On the asymptotic behavior of the eigenvalue distribution of block correlation matrices of high-dimensional time series

Philippe Loubaton

Laboratoire d'Informatique Gaspard Monge, UMR 8049, Université Paris-Est Marne la Vallée
5 Bd. Descartes, Cite Descartes, 77454 Marne la Vallée Cedex 2, France

Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss, 7, Parc Mediterrani de la Tecnologia
08860 Castelldefels, Spain

We consider linear spectral statistics built from the block-normalized correlation matrix of a set of M mutually independent scalar time series. This matrix is composed of $M \times M$ blocks that contain the sample cross correlation between pairs of time series. In particular, each block has size $L \times L$ and contains the sample cross-correlation measured at L consecutive time lags between each pair of time series. Let N denote the total number of consecutively observed windows that are used to estimate these correlation matrices. We analyze the asymptotic regime where $M, L, N \to +\infty$ while $ML/N \to c_\star$, $0 < c_\star < \infty$. We study the behavior of linear statistics of the eigenvalues of this block correlation matrix under these asymptotic conditions and show that the empirical eigenvalue distribution converges to a Marcenko-Pastur distribution. Our results are potentially useful in order to address the problem of testing whether a large number of time series are uncorrelated or not.

Keywords: Large random matrices; Stieltjes transform; Correlated time series; Sample block correlation matrices

1. Introduction

1.1. *Problem addressed and motivation*

We consider a set of M jointly stationary zero mean complex-valued scalar time series, denoted as $y_{1,n}, \ldots, y_{M,n}$, where $n \in \mathbb{Z}$. We assume that the joint distribution

2

of $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$ is the circularly symmetric complex Gaussian law[1]. In this paper, we study the behaviour of linear statistics of the eigenvalues of a certain large random matrix built from the available data when the M time series $(y_m)_{m=1,\ldots,M}$ are uncorrelated (i.e. independent), assuming that both the number of available samples and the number of series are large. Our results are potentially useful in order to address the problem of testing whether a large number of time series are uncorrelated or not.

In order to introduce the large random matrix models that we will address in the following, we consider a column vector gathering L consecutive observations of the mth time series starting at time n, namely

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

and from this build an ML-dimensional column vector

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

We will denote by $\mathcal{R}_L$ the $ML \times ML$ covariance matrix of this random vector, i.e. $\mathcal{R}_L = \mathbb{E}\left[\mathbf{y}_n^L \left(\mathbf{y}_n^L\right)^H \right]$ where $(\cdot)^H$ stands for transpose conjugate. This matrix is sometimes referred to as the spatio-temporal covariance matrix. Clearly, the M series $(y_m)_{m=1,\ldots,M}$ are uncorrelated, to be referred to as the hypothesis H$_0$ in the following, if and only if, for each integer L, matrix $\mathcal{R}_L$ is block-diagonal, namely

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

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

$$\{\mathcal{R}_{m,L}\}_{k,k'} = r_m\left(k - k'\right). \tag{1.1}$$

Here, $r_m\left(k\right)$, $k \in \mathbb{Z}$, is the covariance sequence of the mth time series, defined as

$$r_m\left(k\right) = \int_0^1 \mathcal{S}_m\left(\nu\right) \mathrm{e}^{2\pi\mathrm{i}\nu k} d\nu \tag{1.2}$$

where, for each m, $\mathcal{S}_m$ represents the spectral density of $(y_{m,n})_{n\in\mathbb{Z}}$. We will denote by $\mathcal{R}_{\mathrm{corr},L}$ the block correlation matrix defined by

$$\mathcal{R}_{\mathrm{corr},L} = \mathcal{B}_L^{-1/2} \mathcal{R}_L \mathcal{B}_L^{-1/2} \tag{1.3}$$

where

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

---

[1] Any finite linear combination $z = \sum_{m=1}^M \sum_{j=1}^J \alpha_j y_{m,n_j}$ of the random variables $((y_{m,n})_{n\in\mathbb{Z}})_{m=1,\ldots,M}$ is distributed according to the distribution $\mathcal{N}_{\mathbb{C}}(0, \delta^2)$, i.e. $\mathrm{Re}z$ and $\mathrm{Im}z$ are independent and $\mathcal{N}(0, \delta^2/2)$ distributed, where $\delta^2 > 0$ is the corresponding variance

Consequently, $\mathcal{R}_L$ is block diagonal for each L if and only if $\mathcal{R}_{\mathrm{corr},L} = \mathbf{I}_{ML}$ for each L.

A possible way to test whether the time series $(y_m)_{m=1,\dots,M}$ are uncorrelated thus consists in estimating $\mathcal{R}_{\mathrm{corr},L}$ for a suitable value of L, and subsequently comparing the corresponding estimate with $\mathbf{I}_{ML}$. We will assume from now on that, for each $m = 1, \dots, M$, the observations $y_{m,1}, \dots, y_{m,N+L-1}$ are available where N represents the number of observations that are averaged to build the test statistic for each time lag. In the following, we consider the standard sample estimate $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ defined by

$$\widehat{\mathcal{R}}_{\mathrm{corr},L} = \widehat{\mathcal{B}}_L^{-1/2} \widehat{\mathcal{R}}_L \widehat{\mathcal{B}}_L^{-1/2} \tag{1.4}$$

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

$$\widehat{\mathcal{R}}_L = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n^L \left(\mathbf{y}_n^L\right)^H \tag{1.5}$$

and where $\widehat{\mathcal{B}}_L$ is the corresponding block diagonal

$$\widehat{\mathcal{B}}_L = \mathrm{Bdiag}(\widehat{\mathcal{R}}_L) = \begin{pmatrix} \widehat{\mathcal{R}}_{1,L} & & \\ & \ddots & \\ & & \widehat{\mathcal{R}}_{M,L} \end{pmatrix}. \tag{1.6}$$

with $\widehat{\mathcal{R}}_{m,L}$, $m = 1, \dots, M$, denoting the corresponding $L \times L$ diagonal blocks. The expression (1.5) of $\widehat{\mathcal{R}}_L$ explains why we assume that $N+L-1$ samples are available, because if the sample size had been defined as N, $\widehat{\mathcal{R}}_L$ should have been defined by $\widehat{\mathcal{R}}_L = \frac{1}{N-L} \sum_{n=1}^{N-L} \mathbf{y}_n^L \left(\mathbf{y}_n^L\right)^H$, which would have complicated the notations. In any case, in the asymptotic regime considered in the paper, the ratio $\frac{L}{N}$ converges towards 0. Therefore, the actual sample size $N+L-1$ can be written as $N(1+o(1))$. Changing N with $N+L-1$ does therefore not modify the significance of the results of this paper.

Remark 1.1. A relevant question here is how to choose the lag parameter L. On the one hand, L should be sufficiently large, because this allows to identify correlations among samples in different time series that are well spaced in time. For instance, two time series chosen as copies of the same temporally white noise with a relative delay higher than L lags will be perceived as uncorrelated by examination of $\widehat{\mathcal{R}}_{corr,L}$, which is of course far from true. On the other hand, L should be chosen sufficiently low so that $ML/N \ll 1$ in order to make the estimation error $\|\widehat{\mathcal{R}}_{\mathrm{corr},L} - \mathbf{I}_{ML}\|$ reasonably low under the hypothesis H_0. If the number M of time series is large and that the number of observations N is not unlimited, the condition $ML/N \ll 1$ requires the selection of a small value for L. Such a choice may thus reduce drastically the efficiency of the uncorrelation tests based on $\|\widehat{\mathcal{R}}_{\mathrm{corr},L} - \mathbf{I}_{ML}\|$. Finding statistics having a well defined behaviour under H_0 when ML and N are

4

of the same order of magnitude would allow to consider larger values of L, thus improving the performance of the corresponding tests.

In this paper, we propose to study the behavior of spectral statistics built from the eigenvalues of $\widehat{\mathcal{R}}_{\text{corr},L}$, which will be denoted by $(\hat{\lambda}_{k,N})_{k=1,\ldots,ML}$. More specifically, we will consider statistics of the form[2]

$$\widehat{\phi}_N = \frac{1}{ML}\text{Tr}\left[\phi\left(\widehat{\mathcal{R}}_{\text{corr},L}\right)\right] = \frac{1}{ML}\sum_{k=1}^{ML}\phi\left(\hat{\lambda}_{k,N}\right) \qquad (1.7)$$

where ϕ is assumed to be a suitable function, and will study the behaviour of $\widehat{\phi}_N$ under H_0 in asymptotic regimes where M, N, L converge towards $+\infty$ in such a way that $c_N = \frac{ML}{N}$ converges towards a non zero constant $c_* \in (0, +\infty)$.

The main result of this paper establishes the asymptotic conditions under which $\widehat{\phi}_N$ converges almost surely towards the integral of ϕ with respect to the Marcenko-Pastur distribution. In order to analyze the asymptotic behavior of the above class of statistics, we use large random matrix methods that relate the quantity $\widehat{\phi}_N$ with the empirical eigenvalue distribution of $\widehat{\mathcal{R}}_{\text{corr},L}$, denoted as

$$d\hat{\mu}_N(\lambda) = \frac{1}{ML}\sum_{k=1}^{ML}\delta_{\lambda-\hat{\lambda}_{k,N}}, \qquad (1.8)$$

that is

$$\widehat{\phi}_N = \int \phi(\lambda)\,d\hat{\mu}_N(\lambda).$$

We will establish the behavior of $\widehat{\phi}_N$ by studying the empirical eigenvalue distribution $d\hat{\mu}_N(\lambda)$.

Definition 1. Let $\mu_{mp,d}$ denote the Marcenko-Pastur distribution of parameter d. We recall that for each $d > 0$, $\mu_{mp,d}$ is the limit of the empirical eigenvalue distribution of a large random matrix $\frac{1}{K}\mathbf{X}\mathbf{X}^H$ where $\mathbf{X}$ is a $J \times K$ random matrix with zero mean unit variance i.i.d. entries and where both J and K converge towards $+\infty$ in such a way that $\frac{J}{K} \to d$.

We will prove that, under certain asymptotic assumptions, the statistic $\widehat{\phi}_N$ can be described (up to some error terms) as the integral of $\phi(\lambda)$ with respect to Marchenko-Pastur distribution of parameter c_N, in the sense that

$$\widehat{\phi}_N - \int_{\mathbb{R}^+}\phi(\lambda)\,d\mu_{mp,c_N}(\lambda) \to 0 \qquad (1.9)$$

---

[2]The application of a function ϕ to a Hermitian matrix should be understood as directly applied to its eigenvalues.

almost surely. We will also characterize the rate of convergence to zero of the corresponding error term in (1.9). This result will establish the conditions under which we can test whether the M time series $y_1, \ldots, y_M$ are uncorrelated by comparing linear spectral statistics $\widehat{\phi}_N$ with the corresponding limits under H_0 as established above.

1.2. *On the literature*

Testing whether M time series are uncorrelated is an important problem that has been extensively addressed in the past. Apart from a few works devoted to the case where the number of time series M converges towards $+\infty$ (see below), the vast majority of published papers assumed that M is a fixed integer. In this context, we can first mention spectral domain approaches based on the observation that the M time series $(y_{1,n})_{n \in \mathbb{Z}}, \ldots, (y_{M,n})_{n \in \mathbb{Z}}$ are uncorrelated if and only the spectral coherence matrix of the M–variate time series $(\mathbf{y}_n)_{n \in \mathbb{Z}}$, where $\mathbf{y}_n = (y_{1,n}, \ldots, y_{M,n})^T$, is reduced to $\mathbf{I}_M$ at each frequency. Some examples following this approach are [40], [39], [9], [10]. A number of papers also proposed to develop lag domain approaches, e.g. [17], [18], [8], [23] which considered test statistics based on empirical estimates of the autocorrelation coefficients between the residuals of the various time series. See also [11] for a more direct approach.

We next review the very few existing works devoted to the case where the number M of time series converges towards $+\infty$. We are just aware of papers addressing the case where the observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent identically distributed (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 work, these papers assume that 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 $\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^H)$ is diagonal. Therefore, it is reasonable to consider test statistics that are functionals of the sample covariance matrix $\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^H$. In particular, when the observations are i.i.d. Gaussian random vectors, the generalized likelihood ratio test (GLRT) consists in comparing the test statistics $\log \det(\widehat{\mathcal{R}}_{\mathrm{corr}})$ to a threshold, where $\widehat{\mathcal{R}}_{\mathrm{corr}} = \widehat{\mathcal{R}}_{\mathrm{corr},1}$ represents the sample correlation matrix. [19] proved that under H_0, the empirical eigenvalue distribution of $\widehat{\mathcal{R}}_{\mathrm{corr}}$ converges almost surely towards the Marcenko-Pastur distribution $\mu_{mp,d}$ and therefore, that $\frac{1}{M} \mathrm{Tr}(\phi(\widehat{\mathcal{R}}_{\mathrm{corr}}))$ converges towards $\int \phi(\lambda) d\mu_{mp,d}(\lambda)$ for each bounded continuous function ϕ. In the Gaussian case, [20] also established a central limit theorem (CLT) for $\log \det(\widehat{\mathcal{R}}_{\mathrm{corr}})$ under H_0 using the moment method. [7] remarked that, in the Gaussian real case, $(\det(\widehat{\mathcal{R}}_{\mathrm{corr}}))^{M/2}$ is the product of independent beta distributed random variables. Therefore, $\log \det(\widehat{\mathcal{R}}_{\mathrm{corr}})$ appears as the sum of independent random variables, thus deducing the CLT. We finally mention [30] in which a CLT on linear statistics of the eigenvalues of $\widehat{\mathcal{R}}_{\mathrm{corr}}$ is established in the Gaussian case using large random matrix

techniques when the covariance matrix $\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^H)$ is not necessarily diagonal. This allows to study the asymptotic performance of the GLRT under certain class of alternatives.

Regarding the asymptotic behaviour of the empirical eigenvalue distribution of the complete matrix $\widehat{\mathcal{R}}_L$, it seems relevant to highlight the work in [26] and [27], which also addressed in the asymptotic regime considered in the present paper. More specifically, [26] assumed that the M mutually independent time series $y_1, \ldots, y_M$ are i.i.d. Gaussian and established that the empirical eigenvalue distribution of $\widehat{\mathcal{R}}_L$ converges towards the Marcenko Pastur distribution μ_{mp,c_*}. Moreover, if $L = \mathcal{O}(N^\beta)$ with $\beta < 2/3$, it is established that almost surely, for N large enough, all the eigenvalues of $\widehat{\mathcal{R}}_L$ are located in a neighbourhood of the support of μ_{mp,c_*}. In [27], the mutually independent time series $y_1, \ldots, y_M$ are no longer assumed i.i.d. and it is established that the empirical eigenvalue distribution has a deterministic behaviour. The corresponding deterministic equivalent is characterized, and some results on the corresponding speed of convergence are given. As it will appear below, the present paper uses extensively in Sections 4 and 5 the tools developed in [27].

We also mention [29], which developed large random matrix methods in order to test the hypothesis H_0. However, the approach used in [29] is based on the study of the asymptotic behaviour of the empirical eigenvalue distribution of a frequency smoothed estimator of the spectral coherence matrix. While the techniques developed in [29] appear in general completely different from the technical content of the present paper, we mention that our Section 3 was inspired by Section 4.1 in [29], even though the technical problem solved in section 3 appears harder to solve than that in [29, Section 4.1].

We finally point out that a number of previous works addressed the behaviour of the estimated auto-covariance matrix $\hat{\mathbf{R}}_x(\tau) = \frac{1}{N} \sum_{n=1}^{N-\tau} \mathbf{x}_{n+\tau} \mathbf{x}_n^H$ of a M dimensional time series $\mathbf{x} = (\mathbf{x}_n)_{n \in \mathbb{Z}}$ at a given lag τ in the asymptotic regime where $\frac{M}{N} \to d$ with $d > 0$. We can mention [21], [24], [25], [4], [32], which, under various assumptions on $\mathbf{x}$, study the behaviour of the empirical eigenvalue distribution of $\hat{\mathbf{R}}_x(\tau) + \hat{\mathbf{R}}_x^H(\tau)$, $\hat{\mathbf{R}}_x(\tau)\hat{\mathbf{R}}_x^H(\tau)$, symmetric polynomials of $(\hat{\mathbf{R}}_x(\tau), \hat{\mathbf{R}}_x^H(\tau))$, or of $\hat{\mathbf{R}}_x(\tau)$. We also mention the work in [28], where the asymptotic behaviour of the singular values distribution of the estimated auto-covariance matrix between finite dimensional past and future of $\mathbf{x}$ (which, up to the end effects, depend on matrices $(\hat{\mathbf{R}}_x(\tau))_{\tau=1,\ldots,K}$ for a fixed integer K) is studied when $\frac{M}{N} \to d$ with $d > 0$. These contributions are not directly related to the present paper in that they study the properties of $\hat{\mathbf{R}}_x(\tau)$ for a single value of τ (or for a finite number of values of τ in [28]) when M and N are of the same order of magnitude, while our random matrix model depends, up to a block Toeplitzification of matrix $\widehat{\mathcal{R}}_L$, on $(\hat{\mathbf{R}}_y(\tau))_{\tau=0\ldots,L}$, where, this time, M, N, L converge towards $+\infty$ in such a way that $\frac{ML}{N} \to c_\star$.

1.3. *Assumptions*

Assumption 1. The complex scalar time series y_m, $m \geq 1$, are mutually independent, stationary, zero mean and circularly symmetric Gaussian distributed with autocovariance sequence $r_m = (r_m(k))_{k \in \mathbb{N}}$ defined as $r_m(k) = \mathbb{E}[y_{m,n+k} y_{m,n}^*]$ and associated spectral densities $(\mathcal{S}_m(\nu))_{m \geq 1}$.

Assumption 2. All along the paper, we assume that $M \to +\infty, N \to +\infty$ in such a way that $c_N = \frac{ML}{N} \to c_\star$, where $0 < c_\star < +\infty$, and that $L = L(N) = \mathcal{O}(N^\beta)$ for some constant $\beta \in (0,1)$. In order to shorten the notations, $N \to +\infty$ should be understood as the above asymptotic regime.

We will need that the spectral densities are bounded above and below uniformly in M, namely

Assumption 3. The spectral densities are such that

$$\sup_{m \geq 1} \max_{\nu \in [0,1]} \mathcal{S}_m(\nu) = s_{max} < +\infty \tag{1.10}$$

$$\inf_{m \geq 1} \min_{\nu \in [0,1]} \mathcal{S}_m(\nu) = s_{min} > 0. \tag{1.11}$$

Note that, for each $m = 1, \ldots, M$, the matrix $\mathcal{R}_{m,L}$ can be seen as an $L \times L$ diagonal block of an infinite Toeplitz matrix with symbol $\mathcal{S}_m(\nu)$. Therefore, Assumption 3 directly implies that, for each N, these matrices verify $s_{min}\mathbf{I}_L \leq \mathcal{R}_{m,L} \leq s_{max}\mathbf{I}_L$. This property will be used a number of times throughout the text.

Let us denote by $\mathbf{r}_M$ the M-dimensional sequence of covariances, namely

$$\mathbf{r}_M(k) = [r_1(k), \ldots, r_M(k)]^T \tag{1.12}$$

where $r_m(k)$, $m = 1, \ldots, M$ are defined in (1.2). We can consider the sequence of Euclidean norms $\{\|\mathbf{r}_M(k)\|\}_{k \in \mathbb{Z}}$. At some points, we will need the corresponding series to be of order $\mathcal{O}(\sqrt{M})$.

Assumption 4. The multivariate covariance sequence $\mathbf{r}_M$ defined in (1.12) is such that

$$\sup_{M \geq 1} \frac{1}{\sqrt{M}} \sum_{k \in \mathbb{Z}} \|\mathbf{r}_M(k)\| < +\infty.$$

On the other hand, we will also need to impose a certain rate of decayment of $\sup_{m \geq 1} \sum_{|k| \geq n+1} |r_m(k)|$ when $n \to +\infty$. To that effect, we introduce the weighting sequence $(\omega(n))_{n \in \mathbb{Z}}$ defined as

$$\omega(n) = (1 + |n|)^\gamma$$

8

where $\gamma \geq 0$ is given. This sequence belongs to the class of strong Beurling weights (see [37], Chapter 5), which are functions ω on $\mathbb{Z}$ with the properties: (i) $\omega(n) \geq 1$, (ii) $\omega(n) = \omega(-n)$, (iii) $\omega(m+n) \leq \omega(m)\omega(n)$ for all $m, n \in \mathbb{Z}$ and (iv) $n^{-1} \log \omega(n) \to 0$ as $n \to \infty$. We define ℓ_ω as the Banach space of two sided sequences $a = (a(n))_{n \in \mathbb{Z}}$ such that

$$\|a\|_\omega = \sum_{n=-\infty}^\infty \omega(n)\,|a(n)| = \sum_{n=-\infty}^\infty (1+|n|)^\gamma\,|a(n)| < +\infty.$$

When $\gamma = 0$, $\omega(n) = 1$ for each n, and ℓ_ω coincides with the Wiener algebra $\ell_1 = \{a = (a(n))_{n \in \mathbb{Z}}, \|a\|_1 < +\infty\}$. For each $\gamma \geq 0$, it holds that $\|a\|_1 \leq \|a\|_\omega$, and that ℓ_ω is included in ℓ_1. The function $\sum_{n \in \mathbb{Z}} a(n)e^{2i\pi n\nu}$ is thus well defined and continuous on $[0,1]$, and we will identify the sequence a to the above function. In particular, with a certain abuse of notation, $\sum_{n \in \mathbb{Z}} a(n)e^{2i\pi n\nu}$ will be denoted by $a(e^{2i\pi\nu})$ in the following. We can of course define the convolution product of sequences in ℓ_ω, namely

$$(a_1 * a_2)(n) = \sum_{m \in \mathbb{Z}} a_1(m)a_2(n-m)$$

which has the property that $\|a_1 * a_2\|_\omega \leq \|a_1\|_\omega \|a_2\|_\omega$, and therefore $a_1 * a_2 \in \ell_\omega$. Under the convolution product, we can see ℓ_ω as an algebra (the Beurling algebra) associated with the weight ω.

Assumption 5. For some $\gamma_0 > 0$, the covariance sequence r_m defined in (1.2) belongs to ℓ_{ω_0} for each m, where $\omega_0(n) = (1+|n|)^{\gamma_0}$. Moreover, it is assumed that

$$\sup_{m \geq 1} \|r_m\|_{\omega_0} < \infty. \tag{1.13}$$

Note that the fact that $r_m \in \ell_{\omega_0}$ implies that, for each $0 \leq \gamma < \gamma_0$, we have $r_m \in \ell_\omega$, where $\omega(n) = (1+|n|)^\gamma$. Moreover, (1.13) allows to control uniformly w.r.t. m of the remainder $\sum_{|k| \geq n+1} |r_m(k)|$. Indeed, observe that we can write

$$\|r_m\|_{\omega_0} \geq \sum_{|k| \geq n+1} (1+|k|)^{\gamma_0} |r_m(k)| \geq n^{\gamma_0} \sum_{|k| \geq n+1} |r_m(k)|.$$

Therefore, (1.13) implies that

$$\sup_{m \geq 1} \sum_{|k| \geq n+1} |r_m(k)| \leq \frac{\kappa}{n^{\gamma_0}} \tag{1.14}$$

for some constant κ.

In order to provide some insights on the significance of Assumptions 4 and 5, we provide examples and counterexamples. If there exists $\gamma > \gamma_0$ for which $\sup_m |r_m(n)| \leq \frac{\kappa}{n^{1+\gamma}}$ for each n large enough and $\sup_m |r_m(0)| < +\infty$, then, Assumptions 4 and 5 hold. If one of the time series is such that $\sum_n |r_m(n)| = +\infty$,

then neither Assumption 4 nor Assumption 5 hold true (we recall (see [27]) that Assumption 4 implies that for each m, $\sum_{n\in\mathbb{Z}}|r_m(n)| < +\infty)$) . Finally, if one of the time series (say y_1) verifies $|r_1(n)| \sim_{n\to+\infty} \frac{\kappa}{n(\log n)^{1+\delta}}$ for $\delta > 0$ while $\sum_{m\geq 2}|r_m(n)| \leq \frac{\kappa}{n^{1+\gamma}}$ for $\gamma > \gamma_0$, then Assumption 4 holds, but Assumption 5 does not hold.

1.4. *Main Result*

The main objective of this paper is to establish the asymptotic conditions that guarantee that we can approximate the original statistic in $\widehat{\phi}_N$ by the corresponding integral with respect to the Marchenko-Pastur distribution as in (1.9). To that effect, we will introduce two intermediate quantities that will provide some refined approximations of the original statistic $\widehat{\phi}_N$.

In order to introduce the first intermediate quantity, we need to consider the matrix

$$\overline{\mathcal{R}}_{\mathrm{corr},L} = \mathcal{B}_L^{-1/2}\,\widehat{\mathcal{R}}_L\,\mathcal{B}_L^{-1/2}. \tag{1.15}$$

Note that $\overline{\mathcal{R}}_{\mathrm{corr},L}$ is matrix defined in the same way as $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ by replacing the estimated block-diagonal autocorrelation matrix $\widehat{\mathcal{B}}_L = \mathrm{Bdiag}(\widehat{\mathcal{R}}_L)$ by its true value $\mathcal{B}_L = \mathrm{Bdiag}(\mathcal{R}_L)$, which in fact coincides with $\mathcal{R}_L$ (we are assuming independent sequences). We define $\overline{\phi}_N$ as the modified linear statistic

$$\overline{\phi}_N = \frac{1}{ML}\sum_{k=1}^{ML} \phi(\overline{\lambda}_{k,N}) = \int_{\mathbb{R}^+} \phi(\lambda)d\overline{\mu}_N(\lambda) \tag{1.16}$$

where $(\overline{\lambda}_{k,N})_{k=1,\dots,ML}$ are the eigenvalues of matrix $\overline{\mathcal{R}}_{\mathrm{corr},L}$ and where $\overline{\mu}_N(\lambda)$ is the associated empirical eigenvalue distribution.

In order to introduce the second intermediate quantity, we recall that, given an integer K, a $K\times K$ matrix-valued positive measure $\boldsymbol{\mu}$ is a σ–additive function from the Borel sets of $\mathbb{R}$ onto the set of all positive definite $K\times K$ matrices (see e.g. [36, Chapter 1] for more details).

Definition 2. We denote by $\mathcal{S}_{ML}(\mathbb{R}^+)$ the set of all $ML\times ML$ matrix valued functions defined on $\mathbb{C}\setminus\mathbb{R}^+$ by

$$\mathcal{S}_{ML}(\mathbb{R}^+) = \left\{\int_{\mathbb{R}^+} \frac{1}{\lambda - z}\,d\boldsymbol{\mu}(\lambda)\right\}$$

where $\boldsymbol{\mu}$ is a positive $ML\times ML$ matrix-valued measure carried by $\mathbb{R}^+$ satisfying $\boldsymbol{\mu}(\mathbb{R}^+) = \mathbf{I}_{ML}$.

We will next introduce a deterministic scalar measure $\mu_N(\lambda)$ that will allow us to describe the asymptotic behavior of the modified statistic $\overline{\phi}_N$. To that effect, we

need to introduce some operators that were originally used in [27], which inherently depend on the covariance sequences $(r_m)_{m \geq 1}$. In order to introduce these operators, for $\nu \in [0, 1]$ and $R \in \mathbb{N}$, we define the column vector

$$\mathbf{d}_R(\nu) = \left(1, \mathrm{e}^{2i\pi\nu}, \ldots, \mathrm{e}^{2i\pi(R-1)\nu}\right)^T \tag{1.17}$$

and let $\mathbf{a}_R(\nu)$ denote the corresponding normalized vector

$$\mathbf{a}_R(\nu) = \frac{1}{\sqrt{R}}\mathbf{d}_R(\nu). \tag{1.18}$$

With these two definitions, we are now able to introduce the Toeplitzation operators used to define the above deterministic measure $\mu_N(\lambda)$.

Definition 3. For a given squared matrix $\mathbf{M}$ with dimensions $R \times R$, we define $\Psi_K^{(m)}(\mathbf{M})$, $m = 1, \ldots, M$, as the $K \times K$ Toeplitz matrix given by

$$\Psi_K^{(m)}(\mathbf{M}) = \int_0^1 \mathcal{S}_m(\nu)\,\mathbf{a}_R^H(\nu)\,\mathbf{M}\mathbf{a}_R(\nu)\,\mathbf{d}_K(\nu)\,\mathbf{d}_K^H(\nu)\,d\nu.$$

The above operator is the key building block that defines Ψ and $\overline{\Psi}$, which are the ones that determine the master equations that define $\mu_N(\lambda)$.

Definition 4. Consider an $N \times N$ matrix $\mathbf{M}$. We define $\Psi(\mathbf{M})$ as an $ML \times ML$ block diagonal matrix with mth diagonal block given by $\Psi_L^{(m)}(\mathbf{M})$. Finally, consider an $ML \times ML$ matrix $\mathbf{M}$, and let $\mathbf{M}_{m,m}$ denote its mth $L \times L$ diagonal block. We define $\overline{\Psi}(\mathbf{M})$ as the $N \times N$ matrix given by

$$\overline{\Psi}(\mathbf{M}) = \frac{1}{M} \sum_{m=1}^{M} \Psi_N^{(m)}(\mathbf{M}_{m,m}). \tag{1.19}$$

Having now introduced the above operators, we are now ready to present the master equations that define the deterministic measure $\mu_N(\lambda)$. Consider a $z \in \mathbb{C}^+$ and the following pair of equations in $\mathbf{T}_N(z), \widetilde{\mathbf{T}}_N(z)$:

$$\mathbf{T}_N(z) = -\frac{1}{z}\left(\mathbf{I}_{ML} + \mathcal{B}_L^{-1/2}\Psi\left(\widetilde{\mathbf{T}}_N^T(z)\right)\mathcal{B}_L^{-1/2}\right)^{-1} \tag{1.20}$$

$$\widetilde{\mathbf{T}}_N(z) = -\frac{1}{z}\left(\mathbf{I}_N + c_N\overline{\Psi}^T\left(\mathcal{B}_L^{-1/2}\mathbf{T}_N(z)\mathcal{B}_L^{-1/2}\right)\right)^{-1}. \tag{1.21}$$

We will see that there exists a unique pair of solutions $(\mathbf{T}_N(z), \widetilde{\mathbf{T}}_N(z))$ to the above equations in the set $\mathcal{S}_{ML}(\mathbb{R}^+) \times \mathcal{S}_N(\mathbb{R}^+)$. We will denote as $\boldsymbol{\mu}_N(\lambda)$ the matrix valued measure with Stieltjes transform $\mathbf{T}(z)$ and μ_N the probability measure

$$\mu_N(\lambda) = \frac{1}{ML}\mathrm{Tr}(\boldsymbol{\mu}_N(\lambda)). \tag{1.22}$$

With this, we have now all the ingredients to present the main result of this paper.

Theorem 1.1. *Let Assumptions 1-5 hold true. Then, $\hat{\mu}_N(\lambda)$ converges weakly almost surely to $\mu_{mp,c_\star}(\lambda)$. Furthermore:*

(i) Consider $\widehat{\phi}_N$ and $\overline{\phi}_N$ defined in (1.7) and (1.16) respectively and assume that ϕ is well defined and smooth on a open subset containing $[0, +\infty)$ to be defined in section 3. For every small enough $\epsilon > 0$, there exists a $\gamma > 0$ independent of N such that

$$\mathbb{P}\left(\left|\widehat{\phi}_N - \overline{\phi}_N\right| > N^\epsilon \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right)\right) < \exp(-N^\gamma) \tag{1.23}$$

for all N sufficiently large.

(ii) Let $\beta < 4/5$ and assume that ϕ is a smooth function with compact support. Then, for every small $\epsilon > 0$ there exists a $\gamma > 0$ independent of N such that

$$\mathbb{P}\left(\left|\overline{\phi}_N - \int_{\mathbb{R}^+} \phi(\lambda) d\mu_N(\lambda)\right| > N^\epsilon \max\left(\frac{1}{M\sqrt{L}}, \frac{1}{M^2}\right)\right) < \exp(-N^\gamma) \tag{1.24}$$

for all N sufficiently large.

(iii) Consider the Marchenko-Pastur distribution with parameter $c_N = \frac{ML}{N}$ as given in Definition 1. Then, for every $\gamma < \gamma_0$, $\gamma \neq 1$ and every compactly supported smooth function ϕ, we have

$$\left|\int_{\mathbb{R}^+} \phi(\lambda) d\mu_N(\lambda) - \int_{\mathbb{R}^+} \phi(\lambda) d\mu_{mp,c_N}(\lambda)\right| < \kappa \frac{1}{L^{2\min(\gamma,1)}} \tag{1.25}$$

for some universal constant $\kappa > 0$.

The above theorem basically establishes three levels of approximation of the original linear spectral statistic $\widehat{\phi}_N$ and provides the speed of convergence to zero of the corresponding error terms. In particular, it is interesting to observe that the error term in (1.25) becomes the dominant one as soon as $\beta < 1/3$ if $\gamma_0 > 1$. Note that the situation where β is small (or, equivalently, $L \ll M$) is the most relevant asymptotic scenario. Otherwise, the ratio M/N converges quickly towards 0, which, in practice represents situations in which $M \ll N$. Therefore, it may be possible to choose a reasonably large value of L such that $\frac{ML}{N} \ll 1$. In this context, the simpler asymptotic regime where M, N, L converge towards $+\infty$ in such a way that $\frac{ML}{N} \to 0$ may be relevant.

As a consequence of all the above, we observe that when $\beta < 1/3$ and $\gamma_0 > 1$, the dominant error incurred by approximating the linear spectra statistic $\widehat{\phi}_N$ as an integral with respect to the Marchenko-Pastur law is in fact an unknown deterministic term as established in (1.25).

12

1.5. *Outline of the proof of Theorem 1.1*

In this section, we provide some detail on the strategy that is followed in the proof of Theorem 1.1. In order to present the main steps, we first review the concept of stochastic domination introduced in [12] and slightly adapted in [29]. We summarize next the formulation in [29]. This definition will allow to denote the convergence of (1.23) and (1.24) in a more compact and convenient way. More details can be found in [29].

Definition 5 (Stochastic Domination). Consider two families of non-negative random variables, namely $X = \{X^{(N)}(u),\ N \in \mathbb{N},\ u \in U^{(N)}\}$ and $Y = \{Y^{(N)}(u),\ N \in \mathbb{N},\ u \in U^{(N)}\}$, where $U^{(N)}$ is a set that may depend on N. We say that X is stochastically dominated by Y and write $X \prec Y$ if, for all small $\epsilon > 0$, there exists some $\gamma > 0$ depending on ϵ such that

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

for each large enough $N > N_0(\epsilon)$.

On the other hand, we will say that a family of events $\Omega = \Omega^{(N)}(u)$ holds with exponentially high (resp. small) probability if there exist N_0 and $\gamma > 0$ such that, for any $N \geq N_0$, $\mathbb{P}(\Omega^{(N)}(u)) > 1 - \exp(-N^\gamma)$ (resp. $\mathbb{P}(\Omega^{(N)}(u)) < \exp(-N^\gamma)$) for each $u \in U^{(N)}$.

It can be seen that $\prec$ satisfies the usual arithmetic properties of order relations. In particular, given four families of non-negative random variables X_1, X_2, Y_1, Y_2 such that $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1 X_2 \prec Y_1 Y_2$ (see Lemma 2.1 in [29]).

The proof of Theorem 1.1 is developed in Sections 2 to 5. The main steps are outlined in what follows.

(i) Before we begin with the proper technical content of the paper, we close the present section with some useful properties and technical results that will become useful in the rest of the paper.

(ii) Section 2 provides some preliminary results on the asymptotic behavior of the sample estimate of the spatio-temporal covariance matrix $\widehat{\mathcal{R}}_L$ and its $L \times L$ diagonal blocks. The objective is to show that the eigenvalue behavior of $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ can be studied by examining the eigenvalue behavior of the matrix $\overline{\mathcal{R}}_{\mathrm{corr},L}$. More specifically, we will first prove that $\|\widehat{\mathcal{R}}_L\|$ is bounded with an exponentially large probability (recall that $\|\cdot\|$ denotes spectral norm) and that $\|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$ for each $m = 1, \ldots, M$, where we recall that $\widehat{\mathcal{R}}_{m,L}$ and $\mathcal{R}_{m,L}$ denote the mth diagonal block of $\widehat{\mathcal{R}}_L$ and $\mathcal{R}_L$ respectively (see (1.1) and (1.6)). This will immediately

imply that

$$\|\widehat{\mathcal{R}}_{\mathrm{corr},L} - \overline{\mathcal{R}}_{\mathrm{corr},L}\| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right). \tag{1.26}$$

(iii) Section 3 is devoted to the proof of (1.23), which basically quantifies the influence of replacing $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ with $\overline{\mathcal{R}}_{\mathrm{corr},L}$ in the corresponding linear spectral statistics. More specifically, by exploiting the Helffer-Sjöstrand formula in combination with the preliminary results in Section 2, Theorem 3.1 establishes that

$$\left|\widehat{\phi}_N - \overline{\phi}_N\right| \prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right) \tag{1.27}$$

Notice that (1.23) implies that

$$\left|\widehat{\phi}_N - \overline{\phi}_N\right| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right)$$

Therefore, (1.27) appears as a stronger result. As shown in Section 3, its proof is demanding.

(iv) Section 4 studies the error term $\overline{\phi}_N - \int \phi(\lambda)d\mu_N(\lambda)$. First, this section shows that, for any smooth function ϕ with domain containing $[0, +\infty)$ the study of the modified statistic $\overline{\phi}_N$ can be reduced to the study of the corresponding expectation, $\mathbb{E}\overline{\phi}_N$. More specifically, we establish that

$$\left|\overline{\phi}_N - \mathbb{E}\overline{\phi}_N\right| \prec \frac{1}{M\sqrt{L}}.$$

The remaining error term $\mathbb{E}\overline{\phi}_N - \int \phi(\lambda)d\mu_N(\lambda)$ will be characterized by adapting the tools developed in [27], which was devoted to the study of the empirical eigenvalue distribution of matrix $\widehat{\mathcal{R}}_L$. The main difference between the matrix model considered here and the one in [27] is the fact that here the matrix $\widehat{\mathcal{R}}_L$ is multiplied on both sides by the block diagonal deterministic matrix $\mathcal{B}_L^{-1/2}$, see further (1.15). This multiplication on both sides introduces some modifications in the master equation that defines $\mathbf{T}_N(z)$, which is obviously different from the one in [27]. Other than that, the strategy of the proof will follow [27] almost verbatim, and will mostly be omitted. First, we will establish the almost sure weak convergence of $\bar{\mu}_N - \mu_N$ towards zero (cf. Proposition 4.3). Then, by additionally imposing $\beta < 4/5$ in Assumption 2 and assuming that ϕ is compactly supported, we will be able to conclude that

$$\left|\mathbb{E}\overline{\phi}_N - \int \phi(\lambda)d\mu_N(\lambda)\right| \leq \kappa\frac{1}{M^2} \tag{1.28}$$

for some universal constant $\kappa > 0$. This will directly imply (1.24).

(v) Section 5 finally shows that the deterministic sequence of probability measures $(\mu_N)_{N \geq 1}$ can be approximated by μ_{mp,c_N} in the sense of (1.25) for compactly

supported smooth functions ϕ. A central step in the proof will be to establish that

$$\sup_m \sup_{\nu \in [0,1]} \left| \mathcal{S}_m(\nu)\mathbf{a}_L(\nu)^H \mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu) - 1 \right| = \mathcal{O}\left(\frac{1}{L^{\min(1,\gamma)}}\right), \gamma < \gamma_0, \gamma \neq 1$$

$$= \frac{\log L}{L}, \ \gamma = 1 < \gamma_0$$

(see further Lemma 5.1). This result, proved in Appendix D, is obtained by noting that $\mathbf{a}_L(\nu)^H \mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu)$ can be expressed in terms of the orthogonal Szegö polynomials associated to the measure $\mathcal{S}_m(\nu)d\nu$, and by adapting to our context certain asymptotic related results presented in [37, Chapter 5].

(vi) Section 6 concludes the paper with a numerical validation that confirms the converge rates as established in Theorem 1.1.

The main tool in order to study the linear spectral statistics of the estimated block correlation matrix $\widehat{\mathcal{R}}_{\text{corr},L}$ will be the Stieltjes transform of its empirical eigenvalue distribution defined by (1.8). More specifically, we will denote by $\hat{q}_N(z)$ the Stieltjes transform of $d\hat{\mu}_N(\lambda)$, that is

$$\hat{q}_N(z) = \int_{\mathbb{R}^+} \frac{1}{\lambda - z} d\hat{\mu}_N(\lambda) = \frac{1}{ML} \sum_{k=1}^{ML} \frac{1}{\hat{\lambda}_{k,N} - z}$$

which is well defined for $z \in \mathbb{C}^+$. This function can also be written as $\hat{q}_N(z) = \frac{1}{ML}\text{Tr}\widehat{\mathbf{Q}}_N(z)$ where $\widehat{\mathbf{Q}}_N(z)$ is the resolvent of matrix $\widehat{\mathcal{R}}_{\text{corr},L}$, namely

$$\widehat{\mathbf{Q}}_N(z) = \left(\widehat{\mathcal{R}}_{\text{corr},L} - z\mathbf{I}_{ML}\right)^{-1}. \tag{1.29}$$

Likewise, for $z \in \mathbb{C}^+$, we will respectively denote by $\mathbf{Q}_N(z)$ the resolvent of $\overline{\mathcal{R}}_{\text{corr},L}$ and by $q_N(z)$ the Stieltjes transform associated to its empirical eigenvalue distribution $d\overline{\mu}_N(\lambda)$, namely

$$\mathbf{Q}_N(z) = \left(\overline{\mathcal{R}}_{\text{corr},L} - z\mathbf{I}_{ML}\right)^{-1} \tag{1.30}$$

and

$$q_N(z) = \int_{\mathbb{R}^+} \frac{1}{\lambda - z}d\overline{\mu}_N(\lambda) = \frac{1}{ML}\text{Tr}\mathbf{Q}_N(z).$$

1.6. *Notations*

The set $\mathbb{C}^+$ is composed of the complex numbers with strictly positive imaginary parts. The conjugate of a complex number z is denoted z^*. The conjugate transpose of a matrix $\mathbf{A}$ is denoted $\mathbf{A}^H$ while the conjugate of $\mathbf{A}$ (i.e. the matrix whose entries are the conjugates of the entries of $\mathbf{A}$) is denoted $\mathbf{A}^*$. $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ represent the spectral norm and the Frobenius norm of matrix $\mathbf{A}$, respectively. For a square matrix $\mathbf{A}$, we write $\mathbf{A} > 0$ (resp. $\mathbf{A} \geq 0$) to state that $\mathbf{A}$ is positive definite (resp.

positive semi-definite). If $\mathbf{A}$ and $\mathbf{B}$ are two square matrices, $\mathbf{A} < \mathbf{B}$ (resp. $\mathbf{A} \leq \mathbf{B}$) should be read as $\mathbf{B} - \mathbf{A} > 0$ (resp. $\mathbf{B} - \mathbf{A} \geq 0$). Also, for two general matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ represents the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, i.e. the block matrix whose block (i, j) is $\mathbf{A}_{i,j} \mathbf{B}$. If $\mathbf{A}$ is a square matrix, $\mathrm{Im}(\mathbf{A})$ and $\mathrm{Re}(\mathbf{A})$ represent the Hermitian matrices

$$\mathrm{Im}(\mathbf{A}) = \frac{\mathbf{A} - \mathbf{A}^H}{2i}, \; \mathrm{Re}(\mathbf{A}) = \frac{\mathbf{A} + \mathbf{A}^H}{2}.$$

If $(\mathbf{A}_N)_{N \geq 1}$ (resp. $(\mathbf{b}_N)_{N \geq 1}$) is a sequence of matrices (resp. vectors) whose dimensions increase with N, $(\mathbf{A}_N)_{N \geq 1}$ (resp. $(\mathbf{b}_N)_{N \geq 1}$) is said to be uniformly bounded if $\sup_{N \geq 1} \|\mathbf{A}_N\| < +\infty$ (resp. $\sup_{N \geq 1} \|\mathbf{b}_N\| < +\infty$).

We will let $\mathbf{J}_K$ denote the $K \times K$ shift matrix with ones in the first upper diagonal and zeros elsewhere, namely $\{\mathbf{J}_K\}_{i,j} = \delta_{j-i=1}$. We will denote by $\mathbf{J}_K^{-1}$ its transpose in order to simplify the notation. Likewise, $\mathbf{J}_K^0 = \mathbf{I}_K$ will denote the $K \times K$ identity matrix.

If x is a complex-valued random variable, its expectation is denoted by $\mathbb{E}(x)$ and its variance as

$$\mathrm{Var}(x) = \mathbb{E}(|x|^2) - |\mathbb{E}(x)|^2.$$

The zero-mean random variable $x - \mathbb{E}(x)$ is denoted x°.

In some parts of the paper, we will need to bound quantities by constants that do not depend on the system dimensions nor on the complex variable z. These will be referred to as "nice constants".

Definition 6 (Nice constants and nice polynomials). A nice constant is a positive constant independent of the dimensions L, M, N and the complex variable z. A nice polynomial is a polynomial whose degree is independent from L, M, N, and whose coefficients are nice constants. Throughout the paper, κ and P_1, P_2 will represent a generic nice constant and two generic nice polynomials respectively, whose values may change from one line to another. Finally, $C(z)$ will denote a general term of the form $C(z) = P_1(|z|)P_2(1/\mathrm{Im}z)$.

1.7. *Background on Stieltjes transforms of positive matrix valued measures*

We recall that $\mathcal{S}_K(\mathbb{R}^+)$ denotes the set of all Stieltjes transforms of $K \times K$ positive matrix-valued measures $\boldsymbol{\mu}$ carried by $\mathbb{R}^+$ verifying $\boldsymbol{\mu}(\mathbb{R}^+) = \mathbf{I}_K$. The elements of the class $\mathcal{S}_K(\mathbb{R}^+)$ satisfy the following properties:

Proposition 1.1. *Consider an element* $\mathbf{S}(z) = \int_{\mathbb{R}^+} \frac{d\boldsymbol{\mu}(\lambda)}{\lambda - z}$ *of* $\mathcal{S}_K(\mathbb{R}^+)$. *Then, the following properties hold true:*

16

    (i) $\mathbf{S}$ *is analytic on* $\mathbb{C}^+$.
    (ii) $\mathrm{Im}(\mathbf{S}(z)) \geq 0$ *and* $\mathrm{Im}(z\,\mathbf{S}(z)) \geq 0$ *if* $z \in \mathbb{C}^+$.
    (iii) $\lim_{y \to +\infty} -iy\mathbf{S}(iy) = \mathbf{I}_K$.
    (iv) $\mathbf{S}(z)\mathbf{S}^H(z) \leq \frac{\mathbf{I}_K}{(\mathrm{Im}z)^2}$ *for each* $z \in \mathbb{C}^+$.
    (v) $\int_{\mathbf{R}^+} \lambda\, d\boldsymbol{\mu}(\lambda) = \lim_{y \to +\infty} \mathrm{Re}\left(-iy(\mathbf{I}_K + iy\mathbf{S}(iy))\right)$.

Conversely, if a function $\mathbf{S}(z)$ *satisfy properties (i), (ii), (iii), then* $\mathbf{S}(z) \in \mathcal{S}_K(\mathbb{R}^+)$.

While we have not been able to find a paper in which this result is proved, it has been well known for a long time (see however [15] for more details on (i), (ii), (iii), (v)), as well as Theorem 3 of [1] from which (iv) follows immediately). We however provide an elementary proof of (iv) because it is based on a version of the matrix Schwarz inequality that will be used later. Given a certain $K \times K$ positive matrix measure $\boldsymbol{\mu}$ carried by $\mathbb{R}^+$, we denote by $\mathbb{L}^2(\boldsymbol{\mu})$ the Hilbert space of all K-dimensional row vector-valued functions $\mathbf{u}(\lambda)$ defined on $\mathbb{R}^+$ satisfying $\int_{\mathbb{R}^+} \mathbf{u}(\lambda)\, d\boldsymbol{\mu}(\lambda)\, \mathbf{u}^H(\lambda) < +\infty$ endowed with the scalar product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \int_{\mathbb{R}^+} \mathbf{u}(\lambda)\, d\boldsymbol{\mu}(\lambda)\, \mathbf{v}^H(\lambda).$$

Then, if $\mathbf{U}(\lambda) = (\mathbf{u}_1(\lambda)^T, \dots, \mathbf{u}_{K_u}(\lambda)^T)^T)^T$ and $\mathbf{V}(\lambda) = (\mathbf{v}_1(\lambda)^T, \dots, \mathbf{v}_{K_v}(\lambda)^T)^T)^T$ are matrices with K_u and K_v rows respectively, all of which are elements of $\mathbb{L}^2(\boldsymbol{\mu})$, it holds that

$$\langle \mathbf{U}, \mathbf{V} \rangle \langle \mathbf{V}, \mathbf{V} \rangle^{-1} \langle \mathbf{U}, \mathbf{V} \rangle^H \leq \langle \mathbf{U}, \mathbf{U} \rangle \tag{1.31}$$

where, with some abuse of notation, $\langle \mathbf{U}, \mathbf{V} \rangle$ denotes the matrix defined by $(\langle \mathbf{U}, \mathbf{V} \rangle)_{i,j} = \langle \mathbf{u}_i, \mathbf{v}_j \rangle$. This inequality can be directly proven by considering the $(K_u + K_v) \times K$ matrix $\mathbf{W} = [\mathbf{U}^T, \mathbf{V}^T]^T$ and noting that $\langle \mathbf{W}, \mathbf{W} \rangle$ is positive semi-definite. This implies that its Schur complement is also positive semi-definite, which directly implies (1.31). Now, using (1.31) for $\mathbf{U}(\lambda) = \frac{\mathbf{I}}{\lambda - z}$ and $\mathbf{V} = \mathbf{I}$, and remarking that $|\lambda - z|^2 \geq (\mathrm{Im}z)^2$ for each $\lambda \in \mathbb{R}^+$, we immediately obtain (iv).

1.8. *Further properties of stochastic domination and concentration inequalities*

The following result is a direct consequence of the union bound.

Lemma 1.1. *Let* $X_1, \dots, X_P$ *denote a collection of* $P \in \mathbb{N}$ *families of non-negative random variables, each one defined as* $X_p = \{X_p^{(N)}(u),\ N \in \mathbb{N},\ u \in U^{(N)}\}$. *Let* Y *denote an equivalently defined family of non-negative random variables such that* $X_p \prec Y$ *for each* $p = 1, \dots, P$. *Assuming that* $P \leq N^C$ *for some universal constant* C, *we have*

$$\max_{p=1,\dots,P} X_p \prec Y.$$

On the other hand, we will be using a number of inequalities based on concentration of functions of random Gaussian vectors. More specifically, consider the real-valued function $f(\mathbf{x}, \mathbf{x}^*)$ where $\mathbf{x}$ is a complex N-dimensional variable and where $(\cdot)^*$ denotes complex conjugate. If $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$, $f(\mathbf{x}, \mathbf{x}^*)$ can be interpreted as a function of the $2N$–dimensional $\mathcal{N}(0, \mathbf{I}_{2N})$ vector $\left(\sqrt{2}\mathrm{Re}(\mathbf{x}^T), \sqrt{2}\mathrm{Im}(\mathbf{x}^T)\right)^T$. If f is 1-Lipschitz, there exists a universal constant $C > 0$ such that

$$\mathbb{P}\left[|f(\mathbf{x}, \mathbf{x}^*) - \mathbb{E}f(\mathbf{x}, \mathbf{x}^*)| > t\right] \leq C\exp{-Ct^2}. \tag{1.32}$$

This concentration inequality is well known if f is a function of a $\mathcal{N}(0, \mathbf{I}_N)$ real-valued vector $\mathbf{x}$ ([38, Theorem 2.1.12]). We notice that (1.32) implies that $|f(\mathbf{x}, \mathbf{x}^*) - \mathbb{E}f(\mathbf{x}, \mathbf{x}^*)| \prec 1$ (see [29] for more details).

Finally, we will also make extensive use of the Hanson-Wright inequality, proven in [34] for the subgaussian real-valued case, but easily extended to the complex Gaussian context. If $\mathbf{A}$ denotes an $N \times N$ matrix of complex entries and if $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$, then

$$\mathbb{P}\left[|\mathbf{x}^H\mathbf{A}\mathbf{x} - \mathbb{E}\mathbf{x}^H\mathbf{A}\mathbf{x}| > t\right] \leq 2\exp\left[-C\min\left(\frac{t^2}{\|\mathbf{A}\|_F^2}, \frac{t}{\|\mathbf{A}\|}\right)\right] \tag{1.33}$$

where here again $C > 0$ is a universal constant and where $\|\cdot\|_F$ and $\|\cdot\|$ respectively denote Frobenius and spectral norms.

1.9. *Additional properties of the Toeplitzification operators*

We introduce here some additional properties of the Toeplitzification operators introduced in Definitions 3-4, which will prove useful in the course of the derivations. For a given squared matrix $\mathbf{M}$ with dimensions $R \times R$, the operator $\Psi_K^{(m)}(\mathbf{M})$ in Definition 3 can alternatively be represented as an $K \times K$ Toeplitz matrix with (i, j)th entry equal to

$$\left\{\Psi_K^{(m)}(\mathbf{M})\right\}_{i,j} = \sum_{l=-R+1}^{R-1} r_m(i - j - l)\tau(\mathbf{M})(l) \tag{1.34}$$

or, alternatively, as the matrix

$$\Psi_K^{(m)}(\mathbf{M}) = \sum_{n=-K+1}^{K-1}\left(\sum_{l=-R+1}^{R-1} r_m(n - l)\tau(\mathbf{M})(l)\right)\mathbf{J}_K^{-n} \tag{1.35}$$

where the sequence $\tau(\mathbf{M})(l)$, $-R < l < R$, is defined as

$$\tau(\mathbf{M})(l) = \frac{1}{R}\mathrm{Tr}\left[\mathbf{M}\mathbf{J}_R^l\right]. \tag{1.36}$$

We observe that, with this definition,

$$\sum_{r=-(R-1)}^{R-1} |\tau(\mathbf{M})(r)|^2 \leq \frac{1}{R}\mathrm{Tr}(\mathbf{M}\mathbf{M}^H). \tag{1.37}$$

This inequality can be proven by noting that $\tau(\mathbf{M})(r)$, $r = -R+1, \ldots, R-1$ are the Fourier coefficients of the function $\nu \mapsto \mathbf{a}_R^H(\nu)\mathbf{M}\mathbf{a}_R(\nu)$ so that, by Parseval's identity,

$$\sum_{r=-(R-1)}^{R-1} |\tau(\mathbf{M})(r)|^2 = \int_0^1 \left|\mathbf{a}_R^H(\nu)\mathbf{M}\mathbf{a}_R(\nu)\right|^2 d\nu \leq$$

$$\leq \int_0^1 \mathbf{a}_R^H(\nu)\mathbf{M}\mathbf{M}^H\mathbf{a}_R(\nu)d\nu = \frac{1}{R}\text{Tr}(\mathbf{M}\mathbf{M}^H)$$

where we have used the Cauchy-Schwarz inequality.

We also mention the following property: If $\mathbf{A}$ is a $R \times R$ Toeplitz matrix with entries $\mathbf{A}_{i,j} = a(i-j)$ for some sequence $(a(l))_{l=-(R-1),\ldots,R-1}$, and if $\mathbf{B}$ is another $R \times R$ matrix, we have

$$\frac{1}{R}\text{Tr}(\mathbf{A}\mathbf{B}) = \sum_{l=-(R-1)}^{R-1} a(l)\tau(\mathbf{B})(-l) \tag{1.38}$$

The following properties are easily checked (see [27]).

– Given a square matrix $\mathbf{A}$ of dimension $K \times K$ and a square matrix $\mathbf{B}$ of dimension $R \times R$, we can write

$$\frac{1}{K}\text{Tr}\left[\mathbf{A}\Psi_K^{(m)}(\mathbf{B})\right] = \frac{1}{R}\text{Tr}\left[\Psi_R^{(m)}(\mathbf{A})\mathbf{B}\right] \tag{1.39}$$

– Given a square matrix $\mathbf{M}$ and a positive integer K, we have

$$\left\|\Psi_K^{(m)}(\mathbf{M})\right\| \leq \sup_{\nu \in [0,1]} |\mathcal{S}_m(\nu)| \|\mathbf{M}\|.$$

– Given a square positive definite matrix $\mathbf{M}$ and a positive integer K, the hypothesis $\inf_\nu \mathcal{S}_m(\nu) > 0$ implies that

$$\Psi_K^{(m)}(\mathbf{M}) > 0. \tag{1.40}$$

Consider now the two other linear operators in Definition 4, which respectively operate on $N \times N$ and $ML \times ML$ matrices. If $\mathbf{A}$ and $\mathbf{B}$ are $ML \times ML$ and $N \times N$ matrices, we see directly from (1.39) that

$$\frac{1}{N}\text{Tr}\left[\overline{\Psi}(\mathbf{A})\mathbf{B}\right] = \frac{1}{ML}\text{Tr}\left[\mathbf{A}\Psi(\mathbf{B})\right]. \tag{1.41}$$

We finally conclude this section by two useful propositions that follow directly from [27].

Proposition 1.2. *Let* $\mathbf{\Gamma}^m(z)$, $m = 1, \ldots, M$, *be a collection of* $L \times L$ *matrix-valued complex functions belonging to* $\mathcal{S}_L(\mathbb{R}^+)$ *and define* $\mathbf{\Gamma}(z)$ *as the* $ML \times ML$ *block*

diagonal matrix given by $\mathbf{\Gamma}(z) = \mathrm{diag}\left(\mathbf{\Gamma}^1(z), \ldots, \mathbf{\Gamma}^M(z)\right)$. *Then, for each* $z \in \mathbb{C}^+$, *the matrix* $\mathbf{I}_N + c_N \overline{\mathbf{\Psi}}^T \left(\mathcal{B}_L^{-1/2} \mathbf{\Gamma}(z) \mathcal{B}_L^{-1/2}\right)$ *is invertible, so that we can define*

$$\widetilde{\mathbf{\Upsilon}}(z) = -\frac{1}{z}\left(\mathbf{I}_N + c_N \overline{\mathbf{\Psi}}^T \left(\mathcal{B}_L^{-1/2}\mathbf{\Gamma}(z)\mathcal{B}_L^{-1/2}\right)\right)^{-1}. \tag{1.42}$$

On the other hand, the matrix $\mathbf{I}_{ML} + \mathcal{B}_L^{-1/2}\mathbf{\Psi}\left(\widetilde{\mathbf{\Upsilon}}^T(z)\right)\mathcal{B}_L^{-1/2}$ *is also invertible, and we define*

$$\mathbf{\Upsilon}(z) = -\frac{1}{z}\left(\mathbf{I}_{ML} + \mathcal{B}_L^{-1/2}\mathbf{\Psi}\left(\widetilde{\mathbf{\Upsilon}}^T(z)\right)\mathcal{B}_L^{-1/2}\right)^{-1}. \tag{1.43}$$

Furthermore, $\widetilde{\mathbf{\Upsilon}}(z)$ *and* $\mathbf{\Upsilon}(z)$ *are elements of* $\mathcal{S}_N(\mathbb{R}^+)$ *and* $\mathcal{S}_{ML}(\mathbb{R}^+)$ *respectively. In particular, they are holomorphic on* $\mathbb{C}^+$ *and satisfy*

$$\mathbf{\Upsilon}(z)\mathbf{\Upsilon}^H(z) \le \frac{\mathbf{I}_{ML}}{(\mathrm{Im}z)^2}, \ \widetilde{\mathbf{\Upsilon}}(z)\widetilde{\mathbf{\Upsilon}}^H(z) \le \frac{\mathbf{I}_N}{(\mathrm{Im}z)^2}. \tag{1.44}$$

Moreover, there exist two nice constants η *and* $\widetilde{\eta}$ *such that*

$$\mathbf{\Upsilon}(z)\mathbf{\Upsilon}^H(z) \ge \frac{(\mathrm{Im}z)^2}{16(\eta^2 + |z|^2)^2}\mathbf{I}_{ML} \tag{1.45}$$

$$\widetilde{\mathbf{\Upsilon}}(z)\widetilde{\mathbf{\Upsilon}}^H(z) \ge \frac{(\mathrm{Im}z)^2}{16(\widetilde{\eta}^2 + |z|^2)^2}\mathbf{I}_N. \tag{1.46}$$

Proof. The proof is an easy adaptation of the proof of Lemma 4.1 in [27]. More precisely, if we replace in this Lemma matrix $\mathrm{Bdiag}(\mathbb{E}\mathbf{Q}(z))$ by $\mathbf{\Gamma}(z)$ and matrices $(\mathbf{R}(z), \widetilde{\mathbf{R}}(z))$ by $(\mathbf{\Upsilon}(z), \widetilde{\mathbf{\Upsilon}}(z))$, it is easy to check that the arguments of the proof of Lemma 4.1 in [27] can be extended to the particular context considered in the present paper. $\qquad\square$

In order to state the next result, we consider two $ML \times ML$ block diagonal matrices $\mathbf{S}, \mathbf{T}$ and two $N \times N$ matrices $\widetilde{\mathbf{S}}, \widetilde{\mathbf{T}}$. We also assume that $\mathbf{S}, \mathbf{T}, \widetilde{\mathbf{S}}, \widetilde{\mathbf{T}}$ are full rank matrices. For each fixed z, we define the linear operator Φ on the set of all $ML \times ML$ matrices by

$$\Phi(\mathbf{X}) = z^2 c_N \mathbf{S}\mathbf{\Psi}\left(\widetilde{\mathbf{S}}^T\overline{\mathbf{\Psi}}(\mathbf{X})\widetilde{\mathbf{T}}^T\right)\mathbf{T}. \tag{1.47}$$

Note that the operator Φ of course depends on $\mathbf{S}, \mathbf{T}, \widetilde{\mathbf{S}}, \widetilde{\mathbf{T}}, M, L, N$ and z. We also define the following linear operators on the set of all $ML \times ML$ Hermitian matrices:

$$\Phi_{\mathbf{T}^H}(\mathbf{X}) = |z|^2 c_N \mathbf{T}^H \mathbf{\Psi}\left(\widetilde{\mathbf{T}}^*\overline{\mathbf{\Psi}}(\mathbf{X})\widetilde{\mathbf{T}}^T\right)\mathbf{T} \tag{1.48}$$

$$\Phi_{\mathbf{S}}(\mathbf{X}) = |z|^2 c_N \mathbf{S}\mathbf{\Psi}\left(\widetilde{\mathbf{S}}^T\overline{\mathbf{\Psi}}(\mathbf{X})\widetilde{\mathbf{S}}^*\right)\mathbf{S}^H. \tag{1.49}$$

We remark that both operators are positive in the sense that if $\mathbf{X} \ge 0$, then $\Phi_{\mathbf{S}}(\mathbf{X}) \ge 0$ and $\Phi_{\mathbf{T}^H}(\mathbf{X}) \ge 0$. Let $\Phi^{(1)}(\mathbf{X}) = \Phi(\mathbf{X})$ and recursively define $\Phi^{(n+1)}(\mathbf{X}) = \Phi\left(\Phi^{(n)}(\mathbf{X})\right)$ for $n \ge 1$. Then, the following result holds.

Proposition 1.3. *For any two L-dimensional column vectors $\mathbf{a}$, $\mathbf{b}$ and for each $m = 1, \ldots, M$, the inequality*

$$\left| \mathbf{a}^H \left(\Phi^{(n)} (\mathbf{X}) \right)_{m,m} \mathbf{b} \right| \leq \left[\mathbf{a}^H \left(\Phi_{\mathbf{S}}^{(n)} \left(\mathbf{X}\mathbf{X}^H \right) \right)_{m,m} \mathbf{a} \right]^{1/2} \left[\mathbf{b}^H \left(\Phi_{\mathbf{T}^H}^{(n)} \left(\mathbf{I}_{ML} \right) \right)_{m,m} \mathbf{b} \right]^{1/2}$$
(1.50)

holds, where $(\mathbf{A})_{m,m}$ denotes the mth $L \times L$ diagonal block of $\mathbf{A}$. Moreover, if there exist two $ML \times ML$ positive definite matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ such that

$$\lim_{n \to +\infty} \Phi_{\mathbf{S}}^{(n)} (\mathbf{Y}_1) \to 0 \tag{1.51}$$

$$\lim_{n \to +\infty} \Phi_{\mathbf{T}^H}^{(n)} (\mathbf{Y}_2) \to 0 \tag{1.52}$$

then, for each $ML \times ML$ matrix $\mathbf{X}$,

$$\lim_{n \to +\infty} \Phi^{(n)} (\mathbf{X}) \to 0 \tag{1.53}$$

If, moreover, $\sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} (\mathbf{Y}_1) < +\infty$ and $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} (\mathbf{Y}_2) < +\infty$, then, for each $ML \times ML$ hermitian matrix $\mathbf{Y}$, the two series $\sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} (\mathbf{Y})$ and $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} (\mathbf{Y})$ are convergent. Finally, for each $ML \times ML$ matrix $\mathbf{X}$, $\sum_{n=0}^{+\infty} \Phi^{(n)} (\mathbf{X})$ is also convergent, and we have

$$\left\| \sum_{n=0}^{+\infty} \Phi^{(n)} (\mathbf{X}) \right\| \leq \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} \left(\mathbf{X}\mathbf{X}^H \right) \right\|^{1/2} \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} \left(\mathbf{I}_{ML} \right) \right\|^{1/2} \tag{1.54}$$

as well as

$$\left\| \sum_{n=0}^{+\infty} \Phi^{(n)} (\mathbf{X}) \right\| \leq \| \mathbf{X} \| \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} \left(\mathbf{I}_{ML} \right) \right\|^{1/2} \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} \left(\mathbf{I}_{ML} \right) \right\|^{1/2}. \tag{1.55}$$

Proof. Inequality (1.50) is established in Section 5 of [27]. We now prove (1.53). For this, we first remark that since matrices $(\mathbf{Y}_i)_{i=1,2}$ are positive definite, there exist $\alpha_1 > 0$ and $\alpha_2 > 0$ such that $\mathbf{Y}_i \geq \alpha_i \mathbf{I}_{ML}$ for $i = 1, 2$. As the operators $\Phi_{\mathbf{S}}$ and $\Phi_{\mathbf{T}^H}$ are positive, it holds that $\Phi_{\mathbf{S}}^{(n)}(\mathbf{Y}_1) \geq \alpha_1 \Phi_{\mathbf{S}}^{(n)}(\mathbf{I}_{ML})$ and $\Phi_{\mathbf{T}^H}^{(n)}(\mathbf{Y}_2) > \alpha_2 \Phi_{\mathbf{T}^H}^{(n)}(\mathbf{I}_{ML})$ for each n. Therefore, conditions (1.51) and (1.52) imply that $\Phi_{\mathbf{S}}^{(n)}(\mathbf{I}_{ML}) \to 0$ and $\Phi_{\mathbf{T}^H}^{(n)}(\mathbf{I}_{ML}) \to 0$. If $\mathbf{X}$ is a generic $ML \times ML$ matrix, the inequality $\mathbf{X}\mathbf{X}^H \leq \|\mathbf{X}\|^2 \mathbf{I}_{ML}$ implies that $\Phi_{\mathbf{S}}^{(n)}(\mathbf{X}\mathbf{X}^H) \leq \|\mathbf{X}\|^2 \Phi_{\mathbf{S}}^{(n)}(\mathbf{I}_{ML})$. Therefore, we deduce that for each matrix $\mathbf{X}$, $\Phi_{\mathbf{S}}^{(n)}(\mathbf{X}\mathbf{X}^H) \to 0$ when $n \to +\infty$. The inequality in (1.50) thus leads to (1.53). Using similar arguments, we check that the convergence $\sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} (\mathbf{Y}_1)$ and $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} (\mathbf{Y}_2)$ implies the convergence of $\sum_{n=0}^{+\infty} \Phi_{\mathbf{S}}^{(n)} (\mathbf{Y})$ and $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}^H}^{(n)} (\mathbf{Y})$ for each positive matrix $\mathbf{Y}$. If $\mathbf{Y}$ is not positive, it is sufficient to remark that $\mathbf{Y}$ can be written as the difference of 2 positive matrices to conclude to the convergence of the above two series. We finally consider

a general matrix $\mathbf{X}$, and establish that $\sum_{n=0}^{+\infty} \Phi^{(n)}(\mathbf{X})$ is convergent. For this, we remark that (1.50) implies that for each m and each k, the inequality

$$\sum_{n=0}^{k} \left| \mathbf{a}^H \left(\Phi^{(n)}(\mathbf{X}) \right)_{m,m} \mathbf{b} \right| \leq$$

$$\leq \left[\mathbf{a}^H \left(\sum_{n=0}^{k} \left(\Phi_{\mathbf{S}}^{(n)}(\mathbf{X}\mathbf{X}^H) \right)_{m,m} \right) \mathbf{a} \right]^{1/2} \left[\mathbf{b}^H \left(\sum_{n=0}^{k} \left(\Phi_{\mathbf{T}^H}^{(n)}(\mathbf{I}_{ML}) \right)_{m,m} \right) \mathbf{b} \right]^{1/2} \tag{1.56}$$

holds. This implies that

$$\sum_{n=0}^{+\infty} \left| \mathbf{a}^H \left(\Phi^{(n)}(\mathbf{X}) \right)_{m,m} \mathbf{b} \right| < +\infty$$

and that the series $\sum_{n=0}^{+\infty} \Phi^{(n)}(\mathbf{X})$ is convergent. The result in (1.54) is obtained by taking the limit in the inequality (1.56), while (1.55) is an immediate consequence of (1.54). □

2. Preliminary results on the empirical estimates $\widehat{\mathcal{R}}_L$ and $\widehat{\mathcal{R}}_{m,L}$

Consider again the sample block correlation matrix, namely $\widehat{\mathcal{R}}_{\mathrm{corr},L} = \widehat{\mathcal{B}}_L^{-1/2} \widehat{\mathcal{R}}_L \widehat{\mathcal{B}}_L^{-1/2}$, where we recall that $\widehat{\mathcal{B}}_L = \mathrm{Bdiag}(\widehat{\mathcal{R}}_L)$. In this section, we will show that we can replace the block diagonal sample covariance matrix $\widehat{\mathcal{B}}_L$ by the true matrix $\mathcal{B}_L = \mathcal{R}_L$ without altering the asymptotic behavior of the empirical eigenvalue distribution of $\overline{\mathcal{R}}_{\mathrm{corr},L} = \mathcal{B}_L^{-1/2} \widehat{\mathcal{R}}_L \mathcal{B}_L^{-1/2}$.

For this, we proceed in three steps. First, in Section 2.1, we prove that the spectral norm of $\widehat{\mathcal{R}}_L$ is bounded with exponentially high probability. Then, using similar arguments, we show in Section 2.2 that $\|\widehat{\mathcal{B}}_L - \mathcal{B}_L\| \prec \max(M^{-1/2}, L^{-\gamma_0})$. Finally, in Section 2.3 we establish that $\|\widehat{\mathcal{B}}_L^{-1/2} - \mathcal{B}_L^{-1/2}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$ using Hermitian matrix perturbation results. The fact that $\|\widehat{\mathcal{R}}_L\|$ is bounded with exponentially high probability will immediately imply that

$$\left\| \widehat{\mathcal{R}}_{\mathrm{corr},L} - \overline{\mathcal{R}}_{\mathrm{corr},L} \right\| \prec \max \left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}} \right). \tag{2.1}$$

We will write the normalized observations as $\mathbf{w}_{n,N} = \frac{1}{\sqrt{N}} \mathbf{y}_n^L$, where $n = 1, \dots, N$ and

$$\mathbf{W}_N = [\mathbf{w}_{1,N}, \dots, \mathbf{w}_{N,N}]. \tag{2.2}$$

Therefore $\widehat{\mathcal{R}}_L$ coincides with $\widehat{\mathcal{R}}_L = \mathbf{W}_N \mathbf{W}_N^H$. In the following, we will often drop the index N, and will denote $\mathbf{W}_N, \mathbf{w}_{j,N}, \mathbf{Q}_N, \dots$ by $\mathbf{W}, \mathbf{w}_j, \mathbf{Q}, \dots$ in order to simplify the notations.

2.1. *Control of the largest eigenvalue of $\widehat{\mathcal{R}}_L$*

The approach we follow is based on the observation that it is possible to add a bounded matrix to $\mathbf{W}_N\mathbf{W}_N^H$ to produce a block Toeplitz matrix. Controlling the largest eigenvalue of $\mathbf{W}_N\mathbf{W}_N^H$ becomes therefore equivalent to controlling the largest eigenvalue of the block Toeplitz matrix, a problem that can be solved by studying the supremum over the frequency interval of the spectral norm of the corresponding symbol.

2.1.1. *Modifying $\mathbf{W}_N\mathbf{W}_N^H$ into a block Toeplitz matrix*

In order to present this result, it is more convenient to reorganize the rows of matrix $\mathbf{W}_N$. For this, we define for each n the M dimensional random vector $\mathbf{y}_n$ defined by

$$\mathbf{y}_n = \begin{pmatrix} y_{1,n} \\ \vdots \\ y_{M,n} \end{pmatrix}. \tag{2.3}$$

$(\mathbf{y}_n)_{n\in\mathbb{Z}}$ is thus an M–dimensional stationary random sequence whose spectral density matrix $\mathbf{S}(\nu)$ coincides with the diagonal matrix $\mathbf{S}(\nu) = \mathrm{Diag}(\mathcal{S}_1(\nu), \ldots, \mathcal{S}_M(\nu))$. We next consider the $ML \times N$ matrix $\mathcal{W}_N$, which is defined as

$$\mathcal{W}_N = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{N-1} & \mathbf{y}_N \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_N & \mathbf{y}_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \mathbf{y}_{N+L-2} & \mathbf{y}_{N+L-1} \end{pmatrix}. \tag{2.4}$$

Observe that $\mathcal{W}_N$ can be obtained by simple permutation of the rows of $\mathbf{W}_N$ and consequently $\mathcal{W}_N\mathcal{W}_N^H$ and $\mathbf{W}_N\mathbf{W}_N^H$ have the same eigenvalues. In particular, they have the same spectral norm. For this reason, we may focus on the behavior of $\mathcal{W}_N$ from now on.

We define matrices $\mathcal{W}_{N,1}$ and $\mathcal{W}_{N,2}$ as the $ML \times (N-L+1)$ and $ML \times (L-1)$ matrices such that $\mathcal{W}_N = (\mathcal{W}_{N,1}, \mathcal{W}_{N,2})$. In particular, matrix $\mathcal{W}_{N,2}$ is given by

$$\mathcal{W}_{N,2} = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{y}_{N-L+2} & \mathbf{y}_{N-L+3} & \cdots & \mathbf{y}_{N-1} & \mathbf{y}_N \\ \mathbf{y}_{N-L+3} & \mathbf{y}_{N-L+4} & \cdots & \mathbf{y}_N & \mathbf{y}_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{N+1} & \mathbf{y}_{N+2} & \cdots & \mathbf{y}_{N+L-2} & \mathbf{y}_{N+L-1} \end{pmatrix}. \tag{2.5}$$

We now express $\mathcal{W}_{N,2}$ as $\mathcal{W}_{N,2} = \mathcal{W}_{N,2,1} + \mathcal{W}_{N,2,2}$ where $\mathcal{W}_{N,2,1}$ is the upper block triangular matrix given by

$$\mathcal{W}_{N,2,1} = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{y}_{N-L+2} & \mathbf{y}_{N-L+3} & \cdots & \mathbf{y}_{N-1} & \mathbf{y}_N \\ \mathbf{y}_{N-L+3} & \mathbf{y}_{N-L+4} & \cdots & \mathbf{y}_N & 0 \\ \mathbf{y}_{N-L+4} & \cdots & \mathbf{y}_N & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_N & 0 & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{2.6}$$

and where $\mathcal{W}_{N,2,2}$ is the lower block triangular matrix defined by

$$\mathcal{W}_{N,2,2} = \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{y}_{N+1} \\ 0 & \vdots & 0 & \mathbf{y}_{N+1} & \mathbf{y}_{N+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{y}_{N+1} \cdots & \cdots & \mathbf{y}_{N+L-2} \\ \mathbf{y}_{N+1} & \mathbf{y}_{N+2} & \cdots & \mathbf{y}_{N+L-2} & \mathbf{y}_{N+L-1} \end{pmatrix}. \tag{2.7}$$

In other words, matrix $\mathcal{W}_{N,2,1}$ is obtained by replacing in $\mathcal{W}_{N,2}$ vectors $\mathbf{y}_{N+1}, \ldots, \mathbf{y}_{N+L-1}$ by $\mathbf{0}, \ldots, \mathbf{0}$ while $\mathcal{W}_{N,2,2}$ is obtained by replacing in $\mathcal{W}_{N,2}$ vectors $\mathbf{y}_{N-L+2}, \ldots, \mathbf{y}_N$ by $\mathbf{0}, \ldots, \mathbf{0}$. We also define $\mathcal{W}_{N,0}$ as the $ML \times (L-1)$ lower block triangular matrix given by

$$\mathcal{W}_{N,0} = \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{y}_1 \\ 0 & \vdots & 0 & \mathbf{y}_1 & \mathbf{y}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{y}_1 \cdots & \cdots & \mathbf{y}_{L-2} \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{L-2} & \mathbf{y}_{L-1} \end{pmatrix}. \tag{2.8}$$

We finally introduce the $ML \times (N+L-1)$ block Hankel matrix $\widetilde{\mathcal{W}}_N$ defined by

$$\widetilde{\mathcal{W}}_N = (\mathcal{W}_{N,0}, \mathcal{W}_{N,1}, \mathcal{W}_{N,2,1}). \tag{2.9}$$

It is easy to check that $\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H$ is the block Toeplitz matrix whose $M \times M$ blocks $\left((\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H)_{k,l} \right)_{k,l=1,\ldots,L}$ are given by

$$(\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H)_{k,l} = \widehat{\mathbf{R}}_{k-l}$$

where the $M \times M$ matrices $(\widehat{\mathbf{R}}_l)_{l=-(L-1),\ldots,L-1}$ are defined by

$$\widehat{\mathbf{R}}_l = \frac{1}{N} \sum_{n=1}^{N-l} \mathbf{y}_{n+l} \mathbf{y}_n^H$$

for $l \geq 0$ and $\widehat{\mathbf{R}}_l = \widehat{\mathbf{R}}_{-l}^H$ for $l \leq 0$. In other words, for each l, $\widehat{\mathbf{R}}_l$ is the standard empirical biased estimate of the autocovariance matrix at lag l of the multivariate time series $(\mathbf{y}_n)_{n \in \mathbb{Z}}$.

Matrix $\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H$ also coincides with the block Toeplitz matrix associated to the symbol $\widehat{\mathbf{S}}(\nu)$ defined by

$$\widehat{\mathbf{S}}(\nu) = \sum_{l=-(L-1)}^{L-1} \widehat{\mathbf{R}}_l \mathrm{e}^{-2i\pi l\nu} \tag{2.10}$$

so that we can write

$$\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H = \int_0^1 \mathbf{d}_L(\nu) \mathbf{d}_L^H(\nu) \otimes \widehat{\mathbf{S}}(\nu) \, d\nu. \tag{2.11}$$

The $M \times M$ matrix $\widehat{\mathbf{S}}(\nu)$ coincides with a lag window estimator of the spectral density of $(\mathbf{y}_n)_{n \in \mathbb{Z}}$. Evaluating the spectral norm of $\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H$ is easier than that of $\mathcal{W}_N \mathcal{W}_N^H$, because the spectral norm of $\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H$ is upper bounded by $\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}(\nu)\|$, a term that can be controlled using a discretization in the frequency domain and the epsilon net argument in $\mathbb{C}^M$ (see e.g. [38] for an introduction to the concept of epsilon net). In the reminder of this section, we first prove that $\|\mathcal{W}_N \mathcal{W}_N^H - \widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H\|$ is bounded with exponentially high probability and then establish that $\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}(\nu)\|$, and thus $\|\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H\|$ is also bounded with exponentially high probability.

We first state the following lemma, which will allow to reduce various suprema on the interval $[0,1]$ to the corresponding suprema on a finite grid of the same interval. This result is adapted from Zygmund [42], and was used in [41].

Lemma 2.1. *Let* $h(\nu) = \sum_{l=-(L-1)}^{L-1} h_l \mathrm{e}^{-2i\pi l\nu}$ *an order* $L-1$ *real valued trigonometric polynomial. Then, for each* $\nu_0 \in [0,1]$, $\delta > 0$, $K \geq 2(1+\delta)(L-1)$, *we define* $\nu_k = \nu_0 + k/K$ *for* $k = 0, \ldots, K$. *Then, it holds that*

$$\max_{\nu \in [0,1]} |h(\nu)| \leq \left(1 + \frac{1}{\delta}\right) \max_{k=0,\ldots,K} |h(\nu_k)|. \tag{2.12}$$

We now compare the spectral norms of $\mathcal{W}_N \mathcal{W}_N^H$ and $\widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H$.

Proposition 2.1. *Let* α *denote a large enough constant. Under Assumptions 1-3 and 5, it holds that*

$$\mathbb{P}\left(\|\mathcal{W}_N \mathcal{W}_N^H - \widetilde{\mathcal{W}}_N \widetilde{\mathcal{W}}_N^H\| > \alpha\right) \leq \kappa_1 L \exp\left(-\kappa_2 M \alpha\right). \tag{2.13}$$

for two nice constants κ_1 *and* κ_2.

Proof. We drop all the subindexes N from all the matrices for clarity of exposition. Matrix $\mathcal{W}\mathcal{W}^H$ is equal to $\mathcal{W}\mathcal{W}^H = \mathcal{W}_1\mathcal{W}_1^H + (\mathcal{W}_{2,1} + \mathcal{W}_{2,2})(\mathcal{W}_{2,1} + \mathcal{W}_{2,2})^H$ while $\widetilde{\mathcal{W}}\widetilde{\mathcal{W}}^H = \mathcal{W}_0\mathcal{W}_0^H + \mathcal{W}_1\mathcal{W}_1^H + \mathcal{W}_{2,1}\mathcal{W}_{2,1}^H$. Therefore,

$$\mathcal{W}\mathcal{W}^H - \widetilde{\mathcal{W}}\widetilde{\mathcal{W}}^H = \mathcal{W}_{2,2}\mathcal{W}_{2,2}^H + \mathcal{W}_{2,2}\mathcal{W}_{2,1}^H + \mathcal{W}_{2,1}\mathcal{W}_{2,2}^H - \mathcal{W}_0\mathcal{W}_0^H.$$

In order to establish (2.13), we have to show that $\mathbb{P}(\|\mathcal{W}_{2,i}\mathcal{W}_{2,j}^H\| > \alpha)$, $i,j = 1,2$, and $\mathbb{P}(\|\mathcal{W}_0\mathcal{W}_0^H\| > \alpha)$ decrease at the same rate as the right hand side of (2.13). We just establish this property for matrix $\mathcal{W}_0\mathcal{W}_0^H$, or equivalently for matrix $\widetilde{\mathcal{W}}_0\widetilde{\mathcal{W}}_0^H$, where $\widetilde{\mathcal{W}}_0$ is defined as

$$\widetilde{\mathcal{W}}_0 = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{y}_1 & 0 & \dots & 0 & 0 \\ \mathbf{y}_2 & \mathbf{y}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ \mathbf{y}_{L-2} & \mathbf{y}_{L-3} & \ddots & \mathbf{y}_1 & 0 \\ \mathbf{y}_{L-1} & \mathbf{y}_{L-2} & \dots & \mathbf{y}_2 & \mathbf{y}_1 \end{pmatrix}.$$

It is easily seen that $\widetilde{\mathcal{W}}_0$ can be expressed as

$$\widetilde{\mathcal{W}}_0 = \sqrt{\frac{L}{N}} \int_0^1 \mathbf{d}_{L-1}(\nu)\mathbf{d}_{L-1}^H(\nu) \otimes \boldsymbol{\xi}_{L,y}(\nu)\, d\nu \qquad (2.14)$$

where $\boldsymbol{\xi}_{L,y}(\nu)$ is an M-dimensional column vector defined as $\boldsymbol{\xi}_{L,y}(\nu) = \frac{1}{\sqrt{L}}\sum_{l=0}^{L-2} \mathbf{y}_{l+1}\mathrm{e}^{-2i\pi l\nu}$. The matrix version of the Cauchy-Schwarz inequality in (1.31) with $\mathbf{U}(\nu) = \sqrt{\frac{L}{N}}\mathbf{d}_{L-1}(\nu) \otimes \boldsymbol{\xi}_{L,y}(\nu)$ and $\mathbf{V}(\nu) = \mathbf{d}_{L-1}(\nu)$ leads immediately to

$$\widetilde{\mathcal{W}}_0\widetilde{\mathcal{W}}_0^H \leq \frac{L}{N} \int_0^1 \mathbf{d}_{L-1}(\nu)\mathbf{d}_{L-1}^H(\nu) \otimes \boldsymbol{\xi}_{L,y}(\nu)\boldsymbol{\xi}_{L,y}^H(\nu)\, d\nu.$$

From this, we obtain immediately that

$$\|\widetilde{\mathcal{W}}_0\widetilde{\mathcal{W}}_0^H\| \leq \sup_{\nu\in[0,1]} \frac{L}{N} \|\boldsymbol{\xi}_{L,y}(\nu)\|^2.$$

Next, observe that $\nu \to \frac{L}{N}\|\boldsymbol{\xi}_{L,y}(\nu)\|^2$ is a real valued trigonometric polynomial of order $L-2$. Therefore, if K, δ and the points $(\nu_k)_{k=0,\dots,K}$ are given as in Lemma 2.1, it holds that

$$\left\|\widetilde{\mathcal{W}}_0\widetilde{\mathcal{W}}_0^H\right\| \leq \left(1 + \frac{1}{\delta}\right) \sup_{k=0,\dots,K} \frac{L}{N}\|\boldsymbol{\xi}_{L,y}(\nu_k)\|^2.$$

Noting that $K = \mathcal{O}(L)$, it is sufficient to evaluate $\mathbb{P}\left(\frac{L}{N}\|\boldsymbol{\xi}_{L,y}(\nu)\|^2 > \eta\right)$ for some fixed ν and some well chosen constant η, and then use the union bound. Observe first that we can express

$$\frac{L}{N}\|\boldsymbol{\xi}_{L,y}(\nu)\|^2 = \frac{ML}{N}\frac{1}{M}\sum_{m=1}^M |\xi_{L,y_m}(\nu)|^2$$

where $\xi_{L,y_m}(\nu)$, $m = 1, \ldots, M$ are components of $\boldsymbol{\xi}_{L,y}(\nu)$. These are mutually independent complex Gaussian random variables, so that we can use the Hanson-Wright inequality in order to establish an exponential concentration inequality on $\frac{L}{N}\|\xi_{L,y}(\nu)\|^2$.

In order to use (1.33), we remark that for each m, $\xi_{L,y_m}(\nu)$ can be written as $\xi_{L,y_m}(\nu) = \left(\mathbb{E}|\xi_{L,y_m}(\nu)|^2\right)^{1/2} x_m$ where $x_1, \ldots, x_M$ are $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. random variables. If $\mathbf{x} = (x_1, \ldots, x_M)$, $\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2$ can be written as

$$\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2 = \mathbf{x}^H \boldsymbol{\Xi}(\nu)\mathbf{x}$$

where $\boldsymbol{\Xi}(\nu)$ is the $M \times M$ diagonal matrix with mth diagonal entry equal to

$$[\boldsymbol{\Xi}(\nu)]_{m,m} = \frac{1}{M}\mathbb{E}|\xi_{L,y_m}(\nu)|^2.$$

In order to evaluate $\|\boldsymbol{\Xi}(\nu)\|$ and $\|\boldsymbol{\Xi}(\nu)\|_F^2$, we have to study the behaviour of $\mathbb{E}|\xi_{L,y_m}(\nu)|^2$, i.e. the expectation of the periodogram of the sequence $y_{m,1}, \ldots, y_{m,L-1}$. The following result establishes that the diagonal entries of this matrix are equal to scaled versions of the spectral densities $\frac{1}{M}\mathcal{S}_m(\nu), m = 1, \ldots, M$, up to an error that decays as $\mathcal{O}\left(\frac{1}{ML^{\min(1,\gamma_0)}}\right)$.

Lemma 2.2. *Under Assumptions 1 and 5, $\mathbb{E}|\xi_{L,y_m}(\nu)|^2$ can be written as* $\mathbb{E}|\xi_{L,y_m}(\nu)|^2 = \mathcal{S}_m(\nu) + \epsilon_{m,L}(\nu)$ *where* $\epsilon_{m,L}(\nu)$ *verifies*

$$|\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{(L-1)^{\min(1,\gamma_0)}} \tag{2.15}$$

for each m and for some nice constant κ.

Lemma 2.2 is proved in Appendix A.

This lemma implies that there exists a nice constant κ for which $\mathbb{E}|\xi_{L,y_m}(\nu)|^2 \leq \kappa$ for each ν and each m and $L > 1$. Therefore, if $\boldsymbol{\Xi}(\nu)$ is the above mentioned diagonal matrix, $\boldsymbol{\Xi}(\nu)$ verifies $\|\boldsymbol{\Xi}(\nu)\| \leq \frac{\kappa}{M}$ and $\|\boldsymbol{\Xi}(\nu)\|_F^2 \leq \frac{\kappa^2}{M}$. Consider a nice constant $\eta > 2\kappa$. Then,

$$\mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2 > \eta\right) \leq \mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2 - \mathbb{E}|\xi_{L,y_m}(\nu)|^2 > \eta - \kappa\right)$$

$$\leq \mathbb{P}\left(\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2 - \mathbb{E}|\xi_{L,y_m}(\nu)|^2 > \eta/2\right).$$

As $\min\left(\frac{M(\eta/2)}{\kappa}, \frac{M(\eta/2)^2}{\kappa^2}\right) = \frac{M(\eta/2)}{\kappa}$ and $ML/N \to c_\star$, the Hanson-Wright inequality leads to

$$\mathbb{P}\left(\frac{ML}{N}\frac{1}{M}\sum_{m=1}^{M}|\xi_{L,y_m}(\nu)|^2 > \eta\right) \leq \kappa_1 \exp(-M\kappa_2\,\eta)$$

for some nice constants κ_1 and κ_2. Recalling that $K = \mathcal{O}(L)$, and using the union bound to evaluate $\mathbb{P}\left(\sup_{k=0,\dots,K} \frac{L}{N}\|\boldsymbol{\xi}_{K,y}(\nu_k)\|^2 > \eta\right)$, we have shown that, if α is large enough, there exist two nice constants κ_1 and κ_2 such that

$$\mathbb{P}(\|\widetilde{\mathcal{W}}_0 \widetilde{\mathcal{W}}_0^H\| > \alpha) \leq L\,\kappa_1 \exp(-M\kappa_2\alpha).$$

Following the same approach to evaluate the other terms $\|\mathcal{W}_{N,i}\mathcal{W}_{N,j}^H\|$, we can conclude that (2.13) is established. □

As a consequence of Proposition 2.1, the evaluation of $\mathbb{P}(\|\mathcal{W}_N\mathcal{W}_N^H\| > \alpha)$ can be alternatively formulated in terms of the evaluation of $\mathbb{P}(\|\widetilde{\mathcal{W}}_N\widetilde{\mathcal{W}}_N^H\| > \alpha)$.

2.1.2. *Controlling the spectral norm of* $\widetilde{\mathcal{W}}_N\widetilde{\mathcal{W}}_N^H$

In order to establish the fact that $\|\widetilde{\mathcal{W}}_N\widetilde{\mathcal{W}}_N^H\|$ is bounded with exponentially large probability, we use the expression in (2.11) and remark that

$$\left\|\widetilde{\mathcal{W}}_N\widetilde{\mathcal{W}}_N^H\right\| \leq \sup_{\nu\in[0,1]} \|\widehat{\mathbf{S}}(\nu)\|.$$

In the following, we thus control the spectral norm of $\widehat{\mathbf{S}}(\nu)$. In particular, we have the following result.

Proposition 2.2. *If α is a large enough constant, under Assumptions 1-3 and 5, it holds that*

$$\mathbb{P}\left(\sup_{\nu\in[0,1]} \left\|\widehat{\mathbf{S}}(\nu)\right\| > \alpha\right) < \kappa_1 L \exp\left(-\kappa_2 M\alpha\right) \tag{2.16}$$

for some nice constants κ_1 and κ_2.

Proof. We denote by $\widehat{\mathbf{S}}^\circ(\nu)$ the centered matrix $\widehat{\mathbf{S}}^\circ(\nu) = \widehat{\mathbf{S}}(\nu) - \mathbb{E}\widehat{\mathbf{S}}(\nu)$. We first notice that

$$\sup_\nu \|\widehat{\mathbf{S}}(\nu)\| \leq \sup_\nu \|\mathbb{E}\widehat{\mathbf{S}}(\nu)\| + \sup_\nu \|\widehat{\mathbf{S}}^\circ(\nu)\|$$

and work on the two terms separately. First, we prove that $\sup_\nu \|\mathbb{E}\widehat{\mathbf{S}}(\nu)\|$ is bounded. Indeed, it is clear that $\mathbb{E}(\widehat{\mathbf{S}}(\nu)) = \sum_{l=-(L-1)}^{L-1}(1 - \frac{|l|}{L})\mathbf{R}(l)\mathrm{e}^{-2i\pi l\nu}$ where $\mathbf{R}(l) = \mathbb{E}(\mathbf{y}_{n+l}\mathbf{y}_n^H)$ is the autocovariance matrix of $\mathbf{y}_n$ at lag l. Since the components of $\mathbf{y}_n$ are independent time series, matrix $\mathbf{R}(l)$ coincides with $\mathbf{R}(l) = \mathrm{Diag}\left((r_m(l))_{m=1,\dots,M}\right)$. Therefore,

$$\|\mathbb{E}\widehat{\mathbf{S}}(\nu)\| \leq \sup_{m=1,\dots,M} \sum_{l=-(L-1)}^{L-1} |r_m(l)| \leq \sup_{m\geq 1} \sum_{l\in\mathbb{Z}} |r_m(l)|.$$

28

Condition (1.13) thus implies that $\sup_\nu \|\mathbb{E}\widehat{\mathbf{S}}(\nu)\| < +\infty$. Therefore, in order to establish (2.16), we need to study $\mathbb{P}(\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| > \alpha)$ for α sufficiently large.

We first show that the study of the supremum of $\|\widehat{\mathbf{S}}^\circ(\nu)\|$ over $[0,1]$ can be reduced to the supremum over a discrete grid with $\mathcal{O}(L)$ elements. The idea is to make use Lemma 2.1 by conveniently expressing $\|\widehat{\mathbf{S}}^\circ(\nu)\|$ in terms of trigonometric polynomials.

Lemma 2.3. *We consider δ, K, and $(\nu_k)_{k=0,\ldots,K}$ as in Lemma 2.1. Then, the following result holds:*

$$\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| \leq \left(1 + \frac{1}{\delta}\right) \sup_{k=0,\ldots,K} \|\widehat{\mathbf{S}}^\circ(\nu_k)\|. \tag{2.17}$$

Proof. We will first verify that

$$\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| = \sup_{\nu \in [0,1], \mathbf{h} \in \mathbb{S}^{M-1}} \left| \mathbf{h}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h} \right| \tag{2.18}$$

where $\mathbb{S}^{M-1}$ is the unit sphere in $\mathbb{C}^M$. We remark that, because of the continuity of the spectral norm as well as the continuity of both true and estimated spectral densities, there exists a certain $\widehat{\nu}$ that achieves the supremum on the left hand side of (2.18), that is $\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| = \|\widehat{\mathbf{S}}^\circ(\widehat{\nu})\|$. Moreover, for such given $\widehat{\nu}$, there exists a $\mathbf{h}_{\widehat{\nu}} \in \mathbb{S}^{M-1}$ for which $\|\widehat{\mathbf{S}}^\circ(\widehat{\nu})\| = |\mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\widehat{\nu})\mathbf{h}_{\widehat{\nu}}|$. In other words, $\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\|$ coincides with $|\mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\widehat{\nu})\mathbf{h}_{\widehat{\nu}}|$. Hence, we obtain that the left hand side of (2.18) is less than the right hand side of (2.18). The converse inequality is obvious.

Using a similar continuity argument, we can readily see that

$$\sup_{\nu \in [0,1], \mathbf{h} \in \mathbb{S}^{M-1}} |\mathbf{h}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}| = |\mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\widehat{\nu})\mathbf{h}_{\widehat{\nu}}|$$

also coincides with $\sup_{\nu \in [0,1]} |\mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}_{\widehat{\nu}}|$. The function $\nu \to \mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}_{\widehat{\nu}}$ is a real valued trigonometric polynomial of order $L-1$. Therefore, Lemma 2.1 implies that

$$\sup_{\nu \in [0,1]} \left| \mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}_{\widehat{\nu}} \right| \leq \left(1 + \frac{1}{\delta}\right) \sup_{k=0,\ldots,K} \left| \mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu_k)\mathbf{h}_{\widehat{\nu}} \right|.$$

Since $|\mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu_k)\mathbf{h}_{\widehat{\nu}}| \leq \|\widehat{\mathbf{S}}^\circ(\nu_k)\|$, we have shown that

$$\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| = \sup_{\nu \in [0,1]} \left| \mathbf{h}_{\widehat{\nu}}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}_{\widehat{\nu}} \right| \leq \left(1 + \frac{1}{\delta}\right) \sup_{k=0,\ldots,K} \|\widehat{\mathbf{S}}^\circ(\nu_k)\|.$$

This establishes (2.17). $\qquad\square$

We now complete the proof of (2.16) in Proposition 2.2. The union bound leads to

$$\mathbb{P}\left(\sup_{\nu \in [0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| > \alpha_N\right) \leq \sum_{k=0}^{K} \mathbb{P}\left(\|\widehat{\mathbf{S}}^\circ(\nu_k)\| > \frac{\delta}{1+\delta}\,\alpha_N\right). \tag{2.19}$$

Thus, we only need to evaluate $\mathbb{P}(\|\widehat{\mathbf{S}}^\circ(\nu)\| > \eta_N)$, where ν is a fixed frequency and where $\eta_N = \frac{\delta}{1+\delta}\,\alpha_N$. For this, we use the epsilon net argument in $\mathbb{C}^M$. We recall that an epsilon net $\mathcal{N}_\epsilon$ of $\mathbb{C}^M$ is a finite set of unit norm vectors of $\mathbb{C}^M$ having the property that for each $\mathbf{g} \in \mathbb{S}^{M-1}$, there exists an $\mathbf{h} \in \mathcal{N}_\epsilon$ such that $\|\mathbf{g} - \mathbf{h}\| \leq \epsilon$. It is well known that the cardinal $|\mathcal{N}_\epsilon|$ is upper bounded by $\left(\frac{\kappa}{\epsilon}\right)^{2M}$ for some nice constant κ. We consider such an epsilon net $\mathcal{N}_\epsilon$ and denote by $\widehat{\mathbf{h}}$ a vector of $\mathbb{S}^{M-1}$ for which $\|\widehat{\mathbf{S}}^\circ(\nu)\| = |\widehat{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widehat{\mathbf{h}}|$, and consider a vector $\widetilde{\mathbf{h}} \in \mathcal{N}_\epsilon$ such that $\|\widehat{\mathbf{h}} - \widetilde{\mathbf{h}}\| \leq \epsilon$. We express $\widetilde{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widetilde{\mathbf{h}}$ as

$$\widetilde{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widetilde{\mathbf{h}} = \left(\widehat{\mathbf{h}} + \widetilde{\mathbf{h}} - \widehat{\mathbf{h}}\right)^H \widehat{\mathbf{S}}^\circ(\nu) \left(\widehat{\mathbf{h}} + \widetilde{\mathbf{h}} - \widehat{\mathbf{h}}\right).$$

Using the triangular inequality, we obtain that

$$\left|\widetilde{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widetilde{\mathbf{h}}\right| \geq \left|\widehat{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widehat{\mathbf{h}}\right| - 2\left|(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H \widehat{\mathbf{S}}^\circ(\nu)\widehat{\mathbf{h}}\right| - \left|(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H \widehat{\mathbf{S}}^\circ(\nu)(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H\right|.$$

Since $\widetilde{\mathbf{h}} \in \mathbb{S}^{M-1}$ and $\|\widehat{\mathbf{h}} - \widetilde{\mathbf{h}}\| \leq \epsilon$, we can write

$$\left|(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H \widehat{\mathbf{S}}^\circ(\nu)\widehat{\mathbf{h}}\right| \leq \|\widehat{\mathbf{S}}^\circ(\nu)(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})\| \leq \epsilon\|\widehat{\mathbf{S}}^\circ(\nu)\|$$

together with $\left|(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H \widehat{\mathbf{S}}^\circ(\nu)(\widetilde{\mathbf{h}} - \widehat{\mathbf{h}})^H\right| \leq \epsilon^2\|\widehat{\mathbf{S}}^\circ(\nu)\|$. This implies that

$$\left|\widetilde{\mathbf{h}}^H \widehat{\mathbf{S}}^\circ(\nu)\widetilde{\mathbf{h}}\right| \geq (1 - 2\epsilon - \epsilon^2)\|\widehat{\mathbf{S}}^\circ(\nu)\|.$$

In the following, we assume that ϵ satisfies $1 - 2\epsilon - \epsilon^2 > 0$. Therefore, using again the union bound, we obtain that

$$\mathbb{P}\left(\|\widehat{\mathbf{S}}^\circ(\nu)\| > \eta_N\right) \leq \sum_{\mathbf{h} \in \mathcal{N}_\epsilon} \mathbb{P}\left(\left|\mathbf{h}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}\right| \geq (1 - 2\epsilon - \epsilon^2)\eta_N\right). \qquad (2.20)$$

In order to evaluate $\mathbb{P}(|\mathbf{h}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}| \geq (1 - 2\epsilon - \epsilon^2)\eta_N)$ for each unit norm vector $\mathbf{h}$, we denote by z_n the scalar time series defined by $z_n = \mathbf{h}^H \mathbf{y}_n$. Then, the quadratic form $\mathbf{h}^H \widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}$ coincides with $\widehat{s}_z(\nu) - \mathbb{E}\widehat{s}_z(\nu)$ where $\widehat{s}_z(\nu)$ represents the lag-window estimator of the spectral density of z defined by $\widehat{s}_z(\nu) = \sum_{l=-(L-1)}^{L-1} \widehat{r}_z(l)\mathrm{e}^{-2i\pi l\nu}$. Here, $\widehat{r}_z(l)$ is the standard empirical estimate of the autocovariance coefficient of z at lag l. We denote by $\mathbf{z}$ the N–dimensional vector $\mathbf{z} = (z_1, \ldots, z_N)^T$. As is well known, $\widehat{s}_z(\nu)$ can be expressed as

$$\widehat{s}_z(\nu) = \int_0^1 w(\nu - \mu)\frac{1}{N}\left|\sum_{n=0}^{N-1} z_{n+1}\mathrm{e}^{-2i\pi n\mu}\right|^2 d\mu \qquad (2.21)$$

where $w(\mu)$ is the Fourier transform of the rectangular window $\mathbb{I}_{l \in \{-(L-1),\ldots,L-1\}}$. The expression in (2.21) can also be written as a quadratic form of vector $\mathbf{z}$:

$$\widehat{s}_z(\nu) = \mathbf{z}^H \left(\frac{1}{N}\int_0^1 w(\nu - \mu)\,\mathbf{d}_N(\mu)\mathbf{d}_N^H(\mu)\,d\mu\right)\mathbf{z}. \qquad (2.22)$$

where we recall that $\mathbf{d}_N(\mu)$ is defined by (1.17). If $\mathbf{R}_z$ represents the covariance matrix of vector $\mathbf{z}$, $\mathbf{z}$ can be written as $\mathbf{z} = \mathbf{R}_z^{1/2}\mathbf{x}$ for some $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_N)$ distributed random vector $\mathbf{x}$. Therefore, if we denote by $\boldsymbol{\Omega}$ the $N \times N$ matrix defined by

$$\boldsymbol{\Omega} = \mathbf{R}_z^{1/2}\, \frac{1}{N} \int_0^1 w(\nu - \mu)\, \mathbf{d}_N(\mu)\mathbf{d}_N^H(\mu)\, d\mu\, \mathbf{R}_z^{1/2},$$

the quantity $\widehat{s}_z(\nu) - \mathbb{E}\widehat{s}_z(\nu)$ can be written as $\widehat{s}_z(\nu) - \mathbb{E}\widehat{s}_z(\nu) = \mathbf{x}^H\boldsymbol{\Omega}\mathbf{x} - \mathbb{E}\mathbf{x}^H\boldsymbol{\Omega}\mathbf{x}$. Therefore,

$$P\left(|\widehat{s}_z(\nu) - \mathbb{E}\widehat{s}_z(\nu)| > (1 - 2\epsilon - \epsilon^2)\eta\right)$$

can be evaluated using the Hanson-Wright inequality (1.33). This requires the evaluation of the spectral and the Frobenius norm of $\boldsymbol{\Omega}$. Observe that we can express $\boldsymbol{\Omega} = \mathbf{R}_z^{1/2}\boldsymbol{\Omega}_w\mathbf{R}_z^{1/2}$ where $\boldsymbol{\Omega}_w$ is a Toeplitz matrix defined as

$$\boldsymbol{\Omega}_w = \frac{1}{N} \int_0^1 w(\nu - \mu)\, \mathbf{d}_N(\mu)\mathbf{d}_N^H(\mu)\, d\mu.$$

It is easy to check that the spectral norm of $\mathbf{R}_z$ is uniformly bounded. Moreover, the spectral norm of $\boldsymbol{\Omega}_w$ is bounded by $\frac{1}{N}\sup_\nu |w(\nu)| = L/N$. Therefore, $\|\boldsymbol{\Omega}\| \leq \kappa\frac{L}{N}$ for some nice constant κ. In order to evaluate the Frobenius norm of $\boldsymbol{\Omega}$, observe that $\boldsymbol{\Omega}_w$ is band Toeplitz matrix with entries given by $(\boldsymbol{\Omega}_w)_{k,l} = \frac{1}{N}e^{2i\pi(k-l)\nu}\mathbb{I}_{|k-l|\leq L-1}$. Therefore, $\|\Omega_w\|_F^2 \leq \kappa\frac{L}{N}$, which implies that $\|\boldsymbol{\Omega}\|_F^2 \leq \kappa\frac{L}{N}$. Consequently, the Hanson-Wright inequality in (1.33) implies that, if η is large enough,

$$\mathbb{P}\left(\left|\mathbf{h}^H\widehat{\mathbf{S}}^\circ(\nu)\mathbf{h}\right| \geq (1 - 2\epsilon - \epsilon^2)\eta\right) \leq \kappa_1 \exp\left(-\kappa_2 M\eta\right)$$

where we have introduced two nice constants κ_1 and κ_2. Recalling that $|\mathcal{N}_\epsilon| \leq \left(\frac{\kappa}{\epsilon}\right)^{2M}$, the union bound (2.20) implies that

$$\mathbb{P}\left(\|\widehat{\mathbf{S}}^\circ(\nu)\| > \eta\right) \leq \left(\frac{\kappa}{\epsilon}\right)^{2M} \kappa_1 \exp\left(-\kappa_2 M\alpha\right)$$

The right hand side of the above inequality can clearly be bounded by $\kappa_3 \exp\left(-\kappa_4 M\alpha\right)$ for α large enough, where κ_3 and κ_4 are two new nice constants. Finally, (2.19) leads to

$$\mathbb{P}\left(\sup_{\nu\in[0,1]} \|\widehat{\mathbf{S}}^\circ(\nu)\| > \alpha/2\right) \leq \kappa_1 L \exp\left(-\kappa_2 M\alpha\right)$$

for N sufficiently large and two nice constants κ_1, κ_2. This completes the proof of Proposition 2.2. $\square$

As a direct sequence of Propositions 2.1 and 2.2, we have the following corollary.

Corollary 2.1. *For each α larger than a certain positive constant, then, it holds that*

$$\mathbb{P}(\|\mathcal{W}_N\mathcal{W}_N^H\| > \alpha) \leq \kappa_1\, L \exp(-\kappa_2 M\alpha) \tag{2.23}$$

$$\mathbb{P}(\|\widehat{\mathcal{R}}_L\| > \alpha) \leq \kappa_1\, L \exp(-\kappa_2 M\alpha) \tag{2.24}$$

for some nice constants κ_1 and κ_2. Moreover, $\|\widehat{\mathcal{R}}_L\|$ satisfies

$$\|\widehat{\mathcal{R}}_L\| \prec 1. \tag{2.25}$$

2.2. Evaluation of the behaviour of $\|\mathbf{Bdiag}(\widehat{\mathcal{R}}_L) - \mathbf{Bdiag}(\mathcal{R}_L)\|$

Recall that $\mathcal{R}_{m,L}$, $m = 1, \ldots, M$, denote the $L \times L$ diagonal blocks of the matrix $\mathrm{Bdiag}(\mathcal{R}_L)$. We will denote by $\widehat{\mathcal{R}}_{m,L}$ the mth $L \times L$ diagonal block of $\widehat{\mathcal{R}}_L$. In this section, we establish that

$$\|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right). \tag{2.26}$$

Note first that we can express $\widehat{\mathcal{R}}_{m,L}$ as the empirical estimate of $\mathcal{R}_{m,L}$, that is

$$\widehat{\mathcal{R}}_{m,L} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_{m,n}^L \left(\mathbf{y}_{m,n}^L\right)^H$$

or equivalently by $\widehat{\mathcal{R}}_{m,L} = \mathbf{W}_N^m \left(\mathbf{W}_N^m\right)^H$ where $\mathbf{W}_N^m$ is the $L \times N$ matrix defined by

$$\mathbf{W}_N^m = \frac{1}{\sqrt{N}} \left(\mathbf{y}_{m,1}^L, \ldots, \mathbf{y}_{m,N}^L\right).$$

The arguments used in this section are based on the techniques used in Section 2.1. Therefore, we just provide a sketch of proof of (2.26) based on the same two steps as above: first, we approximate $\mathbf{W}_N^m \left(\mathbf{W}_N^m\right)^H$ with a Toeplitz matrix and then study the equivalent Toeplitz version of (2.26).

2.2.1. Modifying $\mathbf{W}_N^m \left(\mathbf{W}_N^m\right)^H$ into a Toeplitz matrix

We prove here that $\mathbf{W}_N^m \left(\mathbf{W}_N^m\right)^H$ can be approximated as the Toeplitz matrix $\widetilde{\mathbf{W}}_N^m (\widetilde{\mathbf{W}}_N^m)^H$ where $\widetilde{\mathbf{W}}_N^m$ is obtained by replacing vectors $(\mathbf{y}_n)_{n=1,\ldots,N}$ by the scalars $(y_{m,n})_{n=1,\ldots,N}$ in the definition of matrix $\mathcal{W}_N$ in (2.9) above. In particular, it holds that

$$\widetilde{\mathbf{W}}_N^m \left(\widetilde{\mathbf{W}}_N^m\right)^H = \int_0^1 \widehat{\mathcal{S}}_m(\nu) \mathbf{d}_L(\nu) \mathbf{d}_L^H(\nu) \, d\nu \tag{2.27}$$

where $\widehat{\mathcal{S}}_m(\nu)$ represents the mth diagonal entry of the lag window estimator (2.10). More specifically, following the proof of Proposition 2.1, we justify that

$$\left\|\mathbf{W}_N^m \left(\mathbf{W}_N^m\right)^H - \widetilde{\mathbf{W}}_N^m \left(\widetilde{\mathbf{W}}_N^m\right)^H\right\| \prec \frac{1}{M}. \tag{2.28}$$

To verify (2.28), we drop the dependence on N of all matrices to simplify the notation and remark that

$$\mathbf{W}^m(\mathbf{W}^m)^H - \widetilde{\mathbf{W}}^m(\widetilde{\mathbf{W}}^m)^H = \mathbf{W}_{2,2}^m(\mathbf{W}_{2,2}^m)^H + \mathbf{W}_{2,2}^m(\mathbf{W}_{2,1}^m)^H + \mathbf{W}_{2,1}^m(\mathbf{W}_{2,2}^m)^H - \mathbf{W}_0^m(\mathbf{W}_0^m)^H$$

32

where the various matrices of the right hand side are obtained by replacing vectors $(\mathbf{y}_n)_{n=1,\dots,N+L-1}$ in the definition of matrices $\mathcal{W}_{2,2}, \mathcal{W}_{2,1}, \mathcal{W}_0$ used in Section 2.1 by the scalars $(y_{m,n})_{n=1,\dots,N+L-1}$. In order to verify (2.28), we just briefly check that

$$\left\| \mathbf{W}_0^m \left(\mathbf{W}_0^m \right)^H \right\| \prec \frac{1}{M}$$

or equivalently (after proper column permutation of $(\mathbf{W}_0^m)$ that

$$\left\| \widetilde{\mathbf{W}}_0^m \left(\widetilde{\mathbf{W}}_0^m \right)^H \right\| \prec \frac{1}{M}$$

where $\widetilde{\mathbf{W}}_0^m$ is defined by

$$\widetilde{\mathbf{W}}_0^m = \sqrt{\frac{L}{N}} \int_0^1 \mathbf{d}_{L-1}(\nu) \mathbf{d}_{L-1}^H(\nu) \xi_{L,y_m}(\nu) \, d\nu.$$

As in Section 2.1, we notice that the matrix-valued Cauchy-Schwarz inequality in (1.31) with $\mathbf{U}(\nu) = \sqrt{\frac{L}{N}} \mathbf{d}_{L-1}(\nu) \xi_{L,y_m}(\nu)$ and $\mathbf{V}(\nu) = \mathbf{d}_{L-1}(\nu)$ implies that

$$\widetilde{\mathbf{W}}_0^m (\widetilde{\mathbf{W}}_0^m)^H \leq \frac{L}{N} \int_0^1 \mathbf{d}_{L-1}(\nu) \mathbf{d}_{L-1}^H(\nu) |\xi_{L,y_m}(\nu)|^2 \, d\nu.$$

This allow us to establish that

$$\left\| \widetilde{\mathbf{W}}_0^m (\widetilde{\mathbf{W}}_0^m)^H \right\| \leq \sup_{\nu \in [0,1]} \frac{L}{N} |\xi_{L,y_m}(\nu)|^2.$$

By Lemma 2.1 we know that the supremum can be replaced by a maximum over $\mathcal{O}(L)$ points, so that by Lemma 1.1 it is sufficient to establish that

$$\frac{L}{N} |\xi_{L,y_m}(\nu)|^2 \prec \frac{1}{M}$$

for some fixed ν. Following the same reasoning as in Section 2.1, a direct application of the Hanson-Wright inequality shows that $|\xi_{L,y_m}(\nu)|^2 \prec 1$ for any fixed ν, from where the result follows.

2.2.2. *Studying the Toeplitz equivalent of (2.26)*

In order to prove (2.26), it thus remains to establish that

$$\left\| \widetilde{\mathbf{W}}_N^m \left(\widetilde{\mathbf{W}}_N^m \right)^H - \mathcal{R}_{m,L} \right\| \prec \max \left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}} \right)$$

Noting that $\widetilde{\mathbf{W}}_N^m (\widetilde{\mathbf{W}}_N^m)^H - \mathcal{R}_{m,L}$ is the $L \times L$ Toeplitz matrix associated to the symbol $\widehat{\mathcal{S}}_m(\nu) - \mathcal{S}_m(\nu)$, and using Lemma 2.1, it is sufficient to prove that

$$\left| \widehat{\mathcal{S}}_m(\nu) - \mathcal{S}_m(\nu) \right| \prec \max \left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}} \right)$$

for each ν. In order to see this, we write $\widehat{\mathcal{S}}_m(\nu) - \mathcal{S}_m(\nu)$ as

$$\widehat{\mathcal{S}}_m(\nu) - \mathcal{S}_m(\nu) = \widehat{\mathcal{S}}_m(\nu) - \mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) + \mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) - \mathcal{S}_m(\nu).$$

The bias $\mathbb{E}(\widehat{\mathcal{S}}_m(\nu)) - \mathcal{S}_m(\nu)$ is equal to

$$\mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) - \mathcal{S}_m(\nu) = -\sum_{|l| \geq L} r_m(l)\mathrm{e}^{-2i\pi l\nu} - \frac{1}{N} \sum_{l=-(L-1)}^{L-1} |l| r_m(l)\mathrm{e}^{-2i\pi l\nu}.$$

An easy adaptation of the proof of Lemma 2.2 in Appendix A establishes that

$$\left|\mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) - \mathcal{S}_m(\nu)\right| \leq \kappa \left(\frac{1}{L^{\gamma_0}} + \frac{L^{(1-\gamma_0)_+}}{N}\right) \tag{2.29}$$

for some nice constant κ, where $(\cdot)_+ = \max(\cdot, 0)$. This implies that

$$\sup_{m=1,\ldots,M} \sup_{\nu} \left|\mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) - \mathcal{S}_m(\nu)\right| \leq \kappa \max\left(\frac{1}{L^{\gamma_0}}, \frac{1}{M}\right) \tag{2.30}$$

for some nice constant κ. In order to study the term $\widehat{\mathcal{S}}_m(\nu) - \mathbb{E}(\widehat{\mathcal{S}}_m(\nu))$, we remark that it can be written as

$$\widehat{\mathcal{S}}_m(\nu) - \mathbb{E}\left(\widehat{\mathcal{S}}_m(\nu)\right) = \mathbf{e}_m^T \widehat{\mathbf{S}}^\circ(\nu)\mathbf{e}_m$$

where $\mathbf{e}_m$ is the mth vector of the canonical basis of $\mathbb{C}^M$. Using the Hanson-Wright inequality as in Section 2.1, we obtain immediately that for each ν and for each m, there exist two nice constants κ_1 and κ_2 such that

$$\mathbb{P}\left(|\mathbf{e}_m^T \widehat{\mathbf{S}}^\circ(\nu)\mathbf{e}_m| > \alpha_N\right) \leq \kappa_1 \exp(-\kappa_2 M\alpha_N^2)$$

where $(\alpha_N)_{N \geq 1}$ satisfies $\alpha_N \to 0$ and $M\alpha_N^2 \to +\infty$. In particular, the choice $\alpha_N = N^\epsilon/\sqrt{M}$ satisfies this property for all small enough $\epsilon > 0$, which allows to conclude that $|\mathbf{e}_m^T \widehat{\mathbf{S}}^\circ(\nu)\mathbf{e}_m| \prec M^{-1/2}$ for any fixed m and ν. However, noting again that $\mathbf{e}_m^T \widehat{\mathbf{S}}^\circ(\nu)\mathbf{e}_m$ is a real valued trigonometric polynomial, we see by Lemma 2.3 and Lemma 1.1 that $\sup_{m,\nu} |\mathbf{e}_m^T \widehat{\mathbf{S}}^\circ(\nu)\mathbf{e}_m| \prec M^{-1/2}$.

As a consequence of all the above, we have established that $\|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$, which directly implies that

$$\|\mathrm{Bdiag}(\widehat{\mathcal{R}}_L) - \mathrm{Bdiag}(\mathcal{R}_L)\| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right). \tag{2.31}$$

All these results are all the ingredients that we need in order to evaluate the spectral norm of the matrix

$$\mathbf{\Theta}_N = \widehat{\mathcal{R}}_{\mathrm{corr},L} - \overline{\mathcal{R}}_{\mathrm{corr},L} \tag{2.32}$$

which is carried out in the following section.

34

2.3. Evaluation of $\|\mathbf{\Theta}_N\| = \|\widehat{\mathcal{R}}_{\mathbf{corr},L} - \overline{\mathcal{R}}_{\mathbf{corr},L}\|$

We first precise that almost surely, all the matrices $(\hat{\mathcal{R}}_{m,L})_{m=1,\ldots,M(N),N\geq1}$ are invertible. To verify this, we remark that the random variable $\det(\hat{\mathcal{R}}_{m,L})$ is a differentiable function of the $2(N + L - 1)$ entries of the Gaussian vector $(\mathrm{Re}(y_{m,1},\ldots,y_{m,N+L-1}), \mathrm{Im}(y_{m,1},\ldots,y_{m,N+L-1}))$. Therefore, the probability distribution of $\det(\hat{\mathcal{R}}_{m,L})$ is absolutely continuous, and the event $\{\det(\hat{\mathcal{R}}_{m,L}) = 0\}$ has probability 0. Therefore, the union of the above events is also negligible, thus showing the almost sure invertibility of the matrices $(\hat{\mathcal{R}}_{m,L})_{m=1,\ldots,M(N),N\geq1}$.

Using the above definition of $\mathbf{\Theta}_N$, we are able to write

$$\mathbf{\Theta}_N = \widehat{\mathcal{B}}_L^{-1/2}\widehat{\mathcal{R}}_L\widehat{\mathcal{B}}_L^{-1/2} - \mathcal{B}_L^{-1/2}\widehat{\mathcal{R}}_L\mathcal{B}_L^{-1/2}$$
$$= (\widehat{\mathcal{B}}_L^{-1/2} - \mathcal{B}_L^{-1/2})\widehat{\mathcal{R}}_L\widehat{\mathcal{B}}_L^{-1/2} + \mathcal{B}_L^{-1/2}\widehat{\mathcal{R}}_L(\widehat{\mathcal{B}}_L^{-1/2} - \mathcal{B}_L^{-1/2}) \qquad (2.33)$$

We have shown above that $\|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$. Our first objective here is to show that $\|\widehat{\mathcal{R}}_{m,L}^{-1/2} - \mathcal{R}_{m,L}^{-1/2}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$. For this, we use perturbation theory of Hermitian matrices arguments (see e.g. [22, Sec. 2, Ch. 1 and Sec. 1, Ch. 2]) that will also be needed in Section 3.

We first recall that Assumption 3 implies that for each N, matrices $(\mathcal{R}_{m,L})_{m=1,\ldots,M}$ verify $s_{min}\mathbf{I}_L \leq \mathcal{R}_{m,L} \leq s_{max}\mathbf{I}_L$. Therefore, if we denote by $\mathcal{C}$ a simple closed contour included in the half plane $\{\mathrm{Re}(\lambda) > 0\}$ and enclosing the interval $[s_{min}, s_{max}]$, then, $\mathcal{C}$ also encloses the spectrum of the matrices $(\mathcal{R}_{m,L})_{m=1,\ldots,M}$. This in particular implies that matrix $\mathcal{R}_{m,L}^{-1/2}$ can be written as

$$\mathcal{R}_{m,L}^{-1/2} = \frac{1}{2i\pi}\int_{\mathcal{C}_-}\frac{1}{\sqrt{\lambda}}\left(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} d\lambda \qquad (2.34)$$

where $\mathcal{C}_-$ means that the contour is negatively oriented. In the following, we denote by $(\lambda_{k,m})_{k=1,\ldots,K_m}$ the distinct eigenvalues of $\mathcal{R}_{m,L}$, and by $(\Pi_{k,m})_{k=1,\ldots,K_m}$ the orthogonal projection matrices over the corresponding eigenspaces. Therefore, $\mathcal{R}_{m,L}$ can be written as

$$\mathcal{R}_{m,L} = \sum_{k=1}^{K_m}\lambda_{k,m}\,\Pi_{k,m}. \qquad (2.35)$$

We denote by $\mathbf{\Delta}_{m,L}$ the matrix defined by

$$\mathbf{\Delta}_{m,L} = \widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}. \qquad (2.36)$$

In order to investigate $\mathbf{\Delta}_{m,L}$, it will be convenient to introduce a collection of operators $\mathcal{D}_{m,L}(\mathbf{X})$, $m = 1,\ldots,M$, which transform $L \times L$ matrices into $L \times L$ matrices and are defined as

$$\mathcal{D}_{m,L}(\mathbf{X}) = \frac{1}{2\pi i}\int_{\mathcal{C}_-}\frac{1}{\sqrt{\lambda}}\left(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L\right)^{-1}\mathbf{X}\left(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} d\lambda \qquad (2.37)$$

where, as before, $\mathcal{C}_-$ is a negatively oriented simple closed contour on the half plane $\text{Re}\lambda > 0$ enclosing $[s_{min}, s_{max}]$. As seen below, $\mathcal{D}_{m,L}$ can be interpreted as the differential operator of the matrix valued-function $\mathbf{A} \to \mathbf{A}^{-1/2}$ evaluated at $\mathcal{R}_{m,L}$. Note that, using the definitions in (2.35), we can express

$$(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} = \sum_{k=1}^{K_m} \frac{\Pi_{k,m}}{\lambda_{k,m} - \lambda}.$$

Plugging this expression into (2.37) and using the residue theorem, we can trivially check that this operator can also be expressed as

$$\mathcal{D}_{m,L}(\mathbf{X}) = \sum_{k=1}^{K_m} \sum_{l=1}^{K_m} \frac{1}{\sqrt{\lambda_{k,m}}\sqrt{\lambda_{l,m}}(\sqrt{\lambda_{k,m}} + \sqrt{\lambda_{l,m}})} \Pi_{k,m}\mathbf{X}\Pi_{l,m}. \qquad (2.38)$$

We summarize next a number of properties that will be useful about these operators throughout the paper.

Lemma 2.4. *Consider the operator $\mathcal{D}_{m,L}$ as defined in (2.37)-(2.38). Then, for every $L \times L$ matrix $\mathbf{A}$:*

(i) If $\mathbf{B}$ denotes another $L \times L$ matrix,

$$\text{Tr}\left(\mathcal{D}_{m,L}(\mathbf{A})\mathbf{B}\right) = \text{Tr}\left(\mathbf{A}\mathcal{D}_{m,L}(\mathbf{B})\right) \qquad (2.39)$$

(ii) There exists a nice constant $\kappa > 0$ such that

$$\|\mathcal{D}_{m,L}(\mathbf{A})\| \le \kappa\|\mathbf{A}\|. \qquad (2.40)$$

(iii) There exists a nice constant $\kappa > 0$ such that

$$\frac{1}{L}\text{Tr}\left[\mathcal{D}_{m,L}(\mathbf{A})\mathcal{D}_{m,L}^H(\mathbf{A})\right] \le \kappa \frac{1}{L}\text{Tr}(\mathbf{A}\mathbf{A}^H). \qquad (2.41)$$

Proof. The identity in (2.39) follows directly from the definition of $\mathcal{D}_{m,L}$. To see (2.40), simply consider the definition of $\mathcal{D}_{m,L}$ in (2.37) and note that $\sup_{\lambda \in \mathcal{C}} \|(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1}\| \le \kappa$ for some nice constant κ. In order to justify (2.41), we express $\mathcal{D}_{m,L}(\mathbf{A})$ using (2.38) so that, noting that $\Pi_{l,m}\Pi_{l',m} = \Pi_{l,m}\delta_{l-l'}$, we can write

$$\mathcal{D}_{m,L}(\mathbf{A})\mathcal{D}_{m,L}(\mathbf{A})^H = \sum_{k,k',l} \frac{\Pi_{k,m}}{\lambda_{k,m}^{1/2}(\lambda_{k,m}^{1/2} + \lambda_{l,m}^{1/2})}\mathbf{A}\frac{\Pi_{l,m}}{\lambda_{l,m}}\mathbf{A}^H\frac{\Pi_{k',m}}{\lambda_{k',m}^{1/2}\lambda_{l,m}^{1/2}(\lambda_{k',m}^{1/2} + \lambda_{l,m}^{1/2})}$$

Taking the normalized trace, changing the order of the matrices, and using again the fact that $\Pi_{k,m}\Pi_{k',m} = \Pi_{k,m}\delta_{k-k'}$, we obtain

$$\frac{1}{L}\text{Tr}\left[\mathcal{D}_{m,L}(\mathbf{A})\mathcal{D}_{m,L}(\mathbf{A})^H\right] = \sum_{k,l} \frac{1}{L}\text{Tr}\left[\frac{\Pi_{k,m}\mathbf{A}\,\Pi_{l,m}\mathbf{A}^H\Pi_{k,m}}{\lambda_{k,m}\lambda_{l,m}(\lambda_{k,m}^{1/2} + \lambda_{l,m}^{1/2})^2}\right].$$

Using that $\lambda_{k,m} \geq s_{min}$ for each k and m, we obtain immediately that

$$\frac{\Pi_{k,m}\mathbf{A}\,\Pi_{l,m}\mathbf{A}^H\Pi_{k,m}}{\lambda_{k,m}\lambda_{l,m}(\lambda_{k,m}^{1/2} + \lambda_{l,m}^{1/2})^2} \leq \kappa\,\Pi_{k,m}\mathbf{A}\,\Pi_{l,m}\mathbf{A}^H\Pi_{k,m}$$

from where the inequality

$$\frac{1}{L}\mathrm{Tr}\left[\mathcal{D}_{m,L}(\mathbf{A})\mathcal{D}_{m,L}(\mathbf{A})^H\right] \leq \kappa\,\frac{1}{L}\mathrm{Tr}\sum_{k,l}\Pi_{k,m}\mathbf{A}\,\Pi_{l,m}\mathbf{A}^H$$

follows directly. Noting that $\sum_k \Pi_{k,m} = \mathbf{I}_L$, we obtain (2.41). $\qquad\square$

Having introduced these operators, we now formulate a result that will be useful here and in the following sections.

Lemma 2.5. *Under Assumptions 1-3 and 5, it holds that*

$$\widehat{\mathcal{R}}_{m,L}^{-1/2} - \mathcal{R}_{m,L}^{-1/2} = -\mathcal{D}_{m,L}\left(\boldsymbol{\Delta}_{m,L}\right) + \boldsymbol{\Upsilon}_{m,L} \tag{2.42}$$

where the matrix $\boldsymbol{\Upsilon}_{m,L}$, implicitely defined by (2.42), verifies

$$\|\boldsymbol{\Upsilon}_{m,L}\| \prec \max\left(\frac{1}{M}, \frac{1}{L^{2\gamma_0}}\right). \tag{2.43}$$

Proof. See Appendix B. $\qquad\square$

Since $\|\boldsymbol{\Delta}_{m,L}\| = \|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$, we directly observe from the above proposition and (2.40) that $\|\widehat{\mathcal{B}}_L^{-1/2} - \mathcal{B}_L^{-1/2}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$. This of course implies that $\|\widehat{\mathcal{B}}_L^{-1/2}\| \prec 1$. Moreover, using the fact that $\|\widehat{\mathcal{R}}_L\| \prec 1$ (see (2.25)), (2.33) leads to

$$\|\boldsymbol{\Theta}_N\| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right). \tag{2.44}$$

To close this section, we remark that the identity $\widehat{\mathcal{R}}_{m,L}^{-1} - \mathcal{R}_{m,L}^{-1} = -\widehat{\mathcal{R}}_{m,L}^{-1}\boldsymbol{\Delta}_{m,L}\mathcal{R}_{m,L}^{-1}$ leads immediately to $\|\widehat{\mathcal{R}}_{m,L}^{-1} - \mathcal{R}_{m,L}^{-1}\| \prec \max\left(\frac{1}{\sqrt{M}}, \frac{1}{L^{\gamma_0}}\right)$. Using this and (2.24), we obtain the following Proposition.

Proposition 2.3. *Under Assumptions 1-3 and 5, there exists $\alpha_0 > 0$ such that for each $\alpha \geq \alpha_0$, one may find $\epsilon > 0$ and N_0 (both depending on α) such that*

$$\mathbb{P}\left(\|\overline{\mathcal{R}}_{\mathrm{corr},L}\| > \alpha\right) \leq \exp{-N^\epsilon} \tag{2.45}$$

$$\mathbb{P}\left(\|\widehat{\mathcal{R}}_{\mathrm{corr},L}\| > \alpha\right) \leq \exp{-N^\epsilon} \tag{2.46}$$

for each $N \geq N_0$.

From all the above, we can therefore conclude that the spectral behavior of the sample block correlation matrix $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ is equivalent to the spectral behavior of the matrix $\overline{\mathcal{R}}_{\mathrm{corr},L} = \mathcal{B}_L^{-1/2}\widehat{\mathcal{R}}_L\mathcal{B}_L^{-1/2}$.

3. Study of the influence of the estimation of matrices $(\mathcal{R}_{m,L})_{m=1,\ldots,M}$.

In this section, we study the impact of the estimation of matrices $(\mathcal{R}_{m,L})_{m=1,\ldots,M}$ on the asymptotic behaviour of the linear statistics $\widehat{\phi}_N$, defined as

$$\widehat{\phi}_N = \frac{1}{ML}\sum_{k=1}^{ML}\phi(\hat{\lambda}_{k,N}) = \int_{\mathbb{R}^+}\phi(\lambda)d\hat{\mu}_N(\lambda)$$

More specifically, we evaluate the behaviour of $\widehat{\phi}_N - \overline{\phi}_N$ where $\overline{\phi}_N$ is defined in (1.16) by establishing the following result.

Theorem 3.1. *Let Assumptions 1-3 and 5 hold true. Assume that the function ϕ is defined on $(-\delta, +\infty)$ for some $\delta > 0$ and smooth in a neighbourhood of the interval $[0, \alpha_0]$ where α_0 is defined in Proposition 2.3. Then, it holds that*

$$|\widehat{\phi}_N - \overline{\phi}_N| \prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right). \tag{3.1}$$

In order to establish Theorem 3.1, we first mention that Proposition 2.3 implies that it is possible to assume without restriction that ϕ is compactly supported by the interval $[-\delta, \alpha]$ for some $\alpha > \alpha_0$. To justify this claim, we consider $\xi \in (\alpha_0, \alpha)$ and introduce the event $\mathcal{A}_N$ defined by

$$\mathcal{A}_N = \left\{\|\overline{\mathcal{R}}_{\mathrm{corr},L}\| \leq \xi\right\} \cap \left\{\|\widehat{\mathcal{R}}_{\mathrm{corr},L}\| \leq \xi\right\}.$$

Proposition 2.3 implies that there exists a $\eta > 0$ for which $P(\mathcal{A}_N^c) \leq \exp -N^\eta$ for each N large enough. We denote by ϕ_c a smooth function, supported by $[-\delta, \alpha]$, and which coincides with ϕ on the interval $[-\delta/2, \xi]$. Then, it is clear that $\widehat{\phi}_N$ and $\overline{\phi}_N$ coincide with $\widehat{\phi}_{c,N}$ and $\overline{\phi}_{c,N}$ respectively on $\mathcal{A}_N$. For each $\epsilon > 0$, by conditioning on the event $\mathcal{A}_N$ and its complementary $\mathcal{A}_N^c$ we can express

$$\mathbb{P}\left(|\widehat{\phi}_N - \overline{\phi}_N| > N^\epsilon\max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right)\right) =$$

$$= \mathbb{P}\left(|\widehat{\phi}_{c,N} - \overline{\phi}_{c,N}| > N^\epsilon\max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right), \mathcal{A}_N\right) +$$

$$+ \mathbb{P}\left(|\widehat{\phi}_N - \overline{\phi}_N| > N^\epsilon\max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right), \mathcal{A}_N^c\right)$$

where we have used the fact that $\widehat{\phi}_N$ and $\overline{\phi}_N$ respectively coincide with $\widehat{\phi}_{c,N}$ and $\overline{\phi}_{c,N}$ on $\mathcal{A}_N$. Now, for N large enough we can bound the first term of the

above equation by $\mathbb{P}(|\widehat{\phi}_{c,N} - \overline{\phi}_{c,N}| > N^\epsilon \max(M^{-1}, L^{-\gamma_0}))$ and the second term by $\mathbb{P}(\mathcal{A}_N^c) \leq \exp(-N^\eta)$. Therefore, it is sufficient to establish that $|\widehat{\phi}_{c,N} - \overline{\phi}_{c,N}| \prec \max(M^{-1}, L^{-\gamma_0})$ to prove (3.1). For this reason, from now on we assume without loss of generality that ϕ is supported by $[-\delta, \alpha]$.

The main tool that we will use in order to analyze the asymptotic behavior of the linear spectral statistics is the Helffer-Sjöstrand formula for sufficiently regular, compactly supported functions. This formula was already used in the large random matrices literature, see e.g. [2], [3], [31]. In order to introduce this tool, assume that $\phi(\lambda)$ is compactly supported and of class $\mathcal{C}^{k+1}$ for a certain integer k, and denote by $\Phi_k(\phi) : \mathbb{C} \to \mathbb{C}$ the function of complex variable

$$\Phi_k(\phi)(x + iy) = \sum_{l=0}^{k} \frac{(iy)^l}{l!} \phi^{(l)}(x)\rho(y) \tag{3.2}$$

where $\rho : \mathbb{R} \to \mathbb{R}^+$ is a smooth, compactly supported function (to fix the ideas, we assume that the support of ρ is $[-2, 2]$) that takes the value 1 in a neighbourhood of zero. Now, taking $z = x + iy$, we see that the function $\Phi_k(\phi)(z)$ is compactly supported on the complex plane, and therefore by [35, Lemma 20.3] we have

$$\int \phi(\lambda)d\mu(\lambda) = \frac{1}{\pi}\text{Re}\int_{\mathbb{C}^+} dx\,dy\,\overline{\partial}\Phi_k(\phi)(z)s_\mu(z)$$

where μ is a probability measure, $s_\mu(z)$ its Stieltjes transform and where we define

$$\overline{\partial}\Phi_k(\phi)(z) = \frac{\partial\Phi_k(\phi)(x + iy)}{\partial x} + i\frac{\partial\Phi_k(\phi)(x + iy)}{\partial y}.$$

In particular, according to the definition of $\Phi_k(\phi)(z)$ in (3.2), we can see that

$$\overline{\partial}\Phi_k(\phi)(z) = \frac{(iy)^k}{k!}\phi^{(k+1)}(x)$$

when y belongs to a neighbourhood of zero where $\rho(y) = 1$. The regularity of ϕ will allow us to bound quantities of the form $|\overline{\partial}\Phi_k(\phi)(z)y^{-k}|$ when y is in a neighbourhood of zero.

Consider now the two resolvents $\widehat{\mathbf{Q}}_N(z)$ and $\mathbf{Q}_N(z)$ defined in (1.29) and (1.30) respectively. Recall that their normalized trace is equal to the Stieltjes transform of the empirical eigenvalue distribution of $\widehat{\mathcal{R}}_{\text{corr},L}$ and $\overline{\mathcal{R}}_{\text{corr},L}$ respectively. Hence, a direct application of the Helffer-Sjöstrand formula to our problem leads to the identity

$$\widehat{\phi}_N - \overline{\phi}_N = \frac{1}{\pi}\text{Re}\int_{\mathcal{D}} dx\,dy\,\overline{\partial}\Phi_k(\phi)(z)\left(\frac{1}{ML}\text{Tr}(\widehat{\mathbf{Q}}_N(z)) - \frac{1}{ML}\text{Tr}(\mathbf{Q}_N(z))\right) \tag{3.3}$$

where $\mathcal{D}$ is defined by $\mathcal{D} = [-\delta, \alpha] \times [0, 2]$ and where k is large enough. Before going into the details of the proof of Theorem 3.1, we first present the main steps of the proof. In what follows, we will omit the dependence on N and z in all the matrices in order to simplify the notation.

We recall that $\boldsymbol{\Theta}$ is the matrix defined in (2.32) and remark that, by the definition of resolvents, we can write

$$\widehat{\mathbf{Q}} - \mathbf{Q} = -\mathbf{Q}\boldsymbol{\Theta}\widehat{\mathbf{Q}} = -\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q} + \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}\boldsymbol{\Theta}\widehat{\mathbf{Q}}$$

Therefore, (3.3) can also be written as

$$\widehat{\phi}_N - \overline{\phi}_N = -\frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{ML}(\mathrm{Tr}\mathbf{Q}^2\boldsymbol{\Theta})$$
$$+ \frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\mathrm{Tr}(\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}\boldsymbol{\Theta}\widehat{\mathbf{Q}}). \qquad (3.4)$$

Having established these basic facts, the proof of Theorem 3.1 proceeds as follows:

(1) The first step of the proof consists in showing that, by virtue of (2.44), the second term of (3.4) can be disregarded from the evaluation, in the sense that

$$\left|\frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\mathrm{Tr}(\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}\boldsymbol{\Theta}\widehat{\mathbf{Q}})\right| \prec \max\left(\frac{1}{M}, \frac{1}{L^{2\gamma_0}}\right) \qquad (3.5)$$

We therefore just need to evaluate the first term of the right hand side of (3.4).

(2) In the second step, it is proved that $\boldsymbol{\Theta}$ can be written as

$$\boldsymbol{\Theta} = \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2}\right)\mathcal{B}^{1/2}\overline{\mathcal{R}}_{\mathrm{corr}} + \overline{\mathcal{R}}_{\mathrm{corr}}\mathcal{B}^{1/2}\left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2}\right) + \boldsymbol{\Theta}_2 \qquad (3.6)$$

where $\|\boldsymbol{\Theta}_2\| \prec \max(M^{-1}, L^{-2\gamma_0})$. This will imply that the contribution of $\boldsymbol{\Theta}_2$ to the first term of the right hand side of (3.4) can be omitted.

(3) If we take $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta} - \boldsymbol{\Theta}_2$, the purpose of the third step is to establish that

$$\left|\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\mathrm{Tr}(\mathbf{Q}^2\boldsymbol{\Theta}_1)\right| \prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right). \qquad (3.7)$$

For this, we will just verify that

$$\left|\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\mathrm{Tr}\left[\mathbf{Q}^2(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2})\mathcal{B}^{1/2}\overline{\mathcal{R}}_{\mathrm{corr}}\right]\right| \prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right) \qquad (3.8)$$

(note that the second term in (3.6) can be handled similarly). The proof of (3.8) is demanding. Using Lemma 2.5, we only need to show that

$$\left|\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{M}\sum_{m=1}^{M}\frac{1}{L}\mathrm{Tr}\left[\mathcal{D}_{m,L}(\boldsymbol{\Delta}_{m,L})\mathcal{R}_{m,L}^{1/2}(\mathbf{Q} + z\mathbf{Q}^2)_{m,m}\right]\right|$$
$$\prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right) \qquad (3.9)$$

where $\mathbf{Q}_{m,m}$ denotes the mth $L \times L$ diagonal block of $\mathbf{Q}$, where the operator $\mathcal{D}_{m,L}$ is defined in (2.37)-(2.38) and where $\boldsymbol{\Delta}_{m,L}$ is defined in (2.36). We will only establish that

$$\left|\int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z)\frac{1}{M}\sum_{m=1}^{M}\frac{1}{L}\mathrm{Tr}\left[\mathcal{D}_{m,L}(\boldsymbol{\Delta}_{m,L})\mathcal{R}_{m,L}^{1/2}\mathbf{Q}_{m,m}\right]\right| \prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}}\right) \qquad (3.10)$$

because the term due to $z(\mathbf{Q}^2)_{m,m}$ can be handled similarly. In order to show this, we rely on the fact that, up to a term stochastically dominated by $\frac{1}{M}$, it is possible to replace matrices $(\widehat{\mathcal{R}}_{m,L})_{m=1,\ldots,M}$ in (3.10) by their Toeplitz approximations $\widehat{\mathcal{R}}_{m,L}^t = \widetilde{\mathbf{W}}_N^m (\widetilde{\mathbf{W}}_N^m)^H$, $m = 1, \ldots, M$ introduced in (2.27) of Section 2.2. The upper bound in (2.29) will imply that the contribution of the bias of the Toeplitz estimates of $(\mathcal{R}_{m,L})_{m=1,\ldots,M}$ to (3.10) is a term of order $\mathcal{O}(\max{(M^{-1}, L^{-\gamma_0})})$. At this point, it will remain to study the term ζ defined by

$$\zeta = \int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \mathrm{Tr}\left[\mathcal{D}_{m,L}\left(\widehat{\mathcal{R}}_{m,L}^t - \mathbb{E}(\widehat{\mathcal{R}}_{m,L}^t)\right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m}\right].$$
(3.11)

Recall that $\hat{r}_m(l) = \frac{1}{N} \sum_{n=1}^{N-l} y_m(n+l) y_m^*(n)$ and $\hat{r}_m(-l) = \hat{r}_m^*(l)$ for $l \geq 0$ represent the empirical estimate of the autocovariance sequence of y_m at lag l, and consider $\hat{r}_m^\circ(l) = \hat{r}_m - \mathbb{E}\hat{r}_m(l)$. With these definitions and using (2.39) and (1.38, the term ζ can be re-written as

$$\zeta = \int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \sum_{u=-(L-1)}^{L-1} \hat{r}_m^\circ(u)\, \tau\left(\mathcal{D}_{m,L}\left(\mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m}\right)\right)(-u)$$
(3.12)

where we recall that if $\mathbf{A}$ is a $L \times L$ matrix, $\tau(\mathbf{A})(u)$ is defined by $\tau(\mathbf{A})(u) = \frac{1}{L}\mathrm{Tr}(\mathbf{A}\mathbf{J}_L^u)$, see (1.36). This way of expressing ζ will be the key to showing that

$$|\zeta| \prec \frac{1}{M}$$
(3.13)

which will complete the proof of Theorem 3.1. This will be shown in two final steps.

– We check that $\mathbb{E}(\zeta) = \mathcal{O}(M^{-1})$. To verify this, we use that the Nash-Poincaré inequality and obtain that $\mathrm{Var}\left[\tau\left(\mathcal{D}_{m,L}(\mathcal{R}_{m,L}^{1/2}\mathbf{Q}_{m,m})\right)(-u)\right] = \mathcal{O}(N^{-1})$. Since $\mathrm{Var}(\hat{r}_m(u))$ is also a term of order $\mathcal{O}(N^{-1})$, we obtain immediately from the Schwartz inequality that $\mathbb{E}(\zeta) = \mathcal{O}(M^{-1})$.

– The most difficult part of the proof consists in establishing that

$$|\zeta - \mathbb{E}(\zeta)| \prec \frac{1}{M}.$$
(3.14)

For this, for each m, we introduce the $(N+L-1)$–dimensional row vector $\mathbf{y}_m = (y_{m,1}, \ldots, y_{m,N+L-1})$, which can be re-written as

$$\mathbf{y}_m = \mathbf{x}_m \mathcal{R}_{m,N+L-1}^{1/2}$$
(3.15)

for some $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_{N+L-1})$-distributed row vector $\mathbf{x}_m$. By using the above definition, we can re-interpret ζ as a function of of the $M(N+L-1)$ i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ entries of vector

$$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_M).$$
(3.16)

If ζ, considered as a function of $(\mathbf{x}, \mathbf{x}^*)$ were a Lipschitz function with constant of order $\mathcal{O}(M^{-1})$, the result in (3.14) would follow from conventional concentration inequalities of Lipschitz functions of Gaussian random vectors (see (1.32)

above). Unfortunately, the terms $\hat{r}_m(u)$ are not Lipschitz functions of $\mathbf{x}$ due to the quadratic dependence on this vector. In any case, it is still true that for each $\epsilon > 0$, the inequality $|\hat{r}_m(u) - \mathbb{E}(\hat{r}_m(u))| \leq \frac{N^\epsilon}{N}$ holds for fixed u and m, except for an event that has exponentially small probability. Therefore, we show that it is possible to replace (for each u and m) $\hat{r}_m(u) - \mathbb{E}(\hat{r}_m(u))$ by a well chosen function, and that the corresponding modification $\tilde{\zeta}$ of ζ is Lipschitz with constant $\frac{N^\epsilon}{M}$. We deduce from this that $|\zeta - \mathbb{E}(\zeta)| \prec \frac{N^\epsilon}{M}$ for each $\epsilon > 0$, a property which will directly imply (3.14).

We now proceed with the three steps of the proof.

Step 1. In order to establish (3.5), we simply notice that

$$\left| \int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z) \frac{1}{ML} \mathrm{Tr}(\mathbf{Q}\mathbf{\Theta}\mathbf{Q}\hat{\mathbf{Q}}) \right| \leq \int_{\mathcal{D}} dx\, dy\, |\bar{\partial}\Phi_k(\phi)(z)| \left| \frac{1}{ML} \mathrm{Tr}(\mathbf{Q}\mathbf{\Theta}\mathbf{Q}\hat{\mathbf{Q}}) \right|.$$

It is clear that if $z \in \mathbb{C}^+$, we can use the item (iv) in Proposition 1.1 to establish that

$$\left| \frac{1}{ML} \mathrm{Tr}(\mathbf{Q}\mathbf{\Theta}\mathbf{Q}\hat{\mathbf{Q}}) \right| \leq \|\mathbf{Q}\|^2 \|\hat{\mathbf{Q}}\| \|\mathbf{\Theta}\|^2 \leq \frac{1}{(\mathrm{Im} z)^3} \|\mathbf{\Theta}\|^2.$$

Since ϕ is smooth by assumption, we can choose $k \geq 3$ to guarantee that the integral $\int_{\mathcal{D}} dx\, dy\, |\bar{\partial}\Phi_k(\phi)(z)| \frac{1}{(\mathrm{Im} z)^3}$ is finite. This, together with (2.44), shows that

$$\left| \int_{\mathcal{D}} dx\, dy\, \bar{\partial}\Phi_k(\phi)(z) \frac{1}{ML} (\mathrm{Tr}\mathbf{Q}\mathbf{\Theta}\mathbf{Q}\hat{\mathbf{Q}}) \right| \leq \kappa \|\mathbf{\Theta}\|^2 \prec \max\left(\frac{1}{M}, \frac{1}{L^{2\gamma_0}} \right)$$

which completes the proof of (3.5).

Step 2. In order to establish (3.6), we take (2.33) as a starting point and express $\widehat{\mathcal{R}}$ as $\widehat{\mathcal{R}} = \mathcal{B}^{1/2} \overline{\mathcal{R}}_{\mathrm{corr}} \mathcal{B}^{1/2}$, that is

$$\mathbf{\Theta} = \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right) \mathcal{B}^{1/2} \overline{\mathcal{R}}_{\mathrm{corr}} \mathcal{B}^{1/2} \widehat{\mathcal{B}}^{-1/2} + \overline{\mathcal{R}}_{\mathrm{corr}} \mathcal{B}^{1/2} \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right)$$

$$= \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right) \mathcal{B}^{1/2} \overline{\mathcal{R}}_{\mathrm{corr}} + \overline{\mathcal{R}}_{\mathrm{corr}} \mathcal{B}^{1/2} \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right) + \mathbf{\Theta}_2$$

where $\mathbf{\Theta}_2$ is given by

$$\mathbf{\Theta}_2 = \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right) \mathcal{B}^{1/2} \overline{\mathcal{R}}_{\mathrm{corr}} \mathcal{B}^{1/2} \left(\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2} \right).$$

As we showed that $\|\widehat{\mathcal{B}}^{-1/2} - \mathcal{B}^{-1/2}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$, $\|\mathbf{\Theta}_2\|$ clearly verifies $\|\mathbf{\Theta}_2\| \prec \max(M^{-1}, L^{-2\gamma_0})$ as expected. Hence, using the inequality $|\frac{1}{ML} \mathrm{Tr}(\mathbf{Q}^2 \mathbf{\Theta}_2)| \leq \frac{\|\mathbf{\Theta}_2\|}{(\mathrm{Im} z)^2}$ together with

$$\int_{\mathcal{D}} dx\, dy\, \left| \bar{\partial}\Phi_k(\phi)(z) \right| \frac{1}{(\mathrm{Im} z)^2} < +\infty$$

42

we obtain that

$$\left| \int_{\mathcal{D}} dx \, dy \, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{ML} \text{Tr}(\mathbf{Q}^2 \mathbf{\Theta}_2) \right| \leq \kappa \, \|\mathbf{\Theta}_2\| \prec \max\left(\frac{1}{M}, \frac{1}{L^{2\gamma_0}} \right).$$

Step 3. We finally establish (3.7), and just verify (3.8) because the contribution of the second term of $\mathbf{\Theta}_1$ can be handled similarly. Using (2.42) and the resolvent identity $\overline{\mathcal{R}}_{\text{corr}} \mathbf{Q} = \mathbf{I} + z\mathbf{Q}$, we obtain immediately that the term on the left hand side of (3.8) can be written as the sum of the term on the left hand side of (3.9) plus a term depending on the matrices $(\mathbf{\Upsilon}_m)_{m=1,\dots,M}$. As this last term is easily seen to be stochastically dominated by $\max(L^{-2\gamma_0}, M^{-1})$, (3.8) becomes equivalent to (3.9).

We now prove (3.10). We first reason that we can replace the matrices $\widehat{\mathcal{R}}_{m,L}$ with their Toeplitz approximations $\widehat{\mathcal{R}}_{m,L}^t$. Indeed, it was shown in (2.28) of Section 2.2 that $\|\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t\| \prec \frac{1}{M}$. We claim that this implies that

$$\left| \int_{\mathcal{D}} dx \, dy \, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \text{Tr} \left[\mathcal{D}_{m,L} \left(\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t \right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] \right| \prec \frac{1}{M}$$

Indeed, a direct use of (2.40) together with the fact that $\|\mathbf{Q}_{m,m}\| \leq (\text{Im} z)^{-1}$ for $z \in \mathbb{C}^+$ shows that

$$\left| \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \text{Tr} \left[\mathcal{D}_{m,L} \left(\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t \right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] \right| \leq \frac{\kappa}{\text{Im} z} \sup_{m=1,\dots,M} \|\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t\|$$

for some nice constant $\kappa > 0$. However, since the integral $\int_{\mathcal{D}} dx \, dy \, |\bar{\partial} \Phi_k(\phi)(z)| \frac{1}{\text{Im} z}$ is finite as long as $k \geq 1$, we readily see that

$$\left| \int_{\mathcal{D}} dx \, dy \, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \text{Tr} \left[\mathcal{D}_{m,L} \left(\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t \right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] \right| \leq$$

$$\leq \kappa \sup_{m=1,\dots,M} \|\widehat{\mathcal{R}}_{m,L} - \widehat{\mathcal{R}}_{m,L}^t\| \prec \frac{1}{M}.$$

Consequently, in order to establish (3.10), it only remains to prove that

$$\left| \int_{\mathcal{D}} dx \, dy \, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \text{Tr} \left[\mathcal{D}_{m,L} \left(\widehat{\mathcal{R}}_{m,L}^t - \mathcal{R}_{m,L} \right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] \right| \prec$$

$$\prec \max\left(\frac{1}{M}, \frac{1}{L^{\gamma_0}} \right). \quad (3.17)$$

Given the Toeplitz structure of $\widehat{\mathcal{R}}_{m,L}^t - \mathcal{R}_{m,L}$, the bound that was established in (2.29) directly implies that $\sup_{m=1,\dots,M} \|\mathbb{E} \widehat{\mathcal{R}}_{m,L}^t - \mathcal{R}_{m,L}\| \leq \kappa (L^{-\gamma_0} + N^{-1} L^{(1-\gamma_0)+})$.

This leads immediately to

$$\left| \int_{\mathcal{D}} dx\, dy\, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \mathrm{Tr} \left[\mathcal{D}_{m,L} \left(\mathbb{E} \left(\widehat{\mathcal{R}}_{m,L}^t \right) - \mathcal{R}_{m,L} \right) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] \right| \leq$$

$$\leq \kappa \left(\frac{1}{L^{\gamma_0}} + \frac{L^{(1-\gamma_0)_+}}{N} \right) \leq \kappa \left(\frac{1}{L^{\gamma_0}} + \frac{1}{M} \right).$$

It thus remains to study ζ defined in (3.11). Noting that $\widehat{\mathcal{R}}_{m,L}^t - \mathbb{E}(\widehat{\mathcal{R}}_{m,L}^t)$ is the $L \times L$ Toeplitz matrix with entries $\hat{r}_m^\circ(i-j) = \hat{r}_m(i-j) - \mathbb{E}\hat{r}_m(i-j)$, $1 \leq i, j \leq L$, we can establish, by virtue of (2.39) and (1.38),

$$\frac{1}{L} \mathrm{Tr} \left[\mathcal{D}_{m,L} (\widehat{\mathcal{R}}_{m,L}^t - \mathbb{E}(\widehat{\mathcal{R}}_{m,L}^t)) \mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m} \right] = \frac{1}{L} \mathrm{Tr} \left[(\widehat{\mathcal{R}}_{m,L}^t - \mathbb{E}(\widehat{\mathcal{R}}_{m,L}^t)) \mathcal{D}_{m,L} (\mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m}) \right]$$

$$= \sum_{u=-(L-1)}^{L-1} \hat{r}_m^\circ(u) \, \tau(\mathcal{D}_{m,L}(\mathcal{R}_{m,L}^{1/2} \mathbf{Q}_{m,m}))(-u).$$

We first study the expectation of ζ and prove that

$$\mathbb{E}(\zeta) = \mathcal{O}\left(\frac{1}{M} \right). \tag{3.18}$$

For this, we first note that, using again (2.39), we can write

$$\mathbb{E}(\zeta) = \int_{\mathcal{D}} dx\, dy\, \bar{\partial} \Phi_k(\phi)(z) \frac{1}{M} \sum_{m=1}^{M} \sum_{u=-(L-1)}^{L-1} \mathbb{E} \left(\hat{r}_m^\circ(u) \frac{1}{L} \mathrm{Tr} \left(\mathbf{Q}_{m,m}^\circ \mathcal{D}_{m,L}(\mathbf{J}_L^{-u}) \mathcal{R}_{m,L}^{1/2} \right) \right)$$

where we recall that $\mathbf{Q}_{m,m}^\circ = \mathbf{Q}_{m,m} - \mathbb{E}(\mathbf{Q}_{m,m})$. By the Cauchy-Schwarz inequality, we have

$$\left| \mathbb{E} \left(\hat{r}_m^\circ(u) \frac{1}{L} \mathrm{Tr} \left(\mathbf{Q}_{m,m}^\circ \mathcal{D}_{m,L}(\mathbf{J}_L^{-u}) \mathcal{R}_{m,L}^{1/2} \right) \right) \right| \leq$$

$$\leq \mathrm{Var}^{1/2}(\hat{r}_m(u)) \, \mathrm{Var}^{1/2} \left(\frac{1}{L} \mathrm{Tr} \left(\mathbf{Q}_{m,m} \mathcal{D}_{m,L}(\mathbf{J}_L^{-u}) \mathcal{R}_{m,L}^{1/2} \right) \right) \tag{3.19}$$

and it is therefore enough to bound these two variances. Regarding $\mathrm{Var}(\hat{r}_m(u))$, we observe that we can write

$$\hat{r}_m(u) = \frac{1}{N} \mathbf{x}_m \mathcal{R}_{m,N+L-1}^{1/2} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix} \mathbf{J}_N^{-u} \left(\mathbf{I}_N, 0 \right) \mathcal{R}_{m,N+L-1}^{1/2} \mathbf{x}_m^H \tag{3.20}$$

from which we deduce immediately that $\mathrm{Var}(\hat{r}_m(u)) \leq \frac{\kappa}{N}$. Regarding the term corresponding to the second variance in (3.19), we first introduce the following lemma, proven in Appendix C.

Lemma 3.1. *Let* $\mathbf{x}_{m,i}$ *the ith entry of vector* $\mathbf{x}_m$ *defined in (3.15) and consider an* $ML \times ML$ *deterministic matrix* $\mathbf{A}$. *For* $z \in \mathbb{C}^+$, *we have*

$$\sum_{m=1}^{M} \sum_{i=1}^{N+L-1} \left| \frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}}{\partial \mathbf{x}_{m,i}} \mathbf{A} \right) \right|^2 \leq \frac{\kappa}{MN} \frac{1+|z|}{(\mathrm{Im}z)^3} \left(1 + \frac{1}{\mathrm{Im}z} \right) \frac{1}{ML} \mathrm{Tr}(\mathbf{A}\mathbf{A}^H) \tag{3.21}$$

44

for some nice constant κ.

As a consequence of the above lemma, we have the following result, which follows from a direct application of the Poincaré-Nash inequality (see further [27, Lemma 3.1]).

Corollary 3.1. *Let $(\mathbf{A}_N)_{N \geq 1}$ denote a sequence of deterministic $ML \times ML$ matrices. Then,*

$$\mathrm{Var}\frac{1}{ML}\mathrm{Tr}\left(\mathbf{A}_N\mathbf{Q}\right) \leq \frac{\kappa}{MN}\frac{1+|z|}{(\mathrm{Im}z)^3}\left(1+\frac{1}{\mathrm{Im}z}\right)\frac{1}{ML}\mathrm{Tr}(\mathbf{A}_N\mathbf{A}_N^H) \qquad (3.22)$$

for some nice constant κ.

We can apply the above corollary to study the second variance term in (3.19) by defining the $ML \times L$ matrix $\mathbf{E}_m$, which is composed of M blocks of dimension $L \times L$, all of which are zero except for the mth one, which is equal to $\mathbf{I}_L$. This means that we can express $\mathbf{Q}_{m,m} = \mathbf{E}_m^H\mathbf{Q}\mathbf{E}_m$. Hence, the use of (3.22) with $\mathbf{A}_N = \mathbf{E}_m\mathcal{D}_{m,L}(\mathbf{J}_L^{-u})\mathcal{R}_{m,L}^{1/2}\mathbf{E}_m^H$ leads immediately to

$$\mathrm{Var}\left(\frac{1}{L}\mathrm{Tr}\left(\mathbf{Q}_{m,m}^{\circ}\mathcal{D}_{m,L}(\mathbf{J}_L^{-u})\mathcal{R}_{m,L}^{1/2}\right)\right) \leq \frac{\kappa}{N}\frac{1+|z|}{(\mathrm{Im}z)^3}\left(1+\frac{1}{\mathrm{Im}z}\right).$$

Using these two bounds in (3.19) we can conclude that

$$|\mathbb{E}(\zeta)| \leq \frac{\kappa}{M}\int_{\mathcal{D}}dx\,dy\,|\bar{\partial}\Phi_k(\phi)(z)|\,\frac{1}{(\mathrm{Im}z)^{3/2}}\left(1+\frac{1}{\mathrm{Im}z}\right)^{1/2}.$$

Noting that ϕ is smooth, we see that the above integral is finite by choosing $k \geq 2$, and consequently (3.18) is proved.

We finally establish that $|\zeta - \mathbb{E}(\zeta)| \prec \frac{1}{M}$ following the approach in [29]. More specifically, we interpret ζ as a function of the $M(N + L - 1)$–dimensional vector $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_M)$ (where we recall that the vectors $(\mathbf{x}_m)_{m=1,\ldots,M}$ are defined by (3.15)) and exchange ζ by a term that is Lipschitz with a relevant Lipschitz constant. For each $\epsilon > 0$, we denote by $\mathcal{A}_{N,\epsilon}$ the composite event

$$\mathcal{A}_{N,\epsilon} = \bigcap_{\substack{m=1,\ldots,M \\ u=-(L-1),\ldots,L-1}}\left\{|\hat{r}_m(u) - \mathbb{E}(\hat{r}_m(u))| < \tfrac{N^\epsilon}{\sqrt{N}}\right\} \qquad (3.23)$$

$$\cap\left(\bigcap_{m=1,\ldots,M}\left\{\tfrac{\|\mathbf{x}_m\|^2}{N+L-1} \leq 2\right\}\right).$$

It is clear that $\left|\frac{\|\mathbf{x}_m\|^2}{N+L-1} - 1\right| \prec \frac{1}{\sqrt{N}}$ and that $|\hat{r}_m(u) - \mathbb{E}(\hat{r}_m(u))| \prec \frac{1}{\sqrt{N}}$. Therefore, the family of events $(\mathcal{A}_{N,\epsilon})_{N \geq 1}$ holds with exponentially high probability, i.e. there exist N_0 and $\eta > 0$ such that $\mathbb{P}\left(\mathcal{A}_{N,\epsilon}^c\right) \leq \exp{-N^\eta}$ for each $N \geq N_0$. Our strategy is to replace ζ with a certain random variable $\tilde{\zeta}_\epsilon$ such that $\tilde{\zeta}_\epsilon = \zeta$ on $\mathcal{A}_{N,\epsilon}$ and which, considered as a function of $\mathbf{x}$, is Lipschitz with constant $\frac{N^\epsilon}{M}$. In order to build $\tilde{\zeta}_\epsilon$, we

consider a smooth function $g(t)$ satisfying $g(t) = t$ if $t \in [-1, 1]$, $g(t) = 0$ if $|t| \geq 2$, $g(t) \geq 0$ if $t \geq 0$, $g(t) \leq 0$ if $t \leq 0$ and $|g(t)| \leq 2|t|$ for each t. We then define $g_{N,\epsilon}(t)$ by

$$g_{N,\epsilon}(t) = \frac{N^\epsilon}{\sqrt{N}} g\left(\frac{\sqrt{N}}{N^\epsilon} t\right) \tag{3.24}$$

It is easy to check that $g_{N,\epsilon}$ verifies:

$$g_{N,\epsilon}(t) = t \text{ if } |t| \leq \frac{N^\epsilon}{\sqrt{N}}, \ g_{N,\epsilon}(t) = 0 \text{ if } |t| \geq \frac{2N^\epsilon}{\sqrt{N}}, \ |g_{N,\epsilon}(t)| \leq \frac{2N^\epsilon}{\sqrt{N}} \text{ for each } t \tag{3.25}$$

and

$$g'_{N,\epsilon}(t) = 0 \text{ if } |t| \geq \frac{2N^\epsilon}{\sqrt{N}}, \ |g'_{N,\epsilon}(t)| \leq \kappa \text{ if } |t| \leq \frac{2N^\epsilon}{\sqrt{N}}. \tag{3.26}$$

We also introduce a $\mathcal{C}_1$ function $\tilde{g}(t)$ verifying $\tilde{g}(t) = 1$ if $t \in [0, 2]$, and $\tilde{g}(t) = 0$ if t does not belong to $[-1, 3]$. We then define $\tilde{\zeta}_\epsilon$ by

$$\tilde{\zeta}_\epsilon = \frac{1}{M} \int_{\mathcal{D}} dx \, dy \, \bar{\partial} \Phi_k(\phi)(z) f(\mathbf{x}, z) \tag{3.27}$$

where $f(\mathbf{x}, z)$ is defined by

$$f(\mathbf{x}, z) = \sum_{m=1}^{M} \tilde{g}\left(\frac{\|\mathbf{x}_m\|^2}{N + L - 1}\right) \left(\sum_{u=-(L-1)}^{L-1} g_{N,\epsilon}(\hat{r}^\circ_m(u)) \, \tau(\mathcal{D}_{m,L}(\mathcal{R}^{1/2}_{m,L}\mathbf{Q}_{m,m}(z))(-u))\right). \tag{3.28}$$

It is clear that $\tilde{\zeta}_\epsilon = \zeta$ on the set $\mathcal{A}_{N,\epsilon}$. In the following, we first establish that, considered as a function of $\mathbf{x}$, $\tilde{\zeta}_\epsilon$ is a Lipschitz function with constant $\frac{N^\epsilon}{M}$. This property will imply that $|\tilde{\zeta}_\epsilon - \mathbb{E}(\tilde{\zeta}_\epsilon)| \prec \frac{N^\epsilon}{M}$. Next, we justify that if $|\tilde{\zeta}_\epsilon - \mathbb{E}(\tilde{\zeta}_\epsilon)| \prec \frac{N^\epsilon}{M}$, then $|\zeta - \mathbb{E}(\zeta)| \prec \frac{N^\epsilon}{M}$. As this property will be true for each $\epsilon > 0$, we will deduce from this that $|\zeta - \mathbb{E}(\zeta)| \prec \frac{1}{M}$ as expected.

In order to show that $\tilde{\zeta}_\epsilon$ is a Lipschitz function with constant $\frac{N^\epsilon}{M}$, we establish that the norm square of the gradient of $\tilde{\zeta}_\epsilon$ is a $\mathcal{O}(\frac{N^{2\epsilon}}{M^2})$ term.

Lemma 3.2. *Under the assumptions of Theorem 3.1, the inequality*

$$\|\nabla\tilde{\zeta}_\epsilon\|^2 = \sum_{m_0=1}^{M} \sum_{i=1}^{N+L-1} \left(\left|\frac{\partial\tilde{\zeta}_\epsilon}{\partial\mathbf{x}_{m_0,i}}\right|^2 + \left|\frac{\partial\tilde{\zeta}_\epsilon}{\partial\mathbf{x}^*_{m_0,i}}\right|^2\right) \leq \kappa \frac{N^{2\epsilon}}{M^2} \tag{3.29}$$

holds true for some nice constant $\kappa > 0$.

We just evaluate the contribution of the derivatives w.r.t. the variables $\mathbf{x}_{m_0,i}$ because the derivatives w.r.t. $\mathbf{x}^*_{m_0,i}$ can be addressed in a similar way. It is clear

46

that

$$\left| \frac{\partial \tilde{\zeta}_\epsilon}{\partial \mathbf{x}_{m_0,i}} \right|^2 \leq \frac{\kappa}{M^2} \int_{\mathcal{D}} dx\, dy\, \left| \bar{\partial} \Phi_k(\phi)(z) \right|^2 \left| \frac{\partial f}{\partial \mathbf{x}_{m_0,i}} \right|^2$$

so that Lemma 3.2 will be established if we choose $k \geq 4$ and prove that

$$\sum_{m_0=1}^{M} \sum_{i=1}^{N} \left| \frac{\partial f}{\partial \mathbf{x}_{m_0,i}} \right|^2 \leq N^{2\epsilon}\, P_1(|z|) P_2\left(\frac{1}{\mathrm{Im} z} \right) \qquad (3.30)$$

where P_1 and P_2 are nice polynomials, with $\deg(P_2) \leq 4$.

We now observe that the derivative $\partial f / \partial \mathbf{x}_{m_0,i}$ is the sum of the followings three terms:

$$T^1_{m_0,i} = \tilde{g}'\left(\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1} \right) \frac{\mathbf{x}^*_{m_0,i}}{N+L-1} \left(\sum_{u=-(L-1)}^{L-1} g_{N,\epsilon}(\hat{r}^\circ_{m_0}(u))\, \tau(\mathcal{D}_{m_0,L}(\mathcal{R}^{1/2}_{m_0,L}\mathbf{Q}_{m_0,m_0}))(-u) \right)$$

$$T^2_{m_0,i} = \tilde{g}\left(\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1} \right) \left(\sum_{u=-(L-1)}^{L-1} g'_{N,\epsilon}(\hat{r}^\circ_{m_0}(u))\, \frac{\partial \hat{r}_{m_0}(u)}{\partial \mathbf{x}_{m_0,i}}\, \tau(\mathcal{D}_{m_0,L}(\mathcal{R}^{1/2}_{m_0,L}\mathbf{Q}_{m_0,m_0}))(-u) \right)$$

$$T^3_{m_0,i} = \sum_{m=1}^{M} \tilde{g}\left(\frac{\|\mathbf{x}_m\|^2}{N+L-1} \right) \left(\sum_{u=-(L-1)}^{L-1} g_{N,\epsilon}(\hat{r}^\circ_m(u))\, \tau(\mathcal{D}_{m,L}(\mathcal{R}^{1/2}_{m_0,L}\frac{\partial \mathbf{Q}_{m,m}}{\partial \mathbf{x}_{m_0,i}}))(-u) \right)$$

We first address the behaviour of $\sum_{m_0,i} |T^1_{m_0,i}|^2$. For this, we first remark that

$$\left| \tilde{g}'\left(\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1} \right) \right|^2 \leq \kappa\, \mathbf{1}_{\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1} \leq 3}.$$

Now, using the inequality $|g_{N,\epsilon}(t)| \leq 2\frac{N^\epsilon}{\sqrt{N}}$ together with the Cauchy-Schwartz inequality, we obtain

$$\left| \sum_{u=-(L-1)}^{L-1} g_{N,\epsilon}(\hat{r}_{m_0}(u) - \mathbb{E}(\hat{r}_{m_0}(u))\, \tau(\mathcal{D}_{m_0,L}(\mathcal{R}^{1/2}_{m_0,L}\mathbf{Q}_{m_0,m_0}))(-u) \right|^2 \leq$$

$$\leq \kappa\, L\, \frac{N^{2\epsilon}}{N} \sum_{u=-(L-1)}^{L-1} \left| \tau(\mathcal{D}_{m_0,L}(\mathcal{R}^{1/2}_{m_0,L}\mathbf{Q}_{m_0,m_0}))(u) \right|^2.$$

Now, the inequality in (1.37) together with (2.41) imply that, for each $L \times L$ matrix $\mathbf{A}$, we have

$$\sum_{u=-(L-1)}^{L-1} |\tau(\mathcal{D}_{m_0,L}(\mathbf{A}))(u)|^2 \leq \kappa\, \frac{1}{L} \mathrm{Tr}(\mathbf{A}\mathbf{A}^H) \qquad (3.31)$$

for some nice constant $\kappa > 0$. Using this for $\mathbf{A} = \mathcal{R}_{m_0,L}^{1/2}\mathbf{Q}_{m_0,m_0}$ leads to the conclusion that

$$\sum_{u=-(L-1)}^{L-1} \left| \tau(\mathcal{D}_{m_0,L}(\mathcal{R}_{m_0,L}^{1/2}\mathbf{Q}_{m_0,m_0}))(u) \right|^2 \leq \kappa \frac{1}{L} \text{Tr}\left(\mathbf{Q}_{m_0,m_0}\mathbf{Q}_{m_0,m_0}^H\right). \qquad (3.32)$$

Therefore, we obtain that

$$\left|T_{m_0,i}^1\right|^2 \leq \frac{\kappa}{(N+L-1)} \frac{LN^{2\epsilon}}{N} \frac{|\mathbf{x}_{m_0,i}|^2}{N+L-1} \mathbf{1}_{\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}\leq 3} \frac{1}{L} \text{Tr}\left(\mathbf{Q}_{m_0,m_0}\mathbf{Q}_{m_0,m_0}^H\right)$$

from where we deduce that

$$\sum_{i=1}^{N+L-1} \left|T_{m_0,i}^1\right|^2 \leq \frac{\kappa}{(N+L-1)} \frac{LN^{2\epsilon}}{N} \frac{1}{L} \text{Tr}\left(\mathbf{Q}_{m_0,m_0}\mathbf{Q}_{m_0,m_0}^H\right)$$

and

$$\sum_{m_0=1}^{M} \sum_{i=1}^{N+L-1} \left|T_{m_0,i}^1\right|^2 \leq \kappa \frac{N^{2\epsilon}}{N} \frac{1}{ML} \text{Tr}\left(\mathbf{Q}\mathbf{Q}^H\right) \leq \kappa \frac{N^{2\epsilon}}{N} \frac{1}{(\text{Im}z)^2}.$$

We now consider $\sum_{m_0,i} |T_{m_0,i}^2|^2$. We first remark that

$$\left| \tilde{g}\left(\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}\right) \right|^2 \leq \kappa \mathbf{1}_{\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}\leq 3}$$

and that $\left|g'_{N,\epsilon}(\hat{r}_{m_0}(u) - \mathbb{E}(\hat{r}_{m_0}(u))\right|^2 \leq \kappa$. Therefore, a direct use of the Cauchy-Schwartz inequality and (3.32) leads us to

$$\left|T_{m_0,i}^2\right|^2 \leq \kappa \mathbf{1}_{\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}\leq 3} \left(\sum_u \left|\frac{\partial \hat{r}_{m_0}(u)}{\partial \mathbf{x}_{m_0,i}}\right|^2\right) \frac{1}{L} \text{Tr}\left(\mathbf{Q}_{m_0,m_0}\mathbf{Q}_{m_0,m_0}^H\right).$$

Using (3.20), we obtain that

$$\sum_{i=1}^{N+L-1} \left|\frac{\partial \hat{r}_{m_0}(u)}{\partial \mathbf{x}_{m_0,i}}\right|^2 \leq \frac{\kappa}{N+L-1} \frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}$$

and eventually that

$$\sum_{m_0,i} \left|T_{m_0,i}^2\right|^2 \leq \kappa \frac{1}{ML} \text{Tr}\left(\mathbf{Q}\mathbf{Q}^H\right) \leq \frac{\kappa}{(\text{Im}z)^2}.$$

We finally study $\sum_{m_0,i} |T_{m_0,i}^3|^2$. Using once more the fact that $\left|g_{N,\epsilon}(t)\right|^2 \leq 4\frac{N^{2\epsilon}}{N}$ and that $\left|\tilde{g}\left(\frac{\|\mathbf{x}_{m_0}\|^2}{N+L-1}\right)\right|^2 \leq \kappa$, Jensen inequality leads immediately that

$$\sum_{m_0,i} |T_{m_0,i}^3|^2 \leq \frac{\kappa N^{2\epsilon}}{N} ML \sum_{m=1}^{M} \sum_{u=-(L-1)}^{L-1} \sum_{m_0,i} \left| \tau\left(\mathcal{D}_{m,L}\left(\mathcal{R}_{m,L}^{1/2}\frac{\partial \mathbf{Q}_{m,m}}{\partial \mathbf{x}_{m_0,i}}\right)\right)(u)\right|^2$$

$$(3.33)$$

48

In order to evaluate $\sum_{m_0,i} |T^3_{m_0,i}|^2$, we use (2.39) and observe that we can write

$$\tau\left(\mathcal{D}_{m,L}\left(\mathcal{R}^{1/2}_{m,L}\frac{\partial\mathbf{Q}_{m,m}}{\partial\mathbf{x}_{m_0,i}}\right)\right)(u) = \frac{1}{L}\mathrm{Tr}\left(\frac{\partial\mathbf{Q}_{m,m}}{\partial\mathbf{x}_{m_0,i}}\mathcal{D}_{m,L}(\mathbf{J}^u_L)\mathcal{R}^{1/2}_{m,L}\right) = \frac{1}{L}\mathrm{Tr}\left(\frac{\partial\mathbf{Q}}{\partial\mathbf{x}_{m_0,i}}\mathbf{A}\right)$$

where $\mathbf{A}$ is the $ML \times ML$ matrix defined by $\mathbf{A} = \mathbf{E}_m\mathcal{D}_{m,L}(\mathbf{J}^u_L)\mathcal{R}^{1/2}_{m,L}\mathbf{E}^H_m$. Lemma 3.1 leads immediately to

$$\sum_{m_0,i}\left|\tau\left(\mathcal{D}_{m,L}\left(\mathcal{R}^{1/2}_{m,L}\frac{\partial\mathbf{Q}_{m,m}}{\partial\mathbf{x}_{m_0,i}}\right)\right)(u)\right|^2 \leq \frac{\kappa}{N}\frac{(1+|z|)}{(\mathrm{Im}z)^3}\left(1+\frac{1}{\mathrm{Im}z}\right)$$

Plugging this into the evaluation (3.33) eventually leads to

$$\sum_{m_0,i}|T^3_{m_0,i}|^2 \leq \kappa\, N^{2\epsilon}\frac{(1+|z|)}{(\mathrm{Im}z)^3}\left(1+\frac{1}{\mathrm{Im}z}\right)$$

This establishes (3.30) and Lemma 3.2. Therefore, we have shown that, considered as a function of $\mathbf{x}$, $\tilde{\zeta}_\epsilon$ is Lipschitz with constant $\kappa\frac{N^\epsilon}{M}$. The Gaussian concentration inequality thus implies that $\left|\tilde{\zeta}_\epsilon - \mathbb{E}(\tilde{\zeta}_\epsilon)\right| \prec \frac{N^\epsilon}{M}$.

It remains to justify that $\left|\tilde{\zeta}_\epsilon - \mathbb{E}(\tilde{\zeta}_\epsilon)\right| \prec \frac{N^\epsilon}{M}$ implies that $|\zeta - \mathbb{E}(\zeta)| \prec \frac{N^\epsilon}{M}$. For this, it is sufficient to follow the proof of [29, Lemma 4.1, p.41].

4. Evaluation of the modified statistic $\overline{\phi}_N$

4.1. *Reduction to the study of the expectation of $\overline{\phi}_N$*

In this short section, we show that we can reduce the study of the statistic $\overline{\phi}_N$ to the study of the expectation $\mathbb{E}(\overline{\phi}_N)$ up to an error that is dominated by $\frac{1}{M\sqrt{L}}$. We express the result in terms of the following proposition.

Proposition 4.1. *Let Assumptions 1-3 and 5 hold true and let ϕ have the same properties as in the statement of Theorem 3.1. Then,*

$$\left|\overline{\phi}_N - \mathbb{E}\left(\overline{\phi}_N\right)\right| \prec \frac{1}{M\sqrt{L}}. \tag{4.1}$$

We devote the rest of the section to the proof of this result. We first reason that, without loss of generality, we can replace the function ϕ by a smooth function that is supported by $[-\delta, \alpha]$ for some $\alpha > \alpha_0$ (see the statement of Theorem 3.1 for a definition of α_0, δ). The justification is the same that we used at the initial steps of the proof of Theorem 3.1 and is therefore omitted. We therefore focus on this class functions for the rest of the proof.

In order to show Proposition 4.1, consider again the Helffer-Sjöstrand representation of $\overline{\phi}_N$ in (3.3), which allows us to write

$$\overline{\phi}_N^\circ = \overline{\phi}_N - \mathbb{E}\overline{\phi}_N = \frac{1}{\pi}\mathrm{Re}\int_{\mathcal{D}} dx\,dy\,\overline{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\mathrm{Tr}\left(\mathbf{Q}_N(z) - \mathbb{E}\mathbf{Q}_N(z)\right). \quad (4.2)$$

Here again, the idea is to consider $\overline{\phi}_N^\circ$ as a function of the $\mathcal{N}_{\mathbb{C}}(0,\mathbf{I}_{M(N+L-1)})$-distributed random vector $\mathbf{x}$ defined in (3.16). We will show that this function is Lipschitz with constant of order $\mathcal{O}((M\sqrt{L})^{-1})$, so that the result follows from conventional concentration results of Gaussian functionals in (1.32) (see also [38, Theorem 2.1.12]).

Indeed, let $\nabla\overline{\phi}_N^\circ$ denote the gradient of $\overline{\phi}_N^\circ$ with respect to $\mathbf{x}$. Then, we can obviously write

$$\left\|\nabla\overline{\phi}_N^\circ\right\|^2 = \sum_{m_0=1}^{M}\sum_{i=1}^{N+L-1}\left|\frac{\partial\overline{\phi}_N}{\partial\mathbf{x}_{m_0,i}}\right|^2 + \left|\frac{\partial\overline{\phi}_N}{\partial\mathbf{x}_{m_0,i}^*}\right|^2$$

where we recall that $\mathbf{x}_{m_0,i}$ denotes the ith entry of $\mathbf{x}_m$, the mth block of $\mathbf{x}$, with dimension $N + L - 1$. A direct use of Lemma 3.1 shows that

$$\sum_{m_0=1}^{M}\sum_{i=1}^{N+L-1}\left|\frac{\partial\overline{\phi}_N}{\partial\mathbf{x}_{m_0,i}}\right|^2 \leq \frac{\kappa}{MN}\int_{\mathcal{D}} dx\,dy\,\left|\overline{\partial}\Phi_k(\phi)(z)\right|^2\frac{1+|z|}{(\mathrm{Im}z)^3}\left(1+\frac{1}{\mathrm{Im}z}\right)$$

for some nice constant κ, where the integral on the right hand side is finite if we select $k \geq 4$, which is always possible because ϕ is smooth. This concludes the proof of Proposition 4.1.

4.2. Weak convergence of $\overline{\mu}_N(\lambda)$ and evaluation of $\mathbb{E}(\overline{\phi}_N)$

The aim of this section is twofold. On the one hand, we will show that $\overline{\mu}_N(\lambda) - \mu_N(\lambda)$ converges weakly almost surely to zero. On the other hand, we will evaluate the convergence of the $\mathbb{E}(\overline{\phi}_N)$ by establishing that, when $\frac{L^{3/2}}{MN} \to 0$ (equivalently $\beta < 4/5$), we have

$$\left|\mathbb{E}\left(\overline{\phi}_N\right) - \int\phi(\lambda)d\mu_N(\lambda)\right| \leq \kappa\frac{L}{MN} \quad (4.3)$$

for some nice constant $\kappa > 0$.

We will address the problem by studying $\frac{1}{ML}\mathrm{Tr}\left(\mathbb{E}\mathbf{Q}_N(z) - \mathbf{T}_N(z)\right)$. We will study this term by conveniently adapting the tools in [27] to the present context. First, we study the master equations that define the matrix function $\mathbf{T}_N(z)$ in the statement of Theorem 1.1 and establish existence and unicity of the solution using again the tools developed in [27, Section 5]. We then establish that, considering a

sequence of $ML \times ML$ deterministic matrices $\mathbf{A}_N$ of uniformly bounded spectral norm, we have

$$\left|\frac{1}{ML}\operatorname{Tr}\left[(\mathbb{E}\mathbf{Q}_N(z) - \mathbf{T}_N(z))\,\mathbf{A}_N\right]\right| \leq \kappa\frac{L}{MN}$$

for z in a certain subset of $\mathbb{C}^+$, assuming that $\beta < 4/5$. Even if the subset where the above inequality holds is not the whole semiplane $\mathbb{C}^+$, it will be sufficient to deduce (4.3) by conveniently adapting the arguments in [2, Lemma 5.5.5].

First of all, we consider here the two asymptotic equivalents $\mathbf{T}(z), \widetilde{\mathbf{T}}(z)$, as the solutions to the equations (1.20)-(1.21).

Proposition 4.2. *There exists a unique pair of functions* $(\mathbf{T}(z), \widetilde{\mathbf{T}}(z)) \in \mathcal{S}_{ML}(\mathbb{R}^+) \times \mathcal{S}_N(\mathbb{R}^+)$ *that satisfy (1.20)–(1.21) for each* $z \in \mathbb{C}^+$. *Moreover, one can find two nice constants* η *and* $\tilde{\eta}$ *such that*

$$\mathbf{T}(z)\mathbf{T}^H(z) \geq \frac{(\operatorname{Im}z)^2}{16(\eta^2 + |z|^2)^2}\mathbf{I}_{ML} \tag{4.4}$$

$$\widetilde{\mathbf{T}}(z)\widetilde{\mathbf{T}}^H(z) \geq \frac{(\operatorname{Im}z)^2}{16(\tilde{\eta}^2 + |z|^2)^2}\mathbf{I}_N. \tag{4.5}$$

The proof follows the steps as the proof of Proposition 5.1 in [27]. To prove existence, we consider the composition of (1.20)–(1.21) as a mapping in the set of $ML \times ML$ block diagonal matrices. Using Proposition 1.2 one can establish that iterating these two equations one can create a sequence of $ML \times ML$ diagonal block matrices with blocks belonging to the class $\mathcal{S}_L(\mathbb{R}^+)$ that has a limit in this set. Then, in a second step, it can be shown that this limit is a solution to the canonical equation. For more details, the reader may refer to the proof of Proposition 5.1 in [27].

The proof of unicity follows the same path that was established in [27]. More specifically, assume that $\mathbf{T}(z), \widetilde{\mathbf{T}}(z)$ and $\mathbf{S}(z), \widetilde{\mathbf{S}}(z)$ are matrices solutions of the system (1.20, 1.21) of equations at point z, and assume that $\mathbf{T}(z)$ and $\mathbf{S}(z)$ have positive imaginary parts. Let $\mathbf{T}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2}\mathbf{T}(z)\mathcal{B}_L^{-1/2}$ and $\mathbf{S}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2}\mathbf{S}(z)\mathcal{B}_L^{-1/2}$. It is easily seen that

$$\mathbf{T}_{\mathcal{B}}(z) - \mathbf{S}_{\mathcal{B}}(z) = \Phi_{\mathcal{B},0}\left(\mathbf{T}_{\mathcal{B}}(z) - \mathbf{S}_{\mathcal{B}}(z)\right) \tag{4.6}$$

where we have defined the operator $\Phi_{\mathcal{B},0}(\mathbf{X})$ as

$$\Phi_{\mathcal{B},0}(\mathbf{X}) = z^2 c_N \mathbf{S}_{\mathcal{B}}(z)\Psi\left(\widetilde{\mathbf{S}}^T(z)\overline{\Psi}(\mathbf{X})\widetilde{\mathbf{T}}^T(z)\right)\mathbf{T}_{\mathcal{B}}(z) \tag{4.7}$$

where $\mathbf{X}$ is an $ML \times ML$ matrix. This operator is the analog of $\Phi_0(\mathbf{X})$ in [27] translated to our current matrix model. Operating like in [27] we write $\Phi_{\mathcal{B},0}^{(1)}(\mathbf{X}) = \Phi_{\mathcal{B},0}(\mathbf{X})$ and recursively define $\Phi_{\mathcal{B},0}^{(n+1)}(\mathbf{X}) = \Phi_{\mathcal{B},0}(\Phi_{\mathcal{B},0}^{(n)}(\mathbf{X}))$ for $n \geq 1$. By (4.6), unicity is proven if we are able to show that $\lim_{n\to\infty}\Phi_{\mathcal{B},0}^{(n)}(\mathbf{X}) = \mathbf{0}$ for every $ML \times$

ML matrix $\mathbf{X}$. Now, using Proposition 1.3 it is easily established that, for any two L-dimensional column vectors $\mathbf{a}$, $\mathbf{b}$, we can write

$$\left| \mathbf{a}^H \left(\Phi_{\mathcal{B},0}^{(n)} (\mathbf{X}) \right)_{m,m} \mathbf{b} \right| \leq \left[\mathbf{a}^H \left(\Phi_{\mathbf{S}_{\mathcal{B}}}^{(n)} \left(\mathbf{X}\mathbf{X}^H \right) \right)_{m,m} \mathbf{a} \right]^{1/2} \left[\mathbf{b}^H \left(\Phi_{\mathbf{T}_{\mathcal{B}}^H}^{(n)} (\mathbf{I}_{ML}) \right)_{m,m} \mathbf{b} \right]^{1/2}$$

(4.8)

where $\Phi_{\mathbf{T}_{\mathcal{B}}^H}$ and $\Phi_{\mathbf{S}_{\mathcal{B}}}$ are the positive operators defined by

$$\Phi_{\mathbf{T}_{\mathcal{B}}^H} (\mathbf{X}) = |z|^2 \, c_N \mathbf{T}_{\mathcal{B}}^H (z) \Psi \left(\widetilde{\mathbf{T}}^* (z) \overline{\Psi} (\mathbf{X}) \widetilde{\mathbf{T}}^T (z) \right) \mathbf{T}_{\mathcal{B}}(z) \tag{4.9}$$

$$\Phi_{\mathbf{S}_{\mathcal{B}}} (\mathbf{X}) = |z|^2 \, c_N \mathbf{S}_{\mathcal{B}}(z) \Psi \left(\widetilde{\mathbf{S}}^T (z) \overline{\Psi} (\mathbf{X}) \widetilde{\mathbf{S}}^* (z) \right) \mathbf{S}_{\mathcal{B}}^H (z). \tag{4.10}$$

Thus, by Proposition 1.3, $\lim_{n\to\infty} \Phi_{\mathcal{B},0}^{(n)}(\mathbf{X}) = \mathbf{0}$ will follow directly if we are able prove that there exist two positive definite matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ such that $\Phi_{\mathbf{T}_{\mathcal{B}}^H}^{(n)} (\mathbf{Y}_1)$ and $\Phi_{\mathbf{S}_{\mathcal{B}}}^{(n)} (\mathbf{Y}_2)$ converge towards $\mathbf{0}$.

Lemma 4.1. *Let $\mathbf{T}(z), \widetilde{\mathbf{T}}(z)$ be a solution to the canonical equation (1.20, 1.21) at point $z \in \mathbb{C}^+$ satisfying $\mathrm{Im}(\mathbf{T}(z)) \geq 0$, and define $\mathbf{T}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2}\mathbf{T}(z)\mathcal{B}_L^{-1/2}$. Let $\mathbf{X}$ be a positive semi definite matrix. Then, it holds that*

$$\Phi_{\mathbf{T}_{\mathcal{B}}}^{(n)} (\mathbf{X}) \to \mathbf{0} \tag{4.11}$$

and

$$\Phi_{\mathbf{T}_{\mathcal{B}}^H}^{(n)} (\mathbf{X}) \to \mathbf{0} \tag{4.12}$$

as $n \to \infty$. Moreover, the series $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B}}}^{(n)} (\mathbf{X})$ and $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B}}^H}^{(n)} (\mathbf{X})$ converge.

Proof. The proof of the lemma follows the same steps as the proof of Lemma 5.4 in [27] and is therefore omitted. $\qquad\square$

As a consequence of all the above, Theorem 1.1 will be a direct implication of the following result.

Proposition 4.3. *We consider a sequence $(\mathbf{A}_N)_{N\geq 1}$ of $ML \times ML$ deterministic matrices such that $\sup_N \|\mathbf{A}_N\| \leq a$ for some nice constant a. Then, for each $z \in \mathbb{C}^+$, we have*

$$\frac{1}{ML}\mathrm{Tr}\left(\mathbf{A}_N(\mathbf{Q}_N(z) - \mathbf{T}_N(z))\right) \to 0 \tag{4.13}$$

almost surely. For any bounded continuous function ϕ we have

$$\left| \frac{1}{ML}\mathrm{Tr}\left(\phi(\overline{\mathcal{R}}_{\mathrm{corr},L})\right) - \int \phi(\lambda)d\mu_N(\lambda) \right| \to 0 \tag{4.14}$$

almost surely.

Assume, in addition, that $\frac{L^{3/2}}{MN} \to 0$, i.e. that $\beta < \frac{4}{5}$. In this case, we have

$$\left| \frac{1}{ML} \mathrm{Tr} \left(\mathbf{A}_N (\mathbb{E} \mathbf{Q}_N(z) - \mathbf{T}_N(z)) \right) \right| \leq C(z) \frac{L}{MN} \tag{4.15}$$

when z belongs to a set E_N defined as

$$E_N = \left\{ z \in \mathbb{C}^+, \frac{L^{3/2}}{MN} P_1(|z|) P_2(1/\mathrm{Im} z) < 1 \right\}$$

and where P_1 and P_2 are two nice polynomials. Finally, for each compactly supported smooth function ϕ, we have

$$\left| \frac{1}{ML} \mathbb{E} \mathrm{Tr} \left(\phi(\overline{\mathcal{R}}_{\mathrm{corr},L}) \right) - \int \phi(\lambda) d\mu_N(\lambda) \right| \leq \kappa \frac{L}{MN} \tag{4.16}$$

for some nice constant $\kappa > 0$.

Proof. The proof of (4.13) can be established by essentially following the approach in [27]. The main idea is to consider the resolvent in (1.30) together with the co-resolvent, defined as

$$\widetilde{\mathbf{Q}}_N(z) = \left(\mathbf{W}_N^H \mathcal{B}_L^{-1} \mathbf{W}_N - z \mathbf{I}_N \right)^{-1}.$$

Using a trivial modification of [27, Lemma 3.1] one can reduce the problem to the study of the expectations $\mathbb{E} \mathbf{Q}_N(z)$ and $\mathbb{E} \widetilde{\mathbf{Q}}_N(z)$. We can then introduce two matrix-valued functions $\mathbf{R}_N(z)$ and $\widetilde{\mathbf{R}}_N(z)$ defined as

$$\widetilde{\mathbf{R}}_N(z) = -\frac{1}{z} \left(\mathbf{I}_N + c_N \overline{\Psi}^T \left(\mathcal{B}_L^{-1/2} \mathbb{E} \mathbf{Q}_N(z) \mathcal{B}_L^{-1/2} \right) \right)^{-1} \tag{4.17}$$

$$\mathbf{R}_N(z) = -\frac{1}{z} \left(\mathbf{I}_{ML} + \mathcal{B}_L^{-1/2} \Psi \left(\widetilde{\mathbf{R}}_N^T(z) \right) \mathcal{B}_L^{-1/2} \right)^{-1} \tag{4.18}$$

which are the analogous of the same quantities in [27, Section 4]. In particular, one can establish that Lemma 4.1 and Proposition 4.3 in [27] also hold true with these new definitions, so that

$$\left| \frac{1}{ML} \mathrm{Tr} \left[\mathbf{A}_N (\mathbb{E} \mathbf{Q}_N(z) - \mathbf{R}_N(z)) \right] \right| \leq C(z) \frac{L}{MN}$$

for all $z \in \mathbb{C}^+$. In order to see this, we need to make explicit use of Assumption 4. At this point, in order to show (4.13) and (4.15) one only needs to evaluate the quantity $\mathbf{R}_N(z) - \mathbf{T}_N(z)$ using the approach in Section 6 of [27], which essentially holds verbatim after replacing the operators $\Phi_1(\mathbf{X})$ and $\Phi_1^t(\mathbf{X})$ with

$$\Phi_{\mathcal{B},1}(\mathbf{X}) = z^2 c_N \mathbf{R}_{\mathcal{B}}(z) \Psi \left(\widetilde{\mathbf{R}}_N^T(z) \overline{\Psi}(\mathbf{X}) \widetilde{\mathbf{T}}_N^T(z) \right) \mathbf{T}_{\mathcal{B}}(z) \tag{4.19}$$

$$\Phi_{\mathcal{B},1}^t(\mathbf{X}) = z^2 c_N \Psi \left(\widetilde{\mathbf{T}}_N^T(z) \overline{\Psi}(\mathbf{T}_{\mathcal{B}}(z) \mathbf{X} \mathbf{R}_{\mathcal{B}}(z)) \widetilde{\mathbf{R}}_N^T(z) \right) \tag{4.20}$$

where now $\mathbf{R}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2} \mathbf{R}_N(z) \mathcal{B}_L^{-1/2}$. In particular, (4.15) will follow the arguments in [27, Section 6.1], which basically requires the application of Montel's

theorem. To see that (4.13) implies (4.14) we need to check that $(\bar{\mu}_N)_{N \geq 1}$ is almost surely tight and $(\mu_N)_{N \geq 1}$ is tight (see [15, Corollary 2.7]). The fact that $(\bar{\mu}_N)_{N \geq 1}$ is almost surely tight follows from the fact that

$$\int_{\mathbb{R}^+} \lambda d\bar{\mu}_N(\lambda) = \frac{1}{ML} \mathrm{Tr} \mathcal{B}_L^{-1/2} \mathbf{W}_N \mathbf{W}_N^H \mathcal{B}_L^{-1/2} = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L} \mathrm{Tr} \left[\mathcal{R}_{m,L}^{-1/2} \hat{\mathcal{R}}_{m,L} \mathcal{R}_{m,L}^{-1/2} \right].$$

The identity in (2.26) implies that

$$\sup_{m=1,\dots,M} \left| \frac{1}{L} \mathrm{Tr} \left[\mathcal{R}_{m,L}^{-1/2} \hat{\mathcal{R}}_{m,L} \mathcal{R}_{m,L}^{-1/2} \right] - 1 \right| \to 0, \; a.s.$$

Therefore, $\int_{\mathbb{R}^+} \lambda d\bar{\mu}_N(\lambda) \to 1$ almost surely, and tightness holds with probability one. To verify that $(\mu_N)_{N \geq 1}$ is tight, we evaluate $\int_{\mathbb{R}^+} \lambda d\boldsymbol{\mu}_N(\lambda)$ using item (v) of Proposition 1.1 and immediately obtain that $\int_{\mathbb{R}^+} \lambda d\boldsymbol{\mu}_N(\lambda) = \mathbf{I}_{ML}$ and $\int_{\mathbb{R}^+} \lambda d\mu_N(\lambda) = 1$, so tightness established.

To establish (4.15) when $\beta < 4/5$, we follow the corresponding arguments in [27, Section 6.2]. Regarding (4.16), it will be a direct consequence of [6, Lemma 5.5.5][3] provided that we are able to show that, given two nice constants C_0, C_0', there exist three nice constants C_1, C_2, C_3 and an integer N_0 such that

$$\left| \frac{1}{ML} \mathrm{Tr} \left(\mathbb{E} \mathbf{Q}_N(z) - \mathbf{T}_N(z) \right) \right| \leq C_2 \frac{L}{MN} \frac{1}{(\mathrm{Im}z)^{C_3}}$$

for all z inside the domain $|\mathrm{Re}z| \leq C_0$, $N^{-C_1} \leq \mathrm{Im}z \leq C_0'$ and $N > N_0$. For this, it is sufficient to to follow the arguments used to establish Theorem 10.1 in [26]. $\square$

Remark 4.1. We notice that (4.16) is just established for compactly supported functions ϕ. In order to extend (4.16) to non compactly supported ϕ, it would be necessary to establish that the support of μ_N is included for each N large enough in a compact subset independent from N. While we feel that this property holds, its proof does not seem obvious. In Section 6 we provide an example of non-compactly supported ϕ for which (4.16) still holds.

5. Approximation by a Marchenko-Pastur distribution

Let us denote by $t_N(z)$ the Stieltjes transform of the Marcenko-Pastur law μ_{mp,c_N} associated to the parameter $c_N = \frac{ML}{N}$. In other words, for each $z \in \mathbb{C}^+$, $t_N(z)$ is the unique solution of the equation

$$t_N(z) = \frac{1}{-z + \frac{1}{1+c_N t_N(z)}} \tag{5.1}$$

---

[3] The statement of [6, Lemma 5.5.5] requires that the function ϕ vanishes on the support of μ_N. However, the reader may check that this assumption is in fact not needed.

for which $\text{Im}(t_N(z)) \geq 0$. If $\mathbf{T}_N(z)$ represents the deterministic equivalent of $\mathbf{Q}_N(z)$, solution of the equations (1.20, 1.21), the following theorem establishes that, for each $\gamma < \gamma_0$, $\gamma \neq 1$, the Stieltjes transform $\frac{1}{ML}\text{Tr}\mathbf{T}_N(z)$ is well approximated by $t_N(z)$, up to an error of order $\mathcal{O}(L^{-2\min(1,\gamma)})$.

The strategy of the proof follows two steps. In a first step, we will establish that the spectral norm of the error between the two Stieltjes transforms $\|\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\|$ is upper bounded by a term that decays as $L^{-\min(1,\gamma)}$ for each $\gamma < \gamma_0, \gamma \neq 1$. In a second stage, this result is used to obtain a refined convergence rate for the normalized trace of the result, so that, in fact

$$\frac{1}{ML}\text{Tr}\left(\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\right) \leq \frac{1}{L^{2\min(\gamma,1)}}P_1(z)P_2\left(\frac{1}{\text{Im}z}\right) \qquad (5.2)$$

for each $z \in \mathbb{C}^+$ and for two nice polynomials $P_1(z)$, $P_2(z)$.

We observe here that a direct application of the above result to the Helffer-Sjöstrand formula implies (1.25) in Theorem 1.1. Indeed, observe that in this case we can write

$$\int_{\mathbb{R}^+} \phi(\lambda)d\mu_N(\lambda) - \int_{\mathbb{R}^+} \phi(\lambda)d\mu_{mp,N}(\lambda) =$$
$$= \frac{1}{\pi}\text{Re}\int_{\mathcal{D}} dx\, dy\, \overline{\partial}\Phi_k(\phi)(z)\frac{1}{ML}\text{Tr}\left(\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\right).$$

If k is taken to be larger than or equal to the degree of P_2 in (5.2), this directly shows (1.25). On the other hand, from the convergence of $\frac{1}{ML}\text{Tr}\left(\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\right)$ for all $z \in \mathbb{C}^+$ to zero together with the fact that both $(\mu_N)_{N\geq 1}$ and $(\mu_{mp,c_N})_{N\geq 1}$ are tight[4], we see that $\mu_N - \mu_{mp,c_N}$ converges weakly to zero. But since μ_{mp,c_N} in turn converges weakly to $\mu_{mp,c_\star}$, the proof of Theorem 1.1 is completed.

Remark 5.1. We again notice that (1.25) is established for compactly supported smooth functions ϕ. As in the context of Remark 4.1, the generalization of (1.25) to non compactly supported functions would need to prove that the support of μ_N is included in a compact independent of N.

We will present the two stages of the proof in two separate subsections that follow. In order to simplify the notation, we will drop from now on the subindex N in all relevant quantities, i.e. $t_N(z), \tilde{t}_N(z), c_N, \mathbf{T}_N(z), \widetilde{\mathbf{T}}_N(z)$, etc.

---

[4]Tightness of $(\mu_N)_{N\geq 1}$ has been established before, whereas tightness of $(\mu_{mp,c_N})_{N\geq 1}$ follows from the fact that $c_N \to c_\star$.

5.1. Bounding the spectral norm $\|\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\|$

The objective of this section is to prove the following result.

Theorem 5.1. *Under Assumptions 2, 3 and 5, there exist two nice polynomials P_1 and P_2 as given in Definition 6, such that for each $\gamma < \gamma_0$, $\gamma \neq 1$, the inequality*

$$\|\mathbf{T}_N(z) - t_N(z)\mathbf{I}_{ML}\| \leq \frac{1}{L^{\min(\gamma,1)}} P_1(|z|) P_2\left(\frac{1}{\mathrm{Im}z}\right) \tag{5.3}$$

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

We devote the rest of this section to the proof of Theorem 5.1. First of all, it is well known that the function $\tilde{t}(z) = ct(z) - \frac{1-c}{z}$ coincides with the Stieltjes transform of the probability measure $c\mu_{mp,c} + (1-c)\delta_0$ and is equal to

$$\tilde{t}(z) = -\frac{1}{z(1 + ct(z))} \tag{5.4}$$

so that $t(z)$ can also be written as

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

Consider here the two matrix-valued functions $\widetilde{\mathbf{T}}_{mp}(z)$ and $\mathbf{T}_{mp}(z)$ defined by

$$\widetilde{\mathbf{T}}_{mp}(z) = -\frac{1}{z}\left(\mathbf{I}_N + c_N \overline{\Psi}^T \left(\mathcal{B}_L^{-1/2} t(z)\mathbf{I}_{ML} \mathcal{B}_L^{-1/2}\right)\right)^{-1} \tag{5.5}$$

$$\mathbf{T}_{mp}(z) = -\frac{1}{z}\left(\mathbf{I}_{ML} + \mathcal{B}_L^{-1/2}\Psi\left(\widetilde{\mathbf{T}}_{mp}^T(z)\right)\mathcal{B}_L^{-1/2}\right)^{-1}. \tag{5.6}$$

According to Proposition 1.2, these functions belong to $\mathcal{S}_N(\mathbb{R}^+)$ and $\mathcal{S}_{ML}(\mathbb{R}^+)$ respectively, and verify the various properties of functions $\widetilde{\mathbf{\Upsilon}}(z)$ and $\mathbf{\Upsilon}(z)$ defined in the statement of that proposition. In order to establish Theorem 5.1, we define $\Delta_{mp}(z)$ by

$$\Delta_{mp}(z) = t(z)\mathbf{I}_{ML} - \mathbf{T}_{mp}(z) \tag{5.7}$$

and express $t(z)\mathbf{I}_{ML} - \mathbf{T}(z)$ as

$$t(z)\mathbf{I}_{ML} - \mathbf{T}(z) = (\mathbf{T}_{mp}(z) - \mathbf{T}(z)) + \Delta_{mp}(z). \tag{5.8}$$

We also define $\mathbf{t}_\mathcal{B}(z)$, $\mathbf{T}_{\mathcal{B},mp}(z)$ and $\Delta_{\mathcal{B},mp}(z)$ by $\mathbf{t}_\mathcal{B}(z) = \mathcal{B}_L^{-1/2} t(z)\mathbf{I}_{ML} \mathcal{B}_L^{-1/2}$, $\mathbf{T}_{\mathcal{B},mp}(z) = \mathcal{B}_L^{-1/2}\mathbf{T}_{mp}(z)\mathcal{B}_L^{-1/2}$ and $\Delta_{\mathcal{B},mp}(z) = \mathcal{B}_L^{-1/2}\Delta_{mp}(z)\mathcal{B}_L^{-1/2}$ respectively. Using the definition of $\mathbf{T}_{mp}$ and $\widetilde{\mathbf{T}}_{mp}$ as well as the canonical equations (1.20, 1.21), we obtain easily that

$$\mathbf{T}_{\mathcal{B},mp}(z) - \mathbf{T}_\mathcal{B}(z) = \Phi_{\mathcal{B},2}\left(\mathbf{t}_\mathcal{B}(z) - \mathbf{T}_\mathcal{B}(z)\right) \tag{5.9}$$

where $\Phi_{\mathcal{B},2}$ is the linear operator acting on $ML \times ML$ matrices defined as

$$\Phi_{\mathcal{B},2}(\mathbf{X}) = cz^2 \mathbf{T}_{\mathcal{B},mp}(z)\Psi\left(\widetilde{\mathbf{T}}_{mp}^T(z)\overline{\Psi}\left(\mathbf{X}\right)\widetilde{\mathbf{T}}^T(z)\right)\mathbf{T}_\mathcal{B}(z). \tag{5.10}$$

Using this definition, we can re-write (5.8) as

$$\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z) = \Phi_{\mathcal{B},2}\left(\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\right) + \Delta_{\mathcal{B},mp}(z). \tag{5.11}$$

Our approach is to use Proposition 1.3 in order to establish that

$$\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z) = \sum_{n=0}^{+\infty} \Phi_{\mathcal{B},2}^{(n)}\left(\Delta_{\mathcal{B},mp}(z)\right) \tag{5.12}$$

and that $\|\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\| \le C(z) \|\Delta_{\mathcal{B},mp}(z)\|$. The identity in (5.3) will then be established if we are able to show that $\|\Delta_{\mathcal{B},mp}(z)\| \le C(z) \frac{1}{L^{\min(1,\gamma)}}$ if $\gamma < \gamma_0$, $\gamma \ne 1$.

We begin by evaluating the spectral norm of $\Delta_{mp}(z)$ and $\Delta_{\mathcal{B},mp}(z)$. For this, we observe that $\widetilde{\mathbf{T}}_{mp}^T(z)$ is given by

$$\widetilde{\mathbf{T}}_{mp}^T(z) = -\frac{1}{z}\left(\mathbf{I}_N + ct(z)\overline{\Psi}\left(\mathcal{B}_L^{-1}\right)\right)^{-1}$$

where we can express $\overline{\Psi}\left(\mathcal{B}_L^{-1}\right)$ as

$$\overline{\Psi}\left(\mathcal{B}_L^{-1}\right) = \int_0^1 \frac{1}{M}\sum_{m=1}^M \mathcal{S}_m(\nu)\mathbf{a}_L^H(\nu)\mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu)\mathbf{d}_N(\nu)\mathbf{d}_N^H(\nu)\,d\nu.$$

Let us denote by $\mathbf{E}_N$ the $N \times N$ matrix defined by

$$\mathbf{E}_N = \int_0^1 \left(\frac{1}{M}\sum_{m=1}^M \epsilon_{m,L}(\nu)\right)\mathbf{d}_N(\nu)\,\mathbf{d}_N^H(\nu)\,d\nu \tag{5.13}$$

where $\epsilon_{m,L}(\nu)$ is defined by

$$\epsilon_{m,L}(\nu) = \mathcal{S}_m(\nu)\mathbf{a}_L^H(\nu)\,\mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu) - 1.$$

It is clear that $\overline{\Psi}\left(\mathcal{B}_L^{-1}\right) = \mathbf{I}_N + \mathbf{E}_N$, so that $\widetilde{\mathbf{T}}_{mp}^T(z)$ can be written as

$$\widetilde{\mathbf{T}}_{mp}^T(z) = \left[-z(1 + ct(z))\left(\mathbf{I}_N + \frac{ct(z)}{1 + ct(z)}\mathbf{E}_N\right)\right]^{-1}$$

or equivalently as

$$\widetilde{\mathbf{T}}_{mp}^T(z) = \tilde{t}(z)\,\mathbf{I}_N\left(\mathbf{I}_N - c\,z\,t(z)\,\tilde{t}(z)\mathbf{E}_N\right)^{-1}$$
$$= \tilde{t}(z)\mathbf{I}_N + czt(z)\tilde{t}^2(z)\mathbf{E}_N\left(\mathbf{I}_N - c\,z\,t(z)\,\tilde{t}(z)\mathbf{E}_N\right)^{-1}.$$

In order to express $\mathbf{T}_{mp}(z)$ in a convenient way, we define $\mathbf{\Gamma}(z)$ as the $ML \times ML$ block diagonal matrix given by

$$\mathbf{\Gamma}(z) = \Psi\left(\mathbf{E}_N\left(\mathbf{I}_{ML} - c\,z\,t(z)\,\tilde{t}(z)\mathbf{E}_N\right)^{-1}\right). \tag{5.14}$$

Using that $\Psi(\mathbf{I}_N) = \mathcal{B}_L$, we obtain

$$\mathbf{T}_{mp}(z) = \left[-z\left((1 + \tilde{t}(z))\mathbf{I}_{ML} + czt(z)\tilde{t}^2(z)\mathcal{B}_L^{-1/2}\mathbf{\Gamma}(z)\mathcal{B}_L^{-1/2}\right)\right]^{-1}$$

or, equivalently,

$$\mathbf{T}_{mp}(z) = t(z) \left(\mathbf{I}_{ML} - c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z)\right)^{-1}$$
$$= t(z)\mathbf{I}_{ML} + t(z)c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z) \left(\mathbf{I}_{ML} - c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z)\right)^{-1}$$

where $\mathbf{\Gamma}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2} \mathbf{\Gamma}(z) \mathcal{B}_L^{-1/2}$. We eventually obtain that

$$\Delta_{mp}(z) = -t(z)c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z) \left(\mathbf{I} - c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z)\right)^{-1}. \tag{5.15}$$

The asymptotic behaviour of $\Delta_{mp}(z)$ depends on the behaviour of matrix $\mathbf{E}_N$, which itself depends on the properties of the terms $(\epsilon_m(\nu))_{m=1,\dots,M}$. The following Lemma, established in the Appendix D, is the key point of the proof of Theorem 5.1.

Lemma 5.1. *For each $\gamma < \gamma_0$, it holds that*

$$\sup_{m \geq 1} \sup_{\nu \in [0,1]} |\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{L^{\min(\gamma,1)}} \tag{5.16}$$

for some nice constant κ (depending on γ) if $\gamma \neq 1$ while if $\gamma = 1$,

$$\sup_{m \geq 1} \sup_{\nu \in [0,1]} |\epsilon_{m,L}(\nu)| \leq \kappa \frac{\log L}{L}. \tag{5.17}$$

In the following, we use Lemma 5.1 for a value of γ as close as possible to γ_0 in order to obtain the fastest speed of convergence for $\sup_{m \geq 1} \sup_{\nu \in [0,1]} |\epsilon_{m,L}(\nu)|$. If $\gamma_0 \leq 1$, $\gamma < \gamma_0 \leq 1$ cannot be equal to 1. If $\gamma_0 > 1$, we will of course consider a value of γ for which $1 < \gamma < \gamma_0$. Therefore, in the following, we assume that $\gamma \neq 1$. If $\gamma_0 \leq 1$, we thus obtain that for each $\gamma < \gamma_0$

$$\sup_{m \geq 1} \sup_{\nu \in [0,1]} |\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{L^\gamma} \tag{5.18}$$

holds, while if $\gamma_0 > 1$,

$$\sup_{m \geq 1} \sup_{\nu \in [0,1]} |\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{L}. \tag{5.19}$$

Noting that $\mathbf{E}_N$ is the $N \times N$ Toeplitz matrix with symbol $\frac{1}{M} \sum_{m=1}^M \epsilon_{m,L}(\nu)$, we immediately infer from this discussion the following corollary.

Corollary 5.1. *If $\gamma_0 \leq 1$, then, for each $\gamma < \gamma_0$, there exists a nice constant κ depending on γ for which $\|\mathbf{E}_N\| \leq \frac{\kappa}{L^\gamma}$. If $\gamma_0 > 1$, there exists a nice constant κ such that $\|\mathbf{E}_N\| \leq \frac{\kappa}{L}$.*

In order to control the norm of $\mathbf{\Gamma}(z)$, we mention that for each $z \in \mathbb{C}^+$, then $c|zt(z)\tilde{t}(z)|^2 < 1$ (see e.g. Lemma 1.1 in [26]). Therefore, the inequalities $|zt(z)\tilde{t}(z)| \leq \frac{1}{\sqrt{c}}$ and $c|zt(z)\tilde{t}(z)| \leq \sqrt{c}$ hold on $\mathbb{C}^+$. Corollary 5.1 thus implies that for L large enough, $\|\mathbf{I}_N - c\,z\,t(z)\,\tilde{t}(z)\mathbf{E}_N\| > 1 - \sqrt{c}\|\mathbf{E}_N\| > \frac{1}{2}$ and

58

$\| \left(\mathbf{I}_N - c\, z\, t(z)\, \tilde{t}(z) \mathbf{E}_N \right)^{-1} \| \, < \, 2$ hold for each $z \in \mathbb{C}^+$. For L large enough, we thus have $\| \mathbf{E}_N \left(\mathbf{I}_N - c\, z\, t(z)\, \tilde{t}(z) \mathbf{E}_N \right)^{-1} \| \leq \frac{\kappa}{L^{\min(\gamma,1)}}$ for some nice constant κ, a property which also implies that $\| \mathbf{\Gamma}(z) \| \leq \frac{\kappa}{L^{\min(\gamma,1)}}$ because if $\widetilde{\mathbf{X}}$ is any $N \times N$ matrix, then $\| \Psi(\widetilde{\mathbf{X}}) \| \leq s_{max} \| \widetilde{\mathbf{X}} \|$, where we recall that s_{max} is an upper bound on the spectral densities (cf. Assumption 3). We also notice that for L large enough, $\| \mathbf{I}_{ML} - c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z) \| > 1 - \| \mathbf{\Gamma}_{\mathcal{B}}(z) \| > \frac{1}{2}$ for each $z \in \mathbb{C}^+$, and therefore, that $\| \left(\mathbf{I}_{ML} - c(zt(z)\tilde{t}(z))^2 \mathbf{\Gamma}_{\mathcal{B}}(z) \right)^{-1} \| < 2$ on $\mathbb{C}^+$. This, in turn, implies that

$$\|\Delta_{mp}(z)\| \leq \frac{C(z)}{L^{\min(\gamma,1)}} \tag{5.20}$$

and also $\|\Delta_{\mathcal{B},mp}(z)\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$ for L large enough, as we wanted to show.

We now establish that (5.12) holds. We first prove that for any $ML \times ML$ block matrix matrix $\mathbf{X}$ the series $\sum_{n=0}^{+\infty} \Phi_{\mathcal{B},2}^{(n)}(\mathbf{X})$ is convergent. For this, we use Proposition 1.3. According to Lemma 4.1, $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B}}^H}^{(n)}(\mathbf{Y}) < +\infty$ for each positive matrix $\mathbf{Y}$. In order to establish a similar property for operator $\Phi_{\mathbf{T}_{\mathcal{B},mp}}$, we notice that a simple calculation leads to the identity

$$\frac{\mathrm{Im}\mathbf{T}_{\mathcal{B},mp}(z)}{\mathrm{Im}z} = \mathcal{B}_L^{-1/2} \mathbf{T}_{mp}(z) \mathbf{T}_{mp}^H(z) \mathcal{B}_L^{-1/2} + \Phi_{\mathbf{T}_{\mathcal{B},mp}} \left(\frac{\mathrm{Im}\mathbf{t}_{\mathcal{B}}}{\mathrm{Im}z} \right)$$

if $z \in \mathbb{C}^+$. This implies that

$$\frac{\mathrm{Im}\mathbf{t}_{\mathcal{B}}(z)}{\mathrm{Im}z} = \mathcal{B}_L^{-1/2} \mathbf{T}_{mp}(z) \mathbf{T}_{mp}^H(z) \mathcal{B}_L^{-1/2} + \frac{\mathrm{Im}\Delta_{\mathcal{B},mp}(z)}{\mathrm{Im}z} + \Phi_{\mathbf{T}_{\mathcal{B},mp}} \left(\frac{\mathrm{Im}\mathbf{t}_{\mathcal{B}}}{\mathrm{Im}z} \right).$$

Noting that $\|\Delta_{\mathcal{B},mp}(z)\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$, Lemma B.1 in [16] implies that

$$\left\| \frac{\mathrm{Im}\Delta_{\mathcal{B},mp}(z)}{\mathrm{Im}z} \right\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}.$$

Proposition 1.2 implies that $\mathbf{T}_{mp}(z)\mathbf{T}_{mp}^H(z) \geq \frac{1}{C(z)}\mathbf{I}_{ML}$ for each $z \in \mathbb{C}^+$. Therefore, if we denote by $\mathbf{Y}_1(z)$ the matrix $\mathbf{Y}_1(z) = \mathcal{B}_L^{-1/2} \mathbf{T}_{mp}(z) \mathbf{T}_{mp}^H(z) \mathcal{B}_L^{-1/2} + \frac{\mathrm{Im}\Delta_{\mathcal{B},mp}(z)}{\mathrm{Im}z}$, then, $\mathbf{Y}_1(z) > \frac{1}{C(z)}\mathbf{I}_{ML} > 0$ if $z \in F_N$ where F_N is a subset of $\mathbb{C}^+$ defined by

$$F_N = \left\{ z \in \mathbb{C}^+, \frac{1}{L^{\min(1,\gamma)}} P_1(|z|) P_2 \left(\frac{1}{\mathrm{Im}z} \right) \leq \kappa \right\} \tag{5.21}$$

for some nice constant κ. Using the same arguments as in [27], we obtain that for each $z \in F_N$, the series $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B},mp}}^{(n)}(\mathbf{Y}_1(z))$ is convergent. Proposition 1.3 implies that for each positive matrix $\mathbf{Y}$, $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B},mp}}^{(n)}(\mathbf{Y}) < +\infty$ and that for each matrix $\mathbf{X}$, the series $\sum_{n=0}^{+\infty} \Phi_{\mathcal{B},2}^{(n)}(\mathbf{X})$ is convergent if $z \in F_N$. Therefore, (5.12) holds true for $z \in F_N$, and

$$\left\| \sum_{n=0}^{+\infty} \Phi_{\mathcal{B},2}^{(n)}(\Delta_{\mathcal{B},mp}(z)) \right\| \leq \|\Delta_{\mathcal{B},mp}(z)\| \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B},mp}}^{(n)}(\mathbf{I}_{ML}) \right\|^{1/2} \left\| \sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B}}}^{(n)}(\mathbf{I}_{ML}) \right\|^{1/2}.$$

It is easy to check that $\sum_{n=0}^{+\infty} \Phi_{\mathbf{T}_{\mathcal{B},mp}}^{(n)}(\mathbf{I}_{ML}) < C(z)\mathbf{I}_{ML}$ for $z \in F_N$. Therefore, we obtain that $\|\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$ for each $z \in F_N$. It remains to evaluate $\|\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\|$ if z does not belong to F_N. For this, we remark that $\|\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\| \leq \|\mathbf{t}_{\mathcal{B}}(z)\| + \|\mathbf{T}_{\mathcal{B}}(z)\| \leq C(z)$. As z does not belong to F_N, the inequality $1 \leq \frac{C(z)}{L^{\min(1,\gamma)}}$ holds for a certain $C(z)$, from which we deduce that $\|\mathbf{t}_{\mathcal{B}}(z) - \mathbf{T}_{\mathcal{B}}(z)\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$ as expected. Since the matrix $\mathcal{B}_L^{-1/2}$ verifies $\mathcal{B}_L^{-1/2} > \frac{1}{\sqrt{s_{min}}}\mathbf{I}_{ML}$, we obtain (5.3) for each $z \in \mathbb{C}^+$.

5.2. Bounding the term $\frac{1}{ML}\mathrm{Tr}\left(t_N(z)\mathbf{I}_{ML} - \mathbf{T}_N(z)\right)$

We begin by considering the identity in (5.11), and obtain that

$$t(z)\mathbf{I}_{ML} - \mathbf{T}(z) = \mathcal{B}_L^{1/2}\Phi_{\mathcal{B},2}\left(\mathcal{B}_L^{-1/2}(t(z)\mathbf{I}_{ML} - \mathbf{T}(z))\mathcal{B}_L^{-1/2}\right)\mathcal{B}_L^{1/2} + \Delta_{mp}(z) \quad (5.22)$$

which directly implies that

$$t(z) - \frac{1}{ML}\mathrm{Tr}(\mathbf{T}(z)) = \frac{1}{ML}\mathrm{Tr}\left(\Phi_{\mathcal{B},2}\left(\mathcal{B}_L^{-1/2}(t(z)\mathbf{I}_{ML} - \mathbf{T}(z))\mathcal{B}_L^{-1/2}\right)\mathcal{B}_L\right) +$$
$$+ \frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z)). \quad (5.23)$$

We introduce the operator $\Phi_{\mathcal{B},2}^t$ defined by the property that, for any two $ML \times ML$ matrices $\mathbf{X}$, $\mathbf{Y}$, we have

$$\frac{1}{ML}\mathrm{Tr}\left(\mathbf{X}\Phi_{\mathcal{B},2}(\mathbf{Y})\right) = \frac{1}{ML}\mathrm{Tr}\left(\mathbf{Y}\Phi_{\mathcal{B},2}^t(\mathbf{X})\right). \quad (5.24)$$

This can be seen as a transpose operator of $\Phi_{\mathcal{B},2}$. Using (1.41) it can be expressed in closed form as

$$\Phi_{B,2}^t(\mathbf{X}) = cz^2\Psi\left(\widetilde{\mathbf{T}}^T\overline{\Psi}(\mathbf{T}_{\mathcal{B}}\mathbf{X}\mathbf{T}_{\mathcal{B},mp})\widetilde{\mathbf{T}}_{mp}^T\right).$$

Using (5.24), the expression in (5.23) can be rewritten as

$$t(z) - \frac{1}{ML}\mathrm{Tr}(\mathbf{T}(z)) = \frac{1}{ML}\mathrm{Tr}\left((t(z)\mathbf{I}_{ML} - \mathbf{T}(z))\mathcal{B}_L^{-1/2}\Phi_{\mathcal{B},2}^t(\mathcal{B}_L)\mathcal{B}_L^{-1/2}\right)$$
$$+ \frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z)). \quad (5.25)$$

In order to simplify (5.25), we observe that there exists $C(z) = P_1(|z|)P_2(\frac{1}{\mathrm{Im}z})$ for some nice polynomials P_1 and P_2 such that

$$\|\mathbf{T}_{\mathcal{B}}(z) - t(z)\mathcal{B}_L^{-1}\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$$

$$\|\mathbf{T}_{\mathcal{B},mp}(z) - t(z)\mathcal{B}_L^{-1}\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$$

and

$$\|z\widetilde{\mathbf{T}}(z) - z\tilde{t}(z)\mathbf{I}_N\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$$

$$\|z\widetilde{\mathbf{T}}_{mp}(z) - z\tilde{t}(z)\mathbf{I}_N\| \leq \frac{C(z)}{L^{\min(\gamma,1)}}$$

which follow directly from Theorem 5.1 and (5.20). From this, it is easily checked that for each matrix $\mathbf{X}$, $\Phi_{B,2}^t(\mathbf{X})$ can be written as

$$\Phi_{B,2}^t(\mathbf{X}) = c(zt(z)\tilde{t}(z))^2 \, \Psi\left(\overline{\Psi}(\mathcal{B}_L^{-1}\mathbf{X}\mathcal{B}_L^{-1})\right) + \mathbf{\Upsilon}(\mathbf{X}) \tag{5.26}$$

where $\mathbf{\Upsilon}$ is a linear operator verifying

$$\|\mathbf{\Upsilon}(\mathbf{X})\| \leq \kappa \, C(z)\frac{\|\mathbf{X}\|}{L^{\min(\gamma,1)}} \tag{5.27}$$

for each $z \in \mathbb{C}^+$. By (5.26), and using the fact that $\overline{\Psi}(\mathcal{B}_L^{-1}) = \mathbf{I}_N + \mathbf{E}_N$ and that $\Psi(\mathbf{I}_N) = \mathcal{B}_L$, we obtain

$$\Phi_{B,2}^t(\mathcal{B}) = u(z)\, \Psi\left(\overline{\Psi}(\mathcal{B}_L^{-1})\right) + \mathbf{\Upsilon}(\mathcal{B}_L) = u(z)\,(\mathcal{B}_L + \Psi(\mathbf{E}_N)) + \mathbf{\Upsilon}(\mathcal{B}_L) \tag{5.28}$$

where we have introduced the definition $u(z) = c(zt(z)\tilde{t}(z))^2$. We can express the above equation as

$$\Phi_{B,2}^t(\mathcal{B}_L) = u(z)\mathcal{B}_L + \mathbf{\Upsilon}(\mathcal{B}_L) + u(z)\Psi(\mathbf{E}_N). \tag{5.29}$$

Plugging (5.29) into (5.25), we obtain

$$t(z) - \frac{1}{ML}\mathrm{Tr}(\mathbf{T}(z)) = u(z)\left(t(z) - \frac{1}{ML}\mathrm{Tr}(\mathbf{T}(z))\right) + \frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z)) + \delta_1(z)$$

where $\delta_1(z)$ is the error term defined as

$$\delta_1(z) = \frac{1}{ML}\mathrm{Tr}\left((t(z)\mathbf{I}_{ML} - \mathbf{T}(z))\mathcal{B}_L^{-1/2}(\mathbf{\Upsilon}(\mathcal{B}_L) + u(z)\Psi(\mathbf{E}_N))\mathcal{B}_L^{-1/2}\right).$$

We recall (see e.g. [26], Lemma 1.1) that $u(z)$ verifies $1 - |u(z)| > \frac{1}{C(z)}$ on $\mathbb{C}^+$, where $C(z) = P_1(|z|)P_2(\frac{1}{\mathrm{Im}z})$ for some nice polynomials P_1 and P_2. Therefore, we have the inequality

$$\left|t(z) - \frac{1}{ML}\mathrm{Tr}(\mathbf{T}(z))\right| \leq C(z)\left|\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z))\right| + |\delta_1(z)|. \tag{5.30}$$

The bound in (5.3) together with Corollary 5.1, and the properties of operator $\mathbf{\Upsilon}$ imply that $|\delta_1(z)| \leq \frac{C(z)}{L^{2\min(\gamma,1)}}$. As a consequence, in order to complete the proof of (5.2), we only need to establish the following fundamental Lemma.

Lemma 5.2. *Under the above assumptions and for any $z \in \mathbb{C}^+$, we have*

$$\left|\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z))\right| \leq \frac{C(z)}{L^{2\min(\gamma,1)}} \tag{5.31}$$

where $C(z) = P_1(|z|)P_2(\frac{1}{\mathrm{Im}z})$ for two nice polynomials P_1 and P_2.

To justify (5.31), we consider Eq. (5.15) and express $(\mathbf{I}_{ML} - u(z)\mathbf{\Gamma}_{\mathcal{B}}(z))^{-1}$ as

$$(\mathbf{I}_{ML} - u(z)\mathbf{\Gamma}_{\mathcal{B}}(z))^{-1} = \mathbf{I}_{ML} + u(z)\mathbf{\Gamma}_{\mathcal{B}}(z)(\mathbf{I}_{ML} - u(z)\mathbf{\Gamma}_{\mathcal{B}}(z))^{-1}.$$

Hence, $\Delta_{mp}(z)$ can thus be rewritten as

$$\Delta_{mp}(z) = -t(z)u(z)\mathbf{\Gamma}_{\mathcal{B}}(z) + \Delta_{mp,1}(z)$$

where $\Delta_{mp,1}(z)$ is now defined by

$$\Delta_{mp,1}(z) = -t(z)u^2(z)\left(\mathbf{\Gamma}_{\mathcal{B}}(z)\right)^2\left(\mathbf{I}_{ML} - u(z)\mathbf{\Gamma}_{\mathcal{B}}(z)\right)^{-1}.$$

Using the fact that $\|\mathbf{\Gamma}_{\mathcal{B}}(z)\| \leq \frac{\kappa}{L^{\min(\gamma,1)}}$ on $\mathbb{C}^+$, we obtain immediately that $\|\Delta_{mp,1}(z)\| \leq \frac{C(z)}{L^{2\min(\gamma,1)}}$ for each $z \in \mathbb{C}^+$. We finally remark that $\mathbf{\Gamma}_{\mathcal{B}}(z)$ can be written as

$$\mathbf{\Gamma}_{\mathcal{B}}(z) = \mathcal{B}_L^{-1/2}\Psi(\mathbf{E}_N)\mathcal{B}_L^{-1/2} + czt(z)\tilde{t}(z)\Psi\left(\mathbf{E}_N^2(\mathbf{I}_N - czt(z)\tilde{t}(z)\mathbf{E}_N)^{-1}\right). \quad (5.32)$$

The spectral norm of the right hand side of (5.32) is clearly upper bounded by a term such as $\frac{\kappa}{L^{2\min(\gamma,1)}}$ for each $z \in \mathbb{C}^+$. Therefore, $\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z))$ can be written as

$$\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z)) = -t(z)u(z)\frac{1}{ML}\mathrm{Tr}(\Psi(\mathbf{E}_N)\mathcal{B}_L^{-1}) + \delta_2(z) \quad (5.33)$$

where $\delta_2(z)$ verifies $|\delta_2(z)| \leq \frac{C(z)}{L^{2\min(\gamma,1)}}$ for each $z \in \mathbb{C}^+$. Using (1.41), we notice that $\frac{1}{ML}\mathrm{Tr}(\Psi(\mathbf{E}_N)\mathcal{B}_L^{-1})$ is equal to

$$\frac{1}{ML}\mathrm{Tr}(\Psi(\mathbf{E}_N)\mathcal{B}_L^{-1}) = \frac{1}{N}\mathrm{Tr}(\mathbf{E}_N\overline{\Psi}(\mathcal{B}_L^{-1})) = \frac{1}{N}\mathrm{Tr}\left(\mathbf{E}_N(\mathbf{I}_N + \mathbf{E}_N)\right)$$

so that $\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z))$ can in turn be rewritten as

$$\frac{1}{ML}\mathrm{Tr}(\Delta_{mp}(z)) = -t(z)u(z)\frac{1}{N}\mathrm{Tr}(\mathbf{E}_N) + \delta_3(z) \quad (5.34)$$

where $|\delta_3(z)| \leq \frac{C(z)}{L^{2\min(\gamma,1)}}$ for each $z \in \mathbb{C}^+$. We complete the proof of (5.31) by simply noting that

$$\mathrm{Tr}(\mathbf{E}_N) = 0. \quad (5.35)$$

This can be shown by noting that we can express $\mathcal{R}_{m,L}$ as

$$\mathcal{R}_{m,L} = \int_0^1 \mathcal{S}_m(\nu)\mathbf{d}_L(\nu)\mathbf{d}_L^H(\nu)d\nu.$$

As a consequence of this,

$$\int_0^1 \mathcal{S}_m(\nu)\mathbf{a}_L^H(\nu)\mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu)d\nu = \frac{1}{L}\mathrm{Tr}\left[\mathcal{R}_{m,L}^{-1}\mathcal{R}_{m,L}\right] = 1$$

which directly implies that

$$\int_0^1 \epsilon_{m,L}(\nu)d\nu = \int_0^1 \mathcal{S}_m(\nu)\mathbf{a}_L^H(\nu)\mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu)d\nu - 1 = 0.$$

62

However, from the definition of $\mathbf{E}_N$ we see that

$$\mathrm{Tr}\left(\mathbf{E}_N\right) = \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \epsilon_{m,L}(\nu)d\nu = 0$$

which completes the proof.

6. Numerical Validation

The aim of this section is to validate the asymptotic study carried out above via simulations. To that effect, we consider a simple example in which the M independent time series are all autoregressive processes of order one with parameter ρ and unit power. By this, we mean that we generate each time series independently by the recursion $y_{m,n+1} = \rho y_{m,n} + e_{m,n}$ where $e_{m,n} \sim \mathcal{N}_{\mathbb{C}}(0, 1 - |\rho|^2)$.

Let us first compare the empirical eigenvalue distribution of the sample cross correlation matrix $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ with the measure μ_N and the Marchenko-Pastur distribution with parameter c_N. Figure 1 represents the histogram of the eigenvalues of $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ together with the Marchenko-Pastur distribution μ_{mp,c_N} for different values of M, N, L. In general terms, the Marchenko-Pastur approximation provides a relatively good approximation of the actual eigenvalue density. In general terms, we observe that the Marchenko-Pastur law is a very good approximation of the actual empirical eigenvalue distribution, even for relatively low values of M, L.

Next, consider a correlation detection test statistic consisting of the sum of the squared value of all the off-diagonal entries of $\widehat{\mathcal{R}}_{\mathrm{corr},L}$. As mentioned in the introduction, this is reasonable test since under H_0 the true cross-correlation matrix $\mathcal{R}_{\mathrm{corr},L}$ is equal to an identity. This corresponds to a linear spectral statistic of $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ built with the function $\phi(\lambda) = (\lambda - 1)^2$.

Remark 6.1. We observe that this function is not compactly supported so that in principle the asymptotic rates predicted in items (ii) and (iii) of Theorem 1.1 are not guaranteed to hold. However, we claim here that these two items still hold for the choice $\phi(\lambda) = (\lambda - 1)^2$. Indeed, consider first item (ii) in the statement of this theorem. The only point in the proof of this item where the hypothesis of compactly supported $\phi(\lambda)$ is used is in order to establish (4.16). However, for this choice of $\phi(\lambda)$ it is possible to compute $\frac{1}{ML}\mathbb{E}\mathrm{Tr}\left(\phi(\overline{\mathcal{R}}_{\mathrm{corr},L})\right)$ in closed form as well as $\int \phi(\lambda)d\mu_N(\lambda)$, and to establish that

$$\frac{1}{ML}\mathbb{E}\mathrm{Tr}\left(\phi(\overline{\mathcal{R}}_{\mathrm{corr},L})\right) = \int \phi(\lambda)d\mu_N(\lambda) \tag{6.1}$$

so that (4.16) is, in fact, trivial. Indeed, the quantity on the left hand side can be computed by using conventional formulas on the expectation of four Gaussian random vectors, whereas the quantity on the right hand side can be evaluated by
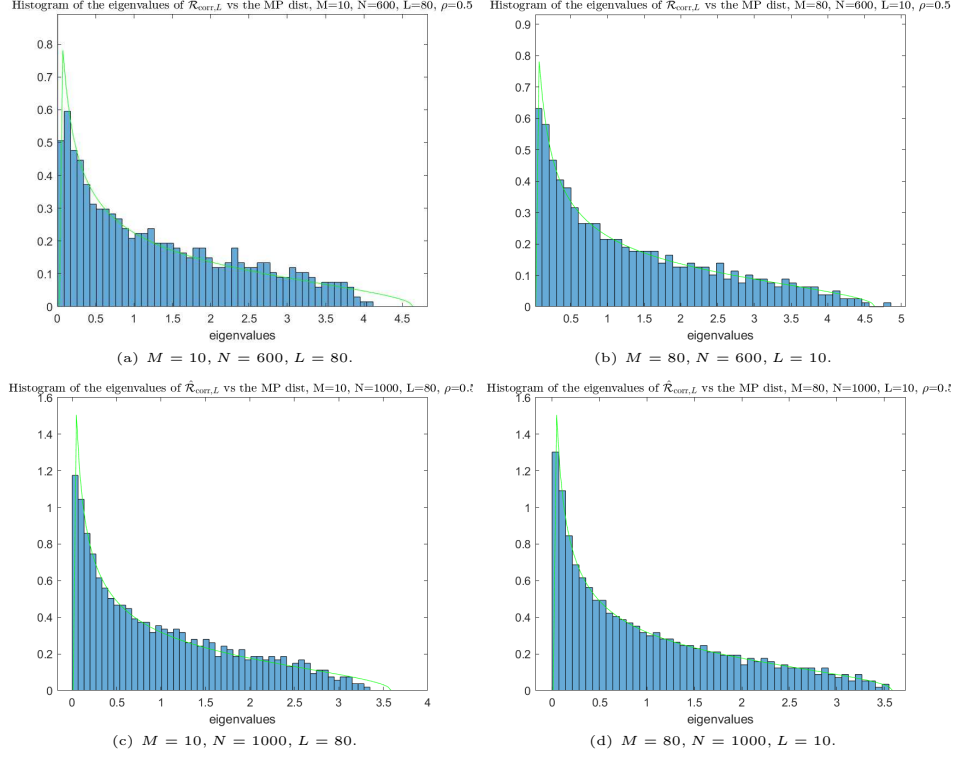
Fig. 1. Histogram of the eigenvalues of $\widehat{\mathcal{R}}_{\mathrm{corr},L}$ and Marchenko-Pastur law for different values of M, N, L with $\rho = 0.5$. Upper plots correspond to a situation where $c_N > 1$ whereas lower plots deploy the case $c_N < 1$.

relating the second order moment of the measure $\mu_N(\lambda)$ with its Stieltjes transform. In both cases, we can establish that both quantities are equal to

$$\int \phi(\lambda) d\mu_N(\lambda) = c_N + c_N \frac{1}{ML} \mathrm{Tr} \left(\mathcal{B}_L^{-1} \mathbf{\Psi}(\mathbf{E}_N) \right) \tag{6.2}$$

where we recall that $\mathbf{E}_N$ is defined in (5.13). Regarding item (iii) in Theorem 1.1, we simply need to observe that $\int \phi(\lambda) d\mu_{mp,c_N}(\lambda) = c_N$, so that

$$\int \phi(\lambda) d\mu_N(\lambda) - \int \phi(\lambda) d\mu_{mp,c_N}(\lambda) = c_N \frac{1}{ML} \mathrm{Tr} \left(\mathcal{B}_L^{-1} \mathbf{\Psi}(\mathbf{E}_N) \right). \tag{6.3}$$

Consequently, a direct application of Corollary 5.1 thus leads to the conclusion that (1.25) also holds for this particular choice of $\phi(\lambda)$. We may therefore consider this statistic to validate the results of the paper.

In order to assess the error between $\hat{\phi}_N$ and the corresponding integral of $\phi(\lambda)$ with respect to the Marchenko-Pastur distribution, we considered here a set of 10^4 realizations of the multivariate autoregressive process described above. In each

experiment, we fixed the three parameters c_*, N and β and considered a set of $M = [(c_* N)^{1-\beta}]$ independent time series, where $[x]$ here denotes the integer that is closest to x. The number of time lags was therefore fixed to $L = [(c_* N)^{\beta}]$. Figure 2 represents the error between $\hat{\phi}_N$ and its corresponding asymptotic limit as a function of β for different values of N. The errors are represented as the square root of the empirical mean of the corresponding normalized difference, averaged over the 10^4 realizations. The plots on the left hand side represent the total error $\hat{\phi}_N - \int \phi(\lambda) d\mu_{mp,N}$ whereas plot on the right hand side represent the two main constituent errors, namely: "Error 1" (solid lines) represents the square root of the empirical mean of the square of $\hat{\phi}_N - \int \phi(\lambda) d\mu_N$, and "Error 2" (dotted lines) represents $\int \phi(\lambda) d\mu_N - \int \phi(\lambda) d\mu_{mp,N}$ as given in (6.3).

These numerical results tend to confirm the fact that the error between the considered statistic and its asymptotic deterministic approximation tends to be dominated by two different phenomena depending on whether $M \ll L$ (large β) or $M \gg L$ (small β). In the fist case, the main contribution to the error corresponds to the term $\hat{\phi}_N - \int \phi(\lambda) d\mu_N$ (Error 1). We recall that, since the correlation sequence considered here decays exponentially to zero, this error term is dominated by $N^{-(1-\beta)}$, which in particular increases with β. Conversely, when $M \gg L$ (small β), the error is dominated by the difference between the two measures μ_N and $\mu_{mp,N}$. We have seen that this error term is dominated by a term of order $N^{-2\beta}$, which in particular decreases with β. Observe also that the optimum choice of β appears to be close to $1/3$, which corresponds to the case where the two error rates coincide.

Acknowledgments

References

[1] D. Alpay, E. Tsekanovskii, "Subclasses of Herglotz-Nevanlinna matrix-valued functions and linear systems", Proc. Int. Conf. on Dynamical Systems and Differential Equations, May 18-21, 2000, Atlanta, USA.

[2] G.W. Anderson, A. Guionnet, O. Zeitouni, "An Introduction to Random Matrices", *Cambridge Studies in Advanced Mathematics*, vol. 118, Cambridge University Press, 2010.

[3] G.W. Anderson, "Convergence of the largest singular value of a polynomial in independent Wigner matrices", Annals of Proba., 2013, vol. 41, No. 3B, 2103-2181.

[4] M. Bhattachargee, A. Bose, "Large sample behaviour of high-dimensional autocovariance matrices", Ann. of Stat., vol. 44, no. 2, pp. 598-628, 2016a.

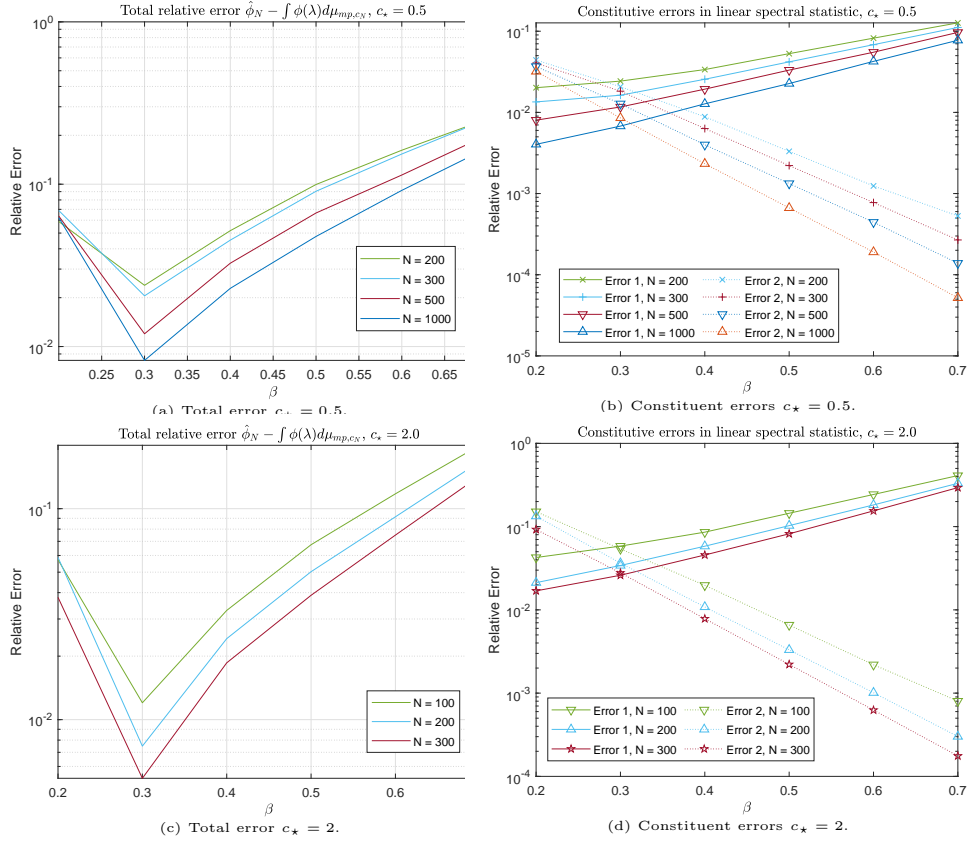[5] D.R. Brillinger, "Time Series, Data Analysis and Theory", Classics in Applied Mathematics 36, SIAM, 2001.

Fig. 2. Evolution of the error of $\hat{\phi}_N$ with respect to the Marchenko-Pastur limit as a function of β. On the left hand side we represent the square root of the empirical mean of the square of $\hat{\phi}_N - \int \phi(\lambda) d\mu_{mp,N}$ over 10^4 realizations of the statistic. On the right hand side, we represent the two main constituent errors. Error 1 (solid lines) represents the square root of the empirical mean of the square of $\hat{\phi}_N - \int \phi(\lambda) d\mu_N$. Error 2 (dotted lines) represents $\int \phi(\lambda) d\mu_N - \int \phi(\lambda) d\mu_{mp,N}$. Upper plots correspond to a situation where $c_\star = 0.5$ whereas lower plots deploy the case $c_\star = 2$.

[6] M. Capitaine, C. Donati-Martin, "Strong asymptotic freeness of Wigner and Wishart matrices", *Indiana Univ. Math. Journal*, vol. 25, pp. 295-309, 2007

[7] H. Dette, N. Dörnemann, "Likelihood ratio tests for many groups in high dimensions", J. Mult. Anal., 178 (2020) 104605.

[8] P. Duchesne, R. Roy, "Robust tests for independence of two time series", Statistica Sinica, pp. 827–852, 2003.

[9] M. Eichler, "A frequency-domain based test for non-correlation between stationary time series", Metrika, vol. 65, no. 2, pp. 133–157, 2007.

[10] M. Eichler, "Testing nonparametric and semiparametric hypotheses in vector stationary processes", J. of Multi. Anal., vol. 99, no. 5, pp. 968–1009, 2008.

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

66

van Eeden.

[12] L. Erdös and A. Knowled and H.T. Yao, "Averaging Fluctuations in Resolvents of Random Band Matrices", Ann. Henri Poincaré, vol. 14, pp. 1837–1926, 2013.

[13] Ya. L. Geronimus, "Polynomials Orthogonal on a Circle and Interval", Pergamon Press, 1960.

[14] U. Haagerup, S. Thorbjornsen, "A new application of random matrices: $\text{Ext}(C^*_{red}(F_2))$ is not a group", *Annals of Mathematics*, vol. 162, no. 2, 2005.

[15] W. Hachem, P. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices", Annals of Applied Probability, 17(3):875–930, 2007.

[16] W. Hachem, P. Loubaton, J. Najim, P. Vallet, "On bilinear forms based on the resolvent of large random matrices", Annales de l'Institut Henri Poincaré, Prob. Stats., vol. 45, no. 1, 2013, pp. 36-63.

[17] L.D. Haugh, "Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach", J. of the Am. Stat. Asso., vol. 71, no. 354, pp. 378–385, 1976

[18] Y. Hong, "Testing for independence between two covariance stationary time series", Biometrika, vol. 83, no. 3, 1996

[19] T. Jiang, "The limiting distributions of eigenvalues of sample correlation matrices", Sankhya: The Indian Journal of Statistics, vol. 66, pp. 35-48, 2004.

[20] T. Jiang, F. Yang, "Central limit theorems for classical likelihood ratio tests for high-dimensional distributions", Ann. of Stats., vol. 41, no. 4, pp. 2029-2074, 2013.

[21] B. Jin, C. Wang, Z.D. Bai, K.K. Nair, M. Harding, "Limited spectral distribution of a symmetrised auto-cross covariance matrices", Ann. of Proba., vol. 24, no. 3, pp. 1199-1225, 2014.

[22] T. Kato, "Perturbation Theory for Linear Operators", Classics in Mathematics, Springer, Reprint of the second edition, 1995.

[23] E. Kim, S. Lee, "A test for independence of two stationary infinite order autoregressive processes", Ann. Inst. Stat. Math., vol. 57, no. 1, pp. 105–127, 2005.

[24] Z. Li, G. Pan, J. Yao, "On singular value distribution of large-dimensional autocovariance matrices", J. Multivariate Analysis, vol. 137, pp. 119-140, May 2015.

[25] H. Liu, A. Aue, D. Paul, "On the Marcenko-Pastur law for linear time series", Ann. of Stat., vol. 43, no. 2, pp. 675-712, 2015.

[26] P. Loubaton, "On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices", J. of Theoretical Probability, vol. 29, no. 4, pp. 1339-1443, December 2016

[27] P. Loubaton, X. Mestre, "Spectral convergence of large block-Hankel Gaussian random matrices", Colombo F., Sabadini I., Struppa D., Vajiac M. (eds) Advances in Complex Analysis and Operator Theory. Trends in Mathematics. Birkhaüser, Cham, 2017, available on Arxiv (arXiv:1704.06651).

[28] P. Loubaton, D. Tieplova, "On the behaviour of large autocovariance matrices between the past and the future", to appear in Random Matrix: Theory and Applications, https://doi.org/10.1142/S2010326321500210.

[29] P. Loubaton, A. Rosuel, "Large random matrix approach for testing independence of a large number of Gaussian time series", preprint arXiv:2007.08806.

[30] X. Mestre, P. Vallet, "Correlation tests and linear spectral statistics of the sample correlation matrix", IEEE Trans. Info. Theory, vol. 63, no. 7, 2017.

[31] J. Najim, J. Yao. "Gaussian fluctuations for linear spectral statistics of large random covariance matrices" . Annals of Applied Probability, vol. 26(3), 2016.

[32] M. A Nowak, W. Tarnowski, "Spectra of large time-lagged correlation matrices from random matrix theory", Journal of Statistical Mechanics: Theory and Experiment,

vol. 2017, June 2017.

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

[34] M. Rudelson, R. Vershynin, "Hanson-Wright inequality and sub-gaussian concentration", Electronic Communications in Probability, vol. 18, 2013.

[35] W. Rudin, "Real and Complex Analysis", Mc Graw Hill, 1966.

[36] Yu.A. Rozanov, "Stationary Random Processes", Holden Day, 1967.

[37] B. Simon, "Orthogonal Polynomials on the Unit Circle, Part 1: Classical Theory", American Mathematical Society Colloquium Publications, Vol. 54, Part 1, 2005.

[38] T. Tao, "Topics in Random Matrix Theory", Graduate Studies in Mathematics 132, 2011.

[39] M. Taniguchi, M.L. Puri, M. Kondo, "Nonparametric approach for non-Gaussian vector stationary processes", J. of Mult. Anal., vol. 56, no. 2, pp. 259–283, 1996.

[40] G. Wahba, "Some tests of independence for stationary multivariate time series", J. Royal Stat. Soc.: Series B (Methodological), vol. 33, no. 1, 1971

[41] H. Xiao, W.B. Wu, "Covariance matrix estimation for stationary time series", Ann. of Statistics, vol. 40, pp. 466-493, 2012.

[42] A. Zygmund, "Trigonometric Series", Third Edition, Cambridge University Press, 2002.

APPENDICES

A. Proof of Lemma 2.2

A classical calculation (see e.g. Theorem 4.3.2 in [5] in the non Gaussian case) leads to

$$\mathbb{E}|\xi_{L,y_m}(\nu)|^2 = \sum_{-(L-2)}^{L-2} (1 - |l|/L) r_m(l) e^{-2i\pi l\nu}.$$

Taking into account that $\mathcal{S}_m(\nu) = \sum_l r_m(l) e^{-2i\pi l\nu}$, we obtain immediately that

$$\mathbb{E}|\xi_{L,y_m}(\nu)|^2 = \mathcal{S}_m(\nu) + \epsilon_{m,L}(\nu)$$

where $\epsilon_{m,L}(\nu)$ is defined by

$$\epsilon_{m,L}(\nu) = - \sum_{|l|\geq L-1} r_m(l) e^{-2i\pi l\nu} - \frac{1}{L} \sum_{-(L-2)}^{L-2} |l| r_m(l) e^{-2i\pi l\nu}.$$

It is clear that

$$|\epsilon_{m,L}(\nu)| \leq \sum_{|l|\geq L-1} |r_m(l)| + \frac{1}{L} \sum_{-(L-2)}^{L-2} |l||r_m(l)|.$$

Using the bound in (1.14) we directly obtain an upper bound of the first term, namely

$$\sum_{|l|\geq L-1} |r_m(l)| \leq \frac{\kappa}{(L-1)^{\gamma_0}}.$$

If $\gamma_0 \geq 1$, $\sum_{-(L-2)}^{L-2} |l||r_m(l)| \leq \|r_m\|_{\omega_0}$ and it holds that $\frac{1}{L}\sum_{-(L-2)}^{L-2} |l||r_m(l)| \leq \frac{\kappa}{L}$. Therefore, if $\gamma_0 \geq 1$, we obtain that

$$|\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{L}.$$

If $\gamma_0 < 1$, we equivalently have

$$\sum_{-(L-2)}^{L-2} |l||r_m(l)| \leq L^{1-\gamma_0}\|r_m\|_{\omega_0}.$$

Therefore, the inequality

$$\frac{1}{L}\sum_{-(L-2)}^{L-2} |l||r_m(l)| \leq \frac{\kappa}{(L-1)^{\gamma_0}}$$

holds, as well as

$$|\epsilon_{m,L}(\nu)| \leq \frac{\kappa}{L^{\gamma_0}}.$$

This completes the proof of Lemma 2.2.

B. Proof of Lemma 2.5

In order to establish (2.42), we first recall that $\|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| \prec \max(M^{-1/2}, L^{-\gamma_0})$. We consider some $\delta > 0$ for which $N^\delta \max(M^{-1/2}, L^{-\gamma_0}) \to 0$ and introduce the event $\mathcal{E}_N$ defined by

$$\mathcal{E}_N = \left\{ \max_{m=1,\dots,M} \|\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\| < N^\delta \max(M^{-1/2}, L^{-\gamma_0}) \right\} \tag{B.1}$$

Then, the event $\mathcal{E}_N$ holds with exponentially high probability. In order to establish (2.43), we have to evaluate $\mathbb{P}(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0}))$ for each $\epsilon > 0$. For this, we express $\mathbb{P}(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0}))$ as

$$\mathbb{P}\left(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0}), \mathcal{E}_N\right) + \mathbb{P}\left(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0}), \mathcal{E}_N^c\right).$$

Therefore, it holds that

$$\mathbb{P}(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0})) \leq$$
$$\leq \mathbb{P}(\mathcal{E}_N^c) + \mathbb{P}\left(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1}, L^{-2\gamma_0}), \mathcal{E}_N\right).$$

In order to establish (2.43), we thus just need to prove that there exists a $\gamma > 0$ such that $\mathbb{P}\left(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max(M^{-1/2}, L^{-\gamma_0}), \mathcal{E}_N\right) \leq \exp(-N^\gamma)$ for each N large enough. For this, we remark that for each N large enough, on $\mathcal{E}_N$, all the eigenvalues of matrices $\widehat{\mathcal{R}}_{m,L}$ are enclosed by the contour $\mathcal{C}$. Therefore, on $\mathcal{E}_N$, the equality

$$\widehat{\mathcal{R}}_{m,L}^{-1/2} = \frac{1}{2i\pi} \int_{\mathcal{C}_-} \frac{1}{\sqrt{\lambda}} \left(\widehat{\mathcal{R}}_{m,L} - \lambda \mathbf{I}_L\right)^{-1} d\lambda \tag{B.2}$$

holds. We note here that $\left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1}$ can be written as

$$\left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} = (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} +$$
$$- \left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} \left(\widehat{\mathcal{R}}_{m,L} - \mathcal{R}_{m,L}\right) (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1}$$

so that, by iterating this formula, we obtain

$$\left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} = (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} - (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} +$$
$$+ \left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1}.$$

We deduce from this expression together with (2.34) and (B.2) that on $\mathcal{E}_N$ we can write

$$\widehat{\mathcal{R}}_{m,L}^{-1/2} - \mathcal{R}_{m,L}^{-1/2} = -\frac{1}{2i\pi} \int_{\mathcal{C}_-} \frac{1}{\sqrt{\lambda}} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} d\lambda +$$
$$+ \frac{1}{2i\pi} \int_{\mathcal{C}_-} \frac{1}{\sqrt{\lambda}} \left(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L\right)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} \mathbf{\Delta}_{m,L} (\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1} d\lambda.$$
$$(\text{B.3})$$

Now, it is clear that on the contour $\mathcal{C}$, $|\frac{1}{\sqrt{\lambda}}|$ and the spectral norm of $(\mathcal{R}_{m,L} - \lambda\mathbf{I}_L)^{-1}$ are upper bounded by a nice constant. This property also holds for $(\widehat{\mathcal{R}}_{m,L} - \lambda\mathbf{I}_L)^{-1}$ on the event $\mathcal{E}_N$. Therefore, on $\mathcal{E}_N$, the spectral norm of the second term on the right hand side of (B.3) is upper bounded by $\kappa\|\mathbf{\Delta}_{m,L}\|^2$, which is stochastically dominated by $\max(M^{-1}, L^{-2\gamma_0})$. This, in turn, establishes that there exists a $\gamma > 0$ such that $\mathbb{P}\left(\|\mathbf{\Upsilon}_{m,L}\| > N^\epsilon \max\left(M^{-1}, L^{-2\gamma_0}\right), \mathcal{E}_N\right) \leq \exp(-N^\gamma)$ for each N large enough. This completes the proof of Lemma 2.5.

C. Proof of Lemma 3.1

We first express matrix $\mathbf{W}_N^m$ in terms of vector $\mathbf{x}_m$. For this, we observe that for each $l = 1, \ldots, L$, the N–dimensional vector $(\mathbf{y}_{m,l}, \ldots, \mathbf{y}_{m,N+l-1})$ can be written as

$$(\mathbf{y}_{m,l}, \ldots, \mathbf{y}_{m,N+l-1}) = \mathbf{y}_m \mathbf{J}_{N+L-1}^{-(l-1)} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix}$$
$$= \mathbf{x}_m \mathcal{R}_{m,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(l-1)} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix}$$

Therefore, matrix $\mathbf{W}_N^m$ can be written as

$$\mathbf{W}_N^m = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{x}_m \mathcal{R}_{m,N+L-1}^{1/2} \\ \vdots \\ \mathbf{x}_m \mathcal{R}_{m,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(L-1)} \end{pmatrix} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix} \qquad (\text{C.1})$$

70

We recall that $\mathbf{W}_N$ is the matrix $\mathbf{W}_N = \left((\mathbf{W}_N^1)^T, \ldots, (\mathbf{W}_N^L)^T \right)^T$, and that $\widehat{\mathcal{R}}_L = \mathbf{W}_N \mathbf{W}_N^H$. Using this notation, we can write

$$
\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} = -\mathbf{Q}_N(z) \, \mathcal{B}_L^{-1/2} \, \frac{\partial \widehat{\mathcal{R}}_L}{\partial \mathbf{x}_{m_0,i}} \, \mathcal{B}_L^{-1/2} \, \mathbf{Q}_N(z)
$$

$$
= -\frac{1}{\sqrt{N}} \mathbf{Q}_N(z) \, \mathcal{B}_L^{-1/2} \, \mathbf{E}_{m_0} \begin{pmatrix} \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \\ \vdots \\ \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(L-1)} \end{pmatrix} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix} \mathbf{W}_N^H \mathcal{B}_L^{-1/2} \, \mathbf{Q}_N(z)
$$

where we recall that $\mathbf{E}_{m_0}$ is an $ML \times L$ selection matrix with entries $(\mathbf{E}_{m_0})_{i,j} = \delta_{i=(m_0-1)M+j}$ and where $\mathbf{e}_i$ denotes the ith column of $\mathbf{I}_{N+L-1}$. We introduce the matrix $\mathbf{H}_N(z)$ defined by

$$
\mathbf{H}_N(z) = \frac{1}{\sqrt{N}} \begin{pmatrix} \mathbf{I}_N \\ 0 \end{pmatrix} \mathbf{W}_N^H \mathcal{B}_L^{-1/2} \mathbf{Q}_N(z) \mathbf{A}_N \mathbf{Q}_N(z) \mathcal{B}_L^{-1/2}.
$$

It is easily seen that

$$
\frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} \mathbf{A}_N \right) = -\frac{1}{ML} \mathrm{Tr} \begin{pmatrix} \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \\ \vdots \\ \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(L-1)} \end{pmatrix} \mathbf{H}_N(z) \, \mathbf{E}_{m_0}.
$$

If we denote by $\mathbf{f}_l^{m_0}$ the l-th column of $\mathbf{E}_{m_0}$, we can re-write the above expression as

$$
\left| \frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} \mathbf{A}_N \right) \right|^2 = \frac{1}{M^2} \left| \frac{1}{L} \sum_{l=1}^{L} \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(l-1)} \mathbf{H}_N(z) \, \mathbf{f}_l^{m_0} \right|^2.
$$

Consequently, a direct application of Jensen's inequality leads to

$$
\left| \frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} \mathbf{A}_N \right) \right|^2 \leq \frac{1}{M^2} \frac{1}{L} \sum_{l=1}^{L} (\mathbf{f}_l^{m_0})^H \, \mathbf{H}_N^H(z) \mathbf{J}_{N+L-1}^{(l-1)} \mathcal{R}_{m_0,N+L-1}^{1/2} \mathbf{e}_i \times
$$

$$
\times \, \mathbf{e}_i^H \mathcal{R}_{m_0,N+L-1}^{1/2} \mathbf{J}_{N+L-1}^{-(l-1)} \, \mathbf{H}_N(z) \, \mathbf{f}_l^{m_0}.
$$

Hence, using $\sum_i \mathbf{e}_i \mathbf{e}_i^H = \mathbf{I}_{N+L-1}$ and $\mathbf{J}_{N+L-1}^{(l-1)} \mathcal{R}_{m_0,N+L-1} \mathbf{J}_{N+L-1}^{-(l-1)} \leq \kappa \, \mathbf{I}_{N+L-1}$, we obtain

$$
\sum_{m_0,i} \left| \frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} \mathbf{A}_N \right) \right|^2 \leq \kappa \, \frac{1}{M} \frac{1}{ML} \mathrm{Tr} \left(\mathbf{H}_N^H(z) \mathbf{H}_N(z) \right)
$$

so that, inserting the expression of $\mathbf{H}_N(z)$ above,

$$
\sum_{m_0,i} \left| \frac{1}{ML} \mathrm{Tr} \left(\frac{\partial \mathbf{Q}_N(z)}{\partial \mathbf{x}_{m_0,i}} \mathbf{A}_N \right) \right|^2 \leq
$$

$$
\leq \frac{\kappa}{MN} \frac{1}{ML} \mathrm{Tr} \left(\mathcal{B}_L^{-1/2} \mathbf{Q}_N^H(z) \mathbf{A}_N^H \mathbf{Q}_N^H(z) \mathcal{B}_L^{-1/2} \mathbf{W}_N \mathbf{W}_N^H \mathcal{B}_L^{-1/2} \mathbf{Q}_N(z) \mathbf{A}_N \mathbf{Q}_N(z) \mathcal{B}_L^{-1/2} \right).
$$

Finally, using the resolvent identity $\mathcal{B}_L^{-1/2}\mathbf{W}_N\mathbf{W}_N^H\mathcal{B}_L^{-1/2}\mathbf{Q}_N(z) = \mathbf{I}_{ML} + z\,\mathbf{Q}_N(z)$, we obtain

$$\|\mathbf{Q}_N^H(z)\mathcal{B}_L^{-1/2}\mathbf{W}_N\mathbf{W}_N^H\mathcal{B}_L^{-1/2}\mathbf{Q}_N(z)\| \leq \frac{1}{\mathrm{Im}z}\left(1 + \frac{|z|}{\mathrm{Im}z}\right) \leq \frac{1+|z|}{\mathrm{Im}z}\left(1 + \frac{1}{\mathrm{Im}z}\right)$$

so that (3.21) follows directly from

$$\sum_{m_0,i}\left|\frac{1}{ML}\mathrm{Tr}\left(\frac{\partial\mathbf{Q}_N(z)}{\partial\mathbf{x}_{m_0,i}}\,\mathbf{A}_N\right)\right|^2 \leq \frac{\kappa}{MN}\frac{1+|z|}{\mathrm{Im}z}\left(1 + \frac{1}{\mathrm{Im}z}\right)\frac{1}{ML}\mathrm{Tr}\left(\mathbf{A}_N\mathbf{Q}_N(z)\mathcal{B}_L^{-1}\mathbf{Q}_N^H(z)\mathbf{A}_N^H\right)$$

$$\leq \frac{\kappa}{MN}(1+|z|)\frac{1}{(\mathrm{Im}z)^3}\left(1 + \frac{1}{\mathrm{Im}z}\right)\frac{1}{ML}\mathrm{Tr}\left(\mathbf{A}_N\mathbf{A}_N^H\right).$$

D. Proof of Lemmas 5.1

The proof of Lemma 5.1 follows from the observation that the term $\mathbf{a}_L^H(\nu)\mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu)$ can be expressed in terms of the Szegö orthogonal polynomials associated to the scalar product

$$\langle z^k, z^l\rangle = \int_0^1 \mathcal{S}_m(\nu)e^{2i\pi(k-l)\nu}\,d\nu. \tag{D.1}$$

For each integer l, we introduce the monic orthogonal polynomial $\Phi_l(z)$ defined by

$$\Phi_l^{(m)}(z) = z^l - z^l|\mathrm{sp}(1, z, \ldots, z^{l-1}) \tag{D.2}$$

where the symbol $|A$ stands for the orthogonal projection over the space A in the sense of the scalar product (D.1). We denote by $\sigma_l^{2,m}$ the norm square of $\Phi_l^{(m)}$, and define for each l the normalized orthogonal polynomial $\phi_l^{(m)}(z)$ by

$$\phi_l^{(m)}(z) = \frac{\Phi_l^{(m)}(z)}{\sigma_l^m}. \tag{D.3}$$

It is well known that the sequence $(\sigma_l^{2,m})_{l\geq 0}$ is decreasing, that $\sigma_0^{2,m} = r_m(0)$, and that $\lim_{l\to+\infty}\sigma_l^{2,m} = \sigma^{2,m}$ coincides with $\exp\int_0^1\log\mathcal{S}_m(\nu)d\nu$. It is clear that the normalized orthogonal polynomials satisfy

$$\langle\phi_l^{(m)}, \phi_{l'}^{(m)}\rangle = \int_0^1 \phi_l^{(m)}(e^{2i\pi\nu})\left(\phi_{l'}^{(m)}(e^{2i\pi\nu})\right)^*\mathcal{S}_m(\nu)d\nu = \delta_{l-l'}.$$

In the following, we also denote by $\Phi_l^{(m)*}(z)$ and $\phi_l^{(m)*}(z)$ the degree l polynomials defined by

$$\Phi_l^{(m)*}(z) = z^l\left(\Phi_l^{(m)}(z^{-*})\right)^*, \ \phi_l^{(m)*}(z) = z^l\left(\phi_l^{(m)}(z^{-*})\right)^*.$$

Noting that Φ_l is for each l a monic polynomial, it is clear that $\Phi_l^{(m)*}(z)$ can be written as

$$\Phi_l^{(m)*}(z) = 1 + \sum_{k=1}^l a_{k,l}^{(m)}z^k \tag{D.4}$$

for some coefficients $(a_{k,l}^{(m)})_{k=1,\dots,l}$. Moroever, $\Phi_l^{(m)*}(z)$ coincides with

$$\Phi_l^{(m)*}(z) = 1 - 1|\mathrm{sp}(z, z^2, \dots, z^l)$$

and the l–dimensional vector $\mathbf{a}_l^{(m)} = (a_{1,l}^{(m)}, \dots, a_{l,l}^{(m)})^T$ is given by

$$\begin{pmatrix} 1 \\ \mathbf{a}_l^{(m)} \end{pmatrix} = \sigma_l^{2,m} \mathcal{R}_{m,l+1}^{-T} \mathbf{e}_1 \tag{D.5}$$

where $\mathbf{e}_1$ is the $l + 1$–dimensional vector $\mathbf{e}_1 = (1, 0, \dots, 0)^T$. It is moreover easily checked that

$$y_{m,n} - y_{m,n}|\mathrm{sp}(y_{m,n-1}, \dots, y_{m,n-l}) = y_{m,n} + \sum_{k=1}^{l} a_{k,l}^{(m)*} y_{m,n-k} \tag{D.6}$$

where the orthogonal projection operator is this time defined on the space of all finite second moment complex valued random variables. For more details on these polynomials, we refer the reader to [37] and [13].

The matrix $\mathcal{R}_{m,L}^{-1}$ can be written as

$$\mathcal{R}_{m,L}^{-1} = \mathbf{A}_{m,L} \mathrm{Diag}\left(\frac{1}{\sigma_0^{2,m}}, \dots, \frac{1}{\sigma_{L-1}^{2,m}}\right) \mathbf{A}_{m,L}^H \tag{D.7}$$

where $\mathbf{A}_{m,L}$ is the upper-triangular matrix defined by

$$\mathbf{A}_{m,L} = \begin{pmatrix} 1 & a_{1,1}^{(m)} & a_{2,2}^{(m)} & \cdots & a_{L-1,L-1}^{(m)} \\ 0 & 1 & a_{1,2}^{(m)} & \cdots & a_{L-2,L-1}^{(m)} \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}. \tag{D.8}$$

In order to see this, simply observe that $\mathcal{R}_{m,L}\mathbf{A}_{m,L}$ is lower triangular because of (D.5) and the fact that $\mathcal{R}_{m,l+1}^T = \mathbf{J}_{l+1}\mathcal{R}_{m,l+1}\mathbf{J}_{l+1}$. Since $\mathbf{A}_{m,L}^H$ is also lower triangular, so is the product $\mathbf{A}_{m,L}^H \mathcal{R}_{m,L}\mathbf{A}_{m,L}$. However, matrix $\mathbf{A}_{m,L}^H \mathcal{R}_{m,L}\mathbf{A}_{m,L}$ is also hermitian, which implies that it must be diagonal. Close examination of (D.5) reveals that its diagonal entries are equal to $\sigma_l^{2,m}$ for $l = 0, \dots, L - 1$. Inverting the corresponding equation we obtain (D.7).

Using the above decomposition of the matrix $\mathcal{R}_{m,L}^{-1}$ we immediately obtain that

$$\mathbf{a}_L(\nu)^H \mathbf{A}_{m,L} = \frac{1}{\sqrt{L}}\left(1, e^{-2i\pi\nu}\Phi_1^{(m)*}(e^{2i\pi\nu}), \dots, e^{-2i\pi(L-1)\nu}\Phi_{L-1}^{(m)*}(e^{2i\pi\nu})\right)$$

and consequently

$$\mathbf{a}_L(\nu)^H \mathcal{R}_{m,L}^{-1} \mathbf{a}_L(\nu) = \frac{1}{L}\sum_{l=0}^{L-1} |\phi_l^{(m)*}(e^{2i\pi\nu})|^2. \tag{D.9}$$

We first explain informally why, for each m, $\mathcal{S}_m(\nu)\mathbf{a}_L(\nu)^H \mathcal{R}_{m,L}^{-1}\mathbf{a}_L(\nu) - 1$ converges uniformly towards 0. For this, we need to recall certain results that are summarized next.

Since the spectral densities $\mathcal{S}_m(\nu)$ are uniformly bounded from below, we can define the cepstrum coefficients $(c_m(k))_{k\in\mathbb{Z}}$, namely

$$c_m(k) = \int_0^1 \log \mathcal{S}_m(\nu) e^{2\pi i \nu k} d\nu.$$

We notice that $\lim_{l\to+\infty} \sigma_l^{2,m} = \sigma^{2,m}$ coincides with $\exp c_m(0)$. Assumption 5 and a generalization of the Wiener-Lévy theorem (see e.g. [37]) implies that for each m, $c_m \in \ell_\omega$ for each $\gamma \le \gamma_0$. We define the function $\pi^{(m)}(z)$ given by

$$\pi^{(m)}(z) = \exp - \left(c_m(0)/2 + \sum_{n=1}^{+\infty} c_m(-n)z^n \right).$$

Then, $\pi^{(m)}(z)$ and $\psi^{(m)}(z) = \frac{1}{\pi^{(m)}(z)}$ are analytic in the open unit disk $\mathbb{D}$ and continuous on the closed unit disk. In the following, we denote by $\pi^{(m)}(z) = \sum_{n=0}^{+\infty} \pi^{(m)}(n)z^n$ and $\psi^{(m)}(z) = \sum_{n=0}^{+\infty} \psi^{(m)}(n)z^n$ their expansion in $\mathbb{D}$. Moreover, functions $\nu \to \pi^{(m)}(e^{2i\pi\nu})$ and $\nu \to \psi^{(m)}(e^{2i\pi\nu})$ also belong to ℓ_{ω_0}. To check this, we denote by $(\tilde{c}_m(n))_{n\ge0}$ the one-sided sequence defined by $\tilde{c}_m(0) = c_m(0)/2$ and $\tilde{c}_m(n) = c_m(-n)$ for $n \ge 1$. Then, the sequences $\pi^{(m)}$ and $\psi^{(m)}$ can be written as

$$\pi^{(m)} = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k!}(\tilde{c}_m)^{*(k)}, \ \psi^{(m)} = \sum_{k=0}^{+\infty} \frac{1}{k!}(\tilde{c}_m)^{*(k)}$$

where for a sequence a, $a^{*(k)}$ represents $\underbrace{a * a * \ldots * a}_{k}$. Observe, in particular, that both sequences are one-sided. Now, for each $\gamma \le \gamma_0$, it holds that

$$\|\pi^{(m)}\|_\omega \le \sum_{k=0}^{+\infty} \frac{1}{k!}\|\tilde{c}_m\|_\omega^k = \exp(\|\tilde{c}_m\|_\omega) \le \exp(\|c_m\|_\omega) \qquad (D.10)$$

$$\|\psi^{(m)}\|_\omega \le \sum_{k=0}^{+\infty} \frac{1}{k!}\|\tilde{c}_m\|_\omega^k = \exp(\|\tilde{c}_m\|_\omega) \le \exp(\|c_m\|_\omega). \qquad (D.11)$$

In the following, we also need a version of (D.10, D.11) holding uniformly w.r.t. m. For this, we establish the following lemma, which can be seen as a uniform version of the generalized Wiener-Lévy theorem.

Lemma D.1. *Consider a function $F(z)$ holomorphic in a neighbourhood of the interval $[s_{min}, s_{max}]$ where s_{min} and s_{max} are defined in Assumption 3. Then, for*

each $\gamma < \gamma_0$ and for each m, the function $F \circ \mathcal{S}_m$ belongs to ℓ_ω and [5]

$$\sup_{m \geq 1} \|F \circ \mathcal{S}_m\|_\omega < +\infty. \tag{D.12}$$

Proof. We adapt the proof of the Wiener-Levy theorem in [42] (Theorem 5.2, p. 245). We first claim that if p is an integer such that $p > 1 + \gamma_0$ and if $G(\nu) = \sum_{n \in \mathbb{Z}} g(n) e^{2i\pi n\nu}$ belongs to $\mathcal{C}_p$, then, $g \in \ell_{\omega_0}$, and

$$\|g\|_{\omega_0} \leq \kappa \left(\sup_\nu |G(\nu)| + \sup_\nu |G^{(p)}(\nu)| \right) \tag{D.13}$$

for some constant κ depending only on γ_0. To verify (D.13), we remark that $|G(0)| \leq \sup_\nu |G(\nu)|$. Moreover, for each $n \neq 0$, the integration by parts formula leads to

$$g(n) = \frac{1}{(2i\pi n)^p} \int_0^1 G^{(p)}(\nu) e^{-2i\pi n\nu} \, d\nu$$

and to $|g(n)| \leq \frac{1}{(2\pi)^p} \frac{1}{|n|^p} \sup_\nu |G^{(p)}(\nu)|$. As $p > 1 + \gamma_0$, we obtain immediately that (D.13) holds.

Since F is holomorphic in a neighbourhood of $[s_{min}, s_{max}]$, there exists a $\rho > 0$ for which F is holomorphic in the open disk $\mathbb{D}(s, 2\rho)$ for each $s \in [s_{min}, s_{max}]$. In particular, for each m and each ν, F is holomorphic in $\mathbb{D}(\mathcal{S}_m(\nu), 2\rho)$. We consider a partial sum $\mathcal{S}_{m,n_0}(\nu) = \sum_{k=-n_0}^{n_0} r_m(k) e^{-2i\pi k\nu}$, and claim that for each $\gamma < \gamma_0$, we have

$$\|\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu)\|_\omega = \sum_{|k| \geq (n_0+1)} (1 + |k|)^\gamma |r_m(k)| \leq \frac{\kappa}{n_0^{\gamma_0 - \gamma}} \tag{D.14}$$

for some nice constant κ. To justify (D.14), we remark that

$$\|r_m\|_{\omega_0} \geq \sum_{|k| \geq (n_0+1)} (1+|k|)^{\gamma_0} |r_m(k)| \geq n_0^{\gamma_0 - \gamma} \sum_{|k| \geq (n_0+1)} (1+|k|)^\gamma |r_m(k)| = n_0^{\gamma_0 - \gamma} \|\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu)\|_\omega.$$

Assumption 5 implies that $\sup_m \|r_m\|_{\omega_0} < +\infty$. This leads immediately to (D.14). We choose n_0 in such a way that $\frac{\kappa}{n_0^{\gamma_0 - \gamma}} \leq \frac{\rho}{2}$, and notice that (D.14) leads to $\sup_\nu |\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu)| \leq \frac{\rho}{2}$ for each m. Therefore, the circle $\mathbb{C}(\mathcal{S}_{m,n_0}(\nu), \rho)$ with center $\mathcal{S}_{m,n_0}(\nu)$ and radius ρ is included into $\mathbb{D}(\mathcal{S}_m(\nu), 2\rho)$, and $\mathcal{S}_m(\nu)$ belongs to the disk $\mathbb{D}(\mathcal{S}_{m,n_0}(\nu), \rho)$. The Cauchy formula implies that

$$(F \circ \mathcal{S}_m)(\nu) = \frac{1}{2\pi} \int_0^{2\pi} \frac{F(\mathcal{S}_{m,n_0}(\nu) + \rho e^{i\theta})}{\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu) - \rho e^{i\theta}} \, \rho \, e^{i\theta} \, d\theta. \tag{D.15}$$

---

[5]We make the slight abuse of notation by identifying the ω-norm of a function on the unit circle as the corresponding norm of its Fourier coefficient sequence.

Since $|\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu)| \le \frac{\rho}{2}$, it holds that

$$\frac{\rho e^{i\theta}}{\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu) - \rho e^{i\theta}} = -\sum_{k=0}^{+\infty} \rho^{-k} e^{-ik\theta} \left(\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu)\right)^k$$

and that

$$\left\|\frac{\rho e^{i\theta}}{\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu) - \rho e^{i\theta}}\right\|_{\omega} \le \sum_{k=0}^{+\infty} \rho^{-k} \|\mathcal{S}_m - \mathcal{S}_{m,n_0}\|_{\omega}^k \le 2.$$

Using (D.13), it is easy to check that $G_m(\nu,\theta)$ defined by $G_m(\nu,\theta) = F(\mathcal{S}_{m,n_0}(\nu) + \rho e^{i\theta})$ verifies

$$\sup_{m,\theta,\nu} \|G_m(\nu,\theta)\|_{\omega} \le \kappa$$

for each $\gamma \le \gamma_0$ for some nice constant κ. We thus obtain that for some nice constant κ, it holds that

$$\left\|\frac{F(\mathcal{S}_{m,n_0}(\nu) + \rho e^{i\theta})}{\mathcal{S}_m(\nu) - \mathcal{S}_{m,n_0}(\nu) - \rho e^{i\theta}} \rho\, e^{i\theta}\right\|_{\omega} \le \kappa$$

for each $\gamma < \gamma_0$, each m and each θ. (D.15) thus implies (D.12). The proof of Lemma D.1 is thus complete. $\qquad\square$

The use of Lemma D.1 for $f(x) = \log x$ shows that

$$\sup_m \|c_m\|_{\omega} < +\infty \tag{D.16}$$

for each $\gamma < \gamma_0$. Therefore, (D.10, D.11) imply that

$$\sup_m \|\pi^{(m)}\|_{\omega} \le \kappa, \quad \sup_m \|\psi^{(m)}\|_{\omega} \le \kappa. \tag{D.17}$$

It also holds that $\mathcal{S}_m(\nu) = \left|\psi^{(m)}(e^{2i\pi\nu})\right|^2$ and therefore $\psi^{(m)}(z)$ coincides with the outer spectral factor of $\mathcal{S}_m$ in the sense that both $\psi^{(m)}(z)$ and $\frac{1}{\psi^{(m)}(z)} = \pi^{(m)}(z)$ are analytic in the unit disc. Theorem 5.1.8 in [37] leads to the conclusion that $\|\phi_l^{(m)*} - \pi^{(m)}\|_{\omega} \to 0$ when $l \to +\infty$, a result which implies that

$$\sup_{\nu} \left|\phi_l^{(m)*}(e^{2i\pi\nu}) - \pi^{(m)}(e^{2i\pi\nu})\right| \to 0. \tag{D.18}$$

Given the fact that $\mathcal{S}_m(\nu) = \left|\frac{1}{\pi^{(m)}(e^{2i\pi\nu})}\right|^2$, (1.10) and (1.11) allow us to conclude that

$$0 < \inf_m \inf_{\nu} |\pi^{(m)}(e^{2i\pi\nu})| \le \sup_m \sup_{\nu} |\pi^{(m)}(e^{2i\pi\nu})| < +\infty. \tag{D.19}$$

Therefore, (D.18) leads to $\sup_{\nu} |\frac{1}{\pi^{(m)}(e^{2i\pi\nu})}\phi_l^{(m)*}(e^{2i\pi\nu}) - 1| \to 0$, and to $\sup_{\nu} \left||\frac{1}{\pi^{(m)}(e^{2i\pi\nu})}|^2|\phi_l^{(m)*}(e^{2i\pi\nu})|^2 - 1\right| \to 0$, or equivalently, to

$$\sup_{\nu} \left|\mathcal{S}_m(\nu)|\phi_l^{(m)*}(e^{2i\pi\nu})|^2 - 1\right| \to 0. \tag{D.20}$$

76

This, in turn, implies that

$$\sup_{\nu} \left| \mathcal{S}_m(\nu) \frac{1}{L} \sum_{l=1}^{L} |\phi_l^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \to 0 \qquad (D.21)$$

when $L \to +\infty$ as expected. In order to complete the proof of Lemma 5.1, we have thus to prove that (D.21) holds uniformly w.r.t. m, and to evaluate the rate of convergence. For this, we can follow the proof of Theorem 5.1.8 in [37], adapting the corresponding arguments to our particular context.

Theorem 5.1.8 in [37] follows from general results concerning Wiener-Hopf operators defined on the Wiener algebra ℓ_1. As explained below, we will show that $\sup_m \|\phi_l^{(m)*} - \pi^{(m)}\|_1 \to 0$, and will only use that $\sup_m \|r_m\|_\omega < +\infty$ and $\sup_m \|c_m\|_\omega < +\infty$ for each $\gamma < \gamma_0$ in order to obtain an upper bound of the above term. In the following, we denote by $C^{(m)}$ the operator defined on the Wiener algebra ℓ_1 by

$$C^{(m)} a = \overline{r}_m * a$$

where $\overline{r}_m$ is the sequence defined by $\overline{r}_m(n) = r_m(-n)$ for each $n \in \mathbb{Z}$. $C^{(m)}$ can alternatively be defined in the Fourier transform domain as the multiplication operator

$$\sum_{n \in \mathbb{Z}} a(n)e^{2i\pi n\nu} \to \mathcal{S}_m(\nu) \sum_{n \in \mathbb{Z}} a(n)e^{2i\pi n\nu}.$$

It is well known that $\|C^{(m)}\|_1 = \|\overline{r}_m\|_1 = \|r_m\|_1$. As $\mathcal{S}_m(\nu) = |\psi^{(m)}(e^{2i\pi\nu})|^2$, the operator $C^{(m)}$ can be factorized as $C^{(m)} = L^{(m)}U^{(m)} = U^{(m)}L^{(m)}$ where $U^{(m)}$ and $L^{(m)}$ represent the multiplication operators by $\psi^{(m)}(e^{2i\pi\nu})$ and $\left(\psi^{(m)}(e^{2i\pi\nu})\right)^*$ defined on ℓ_1 respectively. We denote by P_+ the projection operator defined on ℓ_1 by

$$P_+ (\{a(n), n \in \mathbb{Z}\}) = \{a(n), n \geq 0\}$$

or equivalently in the Fourier transform domain by

$$P_+ \left(\sum_{n \in \mathbb{Z}} a(n)e^{2i\pi n\nu} \right) = \sum_{n=0}^{+\infty} a(n)e^{2i\pi n\nu}.$$

The operator P_- is defined by $P_- = I - P_+$. The operator $U^{(m)}$ is called upper triangular in the sense that $P_-U^{(m)}P_+ = 0$ while $L^{(m)}$ is lower triangular because $P_+L^{(m)}P_- = 0$. Moreover, as $\pi^{(m)} = \frac{1}{\psi^{(m)}}$ belongs to ℓ_1 and $\pi^{(m)}(e^{2i\pi\nu}) = \sum_{n=0}^{+\infty} \pi^{(m)}(n) e^{2i\pi n\nu}$, the operators $U^{(m)}$ and $L^{(m)}$ are invertible, and $(U^{(m)})^{-1}$ and $(L^{(m)})^{-1}$ are upper triangular and lower triangular respectively. In the Fourier domain, $(U^{(m)})^{-1}$ and $(L^{(m)})^{-1}$ correspond respectively to the multiplication operator by $\pi^{(m)}(e^{2i\pi\nu})$ and $(\pi^{(m)}(e^{2i\pi\nu}))^*$ These properties imply that

the factorization $C^{(m} = L^{(m)}U^{(m)} = U^{(m)}L^{(m)}$ is a Wiener-Hopf factorization. In the following, we denote by $T^{(m)}$ the Toeplitz operator defined on ℓ_1 by

$$T^{(m)} = P_+ C^{(m)} P_+. \tag{D.22}$$

It is clear that if $j \geq 0$ and if δ_j is the sequence δ_j defined by $\delta_j(n) = \delta_{n-j}$, then, $< \delta_i, T^{(m)}\delta_j >$, defined as $\left(T^{(m)}\delta_j\right)(i)$, is equal to $r_m(j-i)$. Therefore, the matrix representation of $T^{(m)}$ in the basis $(\delta_j)_{j\geq 0}$ is the infinite matrix $\mathcal{R}^T_{m,\infty}$. Theorem 5.1.1 in [37] implies that, considered as an operator defined on $\mathrm{Range}(P_+)$, $T^{(m)}$ is invertible, i.e. that for each $a \in \mathrm{Range}(P_+)$, there exists a unique $b \in \mathrm{Range}(P_+)$ such that $T^{(m)}b = a$. $\left(T^{(m)}\right)^{-1} b$ is of course defined as a. If an element a does not belong to $\mathrm{Range}(P_+)$, $\left(T^{(m)}\right)^{-1} a$ is defined as $\left(T^{(m)}\right)^{-1} P_+ a$. We also notice that $\left(T^{(m)}\right)^{-1} = P_+ \left(U^{(m)}\right)^{-1} P_+ \left(L^{(m)}\right)^{-1} P_+$. For each $n \geq 1$, we denote by Q_n the projection operator defined by

$$Q_n \left(\{a(l), l \in \mathbb{Z}\}\right) = \{a(l), 0 \leq l \leq n\} \tag{D.23}$$

or equivalently by

$$Q_n \left(\sum_{l\in\mathbb{Z}} a(l)e^{2i\pi l\nu}\right) = \sum_{l=0}^{n} a(l)e^{2i\pi l\nu}.$$

We also introduce the truncated Toeplitz operator $T_n^{(m)}$ defined by

$$T_n^{(m)} = Q_n C^{(m)} Q_n = Q_n T^{(m)} Q_n. \tag{D.24}$$

We note that in the basis $(\delta_j)_{j=0,\ldots,n}$, the matrix representation of $T_n^{(m)}$ is the matrix $\mathcal{R}^T_{m,n+1}$. We now introduce the projection operator R_n defined by $R_n = P_+ - Q_n$, and state the following Lemma which appears as an immediate consequence of Theorem 5.1.2 and Theorem 5.1.3 in [37].

Lemma D.2. *For each $n \geq 0$, it holds that $R_n L^{(m)} Q_n = R_n L^{-(m)} Q_n = Q_n U^{(m)} R_n = Q_n U^{-(m)} R_n = 0$. Moreover, there exists an integer n_0 independent of m such that for each $n \geq n_0$, $T_n^{(m)}$, considered as an operator defined on $\mathrm{Range}(Q_n)$, is invertible, in the sense that for each $a \in \mathrm{Range}(Q_n)$, it exists a unique $b \in \mathrm{Range}(Q_n)$, defined as $(T_n^{(m)})^{-1}a$, such that $T_n^{(m)}b = a$. If $a \in \mathrm{Range}(P_+)$, $(T_n^{(m)})^{-1}a$ is defined as $(T_n^{(m)})^{-1}a = (T_n^{(m)})^{-1}Q_n a$. Moreover, there exists a nice constant α such that, for each $n \geq n_0$ and each $a \in \mathrm{Range}(P_+)$, the inequality*

$$\left\|\left(T_n^{(m)}\right)^{-1} a\right\|_1 \leq \alpha \, \|a\|_1 \tag{D.25}$$

holds.

Proof. We just verify that $R_n L^{(m)} Q_n = 0$, and omit the proof of the three other identities. For this, we have just to check that if $a(e^{2i\pi\nu}) = \sum_{l=0}^{n} a(n)e^{2i\pi l\nu}$, then

78

$\left(\psi^{(m)}(e^{2i\pi\nu})\right)^* a(e^{2i\pi\nu})$ can be written as

$$\left(\psi^{(m)}(e^{2i\pi\nu})\right)^* a(e^{2i\pi\nu}) = \sum_{l=-\infty}^{n} b(l)e^{2i\pi l\nu}$$

for some coefficients $(b(l))_{l=-\infty,\dots,n}$. This, of course, holds true because $\left(\psi^{(m)}(e^{2i\pi\nu})\right)^* = \sum_{l=0}^{\infty} (\psi^{(m)}(l))^* e^{-2i\pi l\nu}$.

In order to be able to use Theorem 5.1.2 in [37], we establish that it exists an integer n_0 such that $\|P_-(L^{(m)})^{-1}R_n U^{(m)}\|_1 \leq \frac{1}{2}$ and $\|R_n(U^{(m)})^{-1}P_- L^{(m)}\|_1 \leq \frac{1}{2}$ for each $n \geq n_0$ and for each m. If $a \in \ell_1$, we evaluate $P_-(L^{(m)})^{-1}R_n U^{(m)}a$ in the Fourier transform domain, and denote $x_n^{(m)}(e^{2i\pi\nu})$ the function defined by $x_n^{(m)}(e^{2i\pi\nu}) = R_n\psi^{(m)}(e^{2i\pi\nu})a(e^{2i\pi\nu})$, which, of course, can be written as $x_n^{(m)}(e^{2i\pi\nu}) = \sum_{l=n+1}^{+\infty} x_n^{(m)}(l)e^{2i\pi l\nu}$. The operation of $(L^{(m)})^{-1}$ is equivalent to the multiplication by $(\pi^{(m)}(e^{2i\pi\nu}))^*$ in the Fourier transform domain, which is associated to a left-sided series. Therefore,

$$P_-\left(\pi^{(m)}(e^{2i\pi\nu})\right)^* x_n^{(m)}(e^{2i\pi\nu}) = P_-\left[\sum_{l=n+1}^{+\infty} \left(\pi^{(m)}(l)\right)^* e^{-2i\pi l\nu} x_n^{(m)}(e^{2i\pi\nu})\right].$$

The norm of the right hand side can be bounded as

$$\left\|P_-\left[\sum_{l=n+1}^{+\infty}\left(\pi^{(m)}(l)\right)^* e^{-2i\pi l\nu} x_n^{(m)}(e^{2i\pi\nu})\right]\right\|_1 \leq \left\|\sum_{l=n+1}^{+\infty}\left(\pi^{(m)}(l)\right)^* e^{-2i\pi l\nu}\right\|_1 \|\psi^{(m)}\|_1 \|a\|_1$$

or equivalently,

$$\left\|P_-(L^{(m)})^{-1}R_n U^{(m)}\right\|_1 \leq \left(\sum_{l=n+1}^{+\infty} |\pi^{(m)}(l)|\right) \|\psi^{(m)}\|_1.$$

The bound in (D.17) implies that $\sup_m \|\psi^{(m)}\|_1 \leq \kappa$ and that $\sup_m \|\pi^{(m)}\|_\omega \leq \kappa$ for some nice constant κ. It is therefore clear that for each $\gamma < \gamma_0$ and for each m, we have

$$\kappa \geq \|\pi^{(m)}\|_\omega \geq \sum_{l=n+1}^{+\infty} (1+l)^\gamma |\pi^{(m)}(l)| \geq (1+n)^\gamma \sum_{l=n+1}^{+\infty} |\pi^{(m)}(l)|.$$

We conclude from this that

$$\sum_{l=n+1}^{+\infty} |\pi^{(m)}(l)| \leq \frac{\kappa}{n^\gamma} \tag{D.26}$$

and therefore

$$\left\|P_-(L^{(m)})^{-1}R_n U^{(m)}\right\|_1 \leq \frac{\kappa}{n^\gamma} \tag{D.27}$$

for some nice constant κ. It can be shown similarly that

$$\left\|R_n(U^{(m)})^{-1}P_- L^{(m)}\right\|_1 \leq \frac{\kappa}{n^\gamma}. \tag{D.28}$$

This implies that it exists an integer n_0 such that $\|P_-(L^{(m)})^{-1}R_nU^{(m)}\|_1 \leq \frac{1}{2}$ and $\|R_n(U^{(m)})^{-1}P_-L^{(m)}\|_1 \leq \frac{1}{2}$ for each $n \geq n_0$ and for each m. Therefore, Theorem 5.1.2 in [37] implies that for each $n \geq n_0$ and for each m, it holds that $T_n^{(m)}$ is invertible and that for each $a \in \text{Range}(Q_n)$, it holds that $\|(T_n^{(m)})^{-1}a\| \leq \alpha_{m,n}\|a\|_1$ where $\alpha_{m,n}$ is given by

$$\alpha_{m,n} = \left\|(L^{(m)})^{-1}(U^{(m)})^{-1}\right\|_1$$
$$+ 2\max\left(\left\|(U^{(m)})^{-1}\right\|_1, \left\|(L^{(m)})^{-1}\right\|_1\right)\left(\left\|P_-(L^{(m)})^{-1}\right\|_1 + \left\|R_n(U^{(m)})^{-1}\right\|_1\right).$$

The bounds in (D.17) imply that for each m and n, $\alpha_{m,n} \leq \alpha$ for some nice constant α. Therefore, $\|(T_n^{(m)})^{-1}a\| \leq \alpha\|a\|_1$ for each $n \geq n_0$, for each m, and for each $a \in \text{Range}(Q_n)$. If $a \in \text{Range}(P_+)$, $(T_n^{(m)})^{-1}a$ is equal to $(T_n^{(m)})^{-1}Q_na$. Therefore, $\|(T_n^{(m)})^{-1}a\|_1 \leq \alpha\|Q_na\|_1 \leq \alpha\|a\|_1$. This completes the proof of the lemma. $\qquad\square$

Lemma D.2 and Theorem 5.1.3 in [37] imply the following corollary.

Corollary D.1. *For each integer m and for each $a \in \text{Range}(P_+)$, it holds that*

$$\lim_{n \to +\infty}\left\|(T_n^{(m)})^{-1}a - (T^{(m)})^{-1}a\right\|_1 = 0. \tag{D.29}$$

Proof. (D.25) implies that $T_n^{(m)}$ is invertible for each $n \geq n_0$. We use the observation that $(T_n^{(m)})^{-1}T_n^{(m)} = Q_n$. Therefore, the operator $(T_n^{(m)})^{-1} - (T^{(m)})^{-1}$ can be written as

$$(T_n^{(m)})^{-1} - (T^{(m)})^{-1} = (T_n^{(m)})^{-1}\left(T^{(m)} - T_n^{(m)}\right)(T^{(m)})^{-1} + (Q_n - I)(T^{(m)})^{-1}.$$

We conclude from this and (D.25) that for each $n \geq n_0$, it holds that

$$\|(T_n^{(m)})^{-1}a - (T^{(m)})^{-1}a\|_1 \leq \alpha\|(T^{(m)} - T_n^{(m)})(T^{(m)})^{-1}a\|_1 + \|(T^{(m)})^{-1}a - Q_n(T^{(m)})^{-1}a\|_1. \tag{D.30}$$

It is clear that $\|(T^{(m)})^{-1}a - Q_n(T^{(m)})^{-1}a\|_1 \to 0$ when $n \to +\infty$. Moreover, for each $b \in \text{Range}(P_+)$, $(T^{(m)} - T_n^{(m)})b$ can be expressed as

$$(T^{(m)} - T_n^{(m)})b = -\left(Q_nC^{(m)}(Q_n - P_+)b + (Q_n - P_+)C^{(m)}P_+b\right). \tag{D.31}$$

From this, we obtain immediately that for each m, $\|(T^{(m)} - T_n^{(m)})b\|_1 \to 0$ when $n \to +\infty$. Taking $b = (T^{(m)})^{-1}a$ leads to (D.29). $\qquad\square$

Corollary D.1 implies that for each m, $\|(T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0\|_1$ converges towards 0 when $n \to +\infty$. Since the matrix representation of $T_n^{(m)}$ in the basis $(\delta_j)_{j=0,\ldots,n}$ coincides with matrix $\mathcal{R}_{m,n+1}^T$, (D.5) implies that $(T_n^{(m)})^{-1}\delta_0$ coincides with the sequence $\frac{1}{\sigma_n^m}(1, a_{1,n}^{(m)}, \ldots, a_{n,n}^{(m)}, 0, \ldots)$ whose Fourier transform coincides with $\frac{1}{\sigma_n^m}\phi_n^{*(m)}(e^{2i\pi\nu})$. Therefore, the Fourier transform of the ℓ_1 sequence

$(T^{(m)})^{-1}\delta_0$ is the limit of $\frac{1}{\sigma_n^m}\phi_n^{*(m)}(e^{2i\pi\nu})$ in the ℓ_1 metric. Theorem 5.1.8 in [37] implies that for each $\gamma < \gamma_0$ and for each m, $\|\phi_n^{(m)*} - \pi^{(m)}\|_\omega \to 0$, and therefore that $\|\phi_n^{(m)*} - \pi^{(m)}\|_1 \to 0$ as $n \to +\infty$. As it is well known that $\sigma_n^m \to \sigma^m = \exp\frac{c_m(0)}{2}$, this discussion leads to the conclusion that for each m,

$$(T^{(m)})^{-1}\delta_0 = \frac{1}{\sigma^m}\pi^{(m)}. \tag{D.32}$$

In the following, we establish the following proposition.

Proposition D.1. *If $\gamma < \gamma_0$, there exist an integer n_1 and a nice constant κ such that*

$$\sup_{m \geq 1}\|(T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0\|_1 \leq \frac{\kappa}{n^\gamma} \tag{D.33}$$

for each $n \geq n_1$.

Proof. In order to establish (D.33), we use (D.30) and (D.31) for $a = \delta_0$ and $b = (T^{(m)})^{-1}\delta_0 = \frac{1}{\sigma^m}\pi^{(m)}$. We first evaluate $\|(T^{(m)})^{-1}\delta_0 - Q_n(T^{(m)})^{-1}\delta_0\|_1$, or equivalently $\frac{1}{\sigma^m}\sum_{k=n+1}^{+\infty}|\pi^{(m)}(n)|$. In order to check that $\sup_m \frac{1}{\sigma^m} < +\infty$, we notice that (1.11) implies that $\inf_m c_0(m) > -\infty$, and that $\inf_m \exp\frac{c_0(m)}{2} > 0$. Therefore, it holds that $\sup_m \frac{1}{\sigma^m} < +\infty$. The bound in (D.26) thus implies that for each $n \geq n_0$ and for each m, it holds that

$$\left\|(T^{(m)})^{-1}\delta_0 - Q_n(T^{(m)})^{-1}\delta_0\right\|_1 \leq \frac{\kappa}{n^\gamma}$$

for some nice constant κ. It remains to control $\|(T^{(m)} - T_n^{(m)})(T^{(m)})^{-1}\delta_0\|_1$. As $\sup_m \frac{1}{\sigma^m} < +\infty$, it is sufficient to study $\|(T^{(m)} - T_n^{(m)})\pi^{(m)}\|_1$. For this, we use (D.31) for $b = \pi^{(m)}$, and obtain that

$$\left\|(T^{(m)} - T_n^{(m)})\pi^{(m)}\right\|_1 \leq \left\|C^{(m)}\right\|_1\left\|\pi^{(m)} - Q_n\pi^{(m)}\right\|_1 + \left\|(P_+ - Q_n)C^{(m)}\pi^{(m)}\right\|_1. \tag{D.34}$$

The bound in (D.26) implies that the first term of the right hand side of (D.34) is upper bounded by $\frac{\kappa}{n^\gamma}$ for some nice constant κ for each n and each m. The second term of the right hand side of (D.34) is given by

$$\left\|(P_+ - Q_n)C^{(m)}\pi^{(m)}\right\|_1 = \sum_{k=n+1}^{\infty}\left|\left(C^{(m)}\pi^{(m)}\right)(k)\right|$$

where it holds that

$$\left(C^{(m)}\pi^{(m)}\right)(k) = \sum_{l=0}^{+\infty}\overline{r}_m(k-l)\,\pi^{(m)}(l).$$

Therefore,

$$\sum_{k=n+1}^{\infty}\left|\left(C^{(m)}\pi^{(m)}\right)(k)\right| \leq \sum_{k=n+1}^{+\infty}\sum_{l=0}^{+\infty}|r_m(k-l)||\pi^{(m)}(l)|.$$

We express the right hand side of the above inequality as

$$\sum_{k=n+1}^{+\infty} \sum_{l=0}^{+k} |r_m(k-l)||\pi^{(m)}(l)| + \sum_{k=n+1}^{+\infty} \sum_{l=k+1}^{+\infty} |r_m(k-l)||\pi^{(m)}(l)|$$

or equivalently as

$$\sum_{k=n+1}^{+\infty} \sum_{u+v=k,u\geq 0,v\geq 0} |r_m(u)||\pi^{(m)}(v)| + \sum_{k=n+1}^{+\infty} \sum_{u+v=k,u\leq -1,v\geq 0} |r_m(u)||\pi^{(m)}(v)|.$$

It is clear that

$$\sum_{k=n+1}^{+\infty} \sum_{u+v=k,u\geq 0,v\geq 0} |r_m(u)||\pi^{(m)}(v)| \leq \left(\sum_{l=0}^{+\infty} |\pi^{(m)}(l)| \right) \left(\sum_{k=[(n+1)/2]}^{+\infty} |r_m(k)| \right) +$$

$$\left(\sum_{k=0}^{+\infty} |r_m(k)| \right) \left(\sum_{l=[(n+1)/2]}^{+\infty} |\pi^{(m)}(l)| \right)$$

and that

$$\sum_{k=n+1}^{+\infty} \sum_{u+v=k,u\leq -1,v\geq 0} |r_m(u)||\pi^{(m)}(v)| \leq \left(\sum_{k\leq -1} |r_m(k)| \right) \left(\sum_{l=n+1}^{+\infty} |\pi^{(m)}(l)| \right).$$

Using the fact that that $\sup_m \|r_m\|_\omega < +\infty$, we obtain, using the same arguments as in (D.26), that

$$\sup_m \sum_{l=n+1}^{+\infty} |\pi^{(m)}(l)| < \frac{\kappa}{n^\gamma}$$

for some nice constant κ. We have thus shown that

$$\sup_m \|(P_+ - Q_n)C^{(m)}\pi^{(m)}\|_1 \leq \frac{\kappa}{n^\gamma}$$

and this completes the proof of Proposition D.1. $\qquad\square$

Proposition D.1 immediately allows to study the behaviour of $\|\phi_n^{*(m)} - \pi^{(m)}\|_1$ when $n \to +\infty$.

Corollary D.2. *If $\gamma < \gamma_0$, it exists an integer n_2 and a nice constant κ for which*

$$\|\phi_n^{(m)*} - \pi^{(m)}\|_1 \leq \frac{\kappa}{n^\gamma} \tag{D.35}$$

for each $n \geq n_2$ and each m.

Proof. $\phi_n^{(m)*} - \pi^{(m)}$ coincides with $\sigma_n^m (T_n^{(m)})^{-1}\delta_0 - \sigma^m (T^{(m)})^{-1}\delta_0$, which can also be written as

$$\phi_n^{(m)*} - \pi^{(m)} = \sigma_n^m \left((T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0 \right) + (\sigma_n^m - \sigma^m)(T^{(m)})^{-1}\delta_0$$

82

or equivalently as

$$\phi_n^{(m)*} - \pi^{(m)} = \sigma_n^m \left((T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0 \right) + (\sigma_n^m - \sigma^m)\frac{\pi^{(m)}}{\sigma^m}. \qquad (D.36)$$

We notice that $\sigma_n^m = \langle (T_n^{(m)})^{-1}\delta_0, \delta_0 \rangle^{-1}$ and that $\sigma^m = \langle (T^{(m)})^{-1}\delta_0, \delta_0 \rangle^{-1}$. We express $\sigma_n^m - \sigma^m$ as

$$\sigma_n^m - \sigma^m = \sigma_n^m \sigma^m \left(\frac{1}{\sigma^m} - \frac{1}{\sigma_n^m} \right) = \sigma_n^m \sigma^m \left\langle (T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0), \delta_0 \right\rangle.$$

Noting that $\sup_{m,n} \sigma_n^m \leq \sup_m r_0(m) < +\infty$, we obtain that for each n large enough and for each m, the inequality

$$\sigma_n^m - \sigma^m \leq \kappa \| (T_n^{(m)})^{-1}\delta_0 - (T^{(m)})^{-1}\delta_0 \|_1 \leq \frac{\kappa}{n^\gamma}$$

holds for some nice constant κ. (D.35) thus follows immediately from Proposition D.1. $\qquad\square$

We finally complete the proof of Lemma 5.1. (D.35) implies that

$$\sup_m \sup_\nu |\phi_n^{(m)*}(e^{2i\pi\nu}) - \pi^{(m)}(e^{2i\pi\nu})| \leq \frac{\kappa}{n^\gamma}$$

for each $n \geq n_2$. Using (D.19) and $\mathcal{S}_m(\nu) = \frac{1}{|\pi^{(m)}(e^{2i\pi\nu})|^2}$, we obtain that

$$\sup_m \sup_\nu \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \leq \frac{\kappa}{n^\gamma} \qquad (D.37)$$

for each $n \geq n_2$. We recall that $\epsilon_m(\nu)$ is equal to

$$\epsilon_{m,L}(\nu) = \frac{1}{L} \sum_{n=0}^{L-1} \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1.$$

Therefore,

$$|\epsilon_{m,L}(\nu)| \leq \frac{1}{L} \sum_{n=0}^{L-1} \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right|.$$

We express the right hand side as

$$\frac{1}{L} \sum_{n=0}^{n_2-1} \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| + \frac{1}{L} \sum_{n=n_2}^{L} \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right|$$

and handle the two terms separately. On the one hand, (D.37) implies that

$$\frac{1}{L} \sum_{n=n_2}^{L} \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \leq \kappa \frac{1}{L} \sum_{n=n_2}^{L} \frac{1}{n^\gamma}.$$

If $\gamma > 1$, $\sum_{n=n_2}^{L} \frac{1}{n^\gamma}$ is a bounded term, and we obtain that

$$\sup_m \sup_\nu \frac{1}{L} \sum_{n=n_2}^{L} \left| \mathcal{S}_m(\nu)|\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \leq \frac{\kappa}{L}.$$

If $\gamma = 1$, the above term is bounded by $\kappa \frac{\log L}{L}$, and if $0 < \gamma < 1$, it holds that

$$\sum_{n=n_2}^{L} \frac{1}{n^\gamma} \leq \kappa L^{1-\gamma}$$

and that

$$\sup_m \sup_\nu \frac{1}{L} \sum_{n=n_2}^{L} \left| \mathcal{S}_m(\nu) |\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \leq \frac{\kappa}{L^\gamma}.$$

We finally justify that there exists a nice constant κ such that

$$\sup_m \sup_\nu \sum_{n=0}^{n_2-1} \left| \mathcal{S}_m(\nu) |\phi_n^{(m)*}(e^{2i\pi\nu})|^2 - 1 \right| \leq \kappa.$$

Indeed, since n_2 is a fixed integer, we have just to verify that for each $n \leq n_2$, $\sup_m \sup_\nu |\phi_n^{(m)*}(e^{2i\pi\nu})| < +\infty$. For this, we recall that the non normalized polynomials $\Phi_n^{(m)}$ and $\Phi_n^{(m)*}$ verify the relation the well known recursion formula

$$\Phi_{n+1}^{(m)}(z) = z\,\Phi_n^{(m)}(z) - \alpha_n^{(m)}\,\Phi_n^{(m)*}(z) \tag{D.38}$$

$$\Phi_{n+1}^{(m)*}(z) = \Phi_n^{(m)*}(z) - \alpha_n^{(m)*}\,z\,\Phi_n^{(m)}(z). \tag{D.39}$$

Here, $(\alpha_m(n))_{n\geq 0}$ are the reflection coefficients sequence associated to autocovariance $(r_m(n))_{n\in\mathbb{Z}}$, also called in [37] the Verblunsky coefficients. For each n, it holds that $|\alpha_m(n)| < 1$. It is obvious that $\|\Phi_n^{(m)}\|_1 = \|\Phi_n^{(m)*}\|_1$. Therefore, (D.38) implies that

$$\|\Phi_{n+1}^{(m)*}\|_1 \leq (1 + |\alpha_m(n)|)\|\Phi_n^{(m)*}\|_1 \leq 2\|\Phi_n^{(m)*}\|_1.$$

Noting that $\|\Phi_0^{(m)*}\|_1 = 1$, we obtain that $\|\Phi_n^{(m)*}\|_1 \leq 2^n$, and that $\sup_m \sup_\nu |\Phi_n^{(m)*}(e^{2i\pi\nu})| \leq 2^n$. As $\inf_{m,n} \sigma_n^m > 0$, the normalized polynomials verify $\sup_m \sup_\nu |\phi_n^{(m)*}(e^{2i\pi\nu})| < +\infty$. This completes the proof of Lemma 5.1.