Large random matrix approach for testing independence of a large number of Gaussian time series

Philippe Loubaton* **and Alexis Rosuel***

Laboratoire d'Informatique Gaspard Monge, UMR 8049,
Université Paris-Est Marne la Vallée, France
e-mail: alexis.rosuel@univ-eiffel.fr; philippe.loubaton@univ-eiffel.fr

Abstract: The asymptotic behaviour of Linear Spectral Statistics (LSS) of the smoothed periodogram estimator of the spectral coherency matrix of a complex Gaussian high-dimensional time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ with independent components is studied under the asymptotic regime where both the dimension M of $\mathbf{y}$ and the smoothing span of the estimator grow to infinity at the same rate. It is established that the estimated spectral coherency matrix is close from the sample covariance matrix of an independent identically $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ distributed sequence, and that its empirical eigenvalue distribution converges towards the Marcenko-Pastur distribution. This allows to conclude that each LSS has a deterministic behaviour that can be evaluated explicitely. Using concentration inequalities, it is shown that the order of magnitude of the deviation of each LSS from its deterministic approximation is of the order of $\frac{M}{N}$ where N is the sample size. Numerical simulations suggest that these results can be used to test whether a large number of time series are uncorrelated or not.

MSC 2010 subject classifications: Primary 60B20, 62H15; secondary 62M15.
Keywords and phrases: Random Matrices, Spectral Analysis, High Dimensional Statistics, Time Series, Independence Test.

Contents

1 Introduction 2
 1.1 The addressed problem and the results 2
 1.2 Motivation 4
 1.3 On the literature 5
 1.4 General approach 7
 1.5 Assumptions and general notations 9
 1.6 Overview of the paper 11
2 Useful technical tools 11
 2.1 Stochastic domination 11
 2.2 Concentration of the smallest and largest eigenvalues of a Gaussian random matrix 13
 2.3 Concentration of functionals of Gaussian entries 14

*This work is funded by ANR Project HIDITSA, reference ANR-17-CE40-0003.

2.4 Hanson-Wright inequality 15
3 Stochastic representations of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$ 15
3.1 Step 1: Stochastic representation of $\tilde{\mathbf{C}}$ 16
3.2 Step 2: Estimates for $\hat{s}_m(\nu)$ 23
3.3 Step 3: Stochastic representation of $\hat{\mathbf{C}}$ 24
4 Control of Linear Spectral Statistics of $\hat{\mathbf{C}}$ 25
4.1 Step 1: proof of (4.2) 26
4.1.1 Reduction to the study of $\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \overline{\partial}\Phi_k(f)(z) \sum_{m=1}^{M} (\mathbf{Q}+z\mathbf{Q}^2)_{mm} \left(1 - \frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)$ 32
4.1.2 Proof of Lemma 4.1 and Lemma 4.3 39
4.2 Step 2: Lipschitz argument 47
4.3 Step 3: Extension to $\nu \in [0,1]$ 53
5 Applications and numerical simulation 55
5.1 Definition of the test statistics 55
5.2 Signal model for the alternative 56
A Appendix 61
A.1 Proof of Lemma A.1 61
A.2 Proof of Lemma 3.1 64
A.3 Proof of Lemma 3.3 67
A.4 Proof of Lemma A.5 67
References 68

1. Introduction

1.1. The addressed problem and the results

We consider a $M-$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

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

[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

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

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

where $\text{diag}(\hat{\mathbf{S}}(\nu)) = \hat{\mathbf{S}}(\nu) \odot \mathbf{I}_M$, 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 $c \in (0,1)$. It is established that if $\mu_{MP}^{(c)}$ represents the Marcenko-Pastur distribution with parameter $c < 1$ defined by

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

then, for each function f defined on $\mathbb{R}^+$ with enough continuous derivatives 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:

$$\mathbb{P}\left[\sup_{\nu\in[0,1]} \left| \frac{1}{M}\text{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f\, \mathrm{d}\mu_{MP}^{(c_N)} \right| > N^\epsilon \frac{B}{N}\right] \leq \exp -N^\gamma \tag{1.4}$$

In other words, under $\mathcal{H}_0$, uniformly w.r.t. the frequency ν, $\frac{1}{M}\text{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)$ behaves as $\int_{\mathbb{R}} f\, \mathrm{d}\mu_{MP}^{(c_N)}$, and with high probability, the order of magnitude of the corresponding error is not larger than $\frac{B}{N} = \mathcal{O}(\frac{1}{N^{1-\alpha}})$. 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 $(\boldsymbol{\xi}_{\mathbf{y}}(\nu + b/N))_{b=-B/2,\ldots,B/2}$. Classical time series analysis results suggest that the vectors $(\boldsymbol{\xi}_{\mathbf{y}}(\nu+b/N))_{b=-B/2,\ldots,B/2}$ appear as "nearly" i.i.d. zero mean complex random vectors with covariance matrix $\mathbf{S}(\nu)$ where $\mathbf{S}(\nu) = \text{Diag}\,(s_1(\nu), \ldots, s_M(\nu))$ and $(s_m)_{m=1,\ldots,M}$ represent the spectral densities of the scalar time series $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$. $\hat{\mathbf{C}}(\nu)$ 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. [21]), it is not surprising that $\frac{1}{M}\text{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right)$ 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. [26], [1]) and from frequency domain time series analysis techniques (see e.g. [4]).

1.2. Motivation

This paper 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. 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. In order 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 from $\mathbf{I}_M$ for each ν if both $\frac{B}{N}$ and $\frac{M}{B}$ are small enough. However, if M is large and that 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. In order to solve this issue, we propose to choose B of the same order of magnitude than M. In this case, $\hat{\mathbf{C}}(\nu)$ has of course no reason to be close from $\mathbf{I}_M$ 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 than M, then it seems reasonable to test that the components of $\mathbf{y}$ are uncorrelated by comparing

$$\sup_{\nu} \left| \frac{1}{M} \mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu) \right) - \int_{\mathbb{R}^+} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right|$$

to a well chosen threshold for some smooth function f. We also notice 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. This is because the quantity of interest of the corresponding papers is very often the empirical covariance matrix

$$\hat{\mathbf{R}}_N = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^* \tag{1.5}$$

However, when the multivariate time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is not supposed to be i.i.d., the study of $\hat{\mathbf{R}}_N$ is of course not sufficient in order to test that $(y_{1,n})_{n\in\mathbb{Z}},\dots,(y_{M,n})_{n\in\mathbb{Z}}$ are uncorrelated or not. The construction of test statistics that are functions of the spectral density and spectral coherency estimates (1.1,1.3) appears much more relevant in this context. However, it is clear that it would be hard to expect the derivation of positive results concerning the asymptotic behaviour of functionals of these estimates when $\frac{M}{N}$ does not converge towards 0. We finally remark that our asymptotic regime is similar to the regime studied in [3], devoted to the use of shrinkage methods to improve the performance of the spectral density estimate (1.1) when M is large. We thus refer the reader to [3] for other arguments motivating this regime.

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 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 when the sample size increases. We first review a few related examples of previous works. Two classes of methods were mainly studied. The first class uses lag domain approaches based on the observation that M jointly stationary time series $(y_{1,n})_{n\in\mathbb{Z}},\dots,(y_{M,n})_{n\in\mathbb{Z}}$ 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

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

is block diagonal. The lag domain approach was in particular used in [16] in conjunction with a prewithening of each time series. In this paper, $M=2$ and the two time series are supposed to be ARMA series. The coefficients of the ARMA models are estimated, as well as their corresponding innovation sequences $u_{1,n}$ and $u_{2,n}$ for $n=1,\dots,N$, and the cross correlation coefficients between u_1 and u_2 at lags $-(L-1),\dots,(L-1)$ where L is a fixed integer. If $\hat{r}_{1,2}(l)$ represents the estimated cross correlation coefficient at lag l defined by

$$\hat{r}_{1,2}(l) = \frac{\sum_{t=l+1}^{N} \hat{u}_{1,t}\hat{u}_{2,t-l}}{\left(\sum_{t=1}^{N} \hat{u}_{1,t}^2\right)^{1/2}\left(\sum_{t=1}^{N} \hat{u}_{2,t}^2\right)^{1/2}}$$

where $\hat{u}_{i,n}$ represents the estimator of $u_{i,n}$, then [16] introduced the statistics

$$Q_L = N\sum_{l=-(L-1)}^{L-1} \hat{r}_{1,2}^2(l)$$

and established that it is asymptotically chi-square distributed under $\mathcal{H}_0$. This idea has been extended and improved in the following decades, see e.g. [23],

[24], [18]. [19] generalized the approach of [16] to non ARMA time series. The main idea was to replace the estimated innovation processes by order p_N prediction errors where p_N converges towards ∞ at certain rate when $N \to +\infty$, and to replace the fixed integer L by a sequence $(L(N))_{N \geq 1}$ converging towards ∞ at a well chosen rate. The test statistics Q_L was also modified by introducing a kernel, a deterministic recentering term, as well as a well chosen renormalization. Duchesne and Roy [7] robustified Hong's statistics by applying the methodology of [24]. We finally mention the more direct approach of [11], valid when the two time series are multivariate, and that did not use any prewhitening of the various time series.

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. [30] studied hypothesis testing based on the nonparametric estimate $\hat{\mathbf{C}}(\nu)$ defined in equation (1.3) of $\mathbf{C}(\nu)$. Motivated by the generalized likelihood ratio test (GLRT) in the case of Gaussian i.i.d. time series and the Hadamard inequality, [30] remarked that for each ν, $\log\det(\mathbf{C}(\nu)) \leq 0$ and the equality holds if and only if $\mathbf{C}(\nu) = \mathbf{I}_M$. [30] thus considered the statistics

$$\sum_{i=1}^{P} \log\det\left(\hat{\mathbf{C}}(\nu_i)\right)$$

where $\nu_1, \ldots, \nu_P$ are fixed frequencies, and computed its asymptotic distribution under $\mathcal{H}_0$. When $M = 2$, [9] derived the asymptotic distribution of the statistics

$$\sum_{k=0}^{N-1} \left|\bar{\mathbf{C}}_{1,2}(k/N)\right|^2$$

under $\mathcal{H}_0$, where $\bar{\mathbf{C}}$ is a windowed frequency-smoothed periodogram estimate of the corresponding spectral coherency matrix. [28] studied more general class of tests of the form

$$\mathcal{H}_0 : \int_0^1 K(\mathbf{S}(\nu))\,\mathrm{d}\nu = \kappa$$

where K is a certain functional. [28] derived the asymptotic distribution of $\int_0^1 K(\tilde{\mathbf{S}}(\nu))\,\mathrm{d}\nu - \kappa$ where $\tilde{\mathbf{S}}(\nu)$ is a lag window estimator of $\mathbf{S}(\nu)$. These general results were used to test that 2 time series are uncorrelated. In the same vein, we also mention [10].

We finally review the very few existing works devoted to the case where the number M of time series converges towards $+\infty$. Apart [3] and [5] presented below, we are just aware of papers addressing the case 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 $d \in (0, 1)$. In particular, in contrast with the asymptotic regime considered in the present paper, 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 (1.5). In particular, when the observations are Gaussian random vectors, the generalized likelihood ratio test 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. [21] 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 $\frac{1}{M}\operatorname{Tr}\left(f(\hat{\mathbf{C}}_N)\right)$ converges towards $\int f d\mu_{MP}^{(d)}$ for each bounded continuous function f. In the Gaussian case, [22] 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, [6] 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 [25] 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 to study the asymptotic performance of the GLRT under certain class of alternatives. We also mention that [20] 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 Gumble distribution, which, of course, allows to test the hypothesis $\mathcal{H}_0$. This first contribution was extended later in a number of works, in particular in [5] who considered the case where the samples $\mathbf{y}_1, \ldots, \mathbf{y}_N$ have some specific correlation pattern. We finally cite [3] which proposed to use 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. While the topics addressed in [3] are different from the main purpose of this paper, [3] introduced the asymptotic regime that we consider here, except that $\frac{B^{3/2}}{N}$ is supposed to converge towards 0 in [3]. When $B = \mathcal{O}(N^\alpha)$, this condition is equivalent to $\alpha < 2/3$, while we rather study situations where $\alpha > 1/2$.

1.4. General approach

In order to establish (1.4), we use the following approach:

- We first study the behaviour of the modified sample spectral coherency matrix $\tilde{\mathbf{C}}(\nu)$ defined by

$$\tilde{\mathbf{C}}(\nu) = \operatorname{diag}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} \hat{\mathbf{S}}(\nu) \operatorname{diag}\left(\mathbf{S}(\nu)\right)^{-\frac{1}{2}} \tag{1.6}$$

  In other words, $\tilde{\mathbf{C}}(\nu)$ is obtained from $\hat{\mathbf{C}}(\nu)$ by replacing the estimated diagonal matrix $\operatorname{diag}\left(\hat{\mathbf{S}}(\nu)\right)$ by its true value $\operatorname{diag}\left(\mathbf{S}(\nu)\right)$. Using classical results of [4], we establish that for each ν, $\tilde{\mathbf{C}}(\nu)$ can be represented as

$$\tilde{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}(\nu) \tag{1.7}$$

where $\mathbf{X}(\nu)$ is a $M \times (B+1)$ 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 $N \in \mathbb{N}$:

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

We deduce from this that for each ϵ and each smooth enough function f,

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

for each $\nu \in [0,1]$. Classical concentration results of LSS of $\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}$ lead immediately to

$$\mathbb{P}\left[\left|\frac{1}{M}\mathrm{Tr}\left(f(\tilde{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)}\right| > N^\epsilon \frac{B}{N}\right] \leq \exp -N^\gamma \tag{1.8}$$

for each ν.

- We next establish that for each ν, $\hat{\mathbf{C}}(\nu)$ can be written as

$$\hat{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \mathbf{\Delta}(\nu) \tag{1.9}$$

where $\mathbf{\Delta}(\nu)$ verifies the concentration inequality

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

for each $\epsilon > 0$, where γ does depend on ϵ but not on ν. If $\alpha < 2/3$, the term $\frac{1}{\sqrt{B}}$ dominates $\frac{B}{N}$, and the estimation of the true spectral densities $(s_m)_{m=1,\dots,M}$ has an impact on the spectral norm of the error matrix $\mathbf{\Delta}(\nu)$. However, due to some subtle effects, it eventually turns out that for each ν,

$$\mathbb{P}\left[\left|\frac{1}{M}\mathrm{Tr}\left(f(\hat{\mathbf{C}}(\nu))\right) - \int_{\mathbb{R}^+} f \,\mathrm{d}\mu_{MP}^{(c_N)}\right| > N^\epsilon \frac{B}{N}\right] \leq \exp -N^\gamma \tag{1.10}$$

still holds provided the function f is smooth enough.

- We eventually show that with high probability, $\nu \to \hat{\mathbf{C}}(\nu)$ is Lipschitz with constant $\mathcal{O}(N^\beta)$ for some constant β. Using this property in conjunction with a discretization in the frequency domain, we eventually deduce (1.4) from (1.10).

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:*

$$\inf_{m\geq 1} \inf_{\nu\in[0,1]} |s_m(\nu)| > 0 \tag{1.11}$$

and

$$\sup_{m\geq 1} \sum_{u\in\mathbb{Z}} (1+|u|^2)|r_{m,u}| < +\infty \tag{1.12}$$

Assumption (1.12) of course implies that the spectral densities $(s_m)_{m\geq 1}$ are C^2 and that

$$\sup_{m\geq 1} \sup_{\nu\in[0,1]} |s_m^{(i)}(\nu)| < +\infty \tag{1.13}$$

for $i = 0, 1, 2$ ($s_m^{(i)}$ represents the derivative of order i of s_m).

Notations. A zero mean complex valued random vector $\mathbf{y}$ is said to be $\mathcal{N}_{\mathbb{C}}(0,\Sigma)$ distributed if $\mathbb{E}(\mathbf{y}\mathbf{y}^*) = \Sigma$ 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 $\mathbf{A}$ 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 $\operatorname{Re} \mathbf{A} = (\mathbf{A} + \mathbf{A}^*)/2$ and $\operatorname{Im} \mathbf{A} = (\mathbf{A} - \mathbf{A}^*)/2$ where $\mathbf{A}^*$ is the conjugate transpose of matrix $\mathbf{A}$.

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

We recall that $\mathbf{S}(\nu)$ represents the $M \times M$ diagonal matrix $\mathbf{S}(\nu) = \operatorname{Diag}(s_1(\nu), \ldots, s_M(\nu))$. 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 positive a constant that does not depend on the frequency ν, on the time series index m, as well as on the dimensions B, M and N. C will represent a generic notation for a nice constant, and the value of C 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 a $M \times M$ Hermitian matrix $\mathbf{A}$ with spectral decomposition $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$ where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_m, m = 1, \ldots, M)$ 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 $\operatorname{Im} z > 0$.

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

$$s_\mu(z) = \int \frac{\mathrm{d}\mu(\lambda)}{\lambda - z} \tag{1.14}$$

We recall that

$$|s_\mu(z)| \leq \frac{1}{\operatorname{Im} z} \tag{1.15}$$

for each $z \in \mathbb{C}^+$.

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

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

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

$$\mathbf{Q}(z) = (\mathbf{A} - z\mathbf{I}_M)^{-1} \tag{1.16}$$

We finally mention the following useful control for the norm $\mathbf{Q}$:

$$\|\mathbf{Q}\| \leq \frac{1}{\mathrm{Im}\, z} \tag{1.17}$$

1.6. Overview of the paper

We first recall in Section 2 useful technical tools: the concept of stochastic domination adapted from [12] which allows to considerably simplify the exposition of the following results, as well as known concentration inequalities expressed using the stochastic domination framework. We establish in Section 3 the stochastic representations (1.7) and (1.9) of $\tilde{\mathbf{C}}(\nu)$ and $\hat{\mathbf{C}}(\nu)$. In Section 4, we finally prove (1.10) and establish the Lipschitz properties that allow to deduce (1.4).

2. Useful technical tools

2.1. Stochastic domination

We now present the concept of stochastic domination introduced in [12]. A nice introduction to this tool can also be found in the lecture notes [2].

Definition 2.1. ***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

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^\epsilon Y^{(N)}(u)\right] \leq \exp -N^\gamma \tag{2.1}$$

for each large enough $N > N_0(\epsilon)$. If X is stochastically dominated by Y we use the notation $X^{(N)}(u) \prec Y^{(N)}(u)$. In order 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 2.1. *Suppose $(X_N)_{N \in \mathbb{N}}$ is a sequence of random variables, satisfying $X_N \prec N^{\epsilon}$ for any $\epsilon > 0$. It turns out that this precisely means that $X_N \prec 1$. 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*

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

which goes to zero exponentially since $X_N \prec N^{\epsilon}$ for the ϵ chosen. This argument will be used in the proof of Lemma 4.1.

Lemma 2.1. *Take four families of non negative random variables X_1, X_2, Y_1 and Y_2 defined as in Definition 2.1. Then the following holds:*

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

Proof. We consider $\epsilon > 0$ and introduce the quantities $\gamma_1 > 0$ and $\gamma_2 > 0$ such that

$$\mathbb{P}[X_1 > N^{\epsilon} Y_1] \leq \exp -N^{\gamma_1}, \quad \mathbb{P}[X_2 > N^{\epsilon} Y_2] \leq \exp -N^{\gamma_2}$$

for each N large enough. Then, there exists a $\gamma > 0$ such that for N large enough:

$$\begin{aligned}\mathbb{P}[X_1 + X_2 > N^{\epsilon}(Y_1 + Y_2)] &\leq \mathbb{P}[X_1 > N^{\epsilon} Y_1] + \mathbb{P}[X_2 > N^{\epsilon} Y_2] \\ &\leq \exp -N^{\gamma_1} + \exp -N^{\gamma_2} \\ &\leq \exp -N^{\gamma}\end{aligned}$$

This proves the first assertion of Lemma 2.1.

Moreover, there exists another quantities $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ and $\tilde{\gamma}$ such that for large enough N:

$$\begin{aligned}\mathbb{P}[X_1 X_2 > N^{\epsilon} Y_1 Y_2] &\leq \mathbb{P}[X_1 X_2 > N^{\epsilon} Y_1 Y_2, X_1 < N^{\epsilon/2} Y_1] + \mathbb{P}[X_1 > N^{\epsilon/2} Y_1] \\ &\leq \mathbb{P}[X_2 > N^{\epsilon/2} Y_2] + \mathbb{P}[X_1 > N^{\epsilon/2} Y_1] \\ &\leq \exp -N^{\tilde{\gamma}_1} + \exp -N^{\tilde{\gamma}_2} \\ &\leq \exp -N^{\tilde{\gamma}}\end{aligned}$$

which proves the second assertion. □

Remark 2.2. *Note that Definition 2.1 is slightly different from the original one [12] which states that the left hand side of* (2.1) *should be bounded by a quantity of order N^{-D} for any finite $D > 0$. In the present paper, all the random variables are Gaussian, and exponential concentration rates can be achieved.*

2.2. Concentration of the smallest and largest eigenvalues of a Gaussian random matrix

In this paper we will at multiple occasion use the concentration of the smallest and largest eigenvalues of empirical covariance matrix of iid $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ random vectors. [13] proved that if $M = M(N)$ and $B = B(N)$ follow the Assumption 1.3, for any $M \times (B+1)$ matrix $\mathbf{X}_N$ with iid $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, we have for any $\epsilon > 0$

$$\mathbb{P}\left[\lambda_M\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) < (1-\sqrt{c})^2 - \epsilon\right] \leq (B+1)\exp -C(B+1)\epsilon^2 \quad (2.2)$$

$$\mathbb{P}\left[\lambda_1\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) > (1+\sqrt{c})^2 + \epsilon\right] \leq (B+1)\exp -C(B+1)\epsilon^2 \quad (2.3)$$

for some universal constant C.

Consider for $\epsilon > 0$, the ϵ–expansion of the support of the Marchenko-Pastur distribution:

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

and the event:

$$\Lambda_{N,\epsilon} = \left\{\sigma\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) \subset \mathrm{Supp}\,\mu_{MP}^{(c)} + \epsilon\right\} \quad (2.4)$$

It is clear that using (2.2) and (2.3), $\Lambda_{N,\epsilon}$ 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.

Eventually, the following (weaker) statement is a simple consequence of the equations (2.2) and (2.3), which will sometimes be enough in the following:

$$\lambda_1\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right) + \frac{1}{\lambda_M\left(\frac{\mathbf{X}_N\mathbf{X}_N^*}{B+1}\right)} \prec 1 \quad (2.5)$$

Indeed, fix $\epsilon > 0$. There exist a $N_0(\epsilon)$ large enough such that for any $N \geq N_0$, $N^\epsilon > (1+\sqrt{c})^2 + \frac{N^\epsilon}{2}$. Therefore

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

which decay in the order $\exp -CN^{\gamma}$ for some $\gamma > 0$. The proof for the three other quantities are similar.

We finally notice that if we consider a family $\mathbf{X}_N(u) \in \mathbb{C}^{M\times(B+1)}$ with iid $\mathcal{N}_{\mathbb{C}}(0,1)$ entries, $u \in U^{(N)}$, where $U^{(N)}$ is a certain set possibly depending on N, then (2.2) and (2.3) hold for each $u \in U^{(N)}$ because the constant C in (2.2) and (2.3) is universal. This implies that the stochastic domination (2.5) is still verified by the family $\mathbf{X}_N(u)$, $u \in U^{(N)}$. Moreover, the family of events $\Lambda_{N,\epsilon}(u)$ defined by (2.4) when $\mathbf{X}_N$ is replaced by $\mathbf{X}_N(u)$ still holds with exponentially high probability.

2.3. Concentration of functionals of Gaussian entries

It is well known (see e.g. [29, 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:

$$\mathbb{P}\left[|f(\mathbf{X}) - \mathbb{E}f(\mathbf{X})| > t\right] \leq C \exp -Ct^2 \tag{2.6}$$

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

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

Within the stochastic domination framework, the concentration inequality (2.6) 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 $\mathbf{X}_N(u) \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$.

2.4. Hanson-Wright inequality

The Hanson-Wright inequality [27] is useful to control deviations of a quadratic form from its expectation. While it is proved in the real case in [27], 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

$$\mathbb{P}[|\mathbf{X}^*\mathbf{A}\mathbf{X} - \mathbb{E}\mathbf{X}^*\mathbf{A}\mathbf{X}| > t] \leq 2\exp -C\min\left(\frac{t^2}{\|\mathbf{A}\|_F^2}, \frac{t}{\|\mathbf{A}\|}\right) \tag{2.7}$$

We now write (2.7) 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:

$$\mathbb{P}\left[|\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u) - \mathbb{E}\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u)| > N^\epsilon\|\mathbf{A}_N(u)\|_F\right] \leq 2\exp -CN^\epsilon \tag{2.8}$$

We can therefore rewrite (2.8) as the following stochastic domination:

$$|\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u) - \mathbb{E}\mathbf{X}_N^*(u)\mathbf{A}_N(u)\mathbf{X}_N(u)| \prec \|\mathbf{A}_N(u)\|_F \tag{2.9}$$

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 iid Gaussian random vectors, and to control the order of magnitude of the corresponding errors. This is the objective of the following result.

Theorem 3.1. *Under Assumptions 1.1, 1.2, 1.3 and 1.4, for any $\nu \in [0,1]$, it exists a $M \times (B+1)$ 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:*

$$\tilde{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}_N(\nu), \quad \|\tilde{\mathbf{\Delta}}_N(\nu)\| \prec \frac{B}{N} \tag{3.1}$$

$$\hat{\mathbf{C}}_N(\nu) = \frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1} + \mathbf{\Delta}_N(\nu), \quad \|\mathbf{\Delta}_N(\nu)\| \prec \frac{1}{\sqrt{B}} + \frac{B}{N} \tag{3.2}$$

Remark 3.1. *Therefore, up to small additive perturbations, $\tilde{\mathbf{C}}_N(\nu)$ and $\hat{\mathbf{C}}_N(\nu)$ appear as empirical covariance matrices of iid $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ 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 iid $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ random vectors.*

Remark 3.2. *In the following, we will often omit to mention that the various matrices under consideration depend on N and ν. Matrices $\hat{\mathbf{C}}_N(\nu), \tilde{\mathbf{C}}_N(\nu), \mathbf{X}_N(\nu), \boldsymbol{\Delta}_N(\nu), \dots$ will therefore be denoted by $\hat{\mathbf{C}}(\nu), \tilde{\mathbf{C}}(\nu), \mathbf{X}(\nu), \boldsymbol{\Delta}(\nu), \dots$ or $\hat{\mathbf{C}}, \tilde{\mathbf{C}}, \mathbf{X}, \boldsymbol{\Delta}, \dots$*

The proof of Theorem 3.1 will proceed in three steps: first we provide the result for matrix $\tilde{\mathbf{C}}(\nu)$, then control the deviations between $\text{diag}(\mathbf{S}(\nu))^{-\frac{1}{2}}$ and $\text{diag}(\hat{\mathbf{S}}(\nu))^{-\frac{1}{2}}$, and eventually extend the stochastic representation of $\tilde{\mathbf{C}}(\nu)$ to $\hat{\mathbf{C}}(\nu)$.

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

In order to establish (3.1), we prove the following Proposition.

Proposition 3.1. *Under Assumptions 1.1, 1.2, 1.3 and 1.4, for any $\nu \in [0,1]$, there exist a $M \times (B+1)$ random matrix $\mathbf{X}_N(\nu)$ with $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. entries, and another matrix $\boldsymbol{\Gamma}_N(\nu)$ such that:*

$$\tilde{\mathbf{C}}_N(\nu) = \frac{(\mathbf{X}_N(\nu) + \boldsymbol{\Gamma}_N(\nu))(\mathbf{X}_N(\nu) + \boldsymbol{\Gamma}_N(\nu))^*}{B+1} \tag{3.3}$$

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

$$\frac{\|\boldsymbol{\Gamma}_N(\nu)\|^2}{B+1} \prec \frac{B^2}{N^2} \tag{3.4}$$

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

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\xi}_{\mathbf{y}}(\nu - \frac{B}{2N}), \dots, \boldsymbol{\xi}_{\mathbf{y}}(\nu + \frac{B}{2N}) \right) \tag{3.5}$$

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:

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

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

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

By Lemma A.1 in Appendix, we have for b and $b_1 \neq b_2$:

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

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

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

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

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 applications $\nu \mapsto s_m(\nu)$ specified in Assumption 1.4 implies that there exists quantities ϵ_m such that:

$$\begin{aligned}\operatorname{diag}&\left(s_m(\nu + \frac{b}{N}) : b = -B/2, \ldots, B/2\right) \\ &= s_m(\nu)\mathbf{I}_{B+1} + \operatorname{diag}\left(\epsilon_m(\nu + \frac{b}{N}) : b = -B/2, \ldots, B/2\right)\end{aligned}$$

where:

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

for some nice constant C. Therefore, if we define matrix $\boldsymbol{\Phi}_m$ as:

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

then the following relations hold:

$$\mathbb{E}[\boldsymbol{\omega}_m^* \boldsymbol{\omega}_m] = s_m\left(\mathbf{I}_{B+1} + \boldsymbol{\Phi}_m\right), \qquad \sup_{m \geq 1, b_1, b_2} |\boldsymbol{\Phi}_{m_{b_1, b_2}}| \leq \frac{C}{N} \tag{3.6}$$

The spectral norm of $\boldsymbol{\Phi}_m$ can be roughly bounded by the following inequality:

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

Using the Gaussianity of vector $\boldsymbol{\omega}_m$ and the expression (3.6), we obtain that $\boldsymbol{\omega}_m$ can be represented as

$$\boldsymbol{\omega}_m = \sqrt{s_m}\,\mathbf{x}_m\,\left(\mathbf{I} + \boldsymbol{\Phi}_m\right)^{1/2}, \quad \mathbf{x}_m \sim \mathcal{N}_{\mathbb{C}}(0, I_{B+1}) \tag{3.7}$$

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 $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$. Consider the eigenvalue / eigenvector decomposition of Hermitian matrix $\mathbf{\Phi}_m$, i.e.

$$\mathbf{\Phi}_m = \mathbf{U}_m \mathbf{\Lambda}_m \mathbf{U}_m^*$$

where $\mathbf{\Lambda}_m$ is the diagonal matrix of the eigenvalues of matrix $\mathbf{\Phi}_m$. It is clear that the entries of $\mathbf{\Lambda}_m$ are $\mathcal{O}(B/N)$ terms, and that matrix $(\mathbf{I}+\mathbf{\Lambda}_m)^{1/2}$ can be written as:

$$(\mathbf{I}+\mathbf{\Lambda}_m)^{1/2} = \mathbf{I} + \mathbf{\Psi}_m, \quad \mathbf{\Psi}_m = \mathrm{diag}(\psi_{m,b} : b = -B/2, \ldots, B/2) \tag{3.8}$$

where $\psi_{m,b}$ verifies

$$\sup_{m\geq 1} \sup_{b=-B/2,\ldots,B/2} |\psi_{m,b}| \leq C\frac{B}{N}$$

Therefore, it holds that:

$$\boldsymbol{\omega}_m = \sqrt{s_m}\mathbf{x}_m\left(\mathbf{I} + \mathbf{U}_m\mathbf{\Psi}_m\mathbf{U}_m^*\right) = \sqrt{s_m}\left(\mathbf{x}_m + \mathbf{x}_m\mathbf{U}_m\mathbf{\Psi}_m\mathbf{U}_m^*\right)$$

We denote by $\mathbf{X}$ and $\mathbf{\Gamma}$ the $M \times (B+1)$ matrices with rows $(\mathbf{x}_m)_{m=1,\ldots,M}$, and $(\mathbf{x}_m\mathbf{U}_m\mathbf{\Psi}_m\mathbf{U}_m^*)_{m=1,\ldots,M}$ respectively. Then, it holds that

$$\mathbf{\Sigma} = \mathrm{diag}\left(\sqrt{s_m}, m=1,\ldots,M\right)(\mathbf{X}+\mathbf{\Gamma}) \tag{3.9}$$

where we recall that $\mathbf{\Sigma}$ is defined by (3.5). We recall the definition of matrix $\tilde{\mathbf{C}}$ given by

$$\begin{aligned}
\tilde{\mathbf{C}} &= \mathrm{diag}(\sqrt{s_m}, m=1,\ldots,M)^{-1/2}\, \hat{\mathbf{S}}\, \mathrm{diag}(\sqrt{s_m}, m=1,\ldots,M)^{-1/2} \qquad (3.10)\\
&= \mathrm{diag}(\sqrt{s_m}, m=1,\ldots,M)^{-1/2}\frac{\mathbf{\Sigma}\mathbf{\Sigma}^*}{B+1}\mathrm{diag}(\sqrt{s_m}, m=1,\ldots,M)^{-1/2}
\end{aligned}$$

The representation (3.9) implies that $\tilde{\mathbf{C}}$ can also be written as

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

This completes the proof of (3.3). It remains to show (3.4). We denote by $\mathbf{Z}$ the $M \times M$ matrix $\mathbf{Z} = \frac{1}{B+1}\mathbf{\Gamma}\mathbf{\Gamma}^*$. As $\|\mathbf{Z}\|$ verifies

$$\|\mathbf{Z}\| \leq \|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| + \|\mathbb{E}\mathbf{Z}\|$$

it is enough to prove the two following facts:

$$\|\mathbb{E}\mathbf{Z}\| \leq C\,\frac{B^2}{N^2} \tag{3.11}$$

$$\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| \prec \frac{B^2}{N^2} \tag{3.12}$$

We start with (3.11). Using the decomposition of the rows of $\mathbf{\Gamma}$ we have:

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}_{i,j}] &= \frac{1}{B+1}\mathbb{E}[\mathbf{\Gamma}\mathbf{\Gamma}^*]_{i,j} \\
&= \frac{1}{B+1}\mathbb{E}[\mathbf{x}_i\mathbf{U}_i\mathbf{\Psi}_i\mathbf{U}_i^*\mathbf{U}_j\mathbf{\Psi}_j^*\mathbf{U}_j^*\mathbf{x}_j^*] \\
&= \frac{1}{B+1}\mathbb{E}[\mathrm{tr}\ \mathbf{x}_i\mathbf{U}_i\mathbf{\Psi}_i\mathbf{U}_i^*\mathbf{U}_j\mathbf{\Psi}_j^*\mathbf{U}_j^*\mathbf{x}_j^*] \\
&= \frac{1}{B+1}\mathrm{tr}\ (\mathbf{U}_j\mathbf{\Psi}_j^*\mathbf{U}_j^*\mathbb{E}[\mathbf{x}_j^*\mathbf{x}_i]\mathbf{U}_i\mathbf{\Psi}_i\mathbf{U}_i^*) \\
&= \delta_{ij}\frac{1}{B+1}\mathrm{tr}\ \mathbf{\Psi}_i\mathbf{\Psi}_j^*
\end{aligned}$$

so that it is clear that $\mathbb{E}[\mathbf{Z}]$ is the diagonal matrix with diagonal entries $(\frac{1}{B+1}\sum_{b=-B/2}^{B/2}|\psi_{m,b}|^2)_{m=1,\dots,M}$. By the estimation in equation (3.8), we easily have (3.11).

It remains to prove (3.12). 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 (2.9). For this, we need to express $\mathbf{g}^*\mathbf{Z}\mathbf{g}$ as a quadratic form of a certain complex Gaussian random vector with iid 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 (recall that $\mathbf{\Psi}_m(\nu)$ is real, so $\mathbf{\Psi}_m(\nu)\mathbf{\Psi}_m(\nu)^*=\mathbf{\Psi}_m(\nu)^2$)

$$\mathbf{G}(\nu)=\mathbb{E}[\mathbf{z}\mathbf{z}^*]=\frac{1}{B+1}\sum_{m=1}^{M}|\mathbf{g}_m|^2\mathbf{U}_m(\nu)\mathbf{\Psi}_m(\nu)^2\mathbf{U}_m(\nu)^*$$

Therefore, $\mathbf{z}$ can be written $\mathbf{z}=\mathbf{G}^{1/2}\mathbf{w}$ 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 (2.9) can now be applied:

$$|\mathbf{w}^*\mathbf{G}\mathbf{w}-\mathbb{E}\mathbf{w}^*\mathbf{G}\mathbf{w}|\prec\|\mathbf{G}\|_F \tag{3.13}$$

Since $\sum_{m=1}^{M}|\mathbf{g}_m|^2=1$, it is clear that $\|\mathbf{G}\|\leq\frac{1}{B+1}\sup_{m=1,\dots,M}\sup_{b=-B/2,\dots,B/2}\psi_{m,b}^2$. Therefore, (3.8) and the rough evaluation $\|\mathbf{G}\|_F^2\leq(B+1)\|\mathbf{G}\|^2$ leads to

$$\|\mathbf{G}\|\leq C\,\frac{1}{B+1}\left(\frac{B}{N}\right)^2,\quad \|\mathbf{G}\|_F^2\leq C\,\frac{1}{B+1}\left(\frac{B}{N}\right)^4 \tag{3.14}$$

The substitution of (3.14) in equation (3.13) gives the following control of $\mathbf{g}^*\mathbf{Z}\mathbf{g}-\mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}$:

$$|\mathbf{g}^*\mathbf{Z}\mathbf{g}-\mathbb{E}\mathbf{g}^*\mathbf{Z}\mathbf{g}|\prec\frac{1}{\sqrt{B}}\left(\frac{B}{N}\right)^2 \tag{3.15}$$

Consider $\epsilon > 0$, and an ϵ–net N_ϵ of $\mathbb{C}^M$, that is a set of $\mathbb{C}^M$ unit norm vectors $\{\mathbf{h}_k : k = 1, \dots, \mathcal{K}\}$ such that for each unit norm vector $\mathbf{u} \in \mathbb{C}^M$, it exists a vector $\mathbf{h} \in N_\epsilon$ for which $\|\mathbf{u} - \mathbf{h}\| \leq \epsilon$. It is well known that the cardinal of N_ϵ is bounded by $C\left(\frac{1}{\epsilon}\right)^{2M}$ where C 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 $t > 0$,

$$\{\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| > t\} \subset \cup_{h \in N_\epsilon}\{|\mathbf{h}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}| > Ct\}$$

for some nice constant C. Using the union bound, we obtain that

$$\mathbb{P}\left[\|\mathbf{Z} - \mathbb{E}\mathbf{Z}\| > t\right] \leq \sum_{h \in N_\epsilon} \mathbb{P}\left[|\mathbf{h}^*(\mathbf{Z} - \mathbb{E}\mathbf{Z})\mathbf{h}| > Ct\right] \tag{3.16}$$

Here, we would like to use equation (3.15) to conclude. By the definition of $\prec$, (3.15) is valid uniformly on any set of vector with cardinality polynomial in N. Here the set N_ϵ is strictly bigger: it's cardinal is $\mathcal{O}(\epsilon^{-2M})$ and therefore exponential in M. As a consequence, we have to accept to loose 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 precisely, write again (3.15) 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 number such that $t_N \geq B^2/N^2$. Using the estimates (3.14) 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 (2.7) provides:

$$\mathbb{P}\left[|\mathbf{g}^*\left[\mathbf{Z} - \mathbb{E}\mathbf{Z}\right)]\,\mathbf{g}| > t_N\right] \leq 2\,\exp\left\{-CB\frac{t_N}{(B/N)^2}\right\}$$

Eventually, the union bound on N_ϵ gives:

$$\sum_{h\in N_\epsilon} \mathbb{P}\left[|\mathbf{h}^*(\mathbf{Z}-\mathbb{E}\mathbf{Z})\mathbf{h}| > Ct_N\right] \leq 2\,\exp\left\{-CB\frac{t_N}{(B/N)^2} + 2CM\log 1/\epsilon\right\}$$

If we take $t_N = N^{\epsilon'}(B^2/N^2)$, then, it exists $\gamma > 0$ such that

$$\exp\left\{-CB\frac{t_N}{(B/N)^2} + 2CM\log 1/\epsilon\right\} \leq \exp -N^\gamma$$

holds for each N large enough. (3.16) thus implies (3.12). This completes the proof of (3.3). □

Corollary 3.1 is a rewriting of Proposition 3.1 in a more concise way. Define:

$$\tilde{\mathbf{\Delta}} = \frac{\mathbf{X}\mathbf{\Gamma}^* + \mathbf{\Gamma}\mathbf{X}^* + \mathbf{\Gamma}\mathbf{\Gamma}^*}{B+1} \tag{3.17}$$

Corollary 3.1. *For any $\nu \in [0,1]$, $\tilde{\mathbf{C}}(\nu)$ can be written as*

$$\tilde{\mathbf{C}}(\nu) = \frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1} + \tilde{\mathbf{\Delta}}(\nu) \tag{3.18}$$

where the family of random variable $\|\tilde{\mathbf{\Delta}}(\nu)\|$, $\nu \in [0,1]$ verifies

$$\|\tilde{\mathbf{\Delta}}\| \prec \frac{B}{N} \tag{3.19}$$

Proof. Let $\nu \in [0,1]$. By the definition (3.17) of $\tilde{\mathbf{\Delta}}$, we indeed have:

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

By equation (3.4) from Theorem 3.1 and equation (2.5) from Paragraph 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 (3.17):

$$\|\tilde{\mathbf{\Delta}}\| \prec \frac{B}{N} + \frac{B}{N} + \frac{B^2}{N^2}$$

□

We now take benefit of Corollary 3.1 to analyse the location of the eigenvalues of matrices $\tilde{\mathbf{C}}$ and $\hat{\mathbf{S}}$. In order to formulate the corresponding result, we define some notations. We introduce the event $\Lambda_\epsilon^{\tilde{\mathbf{C}}}(\nu)$ defined by

$$\Lambda_\epsilon^{\tilde{\mathbf{C}}}(\nu) = \{\sigma(\tilde{\mathbf{C}}(\nu)) \subset \operatorname{Supp}\mu_{MP}^{(c)} + \epsilon\} \tag{3.20}$$

We remark that $\Lambda_\epsilon^{\tilde{\mathbf{C}}}(\nu)$ also depends on N, but again omit to mention this in order to simplify the notations. We also denote by $\mathbf{D}$ and $\hat{\mathbf{D}}$ the matrices $\mathbf{D} = \mathbf{D}(\nu) := \mathrm{diag}(\mathbf{S}(\nu))^{\frac{1}{2}}$ and $\hat{\mathbf{D}} = \hat{\mathbf{D}}(\nu) := \mathrm{diag}(\hat{\mathbf{S}}(\nu))^{\frac{1}{2}}$. 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:

$$\Lambda_\epsilon^{\hat{\mathbf{S}}}(\nu) = \left\{ \sigma(\hat{\mathbf{S}}(\nu)) \subset \mathrm{Supp}\, \mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon \right\} \tag{3.21}$$

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 3.2. *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 (3.18) 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 into

$$\left\{ \lambda_1(\frac{\mathbf{X}\mathbf{X}^*}{B+1}) > (1+\sqrt{c})^2 + \epsilon/2 \right\} \cup \left\{ \|\tilde{\mathbf{\Delta}}\| > \epsilon/2 \right\}$$

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 (2.3) and (3.19) 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 (1.6) 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 $\hat{\mathbf{S}}$ verify

$$\underline{s}\, \lambda_M(\tilde{\mathbf{C}}) \leq \lambda_k(\hat{\mathbf{S}}) \leq \bar{s}\, \lambda_1(\tilde{\mathbf{C}})$$

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 3.3. *Corollary 3.2 implies the following weaker property, which will be useful:*

$$\|\hat{\mathbf{S}}(\nu)\| \prec 1 \tag{3.22}$$

Before ending the section and proving Theorem 3.1, we need some stochastic control on the diagonal elements of $\hat{\mathbf{S}}$ in order to evaluate $\boldsymbol{\Theta}$ defined by

$$\boldsymbol{\Theta} := \hat{\mathbf{C}} - \tilde{\mathbf{C}} \tag{3.23}$$

Using the definition of $\hat{\mathbf{C}}$ from (1.3) and $\tilde{\mathbf{C}}$ from (1.6), $\boldsymbol{\Theta}$ can be written as

$$\boldsymbol{\Theta} = (\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2})\hat{\mathbf{S}}\hat{\mathbf{D}}^{-1/2} + \mathbf{D}^{-1/2}\hat{\mathbf{S}}(\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}) \tag{3.24}$$

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 $\|\boldsymbol{\Theta}\|$. Define

$$\hat{s}_m(\nu) := \hat{\mathbf{S}}_{m,m}(\nu)$$

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 3.2 is to establish stochastic domination results for $\hat{s}_m$, $\|\hat{\mathbf{D}}^{-1/2}\|$ and $\|\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}\|$.

3.2. Step 2: Estimates for $\hat{s}_m(\nu)$

For all this section, we write $s_m(\nu) := s_m$, $\mathbf{D}(\nu) := \mathbf{D}$, etc in order to simplify the notations. Define as in (3.21) the following quantity

$$\Lambda_{\epsilon}^{\hat{\mathbf{D}}}(\nu) = \{\sigma(\hat{\mathbf{D}}(\nu)) \subset [\underline{s}, \bar{s}] + \epsilon\} \tag{3.25}$$

Lemma 3.1. *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 A.2. □

Roughly speaking, this ensure 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 3.2. *The family of random variables $(|\hat{s}_m(\nu)| + \frac{1}{|\hat{s}_m(\nu)|})_{m=1,\ldots,M}$, $\nu \in [0,1]$, verifies*

$$\left(|\hat{s}_m| + \frac{1}{|\hat{s}_m|}\right) \prec 1$$

Proof. Immediate from Lemma 3.1. □

Lemma 3.3. *The set of random variable* $(|\hat{s}_m(\nu)^{-1/2} - s_m(\nu)^{-1/2}|)_{m=1,\dots,M}$ *and* $(|\sqrt{\frac{s_m(\nu)}{\hat{s}_m(\nu)}} - 1|)_{m=1,\dots,M}$, $\nu \in [0,1]$, *verifies*

$$|\hat{s}_m^{-1/2} - s_m^{-1/2}| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2}, \quad \left|\sqrt{\frac{s_m}{\hat{s}_m}} - 1\right| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2} \tag{3.26}$$

Proof. See Appendix A.3 □

3.3. Step 3: Stochastic representation of $\hat{\mathbf{C}}$

We are now in position to prove the result concerning $\hat{\mathbf{C}}$ of Theorem 3.1.

Proof. We have to control the operator norm of:

$$\boldsymbol{\Delta} = \hat{\mathbf{C}} - \frac{\mathbf{X}\mathbf{X}^*}{B+1} = \hat{\mathbf{C}} - \tilde{\mathbf{C}} + \tilde{\mathbf{C}} - \frac{\mathbf{X}\mathbf{X}^*}{B+1} = \boldsymbol{\Theta} + \tilde{\boldsymbol{\Delta}} \tag{3.27}$$

The operator norm of $\|\tilde{\boldsymbol{\Delta}}\|$ has already been proved in Corollary 3.1 to satisfy $\|\tilde{\boldsymbol{\Delta}}\| \prec (B/N)$. Moreover, recall that $\boldsymbol{\Theta}$ can be written as a function of $\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}$ in (3.24), so that one can use Lemma 3.2 and Lemma 3.3 to dominate each term and get:

$$\|\boldsymbol{\Theta}\| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2} \tag{3.28}$$

Summing the estimate of $\boldsymbol{\Theta}$ and the one of $\tilde{\boldsymbol{\Delta}}$, one get:

$$\|\boldsymbol{\Delta}\| \prec \frac{1}{\sqrt{B}} + \frac{B^2}{N^2} + \frac{B}{N}$$

which is the desired result. □

As a consequence, we state here Corollary 3.3 about the localization of the eigenvalues of $\hat{\mathbf{C}}(\nu)$.

Corollary 3.3. *For each* $\epsilon > 0$, *we define* $\Lambda_\epsilon^{\hat{\mathbf{C}}}(\nu)$ *as the event*

$$\Lambda_\epsilon^{\hat{\mathbf{C}}}(\nu) = \left\{\sigma(\hat{\mathbf{C}}(\nu)) \subset \operatorname{Supp} \mu_{MP}^{(c)} + \epsilon\right\} \tag{3.29}$$

Then, the family of events $\Lambda_\epsilon^{\hat{\mathbf{C}}}(\nu), \nu \in [0,1]$ *holds with exponentially high probability.*

Proof. We simply write:

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1} - \|\boldsymbol{\Delta}\|\mathbf{I}_M \le \hat{\mathbf{C}} \le \frac{\mathbf{X}\mathbf{X}^*}{B+1} + \|\boldsymbol{\Delta}\|\mathbf{I}_M$$

and use the same arguments as in the proof of Corollary 3.2. □

4. Control of Linear Spectral Statistics of $\hat{\mathbf{C}}$

In the following, we consider LSS for function f satisfying the following assumptions.

Assumption 4.1. *f is defined on $\mathbb{R}_+$ and there exist some $\epsilon > 0$ such that its restriction on* $\mathrm{Supp}_{MP}^{(c)} + \epsilon$ *is C^p where p is some integer $p \geq 8$.*

We now state the main result, which controls with exponentially high probability the maximum deviation rate of any LSS of $\hat{\mathbf{C}}(\nu)$.

Theorem 4.1. *Let f be an application satisfying the conditions of Assumption 4.1. Then, under Assumptions 1.1, 1.2, 1.3 and 1.4, we have:*

$$\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right| \prec \frac{B}{N} \tag{4.1}$$

Remark 4.1. *This theorem ensures that under $\mathcal{H}_0$, $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ and $\int f \, \mathrm{d}\mu_{MP}^{(c_N)}$ are $\mathcal{O}(1)$ terms, but their dominant behaviour are the same so that the left hand side of (4.1) becomes of reduced order $\mathcal{O}_P(\frac{B}{N})$. We expect that under some alternative hypothesis, the order of magnitude of the l.h.s. of (4.1) will be larger, typically of order 1.*

Remark 4.2. *We also notice that (4.1) is stronger than the property that the family of random variables $\left| \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right|, \nu \in [0,1]$ verifies*

$$\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right| \prec \frac{B}{N} \tag{4.2}$$

Indeed, in (4.2), the stochastic domination holds for each ν, but not uniformly on $\nu \in [0,1]$. More precisely, (4.2) is equivalent to for any ϵ, there exist $N_0(\epsilon)$ and $\alpha > 0$ such that for any $N \geq N_0(\epsilon)$

$$\sup_{\nu \in [0,1]} \mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right| \geq N^{\epsilon} \frac{B}{N} \right] \leq e^{-CN^{\alpha}} \tag{4.3}$$

whereas (4.1) is equivalent to

$$\mathbb{P}\left[\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}_N(\nu)) - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{MP}^{(c_N)} \right| \geq N^{\epsilon} \frac{B}{N} \right] \leq e^{-CN^{\alpha}} \tag{4.4}$$

which are not equivalent since the set $[0,1]$ is not finite, so the union bound cannot be applied directly.

Outline of the proof. We will first prove (4.2) using Gaussian concentration inequalities. The extension to the supremum over any finite grid will then be immediate since the concentration results rates are exponential. Eventually, we will extend the supremum to the whole interval $[0,1]$ by a Lipschitz argument on the application $\nu \mapsto \hat{\mathbf{C}}(\nu)$.

4.1. Step 1: proof of (4.2)

The quantity of interest in equation (4.1) will be split into the three following terms:

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))\right| + \left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}\right)\right| + \left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|$$

As we shall see, the evaluations of the second and third terms are rather easy, but the first term is more demanding.

Proposition 4.1. *The family of random variables* $\left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1}\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|, \nu \in [0,1]$ *verifies*

$$\left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1}\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| \prec \frac{1}{M} << \frac{B}{N} \tag{4.5}$$

where we recall that $\frac{1}{M} << \frac{B}{N}$ *because* $\alpha > \frac{1}{2}$.

Proof. As usual, we omit the dependence with respect to ν. Since f is well defined on $\mathbb{R}_+$, the quantity tr $f(\mathbf{X}\mathbf{X}^*/(B+1))$ is also well defined. First, we need to localize the eigenvalues of $\frac{\mathbf{X}\mathbf{X}^*}{B+1}$ in order to work on an appropriate compact. Let $\kappa > 0$ and let $\chi : \mathbb{R} \to \mathbb{R}$ be a C^∞ application such that:

$$\chi(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \mathrm{Supp}\,\mu_{MP}^{(c)} + \kappa \\ 0 & \text{if } \lambda \notin \mathrm{Supp}\,\mu_{MP}^{(c)} + 2\kappa \end{cases} \tag{4.6}$$

and $\bar{f} : \mathbb{R} \to \mathbb{R}$ defined by $\bar{f} = f \times \chi$. Note that $\bar{f}$ is compactly supported. We write:

$$\begin{aligned}\left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| &\leq \left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) - \frac{1}{M}\mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right| \\ &+ \left|\frac{1}{M}\mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) - \mathbb{E}\frac{1}{M}\mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right| \\ &+ \left|\frac{1}{M}\mathbb{E}\mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) - \int_{\mathbb{R}} \bar{f}\,\mathrm{d}\mu_{MP}^{(c_N)}\right| + \left|\int_{\mathbb{R}} \bar{f}\,\mathrm{d}\mu_{MP}^{(c_N)} - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right|\end{aligned} \tag{4.7}$$

For M, B large enough, $(1-\sqrt{c_N})^2 \geq (1-\sqrt{c})^2 - \kappa$ and $(1+\sqrt{c_N})^2 \leq (1+\sqrt{c})^2 + \kappa$, i.e.

$$[(1-\sqrt{c_N})^2, (1+\sqrt{c_N})^2] \subset \mathrm{Supp}\,\mu_{MP}^{(c)} + \kappa$$

interval on which χ is constant equal to 1. Therefore, for N large enough, f and $\bar{f}$ coincide on $[(1-\sqrt{c_N})^2, (1+\sqrt{c_N})^2]$. Therefore,

$$\int_{\mathbb{R}} \bar{f}\,\mathrm{d}\mu_{MP}^{(c_N)} = \int_{(1-\sqrt{c_N})^2}^{(1+\sqrt{c_N})^2} f(\lambda)\chi(\lambda)\,\mathrm{d}\mu_{MP}^{(c_N)} = \int_{(1-\sqrt{c_N})^2}^{(1+\sqrt{c_N})^2} f(\lambda)\,\mathrm{d}\mu_{MP}^{(c_N)} = \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$$

so the last term in the right hand side of (4.7) vanishes.

In order to evaluate the third term of the right hand side of (4.7), we notice that since $\mathrm{Supp}(\bar{f})$ is compact and $\bar{f}$ is C^p for $p \geq 8$, Theorem 6.2 of [14] proves[2] that the following results holds:

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

Recall the definition of Λ_κ from (2.4) which holds with exponentially high probability. The first term in the right hand side of (4.7) can be handled as follows: on the event Λ_κ, $\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right)$ and $\mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right)$ coincides. Therefore, we have for any $q \geq 0$

$$\begin{aligned}
&\mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) - \frac{1}{M} \mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right| > N^{-q} \right] \\
&\qquad \leq \mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) - \frac{1}{M} \mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right| > N^{-q}, \Lambda_\kappa \right] + \mathbb{P}\left[\Lambda_\kappa^c \right] \\
&\qquad = \mathbb{P}\left[\Lambda_\kappa^c \right]
\end{aligned}$$

which decays to zero exponentially fast in N. Therefore, we showed that:

$$\left| \frac{1}{M} \mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) - \frac{1}{M} \mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right| \prec \frac{1}{N^q}$$

for any $q \in \mathbb{N}$. We obtain in particular that

$$\left| \frac{1}{M} \mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) - \frac{1}{M} \mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right| \prec \frac{1}{M}$$

It remains to study the second term in the right hand side of (4.7). We will use the concentration result for Lipschitz transformation of Gaussian entries from Paragraph 2.3. We consider the real valued function ψ defined by $\psi(\mathbf{X}, \mathbf{X}^*) = \mathrm{tr}\ \bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right)$, and establish that it is $\mathcal{O}(1)$–Lipschitz. For this, we evaluate

$$\|\nabla\psi(\mathbf{X}, \mathbf{X}^*)\|^2 = \sum_{i,j} \left| \frac{\partial\psi}{\partial X_{i,j}} \right|^2 + \left| \frac{\partial\psi}{\partial \overline{X_{i,j}}} \right|^2 \tag{4.8}$$

Using classic identities for derivation of Hermitian matrices, the first term in the sum is:

$$\frac{\partial \mathrm{tr}\ \psi}{\partial X_{ij}}(\mathbf{X}, \mathbf{X}^*) = \frac{1}{B+1}\left[\mathbf{X}^* \bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right]_{ji} = \frac{1}{\sqrt{B+1}}\left[\frac{\mathbf{X}^*}{\sqrt{B+1}} \bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \right]_{ji}$$

and similarly:

$$\frac{\partial \mathrm{tr}\ \psi}{\partial \overline{X_{ij}}}(\mathbf{X}, \mathbf{X}^*) = \frac{1}{B+1}\left[\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \mathbf{X} \right]_{ij} = \frac{1}{\sqrt{B+1}}\left[\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} \right) \frac{\mathbf{X}}{\sqrt{B+1}} \right]_{ij}$$

[2]while Theorem 6.2 of [14] is stated with $\bar{f} \in C^\infty$, the proof actually only needs $f \in C^8$

Replacing these expressions in (4.8) gives:

$$\begin{aligned}\|\nabla\psi\|^2(\mathbf{X},\mathbf{X}^*) &= \frac{1}{B+1}\sum_{i,j}\left|\left[\frac{\mathbf{X}^*}{\sqrt{B+1}}\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right]_{ji}\right|^2 \\ &\qquad + \frac{1}{B+1}\sum_{i,j}\left|\left[\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\frac{\mathbf{X}}{\sqrt{B+1}}\right]_{ij}\right|^2 \\ &= 2\frac{1}{B+1}\mathrm{tr}\ \frac{\mathbf{X}^*}{\sqrt{B+1}}\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\bar{f}'\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)^*\frac{\mathbf{X}}{\sqrt{B+1}} \\ &= \frac{2}{B+1}\mathrm{tr}\ |\bar{f}'|^2\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\frac{\mathbf{X}\mathbf{X}^*}{B+1} \\ &= \frac{2M}{B+1}\,\frac{1}{M}\,\mathrm{tr}\ \left(|\bar{f}'|^2\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\end{aligned}$$

As $\bar{f} \in C^p$ for $p \geq 8$ and is compactly supported, the function $\lambda \to \lambda\,|\bar{f}'(\lambda)|^2$ is bounded by some constant, and there exists a nice constant C such that

$$\frac{1}{M}\,\mathrm{tr}\ \left(|\bar{f}'|^2\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) \leq C$$

for each $\mathbf{X}$. Moreover, since the ratio $\frac{M}{B+1} = c_N$ converges towards the finite constant c, we obtain that

$$\|\nabla\psi\|^2 \leq C = \mathcal{O}(1)$$

for some nice constant C. This proves that ψ is $\mathcal{O}(1)$–Lipschitz and Paragraph 2.3 provides:

$$|\psi(\mathbf{X},\mathbf{X}^*) - \mathbb{E}\psi(\mathbf{X},\mathbf{X}^*)| \prec 1$$

Therefore, we have shown that

$$\left|\mathrm{tr}\ \left(\bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right) - \mathbb{E}\,\mathrm{tr}\ \left(\bar{f}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right)\right| \prec 1 \tag{4.9}$$

This completes the proof of $\left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1}\right) - \int_{\mathbb{R}} f\, d\mu_{MP}^{(c_N)}\right| \prec \frac{1}{M}$.

∎

Proposition 4.2. *The family of random variables* $\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}_N(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}_N(\nu)\mathbf{X}_N^*(\nu)}{B+1}\right)\right|, \nu \in [0,1]$ *verifies*

$$\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}\right)\right| \prec \frac{B}{N} \tag{4.10}$$

Proof. Write, for $\epsilon > 0$,

$$\mathbb{P}\left[\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}\right)\right| > N^\epsilon \frac{B}{N}\right]$$
$$\leq \mathbb{P}\left[\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}(\nu)\mathbf{X}^*(\nu)}{B+1}\right)\right| > N^\epsilon \frac{B}{N}, \Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}}\right]$$
$$+ \mathbb{P}\left[(\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}})^c\right]$$

We recall that Λ_ϵ and $\Lambda_\epsilon^{\tilde{\mathbf{C}}}$ are the events defined by (2.4) and (3.20) respectively. It has been proved in Paragraph 2.2 that Λ_ϵ holds with exponentially high probability, and in Corollary 3.2 that $\Lambda_\epsilon^{\tilde{\mathbf{C}}}$ also holds with exponentially high probability. Therefore, it remains to study:

$$\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right| \mathbb{1}(\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}})$$

On the event $\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}}$, the eigenvalues of $\frac{\mathbf{X}\mathbf{X}^*}{B+1}$ and $\tilde{\mathbf{C}}$ are localized in a compact support. Using the representation (3.18) of $\tilde{\mathbf{C}}$, we get:

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1} - \|\tilde{\mathbf{\Delta}}\|\mathbf{I}_M \leq \tilde{\mathbf{C}} \leq \frac{\mathbf{X}\mathbf{X}^*}{B+1} + \|\tilde{\mathbf{\Delta}}\|\mathbf{I}_M$$

As in the context of the proof of Proposition 4.1, we replace f by the C^p compactly supported function $\bar{f} = f\,\chi$ where χ is defined by (4.6). The derivative of $\bar{f}$ is bounded by some nice constant C. Therefore, for any $m \in \{1, \ldots, M\}$ we have:

$$\left|f(\lambda_m(\tilde{\mathbf{C}})) - f(\lambda_m\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right))\right| \mathbb{1}(\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}})$$
$$= \left|\bar{f}(\lambda_m(\tilde{\mathbf{C}})) - \bar{f}(\lambda_m\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right))\right| \mathbb{1}(\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}})$$
$$\leq C\left|\lambda_m(\tilde{\mathbf{C}}) - \lambda_m\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right|$$
$$\leq C\,\|\tilde{\mathbf{\Delta}}\|$$

It remains to use the concentration result on $\|\tilde{\mathbf{\Delta}}\|$ from Theorem 3.1 and get the following bound:

$$\left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right| \mathbb{1}(\Lambda_\epsilon \cap \Lambda_\epsilon^{\tilde{\mathbf{C}}}) \leq C\|\tilde{\mathbf{\Delta}}\| \prec \frac{B}{N}$$

This completes the proof of Proposition 4.2. □

Now, we focus on proving the same kind of stochastic domination for the following quantity:

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))\right|$$

Proposition 4.3. *Under Assumption 4.1, the family of random variables* $\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})(\nu)\right|$, $\nu \in [0,1]$ *verifies*

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})(\nu)\right| \prec \frac{B}{N} \tag{4.11}$$

Remark 4.3. *As we shall see below, the rate* $\frac{B}{N}$ *in* (4.11) *is pessimistic, but easier to establish than a tighter factor. However, we note that the factor* B/N *in the statement of Theorem 4.1 cannot be improved since it appears in* (4.10)*. Therefore, it is not useful to improve the evaluation* (4.11)*.*

Remark 4.4. (4.11) *seems at first surprising . Using the stochastic domination result* (3.28) *for* $\mathbf{\Theta}$ *, we can directly obtain:*

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})(\nu)\right| \prec \frac{B}{N} + \frac{1}{\sqrt{B}}$$

where $\frac{1}{\sqrt{B}}$ *is the limiting speed for* $\alpha < 2/3$. $\frac{1}{\sqrt{B}}$ *comes from the error of estimation of the spectral densities of the* M *scalar times series. Due to subtle effects, it turns out that this error term eventually doesn't contribute in the LSS* (4.11)*. This can be easily understood if* $f(\lambda) = \log\lambda$*. In this case, the left hand side of (4.11) is reduced to*

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})(\nu) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})(\nu) = \frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right)$$

which depends only on the estimators $(\hat{s}_m(\nu))_{m=1,\dots,M}$*. It is then easy to show that (4.11) holds. We just provide a sketch of proof. For this, we first remark that 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* (3.25)*. For each* m*, we expand around* s_m *the logarithm up to the second order, and obtain that*

$$\frac{1}{M}\sum_{m=1}^{M}\left(\log s_m(\nu) - \log \hat{s}_m(\nu)\right) = -\frac{1}{M}\sum_{m=1}^{M}(\hat{s}_m - s_m)\frac{1}{s_m} + \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2}\left(\frac{\hat{s}_m - s_m}{\theta_m}\right)^2 \tag{4.12}$$

where for each m*,* θ_m *is located between* s_m *and* $\hat{s}_m$*. Lemma A.5 allows to conclude that the second term of the right hand side of (4.12) is dominated by* $\frac{1}{B} + \left(\frac{B}{N}\right)^4$*. In order to evaluate the first term of the r.h.s. of (4.12), we first use Lemma A.2 to obtain that*

$$\frac{1}{M}\sum_{m=1}^{M}(\mathbb{E}(\hat{s}_m - s_m))\frac{1}{s_m} = \mathcal{O}\left(\frac{B}{N}\right)^2$$

and finally remark that, by Eq. (A.13),

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{s}_m - \mathbb{E}(\hat{s}_m)}{s_m}$$

can be interpreted as a recentered quadratic form of the MN dimensional vector $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_M^T)^T$. 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 $\frac{1}{B} << \frac{B}{N}$ because $\alpha > \frac{1}{2}$, we eventually obtain (4.11) if $f(\lambda) = \log \lambda$.

Proof. First, as in Proposition 4.1 and Proposition 4.2, it is possible to replace $f \in C^p(\mathbb{R})$ with the function $\bar{f} \in C^p(\mathbb{R})$ supported by $\mathrm{Supp}\, \mu_{MP}^c + 2\kappa$, for some $\kappa > 0$, defined by $\bar{f} = f \times \chi$ where χ is defined by (4.6). In order to simplify the notations, we drop in the following the notation $\bar{f}$ and simply associate f with its compactly supported version.

In order to prove Proposition 4.3, we use the so-called Helffer-Sjöstrand formula to express $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))$ as a quantity depending on the resolvent of $\hat{\mathbf{C}}$ and the resolvent of $\tilde{\mathbf{C}}$. In order to introduce this tool, we remark that the compactly supported function f is of class C^{k+1} for a certain integer k verifying $k+1 \geq 8$, 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

$$\bar{\partial}\Phi_k(f)(x+iy) = \frac{(iy)^k}{k!} f^{(k+1)}(x) \tag{4.13}$$

(a proof of this result can be found in [8] or [17]) if y belongs to the neighbourhood of 0 in which ρ is equal to 1. If μ is a probability measure, with $s_\mu(z)$ representing its Stieltjes transform (which definition is recalled in (1.14)), the Helffer-Sjöstrand formula can be written as

$$\int f \, \mathrm{d}\mu = \frac{1}{\pi} \mathrm{Re} \int_{\mathbb{C}^+} \bar{\partial}\Phi_k(f)(z) s_\mu(z) \, \mathrm{d}x \, \mathrm{d}y \tag{4.14}$$

In order to understand why the integral at the right hand side of (4.14) is well defined, we take, to fix the ideas, $\rho \in C^\infty$ such that $\rho(y) = 1$ for $|y| \leq 1$ and $\rho(y) = 0$ for $|y| > 2$ throughout this paragraph. Using this with the fact that f is compactly supported on the interval $[a_1, a_2]$, with $a_1 = (1-\sqrt{c})^2 - 2\kappa$ and $a_1 = (1+\sqrt{c})^2 + 2\kappa$, it first appears that the integral is in fact over the compact set $\mathcal{D} = \{x + iy : x \in [a_1, a_2], y \in [0, 2]\}$. Moroever, as $|\mathbf{s}_\mu(z)| \leq \frac{1}{y}$ if $z \in \mathcal{D}$ (see (1.15)), (4.13) 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 (4.14) is well defined.

At this point, we have to introduce some new notations. We denote by $\mathbf{Q}(z)$, $\tilde{\mathbf{Q}}(z)$ and $\hat{\mathbf{Q}}(z)$ the resolvents of matrices $\frac{\mathbf{X}\mathbf{X}^*}{B+1}$, $\tilde{\mathbf{C}}$ and $\hat{\mathbf{C}}$ respectively, while $\tilde{\mu}_N$ and $\hat{\mu}_N$ represent the empirical eigenvalue distributions of matrices $\tilde{\mathbf{C}}$ and $\hat{\mathbf{C}}$.

The term $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})$ can be written as

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}) = \int_{\mathbb{R}} f\ \mathrm{d}\hat{\mu}_N - \int_{\mathbb{R}} f\ \mathrm{d}\tilde{\mu}_N$$

Applying (4.14) to the empirical eigenvalue distributions $\hat{\mu}_N$ and $\tilde{\mu}_N$, we obtain that

$$\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}) = \frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\frac{1}{M}(\mathrm{tr}\ \hat{\mathbf{Q}}(z) - \mathrm{tr}\ \tilde{\mathbf{Q}}(z)) \tag{4.15}$$

In the following, we will drop the notation $\mathbf{Q}(z)$, $\tilde{\mathbf{Q}}(z)$, $\hat{\mathbf{Q}}(z)$ and use instead $\mathbf{Q}$, $\tilde{\mathbf{Q}}$ $\hat{\mathbf{Q}}$. Therefore, we can reformulate (4.11) as follows:

$$\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}})\right| \prec \frac{B}{N} \iff \left|\frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} \bar{\partial}\Phi_k(f)(z)\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\}\,\mathrm{d}x\,\mathrm{d}y\right| \prec \frac{B^2}{N}$$

4.1.1. Reduction to the study of $\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\sum_{m=1}^{M}(\mathbf{Q}+z\mathbf{Q}^2)_{mm}\left(1-\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)$

We define

$$\zeta = \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\sum_{m=1}^{M}(\mathbf{Q}+z\mathbf{Q}^2)_{mm}\left(1-\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right) \tag{4.16}$$

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

$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \zeta\right| \prec \frac{B^2}{N} \tag{4.17}$$

It turns out that by Lemma 4.1 and Lemma 4.3 in Paragraph 4.1.2 below, ζ verifies the key properties :

$$|\zeta| \leq |\zeta - \mathbb{E}\zeta| + |\mathbb{E}\zeta| \prec 1 + 1 \prec 1$$

The condition $\alpha \in (1/2, 1)$ implies that $1 \ll B^2/N$. Proposition 4.3 will then follow directly from (4.17).

Plugging in the integral expression of ζ, we get:

$$\left|\int_{\mathcal{D}} \bar{\partial}\Phi_k(f)(z)\text{tr }\{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\}\,\mathrm{d}x\,\mathrm{d}y-\zeta\right|$$
$$=\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\bar{\partial}\Phi_k(f)(z)\left(\text{tr }\{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\}-\sum_{m=1}^{M}(\mathbf{Q}+z\mathbf{Q}^2)_{mm}\left(1-\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right)\right|$$

We recall the definition of $\boldsymbol{\Theta} := \hat{\mathbf{C}} - \tilde{\mathbf{C}}$ from (3.23). We will proceed in three steps:

1.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\left(\text{tr }\{\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}\}-\text{tr }\{\mathbf{Q}^2\boldsymbol{\Theta}\}\right)\right| \prec \frac{B^2}{N} \tag{4.18}$$

2.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\left(\text{tr }\{\mathbf{Q}^2\boldsymbol{\Theta}\}-2\,\text{tr }\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2}-\mathbf{I})\right)\right| \prec \frac{B^2}{N} \tag{4.19}$$

3.
$$\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\bar{\partial}\Phi_k(f)(z)\times\right.$$
$$\left.\left(2\,\text{tr }\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2}-\mathbf{I})-\sum_{m=1}^{M}(\mathbf{Q}+z\mathbf{Q}^2)_{mm}\left(1-\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right)\right| \prec \frac{B^2}{N} \tag{4.20}$$

Step 1. Using the well known identity $\mathbf{A}^{-1}-\mathbf{B}^{-1}=\mathbf{A}^{-1}(\mathbf{B}-\mathbf{A})\mathbf{B}^{-1}$, we express $\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}$ as:
$$\hat{\mathbf{Q}}-\tilde{\mathbf{Q}}=-\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}} \tag{4.21}$$
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

$$\begin{aligned}
&|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}-\text{tr }\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}|\\
&\qquad=|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}-\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}+\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}-\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}+\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}-\text{tr }\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}|\\
&\qquad\leq|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}-\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}|+|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}-\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}|+|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}-\text{tr }\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}|\\
&\qquad:=T_1+T_2+T_3
\end{aligned}$$

The following rough bounds are enough to control T_1 (we used (1.17) to control the norm of the resolvents):

$$T_1=|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}(\hat{\mathbf{Q}}-\tilde{\mathbf{Q}})|=|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}\boldsymbol{\Theta}\hat{\mathbf{Q}}|\leq M\|\tilde{\mathbf{Q}}\|^2\|\hat{\mathbf{Q}}\|\|\|\boldsymbol{\Theta}\|^2\leq\frac{1}{\text{Im}^3 z}M\|\boldsymbol{\Theta}\|^2$$

Concerning T_2 and T_3, we write similarly that $\tilde{\mathbf{Q}}-\mathbf{Q}=-\tilde{\mathbf{Q}}\tilde{\boldsymbol{\Delta}}\mathbf{Q}$, and obtain that

$$T_2=|\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\tilde{\mathbf{Q}}-\text{tr }\tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q}|\leq M\|\tilde{\mathbf{Q}}\|^2\|\mathbf{Q}\|\|\tilde{\boldsymbol{\Delta}}\|\boldsymbol{\Theta}\|\leq\frac{1}{\text{Im}^3 z}M\|\tilde{\boldsymbol{\Delta}}\|\|\boldsymbol{\Theta}\|$$

$$T_3 = |\mathrm{tr}\ \tilde{\mathbf{Q}}\boldsymbol{\Theta}\mathbf{Q} - \mathrm{tr}\ \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}| \leq M\|\tilde{\mathbf{Q}}\|\|\mathbf{Q}\|^2\|\tilde{\boldsymbol{\Delta}}\|\boldsymbol{\Theta}\| \leq \frac{1}{\mathrm{Im}^3 z} M\|\tilde{\boldsymbol{\Delta}}\|\|\boldsymbol{\Theta}\|$$

Plugging these estimations into the left hand side of (4.18), we obtain that

$$\begin{aligned}
&\left|\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \overline{\partial}\Phi_k(f)(z) \left(\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \mathrm{tr}\ \{\mathbf{Q}^2\boldsymbol{\Theta}\}\right)\right| \\
&\qquad\qquad \leq \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial}\Phi_k(f)(z)|(T_1 + T_2 + T_3) \\
&\qquad\qquad \leq M(\|\boldsymbol{\Theta}\|^2 + 2\|\tilde{\boldsymbol{\Delta}}\|\boldsymbol{\Theta}\|) \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial}\Phi_k(f)(z)| \frac{1}{\mathrm{Im}^3 z}
\end{aligned}$$

The use of (4.13) for $k = 3$ leads to the conclusion that

$$\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial}\Phi_k(f)(z)| \frac{1}{\mathrm{Im}^3 z} < +\infty$$

Moreover, the concentration results (3.27) for $\|\boldsymbol{\Theta}\|$ and (3.19) for $\|\tilde{\boldsymbol{\Delta}}\|$ from Proposition 3.1, as well as $\alpha > 1/2$ imply that

$$\|\boldsymbol{\Theta}\|^2 + 2\|\boldsymbol{\Theta}\|\|\tilde{\boldsymbol{\Delta}}\| \prec \frac{1}{B} + \frac{1}{\sqrt{B}}\frac{B}{N} + \left(\frac{B}{N}\right)^3 << \frac{B}{N}$$

This eventually provides:

$$\left|\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y\, \overline{\partial}\Phi_k(f)(z) \left(\mathrm{tr}\ \{\hat{\mathbf{Q}} - \tilde{\mathbf{Q}}\} - \mathrm{tr}\ \{\mathbf{Q}^2\boldsymbol{\Theta}\}\right)\right| \prec \frac{B^2}{N}$$

which proves (4.18) and ends Step 1.

Step 2. We claim that:

$$\left\|\boldsymbol{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\| \prec \frac{B}{N} \tag{4.22}$$

We recall that $\hat{\mathbf{S}}$ can be written using the definition (1.6) of $\tilde{\mathbf{C}}$, and use the decomposition (3.18) of $\tilde{\mathbf{C}}$ from Corollary 3.1. Using these results, we get that

$$\hat{\mathbf{S}} = \mathbf{D}^{1/2}\, \tilde{\mathbf{C}}\, \mathbf{D}^{1/2} = \mathbf{D}^{1/2} \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\boldsymbol{\Delta}}\right) \mathbf{D}^{1/2}$$

Plugging this expression of $\hat{\mathbf{S}}$ into (3.24), we obtain that

$$\begin{aligned}
\mathbf{\Theta} &= (\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2})\hat{\mathbf{S}}\hat{\mathbf{D}}^{-1/2} + \mathbf{D}^{-1/2}\hat{\mathbf{S}}(\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}) \\
&= (\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2})\mathbf{D}^{1/2}\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}}\right)\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} \\
&\qquad\qquad + \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}}\right)\mathbf{D}^{1/2}(\hat{\mathbf{D}}^{-1/2} - \mathbf{D}^{-1/2}) \\
&= (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}}\right)\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} \\
&\qquad\qquad + \left(\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \tilde{\mathbf{\Delta}}\right)(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \\
&:= \mathbf{\Theta}_1 + \mathbf{\Theta}_2
\end{aligned}$$

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:

$$\begin{aligned}
&\left\|\mathbf{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\right\| \\
&\qquad\qquad = \left\|(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\tilde{\mathbf{\Delta}}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\right\| \\
&\qquad\qquad \leq \|\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\|\|\tilde{\mathbf{\Delta}}\|\|\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\|
\end{aligned} \tag{4.23}$$

$\tilde{\mathbf{\Delta}}$ is controlled by (3.19) from Corollary 3.1, and $\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}$ is controlled by (3.26) from Lemma 3.3 (it is a diagonal matrix which elements are stochastically dominated by Lemma 3.3). Moreover, from Lemma 3.3, it holds that $\|\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\| \prec 1$. Combining these estimations into (4.23), one get:

$$\left\|\mathbf{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2}\right\| \prec \left(\frac{1}{\sqrt{B}} + \frac{B^2}{N^2}\right)\frac{B}{N} \prec \frac{B}{N} \tag{4.24}$$

Using (2.5) from Paragraph 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 $(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}$. To check this, we use Lemma 3.3 and write that

$$\begin{aligned}
&\left\|(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right\| \\
&\qquad\qquad = \left\|(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}_M)\right\| \\
&\qquad\qquad \leq \|\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}\|^2\left\|\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right\| \\
&\qquad\qquad \prec \left(\frac{1}{\sqrt{B}} + \frac{B^2}{N^2}\right)^2 \\
&\qquad\qquad \prec \frac{B}{N}
\end{aligned} \tag{4.25}$$

Collecting (4.23), (4.24) and (4.25), we obtain the desired approximation of $\mathbf{\Theta}_1$:

$$\left\|\mathbf{\Theta}_1 - (\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right\| \prec \frac{B}{N} \tag{4.26}$$

Similarly for $\mathbf{\Theta}_2$, one would obtain:

$$\left\|\mathbf{\Theta}_2 - \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right\| \prec \frac{B}{N} \tag{4.27}$$

Combining (4.26) and (4.27), we obtain (4.22). To finish the proof of Step 2, it remains to consider tr $\mathbf{Q}^2\mathbf{\Theta}$ and prove (4.19). Remark that $\mathbf{X}\mathbf{X}^*/(B+1)$ and its resolvent $\mathbf{Q}$ commutes.

$$\begin{aligned}
&\mathrm{tr}\ \mathbf{Q}^2\left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right) \\
&\qquad = \mathrm{tr}\ \mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \mathrm{tr}\ \mathbf{Q}^2\frac{\mathbf{X}\mathbf{X}^*}{B+1}(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}) \\
&\qquad = 2\,\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})
\end{aligned} \tag{4.28}$$

Therefore, using (4.28):

$$\begin{aligned}
&\left\|\mathrm{tr}\ \mathbf{Q}^2\mathbf{\Theta} - 2\,\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\right\| \\
&\quad = \left\|\mathrm{tr}\ \mathbf{Q}^2\mathbf{\Theta} - \mathrm{tr}\ \mathbf{Q}^2\left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\| \\
&\quad \leq M\|\mathbf{Q}\|^2\left\|\mathbf{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\|
\end{aligned} \tag{4.29}$$

so that the left hand side of (4.22) is recognised in the right hand side of (4.29). We can eventually prove (4.19) by following the same idea as in Step 1:

$$\begin{aligned}
&\left|\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\left(\mathrm{tr}\ \{\mathbf{Q}^2\mathbf{\Theta}\} - 2\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\right)\right| \\
&\qquad \leq M\left\|\mathbf{\Theta} - \left((\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I})\frac{\mathbf{X}\mathbf{X}^*}{B+1} + \frac{\mathbf{X}\mathbf{X}^*}{B+1}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})\right)\right\| \\
&\qquad\qquad \times \underbrace{\int_{\mathcal{D}} |\overline{\partial}\Phi_k(f)(z)|\frac{1}{\mathrm{Im}^2 z}\,\mathrm{d}x\,\mathrm{d}y}_{<+\infty} \\
&\qquad \prec \frac{B^2}{N}
\end{aligned}$$

This proves (4.19) 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:

$$\frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q} = \mathbf{I}_M + z\mathbf{Q} \tag{4.30}$$

Using (4.30) one can write:

$$\begin{aligned}\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) &= \mathrm{tr}\ (\mathbf{I} + z\mathbf{Q})\mathbf{Q}(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \\ &= \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \\ &= \sum_{m=1}^{M} \left(\mathbf{Q} + z\mathbf{Q}^2\right)_{mm} \left(\sqrt{\frac{s_m}{\hat{s}_m}} - 1\right) \end{aligned} \tag{4.31}$$

To handle $\sqrt{\frac{s_m}{\hat{s}_m}} - 1$ we use the following Taylor expansion: define the application h by $h(u) = \frac{1}{\sqrt{u}} - 1$, 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:

$$\begin{aligned} h\left(\frac{\hat{s}_m}{s_m}\right) &= h(1) + \left(\frac{\hat{s}_m}{s_m} - 1\right)h'(1) + \frac{1}{2}\left(\frac{\hat{s}_m}{s_m} - 1\right)^2 h''(\theta_m) \\ &= -\frac{1}{2s_m}(\hat{s}_m - s_m) + \frac{1}{2}\frac{h''(\theta_m)}{s_m^2}(\hat{s}_m - s_m)^2 \end{aligned}$$

where θ_m is some random quantity between $\hat{s}_m$ and s_m. Therefore (4.31) becomes

$$\begin{aligned} &\mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I}) \\ &\quad = \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)\mathrm{diag}\left(-\frac{\hat{s}_m - s_m}{2s_m} + \frac{1}{2}\frac{h''(\theta_m)(\hat{s}_m - s_m)^2}{s_m^2} : m \in \{1, \ldots, M\}\right) \end{aligned}$$

Lemma 3.1 implies that the set $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ defined by (3.25) holds with exponentially high probability. Therefore, it is sufficient to study the term $\mathrm{tr}\ \frac{\mathbf{X}\mathbf{X}^*}{B+1}\mathbf{Q}^2(\mathbf{D}^{1/2}\hat{\mathbf{D}}^{-1/2} - \mathbf{I})$ on the event $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$. If $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ holds, θ_m belongs to $[\underline{s}, \overline{s}] + \epsilon$ 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

$$\begin{aligned} &\left|\mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)\mathrm{diag}\left(\frac{1}{2}\frac{h''(\theta_m)(\hat{s}_m - s_m)^2}{s_m^2} : m \in \{1, \ldots, M\}\right)\right| \\ &\qquad\qquad \leq C(\|\mathbf{Q}\| + z\|\mathbf{Q}\|^2)\sum_{m=1}^{M}(\hat{s}_m - s_m)^2 \end{aligned}$$

holds on $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$. Following again the same argument as in Step 1, we obtain that

$$\left| \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y \overline{\partial} \Phi_k(f)(z) \left\{ \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}) \right.\right.$$
$$\left.\left. - \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2) \mathrm{diag}\left(-\frac{\hat{s}_m - s_m}{2 s_m} : m \in \{1, \ldots, M\} \right) \right\} \right|$$
$$\leq C \sum_{m=1}^{M} (\hat{s}_m - s_m)^2 \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial} \Phi_k(f)(z)| \left(\frac{1}{\mathrm{Im}\, z} + \frac{|z|}{\mathrm{Im}^2 z} \right)$$

on $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$. Lemma A.5 in Appendix implies that

$$\sum_{m=1}^{M} (\hat{s}_m - s_m)^2 \prec 1 + \frac{B^5}{N^4} \prec \frac{B^2}{N}$$

Moreover, as $\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y |\overline{\partial} \Phi_k(f)(z)| (\frac{1}{\mathrm{Im}\, z} + \frac{|z|}{\mathrm{Im}^2 z}) < \infty$ for $k \geq 2$, we obtain immediately that

$$\left| \int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y \overline{\partial} \Phi_k(f)(z) \left\{ \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2)(\hat{\mathbf{D}}^{-1/2}\mathbf{D}^{1/2} - \mathbf{I}) \right.\right.$$
$$\left.\left. - \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2) \mathrm{diag}\left(-\frac{\hat{s}_m - s_m}{2 s_m} : m \in \{1, \ldots, M\} \right) \right\} \right|$$
$$\prec \frac{B^2}{N}$$

Therefore the dominant term that remains is the order one quantity from the previous Taylor expansion:

$$\int_{\mathcal{D}} \mathrm{d}x\, \mathrm{d}y \overline{\partial} \Phi_k(f)(z) \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2) \mathrm{diag}\left(\frac{s_m - \hat{s}_m}{2 s_m} : m \in \{1, \ldots, M\} \right)$$

We now claim (see the proof below) that the following holds:

$$\left| \frac{s_m - \hat{s}_m}{s_m} - \left(1 - \frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right| \prec \frac{B}{N} \tag{4.32}$$

which since that:

$$\left| \mathrm{tr}\ (\mathbf{Q} + z\mathbf{Q}^2) \mathrm{diag}\left(\frac{s_m - \hat{s}_m}{2 s_m} - \frac{1}{2}\left(1 - \frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right) \right|$$
$$\leq \frac{M}{2} (\|\mathbf{Q}\| + |z| \|\mathbf{Q}\|^2) \sup_{m=1,\ldots,M} \left| \frac{s_m - \hat{s}_m}{s_m} - \left(1 - \frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right|$$

eventually proves (4.20). To show that (4.32) is indeed true, we recall equation (3.7), and deduce that $\hat{s}_m = \frac{\|\boldsymbol{\omega}_m\|^2}{B+1}$ can be written as

$$\hat{s}_m = s_m \frac{\mathbf{x}_m^* (\mathbf{I} + \boldsymbol{\Phi}_m) \mathbf{x}_m}{B+1} = s_m \frac{\|\mathbf{x}_m\|_2^2}{B+1} + s_m \frac{\mathbf{x}_m^* \boldsymbol{\Phi}_m \mathbf{x}_m}{B+1}$$

where

$$\sup_{m\geq 1}\sup_{i,j}|(\mathbf{\Phi}_m)_{ij}| = \mathcal{O}\left(\frac{1}{N}\right)$$

This gives:

$$\left|(s_m - \hat{s}_m) - s_m\left(1 - \frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right| = \left|s_m \frac{\mathbf{x}_m^* \mathbf{\Phi}_m \mathbf{x}_m}{B+1}\right|$$

Since $\mathbf{x}_m \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_{B+1})$, and

$$\left\|s_m \frac{\mathbf{\Phi}_m}{B+1}\right\|_F^2 = \mathcal{O}\left((B+1)^2 \frac{1}{(N(B+1))^2}\right) = \mathcal{O}\left(\frac{1}{N^2}\right)$$

the complex Hanson-Wright inequality from Paragraph 2.4 can be applied, and provides:

$$\left|s_m \frac{\mathbf{x}_m^* \mathbf{\Phi}_m \mathbf{x}_m}{B+1} - \mathbb{E}\left[s_m \frac{\mathbf{x}_m^* \mathbf{\Phi}_m \mathbf{x}_m}{B+1}\right]\right| \prec \frac{1}{N} \ll \frac{B}{N} \tag{4.33}$$

Moreover, it is clear that

$$\mathbb{E}\left[\frac{\mathbf{x}_m^* \mathbf{\Phi}_m \mathbf{x}_m}{B+1}\right] = \sum_{b=1}^{B+1} \mathbb{E}\left[|X_{mb}|^2\right] \frac{(\mathbf{\Phi}_m)_{bb}}{B+1} \leq \frac{C}{N} \tag{4.34}$$

Combining the variance estimation (4.33) and the bias estimation (4.34) of $s_m \frac{\mathbf{x}_m^* \mathbf{\Phi}_m \mathbf{x}_m}{B+1}$, (4.32) is indeed true.

Up to the Lemma 4.1 and Lemma 4.3, Proposition 4.3 is proved. □

4.1.2. Proof of Lemma 4.1 and Lemma 4.3

We now establish Lemma 4.1 and Lemma 4.3.

Lemma 4.1. *The family of random variables $\zeta(\nu) - \mathbb{E}\zeta(\nu)$, $\nu \in [0,1]$:*

$$|\zeta(\nu) - \mathbb{E}\zeta(\nu)| \prec 1 \tag{4.35}$$

Proof. ζ defined by (4.16) can be written as

$$\begin{aligned}\zeta = &- \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z) \sum_{m=1}^{M} \mathbf{Q}_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right) \\ &\qquad - \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z) \sum_{m=1}^{M} z(\mathbf{Q}^2)_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right) \\ &:= -\zeta_1 - \zeta_2\end{aligned}$$

In the following, we omit to evaluate $|\zeta_1(\nu) - \mathbb{E}(\zeta_1(\nu))|$, and just establish that $|\zeta_2(\nu) - \mathbb{E}(\zeta_2(\nu))| \prec 1$ using the Gaussian concentration inequality from Paragraph 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 $\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} - 1\right)_{m=1,\ldots,M}$ are localized, which holds with exponentially high probability, ζ_2 is a $\mathcal{O}(B^\epsilon)$–Lipschitz function of the entries of matrix $\mathbf{X}$ for any $\epsilon > 0$. Let $0 < \epsilon < \frac{1}{2}$, and define the family of events $A_{m,\epsilon}(\nu)$, $m = 1, \ldots, M$, $\nu \in [0, 1]$ given by

$$A_{m,\epsilon}(\nu) = \left\{ \frac{\|\mathbf{x}_m(\nu)\|_2^2}{B+1} \in \left[1 - \frac{B^\epsilon}{\sqrt{B}}, 1 + \frac{B^\epsilon}{\sqrt{B}} \right] \right\} \tag{4.36}$$

as well as $A_\epsilon(\nu) = \cap_{m=1}^M A_{m,\epsilon}(\nu)$. It is clear that the family of events $A_{m,\epsilon}(\nu)$, $m = 1, \ldots, M$, $\nu \in [0, 1]$ holds with exponentially high probability, and that the same property holds for the family $A_\epsilon(\nu), \nu \in [0, 1]$. We claim that it exists a family of C^∞ functions $(g_{B,\epsilon})_{B\geq 1}$ verifying

$$g_{B,\epsilon}(t) = \begin{cases} t - 1 & \text{if } t \in [1 - \frac{B^\epsilon}{\sqrt{B}}, 1 + \frac{B^\epsilon}{\sqrt{B}}] \\ 0 & \text{if } t \notin [1 - 2\frac{B^\epsilon}{\sqrt{B}}, 1 + 2\frac{B^\epsilon}{\sqrt{B}}] \end{cases}$$

and

$$\sup_t |g_{B,\epsilon}(t)| \leq C \, \frac{B^\epsilon}{\sqrt{B}}, \quad \sup_t |g'_{B,\epsilon}(t)| \leq C \tag{4.37}$$

for each B, where C is a nice constant. Indeed consider $h \in C^\infty$ such that it verifies $|h(t)| \leq 2|t|$ 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 (4.37).

We define $\tilde{\zeta}_{2,\epsilon}$ by

$$\tilde{\zeta}_{2,\epsilon} = \mathrm{Re} \int_{\mathcal{D}} \mathrm{d}x \, \mathrm{d}y \overline{\partial} \Phi_k(f)(z) \sum_{m=1}^M (z\mathbf{Q}^2)_{mm} g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right), \quad (\mathbf{Q}^2)_{mm} = \sum_{k=1}^M Q_{mk} Q_{km}$$

and notice that ζ_2 and $\tilde{\zeta}_{2,\epsilon}$ coincide on the exponentially high probability event $A_\epsilon(\nu)$. We claim that if $|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| \prec B^\epsilon$, then $|\zeta_2 - \mathbb{E}(\zeta_2)| \prec B^\epsilon$. Since ϵ

is arbitrary and $B^\epsilon = \mathcal{O}(N^{\alpha\epsilon})$, Remark 2.1 will imply that $|\zeta_2 - \mathbb{E}(\zeta_2)| \prec 1$. To justify this, we evaluate $\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > B^\epsilon N^\delta\right)$ for each $\delta > 0$. It holds that

$$\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}\right) \leq \mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}, A_\epsilon\right) + \mathbb{P}(A_\epsilon^c)$$

As $\mathbb{P}(A_\epsilon^c)$ converges towards zero exponentially, we have just to consider

$$\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}, A_\epsilon\right)$$

and write, since ζ_2 and $\tilde{\zeta}_{2,\epsilon}$ coincide on A_ϵ,

$$\begin{aligned}\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}, A_\epsilon\right) &= \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}, A_\epsilon\right) \\ &\leq \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > N^{\frac{\epsilon}{\alpha}+\delta} - |E(\zeta_2 - \tilde{\zeta}_{2,\epsilon})|, A_\epsilon\right)\end{aligned}$$

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

$$|\mathbb{E}(\zeta_2 - \tilde{\zeta}_{2,\epsilon})| = \left|\mathbb{E}((\zeta_2 - \tilde{\zeta}_{2,\epsilon})\mathbb{I}_{A_\epsilon^c})\right| \leq \left(\mathbb{E}\left|\zeta_2 - \tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2} (\mathbb{P}(A_\epsilon^c))^{1/2}$$

A rough evaluation of $\left(\mathbb{E}\left|\zeta_2 - \tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2}$ leads to

$$\left(\mathbb{E}\left|\zeta_2 - \tilde{\zeta}_{2,\epsilon}\right|^2\right)^{1/2} \leq C\,M$$

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. For each N large enough, we thus have

$$\begin{aligned}\mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > N^{\frac{\epsilon}{\alpha}+\delta} - |\mathbb{E}(\zeta_2 - \tilde{\zeta}_{2,\epsilon})|, A_\epsilon\right) &\leq \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > N^{\frac{\epsilon}{\alpha}+\delta/2}, A_\epsilon\right) \\ &\leq \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > N^{\frac{\epsilon}{\alpha}+\delta/2}\right)\end{aligned}$$

We have therefore established that

$$\mathbb{P}\left(|\zeta_2 - \mathbb{E}(\zeta_2)| > N^{\frac{\epsilon}{\alpha}+\delta}, A_\epsilon\right) \leq \mathbb{P}\left(|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| > N^{\frac{\epsilon}{\alpha}+\delta/2}\right)$$

which eventually justifies that if $|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| \prec B^\epsilon$, then $|\zeta_2 - \mathbb{E}(\zeta_2)| \prec B^\epsilon$.

Therefore, it remains to prove that $|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| \prec B^\epsilon$. This is true by Lemma 4.2 below. The stochastic domination relation $|\zeta_1 - \mathbb{E}\zeta_1| \prec B^\epsilon$ is proved similarly. The completes the proof of Lemma 4.1. $\square$

Lemma 4.2.

$$|\tilde{\zeta}_{2,\epsilon} - \mathbb{E}(\tilde{\zeta}_{2,\epsilon})| \prec B^{\epsilon}$$

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 $\mathbf{Q} = (\frac{\mathbf{X}\mathbf{X}^*}{B+1} - z\mathbf{I}_M)^{-1}$, and its corresponding derivative with respect to X_{ij} is

$$\frac{\partial Q_{mk}}{\partial X_{ij}} = \frac{-1}{B+1} Q_{mi} (\mathbf{X}^*\mathbf{Q})_{jk}$$

Therefore, we have

$$\begin{aligned}
\frac{\partial (\mathbf{Q}^2)_{mm}}{\partial X_{ij}} &= \sum_{k=1}^{M} \frac{\partial (Q_{mk} Q_{km})}{\partial X_{ij}} \\
&= \sum_{k=1}^{M} \left\{ Q_{mk} \frac{\partial Q_{km}}{\partial X_{ij}} + Q_{km} \frac{\partial Q_{mk}}{\partial X_{ij}} \right\} \\
&= -\left(\frac{(\mathbf{Q}^2)_{mi} (\mathbf{X}^*\mathbf{Q})_{jm}}{B+1} + \frac{Q_{mi} (\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1} \right) \qquad (4.38)
\end{aligned}$$

Moreover it is clear that

$$\frac{\partial}{\partial X_{ij}} \left(g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right) = \delta_{im} \frac{\overline{X_{m,j}}}{B+1} g'_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \qquad (4.39)$$

Collecting the derivatives (4.38) and (4.39) we get:

$$\begin{aligned}
&\frac{\partial}{\partial X_{ij}} \left(\sum_{m=1}^{M} (\mathbf{Q}^2)_{mm} g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right) \\
&= \sum_{m=1}^{M} \left\{ g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \frac{\partial (\mathbf{Q}^2)_{mm}}{\partial X_{ij}} + (\mathbf{Q}^2)_{mm} \frac{\partial}{\partial X_{ij}} g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \right\} \\
&= \sum_{m=1}^{M} \{ -g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \left(\frac{(\mathbf{Q}^2)_{mi} (\mathbf{X}^*\mathbf{Q})_{jm}}{B+1} + \frac{Q_{mi} (\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1} \right) \\
&\qquad + \delta_{im} \frac{\overline{X_{m,j}}}{B+1} g'_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) (\mathbf{Q}^2)_{mm} \} \\
&= \frac{\overline{X_{i,j}}}{B+1} g'_{B,\epsilon} \left(\frac{\|\mathbf{x}_i\|_2^2}{B+1} \right) (\mathbf{Q}^2)_{ii} \\
&\qquad - \sum_{m=1}^{M} g_{B,\epsilon} \left(\frac{\|\mathbf{x}_m\|_2^2}{B+1} \right) \left(\frac{(\mathbf{Q}^2)_{mi} (\mathbf{X}^*\mathbf{Q})_{jm}}{B+1} + \frac{Q_{mi} (\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1} \right) \qquad (4.40)
\end{aligned}$$

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:

$$\frac{\partial \tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}} = \mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\frac{\partial}{\partial X_{ij}}\left(\sum_{m=1}^{M} z(\mathbf{Q}^2)_{mm} g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\right)$$

Plugging in the derivative computed in (4.40) we get:

$$\begin{aligned}\frac{\partial \tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}} = \mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\,z\Bigg\{&\frac{\overline{X_{i,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_i\|_2^2}{B+1}\right)(\mathbf{Q}^2)_{ii}\\ &-\sum_{m=1}^{M} g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right)\Bigg\}\end{aligned}$$

Using the bounds of $g_{B,\epsilon}$ and $g'_{B,\epsilon}$ from inequalities (4.37), 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:

$$\begin{aligned}\left|\frac{\partial \tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 &\leq C\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2\Bigg|\frac{\overline{X_{i,j}}}{B+1}g'_{B,\epsilon}\left(\frac{\|\mathbf{x}_i\|_2^2}{B+1}\right)(\mathbf{Q}^2)_{ii}\\ &\qquad -\sum_{m=1}^{M} g_{B,\epsilon}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}\right)\left(\frac{(\mathbf{Q}^2)_{mi}(\mathbf{X}^*\mathbf{Q})_{jm}}{B+1}+\frac{Q_{mi}(\mathbf{X}^*\mathbf{Q}^2)_{jm}}{B+1}\right)\Bigg|^2\\ &\leq C\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2\left|\frac{\overline{X_{i,j}}}{B+1}(\mathbf{Q}^2)_{ii}\right|^2\,\mathbb{1}\left(\left|\frac{\|\mathbf{x}_i\|_2^2}{B+1}-1\right|\leq\frac{2B^\epsilon}{\sqrt{B}}\right)\\ &\qquad + C\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2\left(\frac{B^\epsilon}{\sqrt{B}}\right)^2\left|\sum_{m=1}^{M}\frac{\left|(\mathbf{Q}^2)_{mi}\right|\left|(\mathbf{X}^*\mathbf{Q})_{jm}\right|}{B+1}\right|^2\\ &\qquad + C\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2\left(\frac{B^\epsilon}{\sqrt{B}}\right)^2\left|\sum_{m=1}^{M}\frac{|Q_{mi}|\left|(\mathbf{X}^*\mathbf{Q}^2)_{jm}\right|}{B+1}\right|^2\\ &:= C(T_{ij}^{(1)}+T_{ij}^{(2)}+T_{ij}^{(3)})\end{aligned}$$

It remains to sum over i,j.

$$
\begin{aligned}
&\sum_{i,j=1}^{M} T_{ij}^{(1)} \\
&= \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\, \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i,j=1}^{M} \left| \frac{\overline{X_{i,j}}}{B+1}(\mathbf{Q}^2)_{ii} \right|^2 \mathbb{1}\left(\left| \frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1 \right| \leq \frac{2B^{\epsilon}}{\sqrt{B}} \right) \\
&\leq \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\, \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2\, \mathbb{1}\left(\left| \frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1 \right| \leq \frac{2B^{\epsilon}}{\sqrt{B}} \right) \sum_{j=1}^{M} \left| \frac{\overline{X_{i,j}}}{B+1} \right|^2 \\
&= \frac{C}{B+1} \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\, \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2 \frac{\|\mathbf{x}_i\|_2^2}{B+1}\, \mathbb{1}\left(\left| \frac{\|\mathbf{x}_i\|_2^2}{B+1} - 1 \right| \leq \frac{2B^{\epsilon}}{\sqrt{B}} \right) \\
&\leq \frac{C}{B+1}\left(1 + \frac{2B^{\epsilon}}{\sqrt{B}}\right) \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\, \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{i=1}^{M} |(\mathbf{Q}^2)_{ii}|^2
\end{aligned}
$$

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:

$$
\begin{aligned}
&\sum_{i,j=1}^{M} T_{ij}^{(2)} \\
&\leq \frac{B^{2\epsilon} M}{(B+1)^3} \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\, \left|\overline{\partial}\Phi_k(f)(z)\right|^2 \sum_{m=1}^{M} \left(\sum_{i=1}^{M} \left|(\mathbf{Q}^2)_{mi}\right|^2 \right) \left(\sum_{j=1}^{M} |(\mathbf{X}^*\mathbf{Q})_{jm}|^2 \right)
\end{aligned}
$$

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,

$$\sum_{m=1}^{M}\left(\sum_{j=1}^{M}|(\mathbf{X}^*\mathbf{Q})_{jm}|^2\right) = \mathrm{tr}\ \mathbf{X}^*\mathbf{Q}\mathbf{Q}^*\mathbf{X}$$
$$= (B+1)\mathrm{tr}\ \left((\mathbf{I}+z\mathbf{Q})\,\mathbf{Q}^*\right) \leq M(B+1)(\|\mathbf{Q}\|+|z|\|\mathbf{Q}\|^2)$$

therefore

$$\sum_{i,j=1}^{M} T_{ij}^{(2)} \leq B^{2\epsilon}\left(\frac{M}{B+1}\right)^2 \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2 \|\mathbf{Q}^4\|(\|\mathbf{Q}\|+|z|\|\mathbf{Q}\|^2)$$

and similarly for $T_{ij}^{(3)}$ one get:

$$\sum_{i,j=1}^{M} T_{ij}^{(3)} \leq B^{2\epsilon}\left(\frac{M}{B+1}\right)^2 \int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2 \|\mathbf{Q}^2\|(\|\mathbf{Q}\|^3+|z|\|\mathbf{Q}\|^4)$$

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:

$$\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 \leq CB^{2\epsilon}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2 (\|\mathbf{Q}\|^4+\|\mathbf{Q}^5\|+|z|\|\mathbf{Q}^6\|)$$

In conjunction with $|z|$ bounded on $\mathcal{D}$ and $\|\mathbf{Q}\| \leq \frac{1}{\mathrm{Im}\,z}$, we obtain that

$$\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 \leq CB^{2\epsilon}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\left|\overline{\partial}\Phi_k(f)(z)\right|^2 \left(\frac{1}{\mathrm{Im}^4 z}+\frac{1}{\mathrm{Im}^5 z}+\frac{1}{\mathrm{Im}^6 z}\right) \tag{4.41}$$

This integral at the right hand side of (4.41) is finite as soon as $k \geq 3$. Therefore, we proved that:

$$\sum_{i,j}\left|\frac{\partial\tilde{\zeta}_{2,\epsilon}}{\partial X_{ij}}\right|^2 = \mathcal{O}(B^{2\epsilon})$$

as expected. □

It remains to study $\mathbb{E}[\zeta]$, and establish the following Lemma.

Lemma 4.3.

$$|\mathbb{E}\zeta| = \mathcal{O}(1)$$

Proof. As in the proof of Lemma 4.1, we only consider

$$\mathbb{E}[\zeta_2] = \mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,\overline{\partial}\Phi_k(f)(z)\sum_{m=1}^{M} z\mathbb{E}\left[(\mathbf{Q}^2)_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right)\right]$$

as $\mathbb{E}[\zeta_1]$ is shown to be also $\mathcal{O}(1)$ with the same argument. Write

$$\mathbb{E}\left[(\mathbf{Q}^2)_{mm}\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right)\right] =$$
$$\mathbb{E}\left[\left((\mathbf{Q}^2)_{mm}-\mathbb{E}[(\mathbf{Q}^2)_{mm}]\right)\left(\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right)\right]+\mathbb{E}[(\mathbf{Q}^2)_{mm}]\underbrace{\mathbb{E}\left[\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right]}_{=0}$$

Apply now the Cauchy-Schwartz inequality on the remaining term to get the following inequality:

$$|\mathbb{E}[\zeta_2]| \leq \mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\overline{\partial}\Phi_k(f)(z)|\sum_{m=1}^{M}|z|\sqrt{\mathrm{Var}(\mathbf{Q}^2)_{mm}}\sqrt{\mathbb{E}\left|\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right|^2} \tag{4.42}$$

$\|\mathbf{x}_m\|_2^2$ is a χ_2 random variable with $2(B+1)$ degrees of freedom, therefore it is clear that:

$$\mathbb{E}\left|\frac{\|\mathbf{x}_m\|_2^2}{B+1}-1\right|^2=\mathcal{O}\left(\frac{1}{B}\right) \tag{4.43}$$

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 $\mathbb{R}^M$. In order to evaluate the variance of $\mathbf{Q}^2_{mm}$, we establish the following Lemma.

Lemma 4.4. *If $\mathbf{A}$ is a deterministic $M\times M$ matrix, and if $h(\mathbf{X})$ is defined by*

$$h(\mathbf{X}):=\mathrm{tr}\ \mathbf{Q}^2\mathbf{A}=\mathrm{tr}\ \left(\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}-z\mathbf{I}_M\right)^{-2}\mathbf{A}\right)$$

then, it holds that

$$\mathrm{Var}(h(\mathbf{X}))\leq\frac{C}{B+1}\left(\frac{1}{\mathrm{Im}^5 z}+\frac{|z|}{\mathrm{Im}^6 z}\right)\mathrm{tr}\ (\mathbf{A}\mathbf{A}^*) \tag{4.44}$$

for some nice constant C.

The proof follows directly from the Nash-Poincaré inequality (see e.g. [26, Proposition 2.1.6] in the Gaussian real case and Eq. (18) in [15] in the complex Gaussian case) that can be used because the entries of matrix $\mathbf{X}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ distributed random variables:

$$\mathrm{Var}\,h(\mathbf{X})\leq\mathbb{E}\|\nabla h\|_2^2(\mathbf{X})$$

where

$$\|\nabla h\|_2^2=\sum_{i,j}\left|\frac{\partial h}{\partial X_{ij}}\right|^2+\left|\frac{\partial h}{\partial\overline{X_{ij}}}\right|^2$$

Standard calculations lead immediately to (4.44)

Applying Lemma 4.4 with $\mathbf{A} = \mathbf{e}_m \mathbf{e}_m^T$ provides (tr $(\mathbf{e}_m \mathbf{e}_m^T)(\mathbf{e}_m \mathbf{e}_m^T)^* = 1$):

$$\mathrm{Var}(\mathbf{Q}^2)_{mm} \leq \frac{C}{B+1}\left(\frac{1}{\mathrm{Im}^5 z} + \frac{|z|}{\mathrm{Im}^6 z}\right) \tag{4.45}$$

for some nice constant C. Using (4.45) in (4.42) we get:

$$\begin{aligned}
|\mathbb{E}\zeta_2| &\leq C\frac{1}{B+1}\mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\overline{\partial}\Phi_k(f)(z)|\sum_{m=1}^{M}|z|\sqrt{\frac{1}{\mathrm{Im}^5 z}+\frac{|z|}{\mathrm{Im}^6 z}} \\
&\leq C\frac{M}{B+1}\mathrm{Re}\int_{\mathcal{D}} \mathrm{d}x\,\mathrm{d}y\,|\overline{\partial}\Phi_k(f)(z)|\left(|z|\frac{1}{\mathrm{Im}^{5/2} z}+|z|^{3/2}\frac{1}{\mathrm{Im}^3 z}\right)
\end{aligned}$$

Taking $k+1 \geq 4$ is enough to make the integral finite. Under Assumption 4.1, this condition is realized, so $|\mathbb{E}[\zeta_2]| = \mathcal{O}(1)$.

□

We are now in position to complete the proof of (4.2). For this, we use Proposition 4.1, Proposition 4.2, and Proposition 4.3, and write that

$$\begin{aligned}
&\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| \\
&\quad\leq \left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu))\right| + \left|\frac{1}{M}\mathrm{tr}\ f(\tilde{\mathbf{C}}(\nu)) - \frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right)\right| \\
&\qquad + \left|\frac{1}{M}\mathrm{tr}\ f\left(\frac{\mathbf{X}\mathbf{X}^*}{B+1}\right) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}\right| \\
&\quad\prec \frac{B}{N} + \frac{B}{N} + \frac{1}{M} \\
&\quad\prec \frac{B}{N}
\end{aligned}$$

as expected.

4.2. Step 2: Lipschitz argument

The following lemma states that with overwhelming probability uniformly over $\nu \in [0,1]$, the application $\nu \mapsto \hat{\mathbf{S}}(\nu)$ is $\mathcal{O}(MN^{3/2})$–Lipschitz.

Proposition 4.4. *It holds that*

$$\sup_{\delta\neq 0}\sup_{\nu\in[0,1]}\frac{\|\hat{\mathbf{S}}(\nu)-\hat{\mathbf{S}}(\nu+\delta)\|}{|\delta|} \prec MN^{3/2} \tag{4.46}$$

Proof. Let $\delta \in \mathbb{R}$ and $\nu \in [0,1]$. As the random variables $(y_{m,n})_{m=1,\ldots,M,n=1,\ldots,N}$ are complex Gaussian and that $\sup_{m\geq 1}\mathbb{E}|y_{m,n}|^2 < +\infty$, the family $(y_{m,n})_{m=1,\ldots,M}$ verifies $y_{m,n} \prec 1$. Therefore, it holds that

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}|y_{m,n}| \prec \sqrt{N} \tag{4.47}$$

For the same reasons, the family $\xi_{y_m}(\nu), m = 1, \dots, M, \nu \in [0,1]$ satisfies.

$$\xi_{y_m}(\nu) \prec 1 \tag{4.48}$$

We also claim that

$$\sup_{\nu \in [0,1]} |\xi_{y_m}(\nu)| \prec 1 \tag{4.49}$$

In order to verify (4.49), we first observe that for any $n \geq 1$, we have the following control:

$$|\exp -2i\pi n\nu - \exp -2i\pi n(\nu+\delta)| \leq 2|\sin \pi n\delta| \leq 2\pi n|\delta|$$

(4.47) implies that

$$\begin{aligned} \sup_{\delta \neq 0} \sup_{\nu \in [0,1]} & \left| \frac{\xi_{y_m}(\nu) - \xi_{y_m}(\nu+\delta)}{\delta} \right| \\ & = \sup_{\delta \neq 0} \sup_{\nu \in [0,1]} \frac{1}{\sqrt{N}} \left| \sum_{n=1}^{N} y_{m,n} \frac{e^{-2i\pi n\nu} - e^{-2i\pi n(\nu+\delta)}}{\delta} \right| \\ & \leq 2\pi N \frac{1}{\sqrt{N}} \sum_{n=1}^{N} |y_{m,n}| \\ & \prec N^{3/2} \end{aligned} \tag{4.50}$$

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 $\epsilon > 0$, then it 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 the set

$$\mathcal{V}_N^p = \left\{ \frac{k}{N^p} : k \in \{0, \dots, N^p - 1\} \right\} \tag{4.51}$$

which cardinal is $|\mathcal{V}_N^p| = N^p$. We notice that (4.48) 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

$$\begin{aligned} & \mathbb{P}(|\xi_{y_m}(\nu_*)| > N^{\epsilon}) \\ & \quad \leq \mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^{\epsilon}}{2} \right) + \mathbb{P}\left(|\xi_{y_m}(\nu_{*,p})| > \frac{N^{\epsilon}}{2} \right) \\ & \quad \leq \mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^{\epsilon}}{2} \right) + \mathbb{P}\left(\sup_{\nu_p \in \mathcal{V}_N^p} |\xi_{y_m}(\nu_p)| > \frac{N^{\epsilon}}{2} \right) \end{aligned} \tag{4.52}$$

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 (4.52) converges exponentially towards 0. In order to evaluate the first term of the r.h.s. of (4.52), we use (4.50), and obtain that

$$\begin{aligned}\mathbb{P}\left(|\xi_{y_m}(\nu_*) - \xi_{y_m}(\nu_{*,p})| > \frac{N^\epsilon}{2}\right) &\leq \mathbb{P}\left[N \frac{1}{\sqrt{N}} \sum_{n=1}^N |y_{m,n}| \geq \frac{\pi}{2|\nu_* - \nu_{*,p}|} N^\epsilon\right] \\ &\leq \mathbb{P}\left[\frac{1}{\sqrt{N}} \sum_{n=1}^N |y_{m,n}| \geq \frac{\pi}{2} N^{p+\epsilon-1}\right]\end{aligned}$$

We choose p so that $p - 1 > 3/2$, and use (4.47) 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 (4.49).

In order to complete the proof of Proposition 4.4, we consider an individual entry $\hat{\mathbf{S}}_{ij}(\nu)$ of $\hat{\mathbf{S}}(\nu)$ for $i, j \leq M$, and write that

$$\begin{aligned}&\left|\hat{\mathbf{S}}_{ij}(\nu) - \hat{\mathbf{S}}_{ij}(\nu+\delta)\right| \\ &\quad = \frac{1}{B+1}\left|\sum_{b=-B/2}^{B/2} \xi_i\left(\nu + \frac{b}{N}\right)\xi_j\left(\nu + \frac{b}{N}\right)^* \right.\\ &\qquad\qquad \left. - \xi_i\left(\nu + \delta + \frac{b}{N}\right)\xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right| \\ &\quad \leq \frac{1}{B+1}\sum_{b=-B/2}^{B/2}\left|\xi_i\left(\nu + \frac{b}{N}\right)\left(\xi_j\left(\nu + \frac{b}{N}\right)^* - \xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right)\right| \\ &\qquad\qquad + \left|\left(\xi_i\left(\nu + \frac{b}{N}\right) - \xi_i\left(\nu + \delta + \frac{b}{N}\right)\right)\xi_j\left(\nu + \delta + \frac{b}{N}\right)^*\right|\end{aligned}$$

Using the estimations (4.49) and (4.50), we get:

$$\sup_{i,j}\sup_{\delta\neq 0}\sup_{\nu\in[0,1]}\left|\frac{\hat{\mathbf{S}}_{ij}(\nu) - \hat{\mathbf{S}}_{ij}(\nu+\delta)}{\delta}\right| \prec N^{3/2} \tag{4.53}$$

and deduce (4.46) from the rough bound

$$\begin{aligned}\sup_{\nu\in[0,1]}\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu+\delta)\| &\leq \sup_{\nu\in[0,1]}\sup_i \sum_j |\hat{\mathbf{S}}_{ij}(\nu) - \hat{\mathbf{S}}_{ij}(\nu+\delta)| \\ &\leq M \sup_{\nu\in[0,1]}\sup_{i,j}|\hat{\mathbf{S}}_{ij}(\nu) - \hat{\mathbf{S}}_{ij}(\nu+\delta)|\end{aligned}$$

□

Combining the eigenvalue localisation result from Corollary 3.2 and the Lipschitz behaviour of $\hat{\mathbf{S}}$ from Proposition 4.4, the following statement holds.

Corollary 4.1. *(ν uniform version of Corollary 3.2.) Denote for $\epsilon > 0$:*

$$\Lambda_\epsilon^{\hat{\mathbf{S}}} = \left\{\forall \nu \in [0,1] : \sigma(\hat{\mathbf{S}}(\nu)) \subset \operatorname{Supp} \mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon\right\}$$
$$\Lambda_\epsilon^{\hat{\mathbf{D}}} = \left\{\forall \nu \in [0,1] : \sigma(\hat{\mathbf{D}}(\nu)) \subset [\underline{s}, \bar{s}] + \epsilon\right\}$$

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 3.2 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 (4.51) and denote by $\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}}$ the event defined by

$$\Lambda_{\epsilon,p}^{\hat{\mathbf{S}}} = \left\{\forall \nu_p \in \mathcal{V}_N^p : \sigma(\hat{\mathbf{S}}(\nu_p)) \subset \operatorname{Supp} \mu_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon\right\}$$

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:

$$\left(\Lambda_\epsilon^{\hat{\mathbf{S}}}\right)^c \subset \left(\Lambda_{\epsilon/2,p}^{\hat{\mathbf{S}}}\right)^c \cup \left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > \epsilon/2 \text{ where } \nu_p^* \in \operatorname*{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\} \tag{4.54}$$

Suppose that $(\Lambda_\epsilon^{\hat{\mathbf{S}}})^c$ is realized, and denote by $\nu^* \in [0,1]$ a frequency such that $\sigma(\hat{\mathbf{S}})(\nu^*) \not\subset \operatorname{Supp}_{MP}^{(c)} \times [\underline{s}, \bar{s}] + \epsilon$. Denote also $\nu_p^* \in \operatorname{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu_p - \nu^*|$. 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 $\left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\| > \epsilon/2 \text{ where } \nu_p^* \in \operatorname{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\}$ is realized

(4.54) is now proved.

We already showed that $\left(\Lambda^{\hat{\mathbf{S}}}_{\epsilon/2,p}\right)^c$ holds with exponentially small probability, and establish now that the set

$$\left\{\exists \nu \in [0,1] : \|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu)\| > \epsilon/2 \text{ where } \nu_p^* \in \operatorname*{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\}$$

has the same property. To justify this claim, we remark that Proposition 4.4 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

$$\left\{\exists \nu \in [0,1],\ \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu_p^*)\|}{|\nu - \nu_p^*|}, > N^\kappa M N^{3/2}, \text{ where } \nu_p^* \in \operatorname*{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\}$$
$$\subset \left\{\exists \nu, \nu' \in [0,1],\ \frac{\|\hat{\mathbf{S}}(\nu) - \hat{\mathbf{S}}(\nu')\|}{|\nu - \nu'|} > N^\kappa M N^{3/2}\right\}$$

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 eventually become smaller than $\epsilon/2$. This proves that

$$\left\{\exists \nu \in [0,1],\ \|\hat{\mathbf{S}}(\nu_p^*) - \hat{\mathbf{S}}(\nu)\| > \epsilon/2 \text{ where } \nu_p^* \in \operatorname*{argmin}_{\nu_p \in \mathcal{V}_N^p} |\nu - \nu_p|\right\}$$

holds with exponentially small probability.

The same argument can be used in order to control $\Lambda^{\hat{\mathbf{D}}}_\epsilon$. This completes the proof of Corollary 4.1. □

We deduce immediately from Corollary 4.1 the following result that can be seen as a refinement of (3.22) and of Lemma 3.1.

Corollary 4.2. *It holds that*

$$\sup_{\nu \in [0,1]} \|\hat{\mathbf{D}}(\nu)^{-1/2}\| \prec 1, \quad \sup_{\nu \in [0,1]} \|\hat{\mathbf{S}}(\nu)\| \prec 1$$

A useful consequence of this is the following corollary, which states that the Lipschitz result holds for $\hat{\mathbf{C}}(\nu)$.

Corollary 4.3. *It holds that*

$$\sup_{\delta\neq 0}\sup_{\nu\in[0,1]}\left\|\frac{\hat{\mathbf{C}}(\nu)-\hat{\mathbf{C}}(\nu+\delta)}{\delta}\right\|\prec MN^{3/2} \tag{4.55}$$

Proof. For more clarity in the following argument, denote $\nu_1=\nu$ and $\nu_2=\nu+\delta$. Recall that $\hat{\mathbf{D}}=\mathrm{diag}\hat{\mathbf{S}}$. Using the definition of $\hat{\mathbf{C}}$ from equation (1.3), we write:

$$\begin{aligned}\hat{\mathbf{C}}(\nu_2)-\hat{\mathbf{C}}(\nu_1)&=\hat{\mathbf{D}}^{-1/2}(\nu_2)\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{D}}^{-1/2}(\nu_1)\hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1)\\&=(\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{D}}^{-1/2}(\nu_1))\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2)\\&\quad+\hat{\mathbf{D}}^{-1/2}(\nu_1)(\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1))\end{aligned}$$

Moreover, we write that

$$\begin{aligned}&\hat{\mathbf{S}}(\nu_2)\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{S}}(\nu_1)\hat{\mathbf{D}}^{-1/2}(\nu_1)\\&\quad=(\hat{\mathbf{S}}(\nu_2)-\hat{\mathbf{S}}(\nu_1))\hat{\mathbf{D}}^{-1/2}(\nu_2)+\hat{\mathbf{S}}(\nu_1)(\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{D}}^{-1/2}(\nu_1))\end{aligned}$$

Therefore, applying the operator norm, we get by the triangle inequality:

$$\begin{aligned}\|\hat{\mathbf{C}}(\nu_2)-\hat{\mathbf{C}}(\nu_1)\|&\leq\|\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_2)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2)\|\\&\quad+\|\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_2)-\hat{\mathbf{S}}(\nu_1)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2)\|\\&\quad+\|\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\|\hat{\mathbf{S}}(\nu_1)\|\|\hat{\mathbf{D}}^{-1/2}(\nu_2)-\hat{\mathbf{D}}^{-1/2}(\nu_1)\|\end{aligned}$$

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 4.4 and Corollary 4.2 immediately imply (4.55). □

Finally, we can write for the spectrum of $\hat{\mathbf{C}}$ the same kind of result as in Corollary 4.1.

Corollary 4.4. *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 4.1 and is thus omitted. □

4.3. Step 3: Extension to $\nu \in [0,1]$

We are now in position to establish Theorem 4.1.

Proof. **Proof of Theorem 4.1.** We recall that (4.2) holds. We consider again the set $\mathcal{V}_N^p$ defined by (4.51), and obtain from (4.2) that the following stochastic domination property

$$\sup_{\nu \in \mathcal{V}_N^p} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)} \right| \prec \frac{B}{N} \tag{4.56}$$

holds.

It remains to extend this result to the supremum over $\nu \in [0,1]$. We consider $\kappa > 0$ and evaluate

$$\mathbb{P}\left[\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)} \right| > N^\kappa \frac{B}{N} \right]$$

We denote by $\nu^* \in [0,1]$ an element where the supremum is achieved, and consider ν_p^* the the closest element of $\mathcal{V}_N^p$ from ν^*. Therefore, one can write:

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\nu \in [0,1]} \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)} \right| > N^\kappa \frac{B}{N} \right] \\
&\quad \leq \mathbb{P}\Bigg[\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu^*)) - \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*)) \right| \\
&\qquad\qquad + \left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)} \right| > N^\kappa \frac{B}{N} \Bigg] \\
&\quad \leq \mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu^*)) - \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*)) \right| > N^\kappa \frac{B}{2N} \right] \\
&\qquad\qquad + \mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)} \right| > N^\kappa \frac{B}{2N} \right]
\end{aligned}$$

$\mathbb{P}\left[\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*)) - \int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)}\right| > N^\kappa \frac{B}{2N}\right]$ converges exponentially towards 0 by (4.56). It thus remains to study $\mathbb{P}\left[\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu^*)) - \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*))\right| > N^\kappa \frac{B}{2N}\right]$. For this, we will of course use Corollary 4.3, and write that

$$\begin{aligned}
&\mathbb{P}\left[\left| \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu^*)) - \frac{1}{M} \mathrm{tr}\ f(\hat{\mathbf{C}}(\nu_p^*)) \right| > N^\kappa \frac{B}{2N} \right] \\
&\qquad = \mathbb{P}\left[\left| \frac{1}{M} \sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu^*))) - f(\lambda_m(\hat{\mathbf{C}}(\nu_p^*))) \right| > N^\kappa \frac{B}{2N} \right]
\end{aligned}$$

By conditioning on the event $\Lambda_\epsilon^{\hat{\mathbf{C}}}$ which by Corollary 4.4 holds with exponentially high probability, we get:

$$\mathbb{P}\left[\frac{1}{M}\left|\sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu^*))) - f(\lambda_m(\hat{\mathbf{C}}(\nu_p^*)))\right| > N^\kappa \frac{B}{2N}\right]$$
$$\leq \mathbb{P}\left[\left|\frac{1}{M}\sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu^*))) - f(\lambda_m(\hat{\mathbf{C}}(\nu_p^*)))\right| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right] + \mathbb{P}\left[\left(\Lambda_\epsilon^{\hat{\mathbf{C}}}\right)^c\right] \tag{4.57}$$

As $\mathbb{P}\left[\left(\Lambda_\epsilon^{\hat{\mathbf{C}}}\right)^c\right]$ holds with exponentially low probability, it remains to study the first term of the right-hand side of (4.57). Since f is differentiable on a neighborhood of $\mathrm{Supp}_{MP}^{(c)}$, there exist some random quantities $(\tilde{\nu}_m)_{1\leq m\leq M}$ between ν^* and ν_p^* such that:

$$\mathbb{P}\left[\left|\frac{1}{M}\sum_{m=1}^{M} f(\lambda_m(\hat{\mathbf{C}}(\nu^*))) - f(\lambda_m(\hat{\mathbf{C}}(\nu_p^*)))\right| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$
$$\leq \mathbb{P}\left[\frac{1}{M}\sum_{m=1}^{M}\left|f'(\lambda_m(\hat{\mathbf{C}}(\tilde{\nu}_m)))\right|\left|\lambda_m(\hat{\mathbf{C}}(\nu^*)) - \lambda_m(\hat{\mathbf{C}}(\nu_p^*))\right| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$

Using the following eigenvalue inequality for Hermitian matrices:

$$\left|\lambda_m(\hat{\mathbf{C}}(\nu^*)) - \lambda_m(\hat{\mathbf{C}}(\nu_p^*))\right| \leq \|\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)\|$$

in conjunction with the fact that $\sup_{1\leq m\leq M}|f'(\lambda_m(\hat{\mathbf{C}}(\tilde{\nu}_m)))|$ is bounded by some nice constant C on the event $\Lambda_\epsilon^{\hat{\mathbf{C}}}$, we get:

$$\mathbb{P}\left[\frac{1}{M}\sum_{m=1}^{M}\left|f'(\lambda_m(\hat{\mathbf{C}}(\tilde{\nu}_m)))\right|\left|\lambda_m(\hat{\mathbf{C}}(\nu^*)) - \lambda_m(\hat{\mathbf{C}}(\nu_p^*))\right| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$
$$\leq \mathbb{P}\left[C\|\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)\| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$

Recall that $|\nu^* - \nu_p^*| < N^{-p}$. Therefore, we have:

$$\mathbb{P}\left[C\|\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)\| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$
$$= \mathbb{P}\left[\left\|\frac{\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)}{\nu^* - \nu_p^*}\right\| > C\frac{1}{|\nu^* - \nu_p^*|}N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$
$$\leq \mathbb{P}\left[\left\|\frac{\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)}{\nu^* - \nu_p^*}\right\| > CN^p N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$$

We choose p large enough such that $MN^{3/2} \ll N^p \frac{B}{N}$. Then, it is clear that

$$\begin{aligned}\mathbb{P}\left[\left\|\frac{\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)}{\nu^* - \nu_p^*}\right\| > CN^p N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right] \\ \leq \mathbb{P}\left[\left\|\frac{\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p^*)}{\nu^* - \nu_p^*}\right\| > MN^{3/2}N^\kappa,\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right] \\ \leq \mathbb{P}\left[\exists \nu, \nu' \in [0,1],\ \left\|\frac{\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu')}{\nu - \nu'}\right\| > MN^{3/2}N^\kappa\right]\end{aligned}$$

Corollary 4.3 thus implies that $\mathbb{P}\left[C\|\hat{\mathbf{C}}(\nu^*) - \hat{\mathbf{C}}(\nu_p)\| > N^\kappa \frac{B}{2N},\ \Lambda_\epsilon^{\hat{\mathbf{C}}}\right]$ converges towards 0 exponentially fast. This completes the proof of Theorem 4.1.

□

5. Applications and numerical simulation

5.1. Definition of the test statistics

In order to test the decorrelation of the signals $(y_m)_{m=1,\ldots,M}$, we consider for $\epsilon > 0$ the statistics T_ϵ defined by

$$T_\epsilon = \sup_{\nu \in [0,1]} \frac{\left|\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\, \mathrm{d}\mu_{MP}^{(c_N)}\right|}{N^\epsilon (B/N)} \tag{5.1}$$

and compare T_ϵ to a certain threshold $\kappa > 0$. Hypothesis $\mathcal{H}_0$ is accepted if $T_\epsilon \leq \kappa$ and rejected otherwise. The test is consistent in the sense that under $\mathcal{H}_0$, Theorem 4.1 ensures that $T_\epsilon \to 0$ almost surely . Therefore,

$$\mathbb{P}_{\mathcal{H}_0}[T_\epsilon > \kappa] \xrightarrow[N\to\infty]{} 0$$

In the following numerical experiments, motivated by [25], we consider $f(\lambda) = \log \lambda$ for which

$$\int f(\lambda)\, \mathrm{d}\mu_{MP}^{(c_N)}(\lambda) = \frac{c_N - 1}{c_N} \log(1 - c_N) + 1$$

(see e.g. [1]) and $f(\lambda) = (\lambda - 1)^2$ where

$$\int f(\lambda)\, \mathrm{d}\mu_{MP}^{(c_N)}(\lambda) = c_N$$

We notice that $f(\lambda) = \log \lambda$ is not defined for $\lambda = 0$, but the results of this paper can still be used by defining $\mathrm{tr}\ \log A = \log \det A = -\infty$ if $A \geq 0$ and $\det A = 0$.

5.2. Signal model for the alternative

We consider the following simple and flexible model:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{A}\mathbf{x}_n + \mathbf{B}\boldsymbol{\epsilon}_n \\ \mathbf{y}_{n+1} &= \mathbf{C}\mathbf{x}_n + \mathbf{D}\boldsymbol{\epsilon}_n \end{aligned} \tag{5.2}$$

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 $\mathbf{B} = \mathbf{I}_M$, $\mathbf{C} = \mathbf{I}_M$, $\mathbf{A}$ is the bidiagonal lower triangular matrix defined by

$$\mathbf{A} = \begin{pmatrix} \theta & 0 & \dots & \dots & \dots & 0 \\ \beta & \theta & 0 & \dots & \dots & 0 \\ 0 & \beta & \theta & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \beta & \theta \end{pmatrix}$$

for $\theta \geq 0$, $\beta \geq 0$, $\theta + \beta < 1$, and $\mathbf{D}$ is the triangular matrix

$$\mathbf{D} = \delta \begin{pmatrix} \gamma^{M-1} & \gamma^{M-2} & \dots & \dots & \dots & 1 \\ 0 & \gamma^{M-1} & \gamma^{M-2} & \dots & \dots & \gamma \\ 0 & 0 & \gamma^{M-1} & \gamma^{M-2} & \dots & \gamma^2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \gamma^{M-2} \\ 0 & \dots & \dots & 0 & 0 & \gamma^{M-1} \end{pmatrix}$$

for $\delta \geq 0$ and $0 \leq \gamma < 1$. As $\theta + \beta < 1$, for each frequency $\nu \in [0, 1]$, matrix $e^{2i\pi\nu}\mathbf{I} - \mathbf{A}$ is invertible, and its inverse $(e^{2i\pi\nu}\mathbf{I} - \mathbf{A})^{-1}$ is the lower triangular matrix given by

$$(e^{2i\pi\nu}\mathbf{I} - \mathbf{A})^{-1} = \frac{1}{e^{2i\pi\nu} - \theta} \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \frac{\beta}{e^{2i\pi\nu}-\theta} & 1 & 0 & \dots & \dots & 0 \\ \left(\frac{\beta}{e^{2i\pi\nu}-\theta}\right)^2 & \frac{\beta}{e^{2i\pi\nu}-\theta} & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \left(\frac{\beta}{e^{2i\pi\nu}-\theta}\right)^{M-1} & \dots & \dots & \dots & \frac{\beta}{e^{2i\pi\nu}-\theta} & 1 \end{pmatrix}$$

Therefore, for each $m = 1, \ldots, M$, the multivariate signal $\mathbf{y}_n$ is given by

$$\mathbf{y}_n = \left[\frac{1}{e^{2i\pi\nu} - \theta}\right] \begin{pmatrix} \epsilon_{1,n} \\ \epsilon_{2,n} + \left[\frac{\beta}{e^{2i\pi\nu}-\theta}\right] \epsilon_{1,n} \\ \vdots \\ \vdots \\ \epsilon_{M,n} + \left[\frac{\beta}{e^{2i\pi\nu}-\theta}\right] \epsilon_{M-1,n} + \ldots + \left[\left(\frac{\beta}{e^{2i\pi\nu}-\theta}\right)^{M-1}\right] \epsilon_{1,n} \end{pmatrix} + \delta \begin{pmatrix} \gamma^{M-1}\epsilon_{1,n} + \ldots + \epsilon_{M,n} \\ \gamma^{M-1}\epsilon_{2,n} + \ldots + \gamma\epsilon_{M,n} \\ \vdots \\ \vdots \\ \gamma^{M-1}\epsilon_{M,n} \end{pmatrix}$$

where if $h(\nu) = \sum_{k\in\mathbb{Z}} h_k e^{-2i\pi k\nu}$ with $\sum_k |h_k|^2 < +\infty$ and $(u_n)_{n\in\mathbb{Z}}$ is an i.i.d. sequence, the notation $[h(\nu)]u_n$ represents $\sum_{k\in\mathbb{Z}} h_k u_{n-k}$. After some calculations, we obtain that the entry (k, l), $k \geq l$, of the spectral density $\mathbf{S}(\nu)$ of $\mathbf{y}$ is given by:

$$\begin{aligned} \mathbf{S}(\nu)_{k,l} = & \\ & \frac{1}{|e^{2i\pi\nu} - \theta|^2} \left(\frac{\beta}{e^{2i\pi\nu} - \theta}\right)^{k-l} \left(1 + \left|\frac{\beta}{e^{2i\pi\nu} - \theta}\right|^2 + \ldots + \left|\frac{\beta}{e^{2i\pi\nu} - \theta}\right|^{2(l-1)}\right) \\ & + \frac{\delta}{e^{2i\pi\nu} - \theta} \sum_{p=0}^{k-l} \gamma^{p+M-1-(k-l)} \left(\frac{\beta}{e^{2i\pi\nu} - \theta}\right)^p \\ & + \frac{\delta}{e^{-2i\pi\nu} - \theta} + \sum_{p=0}^{k-l} \gamma^{p+M-1-(k-l)} \left(\frac{\beta}{e^{-2i\pi\nu} - \theta}\right)^p \\ & + \delta^2 \gamma^{k+l-2} \left(1 + \gamma^2 + \ldots + \gamma^{2(M-k)}\right) \end{aligned}$$

It is clear that the signals $y_1, \ldots, y_M$ are independent if $\beta = \delta = 0$.

We denote by r the ratio defined by

$$r := \frac{\int \|\mathbf{S}_\mathbf{y}(\nu) - \mathrm{diag}\mathbf{S}_\mathbf{y}(\nu)\|_F^2 \,\mathrm{d}\nu}{\int \|\mathbf{S}_\mathbf{y}(\nu)\|_F^2 \,\mathrm{d}\nu} = \frac{\sum_{l\in\mathbb{Z}} \|\mathbf{R}(l) - \mathrm{diag}\mathbf{R}(l)\|_F^2}{\sum_{l\in\mathbb{Z}} \|\mathbf{R}(l)\|_F^2}$$

where $\mathbf{R}(l) := \mathbb{E}[\mathbf{y}_{n+l}\mathbf{y}_n^*]$ is the autocovariance matrix of $\mathbf{y}$ at lag l. We will choose β, δ and γ such that this ratio is constant for different settings in the simulation (e.g. various values of M or choice of f). Indeed, the "strength" of the joint dependence of the observations $\mathbf{y}$ depends on the dimension M.

Figure 1 and Figure 2 represent the value of the maximum deviation over ν of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ from $\int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$ for increasing values of N (the x-axis represents $B := B(N)$). More precisely, if we define

$$\nu^* = \underset{\nu\in[0,1]}{\mathrm{argmax}} \left| \frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu)) - \int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)} \right|$$

then, the value of the test statistics $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}}(\nu))$ is shown for the frequency ν^*. We compare this value under $\mathcal{H}_0$ ($\beta = 0$ and $\delta = 0$, whereas θ and γ are not precised), and under an alternative $\mathcal{H}_1$ (with the same θ as the one used for $\mathcal{H}_0$, but $\beta \neq 0$ and $\delta \neq 0$) such that $r > 0$. On the left of Figure 1 is shown these values with f representing the Frobenius norm test, and on the right of Figure 1 the log det test.

On Figure 1 we show that as r increases, the separation of the test statistics from $\int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$ also increases. For $r = 0$, ie. under $\mathcal{H}_0$, we see that the maximum deviation of $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})$ from $\int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$ seems to converge toward zero, as stated in Theorem 4.1.

On Figure 2, we show the impact of the rate α (recall its definition from Assumption 1.3). As α is close to 1, N becomes closer to B (the limit $\alpha = 1$ corresponds to B, M and N of the same order), so intuitively, the sample size is very limited. On the opposite, as α is close to $1/2$, N is much larger than B and M, which implies that the behaviour of $\hat{\mathbf{C}}$ under $\mathcal{H}_0$ should be better approximated by the Marchenko-Pastur distribution. This is confirmed in the simulations: as α grows, the separation between $\mathcal{H}_0$ and $\mathcal{H}_1$ becomes smaller.

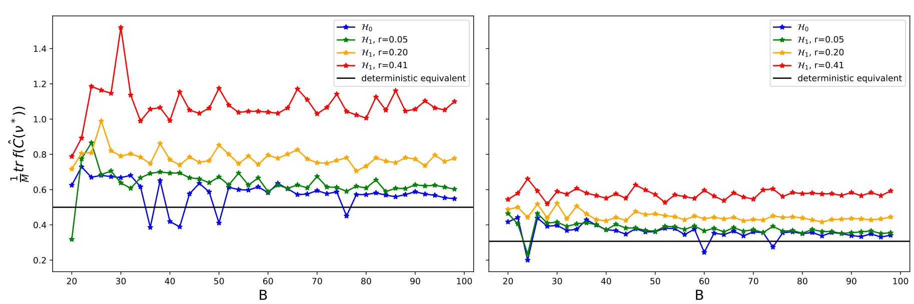

FIG 1. *maximum deviation of* $\frac{1}{M}\mathrm{tr}\ f(\hat{\mathbf{C}})$ *from* $\int_{\mathbb{R}} f\,\mathrm{d}\mu_{MP}^{(c_N)}$ *for various alternatives, with* $\alpha = 0.7$ *and* $c = 0.5$. *Left is the Frobenius Norm Test and right is the Logdet test.*

.

Figure 3 and Figure 4 represent numerical estimation of ROC curves : for various settings of α, θ, c and r, we simulate observations $(\mathbf{y}_n)_{n=1,\dots,N}$ using the state space model (5.2). The probability of false alarm and the probability

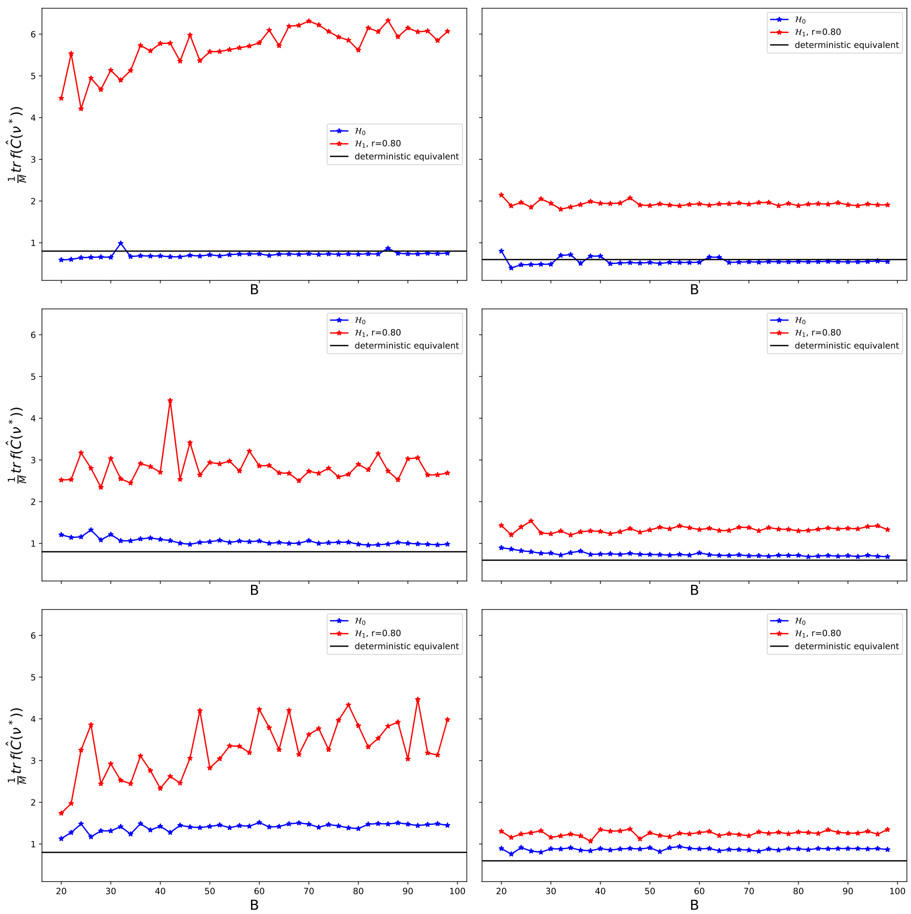

FIG 2. *maximum deviation of* $\frac{1}{M}\operatorname{tr} f(\hat{\mathbf{C}})$ *from* $\int_{\mathbb{R}} f \,\mathrm{d}\mu_{MP}^{(c_N)}$ *for* $c = 0.8$ *and various values of (from top to bottom)* $\alpha = 0.5$, $\alpha = 0.75$, $\alpha = 0.95$, *and from left to right for the Frobenius Norm Test and the Logdet test.*

of detection are estimated via Monte Carlo with 10.000 repetitions. Note that the axis scale is log-log.

On Figure 3 is shown the impact of the alternative factor r. As it increases, the detection becomes easier. On Figure 4 the impact of the rate α is shown. As it is close to $1/2$, N is much larger to M and B and the detection is easier. On Figure 5, the parameter δ is fixed to 0, and the impact of θ is shown. When $\theta = 0$, each time series $(y_{m,n})_{n \in \mathbb{Z}}$ is i.i.d. under the 2 hypotheses, or equivalently, the observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are i.i.d. In this rather simple context, the performance of the test appears of course better than if $\theta \neq 0$. As θ increases, the various time series become more dependent, and the performance decreases.

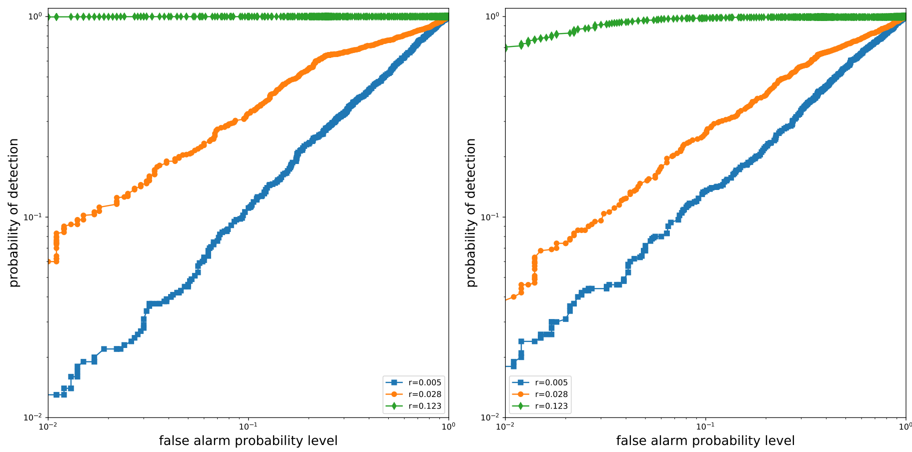

FIG 3. *ROC for various r, with $\alpha = 0.6$, $c = 0.8$ and $M = 33$. Left is Frobenius Norm Test and right is Logdet test.*

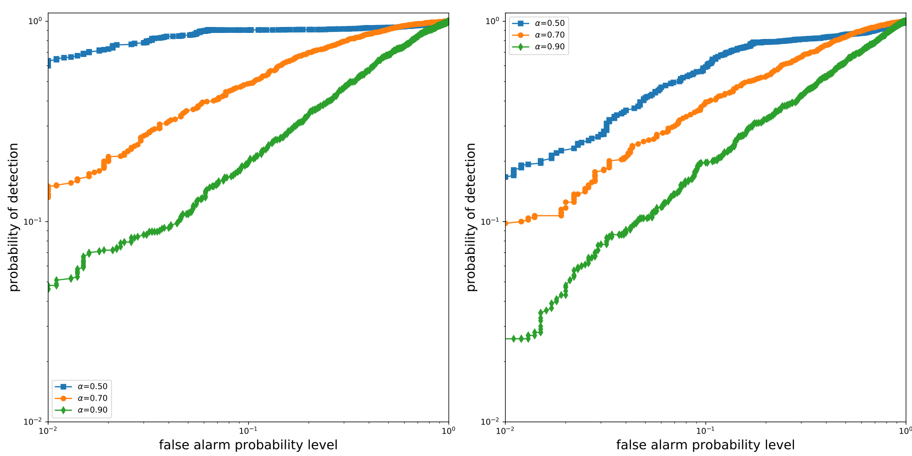

FIG 4. *ROC for various α, with $r = 0.06$, $c = 0.8$ and $M = 33$. Left is Frobenius Norm Test and right is Logdet test.*

We notice that the Frobenius test seems to outperform the log det test in all our simulations. This is in accordance with the conclusions of [25] devoted to the case where the observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are i.i.d.

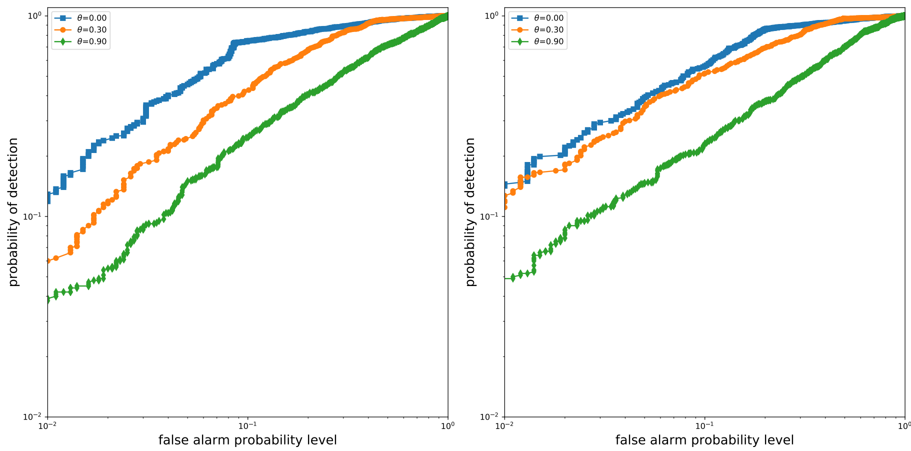

FIG 5. *ROC for various θ with $c = 0.8$, $r = 0.09$ and $M = 33$. Left is Frobenius Norm Test and right is Logdet test.*

Appendix A: Appendix

A.1. Proof of Lemma A.1

Lemma A.1 is a slight variation of Theorem 4.3.2 [4].

Lemma A.1. *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:*

$$\sup_{m\geq 1} |\mathbb{E}\left[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*\right] - s_m(\nu_1)\delta_{\nu_1=\nu_2}| = \mathcal{O}\left(\frac{1}{N}\right) \tag{A.1}$$

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

$$\mathbb{E}[\xi_{y_m}(\nu_1)\xi_{y_m}(\nu_2)^*] = \frac{1}{N} r_{m,0} \sum_{n_1=0}^{N-1} e^{-2i\pi n_1(\nu_2-\nu_1)} + \frac{1}{N}\sum_{u=1}^{(N-1)} r_{m,u} \sum_{n_2=0}^{N-1-u} e^{-2i\pi(u+n_2)\nu_1} e^{2i\pi n_2\nu_2} + \frac{1}{N}\sum_{u=-(N-1)}^{-1} r_{m,u} \sum_{n_2=-u}^{N-1} e^{-2i\pi(u+n_2)\nu_1} e^{2i\pi n_2\nu_2} \tag{A.2}$$

The first term of the right hand side of (A.2) 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 (A.2) is equal to $r_{m,0}\delta_{\nu_1=\nu_2}$.
Consider now the second term of (A.2) (where $u > 0$):

$$\frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} \sum_{n_2=0}^{N-1-u} e^{-2i\pi(u+n_2)\nu_1} e^{2i\pi n_2\nu_2} = \frac{1}{N}\sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u\nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2-\nu_1)} \tag{A.3}$$

The right hand side of (A.3) 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:

$$\sup_{m \geq 1} \left| \frac{1}{N} \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2 - \nu_1)} - \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \right| = \mathcal{O}\left(\frac{1}{N}\right) \tag{A.4}$$

In the case where $\nu_1 \neq \nu_2$, note that $\nu_1 - \nu_2 = k/N$ with $k \neq 0$, therefore:

$$\sum_{n_2=0}^{N-1} e^{-2i\pi n_2(\nu_2 - \nu_1)} = \sum_{n_2=0}^{N-1} e^{-2i\pi n_2 \frac{k}{N}} = 0 \tag{A.5}$$

Using (A.5), one can rewrite the right hand side of (A.3) as

$$\begin{aligned} &\left| \frac{1}{N} \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2 - \nu_1)} \right| \\ &\qquad = \left| -\frac{1}{N} \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=N-u}^{N} e^{-2i\pi n_2(\nu_2 - \nu_1)} \right| \\ &\qquad \leq \frac{1}{N} \sum_{u=1}^{N-1} |u||r_{m,u}| \end{aligned}$$

which, again by Assumption 1.4, provides the bound:

$$\sup_{m \geq 1} \left| \frac{1}{N} \sum_{u=1}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \sum_{n_2=0}^{N-1-u} e^{-2i\pi n_2(\nu_2 - \nu_1)} \right| = \mathcal{O}\left(\frac{1}{N}\right) \tag{A.6}$$

Combining (A.4) and (A.6), the second term of the right hand side of (A.2) 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 (A.2) is similar. Gathering the three terms of equation (A.2) leads to

$$\sup_{m \geq 1} \left| \mathbb{E}[\xi_{y_m}(\nu_1) \xi_{y_m}(\nu_2)^*] - \delta_{\nu_1 = \nu_2} \left(\sum_{u=-(N-1)}^{N-1} r_{m,u} e^{-2i\pi u \nu_1} \right) \right| = \mathcal{O}\left(\frac{1}{N}\right) \tag{A.7}$$

Eventually, 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 (A.7), we obtain equation (A.1) □

A.2. Proof of Lemma 3.1

Proof. Consider the complement of the event $\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)$ and notice that:

$$\Lambda_\epsilon^{\hat{\mathbf{D}}}(\nu)^c \subset \{\exists m \in \{1, \ldots, M\} : \hat{s}_m > \bar{s} + \epsilon\} \cup \{\exists m \in \{1, \ldots, M\} : \hat{s}_m < \underline{s} - \epsilon\} \tag{A.8}$$

We start by proving that the first set of the right handside of (A.8) holds with is 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 A.2 (see below), $|\mathbb{E}\hat{s}_m - s_m| = \mathcal{O}(B^2/N^2)$ so for N large enough, this biais 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 A.3 (see below). The proof for the lower bound is similar. □

It remains to prove Lemma A.2 and Lemma A.3. Concerning the proof of Lemma A.2, we follow the same approach as the one used in Theorem 5.4.2 in [4].

Lemma A.2. *For any $\nu \in [0, 1]$, the following bound holds:*

$$\sup_{m=1,\ldots,M} |\mathbb{E}\hat{s}_m(\nu) - s_m(\nu)| = \mathcal{O}\left(\frac{B^2}{N^2}\right) \tag{A.9}$$

Proof. Let $\nu \in [0,1]$. Inserting $s_m(\nu + \frac{b}{N})$ in (A.9), one can write:

$$
\begin{aligned}
&|\mathbb{E}\hat{s}_m(\nu) - s_m(\nu)| \\
&\quad = \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ \mathbb{E} \left| \xi_{y_m}\left(\nu + \frac{b}{N}\right)\right|^2 - s_m(\nu) \right\} \right| \\
&\quad = \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ \mathbb{E} \left| \xi_{y_m}\left(\nu + \frac{b}{N}\right)\right|^2 - s_m\left(\nu + \frac{b}{N}\right) \right.\right. \\
&\qquad \left.\left. + s_m\left(\nu + \frac{b}{N}\right) - s_m(\nu) \right\} \right| \\
&\quad \leq \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ \mathbb{E} \left| \xi_{y_m}\left(\nu + \frac{b}{N}\right)\right|^2 - s_m\left(\nu + \frac{b}{N}\right) \right\} \right| \\
&\qquad + \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ s_m\left(\nu + \frac{b}{N}\right) - s_m(\nu) \right\} \right|
\end{aligned}
$$

Lemma A.1 provides the following control for the first term of the right-hand side:

$$
\left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ \mathbb{E} \left| \xi_{y_m}\left(\nu + \frac{b}{N}\right)\right|^2 - s_m\left(\nu + \frac{b}{N}\right) \right\} \right| = \mathcal{O}\left(\frac{1}{N}\right) \tag{A.10}
$$

Moreover, by Assumption 1.4, a Taylor-Lagrange expansion of s_m around $\nu + \frac{b}{N}$, provides the existence of a quantity ν_b such that:

$$
s_m\left(\nu + \frac{b}{N}\right) = s_m(\nu) + \frac{b}{N} s_m'(\nu) + \frac{1}{2}\frac{b^2}{N^2} s_m''(\nu_b)
$$

where by Assumption 1.4, $\sup_{m \geq 1} \sup_{\nu \in [0,1]} |s_m''(\nu)| < +\infty$. Therefore, it holds that:

$$
\begin{aligned}
&\left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ s_m\left(\nu + \frac{b}{N}\right) - s_m(\nu) \right\} \right| \\
&\quad = \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} \left\{ \frac{b}{N} s_m'(\nu) + \frac{1}{2}\frac{b^2}{N^2} s_m''(\nu_b) \right\} \right| \\
&\quad \leq \frac{1}{N} |s_m'(\nu)| \left| \frac{1}{B+1} \underbrace{\sum_{b=-B/2}^{B/2} b}_{=0} \right| + \frac{1}{2}\frac{1}{N^2} \left| \frac{1}{B+1} \sum_{b=-B/2}^{B/2} b^2 s_m''(\nu_b) \right| \\
&\quad = \mathcal{O}\left(\frac{B^2}{N^2}\right)
\end{aligned} \tag{A.11}
$$

Combining the estimations (A.10) and (A.11), one get:

$$|\mathbb{E}\hat{s}_m(\nu) - s_m(\nu)| = \mathcal{O}\left(\frac{1}{N}\right) + \mathcal{O}\left(\frac{B^2}{N^2}\right)$$

which is the desired result since $\frac{1}{N} \ll \frac{B^2}{N^2}$ in our asymptotic regime defined in Assumption 1.3. □

Lemma A.3. *The family of random variables* $\sup_{m=1,\dots,M} |\hat{s}_m(\nu) - \mathbb{E}[\hat{s}_m(\nu)]|, \nu \in [0,1]$ *verifies*

$$\sup_{m=1,\dots,M} |\hat{s}_m - \mathbb{E}[\hat{s}_m]| \prec \frac{1}{\sqrt{B}} \tag{A.12}$$

Proof. Let $m \in \{1,\dots,M\}$ and matrix $\mathbf{\Pi}_N$ defined as:

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

where $\mathbf{a}_N = \frac{1}{\sqrt{N}}[1, e^{2i\pi\nu}, \dots, e^{2i\pi\nu(N-1)}]^T$. We also recall that $\mathbf{y}_m$ is defined by $\mathbf{y}_m = (y_{m,1}, \dots, \mathbf{y}_{m,N})^T$. $\hat{s}_m$ can be written in a convenient way we get as

$$\hat{s}_m = \mathbf{y}_m^* \mathbf{\Pi}_N \mathbf{y}_m \tag{A.13}$$

Note that $(B+1)\mathbf{\Pi}_N$ is an orthonormal projection matrix on a $B+1$–dimensional subspace. Its operator norm is therefore 1, which leads to the following equalities:

$$\|\mathbf{\Pi}_N\| = \frac{1}{B+1}, \quad \text{tr } \mathbf{\Pi}_N^2 = \frac{1}{(B+1)}$$

If $\mathbf{R}_m$ represents the covariance matrix of $\mathbf{y}_m$, $\mathbf{y}_m$ can be written as $\mathbf{y}_m = \mathbf{R}_m^{1/2}\mathbf{z}_m$ with $\mathbf{z}_m \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$. Therefore, $\hat{s}_m$ is given by $\hat{s}_m = \mathbf{z}_m^*\mathbf{R}_m^{1/2}\mathbf{\Pi}_N\mathbf{R}_m^{1/2}\mathbf{z}_m$. By Assumption 1.4, $\sup_{m\geq 1} \|\mathbf{R}_m\| < +\infty$, and using the inequality tr $(\mathbf{AB}) \leq \|\mathbf{B}\|$tr $(\mathbf{A})$ for $\mathbf{A}$ a positive semi-definite matrix and $\mathbf{B}$ Hermitian, it holds that:

$$\begin{aligned}
\sup_{m\geq 1} \|\mathbf{R}_m^{1/2}\mathbf{\Pi}_N\mathbf{R}_m^{1/2}\|_F^2 &= \sup_{m\geq 1} \text{tr } \mathbf{R}_m^{1/2}\mathbf{\Pi}_N\mathbf{R}_m\mathbf{\Pi}_N\mathbf{R}_m^{1/2} \\
&= \sup_{m\geq 1} \text{tr } \mathbf{R}_m\mathbf{\Pi}_N\mathbf{R}_m\mathbf{\Pi}_N \\
&\leq \sup_{m\geq 1} \|\mathbf{R}_m\|^2 \text{tr } \mathbf{\Pi}_N\mathbf{\Pi}_N \\
&= \mathcal{O}\left(\frac{1}{B}\right)
\end{aligned}$$

The Hanson-Wright inequality (2.9) provides:

$$\sup_{m\geq 1} |\hat{s}_m - \mathbb{E}\hat{s}_m| \prec \frac{1}{\sqrt{B}}$$

□

A.3. Proof of Lemma 3.3

Proof. These estimates can be proved in a compact way by using the calculus rules available in the stochastic domination framework introduced in Definition 2.1 and proved in Lemma 2.1. Using Lemma 3.2 and Lemma A.4 (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 A.4. *The family of random variables* $(\sup_{m=1,\dots,M}|\hat{s}_m(\nu)-s_m(\nu)|)$, $\nu\in[0,1]$ *verifies*

$$\sup_{m=1,\dots,M}|\hat{s}_m-s_m| \prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}$$

Proof. It is sufficient to check that the family of random variables $(|\hat{s}_m - s_m|)_{m=1,\dots,M}, \nu\in[0,1]$ verifies $|\hat{s}_m-s_m| \prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}$. Using Lemma A.2 and Lemma A.3, we obtain as expected that

$$|\hat{s}_m-s_m| = |s_m-\mathbb{E}\hat{s}_m+\mathbb{E}\hat{s}_m-\hat{s}_m| \leq \underbrace{|s_m-\mathbb{E}\hat{s}_m|}_{\mathcal{O}(\frac{B^2}{N^2})}+\underbrace{|\mathbb{E}\hat{s}_m-\hat{s}_m|}_{\mathcal{O}_{\prec}(\frac{1}{\sqrt{B}})} \prec \frac{1}{\sqrt{B}}+\frac{B^2}{N^2}$$

□

A.4. Proof of Lemma A.5

Lemma A.5. *The set of random variable* $(\sum_{m=1}^{M}|\hat{s}_m(\nu)-s_m(\nu)|^2)$, $\nu\in[0,1]$ *verifies*

$$\sum_{m=1}^{M}|\hat{s}_m-s_m|^2 \prec 1+\frac{B^5}{N^4}$$

Proof. Using Lemma A.4, we have

$$|\hat{s}_m - s_m|^2 \prec \frac{1}{B} + \frac{B^4}{N^4}$$

and summing over $m = 1 \ldots M$, one immediately get:

$$\sum_{m=1}^{M} |\hat{s}_m - s_m|^2 \prec 1 + \frac{B^5}{N^4}$$

□

References

[1] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, vol. 20, Springer, 2010.

[2] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for wigner matrices*, arXiv preprint arXiv:1601.04055, (2016).

[3] H. Böhm and R. von Sachs, *Shrinkage estimation in the frequency domain of multivariate time series*, Journal of Multivariate Analysis, 100 (2009), pp. 913–935.

[4] D. R. Brillinger, *Time series: data analysis and theory*, vol. 36, Siam, 1981.

[5] X. Chen and W. Liu, *Testing independence with high-dimensional correlated samples*, Ann. Stat., 46(2) (2018), pp. 866–895.

[6] H. Dette and N. Dörnemann, *Likelihood ratio tests for many groups in high dimensions*, Journal of Multivariate Analysis, (2020), p. 104605.

[7] P. Duchesne and R. Roy, *Robust tests for independence of two time series*, Statistica Sinica, (2003), pp. 827–852.

[8] E. M. Dyn'kin, *An operator calculus based on the cauchy–green formula*, Zapiski Nauchnykh Seminarov POMI, 30 (1972), pp. 33–39.

[9] M. Eichler, *A frequency-domain based test for non-correlation between stationary time series*, Metrika, 65 (2007), pp. 133–157.

[10] ——, *Testing nonparametric and semiparametric hypotheses in vector stationary processes*, Journal of Multivariate Analysis, 99 (2008), pp. 968–1009.

[11] K. El Himdi, R. Roy, and P. Duchesne, *Tests for non-norrelation of two multivariate time series: A nonparametric approach*, Lecture Notes-Monograph Series, vol. 42, (2003), pp. 397–416.

[12] L. Erdos, A. Knowles, and H.-T. Yau, *Averaging fluctuations in resolvents of random band matrices*, in Annales Henri Poincaré, vol. 14, Springer, 2013, pp. 1837–1926.

[13] U. Haagerup and S. Thorbjornsen, *Random matrices with complex gaussian entries*, Expositiones Mathematicae, 21 (2003), pp. 293–337.

[14] ——, *A new application of random matrices: Is not a group*, Annals of Mathematics, (2005), pp. 711–775.

[15] W. Hachem, O. Kohrunzhiy, P. Loubaton, J. Najim, and L. Pastur, *A new approach for capacity analysis of large dimensional multi-antenna channels*, IEEE Transactions on Information Theory, 54 (2008), pp. 3987–4003.

[16] L. D. Haugh, *Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach*, Journal of the American Statistical Association, 71 (1976), pp. 378–385.

[17] B. Helffer and J. Sjostrand, *Equation de schrödinger avec champ magnétique et équation de harper*, in Schrödinger operators, Springer, 1989, pp. 118–197.

[18] K. E. Himdi and R. Roy, *Tests for noncorrelation of two multivariate arma time series*, Canadian Journal of Statistics, 25 (1997), pp. 233–256.

[19] Y. Hong, *Testing for independence between two covariance stationary time series*, Biometrika, 83 (1996), pp. 615–625.

[20] T. Jiang, *The asymptotic distributions of the largest entries of sample correlation matrices*, Ann. Appl. Probab., 14(2) (2004), pp. 865–880.

[21] ——, *The limiting distributions of eigenvalues of sample correlation matrices*, Sankhya: The Indian Journal of Statistics, 66 (2004), pp. 35–48.

[22] T. Jiang and F. Yang, *Central limit theorems for classical likelihood ratio tests for high-dimensional distributions*, Ann. of Stats., 41 (2013), pp. 2029–2074.

[23] P. D. Koch and S.-S. Yang, *A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function*, Journal of the American Statistical Association, 81 (1986), pp. 533–544.

[24] W. Li and Y. Hui, *Robust residual cross correlation tests for lagged relations in time series*, Journal of Statistical Computation and Simulation, 49 (1994), pp. 103–109.

[25] X. Mestre and P. Vallet, *Correlation tests and linear spectral statistics of the sample correlation matrix*, IEEE Transactions on Information Theory, 63 (2017), pp. 4585–4618.

[26] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, Mathematical Surveys and Monographs, AMS, 2011.

[27] M. Rudelson, R. Vershynin, et al., *Hanson-wright inequality and sub-gaussian concentration*, Electronic Communications in Probability, 18 (2013).

[28] M. Taniguchi, M. L. Puri, and M. Kondo, *Nonparametric approach for non-gaussian vector stationary processes*, Journal of Multivariate Analysis, 56 (1996), pp. 259–283.

[29] T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, 132 (2011).

[30] G. Wahba, *Some tests of independence for stationary multivariate time series*, Journal of the Royal Statistical Society: Series B (Methodological), 33 (1971), pp. 153–166.