. In such a context, it seems more relevant to consider asymptotic regimes for which $\frac{M}{N} \to 0$ and $\frac{M}{B}$ converging towards a positive constant. Below, we will consider the following asymptotic regime.

Assumption 3.4. *$N = N(M)$ and $B = B(M)$ are both functions of M such that, for some $\alpha \in (0,1)$,*

$$M = \mathcal{O}\left(N^{\alpha}\right) \text{ and } \frac{M}{B} \xrightarrow[M \to \infty]{} c \in (0,1)$$

while K is fixed with respect to M.

As $\frac{M}{B}$ does not converge towards 0, the consistency of $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ and $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ is lost. This can be explained in a simple way when $\mathbf{u}_n = 0$ for each n and the signals $((v_{m,n})_{n\in\mathbb{Z}})_{m\geq 1}$ are mutually independent i.i.d. $\mathcal{N}_c(0,\sigma^2)$ distributed sequences. In this context, for each ν, the renormalized Fourier transform vectors $(\boldsymbol{\xi}_{\mathbf{y}}(\nu + b/N))_{b=-B/2,\ldots,B/2}$ are mutually independent $\mathcal{N}_{\mathbb{C}}(0,\sigma^2\mathbf{I})$ random vectors. The spectral density estimate $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ defined by (3.5) thus coincides with the sample covariance matrix of these $(B+1)$ M–dimensional vectors. If B and M are of the same order of magnitude, it cannot be expected that $\|\hat{\mathbf{S}}_{\mathbf{y}}(\nu) - \mathbb{E}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))\|$ converges towards 0 because the true covariance matrix $\mathbb{E}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))$ to be estimated depends on $\mathcal{O}(M^2)$ parameters, and that the number MB of available scalar observations used to estimate $\mathbb{E}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))$ is also $\mathcal{O}(M^2)$. Despite the loss of the convergence of the estimators $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$ and $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, we will see that one can still rely on the high-dimensional structure of these matrices to design relevant test statistics.

3.3 Informal presentation of the proposed test statistic

Mathematical details will reveal later that for each ν, $\hat{\mathbf{C}}(\nu)$ behaves as a spike model covariance matrix, whose eigenvalues are precisely described by [7]. More precisely, we will see that, in some sense, the eigenvalues of $\hat{\mathbf{C}}(\nu)$ that are due to the noise belong to the interval $[\lambda_-, \lambda_+]$ where $\lambda_- = (1-\sqrt{c})^2$ and $\lambda_+ = (1+\sqrt{c})^2$, and that in the presence of signal, some eigenvalues of $\hat{\mathbf{C}}(\nu)$ may be strictly greater than λ_+ if an SNR criteria is respected. For the remainder, we define

$$\mathcal{V}_N = \left\{\frac{k}{N} : k = 0, \ldots, N-1\right\} \tag{3.16}$$

as the set of Fourier frequencies. A natural way to test for $\mathcal{H}_0$ against $\mathcal{H}_1$ is to compute the largest eigenvalue of $\hat{\mathbf{C}}(\nu)$ over the frequencies of $\mathcal{V}_N$, and compare it with λ_+. This leads to the following test statistic:

$$T_\epsilon = \mathbb{1}_{[\lambda^+ + \epsilon, \infty)}\left(\max_{\nu \in \mathcal{V}_N} \lambda_1\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)\right). \tag{3.17}$$

We will prove later that, under proper assumption on the SNR, this test statistic is consistent in the present high-dimensional regime. Before describing the mathematical details leading to consider T_ϵ, we now provide some numerical illustrations of its behaviour. The general settings are as follows. The noise is generated as a Gaussian AR(1) process having spectral density

$$s_m(\nu) = \frac{1}{|1 - \theta \mathrm{e}^{-\mathrm{i}2\pi\nu}|^2}, \tag{3.18}$$

for all $m = 1, \ldots, M$, with $\theta = 0.5$, whereas for the useful signal, we also consider an AR(1) process by choosing $K = 1$ and

$$\mathbf{H}_k = \sqrt{\frac{C}{M}} \beta^k (1, \ldots, 1)^T \tag{3.19}$$

with $\beta = \frac{10}{11}$ and C being a positive constant used to adjust the SNR.

To understand how the test statistics T_ϵ discriminates between $\mathcal{H}_0$ and $\mathcal{H}_1$, we show in Figure 3.3.1 the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for $\nu \in \mathcal{V}_N$ in the presence of signal, and compare it to the threshold λ_+. We see that for some frequencies ν around 0, the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ deviates significantly from λ_+. As we will see later, it is possible to evaluate the asymptotic behaviour of the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, and to establish that it converges towards $\phi(SNR(\nu))$ where ϕ is a certain function, and where $SNR(\nu)$ can be interpreted as a signal-to-noise ratio at frequency ν. $\phi(SNR(\nu))$ is also represented in Figure 3.3.1, and it is seen that it is close to the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$. In Figure 3.3.2, we compare the empirical distribution of T_ϵ under $\mathcal{H}_0$ and $\mathcal{H}_1$ over 10000 repetitions. We see that the distribution of our test statistic T_ϵ is able to discriminate the scenarios where the data $\mathbf{y}_n$ are generated under $\mathcal{H}_0$ or $\mathcal{H}_1$, and that T_ϵ is typically over the threshold λ_+ under $\mathcal{H}_1$.

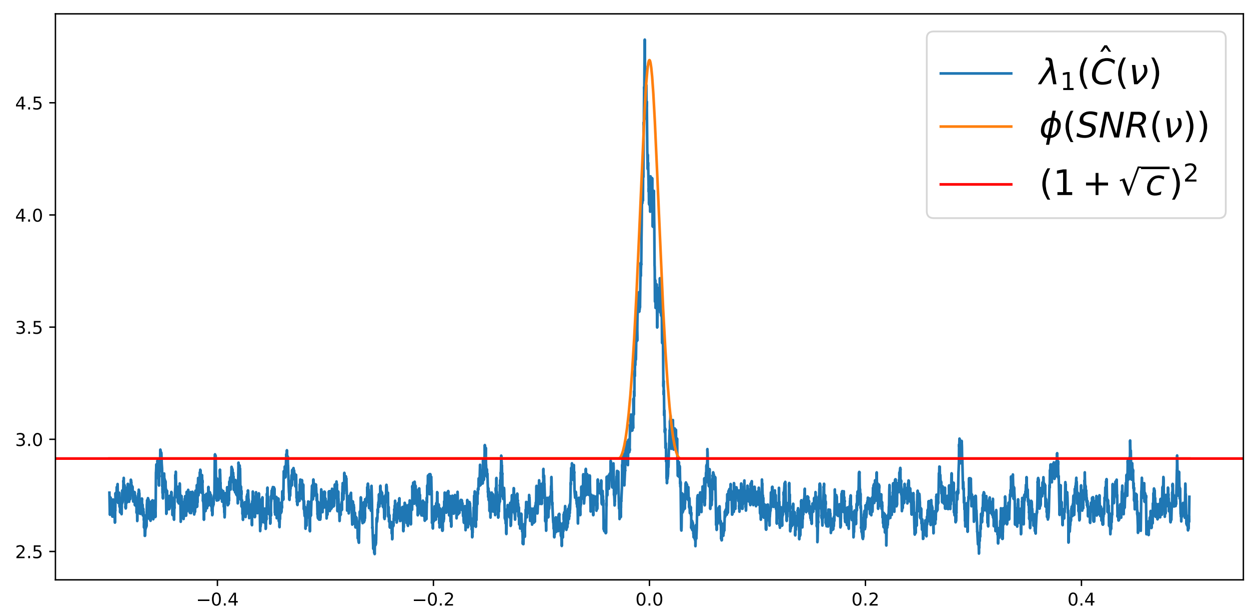


Figure 3.3.1: Largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for $\nu \in \mathcal{V}_N$ vs the threshold $\lambda_+ = (1 + \sqrt{\frac{M}{B+1}})^2$. $M = 60$, $c = 0.5$, $N = 6000$, $\theta = 0.5$, $C = 0.05$

3.4 Approximation results for $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ in the high-dimensional regime

In this section, we present the mathematical details which lead to the test statistic (3.17). More specifically, we provide useful approximation results for $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, which show that $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ behaves

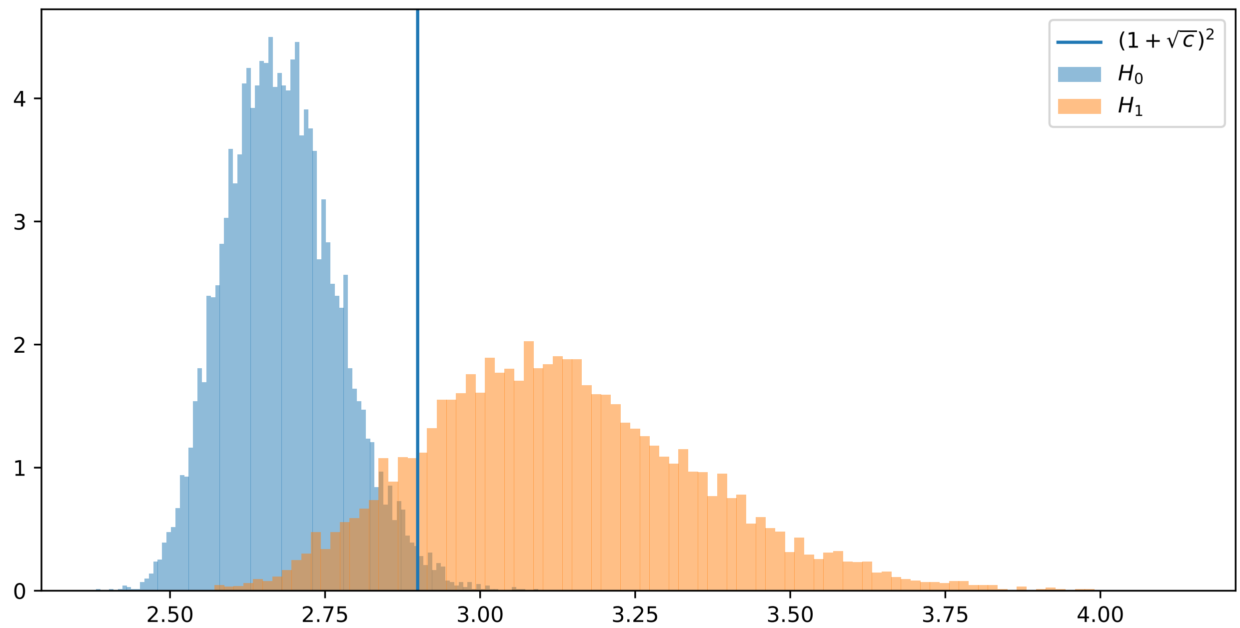


Figure 3.3.2: Histogram of T_ϵ under $\mathcal{H}_0$ and $\mathcal{H}_1$, over 10000 repetitions. $M = 40$, $c = 0.5$, $N = 1000$, $\theta = 0.5$, $C = 0.05$

as a certain Wishart matrix in the high-dimensional regime. These approximation results are the keystone for the study of the behaviour of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ and the detection test proposed in Section 3.5.

We first study separately the signal-free case (i.e. $\mathbf{y}_n = \mathbf{v}_n$) as well as the noise-free case (i.e. $\mathbf{y}_n = \mathbf{u}_n$).

3.4.1 Signal-free case

Let

$$\boldsymbol{\xi}_{\mathbf{v}}(\nu) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{v}_n \mathrm{e}^{-\mathrm{i}2\pi\nu n}$$

denote the discrete (time-limited) Fourier transform of $(\mathbf{v}_n)_{n=0,\ldots,N-1}$, and define the $M \times (B+1)$ matrix $\boldsymbol{\Sigma}_{\mathbf{v}}(\nu)$ as

$$\boldsymbol{\Sigma}_{\mathbf{v}}(\nu) = \frac{1}{\sqrt{B+1}} \left[\boldsymbol{\xi}_{\mathbf{v}} \left(\nu - \frac{B}{2N} \right), \ldots, \boldsymbol{\xi}_{\mathbf{v}} \left(\nu + \frac{B}{2N} \right) \right].$$

The following result, derived in [62], reveals an interesting behaviour of the frequency-smoothed periodogram of the noise.

Theorem 3.1. *Under Assumptions 3.1, 3.2 and 3.4, for all $\nu \in \mathcal{V}_N$, there exists an $M \times (B+1)$ matrix $\mathbf{Z}(\nu)$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries such that*

$$\max_{\nu \in \mathcal{V}_N} \left\| \boldsymbol{\Sigma}_{\mathbf{v}}(\nu) - \frac{1}{\sqrt{B+1}} \mathbf{S}_{\mathbf{v}}(\nu)^{1/2} \mathbf{Z}(\nu) \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

Informally speaking, Theorem 3.1 shows that the random vectors $\frac{1}{\sqrt{B+1}}\boldsymbol{\xi}_{\mathbf{v}}\left(\nu - \frac{B}{N}\right), \ldots, \frac{1}{\sqrt{B+1}}\boldsymbol{\xi}_{\mathbf{v}}\left(\nu + \frac{B}{N}\right)$ asymptotically behave as a family of i.i.d. $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{S}_{\mathbf{v}}(\nu))$

vectors, for all $\nu \in \mathcal{V}_N$. Moreover, if

$$\hat{\mathbf{S}}_{\mathbf{v}}(\nu) := \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_{\mathbf{v}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{\mathbf{v}}\left(\nu + \frac{b}{N}\right)^*$$

denotes the frequency-smoothed periodogram of the noise observations $(\mathbf{v}_n)_{n\in\mathbb{Z}}$, we deduce that $\hat{\mathbf{S}}_{\mathbf{v}}(\nu)$ asymptotically behaves as a complex Gaussian Wishart matrix with covariance matrix $\mathbf{S}_{\mathbf{v}}(\nu)$, thanks to the following corollary.

Corollary 3.1. *Under the assumptions of Theorem 3.1, it holds that*

$$\max_{\nu\in\mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{v}}(\nu) - \mathbf{S}_{\mathbf{v}}(\nu)^{1/2} \frac{\mathbf{Z}(\nu)\mathbf{Z}(\nu)^*}{B+1} \mathbf{S}_{\mathbf{v}}(\nu)^{1/2} \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

Proof. The proof is deferred to Appendix 3.D.1. □

It is worth noticing that Corollary 3.1 implies in particular

$$\max_{\nu\in\mathcal{V}_N} \left\| \mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{v}}(\nu)\right) - \mathbf{S}_{\mathbf{v}}(\nu) \right\| \xrightarrow[M\to\infty]{a.s.} 0$$

and consequently $\mathrm{dg}(\hat{\mathbf{S}}(\nu))$ is a consistent estimator of the noise spectral density $\mathbf{S}_{\mathbf{v}}(\nu)$ in the operator norm sense, at each Fourier frequency $\nu \in \mathcal{V}_N$. This convergence may be directly obtained using Lemma 3.1 in Appendix 3.A and we omit the details since this result is well-known.

3.4.2 Noise-free case

Let

$$\boldsymbol{\xi}_{\mathbf{u}}(\nu) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{u}_n \mathrm{e}^{-\mathrm{i}2\pi\nu n}$$

and let $\boldsymbol{\Sigma}_{\mathbf{u}}(\nu)$ be the $K \times (B+1)$ matrix defined as

$$\boldsymbol{\Sigma}_{\mathbf{u}}(\nu) = \frac{1}{\sqrt{B+1}} \left[\boldsymbol{\xi}_{\mathbf{u}}\left(\nu - \frac{B}{2N}\right), \ldots, \boldsymbol{\xi}_{\mathbf{u}}\left(\nu + \frac{B}{2N}\right) \right].$$

In the same way, we also denote by $\boldsymbol{\xi}_{\boldsymbol{\epsilon}}$ the normalized discrete (time-limited) Fourier transform of $(\boldsymbol{\epsilon}_n)_{n=0,\ldots,N-1}$, and consider the $K \times (B+1)$ matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)$ defined as $\boldsymbol{\Sigma}_{\mathbf{u}}(\nu)$. We then have the following important approximation result.

Theorem 3.2. *Under Assumptions 3.3 and 3.4, it holds that*

$$\max_{\nu\in\mathcal{V}_N} \|\boldsymbol{\Sigma}_{\mathbf{u}}(\nu) - \mathbf{H}(\nu)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

Proof. The proof is deferred to Appendix 3.B. □

As in Theorem 3.1, Theorem 3.2 shows that the random vectors $\boldsymbol{\xi}_{\mathbf{u}}\left(\nu - \frac{B}{N}\right), \ldots, \boldsymbol{\xi}_{\mathbf{u}}\left(\nu + \frac{B}{N}\right)$ asymptotically behave as the i.i.d. vectors $\mathbf{H}(\nu)\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\left(\nu - \frac{B}{N}\right), \ldots, \mathbf{H}(\nu)\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\left(\nu + \frac{B}{N}\right)$, for all $\nu \in \mathcal{V}_N$.

Remark 3.2. *The type of approximation given in Theorem 3.2 is well-known in the low-dimensional regime in which M, K, B are fixed while $N \to \infty$. Indeed, in that case, we have [13, Th. 4.5.2]*

$$\max_{\nu\in[0,1]} \|\boldsymbol{\Sigma}_{\mathbf{u}}(\nu) - \mathbf{H}(\nu)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\|_2 = \mathcal{O}_P\left(\sqrt{\frac{\log(N)}{N}}\right).$$

In the high-dimensional regime where M and B also converge to infinity as described in Assumption 3.4, the result of Theorem 3.2 cannot be obtained from [13, Th. 4.5.2] and thus requires a new study.

We also deduce the following approximation result on the frequency-smoothed periodogram of the signal observations $(\mathbf{u}_n)_{n=0,\dots,N-1}$ given by

$$\hat{\mathbf{S}}_{\mathbf{u}}(\nu) := \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \boldsymbol{\xi}_{\mathbf{u}}\left(\nu + \frac{b}{N}\right) \boldsymbol{\xi}_{\mathbf{u}}\left(\nu + \frac{b}{N}\right)^*.$$

Corollary 3.2. *Under the assumptions of Theorem 3.2, it holds that*

$$\max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{u}}(\nu) - \mathbf{H}(\nu)\mathbf{H}(\nu)^* \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

Proof. The proof is deferred to Appendix 3.D.2. □

As a result of Corollary 3.2, we deduce that the frequency-smoothed periodogram $\hat{\mathbf{S}}_{\mathbf{u}}(\nu)$ is a consistent estimator of the spectral density $\mathbf{S}_{\mathbf{u}}(\nu) = \mathbf{H}(\nu)\mathbf{H}(\nu)^*$ of $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ in the high-dimensional regime, for each $\nu \in \mathcal{V}_N$.

Having characterized the pure noise and pure signal cases, we are now in a position to study the high-dimensional behaviour of the spectral coherence matrix $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$.

3.4.3 The signal-plus-noise case

First, using Corollaries 3.1 and 3.2, we deduce the high-dimensional behaviour of the frequency smoothed periodogram $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$. The following results show that, as it could be expected, the frequency smoothed periodogram essentially behaves as a colored Wishart matrix in the large asymptotic regime.

Proposition 3.1. *For all $\nu \in \mathcal{V}_N$, there exists an $M \times (B+1)$ matrix $\mathbf{X}(\nu)$ with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries such that*

$$\max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{y}}(\nu) - \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0. \tag{3.20}$$

Proof. The proof is deferred to Appendix 3.D.3. □

We finally consider the study of the spectral coherence $\hat{\mathbf{C}}_{\mathbf{y}}(\nu) = \mathrm{dg}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))^{-\frac{1}{2}} \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \mathrm{dg}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)^{-\frac{1}{2}}$. From condition (3.12) in Assumption 3.3 on the SNR, it turns out that (cf. proof of Theorem 3.3 below where the result is shown) that

$$\max_{\nu \in \mathcal{V}_N} \left\| \mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right) - \mathbf{S}_{\mathbf{v}}(\nu) \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0. \tag{3.21}$$

This approximation result regarding the normalization term $\mathrm{dg}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))$ in the SCM naturally leads to the following theorem, which is the key result of this chapter.

Theorem 3.3. *Under Assumptions 3.1, 3.2, 3.3 and 3.4,*

$$\max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{C}}_{\mathbf{y}}(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \right\|_2 \xrightarrow[M\to\infty]{a.s.} 0$$

where

$$\boldsymbol{\Xi}(\nu) = \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \mathbf{H}(\nu)\mathbf{H}(\nu)^* \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} + \mathbf{I}_M.$$

and $\mathbf{X}(\nu)$ is the matrix defined in Proposition 3.1.

Proof. The proof is deferred to Appendix 3.C. □

Let us make a few important comments regarding the result of Theorem 3.3.

First, used in conjunction with Weyl's inequalities [43, Th. 4.3.1], Theorem 3.3 implies in particular that each eigenvalue of the SCM $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ behaves as its counterpart of the Wishart matrix

$$\mathbf{W}(\nu) = \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}}$$

for $\nu \in \mathcal{V}_N$, that is

$$\max_{m=1,\ldots,M} \max_{\nu \in \mathcal{V}_N} \left| \lambda_m \left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu) \right) - \lambda_m \left(\mathbf{W}(\nu) \right) \right| \xrightarrow[M\to\infty]{a.s.} 0. \tag{3.22}$$

Second, Theorem 3.3 has an important consequence regarding the behaviour of linear spectral statistics of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, that is statistics of the type

$$L_{\varphi}(\nu) = \frac{1}{M} \sum_{m=1}^{M} \varphi \left(\lambda_m \left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu) \right) \right) \tag{3.23}$$

where φ belongs to a certain class of functions.

Corollary 3.3. *Let $\varphi \in \mathcal{C}^1\left((0,+\infty)\right)$. Under Assumptions 3.1, 3.2, 3.3 and 3.4, we have*

$$\max_{\nu \in \mathcal{V}_N} \left| L_{\varphi}(\nu) - \int_{\mathbb{R}} \varphi(\lambda) f(\lambda) \mathrm{d}\lambda \right| \xrightarrow[M\to\infty]{a.s.} 0$$

where f is the density of the Marcenko-Pastur distribution given by

$$f(\lambda) = \frac{\sqrt{(\lambda - \lambda^-)(\lambda^+ - \lambda)}}{2\pi c \lambda} \mathbb{1}_{[\lambda^-,\lambda^+]}(\lambda)$$

with $\lambda^{\pm} = (1 \pm \sqrt{c})^2$.

Proof. The proof is deferred to Appendix 3.E. □

Therefore, Corollary 3.3 shows that linear spectral statistics of the SCM converge to the same limit regardless of whether the observations contain only pure noise or signal-plus-noise contributions. This shows that any test statistic solely relying on linear spectral statistics of the SCM is unable to distinguish between the absence or presence of a useful signal, and cannot be consistent in the high-dimensional regime. Nevertheless, in the next section, we will see that we can exploit Theorem 3.3 to build a new test statistic based on the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, which is proved to be consistent in the high-dimensional regime.

Remark 3.3. *Corollary 3.1, Corollary 3.2 and Theorems 3.3 may be interpreted in the context of array processing. Indeed, in the time model (3.7), usually referred to as "wideband", the signal contribution $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ modeled as a linear process, is in general not confined to a low-dimensional subspace (i.e. with dimension less than M). However, in the frequency domain, Corollary 3.1 and Corollary 3.2 show that we can retrieve, in the high-dimensional regime, a "narrowband" model, since the useful signal is confined to a K–dimensional subspace of $\mathbb{C}^M$. Thus, standard narrowband techniques used in array processing for detection may be used, see e.g. [95].*

3.5 A new consistent test statistic

As we have seen in Theorem 3.3 and the related comments, the SCM $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ behaves in the high-dimensional regime as a Wishart matrix with scale $\boldsymbol{\Xi}(\nu) = \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} + \mathbf{I}_M$ being a fixed rank K perturbation of the identity matrix. The behaviour of the eigenvalues for each ν of such a matrix model is well-known since [7] (and other related works such as the well-known BBP-phase-transition [6] or [9]), and the rest of this section is devoted to the application of the results from [7] in our frequency-domain detection context. A crucial point is to choose the particular frequency at which the above mentioned results will be used to obtain information on the behaviour of $\max_{\nu \in \mathcal{V}_N} \lambda_1\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)$. For this, we have first to define some notations. We consider the fundamental function ϕ which already appears in [7]:

$$\phi(x) = \begin{cases} \frac{(x+1)(x+c)}{x} & \text{if } x > \sqrt{c} \\ \lambda^+ & \text{if } x \leq \sqrt{c} \end{cases}$$

where we recall that $\lambda^+ = (1+\sqrt{c})^2$ (see Corollary 3.3). We notice that for all $x > \sqrt{c}$, $\phi(x) > \phi(\sqrt{c}) = \lambda^+$. Define as $\gamma(\nu)$ the maximum eigenvalue of the finite rank perturbation for each ν, that is

$$\gamma(\nu) = \lambda_1\left(\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\right) \tag{3.24}$$

and let $\nu_N^* \in \mathcal{V}_N$ such that

$$\nu_N^* \in \underset{\nu \in \mathcal{V}_N}{\operatorname{argmax}}\, \gamma(\nu).$$

We note that $\gamma(\nu_N^*)$ may be interpreted as a certain SNR metric in the frequency domain. In the following, we study the behaviour of the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*)$, which requires the following additional assumption on $\gamma(\nu_N^*)$.

Assumption 3.5. *There exists $\gamma_\infty \geq 0$ such that*

$$\gamma(\nu_N^*) \xrightarrow[M\to\infty]{} \gamma_\infty.$$

Theorem 3.3 implies that the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*)$ have the same asymptotic behaviour as the corresponding eigenvalues of the matrix $\boldsymbol{\Xi}(\nu_N^*)^{\frac{1}{2}}\frac{\mathbf{X}(\nu_N^*)\mathbf{X}(\nu_N^*)^*}{B+1}\boldsymbol{\Xi}(\nu_N^*)^{\frac{1}{2}}$. Under Assumption 3.5, [6], [7] or [9] immediately imply the following result. Note that since ν_N^* is unknown in practice, this proposition is an intermediate theoretical result that will justify the detection test statistic introduced below.

Proposition 3.2. *Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5, we have*

$$\lambda_1\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*)\right) \xrightarrow[M\to\infty]{a.s.} \phi(\gamma_\infty) \tag{3.25}$$

while

$$\lambda_{K+1}\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*)\right) \xrightarrow[M\to\infty]{a.s.} \lambda^+ \tag{3.26}$$

and

$$\lambda_M\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*)\right) \xrightarrow[M\to\infty]{a.s.} \lambda^-. \tag{3.27}$$

Moreover, if $\gamma_\infty = 0$,

$$\limsup_{M\to\infty} \max_{\nu \in \mathcal{V}_N} \lambda_1\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right) \leq \lambda^+ \quad a.s. \tag{3.28}$$

Proof. It just remains to establish (3.28), see Appendix 3.F. □

Since neither the intrinsic dimensionality K of the useful signal $(\mathbf{u}_n)_{n\in\mathbb{Z}}$ nor the frequency ν_N^* are known in practice, we use the largest eigenvalue of the SCM maximized over all Fourier frequencies as a test statistic. This leads to the test statistic T_ϵ defined previously in (3.17) which we recall here:

$$T_\epsilon = \mathbb{1}_{[\lambda^+ + \epsilon, \infty)} \left(\max_{\nu \in \mathcal{V}_N} \lambda_1 \left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu) \right) \right).$$

It turns out that this test statistic is consistent in the high-dimensional regime, as stated in the following result.

Proposition 3.3. *Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5, and if under Hypothesis $\mathcal{H}_1$,*

$$\gamma_\infty > \sqrt{c}$$

then for all $0 < \epsilon < \phi(\gamma_\infty) - \lambda^+$ and $i \in \{0, 1\}$,

$$\mathbb{P}_i \left(\lim_{M\to\infty} T_\epsilon = i \right) = 1$$

where $\mathbb{P}_i$ is the underlying probability measure under Hypothesis $\mathcal{H}_i$.

Proof. Under Hypothesis $\mathcal{H}_0$, since $\gamma_\infty = 0$, we directly apply (3.28) in Proposition 3.2 to obtain that for all $\epsilon > 0$, $T_\epsilon = 0$ with probability one, for all large M. Under Hypothesis $\mathcal{H}_1$, we get

$$\liminf_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1 \left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu) \right) \geq \lim_{M\to\infty} \lambda_1 \left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu_N^*) \right) = \phi(\gamma_\infty)$$

with probability one. Since by assumption, $\phi(\gamma_\infty) > \lambda^+ + \epsilon$, we deduce that $T_\epsilon = 1$ with probability one for all large M. $\square$

3.6 Simulations

In this section, we provide some numerical illustrations of the approximation results of Section 3.4. We will consider the case where the rank K of the signal is equal to one and then the case where K is strictly greater than one.

3.6.1 Case $K = 1$

As in the numerical simulation presented in Section 3.3, each component of the noise $\mathbf{v}_n$ is generated as a Gaussian AR(1) process with $\theta = 0.5$. The expression of its spectral density s_m for all $m = 1, \dots, M$ is still given in (3.18). The useful signal is generated as an AR(1) process with $K = 1$, $\mathbf{H}_k$ defined by (3.19) and $\beta = \frac{10}{11}$. C is again a positive constant used to tune the SNR. Note that, in this context, the SNR $\gamma(\nu)$ at frequency ν defined in (3.24) takes the form

$$\gamma(\nu) = C \left| \frac{1 - \theta \mathrm{e}^{-\mathrm{i}2\pi\nu}}{1 - \beta \mathrm{e}^{-\mathrm{i}2\pi\nu}} \right|^2.$$

Figures 3.6.1 and 3.6.2 illustrate the signal-free case $C = 0$, and where $(N, M, B) = (20000, 100, 200)$. In Figure 3.6.1, we plot the histogram of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for $\nu = 0$. As predicted by Corollary 3.3 in the signal-free case, the empirical eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ is well approximated by the Marcenko-Pastur distribution with shape parameter $c = 0.5 \approx M/(B+1)$. Figure 3.6.2 further illustrates this convergence, where the cumulative distribution function (cdf) of the

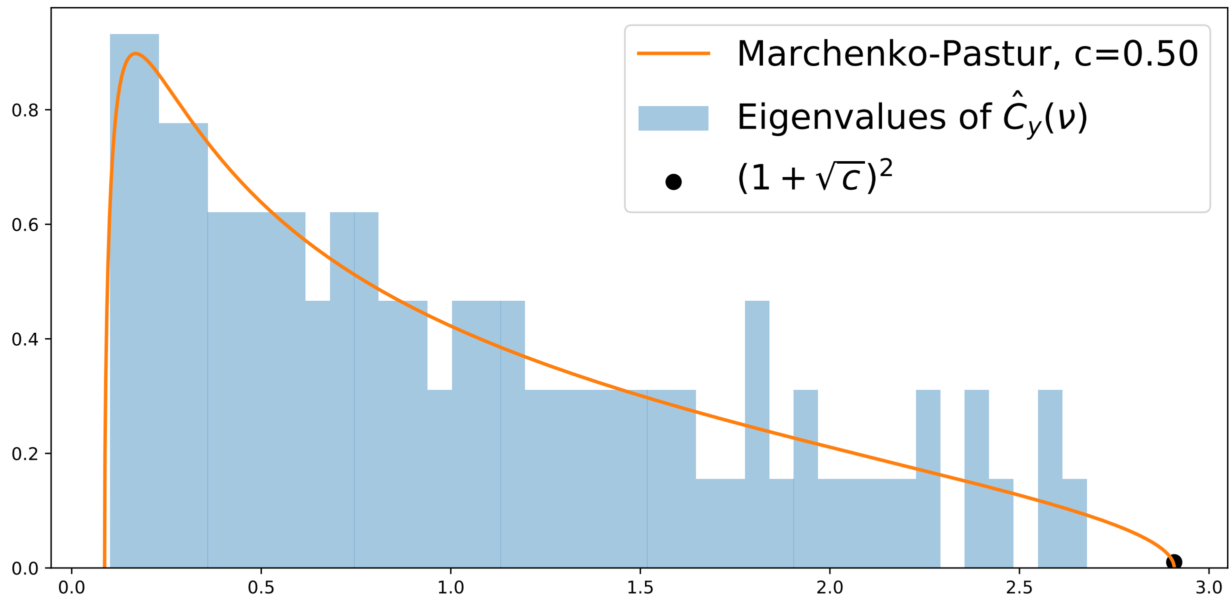


Figure 3.6.1: Eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(0)$ vs the density of the Marcenko-Pastur distribution with parameter $c = 1/2$.

Marcenko-Pastur distribution is plotted against the two following quantities:

$$F_{\min}(t) = \min_{\nu \in \mathcal{V}_N} \frac{1}{M} \sum_{\lambda_i(\hat{\mathbf{C}}(\nu))<t} \delta_{\lambda_i(\hat{\mathbf{C}}(\nu))}$$

$$F_{\max}(t) = \max_{\nu \in \mathcal{V}_N} \frac{1}{M} \sum_{\lambda_i(\hat{\mathbf{C}}(\nu))<t} \delta_{\lambda_i(\hat{\mathbf{C}}(\nu))}.$$

These two functions represent the maximum deviations (from above and below) over the frequencies $\nu \in \mathcal{V}_N$ of the empirical spectral distribution of $\hat{\mathbf{C}}(\nu)$ against the Marcenko-Pastur distribution. As suggested by the uniform convergence in the frequency domain in Corollary 3.3, the Marcenko-Pastur approximation in the high-dimensional regime is reliable over the whole set of Fourier frequencies. Note that the statement of Corollary 3.3 does not exactly match the setting used in Figure 3.6.2, as the test function used here is not in $\mathcal{C}^1((0, +\infty))$.

To illustrate the signal-plus-noise case and the results of Corollary 3.3 and Proposition 3.2, we plot in Figure 3.6.3, the histogram of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for $\nu = 0$, with $\gamma(0) = 2.9$. We see that the largest eigenvalue deviates from the right edge $(1+\sqrt{c})^2$ and is located around the value $\phi(\gamma(0)) = 4.5$, as predicted by Proposition 3.2, while all the other eigenvalues spread as the Marcenko-Pastur distribution, as predicted by Corollary 3.3.

To compare the test statistic (3.17) with other frequency-domain methods based on the SCM, we consider:

- the new test statistic (3.17), denoted as LE (for largest eigenvalue),
- two tests based on LSS of the SCM given by

$$T'_\epsilon = \mathbb{1}_{[\epsilon,+\infty)} \left(\max_{\nu \in \mathcal{V}_N} \left| L_\varphi(\nu) - \int_{\mathbb{R}} \varphi(\lambda) f(\lambda) \mathrm{d}\lambda \right| \right)$$

where L_φ and density f are defined in (3.23) and Corollary 3.3 respectively, and with $\varphi(x) = (x-1)^2$ for the Frobenius norm test (denoted as LSS Frob.) and $\varphi(x) = \log(x)$ for the logdet

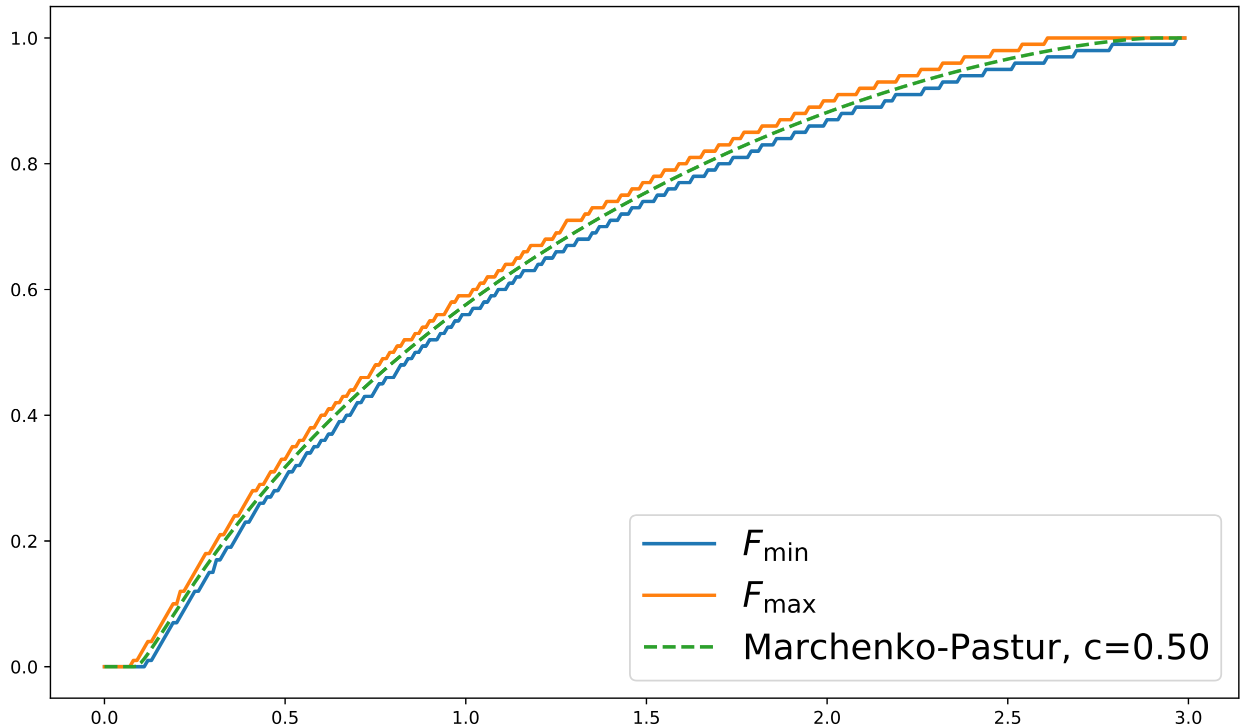


Figure 3.6.2: Uniform convergence of the eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ over $\nu \in \mathcal{V}_N$ toward the Marcenko-Pastur distribution with parameter $c = 1/2$.

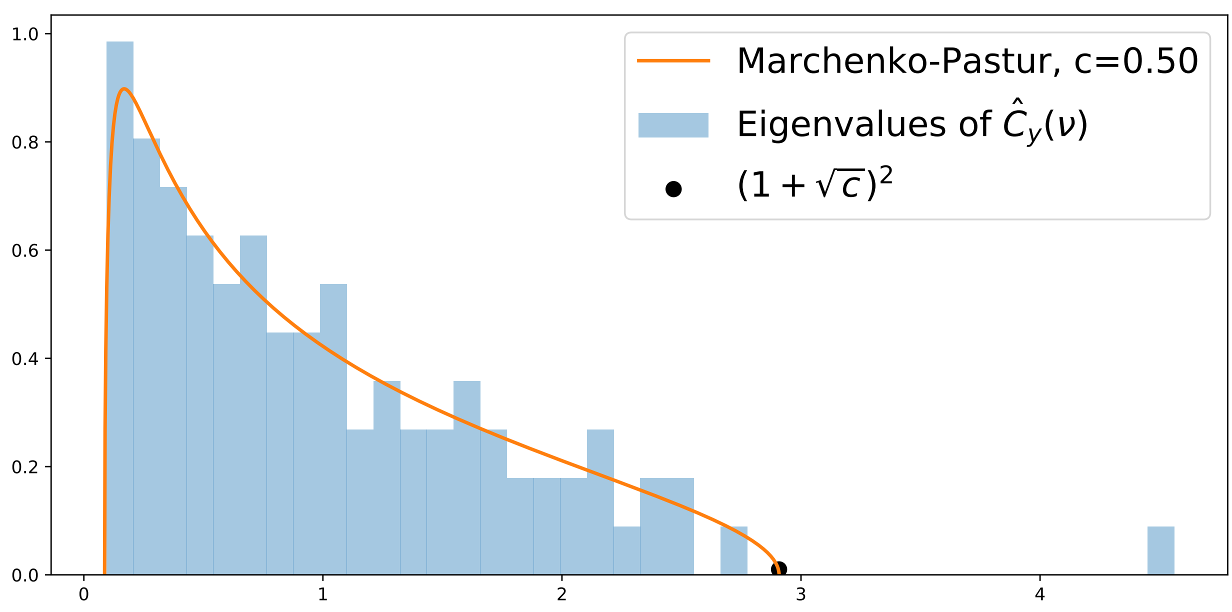


Figure 3.6.3: Eigenvalue distribution of $\hat{\mathbf{C}}(\nu)$ vs Marcenko-Pastur distribution with parameter $c = 1/2$ in the signal case.

test (denoted as LSS logdet),

Table 3.6.1: Power comparison, K=1, $\gamma(\nu_N^*) = 2\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Frob.	LSS logdet	MCC	LE
400	20	40	0.09	0.07	0.06	0.15
1600	40	80	0.15	0.08	0.06	0.37
3600	60	120	0.19	0.08	0.06	0.68
6400	80	160	0.25	0.08	0.06	0.87
10000	100	200	0.26	0.07	0.06	0.96
14400	120	240	0.25	0.06	0.06	0.99
19600	140	280	0.28	0.06	0.06	1.00
25600	160	320	0.30	0.06	0.06	1.00
32400	180	360	0.31	0.06	0.06	1.00

- a test statistic based on the largest off-diagonal entry of the SCM:

$$T_\epsilon^{''} = \mathbb{1}_{[\epsilon,+\infty)}\left(\max_{\nu\in\mathcal{V}_N}\max_{\substack{i,j=1,\ldots,M\\ i<j}}\left|[\hat{\mathbf{C}}_{\mathbf{y}}(\nu)]_{i,j}\right|\right)$$

  denoted as MCC (for Maximum of Cross Coherence),

and where $\epsilon > 0$ is some threshold. In Table 3.6.1, we provide, via Monte-Carlo simulations (10000 draws), the power of each of the four tests, calibrated so that the empirical type I error is equal to 0.05. The results are provided for various values of (N, M, B) chosen so that $M \in \{20, 40, \ldots, 180\}$, $N = M^2$ and $B = 2M$. We set the SNR in the frequency domain as $\max_{\nu\in\mathcal{V}_N}\gamma(\nu) = 2\sqrt{\frac{M}{B}} = 1.41$.

The LE test presents the best detection performance among the four candidates, whereas the MCC test does not seem to be adapted to the detection of this alternative. While it is proved in Corollary 3.3 that the test statistics based on the LSS of $\hat{\mathbf{C}}(\nu)$ cannot asymptotically distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$, they remain sensible to a large variation of a single eigenvalue for finite values of M. Consider for instance the Frobenius LSS test, where the test statistic is based on :

$$\max_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\sum_{m=1}^{M}(\lambda_m(\hat{\mathbf{C}}(\nu))-1)^2 - \int(\lambda-1)^2 f(\lambda)d\lambda\right|$$

where an explicit computation shows that $\int(\lambda-1)^2 f(\lambda)d\lambda = c$. An $\mathcal{O}(1)$ variation of $\lambda_1(\hat{\mathbf{C}}(\nu))$, the largest eigenvalue of $\hat{\mathbf{C}}(\nu)$, will lead to a variation of order $\mathcal{O}(\frac{1}{M})$ of the above term. Therefore, the power of an LSS based test asymptotically converge towards zero, while having non-zero power for finite values of M, as visible in the results of Table 3.6.1.

3.6.2 Case $K > 1$

We finally consider a model which has the flexibility to consider a signal with an arbitrary value of $K \geq 1$. We assume that the matrices $(\mathbf{H}_l)_{l\geq 0}$ satisfy $\mathbf{H}_l = 0$ if $l > L$ for a certain integer L, and that the sequence of $M \times K$ matrices $(\mathbf{H}_l)_{0\leq l\leq L}$ is defined by:

$$\mathbf{H}_l = (C_1\mathbf{w}_{l,1}, \ldots, C_K\mathbf{w}_{l,K})$$

where the vectors $((\mathbf{w}_{l,k})_{l=0,\ldots,L})_{k=1,\ldots,K}$ are generated as independent realisations of M–dimensional vectors uniformly distributed on the unit sphere of $\mathbb{C}^M$ and where the $C_1 \geq C_2 \geq \ldots \geq C_K$ are positive constants used to tune the SNR of each of the K sources at the desired level. Moreover, as the K columns of each matrix $\mathbf{H}_l$ coincide with the realisations of mutually independent random vectors, the columns of $\mathbf{H}(\nu)$ are easily seen to be nearly orthogonal and to nearly share the same norm for each ν if M is large enough. More precisely, for each ν, it holds that $\mathbf{H}(\nu)^*\mathbf{H}(\nu) \to (L+1)\,\mathrm{Diag}(C_1,\ldots,C_K)$ when $M \to +\infty$. As the spectral densities of the components of the noise all coincide with $s(\nu) = \frac{1}{|1-\theta e^{-2i\pi\nu}|^2}$, the non-zero eigenvalues of $\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}$ converge towards $((L+1)C_k/s(\nu))_{k=1,\ldots,K}$ when M increases. Therefore, the signal obtained by this model satisfies Assumption 3.3. Rather than just providing the performance of the test T_ϵ based on the maximum of the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ proposed in this chapter, we compare in the following T_ϵ with $T_{K,\epsilon}$ defined by

$$T_{K,\epsilon} = \mathbb{1}_{[K\lambda^+ + \epsilon,\infty)}\left(\max_{\nu\in\mathcal{V}_N}\sum_{k=1}^{K}\lambda_k\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)\right)$$

which depends on the K largest eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ rather than on the largest one. It is easy to generalize Proposition 3.2 and Proposition 3.3 to study the asymptotic properties of $T_{K,\epsilon}$. More precisely, for each $k = 1,\ldots,K$, we define $\gamma_k(\nu)$ by

$$\gamma_k(\nu) = \lambda_k\left(\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\right) \tag{3.29}$$

and denote $\nu^*_{K,N}$ one of the frequency such that $\max_{\nu\in\mathcal{V}_N}\sum_{k=1}^{K}\gamma_k(\nu) = \sum_{k=1}^{K}\gamma_k(\nu^*_{K,N})$. $\gamma_k(\nu)$ can of course be seen as a generalization of $\gamma(\nu)$ defined by (3.24). Then, under the extra assumption that for $k = 1,\ldots,K$, $\gamma_k(\nu^*_{K,N})$ converges towards a finite limit $\gamma_{k,\infty}$ (a condition which holds in the context of the present experiment because it is easily seen that $\gamma_k(\nu^*_{K,N}) \to (L+1)\,(1+\theta)^2 C_k$), $\lambda_k\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu^*_{K,N})\right)$ converges towards λ^+ if $\gamma_{k,\infty} \leq \sqrt{c}$ and towards $\phi(\gamma_{k,\infty}) > \lambda^+$ if $\gamma_{k,\infty} > \sqrt{c}$. It is easy to check that if $\gamma_\infty = \gamma_{1,\infty} > \sqrt{c}$, then the statistics $T_{K,\epsilon}$ also leads to a consistent test provided $0 < \epsilon < \phi(\gamma_\infty) - \lambda_+$. While in practice the number of sources K is unknown, it is interesting to evaluate the performance provided by $T_{K,\epsilon}$ which can be considered as an ideal reference. Intuitively, $T_{K,\epsilon}$ could lead to a better performance than T_ϵ when $\gamma_{k,\infty} > \sqrt{c}$ for $k = 1,\ldots,K$, because, in this context, if $\hat{\nu}^*_{K,N}$ is a frequency that maximises $\sum_{k=1}^{K}\lambda_k\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)$, then $\liminf_{M\to+\infty}\lambda_k\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu^*_{K,N})\right) > \lambda^+$. Therefore, the K largest eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\hat{\nu}^*_{K,N})$ bring useful information to the detection of the useful signal.

To evaluate numerically the compared performance of T_ϵ and $T_{K,\epsilon}$ when K is known, we first consider the case $K = 2$, $L = 3$, and where $(\gamma_1+\gamma_2)(\nu^*_{2,N}) = 3\sqrt{c}$. Concerning the value of (C_1, C_2), we consider the two following cases: $\frac{C_1}{C_2} = 1$ and $\frac{C_1}{C_2} = 4$. This corresponds respectively to the case where both sources contribute exactly the same on each sensor, and where the first source contributes much more than the second one. Tables 3.6.2, 3.6.3 report the power of the proposed test (LE(1) represents T_ϵ and LE(2) represents $T_{2,\epsilon}$) against the LSS tests and the MCC test, with a type I error fixed at 5%. When $\frac{C_1}{C_2} = 4$, it can be expected that the most powerful source is dominant, and that $\gamma_2(\nu^*_{2,N}) < \sqrt{c}$. Therefore, $\lambda_2\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)$ is likely to stay close to λ^+ for each ν, so that the use of $T_{2,\epsilon}$ should not bring any extra performance. This intuition is confirmed by Table 3.6.2. When $\frac{C_1}{C_2} = 1$, $\gamma_1(\nu^*_{2,N})$ and $\gamma_2(\nu^*_{2,N})$ should be both close to $\frac{3}{2}\sqrt{c}$, thus suggesting that the two largest eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}$ at the maximizing frequency $\hat{\nu}^*_{2,N}$ should also nearly coincide, and should escape from $[\lambda^-, \lambda^+]$. While the second eigenvalue brings here some information, Table 3.6.3 tends to indicate that T_ϵ has better performance than $T_{2,\epsilon}$. In the next experiment, $(\gamma_1+\gamma_2)(\nu^*_{2,N}) = 2\sqrt{c}$. For $\frac{C_1}{C_2} = 4$, the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ is likely to be still dominant for each ν, and Table 3.6.4

Table 3.6.2: Power comparison, $\frac{C_1}{C_2} = 4$, $(\gamma_1 + \gamma_2)(\nu_{2,N}^*) = 3\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Fr.	LSS ld	MCC	LE(1)	LE(2)
100	10	20	0.31	0.18	0.16	0.42	0.37
400	20	40	0.79	0.39	0.45	0.94	0.89
900	30	60	0.94	0.49	0.53	1.00	0.99
1600	40	80	0.98	0.50	0.55	1.00	1.00
2500	50	100	0.99	0.52	0.55	1.00	1.00
3600	60	120	1.00	0.51	0.43	1.00	1.00
4900	70	140	1.00	0.55	0.37	1.00	1.00
6400	80	160	1.00	0.54	0.28	1.00	1.00

Table 3.6.3: Power comparison, $\frac{C_1}{C_2} = 1$, $(\gamma_1 + \gamma_2)(\nu_{2,N}^*) = 3\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Fr.	LSS ld	MCC	LE(1)	LE(2)
100	10	20	0.38	0.22	0.16	0.48	0.46
400	20	40	0.58	0.30	0.30	0.75	0.73
900	30	60	0.67	0.30	0.28	0.91	0.89
1600	40	80	0.74	0.29	0.18	0.96	0.97
2500	50	100	0.79	0.30	0.16	0.99	0.99
3600	60	120	0.79	0.24	0.13	1.00	1.00
4900	70	140	0.85	0.28	0.12	1.00	1.00

confirms the better performance of T_ϵ. When $\frac{C_1}{C_2} = 1$, $\gamma_1(\nu_{2,N}^*)$ and $\gamma_2(\nu_{2,N}^*)$ should be both close to the detectability threshold $\sqrt{c}$, and Table 3.6.5 this time shows that the use of $T_{2,\epsilon}$ leads to some improvement. For comparison, we also report the results of T_ϵ for $C_2 = 0$ in Table 3.6.6.

This discussion tends to indicate that, even when $K > 1$ is assumed known, the use of the maximum over $\mathcal{V}_N$ of the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ does not introduce any significant loss of performance.

3.7 Conclusion

In this chapter, we have studied the statistical behaviour of certain frequency-domain detection test statistics, based on the eigenvalues of a sample estimate of the SCM, in the high-dimensional regime in which both the dimension M of the underlying signals and the number of samples N converge to infinity at certain rates. In particular, we have proved various approximation results showing that the sample SCM asymptotically behaves as a Wishart matrix. These results have been exploited to prove that test statistics based on LSS of the sample SCM are not consistent in the high-dimensional regime. A new test statistic relying on the largest eigenvalue of the sample SCM has also been proposed and proved to be consistent in the high-dimensional regime. Finally, numerical results have demonstrated that this new test statistic provides reasonable performance and outperforms other standard test statistics in situations where the dimension M and the number of samples N are large.

Table 3.6.4: Power comparison, $\frac{C_1}{C_2} = 4$, $(\gamma_1 + \gamma_2)(\nu^*_{2,N}) = 2\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Fr.	LSS ld	MCC	LE(1)	LE(2)
100	10	20	0.15	0.10	0.10	0.21	0.20
400	20	40	0.33	0.15	0.12	0.55	0.50
900	30	60	0.39	0.15	0.17	0.75	0.71
1600	40	80	0.52	0.16	0.14	0.94	0.90
2500	50	100	0.54	0.15	0.14	0.98	0.97
3600	60	120	0.56	0.13	0.13	1.00	0.99
4900	70	140	0.55	0.13	0.10	1.00	1.00
6400	80	160	0.62	0.11	0.10	1.00	1.00

Table 3.6.5: Power comparison, $\frac{C_1}{C_2} = 1$, $(\gamma_1 + \gamma_2)(\nu^*_{2,N}) = 2\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Fr.	LSS ld	MCC	LE(1)	LE(2)
400	20	40	0.17	0.11	0.08	0.27	0.27
1600	40	80	0.18	0.10	0.08	0.45	0.48
3600	60	120	0.15	0.07	0.07	0.58	0.62
6400	80	160	0.16	0.07	0.08	0.69	0.75
10000	100	200	0.13	0.05	0.07	0.76	0.83
14400	120	240	0.10	0.03	0.07	0.82	0.86
19600	140	280	0.09	0.04	0.07	0.86	0.89
25600	160	320	0.10	0.03	0.06	0.89	0.93
32400	180	360	0.09	0.03	0.06	0.87	0.93

Table 3.6.6: Power comparison, $C_2 = 0$, $\gamma(\nu^*_N) = 2\sqrt{\frac{1}{2}}$, type I error = 5%

N	M	B	LSS Fr.	LSS ld	MCC	LE(1)	LE(2)
100	10	20	0.19	0.12	0.12	0.26	0.22
400	20	40	0.43	0.19	0.14	0.66	0.59
900	30	60	0.51	0.19	0.19	0.88	0.83
1600	40	80	0.62	0.20	0.15	0.97	0.95
2500	50	100	0.65	0.18	0.17	0.99	0.99
3600	60	120	0.68	0.16	0.12	1.00	1.00
4900	70	140	0.71	0.16	0.13	1.00	1.00
6400	80	160	0.75	0.17	0.12	1.00	1.00

Appendix

3.A Useful results

In this section, we recall some useful results which will be constantly used in the proofs developed in the following sections.

The first result is based on a Chernoff bound for the χ^2 distribution, and is also a special case of the well-known Hanson-Wright inequality describing the concentration of sub-Gaussian quadratic forms around their means (see [79]).

Lemma 3.1. *Let $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{\Xi}$ a deterministic $n \times n$ complex matrix. Then there exists a constant $\kappa > 0$ independent of n and $\mathbf{\Xi}$ such that for all $t \geq 0$,*

$$\mathbb{P}\left(|\mathbf{z}^* \mathbf{\Xi} \mathbf{z} - \mathbb{E}[\mathbf{z}^* \mathbf{\Xi} \mathbf{z}]| > t\right) \leq 2 \ \exp\left(-\kappa \min\left\{\frac{t^2}{\|\mathbf{\Xi}\|_F^2}, \frac{t}{\|\mathbf{\Xi}\|_2}\right\}\right).$$

The second result describes the behaviour of the largest and the smallest eigenvalues of a standard Wishart matrix.

Lemma 3.2 ([34, Proof of Lemma 7.3])**.** *Let $\mathbf{Z}$ be an $M \times (B+1)$ matrix with i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries. Then under Assumption 3.4, there exists a constant $C > 0$ independent of M, B such that for all $t > 0$,*

$$\mathbb{P}\left(\lambda_1\left(\frac{\mathbf{Z}\mathbf{Z}^*}{B+1}\right) > \left(1 + \sqrt{\frac{M}{B+1}}\right)^2 + t\right) \leq (B+1)\exp\left(-C(B+1)t^2\right)$$

and

$$\mathbb{P}\left(\lambda_M\left(\frac{\mathbf{Z}\mathbf{Z}^*}{B+1}\right) < \left(1 - \sqrt{\frac{M}{B+1}}\right)^2 - t\right) \leq (B+1)\exp\left(-C(B+1)t^2\right).$$

We will mainly use Lemma 3.2 as follows; let $(\mathbf{Z}(\nu))_{\nu \in \mathcal{V}_N}$ be a family of $M \times (B+1)$ random matrices such that $\mathbf{Z}(\nu)$ has i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ (recall the definition of the index set $\mathcal{V}_N$ in (3.16)), then from the union bound

$$\begin{aligned}
\mathbb{P}&\left(\max_{\nu \in \mathcal{V}_N} \lambda_1\left(\frac{\mathbf{Z}(\nu)\mathbf{Z}(\nu)^*}{B+1}\right) > \left(1 + \sqrt{\frac{M}{B+1}}\right)^2 + t\right) \\
&\leq \sum_{\nu \in \mathcal{V}_N} \mathbb{P}\left(\lambda_1\left(\frac{\mathbf{Z}(\nu)\mathbf{Z}(\nu)^*}{B+1}\right) > \left(1 + \sqrt{\frac{M}{B+1}}\right)^2 + t\right) \\
&\leq \exp\Big(-C(B+1)t^2 + \log(N(B+1))\Big).
\end{aligned}$$

Using Assumption 3.4 and the Borel-Cantelli lemma, we deduce that

$$\limsup_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1\left(\frac{\mathbf{Z}(\nu)\mathbf{Z}(\nu)^*}{B+1}\right) \leq (1+\sqrt{c})^2$$

with probability one.

3.B Proof of Theorem 3.2

3.B.1 Reduction to $K = 1$

First, note that we may assume $K = 1$ without loss of generality. Indeed, consider the decomposition

$$\mathbf{u}_n = \sum_{\ell=1}^{K} \mathbf{u}_n^{(\ell)}$$

where $\mathbf{u}_n^{(\ell)} = \sum_{k=0}^{+\infty} \mathbf{h}_{\ell,k}\epsilon_{n-k}^{(\ell)}$ and where $\mathbf{h}_{\ell,k}$ and $\epsilon_n^{(\ell)}$ are the ℓ-th column of $\mathbf{H}_k$ and the ℓ-th entry of $\boldsymbol{\epsilon}_n$ respectively. Moreover, Assumption 3.3 implies that

$$\sup_{M\geq 1} \sum_{k\in\mathbb{Z}} (1+|k|)\|\mathbf{h}_{\ell,k}\|_2 < \infty$$

From the fact that K is fixed with respect to N (Assumption 3.4) and

$$\max_{\nu\in\mathcal{V}_N} \left\|\boldsymbol{\Sigma}_{\mathbf{u}}(\nu) - \mathbf{H}(\nu)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\right\|_2 \leq \sum_{\ell=1}^{K} \max_{\nu\in\mathcal{V}_N} \left\|\boldsymbol{\Sigma}_{\mathbf{u}^{(\ell)}}(\nu) - \mathbf{h}_\ell(\nu)\boldsymbol{\Sigma}_{\epsilon^{(\ell)}}(\nu)\right\|_2$$

where $\boldsymbol{\Sigma}_{\mathbf{u}^{(\ell)}}(\nu)$, $\mathbf{h}_\ell(\nu)$, $\boldsymbol{\Sigma}_{\epsilon^{(\ell)}}(\nu)$ are defined as $\boldsymbol{\Sigma}_{\mathbf{u}}(\nu)$, $\mathbf{H}(\nu)$, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)$ respectively, Theorem 3.2 will be proved if we can show that

$$\max_{\nu\in\mathcal{V}_N} \left\|\boldsymbol{\Sigma}_{\mathbf{u}^{(\ell)}}(\nu) - \mathbf{h}_\ell(\nu)\boldsymbol{\Sigma}_{\epsilon^{(\ell)}}(\nu)\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0$$

for all $\ell = 1, \ldots, K$. Therefore, we assume for the remainder of the proof that

$$\mathbf{u}_n = \sum_{k\in\mathbb{Z}} \mathbf{h}_k \epsilon_{n-k},$$

where

- $(\mathbf{h}_k)_{k\in\mathbb{Z}}$ is a filter, with $\mathbf{h}_k \in \mathbb{C}^M$ and such that

$$\sup_{M\geq 1} \sum_{k\in\mathbb{Z}} (1+|k|)\left\|\mathbf{h}_k\right\|_2 < \infty. \tag{3.30}$$

- $(\epsilon_n)_{n\in\mathbb{Z}}$ is a scalar standard complex Gaussian white noise.

3.B.2 Reduction to $B = 1$

Let $\mathbf{h}(\nu) = \sum_{k\in\mathbb{Z}} \mathbf{h}_k \mathrm{e}^{-\mathrm{i}2\pi\nu k}$ and

$$\xi_\epsilon(\nu) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \epsilon_n \mathrm{e}^{-\mathrm{i}2\pi\nu n}.$$

From (3.30) and Assumption 3.4, a first-order Taylor expansion of $b \mapsto \mathbf{h}\left(\nu + \frac{b}{N}\right)$ at 0 leads to

$$\begin{aligned}\sup_{\nu\in[0,1]} \max_{b\in\{-\frac{B}{2},\dots,\frac{B}{2}\}} \left\| \mathbf{h}(\nu) - \mathbf{h}\left(\nu + \frac{b}{N}\right)\right\|_2 &= \mathcal{O}\left(\frac{B}{N}\right) \\ &= \mathcal{O}\left(\frac{1}{N^{1-\alpha}}\right).\end{aligned}$$

Moreover, from Lemma 3.1 applied to the random vector

$$\mathbf{z} = \left(\xi_\epsilon\left(\nu - \frac{B}{2N}\right), \dots, \xi_\epsilon\left(\nu + \frac{B}{2N}\right)\right)^T \sim \mathcal{N}_{\mathbb{C}^{B+1}}\left(\mathbf{0}, \mathbf{I}_{B+1}\right)$$

and matrix $\boldsymbol{\Xi} = \frac{\mathbf{I}_{B+1}}{B+1}$, there exists some constant κ independent of M such that for all $t \geq 2$,

$$\begin{aligned}\mathbb{P}\left(\max_{\nu\in\mathcal{V}_N} \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_\epsilon\left(\nu + \frac{b}{N}\right)\right|^2 > t\right) &\leq N\mathbb{P}\left(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_\epsilon\left(\frac{b}{N}\right)\right|^2 > t\right) \\ &\leq N\exp\left(-\kappa B\right)\end{aligned}$$

and Borel-Cantelli lemma together with Assumption 3.4 imply

$$\max_{\nu\in\mathcal{V}_N} \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_\epsilon\left(\nu + \frac{b}{N}\right)\right|^2 = \mathcal{O}\left(1\right)$$

with probability one. Defining

$$\boldsymbol{\Sigma}_\epsilon(\nu) = \frac{1}{\sqrt{B+1}}\left(\xi_\epsilon\left(\nu - \frac{B}{2N}\right), \dots, \xi_\epsilon\left(\nu + \frac{B}{2N}\right)\right)$$

as well as

$$\boldsymbol{\Phi}(\nu) = \frac{1}{\sqrt{B+1}}\left[\boldsymbol{\phi}\left(\nu - \frac{B}{2N}\right), \dots, \boldsymbol{\phi}\left(\nu + \frac{B}{2N}\right)\right]$$

with $\boldsymbol{\phi}(\nu) = \mathbf{h}(\nu)\xi_\epsilon(\nu)$, we therefore have the control

$$\begin{aligned}\max_{\nu\in\mathcal{V}_N}\left\|\mathbf{h}(\nu)\boldsymbol{\Sigma}_\epsilon(\nu) - \boldsymbol{\Phi}(\nu)\right\|_2 &\leq \sup_{\nu\in[0,1]} \max_{b\in\{-\frac{B}{2},\dots,\frac{B}{2}\}} \left\|\mathbf{h}(\nu) - \mathbf{h}\left(\nu + \frac{b}{N}\right)\right\|_2 \\ &\quad\times\sqrt{\max_{\nu\in\mathcal{V}_N}\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_\epsilon\left(\nu + \frac{b}{N}\right)\right|^2} \\ &= \mathcal{O}\left(\frac{1}{N^{1-\alpha}}\right) \text{ a.s. } \xrightarrow[M\to\infty]{a.s.} 0.\end{aligned}$$

Finally, since the spectral norm of a matrix is bounded by its Frobenius norm,

$$\begin{aligned}\max_{\nu\in\mathcal{V}_N}\left\|\boldsymbol{\Sigma}_{\mathbf{u}}(\nu) - \boldsymbol{\Phi}(\nu)\right\|_2 &\leq \sqrt{\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left\|\boldsymbol{\xi}_{\mathbf{u}}\left(\nu + \frac{b}{N}\right) - \boldsymbol{\phi}\left(\nu + \frac{b}{N}\right)\right\|_2^2} \\ &\leq \max_{\nu\in\mathcal{V}_N}\left\|\boldsymbol{\xi}_{\mathbf{u}}\left(\nu\right) - \boldsymbol{\phi}\left(\nu\right)\right\|_2.\end{aligned}$$

Theorem 3.2 is proven if we show that

$$\max_{\nu\in\mathcal{V}_N} \|\boldsymbol{\xi}_{\mathbf{u}}(\nu) - \mathbf{h}(\nu)\,\boldsymbol{\xi}_{\epsilon}(\nu)\|_2 \xrightarrow[N\to\infty]{a.s.} 0.$$

3.B.3 Periodization

For all integer n, let $[n]$ denotes the integer contained in $\{0,\ldots,N-1\}$ such that $[n]\equiv n \pmod N$ and define

$$\tilde{\mathbf{u}}_n = \sum_{k\in\mathbb{Z}} \mathbf{h}_k \epsilon_{[n-k]}$$

where $(\tilde{\mathbf{u}}_n)_{n\in\mathbb{Z}}$ represents the circular convolution between $(\mathbf{h}_k)_{k\in\mathbb{Z}}$ and $(\epsilon_n)_{n\in\mathbb{Z}}$. If $\boldsymbol{\xi}_{\tilde{\mathbf{u}}}(\nu) = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\tilde{\mathbf{u}}_n \mathrm{e}^{-\mathrm{i}2\pi n\nu}$, then the equality

$$\boldsymbol{\xi}_{\tilde{\mathbf{u}}}(\nu) = \mathbf{h}(\nu)\xi_{\epsilon}(\nu)$$

holds for all $\nu\in\mathcal{V}_N$. It is straightforward to check that

$$\boldsymbol{\xi}_{\tilde{\mathbf{u}}}(\nu) - \boldsymbol{\xi}_{\mathbf{u}}(\nu) = \boldsymbol{\delta}(\nu) + \check{\boldsymbol{\delta}}(\nu)$$

where

$$\begin{aligned}\boldsymbol{\delta}(\nu) = &\frac{1}{\sqrt{N}}\sum_{k=1}^{N-1}\mathbf{h}_k\sum_{p=1}^{k}\left(\epsilon_{[-p]} - \epsilon_{-p}\right)\mathrm{e}^{-\mathrm{i}2\pi\nu(k-p)} \\ &+ \frac{1}{\sqrt{N}}\sum_{k=N}^{+\infty}\mathbf{h}_k\sum_{p=0}^{N-1}\left(\epsilon_{[p-k]} - \epsilon_{p-k}\right)\mathrm{e}^{-\mathrm{i}2\pi\nu p}\end{aligned}$$

and

$$\begin{aligned}\check{\boldsymbol{\delta}}(\nu) = &\frac{1}{\sqrt{N}}\sum_{k=1}^{N-1}\mathbf{h}_{-k}\sum_{p=1}^{k}\left(\epsilon_{[N+p-1]} - \epsilon_{N+p-1}\right)\mathrm{e}^{-\mathrm{i}2\pi\nu(N-1+p-k)} \\ &+ \frac{1}{\sqrt{N}}\sum_{k=N}^{+\infty}\mathbf{h}_{-k}\sum_{p=0}^{N-1}\left(\epsilon_{[p+k]} - \epsilon_{p+k}\right)\mathrm{e}^{-\mathrm{i}2\pi\nu p}\end{aligned}$$

Theorem 3.2 is proved if we can show that

$$\max_{\nu\in\mathcal{V}_N}\|\boldsymbol{\delta}(\nu)\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.31}$$

and

$$\max_{\nu\in\mathcal{V}_N}\left\|\check{\boldsymbol{\delta}}(\nu)\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0. \tag{3.32}$$

In the remainder, we only prove (3.31) and omit the details for (3.32) whose treatment is similar. To that end, we define

$$\boldsymbol{\delta}_1(\nu) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N-1} \mathbf{h}_k \sum_{p=1}^{k} \left(\epsilon_{[-p]} - \epsilon_{-p}\right) \mathrm{e}^{-\mathrm{i}2\pi\nu(k-p)}$$

$$\boldsymbol{\delta}_2(\nu) = \frac{1}{\sqrt{N}} \sum_{k=N}^{+\infty} \mathbf{h}_k \sum_{p=0}^{N-1} \left(\epsilon_{[p-k]} - \epsilon_{p-k}\right) \mathrm{e}^{-\mathrm{i}2\pi\nu p}.$$

3.B.4 Control of $\boldsymbol{\delta}_1(\nu)$

For $p \in \{1, \ldots, N-1\}$, let

$$z_p(\nu) = \left(\epsilon_{[-p]} - \epsilon_{-p}\right) \mathrm{e}^{\mathrm{i}2\pi\nu p} = \left(\epsilon_{N-p} - \epsilon_{-p}\right) \mathrm{e}^{\mathrm{i}2\pi\nu p}.$$

Then $z_1(\nu), \ldots, z_{N-1}(\nu)$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,2)$ and by rearranging the sums in $\boldsymbol{\delta}_1(\nu)$, we have

$$\boldsymbol{\delta}_1(\nu) = \sum_{p=1}^{N-1} z_p(\nu) \mathbf{g}_p(\nu)$$

with

$$\mathbf{g}_p(\nu) = \frac{1}{\sqrt{N}} \sum_{k=p}^{N-1} \mathbf{h}_k \mathrm{e}^{-\mathrm{i}2\pi k\nu}.$$

Therefore, $\boldsymbol{\delta}_1(\nu) \sim \mathcal{N}_{\mathbb{C}^M}\left(\mathbf{0}, \mathbf{G}(\nu)\right)$ with

$$\mathbf{G}(\nu) = 2 \sum_{p=1}^{N-1} \mathbf{g}_p(\nu) \mathbf{g}_p(\nu)^*.$$

Moreover,

$$\mathbb{E} \left\|\boldsymbol{\delta}_1(\nu)\right\|_2^2 = \mathrm{tr}\ \mathbf{G}(\nu) \leq \frac{2}{N} \sum_{p=1}^{N-1} \left(\sum_{k=p}^{N-1} \left\|\mathbf{h}_k\right\|_2^2 + 2 \sum_{p \leq k < k' \leq N-1} \left\|\mathbf{h}_k\right\|_2 \left\|\mathbf{h}_{k'}\right\|_2 \right)$$

and a straightforward rearrangement together with (3.30) leads to

$$\begin{aligned} \max_{\nu \in [0,1]} \mathbb{E} \left\|\boldsymbol{\delta}_1(\nu)\right\|_2^2 &\leq \frac{2}{N} \sum_{k=1}^{N-1} k \left\|\mathbf{h}_k\right\|_2^2 + \frac{4}{N} \sum_{1 \leq k < k' \leq N-1} \sqrt{k}\sqrt{k'} \left\|\mathbf{h}_k\right\|_2 \left\|\mathbf{h}_{k'}\right\|_2 \\ &= \frac{2}{N} \left(\sum_{k=1}^{N-1} \sqrt{k} \left\|\mathbf{h}_k\right\|_2 \right)^2 = \mathcal{O}\left(\frac{1}{N}\right). \end{aligned}$$

where we used that $k \leq \sqrt{k}\sqrt{k'}$ for $k' \geq k$. Additionally,

$$\max_{\nu \in [0,1]} \left\|\mathbf{G}(\nu)\right\|_2 \leq \max_{\nu \in [0,1]} \mathrm{tr}\ \mathbf{G}(\nu) = \mathcal{O}\left(\frac{1}{N}\right)$$

and

$$\max_{\nu\in[0,1]} \|\mathbf{G}(\nu)\|_F \leq \sqrt{M} \max_{\nu\in[0,1]} \|\mathbf{G}(\nu)\|_2 = \mathcal{O}\left(\frac{\sqrt{M}}{N}\right).$$

Using Lemma 3.1, there exists a constant $\kappa > 0$ independent of $M, (\mathbf{h}_k)_{k\in\mathbb{Z}}$ such that for all $t > 0$,

$$\mathbb{P}\left(\max_{\nu\in\mathcal{V}_N} \left|\|\boldsymbol{\delta}_1(\nu)\|_2^2 - \mathbb{E}\|\boldsymbol{\delta}_1(\nu)\|_2^2\right| > t\right) \leq 2N \max_{\nu\in\mathcal{V}_N} \exp\left(-\kappa \min\left(\frac{t^2}{\|\mathbf{G}(\nu)\|_F^2}, \frac{t}{\|\mathbf{G}(\nu)\|_2}\right)\right).$$

Applying Assumption 3.4 and the Borel-Cantelli lemma, it follows that

$$\max_{\nu\in\mathcal{V}_N} \left|\|\boldsymbol{\delta}_1(\nu)\|_2^2 - \mathbb{E}\|\boldsymbol{\delta}_1(\nu)\|_2^2\right| \xrightarrow[N\to\infty]{a.s.} 0.$$

Finally, we deduce that

$$\begin{aligned}\max_{\nu\in\mathcal{V}_N} \|\boldsymbol{\delta}_1(\nu)\|_2^2 &\leq \max_{\nu\in\mathcal{V}_N} \mathbb{E}\|\boldsymbol{\delta}_1(\nu)\|_2^2 + \max_{\nu\in\mathcal{V}_N} \left|\|\boldsymbol{\delta}_1(\nu)\|_2^2 - \mathbb{E}\|\boldsymbol{\delta}_1(\nu)\|_2^2\right| \\ &\xrightarrow[N\to\infty]{a.s.} 0.\end{aligned}$$

3.B.5 Control of $\boldsymbol{\delta}_2(\nu)$

We first split $\boldsymbol{\delta}_2(\nu)$ in the following two parts

$$\boldsymbol{\delta}_2(\nu) = \boldsymbol{\delta}_{2,1}(\nu) + \boldsymbol{\delta}_{2,2}(\nu)$$

where

$$\boldsymbol{\delta}_{2,1}(\nu) = \frac{1}{\sqrt{N}} \sum_{k=N}^{+\infty} \mathbf{h}_k \sum_{p=0}^{N-1} \epsilon_{[p-k]} \mathrm{e}^{-\mathrm{i}2\pi p\nu}$$

$$\boldsymbol{\delta}_{2,2}(\nu) = \frac{1}{\sqrt{N}} \sum_{k=N}^{+\infty} \mathbf{h}_k \sum_{p=0}^{N-1} \epsilon_{p-k} \mathrm{e}^{-\mathrm{i}2\pi p\nu}.$$

We note that $\boldsymbol{\delta}_{2,1}(\nu)$ only involves the N i.i.d. random variables $\epsilon_0, \ldots, \epsilon_{N-1}$ and that

$$\boldsymbol{\delta}_{2,1}(\nu) = \sum_{p=0}^{N-1} \epsilon_p \tilde{\mathbf{g}}_p(\nu)$$

with $\tilde{\mathbf{g}}_p(\nu)$ defined as

$$\tilde{\mathbf{g}}_p(\nu) = \frac{1}{\sqrt{N}} \sum_{k=N}^{+\infty} \mathbf{h}_k \mathrm{e}^{-\mathrm{i}2\pi\nu[p+k]}.$$

It is clear that

$$\max_{p=1,\ldots,N} \max_{\nu\in[0,1]} \|\tilde{\mathbf{g}}_p(\nu)\|_2 \leq \frac{1}{\sqrt{N}} \sum_{k=N}^{+\infty} \|\mathbf{h}_k\|_2 \leq \frac{1}{N^{3/2}} \sum_{k=N}^{+\infty} k \|\mathbf{h}_k\|_2$$

and from (3.30),

$$\max_{p=1,\dots,N} \max_{\nu\in[0,1]} \|\tilde{\mathbf{g}}_p(\nu)\|_2 = o\left(\frac{1}{N^{3/2}}\right)$$

Thus $\boldsymbol{\delta}_{2,1}(\nu) \sim \mathcal{N}_{\mathbb{C}^M}\left(\mathbf{0}, \tilde{\mathbf{G}}(\nu)\right)$ with $\tilde{\mathbf{G}}(\nu) = \sum_{p=0}^{N-1} \tilde{\mathbf{g}}_p(\nu)\tilde{\mathbf{g}}_p(\nu)^*$ and

$$\max_{\nu\in[0,1]} \operatorname{tr} \tilde{\mathbf{G}}(\nu) = o\left(\frac{1}{N^2}\right)$$

as $M \to \infty$. Using Lemma 3.1 as for the control of $\boldsymbol{\delta}_1(\nu)$ in the previous section, we end up with

$$\max_{\nu\in\mathcal{V}_N} \|\boldsymbol{\delta}_{2,1}(\nu)\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

We now consider the term $\boldsymbol{\delta}_{2,2}(\nu)$, which involves the sequence of random variables $(\epsilon_{-n})_{n\geq 1}$. For all $k \geq N$, set

$$\boldsymbol{\chi}_k = \frac{1}{\sqrt{N}}\mathbf{h}_k \sum_{p=0}^{N-1} \epsilon_{p-k} \mathrm{e}^{-\mathrm{i}2\pi p\nu}$$

and consider the sequence $(\boldsymbol{\chi}_p)_{k\geq N}$. Using Assumption 3.3,

$$\begin{aligned}
\sum_{k=N}^{+\infty} \|\boldsymbol{\chi}_k\|_2 &\leq \sum_{k=N}^{+\infty} \sqrt{k}\|\mathbf{h}_k\|_2 \left| \frac{1}{N} \sum_{p=0}^{N-1} \epsilon_{p-k} \mathrm{e}^{-\mathrm{i}2\pi p\nu} \right| \\
&\leq \left(\sup_{k\geq N} \left| \frac{1}{N} \sum_{p=0}^{N-1} \epsilon_{p-k} \mathrm{e}^{-\mathrm{i}2\pi p\nu} \right| \right) \sum_{k=N}^{+\infty} \sqrt{k}\|\mathbf{h}_k\|_2 \\
&< +\infty \text{ a.s.}
\end{aligned}$$

since for any k, by the gaussianity of the ϵ_{p-k}, $\sup_{\nu\in\mathcal{V}_N} \left| \frac{1}{N} \sum_{p=0}^{N-1} \epsilon_{p-k} \mathrm{e}^{-\mathrm{i}2\pi p\nu} \right|$ converges almost surely towards 0 as $N \to +\infty$ by the law of the large numbers, so it remains almost surely bounded for any finite N. This implies that the family $(\boldsymbol{\chi}_k)_{k\geq N}$ is a.s. absolutely summable. Therefore, we can rearrange the series defining $\boldsymbol{\delta}_{2,2}(\nu)$ and write

$$\boldsymbol{\delta}_{2,2}(\nu) = \sum_{p=1}^{+\infty} \epsilon_{-p} \breve{\mathbf{g}}_p(\nu)$$

with probability one, where this time $\mathbf{g}_p(\nu)$ is defined for all $p \geq 1$ as

$$\breve{\mathbf{g}}_p(\nu) = \begin{cases} \frac{1}{\sqrt{N}} \sum_{k=0}^{p-1} \mathbf{h}_{k+N} \ \mathrm{e}^{-\mathrm{i}2\pi(N+k-p)\nu} & \text{if } p \in \{1,\dots,N\} \\ \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \mathbf{h}_{p+k} \ \mathrm{e}^{-\mathrm{i}2\pi k\nu} & \text{if } p \geq N+1. \end{cases}.$$

Again,

$$\sup_{p\geq 1} \max_{\nu\in[0,1]} \|\breve{\mathbf{g}}_p(\nu)\|_2 = o\left(\frac{1}{N}\right) \tag{3.33}$$

$\boldsymbol{\delta}_{2,2}(\nu) \sim \mathcal{N}_{\mathbb{C}^M}\left(\mathbf{0}, \check{\mathbf{G}}(\nu)\right)$, where

$$\check{\mathbf{G}}(\nu) = \sum_{p=1}^{+\infty} \check{\mathbf{g}}_p(\nu)\check{\mathbf{g}}_p(\nu)^*$$

and such that $\operatorname{tr} \check{\mathbf{G}}(\nu) = o\left(\frac{1}{N}\right)$. Thus, using Lemma 3.1 also yields

$$\max_{\nu \in \mathcal{V}_N} \|\boldsymbol{\delta}_{2,2}(\nu)\|_2 \xrightarrow[M\to\infty]{a.s.} 0.$$

This concludes the proof of Theorem 3.2.

3.C Proof of Theorem 3.3

To prove Theorem 3.3, we need as a preliminary step to study the impact of the renormalization by $\operatorname{dg}(\hat{\mathbf{S}}_{\mathbf{y}}(\nu))^{-\frac{1}{2}}$ in the definition of the SCM.

Lemma 3.3. *Under Assumptions 3.1, 3.3 and 3.4, we have*

$$\max_{\nu \in \mathcal{V}_N} \left\|\operatorname{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right) - \mathbf{S}_{\mathbf{v}}(\nu)\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.34}$$

as well as

$$\max_{\nu \in \mathcal{V}_N} \left\|\operatorname{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.35}$$

Proof. To prove (3.34), we establish successively

$$\max_{\nu \in \mathcal{V}_N} \left\|\operatorname{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right) - \operatorname{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right)\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.36}$$

as well as

$$\max_{\nu \in \mathcal{V}_N} \left\|\operatorname{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right) - \mathbf{S}_{\mathbf{v}}(\nu)\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.37}$$

Using (3.20), we have the bound

$$\max_{\nu \in \mathcal{V}_N} \left\|\operatorname{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right) - \operatorname{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right)\right\|_2 \leq \Delta_1 + \Delta_2,$$

with

$$\begin{aligned}\Delta_1 = &\max_{\nu \in \mathcal{V}_N} \left\|\hat{\mathbf{S}}_{\mathbf{y}}(\nu) - \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}}\right\|_2 \\ &\xrightarrow[M\to\infty]{a.s.} 0,\end{aligned}$$

and

$$\Delta_2 = \max_{\nu \in \mathcal{V}_N} \max_{m=1,\dots,M} \left|\left[\mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} - \mathbf{I}_M\right) \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}}\right]_{m,m}\right|.$$

Denoting $\mathbf{u}_m(\nu) = \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}}\mathbf{e}_m$, where $\mathbf{e}_m$ is the m-th vector of the canonical basis of $\mathbb{C}^M$, as well as

$\mathbf{x}_1(\nu), \ldots, \mathbf{x}_{B+1}(\nu)$ the i.i.d. $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{I}_M)$ column vectors of $\mathbf{X}(\nu)$, we have for all $t > 0$,

$$\mathbb{P}\left(\Delta_2 > t\right) \leq \sum_{\nu \in \mathcal{V}_N} \sum_{m=1}^{M} \mathbb{P}\left(\left|\frac{1}{B+1} \sum_{b=1}^{B+1} \left|\mathbf{u}_m(\nu)^* \mathbf{x}_b(\nu)\right|^2 - \|\mathbf{u}_m(\nu)\|_2^2\right| > t\right).$$

From Assumption 3.1, Assumption 3.2 and condition (3.11) from Assumption 3.3, we have

$$0 < \inf_{M \geq 1} \min_{m=1,\ldots,M} \min_{\nu \in [0,1]} \|\mathbf{u}_m(\nu)\|_2 \leq \sup_{M \geq 1} \max_{m=1,\ldots,M} \max_{\nu \in [0,1]} \|\mathbf{u}_m(\nu)\|_2 < \infty.$$

Using in the statement of Lemma 3.1

$$\mathbf{z} = \left(\mathbf{x}_1(\nu)^T, \ldots, \mathbf{x}_{B+1}(\nu)^T\right)^T \sim \mathcal{N}_{\mathbb{C}^{M(B+1)}}(\mathbf{0}, \mathbf{I}_{M(B+1)})$$

and $\mathbf{\Xi}$ as the $M(B+1) \times M(B+1)$ block-diagonal matrix

$$\mathbf{\Xi} = \frac{\mathbf{I}_{B+1} \otimes (\mathbf{u}_m(\nu)\mathbf{u}_m(\nu)^*)}{B+1}$$

with $\otimes$ denoting the Kronecker product, we obtain

$$\mathbb{P}\left(\Delta_2 > t\right) \leq 2MN \max_{\nu \in \mathcal{V}_N} \exp\left(-C \min\left\{\frac{Bt^2}{\|\mathbf{u}_m(\nu)\|_2^4}, \frac{Bt}{\|\mathbf{u}_m(\nu)\|_2^2}\right\}\right)$$

where $C > 0$ is a constant independent of M, which in turn implies that

$$\Delta_2 \xrightarrow[M \to \infty]{a.s.} 0$$

and that (3.36) holds. To check (3.37), we use Assumption 3.3 eq. (3.12) to get that

$$\begin{aligned}\max_{\nu \in \mathcal{V}_N} \|\mathrm{dg}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^*\right)\|_2 &= \max_{\nu \in \mathcal{V}_N} \max_{m=1,\ldots,M} \|\mathbf{h}_m(\nu)\|_2^2 \\ &\xrightarrow[M \to \infty]{} 0\end{aligned}$$

and from the fact that

$$\mathrm{dg}\left(\mathbf{S}_{\mathbf{y}}(\nu)\right) = \mathrm{dg}\left(\mathbf{H}(\nu)\mathbf{H}(\nu)^*\right) + \mathbf{S}_{\mathbf{v}}(\nu)$$

we obtain (3.37) and, in turn, (3.34).

To prove (3.35), we write (using that $|\sqrt{a} - \sqrt{b}| < \sqrt{|a-b|}$ for $a, b > 0$)

$$\max_{\nu \in \mathcal{V}_N} \left\|\mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\right\|_2 \leq \max_{\nu \in \mathcal{V}_N} \max_{m=1,\ldots,M} \sqrt{\frac{|[\hat{\mathbf{S}}_{\mathbf{y}}(\nu)]_{m,m} - s_m(\nu)|}{[\hat{\mathbf{S}}_{\mathbf{y}}(\nu)]_{m,m}\ s_m(\nu)}}.$$

From Assumption 3.2, there exists $\epsilon > 0$ such that

$$\inf_{M \geq 1} \min_{m=1,\ldots,M} \min_{\nu \in \mathcal{V}_N} s_m(\nu) \geq \epsilon > 0.$$

Using (3.34) and denoting

$$\Delta = \max_{\nu \in \mathcal{V}_N} \left\|\mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right) - \mathbf{S}_{\mathbf{v}}(\nu)\right\|_2$$

we have that

$$\max_{\nu \in \mathcal{V}_N} \left\| \mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \right\|_2 \leq \sqrt{\frac{\Delta}{\epsilon\,(\epsilon - \Delta)}}$$

with probability one for all large M, which proves (3.35). □

We also need the following lemma on the boundedness of the matrix $\hat{\mathbf{S}}_{\mathbf{y}}(\nu)$.

Lemma 3.4. *Under Assumptions 3.1, 3.3 and 3.4, we have*

$$\limsup_{M\to\infty} \max_{\nu \in \mathcal{V}_M} \left\| \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \right\|_2 < \infty$$

with probability one.

Proof. From (3.20), we have

$$\limsup_{M\to\infty} \max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \right\|_2 \leq \limsup_{M\to\infty} \max_{\nu \in \mathcal{V}_N} \left\| \mathbf{S}_{\mathbf{y}}(\nu) \right\|_2 \max_{\nu \in \mathcal{V}_N} \left\| \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \right\|_2.$$

From Assumptions 3.1 and 3.3, it is clear that

$$\sup_{M\geq 1} \max_{\nu \in \mathcal{V}_N} \left\| \mathbf{S}_{\mathbf{y}}(\nu) \right\|_2 < \infty.$$

Finally, from Lemma 3.2 and the remarks below this lemma, we have

$$\limsup_{M\to\infty} \max_{\nu \in \mathcal{V}_N} \left\| \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \right\|_2 < \infty$$

with probability one, and Lemma 3.4 is proved. □

Equipped with Lemmas 3.3 and 3.4, we are now in a position to prove Theorem 3.3. Define

$$\tilde{\Delta} = \max_{\nu \in \mathcal{V}_N} \left\| \mathrm{dg}\left(\hat{\mathbf{S}}_{\mathbf{y}}(\nu)\right)^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \right\|_2$$

and recall the definition of the random matrix $\mathbf{X}(\nu)$ in (3.20). Let us write

$$\hat{\mathbf{C}}_{\mathbf{y}}(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} = \Psi_1(\nu) + \Psi_2(\nu)$$

where the two error terms are defined by:

$$\begin{aligned}
\Psi_1(\nu) &= \hat{\mathbf{C}}_{\mathbf{y}}(\nu) - \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \\
\Psi_2(\nu) &= \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}} - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}}
\end{aligned}$$

which satisfy:

$$\max_{\nu \in \mathcal{V}_N} \left\| \Psi_1(\nu) \right\|_2 \leq \tilde{\Delta} \max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{y}}(\nu) \right\|_2 \left(\tilde{\Delta} + \frac{2}{\sqrt{\min_{\nu \in \mathcal{V}_N} \lambda_M\left(\mathbf{S}_{\mathbf{v}}(\nu)\right)}} \right)$$

and

$$\max_{\nu \in \mathcal{V}_N} \left\| \Psi_2(\nu) \right\|_2 \leq \frac{\max_{\nu \in \mathcal{V}_N} \left\| \hat{\mathbf{S}}_{\mathbf{y}}(\nu) - \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \mathbf{S}_{\mathbf{y}}(\nu)^{\frac{1}{2}} \right\|_2}{\min_{\nu \in \mathcal{V}_N} \lambda_M(\mathbf{S}_{\mathbf{v}}(\nu))}.$$

From Assumption 3.2, we have

$$\inf_{M\geq 1}\min_{\nu\in\mathcal{V}_N}\lambda_M\left(\mathbf{S}_{\mathbf{v}}(\nu)\right)>0.$$

Using Lemmas 3.3 and 3.4, we directly deduce that

$$\max_{\nu\in\mathcal{V}_N}\|\Psi_1(\nu)\|_2\xrightarrow[M\to\infty]{a.s.}0.$$

Likewise, using (3.20), we deduce that

$$\max_{\nu\in\mathcal{V}_N}\|\Psi_2(\nu)\|_2\xrightarrow[M\to\infty]{a.s.}0,$$

which concludes the proof of Theorem 3.3.

3.D Proof of Corollary 3.1, Corollary 3.2 and Proposition 3.1

3.D.1 Proof of Corollary 3.1

Write $\hat{\mathbf{S}}_{\mathbf{v}}(\nu)=\boldsymbol{\Sigma}_{\mathbf{v}}(\nu)\boldsymbol{\Sigma}_{\mathbf{v}}(\nu)^*$, and denote

$$\Delta_{\mathbf{v}}(\nu)=\left\|\boldsymbol{\Sigma}_{\mathbf{v}}(\nu)-\frac{1}{\sqrt{B+1}}\mathbf{S}_{\mathbf{v}}(\nu)^{1/2}\mathbf{Z}(\nu)\right\|_2.$$

Using the fact that for any two matrices $\mathbf{A},\mathbf{B}$ of appropriate dimensions, we have

$$\mathbf{A}\mathbf{A}^*-\mathbf{B}\mathbf{B}^*=(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*+(\mathbf{A}-\mathbf{B})\mathbf{B}^*+\mathbf{B}(\mathbf{A}-\mathbf{B})^*$$

and

$$\|\mathbf{A}\mathbf{B}\|_2\leq\|\mathbf{A}\|_2\|\mathbf{B}\|_2$$

we see that

$$\left\|\hat{\mathbf{S}}_{\mathbf{v}}(\nu)-\frac{1}{B+1}\mathbf{S}_{\mathbf{v}}(\nu)^{1/2}\mathbf{Z}(\nu)\mathbf{Z}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{1/2}\right\|_2\leq\Delta_{\mathbf{v}}(\nu)\left(\Delta_{\mathbf{v}}(\nu)+2\sqrt{\frac{\|\mathbf{S}_{\mathbf{v}}(\nu)\|_2}{B+1}}\|\mathbf{Z}(\nu)\|_2\right).$$

Assumption 3.1 implies that

$$\sup_{M\geq 1}\max_{\nu\in[0,1]}\|\mathbf{S}_{\mathbf{v}}(\nu)\|_2<\infty$$

while from Lemma 3.2 of Appendix 3.A, since $\mathbf{Z}(\nu)$ has i.i.d. complex Gaussian entries,

$$\limsup_{M\to\infty}\max_{\nu\in\mathcal{V}_N}\frac{\|\mathbf{Z}(\nu)\|_2}{\sqrt{B+1}}<\infty \tag{3.38}$$

with probability one. This concludes the proof of Corollary 3.1.

3.D.2 Proof of Corollary 3.2

The proof of Corollary 3.2 is similar to that of Corollary 3.1. Denoting $\Delta_{\mathbf{u}}(\nu)=\|\boldsymbol{\Sigma}_{\mathbf{u}}(\nu)-\mathbf{H}(\nu)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\|_2$, and noticing that $\sup_{M\geq 1}\max_{\nu\in[0,1]}\|\mathbf{H}(\nu)\|_2<\infty$ from Assumption 3.3

eq. (3.11), we obtain that

$$\max_{\nu\in\mathcal{V}_N}\left\|\hat{\mathbf{S}}_{\mathbf{u}}(\nu)-\mathbf{H}(\nu)\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)^*\mathbf{H}(\nu)^*\right\|_2 \leq \max_{\nu\in\mathcal{V}_N}\Delta_{\mathbf{u}}(\nu)\left(\Delta_{\mathbf{u}}(\nu)+2\left\|\mathbf{H}(\nu)\right\|_2\left\|\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\right\|_2\right), \xrightarrow[M\to\infty]{a.s.} 0.$$

Since K is fixed with respect to M from Assumption 3.4, we also have

$$\max_{\nu\in\mathcal{V}_N}\left\|\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)^*-\mathbf{I}_M\right\|_2 \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.39}$$

using Lemma 3.1, which proves Corollary 3.2.

3.D.3 Proof of Proposition 3.1

To prove Proposition 3.1, let us write

$$\mathbf{Y}(\nu)=\mathbf{H}(\nu)\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)+\frac{\mathbf{S}_{\mathbf{v}}(\nu)^{1/2}\mathbf{Z}(\nu)}{\sqrt{B+1}}.$$

Then, from (3.38) and (3.39), we have

$$\limsup_{M\to\infty}\max_{\nu\in\mathcal{V}_N}\left\|\mathbf{Y}(\nu)\right\|_2<\infty$$

with probability one, which implies the following convergence

$$\max_{\nu\in\mathcal{V}_N}\left\|\hat{\mathbf{S}}_{\mathbf{y}}(\nu)-\mathbf{Y}(\nu)\mathbf{Y}(\nu)^*\right\|_2 \leq \max_{\nu\in\mathcal{V}_N}\left(\Delta_{\mathbf{u}}(\nu)+\Delta_{\mathbf{v}}(\nu)\right)\left(\Delta_{\mathbf{u}}(\nu)+\Delta_{\mathbf{v}}(\nu)+2\left\|\mathbf{Y}(\nu)\right\|_2\right) \xrightarrow[M\to\infty]{a.s.} 0.$$

Finally, since the columns of $\sqrt{B+1}\mathbf{Y}(\nu)$ are i.i.d. $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0},\mathbf{S}_{\mathbf{y}}(\nu))$ with $\mathbf{S}_{\mathbf{y}}(\nu)=\mathbf{H}(\nu)\mathbf{H}(\nu)^*+\mathbf{S}_{\mathbf{v}}(\nu)$, it follows that

$$\mathbf{Y}(\nu)=\mathbf{S}_{\mathbf{y}}(\nu)^{1/2}\frac{\mathbf{X}(\nu)}{\sqrt{B+1}}$$

for some $M\times(B+1)$ matrix $\mathbf{X}(\nu)$ having i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries and the proof of Proposition 3.1 is complete.

3.E Proof of Corollary 3.3

We first prove that all the eigenvalues of the SCM asymptotically concentrate in a compact set with probability one for all large M. Indeed, considering

$$\mathbf{W}(\nu)=\mathbf{\Xi}(\nu)^{\frac{1}{2}}\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\mathbf{\Xi}(\nu)^{\frac{1}{2}}$$

defined through Theorem 3.3 and using Lemma 3.2 in conjunction with Borel-Cantelli lemma, we deduce that there exists constants C_1, C_2 such that

$$\liminf_{M\to\infty}\min_{\nu\in\mathcal{V}_N}\lambda_M\left(\mathbf{W}(\nu)\right)\geq C_1(1-\sqrt{c})^2$$

and

$$\limsup_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1\left(\mathbf{W}(\nu)\right) \le C_2(1+\sqrt{c})^2 \tag{3.40}$$

with probability one, where C_1, C_2 satisfy, thanks to Assumption 3.3,

$$0 < C_1 < 1 = \inf_{M\ge 1} \min_{\nu\in\mathcal{V}_N} \lambda_M\left(\boldsymbol{\Xi}(\nu)\right)$$

and

$$\sup_{M\ge 1} \max_{\nu\in\mathcal{V}_N} \lambda_M\left(\boldsymbol{\Xi}(\nu)\right) < C_2 < \infty.$$

Using (3.22), we obtain similarly

$$\liminf_{M\to\infty} \min_{\nu\in\mathcal{V}_N} \lambda_M\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right) \ge C_1(1-\sqrt{c})^2$$

and

$$\limsup_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right) \le C_2(1+\sqrt{c})^2$$

with probability one. Let $0<\epsilon<\frac{C_1}{2}(1-\sqrt{c})^2$ and $h\in\mathcal{C}_c^1(\mathbb{R})$ such that

$$h(\lambda) = \begin{cases} 1 & \text{if } \lambda\in\left[C_1(1-\sqrt{c})^2-\epsilon, C_2(1+\sqrt{c})^2+\epsilon\right] \\ 0 & \text{if } \lambda\notin\left[C_1(1-\sqrt{c})^2-2\epsilon, C_2(1+\sqrt{c})^2+2\epsilon\right] \end{cases}.$$

Then it follows that

$$\max_{\nu\in\mathcal{V}_N} \left|L_\varphi(\nu) - L_{\varphi h}(\nu)\right| \xrightarrow[M\to\infty]{a.s.} 0.$$

Thus, without loss of generality, we may assume for the remainder of the proof that $\varphi\in\mathcal{C}_c^1\left((0,+\infty)\right)$. Using (3.22), we deduce that

$$\max_{\nu\in\mathcal{V}_N} \frac{1}{M}\sum_{m=1}^{M}\left|\varphi\left(\lambda_m\left(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)\right)\right) - \varphi\left(\lambda_m\left(\mathbf{W}(\nu)\right)\right)\right| \xrightarrow[M\to\infty]{a.s.} 0.$$

Next, consider the two functions

$$\hat{m}(z,\nu) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m\left(\mathbf{W}(\nu)\right)-z} = \int_{\mathbb{R}}\frac{\mathrm{d}\hat{\mu}(\lambda,\nu)}{\lambda-z}$$

and

$$\tilde{m}(z,\nu) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)-z} = \int_{\mathbb{R}}\frac{\mathrm{d}\tilde{\mu}(\lambda,\nu)}{\lambda-z}$$

defined for all $z\in\mathbb{C}^+ := \{\zeta\in\mathbb{C} : \mathrm{Im}(\zeta)>0\}$, and where for all Borel set $A\subset\mathbb{R}$,

$$\hat{\mu}(A,\nu) = \frac{1}{M}\sum_{m=1}^{M}\delta_{\lambda_m(\mathbf{W}(\nu))}(A)$$

and

$$\tilde{\mu}(A,\nu) = \frac{1}{M}\sum_{m=1}^{M} \delta_{\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)}(A)$$

denote the empirical eigenvalue distributions of the matrices $\mathbf{W}(\nu)$ and $\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}$ respectively, and δ_x is the Dirac measure at point x. Functions $z \mapsto \hat{m}(z,\nu)$ and $z \mapsto \tilde{m}(z,\nu)$ coincide with the Stieltjes transforms of the measures $\hat{\mu}(.,\nu)$ and $\tilde{\mu}(.,\nu)$ respectively (see [89] for a review of the main properties of the Stieltjes transform). Since

$$\hat{m}(z,\nu) - \tilde{m}(z,\nu) = \frac{1}{M}\mathrm{tr}\,\left((\mathbf{W}(\nu) - z\mathbf{I})^{-1} - \left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} - z\mathbf{I}\right)^{-1}\right)$$

and using the fact that $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B}-\mathbf{A})\mathbf{B}^{-1}$ for non-singular matrices $\mathbf{A},\mathbf{B}$, we have

$$|\hat{m}(z,\nu) - \tilde{m}(z,\nu)| \leq \frac{1}{|\mathrm{Im}(z)|^2}\frac{K}{M}\left\|\mathbf{H}(\nu)\right\|_2^2 \left\|\mathbf{S}_{\mathbf{v}}(\nu)^{-1}\right\|_2 \left\|\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right\|_2$$

it follows from Assumptions 3.2, 3.3, 3.4 and Lemma 3.2 that

$$\max_{\nu\in\mathcal{V}_N} |\hat{m}(z,\nu) - \tilde{m}(z,\nu)| \xrightarrow[M\to\infty]{a.s.} 0 \tag{3.41}$$

for all $z \in \mathbb{C}^+$. In the following, we fix a realization in an event of probability one for which (3.41) holds for all $z \in \mathbb{C}^+$ and consider

$$\nu^* \in \operatorname*{argmax}_{\nu\in\mathcal{V}_N}\left|\int_{\mathbb{R}} \varphi(\lambda)\mathrm{d}\hat{\mu}(\lambda,\nu) - \int_{\mathbb{R}} \varphi(\lambda)\mathrm{d}\tilde{\mu}(\lambda,\nu)\right|.$$

Then $|\hat{m}(z,\nu^*) - \tilde{m}(z,\nu^*)| \to 0$ as $M \to \infty$, for all $z \in \mathbb{C}^+$. From the fact that the pointwise convergence on $\mathbb{C}^+$ of a sequence of Stieltjes transforms is equivalent to the weak convergence of the related sequence of probability measures (see e.g. [89, Ex.2.4.10]), we deduce that

$$\max_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\sum_{m=1}^{M}\left(\varphi\left(\lambda_m\left(\mathbf{W}(\nu)\right)\right) - \varphi\left(\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right)\right)\right| = \left|\int_{\mathbb{R}}\varphi(\lambda)\mathrm{d}\hat{\mu}(\lambda,\nu^*) - \int_{\mathbb{R}}\varphi(\lambda)\mathrm{d}\tilde{\mu}(\lambda,\nu^*)\right| \xrightarrow[M\to\infty]{} 0.$$

To conclude the proof of Corollary 3.3, it remains to prove that

$$\max_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\sum_{m=1}^{M}\varphi\left(\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right) - \int_{\mathbb{R}}\varphi(\lambda)f(\lambda)\mathrm{d}\lambda\right| \xrightarrow[M\to\infty]{a.s.} 0.$$

Consider the decomposition

$$\max_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\sum_{m=1}^{M}\varphi\left(\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right) - \int_{\mathbb{R}}\varphi(\lambda)f(\lambda)\mathrm{d}\lambda\right| \leq \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = \max_{\nu\in\mathcal{V}_N}\left|\frac{1}{M}\sum_{m=1}^{M}\left(\varphi\left(\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right) - \mathbb{E}\left[\varphi\left(\lambda_m\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right)\right)\right]\right)\right|$$

and

$$\Delta_2 = \left| \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left[\varphi\left(\lambda_m \left(\frac{\mathbf{X}(0)\mathbf{X}(0)^*}{B+1} \right) \right) \right] - \int_{\mathbb{R}} \varphi(\lambda) f(\lambda) \mathrm{d}\lambda \right|.$$

Using the concentration inequality of [33, Cor. 1.8(b)], it is straightforward to show that

$$\Delta_1 \xrightarrow[M\to\infty]{a.s.} 0.$$

Moreover, using again the properties of the Stieltjes transform, it can be deduced from e.g. [36] that

$$\Delta_2 \xrightarrow[M\to\infty]{} 0.$$

This concludes the proof of Corollary 3.3.

3.F Proof of Proposition 3.2

Convergences (3.25), (3.26) and (3.27) are straightforward consequences of (3.22) and the results of [7, Th. 1.1] on the behaviour of the largest eigenvalues for the so-called multiplicative spike model random matrices. To prove (3.28), we use the bound

$$\lambda_1\left(\mathbf{W}(\nu)\right) \leq \lambda_1\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right) \lambda_1\left(\boldsymbol{\Xi}(\nu)\right)$$

Then, from the fact that $\gamma_\infty = 0$ and Lemma 3.2, we finally obtain

$$\begin{aligned}\limsup_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1\left(\mathbf{W}(\nu)\right) &\leq \limsup_{M\to\infty} \max_{\nu\in\mathcal{V}_N} \lambda_1\left(\frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1}\right) \\ &\leq \left(1+\sqrt{c}\right)^2.\end{aligned}$$

The proof is concluded by invoking again convergence (3.22).

Chapter 4

Extension to the case of non-vanishing SNR per sensor

4.1 Introduction

In this short chapter, we show that it is possible to slightly weaken one of the assumptions used in Chapter 3. However, we will see that the corresponding results in this chapter are less natural to write and will still require other conditions. In particular, we are interested in Assumption 3.3 equation (3.12) that we recall here.

Assumption 3.3. *The MIMO filter coefficient matrices are such that*

$$\sup_{M\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)\left\|\mathbf{H}_k\right\|_2<\infty$$

and

$$\lim_{M\to\infty}\max_{m=1,\ldots,M}\max_{\nu\in[0,1]}\left\|\mathbf{h}^m(\nu)\right\|_2=0.$$

Intuitively, we recall that $\|\mathbf{h}_m(\nu)\|_2^2=\sum_{k=1}^K|(\mathbf{H})_{m,k}(\nu)|^2$ can be represented as the total power received by the sensor m from the K–dimensional signal at frequency ν. Assumption 3.3 assumes that the power received on each sensor from the signal is asymptotically zero, and no relevant information about the signal can be extracted by each sensor. The detection becomes achievable only when the observations from the M sensors are combined, as shown in the previous chapter. Suppose now that for some reason, the signal power received at some sensors does not converge to zero, so that some of the sensors can receive asymptotically a strictly positive total power, ie. for some $m\geq 1$ and $\nu\in[0,1]$

$$\lim_{m\to+\infty}\|\mathbf{h}_m(\nu)\|_2>0. \tag{4.1}$$

We modify Assumption 3.3 according to this more general context.

Assumption 3.3b. *The MIMO filter coefficient matrices are such that*

$$\sup_{M\geq 1}\sum_{k\in\mathbb{Z}}(1+|k|)\left\|\mathbf{H}_k\right\|_2<\infty$$

Moreover, for each $\nu\in[0,1]$ there exists a sequence $(d_{\mathbf{u},\infty,m}(\nu))_{m\geq 1}$ such that

$$\sup_{\nu\in[0,1]}\operatorname{card}\left\{d_{\mathbf{u},\infty,m}(\nu)\neq 0\right\}<+\infty$$

and

$$\|\mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*) - \mathrm{dg}(d_{\mathbf{u},\infty,m}(\nu), m \in [M])\| \xrightarrow{N\to+\infty} 0$$

Assumption 3.3b means that $\mathbf{D}_{\mathbf{u},\infty}(\nu) := \mathrm{dg}(d_{\mathbf{u},\infty,m}(\nu), m \in [M])$ has a finite number of non-zero entries, so its rank remains finite when $N \to +\infty$. This is not the case of $\mathbf{D}_u(\nu) := \mathrm{dg}(\mathbf{H}(\nu)\mathbf{H}(\nu)^*)$, even though $\mathbf{H}(\nu)\mathbf{H}(\nu)^*$ is of finite rank K. Note that relying on Assumption 3.3b instead of Assumption 3.3 is not in contradiction with (3.14) which states that the SNR $\frac{\mathbb{E}\|\mathbf{u}_n\|_2^2}{\mathbb{E}\|\mathbf{v}_n\|_2^2}$ converges to zero at speed $\frac{1}{M}$, as long as only a finite number of indices m satisfies (4.1).

4.2 Statement of the results under our new assumption

We start by showing that the conclusion of Theorem 3.3 from Chapter 3 is slightly modified under the new Assumption 3.3b.

Theorem 4.1. *Under Assumptions 3.1, 3.2, 3.3b and 3.4, for each* $\nu \in \mathcal{V}_N$*, there exist an* $M \times (B+1)$ *random matrix* $\mathbf{X}(\nu)$ *with iid* $\mathcal{N}_{\mathbb{C}}(0,1)$ *entries such that*

$$\max_{\nu\in\mathcal{V}_N} \left\| \hat{\mathbf{C}}_{\mathbf{y}}(\nu) - \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \frac{\mathbf{X}(\nu)\mathbf{X}(\nu)^*}{B+1} \boldsymbol{\Xi}(\nu)^{\frac{1}{2}} \right\| \xrightarrow[N\to+\infty]{} 0$$

where

$$\boldsymbol{\Xi}(\nu) = \mathbf{I}_M + (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}} (\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)) (\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu))^{-\frac{1}{2}}$$

Proof. For the entire proof, we fix $\nu \in \mathcal{V}_N$ and omit the dependence in ν for all the quantities (for instance $\mathbf{C}_{\mathbf{y}}$ represents $\mathbf{C}_{\mathbf{y}}(\nu)$). The true spectral coherency matrix $\mathbf{C}_{\mathbf{y}}$ is equal to

$$\mathrm{dg}(\mathbf{S}_{\mathbf{y}})^{-\frac{1}{2}}\mathbf{S}_{\mathbf{y}}\mathrm{dg}(\mathbf{S}_{\mathbf{y}})^{-\frac{1}{2}} = (\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-\frac{1}{2}}(\mathbf{H}\mathbf{H}^* + \mathbf{S}_{\mathbf{v}})(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-\frac{1}{2}}$$

Distributing the inner sum, and defining $\mathbf{D} = \mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}}$, we get:

$$(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-\frac{1}{2}}(\mathbf{H}\mathbf{H}^* + \mathbf{S}_{\mathbf{v}})(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^*\mathbf{D}^{-\frac{1}{2}} + \mathbf{S}_{\mathbf{v}}(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-1} \tag{4.2}$$

Using the resolvent identity $(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{B}^{-1} - (\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\mathbf{B}^{-1}$ we obtain that

$$(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-1} = \mathbf{S}_{\mathbf{v}}^{-1} - \mathbf{S}_{\mathbf{v}}^{-1}\mathbf{D}_{\mathbf{u}}(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-1}$$

so denoting $\mathbf{D} = \mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}}$ and multiplying by $\mathbf{S}_{\mathbf{v}}$:

$$\mathbf{S}_{\mathbf{v}}(\mathbf{D}_{\mathbf{u}} + \mathbf{S}_{\mathbf{v}})^{-1} = \mathbf{I}_M - \mathbf{D}_{\mathbf{u}}\mathbf{D}^{-1} = \mathbf{I}_M - \mathbf{D}^{-\frac{1}{2}}\mathbf{D}_{\mathbf{u}}\mathbf{D}^{-\frac{1}{2}} \tag{4.3}$$

Combining (4.2) and (4.3) we get:

$$\begin{aligned}\mathrm{dg}(\mathbf{S})^{-\frac{1}{2}}\mathbf{S}\mathrm{dg}(\mathbf{S})^{-\frac{1}{2}} &= \mathbf{D}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^*\mathbf{D}^{-\frac{1}{2}} + \mathbf{I}_M - \mathbf{D}^{-\frac{1}{2}}\mathbf{D}_{\mathbf{u}}\mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{I}_M + \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{H}\mathbf{H}^* - \mathbf{D}_{\mathbf{u}}\right)\mathbf{D}^{-\frac{1}{2}}\end{aligned}$$

It remains to use the fact that $\sup_{\nu\in[0,1]} \|\mathbf{D}_{\mathbf{u}}(\nu) - \mathbf{D}_{\mathbf{u},\infty}(\nu)\| \to 0$ and follow the same steps as in the proof of Theorem 3.3 to end the proof. ∎

Remark 4.1. *We see that imposing Assumption 3.3 implies that all the entries of* $\mathbf{D}_{\mathbf{u}}$ *converge towards* 0*, so* $\mathbf{D}_{\mathbf{u},\infty}(\nu) = 0$*, and* $\boldsymbol{\Xi}(\nu)$ *becomes equal to* $\mathbf{I}_M + \mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}\mathbf{H}(\nu)\mathbf{H}(\nu)^*\mathbf{S}_{\mathbf{v}}(\nu)^{-\frac{1}{2}}$ *which is exactly the matrix obtained in Theorem 3.3.*

With Theorem 4.1 at hand, we know that the behaviour of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ are governed by the matrix $\boldsymbol{\Xi}(\nu)$, and in particular by the perturbation part that we denote as $\boldsymbol{\Omega}(\nu)$:

$$\boldsymbol{\Omega}(\nu) = \left(\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu)\right)^{-\frac{1}{2}} \left(\mathbf{H}(\nu)\mathbf{H}(\nu)^* - \mathbf{D}_{\mathbf{u},\infty}(\nu)\right) \left(\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu)\right)^{-\frac{1}{2}}$$

We will now state some sufficient conditions on the eigenvalues of the quantities involved in $\boldsymbol{\Omega}(\nu)$ so that asymptotically, as in Chapter 3, some eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ escape the support of the Marcenko-Pastur distribution, which makes the detection of the signal possible. Results on spike models from [7] provide a sufficient condition to detect eigenvalues leaving the Marcenko-Pastur distribution support as long as the perturbation matrix $\boldsymbol{\Omega}(\nu)$ rank remains finite. This is why it was necessary to approximate $\mathbf{D}_{\mathbf{u}}(\nu)$ by $\mathbf{D}_{\mathbf{u},\infty}(\nu)$ to ensure that $\boldsymbol{\Xi}(\nu)$ remains asymptotically a finite rank perturbation of the identity. Denote $\lambda_1 \geq \ldots \geq \lambda_M$ the ordered eigenvalues of $\Omega(\nu)$, and define

$$\begin{aligned} \gamma^+(\nu) &= \lambda_1(\Omega(\nu)) \\ \gamma^-(\nu) &= \lambda_M(\Omega(\nu)) \end{aligned}$$

Let $\nu_N^{+,*}$ and $\nu_N^{-,*}$ such that

$$\begin{aligned} \nu_N^{+,*} &\in \operatorname*{argmax}_{\nu \in \mathcal{V}_N} \gamma^+(\nu) \\ \nu_N^{-,*} &\in \operatorname*{argmin}_{\nu \in \mathcal{V}_N} \gamma^-(\nu) \end{aligned}$$

We now make the following assumption.

Assumption 4.1. *There exist constants γ_∞^+ and γ_∞^- such that*

$$\begin{aligned} \gamma^+(\nu_N^{+,*}) &\xrightarrow[N\to+\infty]{} \gamma_\infty^+ \\ \gamma^-(\nu_N^{-,*}) &\xrightarrow[N\to+\infty]{} \gamma_\infty^- \end{aligned}$$

We motivate the definition of these two quantities. For each ν, the matrix $\boldsymbol{\Omega}(\nu)$ perturbs the Marcenko-Pastur behaviour of the eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ in two possible ways: the smallest (negative) eigenvalue of $\boldsymbol{\Omega}(\nu)$ will potentially create an eigenvalue leaving the support of the Marcenko-Pastur distribution on the left, and symmetrically the largest (positive) eigenvalue of $\boldsymbol{\Omega}(\nu)$ will potentially create an eigenvalue leaving the support of the Marcenko-Pastur distribution on the right. The frequencies ν_N^- and ν_N^+ represent the frequencies for which these perturbations are maximal. In anticipation of the detection problem, these frequencies represent those where the signal is most likely to be detected. We note that $\gamma^+(\nu_N^{+,*})$ and $\gamma^-(\nu_N^{-,*})$ can be interpreted as a certain SNR metric in the frequency domain, as it was the case in Chapter 3, except that we now have two quantities, one for the positive potential spikes and one for the negative potential spikes. In contrast to the result from Chapter 3 it is now possible to observe spikes on both sides of the Marcenko-Pastur distribution. A detection test should therefore monitor eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ leaving the bulk of the distribution not only on the right-side but also on the left-side. Consider the following test statistics:

$$T_\epsilon = \mathbf{1}\left(\sup_{\nu\in\mathcal{V}_N} \lambda_1(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) > \lambda^+ + \epsilon \text{ or } \inf_{\nu\in\mathcal{V}_N} \lambda_M(\hat{\mathbf{C}}_{\mathbf{y}}(\nu)) < \lambda^- - \epsilon\right)$$

where $[M] = \{1, \ldots, M\}$. The following proposition proves the consistency of T_ϵ under $\mathcal{H}_0$ and the alternative $\mathcal{H}_1$.

Proposition 4.1. *For any $\epsilon > 0$, under Assumption 3.3b and Assumption 4.1,*

$$\mathbb{P}_{\mathcal{H}_0}\left(\lim_{N\to\infty} T_\epsilon = 0\right) = 1.$$

Moreover, under $\mathcal{H}_1$, if γ_∞^+ and γ_∞^- are such that $\gamma_\infty^+ > \sqrt{c}$ or $\gamma_\infty^- < -\sqrt{c}$, then for all $0 < \epsilon < \min(\phi(\gamma_\infty^+) - \lambda^+, \lambda^- - \phi(\gamma_\infty^-))$:

$$\mathbb{P}_{\mathcal{H}_1}\left(\lim_{N\to\infty} T_\epsilon = 1\right) = 1,$$

We omit the proof of this proposition at it is the same as the proof of Proposition 3.3 from Chapter 3. The condition $\gamma_\infty^+ > \sqrt{c}$ or $\gamma_\infty^- < -\sqrt{c}$ ensures that the SNR is high enough in the sense that the perturbation matrices $(\mathbf{\Omega}(\nu))_{\nu\in[0,1]}$ have asymptotically, at least for one frequency ν, an eigenvalue greater in absolute value than the threshold $\sqrt{c}$. Intuitively, under this condition, we will observe a spike in the spectrum of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$, which could be located on the right-side or the left-side of the Marcenko-Pastur distribution. As in Chapter 3, the location of the spike is driven by the values γ_∞^+ and γ_∞^-.

4.3 Numerical simulations

4.3.1 Definition of an alternative

We consider the same model as in Chapter 3 except that we now suppose that for exactly one sensor, the SNR remains $\mathcal{O}(1)$ (instead of being $\mathcal{O}(\frac{1}{M})$). The noise $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is generated as a Gaussian AR(1) process having the spectral density

$$s_v(\nu) = \frac{1}{|1-\theta \mathrm{e}^{-\mathrm{i}2\pi\nu}|^2},$$

for all $m = 1, \dots, M$, with $|\theta| < 1$. The useful signal $(\mathbf{u})_{n\in\mathbb{Z}}$, is also modeled as a rank one ($K = 1$) AR(1) process defined by:

$$\mathbf{H}_k = C\beta^k\left(1, \frac{1}{\sqrt{M}}, \dots, \frac{1}{\sqrt{M}}\right)^T \in \mathbb{C}^M$$

with $|\beta| < 1$ and C a constant which can be chosen to manually tune the SNR. This model is very close to the one used in Chapter 3 except that the first component is $\mathcal{O}(1)$ instead of $\mathcal{O}(\frac{1}{\sqrt{M}})$. In this case,

$$\mathbf{H}(\nu)\mathbf{H}^*(\nu) = \frac{C^2}{|1-\beta e^{-2i\pi\nu}|^2}\begin{pmatrix} 1 & \frac{1}{\sqrt{M}}\mathbf{1}_{M-1}^T \\ \frac{1}{\sqrt{M}}\mathbf{1}_{M-1} & \frac{1}{M}\mathbf{1}_{M-1}\mathbf{1}_{M-1}^T \end{pmatrix}$$

where $\mathbf{1}_M = (1, \dots, 1)^T \in \mathbb{R}^M$. Denoting $s_u(\nu) = \frac{C^2}{|1-\beta e^{-2i\pi\nu}|^2}$, this implies that

$$\mathbf{D}_{\mathbf{u},\infty}(\nu) = s_u(\nu)\begin{pmatrix} 1 & \mathbf{0}_{M-1}^T \\ \mathbf{0}_{M-1} & \mathbf{0}_{M-1}\mathbf{0}_{M-1}^T \end{pmatrix}$$

where $\mathbf{0}_M = (0, \dots, 0)^T \in \mathbb{R}^M$. We write:

$$\mathbf{H}(\nu)\mathbf{H}^*(\nu) - \mathbf{D}_{\mathbf{u},\infty}(\nu) = s_u(\nu)\begin{pmatrix} 0 & \frac{1}{\sqrt{M}}\mathbf{1}_{M-1}^T \\ \frac{1}{\sqrt{M}}\mathbf{1}_{M-1} & \frac{1}{M}\mathbf{1}_{M-1}\mathbf{1}_{M-1}^T \end{pmatrix}$$

and $\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu)$ is the diagonal matrix equal to:

$$\mathbf{D}_{\mathbf{u},\infty}(\nu) + \mathbf{S}_{\mathbf{v}}(\nu) = \begin{bmatrix} s_u(\nu) + s_v(\nu) & & & \\ & s_v(\nu) & & \\ & & \ddots & \\ & & & s_v(\nu) \end{bmatrix}$$

Denote $s(\nu) = s_u(\nu) + s_v(\nu)$. With this notation, it is clear that:

$$\mathbf{\Omega}(\nu) = \begin{pmatrix} 0 & \frac{s_u(\nu)}{\sqrt{s_v(\nu)s(\nu)}} \frac{1}{\sqrt{M}} \mathbf{1}_{M-1}^T \\ \frac{s_u(\nu)}{\sqrt{s(\nu)s_v(\nu)}} \frac{1}{\sqrt{M}} \mathbf{1}_{M-1} & \frac{s_u(\nu)}{s_v(\nu)} \frac{1}{M} \mathbf{1}_{M-1} \mathbf{1}_{M-1}^T \end{pmatrix}$$

To find the eigenvalues of $\mathbf{\Omega}(\nu)$, we first need to write the following Lemma 4.1.

Lemma 4.1. *For $a, b \neq 0$, the eigenvalues of $\mathbf{A} \in \mathbb{R}^{M \times M}$ defined by:*

$$\mathbf{A} = \begin{pmatrix} 0 & a\mathbf{1}_{M-1}^T \\ a\mathbf{1}_{M-1} & b\mathbf{1}_{M-1}\mathbf{1}_{M-1}^T \end{pmatrix}$$

are 0 with multiplicity $M-2$ while the largest and smallest eigenvalues are

$$\frac{(M-1)b}{2}\left(1 \pm \sqrt{1 + \frac{4a^2}{(M-1)b^2}}\right)$$

Proof. It can be checked that 0 is an eigenvalue with multiplicity $M-2$, with eigenvectors $(0, -1, 1, 0, \dots, 0), \dots, (0, -1, 0, \dots, 0, 1)$. It remains to find the two remaining eigenvalues. By symmetry of $\mathbf{A}$, it is natural to consider the eigenvector $\mathbf{x} = (v, 1, \dots, 1)$ where $v \neq 0$, and denote $\lambda \neq 0$ its associated eigenvalue. The identity $\mathbf{Ax} = \lambda \mathbf{x}$ provides the following system of equations:

$$\begin{cases} (M-1)a = \lambda v \\ va + (M-1)b = \lambda \\ \vdots \\ va + (M-1)b = \lambda \end{cases}$$

Using $v = \frac{(M-1)a}{\lambda}$, we get:

$$(M-1)\frac{a^2}{\lambda} + (M-1)b = \lambda$$

so

$$\lambda^2 - (M-1)b\lambda - (M-1)a^2 = 0$$

which gives two solutions:

$$\lambda_{\pm} = \frac{(M-1)b \pm \sqrt{(M-1)^2 b^2 + 4(M-1)a^2}}{2} = \frac{(M-1)b}{2}\left(1 \pm \sqrt{1 + \frac{4a^2}{(M-1)b^2}}\right)$$

□

Substituting a (respectively b) by $\frac{1}{\sqrt{M}} \frac{s_u(\nu)}{\sqrt{s(\nu)s_v(\nu)}}$ (respectively $\frac{1}{M}\frac{s_u(\nu)}{s_v(\nu)}$), Lemma 4.1 applied to

$\mathbf{\Omega}(\nu)$ provides:

$$\frac{4a^2}{(M-1)b^2} = 4\frac{s_v(\nu)}{s(\nu)}\frac{M}{(M-1)}$$
$$\frac{(M-1)b}{2} = \frac{(M-1)}{2M}\frac{s_u(\nu)}{s_v(\nu)}$$

so that the eigenvalues of $\mathbf{\Omega}(\nu)$ are 0 with multiplicity $M-2$ and

$$\frac{(M-1)}{2M}\frac{s_u(\nu)}{s_v(\nu)}\left(1 \pm \sqrt{1 + 4\frac{s_v(\nu)}{s_u(\nu)+s_v(\nu)}\frac{M}{(M-1)}}\right)$$

Asymptotically, the two non-zero eigenvalues become the largest (strictly positive) and the smallest (strictly negative) and converge towards:

$$\lambda_+(\nu) = \frac{1}{2}\frac{s_u(\nu)}{s_v(\nu)}\left(1 + \sqrt{1 + 4\frac{s_v(\nu)}{s_u(\nu)+s_v(\nu)}}\right)$$
$$\lambda_-(\nu) = \frac{1}{2}\frac{s_u(\nu)}{s_v(\nu)}\left(1 - \sqrt{1 + 4\frac{s_v(\nu)}{s_u(\nu)+s_v(\nu)}}\right)$$

Note that $\lambda_+(\nu)$ and $\lambda_-(\nu)$ converges to finite non null constants. In the case where for some ν we have $\lambda_+(\nu) > \sqrt{c}$ and $\lambda_-(\nu) < -\sqrt{c}$ we should see two spikes : one on the left associated with λ_- and one on the right associated with λ_+. The other eigenvalues are $\mathcal{O}(\frac{1}{M})$, so they will not create additional spikes.

4.3.2 Numerical results

In Figure 4.3.1 is represented for $N = 20000$, $M = 200$, $B = 2000$, $\theta = 0.7$ and $\beta = \frac{10}{18}$ the eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for $\nu = 0$. On the right graph of Figure 4.3.1, compared to the case under Assumption 3.3, we indeed observe an additional spike on the left-side of the distribution.

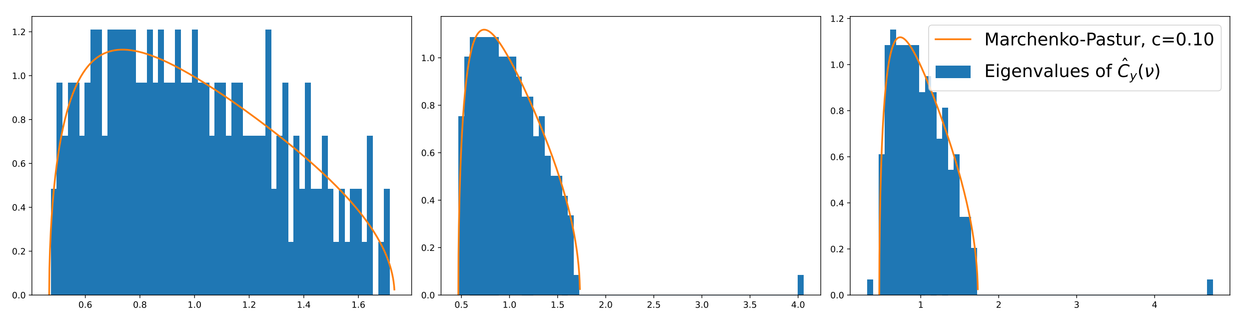


Figure 4.3.1: Eigenvalue distribution of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ ($\nu = 0$) vs the density of the Marcenko-Pastur distribution with parameter $c = 0.1$. $C = 0$ (no signal) on the left, $C = 0.1$ under Assumption 3.3 in the middle, $C = 0.1$ under Assumption 3.3b on the right.

Lastly, in Figure 4.3.2 is represented under Assumption 3.3 and Assumption 3.3b the largest and smallest eigenvalues of $\mathbf{\Omega}(\nu)$ as a function of ν in the case where $\theta = 0.7$, $\beta = \frac{10}{18}$, and $M = 100$. Concerning the largest eigenvalue, we see that the maximum perturbation happens for $\nu = 0$ (which justifies why we studied $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ for this frequency in Figure 4.3.1) and that the results for Assumption 3.3 and Assumption 3.3b approximately coincide. However, we observe a significant difference in the behaviour of the smallest eigenvalue: the smallest eigenvalue is strictly negative under Assumption

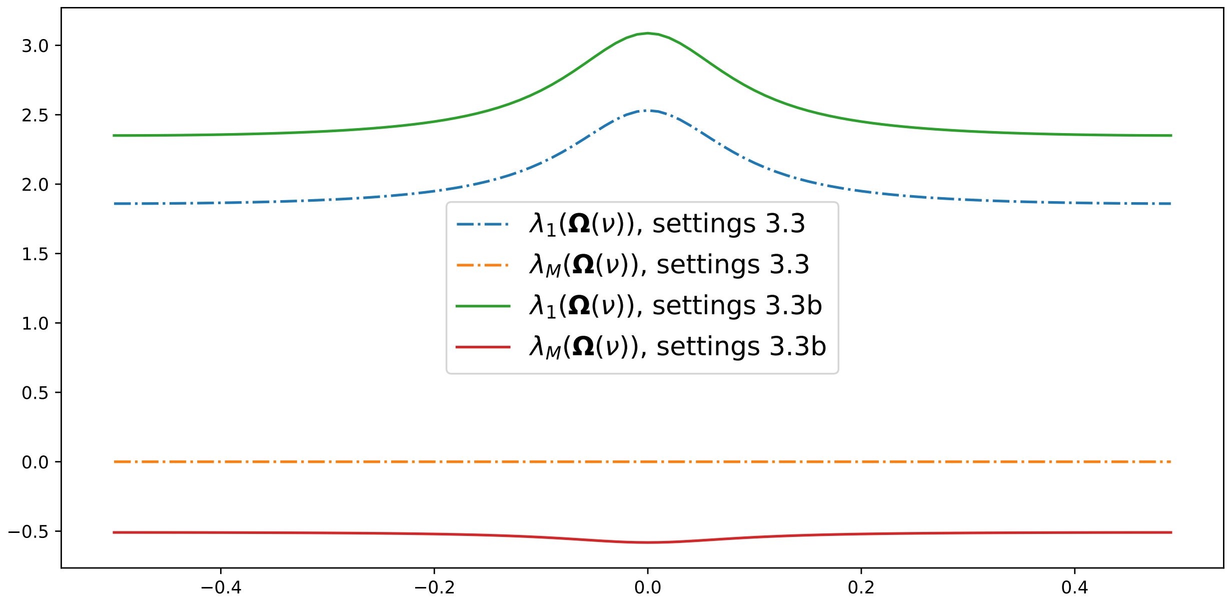


Figure 4.3.2: Largest and smallest eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}}(\nu)$ as a function of ν in Assumption 3.3 and Assumption 3.3b

3.3b while it is null under Assumption 3.3.

Part III

Maximum of the sample spectral coherency estimator

Chapter 5

On the asymptotic distribution of the maximum sample spectral coherence of Gaussian time series in the high dimensional regime

5.1 Introduction

5.1.1 The addressed problem and the results

We consider M jointly stationary complex Gaussian time series $(y_{1,n})_{n\in\mathbb{Z}}, \ldots, (y_{M,n})_{n\in\mathbb{Z}}$ and for all $i, j \in \{1, \ldots, M\}$, we denote by s_{ij} and c_{ij} the spectral density and spectral coherence between $(y_{i,n})_{n\in\mathbb{Z}}$ and $(y_{j,n})_{n\in\mathbb{Z}}$ given respectively by

$$s_{ij}(\nu) = \sum_{u\in\mathbb{Z}} r_{ij}(u)\mathrm{e}^{-\mathrm{i}2\pi u\nu}$$

and

$$c_{ij}(\nu) = \frac{s_{ij}(\nu)}{\sqrt{s_{ii}(\nu)s_{jj}(\nu)}}$$

for all $\nu \in [0, 1]$, where $r_{ij}(u) = \mathbb{E}[y_{i,n+u}\overline{y}_{j,n}]$. Assuming N observations $(y_{1,n})_{n=1,\ldots,N}, \ldots, (y_{M,n})_{n=1,\ldots,N}$ are available for each time series, we consider the frequency smoothed estimate $\hat{s}_{ij}$ of s_{ij} given by

$$\hat{s}_{ij}(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \xi_{y_i}\left(\nu + \frac{b}{N}\right)\overline{\xi_{y_j}\left(\nu + \frac{b}{N}\right)}, \tag{5.1}$$

where B is an even integer representing the smoothing span, and where

$$\xi_{y_i}(\nu) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} y_{i,n}\mathrm{e}^{-2\mathrm{i}\pi(n-1)\nu}$$

denotes the normalized Fourier transform of $(y_{i,n})_{n=1,\dots,N}$. The corresponding sample estimate of the spectral coherence is defined as

$$\hat{c}_{ij}(\nu) = \frac{\hat{s}_{ij}(\nu)}{\sqrt{\hat{s}_{ii}(\nu)\hat{s}_{jj}(\nu)}}.$$

Under the hypothesis

$$\mathcal{H}_0 : (y_{1,n})_{n\in\mathbb{Z}}, \dots, (y_{M,n})_{n\in\mathbb{Z}} \text{ are mutually uncorrelated,}$$

we evaluate the behaviour of the Maximum Sample Spectral Coherence (MSSC) defined by

$$\max_{1\le i<j\le M} \max_{\nu\in\mathcal{G}} |\hat{c}_{ij}(\nu)|$$

where

$$\mathcal{G} := \left\{ k\frac{B+1}{N} : k \in \mathbb{N}, 0 \le k \le \frac{N}{B+1} \right\}$$

is the subset of the Fourier frequencies

$$\mathcal{F} := \left\{ \frac{k}{N} : k \in \mathbb{N}, 0 \le k \le N-1 \right\}$$

with elements spaced by a distance $(B+1)/N$. Our study is conducted in the asymptotic regime where $M = M(N)$ and $B = B(N)$ are both functions of N such that for some $\rho \in (0,1)$, $M \asymp N^\rho$ and $B \asymp N^\rho$ as $N \to \infty$ [1] , while the ratio M/B converges to some constant $c \in (0, +\infty)$. It is established that, under $\mathcal{H}_0$ and proper assumptions on the time series $(y_{1,n})_{n\in\mathbb{Z}}, \dots, (y_{M,n})_{n\in\mathbb{Z}}$, for any $t \in \mathbb{R}$:

$$\mathbb{P}\left((B+1) \max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{ij}(\nu)|^2 \le t + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2} \right) \xrightarrow[N\to+\infty]{} e^{-e^{-t}} \tag{5.2}$$

where

$$\mathcal{I} := \{(i,j,\nu) : i,j \in [M] \text{ such that } i<j,\ \nu \in \mathcal{G}\} \tag{5.3}$$

with $[M] = \{1, \dots, M\}$.

In other words, under proper normalization and centering, $\max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{ij}(\nu)|^2$ follows asymptotically a Gumbel distribution (see [26] or [74] for a general theory of extreme value distributions).

5.1.2 Motivation

This chapter is motivated by the problem of testing the independence of a large number of Gaussian time series. Since hypothesis $\mathcal{H}_0$ can be equivalently formulated as

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |s_{ij}(\nu)|^2 = 0,$$

or by

$$\mathcal{H}_0 : \max_{1\le i<j\le M} \max_{\nu\in[0,1]} |c_{ij}(\nu)|^2 = 0,$$

this suggests to compute consistent estimators of these quantities, and test their closeness to zero.

[1]For two sequences $(x_n)_{n\ge1}, (y_n)_{n\ge1}$, we denote by $x_n \asymp y_n$ if there exists $k_1, k_2 > 0$ such that $k_1|y_n| \le |x_n| \le k_2|y_n|$ for all large n.

Our choice of the high-dimensional regime defined above is motivated as follows. Under mild assumptions on the memory of the time series $((y_{m,n})_{n\in\mathbb{Z}})_{m\geq 1}$, in the low-dimensional regime where $N \to +\infty$ and M is fixed, it can be shown that the sample spectral coherence matrix

$$\hat{\mathbf{C}}(\nu) = (\hat{c}_{i,j}(\nu))_{i,j=1,\ldots,M} \tag{5.4}$$

is a consistent estimate (in spectral norm for instance) of the spectral coherence matrix

$$\mathbf{C}(\nu) = (c_{i,j}(\nu))_{i,j=1,\ldots,M}$$

as long as $B \to +\infty$ and $B/N \to 0$ (up to some additional logarithmic terms). In practice, this asymptotic regime and the underlying predictions are relevant as long as the ratio M/N is small enough. If this condition is not met, test statistics based on $\hat{\mathbf{C}}(\nu)$ may be of delicate use, as the choice of the smoothing span B must meet the constraints $B \gg M$ (because B is supposed to converge towards $+\infty$) as well as $B \ll N$ (because B/N is supposed to converge towards 0). Nowadays, for many practical applications involving high dimensional signals and/or moderate sample size, the ratio M/N may not be small enough to be able to choose B to meet $B \gg M$ and $B \ll N$. In this situation, one may rely on the more relevant high dimensional regime in which M, B, N converge to infinity such that M/B converges to a positive constant while B/N converges to zero.

5.1.3 On the literature

Correlation tests using spectral approaches have been studied in several papers, see e.g. [93], [25] and the references therein.

More recently, an approach similar to the one of this chapter has been explored in [97], where the maximum of the sample spectral coherence, when using lag-window estimates of the spectral density, is studied. In the low-dimensional regime where M is fixed and $N \to \infty$, it is proved that the distribution of such statistic under $\mathcal{H}_0$, after proper centering and normalization, converges to the Gumbel distribution. We also mention other related papers exploring the asymptotic behaviour of various spectral density estimates in the low-dimensional regime: [96], [81], [83], [56] and [58].

In the high-dimensional regime when M is a function of N such that $M := M(N) \to +\infty$, few results on the behaviour of correlation test statistics in the spectral domain are known. [62] proved that under $\mathcal{H}_0$ and mild assumptions on the underlying time series, the empirical eigenvalue distribution of $\hat{\mathbf{C}}(\nu)$ defined in (5.4) converges weakly almost surely towards the Marcenko-Pastur distribution, which can be exploited to build test statistics based on linear spectral statistics of $\hat{\mathbf{C}}(\nu)$. In [78], a consistent test statistic based on the largest eigenvalue of $\hat{\mathbf{C}}(\nu)$ was derived for the problem of detecting the presence of a signal with a low-rank spectral density matrix within a noise with uncorrelated components.

In the asymptotic regime where $\frac{M}{N} \to \gamma$, [69] proposed to test hypothesis $\mathcal{H}_0$ when the components of $\mathbf{y}$ share the same spectral density. In this case, the rows of the $M \times N$ matrix $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ are independent and identically distributed under $\mathcal{H}_0$. [69] established a central limit theorem for linear spectral statistics of the empirical covariance matrix, and deduced from this a test statistics to check whether $\mathcal{H}_0$ holds or not. We notice that the results of [69] are valid in the non-Gaussian case.

More results are available in the case where the time series $(y_{m,n})_{n\in\mathbb{Z}}$, $m \in [M]$, are temporally white. To test the correlation of the M components, one can similarly consider sample estimates of the correlation matrix, and test whether it is close to the identity matrix. Under the asymptotic regime where $\frac{M}{N} \to \gamma \in (0, +\infty)$, [45] showed that the maximum off-diagonal entry of the sample correlation matrix after proper normalization is also asymptotically distributed as Gumbel. The techniques used here for proving (5.2) are partly based on this paper. Other works such as [64] studied the asymptotic distribution of linear spectral statistics of the correlation matrix, [21] focused on the behaviour of the determinant of the correlation matrix, and [15] considered a U-statistic and obtained minimax results over some class of alternatives. Some other papers also explored various classes of alternative $\mathcal{H}_1$, among which is [28], who showed a phase transition phenomena in the behaviour of the largest off-diagonal entry of the correlation matrix driven by the magnitude of the dependence parameter defined in the alternative class $\mathcal{H}_1$. Lastly, [65] studied asymptotic first and second-order behaviour of the largest eigenvalues and associated eigenvectors of the sample correlation matrix under a specific alternative spiked model.

5.2 Main results

5.2.1 Assumptions

Throughout the chapter we rely on the following assumptions.

Assumption 5.1 (Time series)**.** *The time series $(y_{m,n})_{n\in\mathbb{Z}}$, $m \geq 1$, are mutually independent, stationary and zero-mean complex Gaussian distributed* [2]*.*

For each $m \geq 1$, we denote by $r_m = (r_m(u))_{u\in\mathbb{Z}}$ (instead of $r_{m,m}$) the covariance sequence of $(y_{m,n})_{n\in\mathbb{Z}}$, i.e. $r_m(u) = \mathbb{E}[y_{m,n+u}\overline{y_{m,n}}]$, and we formulate the following assumption on $(r_m)_{m\geq 1}$:

Assumption 5.2 (Memory)**.** *The covariance sequences $(r_m)_{m\geq 1}$ satisfy the uniform short memory condition*

$$\sup_{m\geq 1} \sum_{u\in\mathbb{Z}} (1+|u|)|r_m(u)| < +\infty.$$

We denote by $s_m(\nu) = \sum_{u\in\mathbb{Z}} r_m(u)\mathrm{e}^{-\mathrm{i}2\pi\nu}$ (instead of $s_{m,m}(\nu)$) the spectral density of $(y_{m,n})_{n\in\mathbb{Z}}$ at frequency $\nu \in [0,1]$. Assumption 5.2 of course implies that the function s_m is continously differentiable and that

$$\sup_{m\geq 1} \max_{\nu\in[0,1]} s_m(\nu) < +\infty, \quad \sup_{m\geq 1} \max_{\nu\in[0,1]} \left|\frac{\mathrm{d}s_m}{\mathrm{d}\nu}(\nu)\right| < +\infty. \tag{5.5}$$

Finally, as the sample spectral coherence of $(y_{i,n})_{n\in\mathbb{Z}}$ and $(y_{j,n})_{n\in\mathbb{Z}}$ involves a renormalization by the inverse of the estimates of the spectral densities s_i and s_j, we also need that s_i, s_j do not vanish. This is the substance of the next assumption.

Assumption 5.3 (Non-vanishing spectrum)**.** *The spectral densities are uniformly bounded away from zero, that is*

$$\inf_{m\geq 1} \min_{\nu\in[0,1]} s_m(\nu) > 0. \tag{5.6}$$

By Assumptions 5.2 and 5.3, there exist quantities $s_{\min}$ and $s_{\max}$ such that

$$0 < s_{\min} \leq \inf_{m\geq 1} \min_{\nu\in[0,1]} s_m(\nu) \leq \sup_{m\geq 1} \max_{\nu\in[0,1]} s_m(\nu) \leq s_{\max} < +\infty. \tag{5.7}$$

We now formulate the following assumptions on the growth rate of the quantities N, M, B, which describe the high-dimensional regime considered in this chapter.

Assumption 5.4 (Asymptotic regime)**.** *B and M are functions of N such that there exist positive constants $C_1, C_2 \in (0,+\infty)$ and $\rho \in (0,1)$ such that:*

$$C_1 N^\rho \leq B, M \leq C_2 N^\rho$$

and

$$\frac{M}{B} := c_N \xrightarrow[N\to+\infty]{} c \in (0,+\infty).$$

[2] A complex random variable Z is zero-mean complex Gaussian distributed with variance σ^2, denoted as $Z \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$, if $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are i.i.d. $\mathcal{N}(0,\frac{\sigma^2}{2})$ random variables.

Notations Even if the subscript $\cdot_N$ is not always specified, almost all quantities should be remembered to be dependent on N. Moreover, C represents a universal constant (i.e. a positive quantity independent of N, M, B), whose precise value is irrelevant and which may change from one line to another.

5.2.2 Statement of the result

The main result of this chapter, whose proof is deferred to Section 5.4, is given in the following theorem.

Theorem 5.1. *Under Assumptions 5.1 – 5.3, for any $t \in \mathbb{R}$:*

$$\mathbb{P}\left((B+1)\max_{(i,j,\nu)\in\mathcal{I}}|\hat{c}_{ij}(\nu)|^2 \leq t + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2}\right) \xrightarrow[N\to+\infty]{} e^{-e^{-t}}.$$

Thus, Theorem 5.1 states that $\max_{(i,j,\nu)\in\mathcal{I}}|\hat{c}_{ij}(\nu)|^2$, atfer proper normalization and centering, converges in distribution to a type I extreme value distribution, also known as Gumbel distribution. As it will be clear in the proof, the term $\log\frac{M(M-1)}{2}$ is related to the maximum over (i,j) while the term $\log\frac{N}{B+1}$ is related to the maximum over $\nu \in \mathcal{G}$.

We now illustrate numerically the above asymptotic result. Consider M independent AR(1) processes, driven by a standard Gaussian white noise, i.e.

$$\mathbf{y}_n := \begin{pmatrix} y_{1,n} \\ \vdots \\ y_{M,n} \end{pmatrix} = \theta \begin{pmatrix} y_{1,n-1} \\ \vdots \\ y_{M,n-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,n} \\ \vdots \\ \epsilon_{M,n} \end{pmatrix}, \quad \epsilon_{m,n} \overset{i.i.d.}{\sim} \mathcal{N}_{\mathbb{C}}(0,1)$$

with $\theta = 0.6$, and $(N, M) = (20000, 500)$. The smoothed periodogram estimators are computed using $B = 1000$. We independently draw 10000 samples of the time series $(\mathbf{y}_n)_{n\in[N]}$ and compute the associated MSSC $\max_{(i,j,\nu)\in\mathcal{I}_N}|\hat{c}_{ij}(\nu)|^2$. On Figure 5.2.1 are represented the sample cumulative distribution function (cdf) and the histogram of the MSSC against the Gumbel cdf and probability density function (pdf). We indeed observe that the rescaled distribution of $\max_{(i,j,\nu)\in\mathcal{I}_N}|\hat{c}_{ij}(\nu)|^2$ is close to the Gumbel distribution.

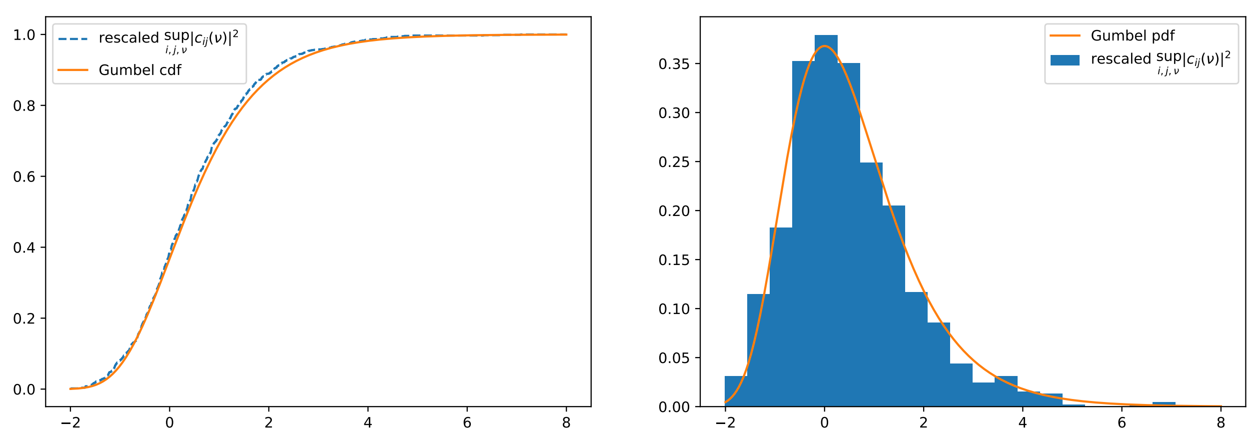


Figure 5.2.1: Sample cdf and histogram of the MSSC as defined in Theorem 5.1 vs Gumbel distribution.

.

5.3 Application to testing

5.3.1 New proposed test statistic

Theorem 5.1 can be used to design a new independence test statistic with a controlled asymptotic level in the proposed high-dimensional regime.

Define q_α the α–quantile of the Gumbel distribution: $q_\alpha = F^{-1}(\alpha)$ where

$$F(x) = \exp(-\exp(-x)).$$

The test statistic $T_N^{(\mathrm{MSSC})}$ defined by

$$T_N^{(\mathrm{MSSC})} = \mathbb{1}\left(\max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{ij}(\nu)|^2 > \frac{q_{1-\alpha} + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2}}{B+1} \right) \tag{5.8}$$

satisfies, as a direct consequence of Theorem 5.1, $\lim_{N\to+\infty} \mathbb{P}[T_N^{(\mathrm{MSSC})} = 1] = \alpha$ under $\mathcal{H}_0$.

5.3.2 Type I error

To test the independence of the signals $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\dots,M}$, we consider the statistic $T_N^{(\mathrm{MSSC})}$ defined in (5.8). On Table 5.3.1 are presented the sample type I errors of $T_N^{(\mathrm{MSSC})}$ with different combinations of sample sizes and dimensions ($\rho = 0.7$ and $\frac{M}{B+1} = 0.5$), when the nominal significant level for all the tests is set at $\alpha = 0.05$, and all statistics are computed from 30000 independent replications. One can see as expected that the type I error of $T_N^{(\mathrm{MSSC})}$ does indeed remain near 5% as M increases.

Table 5.3.1: Sample type I error at 5%

N	B	M	$T_N^{(\mathrm{MSSC})}$
42	20	10	0.021
316	100	50	0.031
659	180	90	0.037
1044	260	130	0.037
1459	340	170	0.040
1901	420	210	0.042
5623	1000	500	0.048
13374	2000	1000	0.051

5.3.3 Power

We now compare the power of our new test statistic against other independence test statistics which are designed to work in the high-dimensional regime. We define the Linear Spectral Statistic (LSS) test from [62] for any $\epsilon > 0$ by

$$T_N^{(\mathrm{LSS})} = \mathbb{1}\left(\sup_{\nu\in[0,1]} \frac{\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)}\right|}{N^{\epsilon}(B/N)} > \kappa_{1-\alpha} \right) \tag{5.9}$$

where $\mu_{MP}^{(c)}$ represents the Marcenko-Pastur distribution with parameter c defined by

$$d\mu_{MP}^{(c)}(\lambda) = \left(1 - \frac{1}{c}\right)_+ \delta_0(\mathrm{d}\lambda) + \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi c\lambda} \mathbb{1}_{[\lambda_-, \lambda_+]}(\lambda)\,\mathrm{d}\lambda$$

where $\lambda_\pm = (1 \pm \sqrt{c})^2$, $(\cdot)_+ := \max(\cdot, 0)$, $c_N := \frac{M}{B+1}$ and f is some function defined on $\mathbb{R}_+$ satisfying regularity assumptions (see more details in [62]). In practice, ϵ will be taken equal to 0.1. It is proven in [62] that under $\mathcal{H}_0$, $T_N^{(\mathrm{LSS})} \to 0$ almost surely in the high-dimensional regime but the exact asymptotic distribution of the LSS test is unknown. Therefore, the detection threshold $\kappa_{1-\alpha}$ for this test is based on a sample quantile of $T_N^{(\mathrm{LSS})}$ under $\mathcal{H}_0$ computed from Monte-Carlo simulation. For fairness comparison, we also use this procedure for the new test statistic $T_N^{(\mathrm{MSSC})}$. More precisely, we compute the sample $(1-\alpha)$–quantile $\kappa_{1-\alpha}$ of a test statistic $T_N^{(\mathrm{LSS})}$ from samples under $\mathcal{H}_0$, and then reject the null hypothesis under $\mathcal{H}_1$ if $T_N^{(LSS)} > \kappa_{1-\alpha}$. It remains to choose a test function f, and we again follow [62] by considering

- the Frobenius test $T_N^{(\mathrm{FROB})}$ when $f(x) = (x-1)^2$

- the logdet test $T_N^{(\mathrm{LOG})}$ when $f(x) = \log x$

It remains to define the alternatives. For this, we consider the following multidimensional $AR(1)$ model:

$$\mathbf{y}_{n+1} = \mathbf{A}\mathbf{y}_n + \boldsymbol{\epsilon}_n \tag{5.10}$$

where $(\boldsymbol{\epsilon}_n)_{n\in\mathbb{Z}}$ is a sequence of independent $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{I})$ distributed random vectors, and $\mathbf{A}$ is a bidiagonal matrix. Three choices of $\mathbf{A}$ ($\mathbf{A}^{(\mathcal{H}_0)}$, $\mathbf{A}^{(\mathcal{H}_{1,\mathrm{loc}})}$, $\mathbf{A}^{(\mathcal{H}_{1,\mathrm{glob}})}$) allows us to define two alternatives:

1. $\mathcal{H}_0$: for $|\theta| < 1$:

$$\mathbf{A}^{(\mathcal{H}_0)} = \begin{pmatrix} \theta & 0 & \dots & \dots & \dots & 0 \\ 0 & \theta & 0 & \dots & \dots & 0 \\ 0 & 0 & \theta & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 0 & \theta \end{pmatrix}$$

so the signals $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\dots,M}$ are mutually independent.

2. $\mathcal{H}_{1,\mathrm{loc}}$: for $|\theta| < 1$ and $\beta \in \mathbb{R}$:

$$\mathbf{A}^{(\mathcal{H}_{1,\mathrm{loc}})} = \begin{pmatrix} \theta & 0 & \dots & \dots & \dots & 0 \\ \beta & \theta & 0 & \dots & \dots & 0 \\ 0 & 0 & \theta & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 0 & \theta \end{pmatrix}$$

so the couple of time series (1,2) is the unique correlated pair of signals.

3. $\mathcal{H}_{1,\text{glob}}$: for $|\theta| < 1$ and $\beta \in \mathbb{R}$:

$$\mathbf{A}^{(\mathcal{H}_{1,\text{glob}})} = \begin{pmatrix} \theta & 0 & \dots & \dots & \dots & 0 \\ \beta & \theta & 0 & \dots & \dots & 0 \\ 0 & \beta & \theta & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \beta & \theta \end{pmatrix}$$

so all the signals are mutually correlated.

We now fix the value of the parameters involved under the three hypotheses. θ will always be taken equal to 0.5. Under $\mathcal{H}_{1,\text{loc}}$, $\beta = 0.1$. Concerning the alternative $\mathcal{H}_{1,\text{glob}}$, more care is required to choose β. Indeed, one can define a measure of total dependence as:

$$r := \frac{\int \|\mathbf{S}(\nu) - \mathrm{dg}\mathbf{S}(\nu)\|_F^2 \,\mathrm{d}\nu}{\int \|\mathbf{S}(\nu)\|_F^2 \,\mathrm{d}\nu} = \frac{\sum_{u\in\mathbb{Z}} \|\mathbf{R}(u) - \mathrm{dg}\mathbf{R}(u)\|_F^2}{\sum_{u\in\mathbb{Z}} \|\mathbf{R}(u)\|_F^2}$$

where $\mathbf{R}(u) := \mathbb{E}[\mathbf{y}_{n+u}\mathbf{y}_n^*]$, $\mathbf{S}(\nu) = \sum_{u\in\mathbb{Z}} \mathbf{R}(u)\mathrm{e}^{-\mathrm{i}2\pi u\nu}$ and dg denotes the diagonal part operator. Clearly, $r = 0$ under $\mathcal{H}_0$, and as $r > 0$ increases, the M–dimensional time series become correlated. We also see that for any fixed value of β, r is increasing with M. It is therefore more desirable to tune $\beta := \beta(M)$ such that r remains constant as M increases. This will enable our tests to be compared against an alternative which does not become asymptotically trivial.

The two alternatives $\mathcal{H}_{1,\text{loc}}$ and $\mathcal{H}_{1,\text{glob}}$ are useful to measure the performance of the independence tests under two different setups. Under $\mathcal{H}_{1,\text{loc}}$, each pair of time series are independent except the pair $(y_{1,n})_{n\in\mathbb{Z}}$,$(y_{2,n})_{n\in\mathbb{Z}}$, whereas under $\mathcal{H}_{1,\text{glob}}$ each time series has a small correlation with every other time series.

In Table 5.3.2 and Table 5.3.3 are presented the sample powers when the type I error is fixed at 5% for the considered tests and the two alternatives. The asymptotic regime is the same as the one considered for Table 5.3.1: $\rho = 0.7$ and $\frac{M}{B+1} = 0.2$. All statistics are computed from 30000 independent replications. We observe that under $\mathcal{H}_{1,\text{glob}}$, with $r = 0.01$, all the tests asymptotically detect the alternative, however with different performances. The LSS test statistics show better power which indicates that they may be more suited to detect alternatives under $\mathcal{H}_{1,\text{glob}}$ than the MSSC test statistics. Under $\mathcal{H}_{1,\text{loc}}$ the results are opposite: the power of $T_N^{(\text{MSSC})}$ rapidly increases to 1 as M increases. These results are not surprising since the MSSC test statistic is designed to detect peaks in the off-diagonal entries of $\hat{\mathbf{C}}(\nu)$ which is exactly the class of alternative considered in $\mathcal{H}_{1,\text{loc}}$. However, when the correlations are spread among all pairs of time series under $\mathcal{H}_{1,\text{glob}}$, the test statistics based on the global behaviour of the eigenvalues of $\hat{\mathbf{C}}(\nu)$ seem more relevant.

In Figure 5.3.1 are represented the ROC for each test under both alternatives. We observe that $T_N^{(\text{FROB})}$ and $T_N^{(\text{LOG})}$ have similar performance and outperform $T_N^{(\text{MSSC})}$ for the alternative $\mathcal{H}_{1,\text{glob}}$, while $T_N^{(\text{MSSC})}$ has better performance for $\mathcal{H}_{1,\text{loc}}$.

5.4 Proof of Theorem 5.1

We will detail in this section the main steps to prove Theorem 5.1, while some details will be left in the Appendix.

Table 5.3.2: Power comparison under $\mathcal{H}_1$ global, type I error = 5%

N	M	B	$T_N^{(\mathrm{FROB})}$	$T_N^{(\mathrm{LOG})}$	$T_N^{(\mathrm{MSSC})}$
42	10	20	0.050	0.049	0.052
316	50	100	0.036	0.042	0.067
659	90	180	0.067	0.065	0.086
1044	130	260	0.142	0.122	0.133
1459	170	340	0.339	0.255	0.214
1901	210	420	0.601	0.462	0.328
2364	250	500	0.836	0.682	0.503
2846	290	580	0.960	0.852	0.672

Table 5.3.3: Power comparison under $\mathcal{H}_1$ local, type I error = 5%

N	M	B	$T_N^{(\mathrm{FROB})}$	$T_N^{(\mathrm{LOG})}$	$T_N^{(\mathrm{MSSC})}$
42	10	20	0.049	0.049	0.061
316	50	100	0.038	0.044	0.352
659	90	180	0.038	0.041	0.881
1044	130	260	0.034	0.038	0.999
1459	170	340	0.034	0.038	1.000
1901	210	420	0.035	0.039	1.000
2364	250	500	0.031	0.039	1.000
2846	290	580	0.032	0.036	1.000

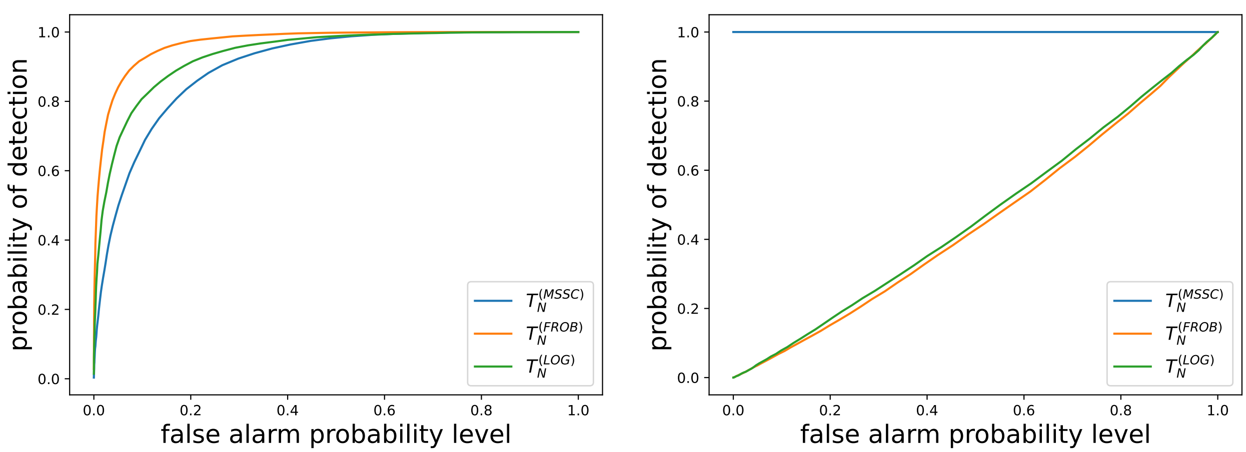


Figure 5.3.1: ROC associated to each test under $\mathcal{H}_1^{(glob)}$ with $r = 0.01$ (left) and $\mathcal{H}_1^{(loc)}$ with $\beta = 0.1$ (right) when $(N, M, B) = (2846, 290, 580)$)

.

5.4.1 General approach

First, we notice that the frequency smoothed estimate $\hat{s}_{i,j}(\nu)$ defined in (5.1) can be written as

$$\hat{s}_{ij}(\nu) = \frac{1}{B+1} \boldsymbol{\xi}_{y_j}(\nu)^* \boldsymbol{\xi}_{y_i}(\nu) \tag{5.11}$$

where

$$\boldsymbol{\xi}_{y_i}(\nu) = \left(\xi_{y_i}\left(\nu - \frac{B}{2N}\right), \ldots, \xi_{y_i}\left(\nu + \frac{B}{2N}\right)\right)^T.$$

This is a sesquilinear form of the finite Fourier transform of the M time series samples $(y_{i,1}, \ldots, y_{i,N})_{i\in[M]}$. To handle the statistical dependence between the components of $\boldsymbol{\xi}_{y_i}(\nu)$, we use the well-known Bartlett decomposition (see for instance [94]) whose procedure is described hereafter.

From Assumptions 5.2 and 5.3, the spectral distribution of $(y_{m,n})_{n\in\mathbb{Z}}$ is absolutely continuous with density s_m being uniformly bounded and bounded away from 0. Therefore, from Wold's Theorem [14, Th. 5.7.1, Th. 5.7.2], each time series $(y_{m,n})_{n\in\mathbb{Z}}$ admits a causal and causally invertible linear representation in terms of its normalized innovation sequence:

$$y_{m,n} = \sum_{k=0}^{+\infty} a_{m,k}\epsilon_{m,n-k}, \tag{5.12}$$

where $(\epsilon_{1,k})_{k\in\mathbb{Z}}, \ldots, (\epsilon_{M,k})_{k\in\mathbb{Z}}$ are mutually independent sequences of $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. random variables, and $(a_{1,k})_{k\in\mathbb{N}}, \ldots, (a_{M,k})_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$ such that if

$$h_m(\nu) = \sum_{k=0}^{+\infty} a_{m,k}e^{-2i\pi k\nu} \tag{5.13}$$

then $|h_m(\nu)|^2 = s_m(\nu)$ and $h_m(\nu)$ coincides with the outer causal spectral factor of $s_m(\nu)$. Define now $\tilde{s}_{ij}(\nu)$, an approximation of $\hat{s}_{ij}(\nu)$, as:

$$\tilde{s}_{ij}(\nu) = \frac{1}{B+1}\sum_{b=-B/2}^{B/2} h_i\left(\nu + \frac{b}{N}\right)\overline{h_j\left(\nu + \frac{b}{N}\right)}\xi_{\epsilon_i}\left(\nu + \frac{b}{N}\right)\overline{\xi_{\epsilon_j}\left(\nu + \frac{b}{N}\right)}$$

or equivalently

$$\tilde{s}_{ij}(\nu) = \boldsymbol{\xi}_{\epsilon_j}(\nu)^*\frac{\boldsymbol{\Pi}_{ij}(\nu)}{B+1}\boldsymbol{\xi}_{\epsilon_i}(\nu) \tag{5.14}$$

where

$$\boldsymbol{\Pi}_{ij}(\nu) = \mathrm{dg}\left(h_i\left(\nu + \frac{b}{N}\right)\overline{h_j\left(\nu + \frac{b}{N}\right)}\right)_{b=-B/2,\ldots,B/2} \tag{5.15}$$

and

$$\boldsymbol{\xi}_{\epsilon_i}(\nu) = \left(\xi_{\epsilon_i}\left(\nu - \frac{B}{2N}\right), \ldots, \xi_{\epsilon_i}\left(\nu + \frac{B}{2N}\right)\right)^T.$$

Instead of working directly with $|\hat{c}_{ij}(\nu)|^2 = \frac{|\hat{s}_{ij}(\nu)|^2}{\hat{s}_i(\nu)\hat{s}_j(\nu)}$, it turns out that it is more convenient to show the limiting Gumbel distribution for $\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)}$ where

$$\begin{aligned}
\sigma_{ij}^2(\nu) &= \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|h_i\left(\nu + \frac{b}{N}\right)\right|^2\left|h_j\left(\nu + \frac{b}{N}\right)\right|^2 \\
&= \frac{1}{B+1}\sum_{b=-B/2}^{B/2} s_i\left(\nu + \frac{b}{N}\right)s_j\left(\nu + \frac{b}{N}\right) \\
&:= \frac{\mathrm{tr}\ \boldsymbol{\Sigma}_{ij}(\nu)}{B+1}
\end{aligned} \tag{5.16}$$

and where

$$\mathbf{\Sigma}_{ij}(\nu) := \mathbf{\Pi}_{ij}^*(\nu)\mathbf{\Pi}_{ij}(\nu)$$
$$= \mathrm{dg}\left(\left|h_i\left(\nu + \frac{b}{N}\right)\right|^2 \left|h_j\left(\nu + \frac{b}{N}\right)\right|^2, b = -\frac{B}{2}, \dots, \frac{B}{2}\right). \quad (5.17)$$

This is the aim of Proposition 5.1 below.

Proposition 5.1 (Gumbel limit for $\max_{(i,j,\nu)\in\mathcal{I}} |\tilde{s}_{ij}(\nu)|^2$). *Under Assumptions 5.1 – 5.3, for any $t \in \mathbb{R}$, we have*

$$\mathbb{P}\left(\max_{(i,j,\nu)\in\mathcal{I}} (B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} \leq t + \log\frac{N}{B+1} + \log\frac{M(M-1)}{2}\right) \xrightarrow[N\to+\infty]{} e^{-e^{-t}}. \quad (5.18)$$

Once equipped with Proposition 5.1, it remains then to show that $\max_{(i,j,\nu)\in\mathcal{I}} \frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)}$ is close enough to $\max_{(i,j,\nu)\in\mathcal{I}} |\hat{c}_{ij}(\nu)|^2$ to prove that these quantities have the same limiting distribution. This result is given by the following Proposition.

Proposition 5.2. *Under Assumptions 5.1 – 5.3, as $N \to \infty$,*

$$\max_{(i,j,\nu)\in\mathcal{I}} (B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} - \max_{(i,j,\nu)\in\mathcal{I}} (B+1)|\hat{c}_{ij}(\nu)|^2 = o_P(1).$$

As Theorem 5.1 is directly obtained by Proposition 5.1, Proposition 5.2 and an application of Slutsky's lemma, the two remaining subsections are devoted to the proofs of Proposition 5.1 and Proposition 5.2.

5.4.2 Proof of Proposition 5.1

To prove Proposition 5.1, the main tool is Lemma A.4 from [45], which is a special case of Poisson approximation from [3]. We rewrite it here for the sake of completeness.

Lemma 5.1. *Let $(X_\alpha)_{\alpha\in\mathcal{I}}$ be a finite collection of Bernoulli random variables, and for each $\alpha \in \mathcal{I}$, let $\mathcal{I}_\alpha \subset \mathcal{I}$ such that $\alpha \in \mathcal{I}_\alpha$. Then,*

$$\left|\mathbb{P}\left(\sum_{\alpha\in\mathcal{I}} X_\alpha = 0\right) - \exp\left(-\sum_{\alpha\in\mathcal{I}} \mathbb{P}(X_\alpha = 1)\right)\right| \leq \Delta_1 + \Delta_2 + \Delta_3$$

where

$$\Delta_1 = \sum_{\alpha\in\mathcal{I}} \sum_{\beta\in\mathcal{I}_\alpha} \mathbb{P}\left(X_\alpha = 1\right) \mathbb{P}\left(X_\beta = 1\right)$$
$$\Delta_2 = \sum_{\alpha\in\mathcal{I}} \sum_{\beta\in\mathcal{I}_\alpha\setminus\{\alpha\}} \mathbb{P}\left(X_\alpha = 1,\, X_\beta = 1\right)$$
$$\Delta_3 = \sum_{\alpha\in\mathcal{I}} \mathbb{E}\left|\mathbb{P}\left(X_\alpha = 1 | \left(X_\beta\right)_{\beta\in\mathcal{I}\setminus\mathcal{I}_\alpha}\right) - \mathbb{P}(X_\alpha = 1)\right|$$

In particular, if for each $\alpha \in \mathcal{I}$, X_α is independent of $\{X_\beta : \beta \in \mathcal{I} \setminus \mathcal{I}_\alpha\}$, then $\Delta_3 = 0$.

Lemma 5.1 is the keystone for the proof of Proposition 5.1, and is a standard tool for analyzing distributions of maxima of dependent random variables. We now prove Proposition 5.1.

Proof. We start by proving (5.18). Define

$$t_N = \sqrt{x + \log\frac{M(M-1)}{2} + \log\frac{N}{B+1}} \tag{5.19}$$

and for $(i,j,\nu) \in \mathcal{I}$ (recall that $\mathcal{I}$ is defined in (5.3), and that it depends on N, but to avoid cumbersome notations we do not recall this dependency) the Bernoulli random variables $X_{ij}(\nu)$ as

$$X_{ij}(\nu) := \mathbb{1}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right). \tag{5.20}$$

Define the set $\mathcal{I}_{(i,j,\nu)}$

$$\mathcal{I}_{(i,j,\nu)} = \{(i',j',\nu) : 1 \le i' < j' \le M, i = i' \text{ or } j = j'\}. \tag{5.21}$$

From (5.14) and under Assumption 5.1, if $(i',j',\nu') \in \mathcal{I}\backslash\mathcal{I}_{(i,j,\nu)}$, then $\tilde{s}_{i'j'}(\nu')$ is independent from $\tilde{s}_{ij}(\nu)$ since we have either

(1) $i' \neq i$, $j' \neq j$, $\nu' = \nu$;

(2) $i' = i$ or $j' = j$, and $\nu' \neq \nu$ (implying $|\nu - \nu'| > \frac{B}{N}$ by assumption), in which case $\left(\boldsymbol{\xi}_{\epsilon_{i'}}(\nu'), \boldsymbol{\xi}_{\epsilon_{j'}}(\nu')\right)$ is independent from $\left(\boldsymbol{\xi}_{\epsilon_i}(\nu), \boldsymbol{\xi}_{\epsilon_j}(\nu)\right)$.

From the definition of $X_{ij}(\nu)$ in (5.20),

$$\mathbb{P}\left((B+1)\max_{(i,j,\nu)\in\mathcal{I}}\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} \le t_N^2\right) = \mathbb{P}\left(\sum_{(i,j,\nu)\in\mathcal{I}} X_{ij}(\nu) = 0\right)$$

which can be estimated by Lemma 5.1 as:

$$\left|\mathbb{P}\left(\sum_{(i,j,\nu)\in\mathcal{I}} X_{ij}(\nu) = 0\right) - e^{-\lambda}\right| \le \Delta_1 + \Delta_2 + \Delta_3$$

where

$$\lambda = \sum_{(i,j,\nu)\in\mathcal{I}} \mathbb{P}\left(X_{ij}(\nu) = 1\right) = \sum_{(i,j,\nu)\in\mathcal{I}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right)$$

and

$$\Delta_1 = \sum_{(i,j,\nu)\in\mathcal{I}} \sum_{(i',j',\nu)\in\mathcal{I}_{(i,j,\nu)}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{i,j}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right)\mathbb{P}\left((B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > t_N^2\right) \tag{5.22}$$

$$\Delta_2 = \sum_{(i,j,\nu)\in\mathcal{I}} \sum_{\substack{(i',j',\nu)\in\mathcal{I}_{(i,j,\nu)} \\ (i',j')\neq(i,j)}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{i,j}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2,\ (B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > t_N^2\right) \tag{5.23}$$

$$\Delta_3 = \sum_{(i,j,\nu)\in\mathcal{I}} \mathbb{E}\left|\mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2 \middle| \left(\tilde{s}_{i'j'}(\nu')\right)_{(i',j',\nu')\in\mathcal{I}\backslash\mathcal{I}_{(i,j,\nu)}}\right) - \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right)\right|. \tag{5.24}$$

We now have to control the four quantities λ, Δ_1, Δ_2 and Δ_3, which requires studying moderate deviations results for

$$\mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right)$$

as well as

$$\mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2,\ (B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > t_N^2\right)$$

for all $(i',j',\nu) \in \mathcal{I}_{(i,j,\nu)}$. The following Proposition 5.3, proved in Appendix 5.C provides exactly this.

Proposition 5.3. *Under Assumptions 5.1 – 5.3, there exists a constant $\eta > 0$ such that for any $C > 0$, we have*

$$\max_{t \in [0, CB^\eta]} \max_{(i,j,\nu)\in\mathcal{I}} \left| \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t^2\right) e^{t^2} - 1 \right| \xrightarrow[N\to\infty]{} 0 \tag{5.25}$$

and

$$\max_{t,s \in [0, CB^\eta]} \max_{\substack{(i,j,\nu)\in\mathcal{I} \\ (i',j',\nu)\in\mathcal{I}_{(i,j,\nu)}}} \left| \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t^2, (B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > s^2\right) \times e^{t^2+s^2} - 1 \right| \xrightarrow[N\to\infty]{} 0. \tag{5.26}$$

First, concerning $\exp(-\lambda)$, since t_N as defined in (5.19) is $\mathcal{O}(\log N)$, one can use Proposition 5.3 to get

$$\begin{aligned}
\exp(-\lambda) &= \exp\left(-\sum_{(i,j,\nu)\in\mathcal{I}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2\right)\right) \\
&= \exp\left(-\frac{N}{B+1}\frac{M(M-1)}{2} e^{-t_N^2}(1+o(1))\right) \\
&\xrightarrow[N\to\infty]{} \exp\left(-\exp(-x)\right).
\end{aligned}$$

We now turn to the control of Δ_1, Δ_2, and Δ_3. Regarding Δ_3, since under Assumption 5.1 the random variables $\tilde{s}_{ij}(\nu)$ and $\tilde{s}_{i'j'}(\nu')$ for $(i',j',\nu') \in \mathcal{I}\backslash\mathcal{I}_{(i,j,\nu)}$ are independent, we clearly have $\Delta_3 = 0$. Consider now (5.22) and (5.23). The aim is to show that $\Delta_1 = o(1)$ and $\Delta_2 = o(1)$ when t_N is defined by (5.19). Using the moderate deviation result (5.25) from Proposition 5.3, and recalling that C represents a universal constant independent of N whose value can change from one line to another, we get:

$$\begin{aligned}
\Delta_1 &\leq \underbrace{|\mathcal{I}|}_{\mathcal{O}(\frac{N}{B}M^2)} \underbrace{\max_{(i,j,\nu)\in\mathcal{I}} |\mathcal{I}_{(i,j,\nu)}|}_{\mathcal{O}(M)} \max_{(i,j,\nu)\in\mathcal{I}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{i,j}(\nu)|^2}{\sigma_{i,j}(\nu)^2} > t_N^2\right)^2 \\
&\leq C\frac{N}{B}M^3 \underbrace{e^{-2t_N^2}}_{\mathcal{O}(\frac{1}{M^4}\frac{B^2}{N^2})} \underbrace{\max_{(i,j,\nu)\in\mathcal{I}} \left(\mathbb{P}\left[(B+1)\frac{|\tilde{s}_{i,j}(\nu)|^2}{\sigma_{i,j}(\nu)^2} > t_N^2\right] e^{t_N^2}\right)^2}_{=1+o(1)} \\
&= \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}$$

Δ_2 is handled similarly with equation (5.26) from Proposition 5.3:

$$\begin{aligned}
\Delta_2 &= \sum_{(i,j,\nu)\in\mathcal{I}} \sum_{(i',j',\nu)\in\mathcal{I}_{(i,j,\nu)}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} > t_N^2,\ (B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > t_N^2\right) \\
&\leq |\mathcal{I}| \max_{(i,j,\nu)\in\mathcal{I}} |\mathcal{I}_{(i,j,\nu)}|\ e^{-2t_N^2} \\
&\quad \times \underbrace{\max_{(i,j,\nu)\in\mathcal{I}} \max_{(i',j',\nu)\in\mathcal{I}_{(i,j,\nu)}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_{i,j}(\nu)|^2}{\sigma_{i,j}(\nu)^2} > t_N^2, (B+1)\frac{|\tilde{s}_{i'j'}(\nu)|^2}{\sigma_{i'j'}^2(\nu)} > t_N^2\right) e^{2t_N^2}}_{=1+o(1)} \\
&= \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}$$

The proof of (5.18) is complete. □

5.4.3 Proof of Proposition 5.2

To prove Proposition 5.2, ie. the fact that $\max_{(i,j,\nu)\in\mathcal{I}} \frac{|\hat{s}_{ij}(\nu)|^2}{\hat{s}_i(\nu)\hat{s}_j(\nu)}$ and $\max_{(i,j,\nu)\in\mathcal{I}} \frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)}$ are close enough in probability, we work separately on the numerator and the denominator. This constitutes the statement of the two following propositions.

Proposition 5.4 (Change of numerator)**.** *Under Assumptions 5.1 – 5.3, there exists $\delta > 0$ such that as $N \to \infty$,*

$$\sqrt{B+1} \max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)| = \mathcal{O}_P(N^{-\delta}). \tag{5.27}$$

The proof is deferred to Appendix 5.B. A consequence of Proposition 5.4 and Proposition 5.1 is that

$$\begin{aligned}
&\sqrt{B+1} \max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu)| \\
&\qquad \leq \sqrt{B+1} \max_{(i,j,\nu)\in\mathcal{I}} |\tilde{s}_{i,j}(\nu)| + \sqrt{B+1} \max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)| \\
&\qquad = \mathcal{O}_P\left(\sqrt{\log N}\right).
\end{aligned} \tag{5.28}$$

Proposition 5.5 (Change of denominator)**.** *Under Assumption 5.2, for any $\epsilon > 0$, as $N \to \infty$,*

$$\max_{(i,j,\nu)\in\mathcal{I}} \left|\hat{s}_i(\nu)\hat{s}_j(\nu) - \sigma_{ij}^2(\nu)\right| = \mathcal{O}_P\left(\frac{B}{N} + \frac{N^\epsilon}{\sqrt{B}}\right). \tag{5.29}$$

Moreover,

$$0 < \inf_{N \geq 1} \min_{(i,j,\nu) \in \mathcal{I}} \sigma_{ij}^2(\nu) \leq \sup_{N \geq 1} \max_{(i,j,\nu) \in \mathcal{I}} \sigma_{ij}^2(\nu) < +\infty \tag{5.30}$$

and

$$\max_{i \in [M]} \max_{\nu \in \mathcal{G}} \frac{1}{\hat{s}_i(\nu)} = O_P(1), \quad \max_{i \in [M]} \max_{\nu \in \mathcal{G}} \hat{s}_i(\nu) = O_P(1). \tag{5.31}$$

The proof is deferred to Appendix 5.A. We recall that for any sequences (a_n) and (b_n), the following inequality holds:

$$\left| \sup_n a_n - \sup_n b_n \right| \leq \sup_n |a_n - b_n|.$$

Therefore, to show that Proposition 5.2 holds, it is enough to show that

$$\max_{(i,j,\nu) \in \mathcal{I}} \left| (B+1) \frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} - (B+1)|\hat{c}_{ij}(\nu)|^2 \right| = o_P(1).$$

This result could be proved by writing the following decomposition:

$$(B+1) \max_{(i,j,\nu) \in \mathcal{I}} \left| \frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} - \frac{|\hat{s}_{ij}(\nu)|^2}{\hat{s}_i(\nu)\hat{s}_j(\nu)} \right| \leq \Psi_3(\Psi_1 + \Psi_2).$$

where

$$\begin{aligned}
\Psi_1 &:= (B+1) \max_{(i,j,\nu) \in \mathcal{I}} \left| |\hat{s}_{ij}(\nu)|^2 - |\tilde{s}_{ij}(\nu)|^2 \right| \hat{s}_i(\nu)\hat{s}_j(\nu) \\
\Psi_2 &:= (B+1) \max_{(i,j,\nu) \in \mathcal{I}} |\hat{s}_{ij}(\nu)|^2 |\hat{s}_i(\nu)\hat{s}_j(\nu) - \sigma_{ij}^2(\nu)| \\
\Psi_3 &:= \max_{(i,j,\nu) \in \mathcal{I}} \frac{1}{\hat{s}_i(\nu)\hat{s}_j(\nu)\sigma_{ij}^2(\nu)}.
\end{aligned}$$

It is clear by (5.18) that

$$\max_{(i,j,\nu) \in \mathcal{I}} (B+1)|\tilde{s}_{ij}(\nu)|^2 = \mathcal{O}_P(\log N).$$

Combining this with Proposition 5.5 and equation (5.27) from Proposition 5.4, there exists $\delta > 0$ such that

$$\begin{aligned}
(B+1) \max_{(i,j,\nu) \in \mathcal{I}} \left| |\hat{s}_{ij}(\nu)|^2 - |\tilde{s}_{ij}(\nu)|^2 \right| \leq \\
\underbrace{\sqrt{B+1} \max_{(i,j,\nu) \in \mathcal{I}} (|\hat{s}_{ij}(\nu)| + |\tilde{s}_{ij}(\nu)|)}_{=\mathcal{O}_P(\sqrt{\log N})} \\
\times \underbrace{\sqrt{B+1} \max_{(i,j,\nu) \in \mathcal{I}} \left| |\hat{s}_{ij}(\nu)| - |\tilde{s}_{ij}(\nu)| \right|}_{=\mathcal{O}_P(N^{-\delta})}
\end{aligned}$$

which is $\mathcal{O}_P(\sqrt{\log N} N^{-\delta})$. Using (5.31), this implies that

$$\Psi_1 = \mathcal{O}_P\left(\sqrt{\log N} N^{-\delta}\right).$$

Similarly, using Proposition 5.5, for any $\epsilon > 0$,

$$\Psi_2 = \mathcal{O}_P\left(\log N\left(\frac{B}{N} + \frac{N^\epsilon}{\sqrt{B}}\right)\right)$$
$$\Psi_3 = \mathcal{O}_P(1).$$

Combining the estimates of Ψ_1, Ψ_2 and Ψ_3 we get that for any $\epsilon > 0$:

$$(B+1)\max_{(i,j,\nu)\in\mathcal{I}}\left|\frac{|\tilde{s}_{ij}(\nu)|^2}{\sigma_{ij}^2(\nu)} - \frac{|\hat{s}_{ij}(\nu)|^2}{\hat{s}_i(\nu)\hat{s}_j(\nu)}\right| =$$
$$\mathcal{O}_P\left(N^{-\delta}\sqrt{\log N} + \log N\left(\frac{B}{N} + \frac{N^\epsilon}{\sqrt{B}}\right)\right).$$

This quantity is $o_P(1)$ if $\frac{N^\epsilon}{\sqrt{B}} = o(1)$ which is satisfied by choosing $\epsilon < \frac{\rho}{2}$ from Assumption (5.4).

Appendix

5.A Proof of Proposition 5.5

Before proving (5.29), the main result of Proposition 5.5, we focus first on proving (5.30) and (5.31). Concerning (5.30), recall that $\sigma_{ij}^2(\nu)$ defined in (5.16) is equal to:

$$\sigma_{ij}^2(\nu) = \frac{1}{B+1} \sum_{b=-B/2}^{B/2} s_i\left(\nu + \frac{b}{N}\right) s_j\left(\nu + \frac{b}{N}\right). \tag{5.32}$$

By Assumption 5.2, it is clear that (5.30) holds. We now focus on proving (5.31). Since by Assumption 5.2 and Assumption 5.3 the true spectral densities $s_i(\nu)$ are far from 0 and $+\infty$, the same result should also hold for the estimators $\hat{s}_i(\nu)$. More precisely, we prove the following lemma.

Lemma 5.2. *Under Assumption 5.2,*

$$\max_{i\in[M]} \max_{\nu\in\mathcal{F}} |\mathbb{E}\hat{s}_i(\nu) - s_i(\nu)| = \mathcal{O}\left(\frac{B}{N}\right) \tag{5.33}$$

Moreover, under Assumption 5.1 and Assumption 5.2, for any $\epsilon > 0$, there exist $\gamma > 0$ and $N_0(\epsilon) \in \mathbb{N}$ such that:

$$\mathbb{P}\left(\max_{i\in[M]} \max_{\nu\in\mathcal{F}} |\hat{s}_i(\nu) - \mathbb{E}[\hat{s}_i(\nu)]| > N^{\epsilon}\frac{1}{\sqrt{B}}\right) \leq \exp\left(-N^{\gamma}\right) \tag{5.34}$$

for $N > N_0(\epsilon)$.

Proof. These results are close to those proved in Lemma A.2 and Lemma A.3 of [62]. We will therefore closely follow their proofs. We start with the bias. By the definition (5.11) of $\hat{s}_i(\nu)$:

$$|\mathbb{E}\hat{s}_i(\nu) - s_i(\nu)| = \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \mathbb{E}\left|\xi_{y_i}\left(\nu + \frac{b}{N}\right)\right|^2 - s_i(\nu) \right|.$$

Inserting $s_i(\nu + \frac{b}{N})$, one can write:

$$|\mathbb{E}\hat{s}_i(\nu) - s_i(\nu)| \leq \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left(\mathbb{E}\left|\xi_{y_i}\left(\nu + \frac{b}{N}\right)\right|^2 - s_i\left(\nu + \frac{b}{N}\right)\right)\right| + \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left(s_i\left(\nu + \frac{b}{N}\right) - s_i(\nu)\right)\right|.$$

[62, Lemma A.1] provides the following control for the first term of the right-hand side under

Assumption 5.2:

$$\max_{\nu\in\mathcal{F}}\max_{i\in[M]}\left|\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\mathbb{E}\left|\xi_{y_i}\left(\nu+\frac{b}{N}\right)\right|^2 - s_i\left(\nu+\frac{b}{N}\right)\right)\right| = \mathcal{O}\left(\frac{1}{N}\right). \tag{5.35}$$

Moreover, by Assumption 5.2, a Taylor expansion of s_i around $\nu+\frac{b}{N}$, provides the existence of a quantity ν_b such that:

$$s_i\left(\nu+\frac{b}{N}\right) = s_i(\nu) + \frac{b}{N}s_i'(\nu_b)$$

where by Assumption 5.2, $\sup_{i\geq 1}\sup_{\nu\in[0,1]}|s_i'(\nu)| < +\infty$. Therefore, uniformly in $\nu\in\mathcal{F}$ and $i\in[M]$:

$$\begin{aligned}\max_{\nu\in\mathcal{F}}\max_{i\in[M]}&\left|\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(s_i\left(\nu+\frac{b}{N}\right)-s_i(\nu)\right)\right| \\ &= \max_{\nu\in\mathcal{F}}\max_{i\in[M]}\left|\frac{1}{B+1}\sum_{b=-B/2}^{B/2}\frac{b}{N}s_i'(\nu_b)\right| = \mathcal{O}\left(\frac{B}{N}\right).\end{aligned} \tag{5.36}$$

Combining the estimates (5.35) and (5.36), one gets:

$$\max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\mathbb{E}\hat{s}_i(\nu)-s_i(\nu)| = \mathcal{O}\left(\frac{1}{N}+\frac{B}{N}\right) = \mathcal{O}\left(\frac{B}{N}\right)$$

which is the desired result.

The second part of the lemma is an extension of a similar result also proved in [62, Lemma A.3] (see also similar results in [10]). Under Assumption 5.1 and Assumption 5.2, they have shown that for any $\nu\in[0,1]$ and for any $\epsilon>0$, there exists $\gamma>0$ such that:

$$\mathbb{P}\left(\max_{i\in[M]}|\hat{s}_i(\nu)-\mathbb{E}[\hat{s}_i(\nu)]| > \frac{N^\epsilon}{\sqrt{B}}\right) \leq \exp -N^\gamma$$

for large enough $N > N_0(\epsilon)$. It remains to extend this concentration result to handle the uniformity over $\nu\in\mathcal{F}$. This is done easily by the union bound. $\square$

We can now prove (5.31). For any $A>0$, inserting $\mathbb{E}[\hat{s}_i(\nu)]$ and $s_i(\nu)$ we can write:

$$\begin{aligned}&\mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}\hat{s}_i(\nu) > A\right)\\ &= \mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\hat{s}_i(\nu)-\mathbb{E}[\hat{s}_i(\nu)]+\mathbb{E}[\hat{s}_i(\nu)]-s_i(\nu)+s_i(\nu)| > A\right)\\ &\leq \mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\hat{s}_i(\nu)-\mathbb{E}[\hat{s}_i(\nu)])\right.\\ &\qquad\qquad \left. > A - \max_{i\in[M]}\max_{\nu\in\mathcal{F}}s_i(\nu) - \max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\mathbb{E}[\hat{s}_i(\nu)]-s_i(\nu)|\right).\end{aligned}$$

By Lemma 5.2 equation (5.33) and Assumption 5.2, for N large enough:

$$A - \max_{i\in[M]}\max_{\nu\in\mathcal{F}}s_i(\nu) - \max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\mathbb{E}[\hat{s}_i(\nu)]-s_i(\nu)| \geq \frac{A}{2}.$$

The deviation result (5.34) from Lemma 5.3 finally provides:

$$\mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}\hat{s}_i(\nu) > A\right) \leq \mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\hat{s}_i(\nu) - \mathbb{E}[\hat{s}_i(\nu)]| > \frac{A}{2}\right) \xrightarrow[N\to+\infty]{} 0.$$

The proof that $\max_{i\in[M]}\max_{\nu\in\mathcal{F}}\frac{1}{\hat{s}_i(\nu)} = \mathcal{O}_P(1)$ is done similarly by considering

$$\mathbb{P}\left(\max_{i\in[M]}\max_{\nu\in\mathcal{F}}\frac{1}{\hat{s}_i(\nu)} > A\right).$$

We now focus on (5.29), and consider the following decomposition:

$$\max_{(i,j,\nu)\in\mathcal{I}}\left|\hat{s}_i(\nu)\hat{s}_j(\nu) - \sigma_{ij}^2(\nu)\right| \leq \max_{(i,j,\nu)\in\mathcal{I}}|\hat{s}_i(\nu)\hat{s}_j(\nu) - s_i(\nu)s_j(\nu)| + \max_{(i,j,\nu)\in\mathcal{I}}\left|s_i(\nu)s_j(\nu) - \sigma_{ij}^2(\nu)\right|.$$

The following two lemmas bound each term of the right hand side, and lead to (5.29).

Lemma 5.3. *Under Assumption 5.2, for any $\epsilon > 0$, as $N \to \infty$,*

$$\max_{(i,j,\nu)\in\mathcal{I}}|\hat{s}_i(\nu)\hat{s}_j(\nu) - s_i(\nu)s_j(\nu)| = \mathcal{O}_P\left(\frac{B}{N} + \frac{N^\epsilon}{\sqrt{B}}\right).$$

Proof. Write

$$\begin{aligned}\max_{(i,j,\nu)\in\mathcal{I}}&|\hat{s}_i(\nu)\hat{s}_j(\nu) - s_i(\nu)s_j(\nu)| \\ &\leq \underbrace{\max_{i\in[M]}\max_{\nu\in\mathcal{F}}|\hat{s}_i(\nu) - s_i(\nu)|}_{=\mathcal{O}_P(\frac{B}{N}+\frac{N^\epsilon}{\sqrt{B}})}\underbrace{\max_{j\in[M]}\max_{\nu\in\mathcal{F}}\hat{s}_j(\nu)}_{=\mathcal{O}_P(1)} \\ &\quad + \underbrace{\max_{i\in[M]}\max_{\nu\in\mathcal{F}}s_i(\nu)}_{=\mathcal{O}(1)}\underbrace{\max_{j\in[M]}\max_{\nu\in\mathcal{F}}|\hat{s}_j(\nu) - s_j(\nu)|}_{=\mathcal{O}_P(\frac{B}{N}+\frac{N^\epsilon}{\sqrt{B}})}\end{aligned}$$

for any $\epsilon > 0$, where each estimate comes from Lemma 5.2 and Assumption 5.2. □

Lemma 5.4. *Under Assumption 5.2, as $N \to \infty$,*

$$\max_{(i,j,\nu)\in\mathcal{I}}\left|\sigma_{ij}^2(\nu) - s_i(\nu)s_j(\nu)\right| = O\left(\frac{B}{N}\right).$$

Proof. By Assumption 5.2, the applications $\nu \mapsto s_i(\nu)$ are C^1, so by Taylor expansion of s_i around $\nu + \frac{b}{N}$, there exist frequencies $\nu_{i,b} \in [\nu, \nu + b/N]$ such that:

$$s_i\left(\nu + \frac{b}{N}\right) = s_i(\nu) + \frac{b}{N}s_i'(\nu_{i,b})$$

where $s_i(\nu)$ and $s_i'(\nu_{i,b})$ satisfies:

$$\sup_{i\geq 1}\max_{\nu\in[0,1]} s_i(\nu) < +\infty, \quad \sup_{i\geq 1}\max_{\nu\in[0,1]}|s_i'(\nu)| < +\infty$$

Recall the expression (5.32) of $\sigma_{ij}^2(\nu)$, and write:

$$\begin{aligned}&\left|\sigma_{ij}^2(\nu) - s_i(\nu)s_j(\nu)\right|\\ &\quad= \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\underbrace{\left(s_i\left(\nu+\frac{b}{N}\right) - s_i(\nu)\right)}_{=\mathcal{O}(B/N)}\underbrace{s_j\left(\nu+\frac{b}{N}\right)}_{=\mathcal{O}(1)} + \right.\\ &\qquad\left.\underbrace{s_i(\nu)}_{=\mathcal{O}(1)}\underbrace{\left(s_j\left(\nu+\frac{b}{N}\right) - s_j(\nu)\right)}_{=\mathcal{O}(B/N)}\right)\end{aligned}$$

where each bound above is uniform over $(i,j,\nu) \in \mathcal{I}$. □

5.B Proof of Proposition 5.4

To prove Proposition 5.4, we need the three following lemmas (Lemma 5.5, Lemma 5.6 and Lemma 5.7), which are exactly or slight modifications of results from [94]. We recall that according to (5.14), $\tilde{s}_{ij}(\nu)$ can be expressed as the following sesquilinear form

$$\tilde{s}_{ij}(\nu) = \boldsymbol{\xi}_{\epsilon_j}(\nu)^* \frac{\boldsymbol{\Pi}_{ij}(\nu)}{B+1}\boldsymbol{\xi}_{\epsilon_i}(\nu),$$

where the random variables $(\epsilon_{j,n})_{\substack{j\in[M]\\ n\in[N]}}$ are independent and indentically distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$. For the remainder, we denote for all $j \in [M]$ by $I_{N,\epsilon_j}(\nu)$ the periodogram of $(\epsilon_{j,n})_{n\in[N]}$ at frequency ν, i.e.

$$I_{N,\epsilon_j}(\nu) = \left|\xi_{\epsilon_j}(\nu)\right|^2.$$

The two following lemmas provide controls for the maximum of $I_{N,\epsilon_j}(\nu)$ over ν and j.

Lemma 5.5. *It holds that*

$$\mathbb{E}\left[\max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu)\right] = \mathcal{O}\left(\log N + \log M\right). \tag{5.37}$$

Proof. By independence and Gaussianity of the observations from the time series ϵ_j, it is well known that the random variables $(I_{N,\epsilon_j}(\nu))$ for $\nu \in \mathcal{F}$ and $j \geq 1$ are independent exponential $\mathcal{E}(1)$ random variables. Therefore, for any $x \geq 0$:

$$\mathbb{P}\left(\max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu) \leq x\right) = (1-e^{-x})^{MN}.$$

Using the change of variable $y = 1 - e^{-x}$:

$$
\begin{aligned}
\mathbb{E}\left[\max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu)\right] &= \int_0^{+\infty} \mathbb{P}\left(\max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu) > x\right) \mathrm{d}x \\
&= \int_0^{+\infty} (1 - (1-e^{-x})^{MN})\, \mathrm{d}x \\
&= \int_0^1 \frac{1-y^{MN}}{1-y}\, \mathrm{d}y \\
&= \sum_{r=0}^{MN-1} \frac{1}{r+1} \\
&= \mathcal{O}\left(\log M + \log N\right).
\end{aligned}
$$

This proves (5.37). □

Under Assumption 5.4, (5.37) simply becomes:

$$
\mathbb{E}\left[\max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu)\right] = \mathcal{O}\left(\log N\right) \tag{5.38}
$$

The following lemma is from [94, Lemma 1] that we rewrite here for the sake of completeness. It allows to extend a control from $\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu)$ to $\max_{\nu\in[0,1]} I_{N,\epsilon_j}(\nu)$.

Lemma 5.6. *There exists a universal constant C such that:*

$$
\max_{j\in[M]}\max_{\nu\in[0,1]} I_{N,\epsilon_j}(\nu) \le C \log N \max_{j\in[M]}\max_{\nu\in\mathcal{F}} I_{N,\epsilon_j}(\nu). \tag{5.39}
$$

A direct consequence of Lemma 5.6 used in Lemma 5.5 is that

$$
\mathbb{E}\left[\max_{j\in[M]}\max_{\nu\in[0,1]} I_{N,\epsilon_j}(\nu)\right] = \mathcal{O}\left(\log N(\log N + \log M)\right). \tag{5.40}
$$

The main argument in the proof of Proposition 5.4 is the following result.

Lemma 5.7. *Define*

$$
R_{N,j}(\nu) = \left|\xi_{y_j}(\nu) - h_j(\nu)\xi_{\epsilon_j}(\nu)\right|. \tag{5.41}
$$

Under Assumptions 5.1–5.2, for any $0 < \delta < \frac{1}{2}$,

$$
\max_{\nu\in\mathcal{F}}\max_{j\in[M]} R_{N,j}(\nu) = \mathcal{O}_P(N^{-\delta}). \tag{5.42}
$$

Proof. We closely follow the proof of Theorem 2b from [94]. To prove (5.42), the Markov inequality shows that it is sufficient to prove that for any $\delta < 1/2$,

$$
\mathbb{E}\left[\max_{\nu\in\mathcal{F}}\max_{j\in[M]} R_{N,j}(\nu)\right] = \mathcal{O}(N^{-\delta}).
$$

We use the linear causal representation (5.12) of y_j to write

$$
\begin{aligned}
\xi_{y_j}(\nu) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} y_{j,n} e^{-2i\pi(n-1)\nu} \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\sum_{u=0}^{+\infty} a_{j,u}\epsilon_{j,n-u}\right) e^{-2i\pi(n-1)\nu}.
\end{aligned}
$$

Since almost surely, for all $j \in [M]$, $n \in [N]$, $(a_{j,u}\epsilon_{j,n-u})_{u\geq 0} \in \ell^2(\mathbb{N})$, we can switch the order of summation and make the change of variable $v = n - u$ to get

$$\xi_{y_j}(\nu) = \frac{1}{\sqrt{N}} \sum_{u=0}^{+\infty} a_{j,u} e^{-2i\pi u\nu} \sum_{v=1-u}^{N-u} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}.$$

Define

$$Z_{N,j,u}(\nu) = \left(\sum_{v=1-u}^{N-u} - \sum_{v=1}^{N} \right) \epsilon_{j,v} e^{-2i\pi(v-1)\nu} \tag{5.43}$$

so that $R_{N,j}(\nu)$ can be rewritten as:

$$R_{N,j}(\nu) = \left| \frac{1}{\sqrt{N}} \sum_{u=0}^{+\infty} a_{j,u} e^{-2i\pi u\nu} Z_{N,j,u}(\nu) \right| \tag{5.44}$$

on which one can take the supremum over $j \in [M]$ and $\nu \in \mathcal{F}$ on each side and arrive at the following inequality:

$$\max_{j\in[M]} \max_{\nu\in\mathcal{F}} R_{N,j}(\nu) \leq \frac{1}{\sqrt{N}} \max_{j\in[M]} \sum_{u=0}^{+\infty} |a_{j,u}| \max_{\nu\in\mathcal{F}} |Z_{N,j,u}(\nu)|$$

where the right hand side is also bounded by:

$$\max_{j\in[M]} \max_{\nu\in\mathcal{F}} R_{N,j}(\nu) \leq \frac{1}{\sqrt{N}} \max_{j_1\in[M]} \sum_{u=0}^{+\infty} |a_{j_1,u}| \max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|.$$

Note that for $u = 0$, $\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)| = 0$, so the sum in fact can be written as starting from 1. For any $\gamma < 1$, the Cauchy-Schwarz inequality provides:

$$\begin{aligned} &\frac{1}{\sqrt{N}} \max_{j_1\in[M]} \sum_{u=1}^{+\infty} |a_{j_1,u}| \max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)| \\ &\qquad \leq \frac{1}{\sqrt{N}} \sqrt{\max_{j_1\in[M]} \sum_{u=1}^{+\infty} u^{2\gamma} |a_{j_1,u}|^2} \sqrt{\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \max_{j_2\in[M],\nu\in\mathcal{F}_N} |Z_{N,j_2,u}(\nu)|^2}. \end{aligned}$$

Taking the expectation (and an application of the Jensen inequality to exchange the expectation and the square root), we get the following bound:

$$\begin{aligned} &\mathbb{E}\left[\max_{j\in[M]} \max_{\nu\in\mathcal{F}} R_{N,j}(\nu)\right] \\ &\qquad \leq \frac{1}{\sqrt{N}} \sqrt{\max_{j_1\in[M]} \sum_{u=1}^{+\infty} u^{2\gamma} |a_{j_1,u}|^2} \sqrt{\mathbb{E}\left[\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right]}. \end{aligned} \tag{5.45}$$

Consider the first term in the right hand side of (5.45). We see that we need to transfer the uniform summability property of the sequences $(r_{j,u})_{u\in\mathbb{N}}$ from Assumption 5.2 to a summability property on the sequences $(a_{j,u})_{u\in\mathbb{N}}$ uniformly over the $j \geq 1$ times series. Fortunately, Lemma D.1 from [60], a generalization of the Wiener-Lévy theorem, provides an answer that we rewrite here for sake of completeness.

Lemma 5.8. *(Lemma D.1, [60]) Consider a function $F(z)$ holomorphic in a neighbourhood of the* *interval* $[s_{\min}, s_{\max}]$ *where* $s_{\min}$ *and* $s_{\max}$ *are defined by* (5.7). *Under Assumption* 5.2, *for each* $\gamma < 1$,

$$\sup_{j \geq 1} \sum_{u \in \mathbb{Z}} (1+|u|)^{\gamma} \left| \int_0^1 (F \circ s_j)(\nu) e^{2i\pi\nu u} \, \mathrm{d}u \right| < +\infty.$$

We now show how Lemma 5.8 can be used to find a summability property on the sequences $(a_{j,u})_{u \in \mathbb{N}}$ uniformly in $j \geq 1$. Take $F(z) = \log z$, which is holomorphic on a neighborhood of $[s_{\min}, s_{\max}]$, so for any $\gamma < 1$,

$$\sup_{j \geq 1} \sum_{u \in \mathbb{Z}} (1+|u|)^{\gamma} |c_{j,u}| < +\infty \tag{5.46}$$

where

$$c_{j,u} = \int_0^1 \log(s_j(\nu)) e^{2i\pi\nu u} \, \mathrm{d}u.$$

It is well known (see [80, Theorem 17.17] and [60]) that the sequence $c_{j,u}$ satisfies

$$h_j(\nu) = \exp\left(\frac{c_{j,0}}{2} + \sum_{u=1}^{+\infty} c_{j,u} e^{-2i\pi\nu u} \right).$$

where we recall that $h(\nu) = \sum_{u \in \mathbb{N}} a_{j,u} e^{-2i\pi u\nu}$ coincides with the outer spectral factor of $s_j(\nu) = |h_j(\nu)|^2$. We therefore see that the sequence of coefficients $(a_{j,u})_{u \in \mathbb{N}}$ are related to $(c_{j,u})_{u \in \mathbb{N}}$, and it can be shown (equation (D.11) in [60]) that for each $\gamma < 1$,

$$\sup_{j \geq 1} \sum_{u \geq 0} (1+|u|)^{\gamma} |a_{j,u}| \leq \sup_{j \geq 1} \exp\left(\sum_{u \geq 0} (1+|u|)^{\gamma} |c_{j,u}| \right)$$

which by (5.46) provides for any $\gamma < 1$:

$$\sup_{j \geq 1} \sum_{u \geq 0} (1+|u|)^{\gamma} |a_{j,u}| < +\infty. \tag{5.47}$$

Returning to (5.45), and using (5.47), we find

$$\sup_{j \geq 1} \sum_{u=1}^{+\infty} |u|^{2\gamma} |a_{j,u}|^2 \leq \sup_{j \geq 1} \left(\sum_{u \geq 0} |u|^{\gamma} |a_{j,u}| \right)^2 < +\infty.$$

Consider now the second term in (5.45). For each $u \in \mathbb{N}$, the quantity $u^{-2\gamma} \sup_{j_2 \in [M], \nu \in \mathcal{F}_N} |Z_{N,j_2,u}(\nu)|^2$ is positive so the monotone convergence theorem allows to exchange the sum and the expectation.

$$\mathbb{E}\left[\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \max_{j_2 \in [M], \nu \in \mathcal{F}} |Z_{N,j_2,u}(\nu)|^2 \right] = \sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2 \in [M], \nu \in \mathcal{F}} |Z_{N,j_2,u}(\nu)|^2 \right]$$

so that equation (5.45) becomes

$$\mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in \mathcal{F}_N} R_{N,j}(\nu) \right] \leq \frac{C}{\sqrt{N}} \sqrt{ \sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2 \in [M], \nu \in \mathcal{F}} |Z_{N,j_2,u}(\nu)|^2 \right] } \tag{5.48}$$

for some universal constant $C < +\infty$. To end the proof of Lemma 5.7, it remains to show that for any $\eta > 0$,

$$\sqrt{\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2 \in [M], \nu \in \mathcal{F}_N} |Z_{N,j_2,u}(\nu)|^2\right]} = \mathcal{O}(N^\eta)$$

which is equivalent to showing that for any $\eta > 0$:

$$\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2 \in [M], \nu \in \mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right] = \mathcal{O}(N^\eta). \tag{5.49}$$

We now see that the behaviour of $\mathbb{E}[\max_{j \in [M]} \max_{\nu \in \mathcal{F}_N} R_{N,j}(\nu)]$ is governed by $\mathbb{E}[\max_{j_2 \in [M], \nu \in \mathcal{F}_N} |Z_{N,j_2,u}(\nu)|^2]$, so it remains to study this quantity. By the triangle inequality:

$$|Z_{N,j,u}(\nu)| \leq \begin{cases} \left|\sum_{v=1-u}^{0} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right| + \left|\sum_{v=N-u+1}^{N} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right| & \text{if } u < N \\ \left|\sum_{v=1-u}^{N-u} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right| + \left|\sum_{v=1}^{N} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right| & \text{if } u \geq N \end{cases}$$

and using the inequality $|a+b|^2 \leq 2(|a|^2 + |b|^2)$:

$$|Z_{N,j,u}(\nu)|^2 \leq \begin{cases} 2\left(\left|\sum_{v=1-u}^{0} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2 + \left|\sum_{v=N-u+1}^{N} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2\right) & \text{if } u < N \\ 2\left(\left|\sum_{v=1-u}^{N-u} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2 + \left|\sum_{v=1}^{N} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2\right) & \text{if } u \geq N. \end{cases} \tag{5.50}$$

In the case $u \geq N$, the two sums can be recognized as N times the periodogram estimator which we defined previously as $I_{N,\epsilon_j}(\nu)$. Using the estimates (5.38) from Lemma 5.5:

$$\begin{aligned} \mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in \mathcal{F}} \left|\sum_{v=1-u}^{N-u} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2\right] &= N\mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in \mathcal{F}} I_{N,\epsilon_j}(\nu)\right] \\ &= \mathcal{O}\left(N \log N\right). \end{aligned} \tag{5.51}$$

The other sum in the case $u \geq N$ is similar.

For $u < N$, the two sums have to be handled with more care for two reasons: the summation is only across u terms (instead of N terms) and the frequency ν is of the form $\frac{k}{N}$ instead of the required form $\frac{k}{u}$ to use the bound from Lemma 5.5 (said differently ν is no more a Fourier frequency for a sample size $u < N$). Therefore, we have to estimate the order of magnitude of $I_{u,\epsilon_j}(\nu)$ for $\nu \in [0,1]$ instead of $\nu \in \mathcal{F}$. Lemma 5.6 and especially equation (5.40) provides this.

$$\begin{aligned} \mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in \mathcal{F}} \left|\sum_{v=1-u}^{0} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2\right] &\leq \mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in [0,1]} \left|\sum_{v=1-u}^{0} \epsilon_{j,v} e^{-2i\pi(v-1)\nu}\right|^2\right] \\ &= \mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in [0,1]} u I_{u,\epsilon_j}(\nu)\right] \\ &= \mathcal{O}\left(u \log u \left(\log u + \log M\right)\right). \end{aligned} \tag{5.52}$$

The second sum in the case $u < N$ is also similar, therefore, collecting (5.51) and (5.52) in (5.50), we get:

$$\mathbb{E}\left[\max_{j \in [M]} \max_{\nu \in \mathcal{F}} |Z_{N,j,u}(\nu)|^2\right] = \begin{cases} \mathcal{O}\left(u \log u \log(uM)\right) & \text{if } u < N \\ \mathcal{O}\left(N \log N\right) & \text{if } u \geq N. \end{cases} \tag{5.53}$$

It remains to use these bounds in the left hand side of (5.49).

$$\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right] = \sum_{u=1}^{N-1} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right] + \sum_{u=N}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right]$$

and using the estimates (5.53),

$$\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right] \le \sum_{u=1}^{N-1} \frac{\log u \log(uM)}{u^{2\gamma-1}} + \sum_{u=N}^{+\infty} \frac{N\log N}{u^{2\gamma}}.$$

It is clear that

$$\sum_{u=1}^{N-1} \frac{\log u \log(uM)}{u^{2\gamma-1}} = \mathcal{O}\left(\frac{\log^2 N}{N^{2(\gamma-1)}}\right)$$
$$\sum_{u=N}^{+\infty} \frac{1}{u^{2\gamma}} = \mathcal{O}\left(\frac{1}{N^{2\gamma-1}}\right)$$

so for any $\gamma < 1$,

$$\sum_{u=1}^{+\infty} \frac{1}{u^{2\gamma}} \mathbb{E}\left[\max_{j_2\in[M],\nu\in\mathcal{F}} |Z_{N,j_2,u}(\nu)|^2\right] = \mathcal{O}\left(N^{2(1-\gamma)}\log^2 N + N^{2(1-\gamma)}\log N\right).$$

This quantity is $\mathcal{O}(N^\eta)$ for any $\eta > 0$, which proves (5.49) and ends the proof. □

Proposition 5.4 can now be proved.

Proof. Write $\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)$ as:

$$\begin{aligned}
\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu) &= \frac{1}{B+1}\sum_{b=-B/2}^{B/2} \xi_{y_i}\left(\nu+\frac{b}{N}\right)\overline{\xi_{y_j}\left(\nu+\frac{b}{N}\right)} \\
&\quad - h_i\left(\nu+\frac{b}{N}\right)\xi_{\epsilon_i}\left(\nu+\frac{b}{N}\right)\overline{h_j\left(\nu+\frac{b}{N}\right)\xi_{\epsilon_j}\left(\nu+\frac{b}{N}\right)} \\
&= \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left(\xi_{y_i}\left(\nu+\frac{b}{N}\right) - h_i\left(\nu+\frac{b}{N}\right)\xi_{\epsilon_i}\left(\nu+\frac{b}{N}\right)\right)\overline{\xi_{y_j}\left(\nu+\frac{b}{N}\right)} \\
&\quad + h_i\left(\nu+\frac{b}{N}\right)\xi_{\epsilon_i}\left(\nu+\frac{b}{N}\right)\left(\overline{\xi_{y_j}\left(\nu+\frac{b}{N}\right)} - \overline{h_j\left(\nu+\frac{b}{N}\right)\xi_{\epsilon_j}\left(\nu+\frac{b}{N}\right)}\right).
\end{aligned}$$

We recognize the quantities $R_{N,i}(\nu)$ that have been bounded in Lemma 5.7. It is now clear that:

$$\max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)| \le \max_{i\in[M],\nu\in\mathcal{F}} R_{N,i}(\nu) \times \left(\max_{i\in[M],\nu\in\mathcal{F}} |\xi_{y_i}(\nu)| + \max_{i\in[M],\nu\in\mathcal{F}} |h_i(\nu)\xi_{\epsilon_i}(\nu)|\right). \tag{5.54}$$

By Lemma 5.5:

$$\max_{i\in[M],\nu\in\mathcal{F}} |\xi_{\epsilon_i}(\nu)| = \mathcal{O}_P\left(\sqrt{\log M + \log N}\right) = \mathcal{O}_P\left(\sqrt{\log N}\right)$$

and in conjunction with Lemma 5.7, for any $\delta < 1/2$,

$$\max_{i\in[M]}\max_{\nu\in\mathcal{F}} |\xi_{y_i}(\nu)| \le \underbrace{\max_{i\in[M]}\max_{\nu\in\mathcal{F}} |h_i(\nu)\xi_{\epsilon_i}(\nu)|}_{\mathcal{O}_P(\sqrt{\log N})} + \underbrace{\max_{i\in[M]}\max_{\nu\in\mathcal{F}} |R_{N,i}(\nu)|}_{\mathcal{O}_P(N^{-\delta})}$$

which is $\mathcal{O}_P\left(\sqrt{\log N}\right)$. Each quantity involved in (5.54) is now estimated, and provides, for any $\delta < 1/2$:

$$\max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)| = \mathcal{O}_P(N^{-\delta}\sqrt{\log N}) = \mathcal{O}_P(N^{-\delta'})$$

for any $\delta' < 1/2$. By Assumption 5.4, $\rho < 1$, ie. $\sqrt{B+1} = o(N^{1/2})$ therefore one can always take $\delta' = \frac{\rho/2+1/2}{2} \in (0, 1/2)$ such that:

$$\sqrt{B+1}\max_{(i,j,\nu)\in\mathcal{I}} |\hat{s}_{ij}(\nu) - \tilde{s}_{ij}(\nu)| = \mathcal{O}_P(N^{-\delta'})$$

and we get (5.27). □

5.C Proof of Proposition 5.3: moderate deviations of $\tilde{s}_{ij}(\nu)$

First, we give two preliminary lemmas regarding the concentration of Gaussian sesquilinear forms.

Lemma 5.9. *Let* $\mathbf{x}, \mathbf{y}$ *independent* $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{I})$ *random vectors and* $\mathbf{A}$ *a non-zero* $M \times M$ *deterministic matrix. For any* $t > 0$,

$$\mathbb{P}\left(|\mathbf{x}^*\mathbf{A}\mathbf{y}| > t\right) = \mathbb{E}\left[\exp\left(-\frac{t^2}{\mathbf{y}^*\mathbf{A}^*\mathbf{A}\mathbf{y}}\right)\right]. \tag{5.55}$$

Moreover, if $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{I})$ *is jointly independent from* $\mathbf{x}$ *and* $\mathbf{y}$, *and* $\mathbf{B}$ *is another non zero* $M \times M$ *deterministic matrix, for any* $t, s > 0$,

$$\mathbb{P}\left(|\mathbf{x}^*\mathbf{A}\mathbf{y}| > t,\ |\mathbf{z}^*\mathbf{B}\mathbf{y}| > s\right) = \mathbb{E}\left[\exp\left(-\frac{t^2}{\mathbf{y}^*\mathbf{A}^*\mathbf{A}\mathbf{y}} - \frac{s^2}{\mathbf{y}^*\mathbf{B}^*\mathbf{B}\mathbf{y}}\right)\right]. \tag{5.56}$$

The proof of Lemma 5.9 is straightforward and therefore omitted.

The next lemma is the Hanson-Wright inequality [79] in the special case of a sesquilinear form.

Lemma 5.10. *Let* $\mathbf{x}, \mathbf{y}$ *be independent* $\mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{I})$ *random variables, and* $\mathbf{A}$ *a deterministic* $M \times M$ *matrix. Then, for any* $t \ge 0$:

$$\mathbb{P}\left(|\mathbf{x}^*\mathbf{A}\mathbf{y} - \mathbb{E}\left[\mathbf{x}^*\mathbf{A}\mathbf{y}\right]| > t\right) \le 2\exp\left(-C\min\left(\frac{t}{\|\mathbf{A}\|}, \frac{t^2}{\|\mathbf{A}\|_F^2}\right)\right)$$

where C *is a universal constant (independent of* t *and* $\mathbf{A}$*).*

To prove Proposition 5.3, we recall from (5.14) that $\tilde{s}_{ij}(\nu)$ may be written as the Gaussian sesquilinear form

$$\tilde{s}_{ij}(\nu) = \boldsymbol{\xi}_{\epsilon_j}(\nu)^* \frac{\mathbf{\Pi}_{ij}(\nu)}{\sqrt{B+1}} \boldsymbol{\xi}_{\epsilon_i}(\nu)$$

where $\boldsymbol{\xi}_{\epsilon_1}(\nu), \ldots, \boldsymbol{\xi}_{\epsilon_M}(\nu)$ are i.i.d. $\mathcal{N}_{\mathbb{C}^{B+1}}(\mathbf{0}, \mathbf{I})$ distributed, and that we denote $\sigma_{ij}^2(\nu) = \frac{1}{B+1}\text{tr } \boldsymbol{\Sigma}_{ij}(\nu)$ with

$$\begin{aligned}\boldsymbol{\Sigma}_{ij}(\nu) &= \boldsymbol{\Pi}_{ij}(\nu)\boldsymbol{\Pi}_{ij}(\nu)^* \\ &= \text{dg}\left(s_i\left(\nu + \frac{b}{N}\right)s_j\left(\nu + \frac{b}{N}\right) : b = -\frac{B}{2}, \ldots, \frac{B}{2}\right).\end{aligned}$$

Note also that thanks to Assumptions 5.2 and 5.3, there exist $s_{\min}, s_{\max} > 0$ such that:

$$0 < s_{\min} \leq \inf_{m\geq 1} \min_{\nu\in[0,1]} s_m(\nu) \leq \sup_{m\geq 1} \max_{\nu\in[0,1]} s_m(\nu) \leq s_{\max} < +\infty$$

and consequently, the following inequality holds:

$$0 < s_{\min}^2 \leq \inf_{N\geq 1} \min_{(i,j,\nu)\in\mathcal{I}} \lambda_{\min}\left(\boldsymbol{\Sigma}_{ij}(\nu)\right) < \sup_{N\geq 1} \max_{(i,j,\nu)\in\mathcal{I}} \lambda_{\max}\left(\boldsymbol{\Sigma}_{ij}(\nu)\right) \leq s_{\max}^2 < +\infty, \tag{5.57}$$

where $\lambda_{\min}\left(\boldsymbol{\Sigma}_{ij}(\nu)\right)$, $\lambda_{\max}\left(\boldsymbol{\Sigma}_{ij}(\nu)\right)$ are respectively the smallest and largest eigenvalue (or diagonal entry) of $\boldsymbol{\Sigma}_{ij}(\nu)$.

In the remainder, to lighten the presentation, we use the multi-index α instead of (i, j, ν) as well as the notation $\cdot_\alpha$ in place of $\cdot_{ij}(\nu)$ so that, for example, $\tilde{s}_{ij}(\nu)$, $\boldsymbol{\Pi}_{ij}(\nu)$, $\boldsymbol{\Sigma}_{ij}(\nu)$ become $\tilde{s}_\alpha$, $\boldsymbol{\Pi}_\alpha$ and $\boldsymbol{\Sigma}_\alpha$ respectively.

From Lemma 5.9, the probabilities appearing in (5.25) and (5.26) in the statement of Proposition 5.3 can be rewritten as

$$\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2} > t^2\right) = \mathbb{E}\left[\exp\left(-\frac{\text{tr } \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}t^2\right)\right] \tag{5.58}$$

and

$$\begin{aligned}&\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2} > t^2, (B+1)\frac{|\tilde{s}_{\alpha'}|^2}{\sigma_{\alpha'}^2} > s^2\right) \\ &\qquad = \mathbb{E}\left[\exp\left(-\left(\frac{\text{tr } \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}t^2 + \frac{\text{tr } \boldsymbol{\Sigma}_{\alpha'}}{\mathbf{w}^*\boldsymbol{\Sigma}_{\alpha'}\mathbf{w}}s^2\right)\right)\right].\end{aligned} \tag{5.59}$$

for some $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}^{B+1}}(\mathbf{0}, \mathbf{I})$. The next two lemmas are dedicated to the study of the concentration of $\text{tr } \boldsymbol{\Sigma}_\alpha / \mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}$ around 1.

Lemma 5.11. *There exists two universal constants C_1, C_2 such that for all $t \in (0, 1)$,*

$$\max_{\alpha\in\mathcal{I}} \mathbb{P}\left(\left|\frac{\text{tr } \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - 1\right| > t\right) \leq C_1 \exp\left(-C_2 B t^2\right). \tag{5.60}$$

Proof. We have

$$\begin{aligned}&\max_{\alpha\in\mathcal{I}} \mathbb{P}\left(\left|\frac{\text{tr } \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - 1\right| > t\right) \\ &\qquad = \max_{\alpha\in\mathcal{I}} \mathbb{P}\left(\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w} - \text{tr } \boldsymbol{\Sigma}_\alpha \in \left[\frac{-t}{1+t}\text{tr } \boldsymbol{\Sigma}_\alpha, \frac{t}{1-t}\text{tr } \boldsymbol{\Sigma}_\alpha\right]^c\right) \\ &\qquad \leq \max_{\alpha\in\mathcal{I}} \mathbb{P}\left(\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w} - \text{tr } \boldsymbol{\Sigma}_\alpha \in \left[-\frac{t}{2}\text{tr } \boldsymbol{\Sigma}_\alpha, \frac{t}{2}\text{tr } \boldsymbol{\Sigma}_\alpha\right]^c\right) \\ &\qquad \leq \max_{\alpha\in\mathcal{I}} \mathbb{P}\left(\left|\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w} - \text{tr } \boldsymbol{\Sigma}_\alpha\right| > \frac{t}{2}\text{tr } \boldsymbol{\Sigma}_\alpha\right).\end{aligned}$$

Since $\mathbf{w} \sim \mathcal{N}_{\mathbb{C}^{B+1}}(\mathbf{0}, \mathbf{I})$, the Hanson-Wright inequality [79] provides that

$$\begin{aligned}\mathbb{P}&\left(|\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w} - \text{tr }\mathbf{\Sigma}_\alpha| > \frac{t}{2}\text{tr }\mathbf{\Sigma}_\alpha\right)\\ &\leq 2\exp\left(-C\min\left(\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\|\mathbf{\Sigma}_\alpha\|}t, \frac{(\text{tr }\mathbf{\Sigma}_\alpha)^2}{\|\mathbf{\Sigma}_\alpha\|_F^2}t^2\right)\right)\\ &\leq 2\exp\left(-Ct^2\min\left(\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\|\mathbf{\Sigma}_\alpha\|}, \frac{(\text{tr }\mathbf{\Sigma}_\alpha)^2}{\|\mathbf{\Sigma}_\alpha\|_F^2}\right)\right)\end{aligned}$$

for some universal constant C, where $\|\mathbf{\Sigma}_\alpha\|$ and $\|\mathbf{\Sigma}_\alpha\|_F$ denote the spectral norm and Frobenius norm of $\mathbf{\Sigma}_\alpha$ respectively. From (5.57), we also have

$$\begin{aligned}\min_{\alpha\in\mathcal{I}}\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\|\mathbf{\Sigma}_\alpha\|} &\geq (B+1)\frac{s_{\text{min}}^2}{s_{\text{max}}^2}\\ \min_{\alpha\in\mathcal{I}}\frac{(\text{tr }\mathbf{\Sigma}_\alpha)^2}{\|\mathbf{\Sigma}_\alpha\|_F^2}&(B+1)\frac{s_{\text{min}}^4}{s_{\text{max}}^4}.\end{aligned}$$

Consequently, we can find some universal constants C_1, C_2 such that for all $N \geq 1$, (5.60) holds. □

Lemma 5.12. *For any* $\beta \in \left(0, \frac{1}{2}\right)$,

$$\max_{\alpha\in\mathcal{I}}\left|\mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right] - 1\right| = \mathcal{O}(B^{-\beta}) \tag{5.61}$$

Proof. Define the event

$$\Omega_{\alpha,N} := \left\{\left|\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}} - 1\right| < \kappa_N\right\}$$

where κ_N is some sequence satisfying $\kappa_N \to 0$ as $N \to +\infty$ and consider the decomposition

$$\left|\mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right] - 1\right| \leq \left|\mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\mathbb{1}_{\Omega_{\alpha,N}}\right] - 1\right| + \mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\mathbb{1}_{\Omega_{\alpha,N}^c}\right]. \tag{5.62}$$

For the first term of the right-hand side of (5.62), the following bound holds:

$$\left|\mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\mathbb{1}_{\Omega_{\alpha,N}}\right] - 1\right| \leq \kappa_N + \mathbb{P}\left(\Omega_{\alpha,N}^c\right). \tag{5.63}$$

Regarding the second term, Cauchy-Schwarz inequality implies that

$$\mathbb{E}\left[\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\mathbb{1}_{\Omega_{\alpha,N}^c}\right] \leq \sqrt{\mathbb{E}\left[\left|\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right|^2\right]}\sqrt{\mathbb{P}\left(\Omega_{\alpha,N}^c\right)}. \tag{5.64}$$

Using (5.57), we have

$$\max_{\alpha\in\mathcal{I}}\frac{\text{tr }\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}} \leq (B+1)\frac{s_{\text{max}}^2}{s_{\text{min}}^2}\frac{1}{\|\mathbf{w}\|^2}.$$

and since $\frac{1}{2\|\mathbf{w}\|^2}$ is distributed as an inverse–χ^2 random variable with $2(B+1)$ degrees of freedom, we have from [75, Appendix A6] that

$$\mathbb{E}\left[\frac{1}{\|\mathbf{w}\|^4}\right] = \mathcal{O}\left(\frac{1}{B^2}\right)$$

which yields

$$\sup_{N\geq 1}\max_{\alpha\in\mathcal{I}}\mathbb{E}\left[\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right|^2\right]<+\infty.$$

Consequently, gathering (5.63) and (5.64) and using Lemma 5.11, we get

$$\begin{aligned}\max_{\alpha\in\mathcal{I}}\left|\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]-1\right| &\leq \kappa_N+C\sqrt{\mathbb{P}\left(\Omega_{\alpha,N}^c\right)}\\ &\leq \kappa_N+C_1\exp\{-C_2\kappa_N^2 B\}\end{aligned}$$

for some universal constants C_1, C_2. Choosing $\kappa_N = B^{-\beta}$ with $\beta\in(0,1/2)$ yields the desired result.

□

Before proving Proposition 5.3, we need one last result on the concentration of $\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}$ around its mean, which is a straightforward consequence of the previous Lemmas 5.11 and 5.12.

Lemma 5.13. *Let $\delta\in(0,\frac{1}{2})$ and $(\epsilon_N)_{N\geq 1}$ some non-negative sequence converging towards 0 as $N\to\infty$ and such that $\epsilon_N B^\delta\to+\infty$. Then, there exist two universal constants C_1, C_2 such that*

$$\max_{\alpha\in\mathcal{I}}\mathbb{P}\left(\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|>\epsilon_N\right)\leq C_1\exp\left(-C_2\epsilon_N^2 B\right).$$

Proof. Write:

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|>\epsilon_N\right)\\ &\qquad\leq\mathbb{P}\left(\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-1\right|>\epsilon_N-\left|1-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|\right).\end{aligned}$$

From Lemma 5.12, there exists a universal constant C such that

$$\max_{\alpha\in\mathcal{I}}\left|1-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|\leq\frac{C}{B^\delta}.$$

Moreover, by assumption on the rate of ϵ_N, we have $\frac{C}{B^\delta}<\frac{\epsilon_N}{2}$ for all large N. Consequently,

$$\max_{\alpha\in\mathcal{I}}\mathbb{P}\left(\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|>\epsilon_N\right)\leq\max_{\alpha\in\mathcal{I}}\mathbb{P}\left(\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-1\right|>\frac{\epsilon_N}{2}\right)\tag{5.65}$$

for all large N. Applying directly Lemma 5.11 to (5.65) allows to conclude the proof.

□

Endowed with Lemmas 5.11, 5.12 and 5.13, we are now in a position to complete the proof of Proposition 5.3.

We first tackle (5.25) and show as a first step that there exists $\eta>0$ such that for any universal constant C,

$$\max_{t\in[0,CB^\eta]}\max_{\alpha\in\mathcal{I}}\left|\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2}>t^2\right)\exp\left(\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}t^2\right]\right)-1\right|=o(1).\tag{5.66}$$

Let $\delta\in\left(0,\frac{1}{2}\right)$ and $(\epsilon_N)_{N\geq 1}$ some non-negative sequence converging to 0 and satisfying $\epsilon_N B^\delta\to+\infty$, and define the event

$$\Theta_{\alpha,N}:=\left\{\left|\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}-\mathbb{E}\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right|<\epsilon_N\right\}\tag{5.67}$$

as in Lemma 5.13. Next, consider the decomposition

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\left(\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\left(\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\mathbb{1}_{\Theta_{\alpha,N}}\right] \\
&\quad + \mathbb{E}\left[\exp\left(-\left(\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\mathbb{1}_{\Theta^c_{\alpha,N}}\right] \\
&:= \Psi_{\alpha,N}(t) + \Delta_{\alpha,N}(t).
\end{aligned} \tag{5.68}$$

On the event $\Theta_{\alpha,N}$, we have

$$\exp\left(-\left(\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right) \in \left[\exp\left(-\epsilon_N t^2\right), \exp\left(\epsilon_N t^2\right)\right]$$

which implies, that:

$$\begin{aligned}
|\Psi_{\alpha,N}(t) - 1| &\le \max\left(1 - \mathbb{P}[\Theta_{\alpha,N}]e^{-\epsilon_N t^2}, \mathbb{P}[\Theta_{\alpha,N}]e^{\epsilon_N t^2} - 1\right) \\
&\le \left|e^{\epsilon_N t^2} - 1\right| + \left(1 - e^{-\epsilon_N t^2}\right) + (e^{\epsilon_N t^2} + e^{-\epsilon_N t^2})\mathbb{P}[\Theta^c_{\alpha,N}].
\end{aligned}$$

Using Lemma 5.13, we further have

$$\begin{aligned}
&\max_{\alpha\in\mathcal{I}} |\Psi_{\alpha,N}(t) - 1| \\
&\qquad \le \left|e^{\epsilon_N t^2} - 1\right| + \left(1 - e^{-\epsilon_N t^2}\right) + (e^{\epsilon_N t^2} + e^{-\epsilon_N t^2})C_1 e^{-C_2\epsilon_N^2 B}
\end{aligned} \tag{5.69}$$

for some universal constants C_1, C_2. Regarding $\Delta_{\alpha,N}(t)$, we clearly have

$$|\Delta_{\alpha,N}(t)| \le \mathbb{P}\left(\Theta^c_{\alpha,N}\right)\exp\left(\mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]t^2\right).$$

Using Lemmas 5.12 and 5.13, for any $\beta \in \left(0, \frac{1}{2}\right)$, there exists a universal constant C_3 such that

$$\max_{\alpha\in\mathcal{I}} |\Delta_{\alpha,N}(t)| \le C_1 \exp\left(-C_2 B\epsilon_N^2 + \left(1 + \frac{C_3}{B^\beta}\right)t^2\right). \tag{5.70}$$

Combining (5.68), (5.69) and (5.70), one gets

$$\begin{aligned}
&\max_{\alpha\in\mathcal{I}} \left|\mathbb{E}\left[\exp\left(-\left(\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\right] - 1\right| \\
&\qquad \le \left|e^{\epsilon_N t^2} - 1\right| + \left(1 - e^{-\epsilon_N t^2}\right) + (e^{\epsilon_N t^2} + e^{-\epsilon_N t^2})C_1 e^{-C_2\epsilon_N^2 B} \\
&\qquad\quad + C_1 \exp\left(-C_2 B\epsilon_N^2 + \left(1 + \frac{C_3}{B^\beta}\right)t^2\right).
\end{aligned}$$

Set $\epsilon_N = B^{-\frac{\delta}{2}}$ so that $\epsilon_N \to 0$ and $\epsilon_N B^\delta \to +\infty$ as required, and let $\eta = \frac{\delta}{8}$. Then, recalling that

$\delta \in \left(0, \frac{1}{2}\right)$, we have

$$\begin{aligned}
\epsilon_N B^{2\eta} &= \frac{1}{B^{\frac{\delta}{4}}} \xrightarrow[N\to+\infty]{} 0, \\
\epsilon_N^2 B &= B^{1-\delta} \xrightarrow[N\to+\infty]{} +\infty, \\
\frac{B^{2\eta}}{\epsilon_N^2 B} &= B^{2\eta+\delta-1} = B^{\frac{5}{4}\delta-1} \xrightarrow[N\to+\infty]{} 0.
\end{aligned}$$

Therefore, for any universal constant C,

$$\max_{t\in[0,CB^\eta]} \max_{\alpha\in\mathcal{I}} \left| \mathbb{E}\left[\exp\left(-\left(\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\right] - 1 \right| \xrightarrow[N\to\infty]{} 0,$$

which, thanks to (5.58), implies (5.66). Finally, using Lemma 5.9, we deduce that

$$\begin{aligned}
&\max_{t\in[0,CB^\eta]} \max_{\alpha\in\mathcal{I}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2} > t^2\right) \left|\exp\left(\mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]t^2\right) - e^{t^2}\right| \qquad (5.71)\\
&\quad\le \max_{t\in[0,CB^\eta]} \max_{\alpha\in\mathcal{I}} \mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2} > t^2\right) \exp\left(\mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]t^2\right) \\
&\qquad\qquad \times \max_{t\in[0,CB^\eta]} \max_{\alpha\in\mathcal{I}} \left|1 - \exp\left(\left(1 - \mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right)\right| \\
&\quad\le (1+o(1))\left(1 - \exp\left(\mathcal{O}\left(\frac{B^{2\eta}}{B^\delta}\right)\right)\right) \\
&\quad\xrightarrow[N\to\infty]{} 0,
\end{aligned}$$

which, combined with (5.58), shows (5.25).

We now turn to (5.26). Since the proof is very similar to the one of (5.25), we only provide the main steps. Using (5.59), we consider the following decomposition:

$$\begin{aligned}
&\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma_\alpha^2} > t, (B+1)\frac{|\tilde{s}_{\alpha'}|^2}{\sigma_{\alpha'}^2} > s\right) \\
&\qquad\qquad \times \exp\left(\mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]t^2 + \mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_{\alpha'}}{\mathbf{w}^*\mathbf{\Sigma}_{\alpha'}\mathbf{w}}\right]s^2\right) \\
&\qquad\qquad\qquad\qquad := \Psi_{\alpha,\alpha',N}(t,s) + \Delta_{\alpha,\alpha',N}(t,s)
\end{aligned}$$

where

$$\begin{aligned}
&\Psi_{\alpha,\alpha',N}(t,s) = \\
&\quad \mathbb{E}\Bigg[\exp\left(-\left(\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_\alpha}{\mathbf{w}^*\mathbf{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right) \\
&\qquad \times \exp\left(-\left(\frac{\operatorname{tr}\,\mathbf{\Sigma}_{\alpha'}}{\mathbf{w}^*\mathbf{\Sigma}_{\alpha'}\mathbf{w}} - \mathbb{E}\left[\frac{\operatorname{tr}\,\mathbf{\Sigma}_{\alpha'}}{\mathbf{w}^*\mathbf{\Sigma}_{\alpha'}\mathbf{w}}\right]\right)s^2\right) \mathbb{1}(\Theta_{\alpha,N}\cap\Theta_{\alpha',N})\Bigg]
\end{aligned}$$

and

$$\Delta_{\alpha,\alpha',N}(t,s) = \mathbb{E}\Bigg[\exp\left(-\left(\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}} - \mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]\right)t^2\right) \times \exp\left(-\left(\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_{\alpha'}}{\mathbf{w}^*\boldsymbol{\Sigma}_{\alpha'}\mathbf{w}} - \mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_{\alpha'}}{\mathbf{w}^*\boldsymbol{\Sigma}_{\alpha'}\mathbf{w}}\right]\right)s^2\right)\mathbb{1}(\Theta^c_{\alpha,N}\cup\Theta^c_{\alpha',N})\Bigg].$$

Using exactly the same arguments as for (5.69) and (5.70) and keeping the same requirements as above regarding the behaviour of sequence $(\epsilon_N)_{N\geq 1}$ and constant η, we may show that

$$\max_{t,s\in[0,CB^\eta]}\max_{\substack{\alpha\in\mathcal{I}\\ \alpha'\in\mathcal{I}_\alpha}}|\Psi_{\alpha,\alpha',N}(t,s)-1| \xrightarrow[N\to\infty]{} 0,$$

as well as

$$\max_{t,s\in[0,CB^\eta]}\max_{\substack{\alpha\in\mathcal{I}\\ \alpha'\in\mathcal{I}_\alpha}}\Delta_{\alpha,\alpha',N}(t,s) \xrightarrow[N\to\infty]{} 0.$$

Consequently,

$$\max_{t,s\in[0,CB^\eta]}\max_{\substack{\alpha\in\mathcal{I}\\ \alpha'\in\mathcal{I}_\alpha}}\Bigg|\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma^2_\alpha} > t, (B+1)\frac{|\tilde{s}_{\alpha'}|^2}{\sigma^2_{\alpha'}} > s\right) \times \exp\left(\mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]t^2 + \mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_{\alpha'}}{\mathbf{w}^*\boldsymbol{\Sigma}_{\alpha'}\mathbf{w}}\right]s^2\right) - 1\Bigg| \xrightarrow[N\to\infty]{} 0.$$

As for (5.71), we also have using a similar bound,

$$\max_{t,s\in[0,CB^\eta]}\max_{\substack{\alpha\in\mathcal{I}\\ \alpha'\in\mathcal{I}_\alpha}}\mathbb{P}\left((B+1)\frac{|\tilde{s}_\alpha|^2}{\sigma^2_\alpha} > t, (B+1)\frac{|\tilde{s}_{\alpha'}|^2}{\sigma^2_{\alpha'}} > s\right) \times \left|\exp\left(\mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_\alpha}{\mathbf{w}^*\boldsymbol{\Sigma}_\alpha\mathbf{w}}\right]t^2 + \mathbb{E}\left[\frac{\mathrm{tr}\ \boldsymbol{\Sigma}_{\alpha'}}{\mathbf{w}^*\boldsymbol{\Sigma}_{\alpha'}\mathbf{w}}\right]s^2\right) - \exp\left(t^2+s^2\right)\right| \xrightarrow[N\to\infty]{} 0.$$

The two previous convergences combined together complete the proof of (5.26).

Bibliography

[1] Hong-Zhi An, Zhao-Guo Chen, and E.J Hannan. “The maximum of the periodogram”. In: *Journal of Multivariate Analysis* 13.3 (1983), pp. 383–400. ISSN: 0047-259X. DOI: https://doi.org/10.1016/0047-259X(83)90017-9. URL: https://www.sciencedirect.com/science/article/pii/0047259X83900179.

[2] Theodore W Anderson. *An introduction to multivariate statistical analysis*. Tech. rep. 1958.

[3] Richard Arratia, Larry Goldstein, and Louis Gordon. “Two moments suffice for Poisson approximations: the Chen-Stein method”. In: *The Annals of Probability* 17.1 (1989), pp. 9–25. URL: https://doi.org/10.1214/aop/1176991491.

[4] Z. D. Bai and Jack W. Silverstein. “No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large-Dimensional Sample Covariance Matrices”. In: *The Annals of Probability* 26.1 (1998), pp. 316–345. ISSN: 00911798. URL: http://www.jstor.org/stable/2652903.

[5] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*. Vol. 20. Springer, 2010.

[6] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. “Phase Transition of the Largest Eigenvalue for Nonnull Complex Sample Covariance Matrices”. In: *The Annals of Probability* 33.5 (2005), pp. 1643–1697. ISSN: 00911798. URL: http://www.jstor.org/stable/3481698.

[7] Jinho Baik and Jack W. Silverstein. “Eigenvalues of large sample covariance matrices of spiked population models”. In: *Journal of Multivariate Analysis* 97.6 (2006), pp. 1382–1408. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2005.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0047259X0500134X.

[8] Florent Benaych-Georges and Antti Knowles. *Lectures on the local semicircle law for Wigner matrices*. 2018. arXiv: 1601.04055 [math.PR].

[9] Florent Benaych-Georges and Raj Rao Nadakuditi. “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices”. In: *Advances in Mathematics* 227.1 (2011), pp. 494–521. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2011.02.007. URL: https://www.sciencedirect.com/science/article/pii/S0001870811000570.

[10] R Yu Bentkus and RA Rudzkis. “On the distribution of some statistical estimates of spectral density”. In: *Theory of Probability & Its Applications* 27.4 (1983), pp. 795–814.

[11] Hilmar Böhm and Rainer von Sachs. “Shrinkage estimation in the frequency domain of multivariate time series”. In: *Journal of Multivariate Analysis* 100.5 (2009), pp. 913–935. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2008.09.009. URL: https://www.sciencedirect.com/science/article/pii/S0047259X08001942.

[12] A.-J. Boonstra and A.-J. van der Veen. “Gain calibration methods for radio telescope arrays”. In: *IEEE Transactions on Signal Processing* 51.1 (2003), pp. 25–38. DOI: 10.1109/TSP.2002.806588.

[13] David R. Brillinger. *Time series: data analysis and theory*. Vol. 36. SIAM, 2001.

[14] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Reprint of the second (1991) edition. Springer, New York, 2006, pp. xvi+577. ISBN: 978-1-4419-0319-8.

[15] T. Tony Cai and Zongming Ma. “Optimal hypothesis testing for high dimensional covariance matrices”. In: *Bernoulli* 19.5B (2013), pp. 2359–2388. ISSN: 13507265. URL: http://www.jstor.org/stable/42919363.

[16] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. “The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations”. In: *The Annals of Probability* 37.1 (2009), pp. 1–47. DOI: 10.1214/08-AOP394. URL: https://doi.org/10.1214/08-AOP394.

[17] Xi Chen and Weidong Liu. “Testing independence with high-dimensional correlated samples”. In: *The Annals of Statistics* 46.2 (2018), pp. 866–894. DOI: 10.1214/17-AOS1571. URL: https://doi.org/10.1214/17-AOS1571.

[18] Romain Couillet and Walid Hachem. “Analysis of the limiting spectral measure of large random matrices of the separable covariance type”. In: *Random Matrices: Theory and Applications* 03.04 (2014), p. 1450016. DOI: 10.1142/S2010326314500166. eprint: https://doi.org/10.1142/S2010326314500166. URL: https://doi.org/10.1142/S2010326314500166.

[19] Herbert T. Davis. “Asymptotic Properties of Certain Functionals of the Periodogram”. In: *Journal of Applied Probability* 11.3 (1974), pp. 578–581. ISSN: 00219002. URL: http://www.jstor.org/stable/3212703.

[20] Richard A. Davis and Thomas Mikosch. “The Maximum of the Periodogram of a Non-Gaussian Sequence”. In: *The Annals of Probability* 27.1 (1999), pp. 522–536. ISSN: 00911798. URL: http://www.jstor.org/stable/2652883.

[21] Holger Dette and Nina Dörnemann. “Likelihood ratio tests for many groups in high dimensions”. In: *Journal of Multivariate Analysis* 178 (2020), p. 104605. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2020.104605. URL: https://www.sciencedirect.com/science/article/pii/S0047259X1930346X.

[22] Pierre Duchesne and Roch Roy. “Robust tests for independence of two time seires”. In: *Statistica Sinica* 13.3 (2003), pp. 827–852. ISSN: 10170405, 19968507. URL: http://www.jstor.org/stable/24307126.

[23] Evsey Mordukhovich Dyn’kin. “An operator calculus based on the Cauchy–Green formula”. In: *Zapiski Nauchnykh Seminarov POMI* 30 (1972), pp. 33–39.

[24] Michael Eichler. “A frequency-domain based test for non-correlation between stationary time series”. In: *Metrika* 65.2 (2007), pp. 133–157. DOI: 10.1007/s00184-006-0065-8. URL: https://doi.org/10.1007/s00184-006-0065-8.

[25] Michael Eichler. “Testing nonparametric and semiparametric hypotheses in vector stationary processes”. In: *Journal of Multivariate Analysis* 99.5 (2008), pp. 968–1009. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2007.06.003. URL: https://www.sciencedirect.com/science/article/pii/S0047259X0700084X.

[26] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events: for insurance and finance*. Vol. 33. Springer Science & Business Media, 2013.

[27] László Erdos, Antti Knowles, and Horng-Tzer Yau. “Averaging fluctuations in resolvents of random band matrices”. In: *Annales Henri Poincaré*. Vol. 14. 8. Springer. 2013, pp. 1837–1926.

[28] Jianqing Fan and Tiefeng Jiang. “Largest entries of sample correlation matrices from equi-correlated normal populations”. In: *The Annals of Probability* 47.5 (2019), pp. 3321–3374. DOI: 10.1214/19-AOP1341. URL: https://doi.org/10.1214/19-AOP1341.

[29] Mark Fiecas and Rainer von Sachs. “Data-driven shrinkage of the spectral density matrix of a high-dimensional time series”. In: *Electronic Journal of Statistics* 8.2 (2014), pp. 2975–3003. DOI: 10.1214/14-EJS977. URL: https://doi.org/10.1214/14-EJS977.

[30] R. A. Fisher. “Tests of Significance in Harmonic Analysis”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 125.796 (1929), pp. 54–59. ISSN: 09501207. URL: http://www.jstor.org/stable/95247.

[31] Mario Forni et al. “The generalized dynamic factor model consistency and rates”. In: *Journal of Econometrics* 119.2 (2004). Dynamic factor models, pp. 231–255. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(03)00196-9. URL: https://www.sciencedirect.com/science/article/pii/S0304407603001969.

[32] Mario Forni et al. “The Generalized Dynamic-Factor Model: Identification and Estimation”. In: *The Review of Economics and Statistics* 82.4 (2000), pp. 540–554. ISSN: 00346535, 15309142. URL: http://www.jstor.org/stable/2646650.

[33] Alice Guionnet and Ofer Zeitouni. “Concentration of the Spectral Measure for Large Matrices”. In: *Electron. Commun. Probab.* 5 (2000), pp. 119–136. DOI: 10.1214/ECP.v5-1026. URL: https://doi.org/10.1214/ECP.v5-1026.

[34] Uffe Haagerup and Steen Thorbjornsen. “Random matrices with complex Gaussian entries”. In: *Expositiones Mathematicae* 21.4 (2003), pp. 293–337. ISSN: 0723-0869. DOI: https://doi.org/10.1016/S0723-0869(03)80036-1. URL: https://www.sciencedirect.com/science/article/pii/S0723086903800361.

[35] Walid Hachem, Philippe Loubaton, and Jamal Najim. “Deterministic Equivalents for Certain Functionals of Large Random Matrices”. In: *The Annals of Applied Probability* 17.3 (2007), pp. 875–930. ISSN: 10505164. URL: http://www.jstor.org/stable/25442865.

[36] Walid Hachem et al. “A New Approach for Mutual Information Analysis of Large Dimensional Multi-Antenna Channels”. In: *IEEE Transactions on Information Theory* 54.9 (2008), pp. 3987–4004. DOI: 10.1109/TIT.2008.928229.

[37] Walid Hachem et al. “On bilinear forms based on the resolvent of large random matrices”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 49. 1. 2013, pp. 36–63. DOI: 10.1214/11-AIHP450. URL: http://www.numdam.org/articles/10.1214/11-AIHP450/.

[38] Larry D. Haugh. “Checking the Independence of Two Covariance-Stationary Time Series: A Univariate Residual Cross-Correlation Approach”. In: *Journal of the American Statistical Association* 71.354 (1976), pp. 378–385. ISSN: 01621459. URL: http://www.jstor.org/stable/2285318.

[39] Bernard Helffer and J Sjostrand. “Equation de Schrödinger avec champ magnétique et équation de Harper”. In: *Schrödinger operators*. Springer, 1989, pp. 118–197.

[40] Khalid El Himdi and Roch Roy. “Tests for Noncorrelation of Two Multivariate ARMA Time Series”. In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 25.2 (1997), pp. 233–256. ISSN: 03195724. URL: http://www.jstor.org/stable/3315734.

[41] Khalid El Himdi, Roch Roy, and Pierre Duchesne. “Tests for Non-Correlation of Two Multivariate Time Series: A Nonparametric Approach”. In: *Lecture Notes-Monograph Series* 42 (2003), pp. 397–416. ISSN: 07492170. URL: http://www.jstor.org/stable/4356250.

[42] Yongmiao Hong. “Testing for Independence Between Two Covariance Stationary Time Series”. In: *Biometrika* 83.3 (1996), pp. 615–625. ISSN: 00063444. URL: http://www.jstor.org/stable/2337513.

[43] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[44] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991. DOI: 10.1017/CBO9780511840371.

[45] Tiefeng Jiang. "The Asymptotic Distributions of the Largest Entries of Sample Correlation Matrices". In: *The Annals of Applied Probability* 14.2 (2004), pp. 865–880. ISSN: 10505164. URL: http://www.jstor.org/stable/4140432.

[46] Tiefeng Jiang. "The Limiting Distributions of Eigenvalues of Sample Correlation Matrices". In: *Sankhyā: The Indian Journal of Statistics (2003-2007)* 66.1 (2004), pp. 35–48. ISSN: 09727671. URL: http://www.jstor.org/stable/25053330.

[47] Tiefeng Jiang and Fan Yang. "Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions". In: *The Annals of Statistics* 41.4 (2013), pp. 2029–2074. DOI: 10.1214/13-AOS1134. URL: https://doi.org/10.1214/13-AOS1134.

[48] Abla Kammoun and Mohamed-Slim Alouini. "No Eigenvalues Outside the Limiting Support of Generally Correlated Gaussian Matrices". In: *IEEE Transactions on Information Theory* 62.7 (2016), pp. 4312–4326. DOI: 10.1109/TIT.2016.2561998.

[49] Noureddine El Karoui. "Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond". In: *The Annals of Applied Probability* 19.6 (2009), pp. 2362–2405. DOI: 10.1214/08-AAP548. URL: https://doi.org/10.1214/08-AAP548.

[50] Nick Klausner, Mahmood R. Azimi-Sadjadi, and Louis L. Scharf. "Detection of Spatially Correlated Time Series From a Network of Sensor Arrays". In: *IEEE Transactions on Signal Processing* 62.6 (2014), pp. 1396–1407. DOI: 10.1109/TSP.2014.2298833.

[51] Paul D. Koch and Shie-Shien Yang. "A Method for Testing the Independence of Two Time Series That Accounts for a Potential Pattern in the Cross-Correlation Function". In: *Journal of the American Statistical Association* 81.394 (1986), pp. 533–544. ISSN: 01621459. URL: http://www.jstor.org/stable/2289246.

[52] Lambert H. Koopmans. *The spectral analysis of time series*. Vol. 22. Probability and Mathematical Statistics, Academic Press, 1995.

[53] M. G. Krein and Adolf Abramovich Nudelman. *The Markov moment problem and extremal problems : ideas and problems of P. L. Cebysev and A. A. Markov and their further development / by M. G. Krein and A. A. Nudelman ; [translated from the Russian by Israel Program for Scientific Translations, translated by D. Louvish]*. English. American Mathematical Society Providence, R.I, 1977, v, 417 p. ISBN: 0821845004.

[54] A. Leshem and A.-J. van der Veen. "Multichannel detection of Gaussian signals with uncalibrated receivers". In: *IEEE Signal Processing Letters* 8.4 (2001), pp. 120–122. DOI: 10.1109/97.911477.

[55] W. K Li and Y. V Hui. "Robust residual cross correlation tests for lagged relations in time series". In: *Journal of Statistical Computation and Simulation* 49.1-2 (1994), pp. 103–109. DOI: 10.1080/00949659408811563. eprint: https://doi.org/10.1080/00949659408811563. URL: https://doi.org/10.1080/00949659408811563.

[56] Zhengyan Lin and Weidong Liu. "On Maxima of Periodograms of Stationary Processes". In: *The Annals of Statistics* 37.5B (2009), pp. 2676–2695. ISSN: 00905364, 21688966. URL: http://www.jstor.org/stable/30243724.

[57] Weidong Liu and Qi-Man Shao. "Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series". In: *The Annals of Statistics* 38.3 (2010), pp. 1913–1935. ISSN: 00905364, 21688966. URL: http://www.jstor.org/stable/20744474.

[58] Weidong Liu and Wei Biao Wu. "Asymptotics of spectral density estimates". In: *Econometric Theory* 26.4 (2010), pp. 1218–1245. ISSN: 02664666, 14694360. URL: http://www.jstor.org/stable/40800879.

[59] Philippe Loubaton. “On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices”. In: *Journal of theoretical probability* 29.4 (2016), pp. 1339–1443. DOI: 10.1007/s10959-015-0614-z. URL: https://doi.org/10.1007/s10959-015-0614-z.

[60] Philippe Loubaton and Xavier Mestre. *On the asymptotic behaviour of the eigenvalue distribution of block correlation matrices of high-dimensional time series*. 2021. arXiv: 2004.07226 [math.PR].

[61] Philippe Loubaton and Xavier Mestre. “Spectral convergence of large block-Hankel Gaussian random matrices”. In: *Advances in Complex Analysis and Operator Theory*. Springer, 2017, pp. 247–280.

[62] Philippe Loubaton and Alexis Rosuel. *Large random matrix approach for testing independence of a large number of Gaussian time series, v4*. 2021. arXiv: 2007.08806 [cs.IT].

[63] V A Marcenko and L A Pastur. “Distribution of eigenvalues for some sets of random matrices”. In: *Mathematics of the USSR-Sbornik* 1.4 (Apr. 1967), pp. 457–483. DOI: 10.1070/sm1967v001n04abeh001994. URL: https://doi.org/10.1070/sm1967v001n04abeh001994.

[64] Xavier Mestre and Pascal Vallet. “Correlation Tests and Linear Spectral Statistics of the Sample Correlation Matrix”. In: *IEEE Transactions on Information Theory* 63.7 (2017), pp. 4585–4618. DOI: 10.1109/TIT.2017.2689780.

[65] David Morales-Jimenez et al. “Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models”. In: *Statistica Sinica* 31.2 (2021), pp. 571–601.

[66] Seiji Nabeya and Katsuto Tanaka. “Approximate Distributions of the Periodogram and Related Statistics under Normality”. In: *Econometric Theory* 2.1 (1986), pp. 33–65. ISSN: 02664666, 14694360. URL: http://www.jstor.org/stable/3532213.

[67] Richard A. Olshen. “Asymptotic Properties of the Periodogram of a Discrete Stationary Process”. In: *Journal of Applied Probability* 4.3 (1967), pp. 508–528. ISSN: 00219002. URL: http://www.jstor.org/stable/3212218.

[68] Hernando Ombao and SÉbastien Van Bellegem. “Evolutionary Coherence of Nonstationary Signals”. In: *IEEE Transactions on Signal Processing* 56.6 (2008), pp. 2259–2266. DOI: 10.1109/TSP.2007.914341.

[69] Guangming Pan, Jiti Gao, and Yanrong Yang. “Testing Independence Among a Large Number of High-Dimensional Random Vectors”. In: *Journal of the American Statistical Association* 109.506 (2014), pp. 600–612. ISSN: 01621459. URL: http://www.jstor.org/stable/24247189.

[70] Leonid Pastur and Mariya Shcherbina. *Eigenvalue Distribution of Large Random Matrices*. Mathematical Surveys and Monographs, AMS, 2011.

[71] Debashis Paul and Jack W. Silverstein. “No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix”. In: *Journal of Multivariate Analysis* 100.1 (2009), pp. 37–57. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2008.03.010. URL: https://www.sciencedirect.com/science/article/pii/S0047259X08001024.

[72] David Ramirez et al. “Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas”. In: *IEEE Transactions on Signal Processing* 59.8 (2011), pp. 3764–3774. DOI: 10.1109/TSP.2011.2146779.

[73] David Ramírez et al. “Detection of Spatially Correlated Gaussian Time Series”. In: *IEEE Transactions on Signal Processing* 58.10 (2010), pp. 5006–5015. DOI: 10.1109/TSP.2010.2053360.

[74] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.

[75] Christian P. Robert. *The Bayesian choice*. second. Springer Texts in Statistics. From decision-theoretic foundations to computational implementation. Springer, New York, 2007, pp. xxvi+602. ISBN: 978-0-387-71598-8.

[76] A. Rosuel et al. "On the Detection of Low-Rank Signal in the Presence of Spatially Uncorrelated Noise: A Frequency Domain Approach". In: *IEEE Transactions on Signal Processing* 69 (2021), pp. 4458–4473. DOI: 10.1109/TSP.2021.3099644.

[77] Alexis Rosuel, Philippe Loubaton, and Pascal Vallet. *On the asymptotic distribution of the maximum sample spectral coherence of Gaussian time series in the high dimensional regime*. 2021. arXiv: 2107.02891 [math.ST].

[78] Alexis Rosuel et al. "On the frequency domain detection of high dimensional time series". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 8782–8786.

[79] Mark Rudelson and Roman Vershynin. "Hanson-Wright inequality and sub-gaussian concentration". In: *Electronic Communications in Probability* 18.none (2013), pp. 1–9. DOI: 10.1214/ECP.v18-2865. URL: https://doi.org/10.1214/ECP.v18-2865.

[80] Walter Rudin. *Real and Complex Analysis*. Higher Mathematics Series. McGraw-Hill Education, 1987. ISBN: 9780070542341.

[81] R Rudzkis. "On the distribution of the maximum deviation of the Gaussian stationary time series spectral density estimate". In: *Lithuanian Mathematical Journal* 25.1 (1985), pp. 18–130.

[82] Josep Sala-Alvarez et al. "Multiantenna GLR Detection of Rank-One Signals With Known Power Spectral Shape Under Spatially Uncorrelated Noise". In: *IEEE Transactions on Signal Processing* 64.23 (2016), pp. 6269–6283. DOI: 10.1109/TSP.2016.2601290.

[83] Xiaofeng Shao and Wei Biao Wu. "Asymptotic Spectral Theory for Nonlinear Time Series". In: *The Annals of Statistics* 35.4 (2007), pp. 1773–1801. ISSN: 00905364. URL: http://www.jstor.org/stable/25464559.

[84] J.W. Silverstein. "Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices". In: *Journal of Multivariate Analysis* 55.2 (1995), pp. 331–339. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.1995.1083. URL: https://www.sciencedirect.com/science/article/pii/S0047259X85710834.

[85] J.W. Silverstein and Z.D. Bai. "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices". In: *Journal of Multivariate Analysis* 54.2 (1995), pp. 175–192. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.1995.1051. URL: https://www.sciencedirect.com/science/article/pii/S0047259X85710512.

[86] J.W. Silverstein and S.I. Choi. "Analysis of the Limiting Spectral Distribution of Large Dimensional Random Matrices". In: *Journal of Multivariate Analysis* 54.2 (1995), pp. 295–309. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.1995.1058. URL: https://www.sciencedirect.com/science/article/pii/S0047259X85710585.

[87] Yiming Sun et al. *Large Spectral Density Matrix Estimation by Thresholding*. 2018. arXiv: 1812.00532 [stat.ME].

[88] Masanobu Taniguchi, Madan L. Puri, and Masao Kondo. "Nonparametric Approach for Non-Gaussian Vector Stationary Processes". In: *Journal of Multivariate Analysis* 56.2 (1996), pp. 259–283. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.1996.0014. URL: https://www.sciencedirect.com/science/article/pii/S0047259X96900147.

[89] Terence Tao. "Topics in random matrix theory". In: *Graduate Studies in Mathematics* 132 (2011).

[90] Pascal Vallet, Philippe Loubaton, and Xavier Mestre. "Improved Subspace Estimation for Multivariate Observations of High Dimension: The Deterministic Signals Case". In: *IEEE Transactions on Information Theory* 58.2 (2012), pp. 1043–1068. DOI: 10.1109/TIT.2011.2173718.

[91] Pascal Vallet, Xavier Mestre, and Philippe Loubaton. "Performance Analysis of an Improved MUSIC DoA Estimator". In: *IEEE Transactions on Signal Processing* 63.23 (2015), pp. 6407–6422. DOI: 10.1109/TSP.2015.2465302.

[92] Sebastian Wagner et al. "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback". In: *IEEE Transactions on Information Theory* 58.7 (2012), pp. 4509–4537. DOI: 10.1109/TIT.2012.2191700.

[93] Grace Wahba. "Some Tests of Independence for Stationary Multivariate Time Series". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 33.1 (1971), pp. 153–166. ISSN: 00359246. URL: http://www.jstor.org/stable/2986017.

[94] A. M. Walker. "Some asymptotic results for the periodogram of a stationary time series". In: *Journal of the Australian Mathematical Society* 5.1 (1965), pp. 107–128. DOI: 10.1017/S1446788700025921.

[95] M. Wax and T. Kailath. "Detection of signals by information theoretic criteria". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33.2 (1985), pp. 387–392. DOI: 10.1109/TASSP.1985.1164557.

[96] Michael B. Woodroofe and John W. Van Ness. "The Maximum Deviation of Sample Spectral Densities". In: *The Annals of Mathematical Statistics* 38.5 (1967), pp. 1558–1569. ISSN: 00034851. URL: http://www.jstor.org/stable/2238971.

[97] Wei Biao Wu and Paolo Zaffaroni. "ASYMPTOTIC THEORY FOR SPECTRAL DENSITY ESTIMATES OF GENERAL MULTIVARIATE TIME SERIES". In: *Econometric Theory* 34.1 (2018), pp. 1–22. DOI: 10.1017/S0266466617000068.

[98] Jianfeng Yao. "A note on a Marcenko–Pastur type theorem for time series". In: *Statistics & Probability Letters* 82.1 (2012), pp. 22–28. URL: https://EconPapers.repec.org/RePEc:eee:stapro:v:82:y:2012:i:1:p:22-28.