Random Matrices: Theory and Applications
2250053 (60 pages)

DOI: 10.1142/S2010326322500538

Spectral measure of empirical autocovariance matrices of high-dimensional Gaussian stationary processes

Arup Bose

Statistics and Mathematics Unit
Indian Statistical Institute, Kolkata
bosearu@gmail.com

Walid Hachem*

LIGM, CNRS, Univ Gustave Eiffel, ESIEE Paris
F-77454 Marne-la-Vallée, France
walid.hachem@univ-eiffel.fr

Received 26 October 2021
Revised 31 May 2022
Accepted 12 July 2022
Published

Consider the empirical autocovariance matrices at given non-zero time lags, based on observations from a multivariate complex Gaussian stationary time series. The spectral analysis of these autocovariance matrices can be useful in certain statistical problems, such as those related to testing for white noise. We study the behavior of their spectral measure in the asymptotic regime where the time series dimension and the observation window length both grow to infinity, and at the same rate. Following a general framework in the field of the spectral analysis of large random non-Hermitian matrices, at first the probabilistic behavior of the small singular values of a shifted version of the autocovariance matrix is obtained. This is then used to obtain the asymptotic behavior of the empirical spectral measure of the autocovariance matrices at any lag. Matrix orthogonal polynomials on the unit circle play a crucial role in our study.

Keywords: High-dimensional times series analysis; large non-Hermitian matrix theory; limit spectral distribution; matrix orthogonal polynomials; multivariate stationary processes; small singular values.

Mathematics Subject Classification 2020: Primary: 60B20; Secondary: 60G57, 62M10, 33C47

*Corresponding author.

1. Background, assumptions and results

1.1. *Background*

Consider a multivariate time-series sequence $(\boldsymbol{x}^{(N)})_{N=1,2,\ldots}$, where for each $N \in \mathbb{N}\backslash\{0\}$, the process $\boldsymbol{x}^{(N)} = (\boldsymbol{x}_k^{(N)})_{k\in\mathbb{Z}}$ is a $\mathbb{C}^N$-valued centered Gaussian stationary process in the discrete time parameter k. Let

$$R_L^{(N)} = \mathbb{E}\boldsymbol{x}_L^{(N)}(\boldsymbol{x}_0^{(N)})^*$$

be the *autocovariance matrix* of $\boldsymbol{x}^{(N)}$ at lag L (throughout this paper, * stands for the conjugate transpose). Let $(n_N)_N$ be an increasing sequence of positive integers such that

$$0 < \liminf_{N\to\infty} \frac{N}{n_N} \le \limsup_{N\to\infty} \frac{N}{n_N} < \infty. \tag{1.1}$$

Assume that for each N, we have the sample $\boldsymbol{x}_0^{(N)}, \ldots, \boldsymbol{x}_{n_N-1}^{(N)}$ of the process $\boldsymbol{x}^{(N)}$. Fixing an integer $L \ge 0$, the *empirical autocovariance matrix* of order L is given by

$$\widehat{R}_L^{(N)} = \frac{1}{n_N}\sum_{\ell=0}^{n_N-1} \boldsymbol{x}_{\ell+L}^{(N)}\boldsymbol{x}_\ell^{(N)*} \quad \in \mathbb{C}^{N\times N},$$

where the sum $\ell + L$ will be taken modulo–n_N. For any matrix $M \in \mathbb{C}^{m\times m}$, let $\{\lambda_0(M), \ldots, \lambda_{m-1}(M)\}$ be its eigenvalues. The spectral measure of $\widehat{R}_L^{(N)}$ is then defined as

$$\mu_N = \frac{1}{N}\sum_{\ell=0}^{N-1} \delta_{\lambda_\ell(\widehat{R}_L^{(N)})}.$$

We are interested in studying this measure as $N \to \infty$. In the field of multivariate time series analysis, it is classically assumed that N is fixed while the observation window length increases to ∞, in which case, under standard sets of assumptions, μ_N converges weakly to the spectral measure of $R_L^{(N)}$ in the almost sure sense. This is no more true in the asymptotic regime that we consider in this paper, where the time series dimension and the window length are both large and of the same order of magnitude.

Note that for $L = 0$, $\widehat{R}_L^{(N)}$ is Hermitian and several results are known for this case under different assumptions. Our aim is to consider the cases $L \ge 1$ in which case $\widehat{R}_L^{(N)}$ are non-Hermitian. Generally speaking, the study of the spectral measure of non-Hermitian matrices is much harder than that for Hermitian matrices. See for example [18, 23, 31]. As an example, it took a tremendous amount of effort from researchers over a long period of time to establish the limit of the empirical spectral measure of the matrix all whose elements are real-valued iid with mean zero and variance 1 (see [38]).

In [7], using the ideas from [23, 31, 40], we identified the limit spectral measure of $\widehat{R}_L^{(N)}$ in the particular case when the time series is a (complex) white noise process. The same setting is considered in [42], but they relax the modulo-n_N summation when constructing $\widehat{R}_L^{(N)}$. However, when the time series is not a white noise, no such result appears to exist in the literature.

Spectral properties of the sample autocovariance matrices in a stationary time series carry information on the process and hence potentially, can be used for statistical inference. Some work in this area were initiated by [2, 3]. See also [6] for a book-level exposition. For example, plots of the empirical spectral measure of the sample autocovariance matrices for different lags can serve as graphical tests for white noise, or for the order of dependence in the time series. Some first results in this vein were proposed in [7]. The nature of the empirical spectral measure of $\widehat{R}_L^{(N)}$ for different values of L reflect the degree of dependence that exists in the underlying process. Theoretical support for these tests rests on the asymptotic behavior of the empirical spectral measure. Moreover, once this behavior is identified, it can potentially be used to develop significance tests for other statistical hypothesis too. This issue will be taken up elsewhere.

For any matrix $M \in \mathbb{C}^{N\times N}$, let $s_0(M) \geq s_1(M) \geq \cdots \geq s_{N-1}(M)$ be its singular values arranged in a non-increasing order. It is well known that for a non-Hermitian matrix, say $M_N \in \mathbb{C}^{N\times N}$, as $N \to \infty$, the behavior of its spectral measure is connected to the probabilistic behavior of the small singular values of the related matrix $M_N - z \triangleq M_N - zI_N$ for $z \in \mathbb{C}\backslash\{0\}$. With this in mind, we shall seek solutions to the following problems:

- The behavior of the smallest singular value $s_{N-1}(\widehat{R}_L^{(N)} - z)$ for an arbitrary $z \in \mathbb{C}\backslash\{0\}$.
- For an arbitrary $\beta \in (0, 1)$, the behavior of the "small" singular values $s_{N-\ell}(\widehat{R}_L^{(N)} - z)$ for $\ell \in \{\lfloor N^\beta\rfloor, \ldots, \lfloor N/2\rfloor\}$ and $z \in \mathbb{C}\backslash\{0\}$. Later this will help in controlling the magnitude of the singular values $s_{N-\ell}$ when the indices ℓ are close to N^β.
- The behavior of μ_N as $N \to \infty$ via the existence of a *deterministic equivalent*.

1.2. *Assumptions*

We shall assume that for every N, $(\boldsymbol{x}_k^{(N)})_k$ is a stationary Gaussian process whose spectral density exists and satisfies some reasonable regularity conditions. As we shall see, this provides the opportunity to use a variety of technical tools. If the time series are not Gaussian, then the situation is way more involved technically but results similar to those in this paper are expected to hold under suitable restrictions on the time series. Let $\mathbb{T}$ denote the unit circle of the complex plane. Let $\mathcal{H}_+^N$ be the set of $N \times N$ Hermitian non-negative matrices. Suppose that for each positive integer N, there is an integrable function $S^{(N)} : \mathbb{T} \to \mathcal{H}_+^N$ such that for each $L \in \mathbb{Z}$,

$$R_L^{(N)} = \frac{1}{2\pi}\int_0^{2\pi} e^{-\imath L\theta} S^{(N)}(e^{\imath\theta})\, d\theta. \tag{1.2}$$

This $S^{(N)}$ is called the *spectral density* of $\{R_L^{(N)}, L \geq 0\}$ or of the corresponding stationary process [30, Chap. 1; 9]. We assume that for each N, $S^{(N)}$ is nontrivial in the sense that for each non-zero $\mathbb{C}^N$-valued polynomial $p(z)$

$$\int_0^{2\pi} p(e^{\imath\theta})^* S^{(N)}(e^{\imath\theta}) p(e^{\imath\theta}) d\theta > 0 \quad \text{(that is, the matrix is positive definite)}.$$

We now turn to the more substantial assumptions on $S^{(N)}$. The first assumption is akin to uniform equicontinuity of $\{S^{(N)}\}$. For $h > 0$, let

$$\boldsymbol{w}(S^{(N)}, h) = \sup_\theta \sup_{|\psi| \leq h} \|S^{(N)}(e^{\imath(\theta+\psi)}) - S^{(N)}(e^{\imath\theta})\|$$

be the modulus of continuity of $S^{(N)}$ with respect to the spectral norm $\|\cdot\|$.

Assumption 1.1. (i) For any $\varepsilon > 0$, there exists $h > 0$ such that

$$\sup_{N \in \mathbb{N}} \boldsymbol{w}(S^{(N)}, h) < \varepsilon.$$

(ii) With $\|S^{(N)}\|_\infty^{\mathbb{T}} = \max_\theta \|S^{(N)}(e^{\imath\theta})\|$,

$$\boldsymbol{M} := \sup_N \|S^{(N)}\|_\infty^{\mathbb{T}} < \infty.$$

(iii) $\inf_N s_{N-1}(R_0^{(N)}) > 0$.

Regarding the last assumption, note that

$$s_{N-1}(R_0^{(N)}) = s_{N-1}\left(\frac{1}{2\pi}\int_0^{2\pi} S^{(N)}(e^{\imath\theta})\, d\theta\right).$$

We allow the spectral density $S^{(N)}(e^{\imath\theta})$ to be singular or close to singular at some points of $\mathbb{T}$, but within the restrictions provided by the two following assumptions. Assumption 1.2 stipulates that $S^{(N)}(e^{\imath\theta})$ can be close to being singular only on a set of frequencies with a small Lebesgue measure and it implies Assumption 1.1 (iii). Assumption 1.3 puts additional constraint on $s_{N-1}(S^{(N)}(e^{\imath\theta}))$. Examples where Assumptions 1.2 and 1.3 are satisfied are provided in Sec. 2. Let Leb$(\cdot)$ denote the Lebesgue measure.

Assumption 1.2. Suppose that for any $\kappa \in (0, 1)$, there exists $\delta > 0$ such that

$$\sup_N \mathrm{Leb}\{z \in \mathbb{T} : s_{N-1}(S^{(N)}(z)) \leq \delta\} \leq \kappa.$$

Assumption 1.3.

$$\frac{1}{N}\int_0^{2\pi} \log s_{N-1}(S^{(N)}(e^{\imath\theta})) d\theta \xrightarrow[N\to\infty]{} 0.$$

1.3. *Results*

Before we state our results, we wish to recall that we have used modulo-n_N summation to construct $\widehat{R}_L^{(N)}$. This is for convenience and helps in the details of the proofs. We believe that the results we establish continue to hold for the sample autocovariance matrices when we define them via the usual summation over all indices that maintain a lag L. See for instance [42] who relax the modulo-n_N summation in the context of the model with i.i.d. processes considered in [7]. These results can be adapted to our model with some work.

The first result is on a probabilistic bound on the magnitude of the smallest singular value of $(\widehat{R}_L^{(N)} - z)$. In problems similar to ours, often the optimal factor at the left-hand side of the bound given below is $N^{-1}t$ instead of our $N^{-3/2}t$ when $\varepsilon > 0$ is fixed. Our weaker bound will serve our purpose. We shall elaborate on this issue during the course of the proof.

Theorem 1.1. *Suppose Assumptions* 1.1–1.3 *hold. Then, for each $z \neq 0$ and arbitrarily small $\varepsilon > 0$, there exists a constant $c_{1.1}$ such that for all small $t > 0$ and for all large N,*

$$\mathbb{P}[s_{N-1}(\widehat{R}_L^{(N)} - z) \leq N^{-3/2}t] \leq \varepsilon t + \exp(-c_{1.1}\varepsilon^2 N).$$

The behavior of the small singular values is handled by Theorem 1.2. The behavior of $s_{N-k}(\widehat{R}_L^{(N)} - z)$ for values of k which are close to N^β is more important. The theorem implies that $s_{N-N^\beta}(\widehat{R}_L^{(N)} - z) \gtrsim N^{\beta/2-1}$ with large probability. Again, though the rate is not optimal, it will be sufficient for our needs. Further comments on this issue will be provided in the course of the proof, see Remark 3.1.

Theorem 1.2. *Suppose Assumptions* 1.1–1.3 *hold. Let $\beta \in (0, 1)$. Then, for each $z \neq 0$, there exist two positive constants $c_{1.2}$ and $C_{1.2}$ such that for all $k \in [\lfloor N^\beta \rfloor, \lfloor N/2 \rfloor]$, and for $N \geq N_0$, where N_0 is independent of k,*

$$\mathbb{P}[s_{N-k-1}(\widehat{R}_L^{(N)} - z) \leq C_{1.2}\sqrt{k}/N] \leq \exp(-c_{1.2}k).$$

We now turn to the large-N behavior of μ_N. For this we rely on the well-known Hermitization technique due to Girko [16] (see [5] for a comprehensive exposition). Let μ be a probability measure on $\mathbb{C}$ that integrates $\log|\cdot|$ near infinity. Then its log-potential $U_\mu(\cdot) : \mathbb{C} \to (-\infty, \infty]$ is defined below. The measure μ can be recovered from $U_\mu(z)$

$$U_\mu(z) = -\int_{\mathbb{C}} \log|w - z| \; \mu(dw).$$

For the empirical spectral measure μ_N, we can write

$$\begin{aligned} U_{\mu_N}(z) &= -\frac{1}{N}\sum_{\ell=0}^{N-1} \log|\lambda_\ell(\widehat{R}_L^{(N)}) - z| = -\frac{1}{2N}\log\det(\widehat{R}_L^{(N)} - z)(\widehat{R}_L^{(N)} - z)^* \\ &= -\int \log t \; \nu_{z,N}(dt), \end{aligned}$$

where $\nu_{z,N}$, the empirical measure (on $\mathbb{R}$) of the singular values of the matrix $\widehat{R}_L^{(N)} - z$ given by

$$\nu_{z,N} = \frac{1}{N}\sum_{\ell=0}^{N-1} \delta_{s_\ell(\widehat{R}_L^{(N)}-z)}.$$

Given a matrix $M \in \mathbb{C}^{N\times N}$, denote its so-called *Hermitized* version as

$$\boldsymbol{H}(M) = \begin{bmatrix} & M \\ M^* & \end{bmatrix}.$$

As is well known, the spectral measure $\check{\nu}_{z,N}$ (on $\mathbb{R}$) of $\boldsymbol{H}(\widehat{R}_L^{(N)} - z)$ is given by

$$\check{\nu}_{z,N} = \frac{1}{2N}\sum_{\ell=0}^{N-1} (\delta_{s_\ell(\widehat{R}_L^{(N)}-z)} + \delta_{-s_\ell(\widehat{R}_L^{(N)}-z)}).$$

The measure $\check{\nu}_{z,N}$ is clearly symmetric in the sense that $\check{\nu}_{z,N}(B) = \check{\nu}_{z,N}(-B)$ for each Borel set $B \subset \mathbb{R}$, and it is the symmetrized version of $\nu_{z,N}$. Observe that $\int \log t \ \nu_{z,N}(dt) = \int \log|t| \ \check{\nu}_{z,N}(dt)$.

Note that all of these are *random measures*. To study the behavior of μ_N, we will find it more convenient to study $\check{\nu}_{z,N}$ instead of $\nu_{z,N}$, since the matrix $\boldsymbol{H}(\widehat{R}_L^{(N)} - z)$ is Hermitian. This approach is formalized in the following general proposition which provides conditions under which the sequence of random measures (μ_N) can be approximated by some sequence $(\boldsymbol{\mu}_N)$ of deterministic measures.

Proposition 1.1. *Assume that for almost every $z \in \mathbb{C}$, the following two conditions hold*:

(1) *With probability one,* $\log|\cdot|$ *is uniformly integrable with respect to* $\{\check{\nu}_{z,N}\}_N$.
(2) *There exists a tight sequence of deterministic symmetric probability measures* $(\check{\boldsymbol{\nu}}_{z,N})_N$ *on* $\mathbb{R}$ *such that for each bounded and continuous function* $f : \mathbb{R} \to \mathbb{R}$

$$\int f \ d\check{\nu}_{z,N} - \int f \ d\check{\boldsymbol{\nu}}_{z,N} \xrightarrow[N\to\infty]{a.s.} 0.$$

Then, there exists a tight sequence of deterministic probability measure $(\boldsymbol{\mu}_N)$ *on* $\mathbb{C}$ *such that for each bounded and continuous function* $f : \mathbb{C} \to \mathbb{C}$, *we have*

$$\int f \ d\mu_N - \int f \ d\boldsymbol{\mu}_N \xrightarrow[N\to\infty]{a.s.} 0.$$

Moreover, the logarithmic potential of $\boldsymbol{\mu}_N$ *is*

$$U_{\boldsymbol{\mu}_N}(z) = -\int \log|t| \ \check{\boldsymbol{\nu}}_{z,N}(dt).$$

This proposition is very close to [5, Lemma 4.3] (see also [11]). The proof of this lemma can be adapted to our situation by considering converging subsequences of $(\check{\boldsymbol{\nu}}_{z,N})_N$.

The uniform integrability condition needed in Proposition 1.1 is ensured by Theorems 1.1 and 1.2 (again, see [5, Sec. 4.2] for the proof details). We are thus left to consider the asymptotic properties of $\check{\nu}_{z,N}$ with a goal to comply with Condition 2 of Proposition 1.1. Classically, the central object that is used for this is the resolvent (in what follows $z \in \mathbb{C}$ is arbitrary)

$$Q^{(N)}(z,\eta) = (\boldsymbol{H}(\widehat{R}_L^{(N)} - z) - \eta I_{2N})^{-1}$$

of $\boldsymbol{H}(\widehat{R}_L^{(N)} - z)$ in the complex variable $\eta \in \mathbb{C}_+ \triangleq \{w \in \mathbb{C} : \Im w > 0\}$. The resolvent $Q^{(N)}(z,\cdot)$ is a typical example of a so-called matrix Stieltjes transform. Before we recall its nature, let us recall that the Stieltjes transform of any probability measure ν on $\mathbb{R}$ is given by

$$g_\nu(\eta) = \int_{\mathbb{R}} \frac{1}{\lambda - \eta} \nu(d\lambda), \quad \eta \in \mathbb{C}\backslash\mathbb{R}.$$

The real and imaginary parts of a square matrix M are, respectively, given as

$$\Re M = \frac{M + M^*}{2} \quad \text{and} \quad \Im M = \frac{M - M^*}{2\imath}.$$

Given an integer $m > 0$, we let $\mathcal{M}_+^m$ denote the set of matrices $M \in \mathbb{C}^{m\times m}$ such that $\Im M > 0$.

The following result can be found in the literature dealing with the moment problem and related topics, see, e.g. [4, pp. 64–65; 15; 19, Proposition 2.2 and Appendix A].

Proposition 1.2 (Matrix Stieltjes transform). *Let $F : \mathbb{C}_+ \to \mathbb{C}^{m\times m}$ be a matrix-valued function. Then, the following facts are equivalent*:

(1) *F is the Stieltjes transform of an $\mathcal{H}_+^m$-valued measure μ on $\mathbb{R}$ such that $\mu(\mathbb{R}) = I_m$.*
(2) *F is analytic, $F(\eta) \in \mathcal{M}_+^m$ for $\eta \in \mathbb{C}_+$, and $-\imath t F(\imath t)$ converges to I_m as $t \to \infty$.*

Such an F is called a *matrix Stieltjes transform*. Let $\mathfrak{S}^m$ denote the set of all such $m \times m$ matrix Stieltjes transforms. It is known that if $F \in \mathfrak{S}^m$, then

(1) $\|F(\eta)\| \le 1/\Im\eta$.
(2) $m^{-1} \operatorname{tr} F(\eta) = g_\nu(\eta)$ for some probability measure ν on $\mathbb{R}$.

An illustration of (2) is provided by the resolvent we just defined: it can be checked that

$$Q^{(N)}(z,\cdot) \in \mathfrak{S}^{2N} \quad \text{and} \quad g_{\check{\nu}_{z,N}}(\eta) = (2N)^{-1} \operatorname{tr} Q^{(N)}(z,\eta).$$

In the remainder of this paper, whenever we write $M \in \mathbb{C}^{2N\times 2N}$ as $M = \left[\begin{smallmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{smallmatrix}\right]$, it is understood that the blocks M_{uv} belong to $\mathbb{C}^{N\times N}$. With this notation, we define the linear operator $\mathscr{T} : \mathbb{C}^{2N\times 2N} \to \mathbb{C}^{2\times 2}$ as

$$\mathscr{T}\left(\begin{bmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{bmatrix}\right) = \begin{bmatrix} \operatorname{tr} M_{00}/n_N & \operatorname{tr} M_{01}/n_N \\ \operatorname{tr} M_{10}/n_N & \operatorname{tr} M_{11}/n_N \end{bmatrix}.$$

Given an integer $L > 0$, we also define the 2×2 Hermitian unitary matrix function U_L on $\mathbb{T}$ as

$$U_L(e^{\imath\theta}) = \begin{bmatrix} & e^{-\imath L\theta} \\ e^{\imath L\theta} & \end{bmatrix}.$$

The next technical result is on the existence of a unique solution for a functional equation. It will be used to show that $(\check{\boldsymbol{\nu}}_{z,N})_N$ approximates $(\check{\nu}_{z,N})_N$. Denote as $\otimes$ the Kronecker product between matrices.

Theorem 1.3. *Let $\Sigma : \mathbb{T} \to \mathcal{H}_+^N$ be a continuous function, and let $z \in \mathbb{C}$. Given a function $M(\eta) \in \mathfrak{S}^{2N}$, the function displayed below is well-defined and belongs to $\mathfrak{S}^{2N}$ as a function of η.*

$$\mathcal{F}_{\Sigma,z}(M(\eta),\eta) = \left(\frac{1}{2\pi}\int_0^{2\pi} (\mathscr{T}((I_2 \otimes \Sigma(e^{\imath\theta}))M(\eta)) + U_L(e^{\imath\theta}))^{-1} \otimes \Sigma(e^{\imath\theta})\, d\theta - \begin{bmatrix} \eta & z \\ \bar{z} & \eta \end{bmatrix} \otimes I_N \right)^{-1}.$$

Moreover, the functional equation in the parameter $\eta \in \mathbb{C}_+$

$$P(z,\eta) = \mathcal{F}_{\Sigma,z}(P(z,\eta),\eta) \tag{1.3}$$

admits a unique solution in the class $P(z,\cdot) \in \mathfrak{S}^{2N}$. Write

$$P(z,\cdot) = \begin{bmatrix} P_{00}(z,\cdot) & P_{01}(z,\cdot) \\ P_{10}(z,\cdot) & P_{11}(z,\cdot) \end{bmatrix} \quad \text{and} \quad \Lambda(dt) = \begin{bmatrix} \Lambda_{00}(dt) & \Lambda_{01}(dt) \\ \Lambda_{10}(dt) & \Lambda_{11}(dt) \end{bmatrix},$$

where $\Lambda(dt)$ is the matrix measure whose matrix Stieltjes transform is $P(z,\cdot)$, where $P_{ii}(z,\cdot)$ is the Stieltjes transform of $\Lambda_{ii}(dt)$. The positive matrix measures Λ_{00} and Λ_{11} are symmetric.

One consequence of this theorem is that the probability measure ζ on $\mathbb{R}$ such that $g_\zeta(\eta) = (2N)^{-1} \operatorname{tr} P(z,\eta)$ is symmetric. In fact, it can be proved that

$$g_\zeta(\eta) = N^{-1} \operatorname{tr} P_{00}(z,\eta) = N^{-1} \operatorname{tr} P_{11}(z,\eta).$$

The first part of the next theorem claims the tightness of $(\check{\boldsymbol{\nu}}_{z,N})_N$. Later we shall need the behavior of the difference $(Q^{(N)}(z,\eta) - G^{(N)}(z,\eta))$. This is captured in the second part. The extra generality obtained by including the matrix D is not necessary for the purposes of this work in which we use only the specific choice of

$D = I$. However this will be useful to derive convergence and related properties of $(\mu_N)_N$ and $(\boldsymbol{\mu}_N)_N$ that we wish to pursue in future.

Theorem 1.4. *Suppose Assumption* 1.1 *holds. Let* $G^{(N)}(z,\eta)$ *be the solution in* $\mathfrak{S}^{2N}$, *of the equation*

$$G^{(N)}(z,\eta) = \mathcal{F}_{S^{(N)},z}(G^{(N)}(z,\eta),\eta),$$

as specified by Theorem 1.3. *Let* $\check{\boldsymbol{\nu}}_{z,N}$ *be the symmetric probability measure on* $\mathbb{R}$ *whose Stieltjes transform is*

$$g_{\check{\boldsymbol{\nu}}_{z,N}}(\eta) = \frac{1}{2N} \operatorname{tr} G^{(N)}(z,\eta).$$

Then, the sequence $(\check{\boldsymbol{\nu}}_{z,N})_N$ *is tight.*

Let $D^{(N)} \in \mathbb{C}^{2N\times 2N}$ *be an arbitrary deterministic matrix such that* $\|D^{(N)}\| = 1$. *Then,*

$$\forall \eta \in \mathbb{C}_+, \quad \frac{1}{2N} \operatorname{tr} D^{(N)}(Q^{(N)}(z,\eta) - G^{(N)}(z,\eta)) \xrightarrow[N\to\infty]{a.s.} 0. \tag{1.4}$$

Taking $D^{(N)} = I$, this theorem shows in particular that $g_{\check{\nu}_{z,N}}(\eta) - g_{\check{\boldsymbol{\nu}}_{z,N}}(\eta) \xrightarrow[N\to\infty]{\text{a.s.}} 0$. Since $(\check{\boldsymbol{\nu}}_{z,N})_N$ is tight, we have the following corollary, whose proof employs well-known facts about scalar Stieltjes transforms (see, e.g. [27; 19, Sec. 2]), and is omitted.

Corollary 1.1. *The sequence* $(\check{\nu}_{z,N})_N$ *of probability measures is tight with probability one. Furthermore, for each bounded and continuous function* $f : \mathbb{R} \to \mathbb{R}$, *we have*

$$\int f d\check{\nu}_{z,N} - \int f d\check{\boldsymbol{\nu}}_{z,N} \xrightarrow[N\to\infty]{a.s.} 0.$$

We can now conclude by characterizing the asymptotic behavior of (μ_N). Theorems 1.1 and 1.2 provide the uniform integrability condition stated in Proposition 1.1 (1), see, e.g. the derivations made in [5, Sec. 4.2]. Condition (2) in the statement of Proposition 1.1 is ensured by Theorem 1.4 via its Corollary 1.1. Therefore, thanks to Proposition 1.1, we obtain the following theorem.

Theorem 1.5. *Suppose Assumptions* 1.1–1.3 *hold. Then, there exists a tight sequence of deterministic probability measures* $(\boldsymbol{\mu}_N)$ *on* $\mathbb{C}$ *such that, for each bounded and continuous function* $f : \mathbb{C} \to \mathbb{C}$,

$$\int f \; d\mu_N - \int f \, d\boldsymbol{\mu}_N \xrightarrow[N\to\infty]{a.s.} 0.$$

The function $\log|\cdot|$ *is integrable with respect to* $\check{\boldsymbol{\nu}}_{z,N}$ *for each* $z \neq 0$, *and the measure* $\boldsymbol{\mu}_N$ *is defined by its logarithmic potential through the identity*

$$U_{\boldsymbol{\mu}_N}(z) = -\int \log|t| \; \check{\boldsymbol{\nu}}_{z,N}(dt).$$

A natural issue is if can we say anything further about the behavior of $(\boldsymbol{\mu}_N)$. In particular, its weak convergence in turn would ensure the weak convergence of (μ_N). An idea that goes back to [13] and which has been frequently used in the literature of non-Hermitian matrices, connects the behavior of $(\boldsymbol{\mu}_N)$ to that of $N^{-1}\operatorname{tr} G_{01}^{(N)}(z,\imath t)$ as $t \searrow 0$. In the present context the details need some significant efforts. Since this is outside the focus of this work, it will be pursued separately.

Incidentally, the above idea was applied in [7] to the particular case where $S^{(N)}(e^{\imath\theta}) = I_N$. The connection of this paper with [7] is provided by the following corollary to Theorem 1.5. This corollary will be proved by showing that $N^{-1}\operatorname{tr} G_{01}^{(N)}(z,\imath t)$ coincides with the analogue of this quantity that was computed in [7].

Corollary 1.2. *Assume that $S^{(N)}(e^{\imath\theta}) = I_N, N/n_N \to \gamma > 0$ as $N \to \infty$, and $L = 1$. Then the random sequence (μ_N) converges weakly in the almost sure sense to the deterministic probability measure $\boldsymbol{\mu}$ on $\mathbb{C}$ described by* [7, Theorem 2].

The remainder of the paper is organized as follows. Examples where Assumptions 1.2 and 1.3 hold, are provided in Sec. 2. Theorems 1.1 and 1.2 are proved in Sec. 3. The results related to the behavior of the singular values, i.e. Theorems 1.3 and 1.4, are proved in Sec. 4. The proof of Corollary 1.2 will be sketched in this section as well.

Notation. The indices of the elements of a vector or a matrix start from zero. Given two integers k and m, we write $[k:m] = \{k, k+1, \ldots, m-1\}$, this set being empty if $k \geq m$. We also write $[m] = [0:m]$. Assume $m > 0$. For $k \in [m]$, we denote as $e_{m,k}$ the kth canonical basis vector of $\mathbb{C}^m$. When there is no ambiguity, we write e_k for $e_{m,k}$. Given a matrix $M \in \mathbb{C}^{m\times n}$, and two sets $\mathcal{I} \subset [m]$ and $\mathcal{J} \subset [n]$, we denote as $M_{\mathcal{I},\mathcal{J}}$ the $|\mathcal{I}| \times |\mathcal{J}|$ sub-matrix of M that is obtained by retaining the rows of M whose indices belong to $\mathcal{I}$ and the columns whose indices belong to $\mathcal{J}$. We also write $M_{\cdot,\mathcal{J}} = M_{[m],\mathcal{J}}$ and $M_{\mathcal{I},\cdot} = M_{\mathcal{I},[n]}$. Given a vector $v \in \mathbb{C}^m$, we denote by $v_{\mathcal{I}}$ the $\mathbb{C}^{|\mathcal{I}|}$ sub-vector obtained by keeping the elements of v whose indices are in $\mathcal{I}$.

For any matrix M, $M > 0$ means that it is positive definite. The column span of a matrix M is denoted by $\operatorname{span}(M)$. The orthogonal projection matrix onto $\operatorname{span}(M)$ (respectively, onto the subspace orthogonal to $\operatorname{span}(M)$) is denoted by Π_M (respectively, $\Pi_M^{\perp}$). The spectral norm of a matrix and the Euclidean norm of a vector are denoted by $\|\cdot\|$. The Hilbert–Schmidt norm of an operator will be denoted $\|\cdot\|_{\mathrm{HS}}$. The notation $M \geq G$ where M and G are Hermitian matrices refers to the semi-definite ordering of such matrices.

We write $\mathbb{R}_+ = [0,\infty)$. Suppose B is a Borel set of $\mathbb{R}^d$. Then $\operatorname{Leb}(B)$ denotes its Lebesgue measure. For any $x \in \mathbb{R}^d$, $\operatorname{dist}(x,B) = \inf_{y\in B} ||x-y||$. For B in a metric space E, $\mathcal{V}_\rho^E(B)$ denotes its closed ρ-neighborhood. If the underlying space E is clear, we simply write $\mathcal{V}_\rho(B)$.

The unit-sphere of $\mathbb{C}^m$ will be denoted as $\mathbb{S}^{m-1}$. The set of vectors of $\mathbb{S}^{m-1}$ that are supported by the (index) set $\mathcal{I} \subset [m]$ will be denoted by $\mathbb{S}^{m-1}_{\mathcal{I}}$.

The probability and the expectation with respect to the law of the vector x will be denoted by $\mathbb{P}_x$ and $\mathbb{E}_x$. The centered and circularly symmetric complex Gaussian distribution with covariance matrix Σ will be denoted by $\mathcal{N}_{\mathbb{C}}(0, \Sigma)$.

2. Examples where Assumptions 1.2 and 1.3 are satisfied

In this section, we provide examples where the spectral density satisfies Assumptions 1.2 and 1.3. Consider the *moving average* $MA(\infty)$ model:

$$\boldsymbol{x}_k^{(N)} = \boldsymbol{\xi}_k^{(N)} + \sum_{\ell \geq 1} A_\ell^{(N)} \boldsymbol{\xi}_{k-\ell}^{(N)}, \tag{2.1}$$

where $\boldsymbol{\xi}^{(N)} = (\boldsymbol{\xi}_k^{(N)})_{k\in\mathbb{Z}}$ is an i.i.d. process with $\boldsymbol{\xi}_k^{(N)} \sim \mathcal{N}_{\mathbb{C}}(0, I_N)$ and $(A_\ell^{(N)})_{\ell\geq 1}$ is a sequence of deterministic matrices which satisfy the minimal requirement $\sum_{\ell\geq 1} \|A_\ell^{(N)}\| < \infty$. The spectral density for this model is

$$S^{(N)}(e^{\imath\theta}) = \left(I_N + \sum_{\ell\geq 1} e^{\imath\ell\theta} A_\ell^{(N)} \right) \left(I_N + \sum_{\ell\geq 1} e^{-\imath\ell\theta} (A_\ell^{(N)})^* \right).$$

Let us look at some particular cases:

(i) First suppose that

$$\limsup_N \sum_{\ell\geq 1} \|A_\ell^{(N)}\| < 1.$$

Then it is obvious that both Assumptions 1.2 and 1.3 are satisfied and $s_{N-1}(S^{(N)}(z))$ remains bounded away from zero when z runs through $\mathbb{T}$.

(ii) Now consider the MA(1) case, so that $A_\ell^{(N)} = 0$ for $\ell \geq 2$, and suppose that we only have $\sup_N \|A_1^{(N)}\| < \infty$. In this case, Assumptions 1.2 and 1.3, written in terms of $A_1^{(N)}$ are, respectively, (the log in the second expression is always integrable),

$$\forall \kappa \in (0,1), \exists \delta > 0, \sup_N \mathrm{Leb}\{z \in \mathbb{T} : \|(z - A_1^{(N)})^{-1}\| \geq 1/\delta\} \leq \kappa, \quad \text{and} \tag{2.2a}$$

$$\frac{1}{N} \int_0^{2\pi} \log \|(e^{\imath\theta} - A_1^{(N)})^{-1}\| \, d\theta \xrightarrow[N\to\infty]{} 0. \tag{2.2b}$$

These conditions are closely connected with the *pseudospectrum* of $A_1^{(N)}$ [39]. The Toeplitz matrices are among the matrices for which the pseudospectra are well-understood. Suppose that $A_1^{(N)} = [a_{k-\ell}^{(N)}]_{0\leq k,\ell\leq N-1}$ is a Toeplitz matrix with the so-called *symbol*

$$f_N(z) = \sum_{k=1-N}^{N-1} a_k^{(N)} z^k.$$

Then, by the Brown and Halmos theorem [8, Theorem 4.29], Assumptions 1.2 and 1.3 are, respectively, satisfied if

$$\forall \kappa \in (0,1), \quad \exists \delta > 0, \quad \limsup_N \mathrm{Leb}\{z \in \mathbb{T} : \mathrm{dist}(e^{\imath\theta}, \mathrm{conv}\, f_N(\mathbb{T})) \le \delta\} \le \kappa,$$

$$\text{and} \quad \frac{1}{N}\int_0^{2\pi} \log \mathrm{dist}(e^{\imath\theta}, \mathrm{conv}\, f_N(\mathbb{T}))\, d\theta \xrightarrow[N\to\infty]{} 0,$$

where conv denotes the convex hull of a set.

(iii) Another particular case of the model in (2.1) is the following. Assume that $N = MK$ where M and K are two positive integers. Assume that the matrices $A_\ell^{(N)}$ are block-diagonal matrices of the form

$$A_\ell^{(N)} = I_M \otimes B_\ell^{(K)},$$

where $B_\ell^{(K)} \in \mathbb{C}^{K\times K}$. We also assume that $\mathfrak{b} = \sup_K \sum_{\ell\ge 1} \|B_\ell^{(K)}\| < \infty$. The process $\boldsymbol{\xi}^{(N)}$ consists of M i.i.d. streams of stationary processes each of dimension K. Let us show that if $M \to \infty$, then Assumption 1.3 is satisfied for this process. Further if K is upper bounded, then Assumption 1.2 is satisfied.

The spectral density of each stream is $\Sigma^{(K)}(e^{\imath\theta}) = C^{(K)}(e^{\imath\theta})C^{(K)}(e^{\imath\theta})^*$, with $C^{(K)}(z) = I_K + \sum_{\ell\ge 1} z^\ell B_\ell^{(K)}$ on the unit-disk. It is well known that (see, e.g. [30, Theorem 6.1]),

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi} |\log|\det C^{(K)}(e^{\imath\theta})|^{\frac{1}{K}}|d\theta \\
&\quad \le \frac{1}{2\pi}\int_0^{2\pi} (\det \Sigma^{(K)}(e^{\imath\theta}))^{\frac{1}{K}} d\theta - \log|\det C^{(K)}(0)|^{\frac{1}{K}} \\
&\quad = \frac{1}{2\pi}\int_0^{2\pi} (\det \Sigma^{(K)}(e^{\imath\theta}))^{\frac{1}{K}} d\theta.
\end{aligned}$$

By Hadamard's inequality, the right-hand side is bounded by $(1+\mathfrak{b})^2$. Thus,

$$\begin{aligned}
|\log s_{K-1}(\Sigma^{(K)}(e^{\imath\theta}))| &\le \left|\sum_{\ell=0}^{K-1} \log s_\ell(\Sigma^{(K)}(e^{\imath\theta}))\mathbb{1}_{s_\ell(\Sigma^{(K)}(e^{\imath\theta}))\le 1}\right| + \log(1+\mathfrak{b})^2 \\
&\le \left|\sum_{\ell=0}^{K-1} \log s_\ell(\Sigma^{(K)}(e^{\imath\theta}))\right| + (K+1)\log(1+\mathfrak{b})^2 \\
&= 2|\log|\det C^{(K)}(e^{\imath\theta})|\,| + (K+1)\log(1+\mathfrak{b})^2,
\end{aligned}$$

and we deduce from the last display that

$$\frac{1}{2\pi}\int_0^{2\pi} |\log s_{K-1}(\Sigma^{(K)}(e^{\imath\theta}))|\, d\theta \le 2K(1+\mathfrak{b})^2 + (K+1)\log(1+\mathfrak{b})^2.$$

Since $S^{(N)}(e^{\imath\theta}) = I_M \otimes \Sigma^{(K)}(e^{\imath\theta})$, Assumption 1.3 is satisfied when $M \to \infty$. Furthermore, given a small $\delta > 0$, we have

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi} \mathbb{1}_{s_{K-1}(\Sigma^{(K)}(e^{\imath\theta}))\le\delta}\, d\theta &\le \frac{1}{|\log\delta|}\frac{1}{2\pi}\int_0^{2\pi} |\log s_{K-1}(\Sigma^{(K)}(e^{\imath\theta}))|\, d\theta \\ &\le \frac{2K(1+\mathfrak{b})^2 + (K+1)\log(1+\mathfrak{b})^2}{|\log\delta|}.\end{aligned}$$

This shows that Assumption 1.2 is satisfied if K is bounded.

3. Small Singular Values: Proofs of Theorems 1.1 and 1.2

3.1. *Outline of the proofs*

In the sequel, we shall most often omit the dependency on N in the notation for simplicity. Writing $n = n_N$, define the matrix

$$X = [x_0 \quad \cdots \quad x_{n-1}] = n^{-1/2}[\boldsymbol{x}_0 \quad \cdots \quad \boldsymbol{x}_{n-1}] \in \mathbb{C}^{N\times n}$$

with $x_k \in \mathbb{C}^N$ being the kth column of X, and consider the $n \times n$ circulant matrix

$$J = \begin{bmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}.$$

Then, the sample autocovariance matrix $\widehat{R}_L$ can be rewritten as $\widehat{R}_L = XJ^LX^*$. Let the so-called $n \times n$ *Fourier matrix* F be defined as

$$\mathsf{F} = \frac{1}{\sqrt{n}}[\exp(2\imath\pi k\ell/n)]_{k,\ell=0}^{n-1}. \tag{3.1}$$

Then $J = \mathsf{F}\Omega\mathsf{F}^*$, where

$$\Omega = \operatorname{diag}(\omega^k)_{k=0}^{n-1} \quad \text{and} \quad \omega = \exp(-2\imath\pi/n).$$

Furthermore, for $k \in [n]$, let

$$y_k = \frac{1}{\sqrt{n}}\sum_{\ell=0}^{n-1} e^{2\imath\pi k\ell} x_\ell$$

be the *discrete Fourier transform* of the finite sequence $(x_0, \dots, x_{n-1})$, and define the $N \times n$ matrix

$$Y = [y_0 \quad \cdots \quad y_{n-1}] = X\mathsf{F} \in \mathbb{C}^{N\times n}.$$

Then, we obviously have

$$\widehat{R}_L = Y\Omega^L Y^*.$$

Note that the columns y_k of Y are "almost" independent, since they are the discrete Fourier transforms applied to a time window of a Gaussian stationary process. This is why we shall heavily rely on the above expression of $\widehat{R}_L$. Let us elaborate on this point before entering the core of the proof. Write

$$y_k = \tilde{y}_k + \check{y}_k, \quad \check{y}_k = \mathbb{E}[y_k | Y_k],$$

the latter being the conditional expectation with respect to the σ-field $\sigma(Y_k)$ generated by the elements of the matrix $Y_k = [y_0 \cdots y_{k-1} \ y_{k+1} \cdots y_{n-1}] \in \mathbb{C}^{N\times(n-1)}$. Write $S_k = S(e^{2\imath\pi k/n})$ for brevity, where we recall that $S(e^{\imath\theta})$ is the matrix spectral density. Due to the Gaussianity, $\tilde{y}_k$ and Y_k will be independent, and we shall prove that

$$\mathbb{E}[\tilde{y}_k\tilde{y}_k^*] \simeq \mathbb{E}[y_k y_k^*] \simeq S_k.$$

For this, as is frequently done in estimation theory for stationary processes, we shall make use of the orthogonal matrix polynomial theory with respect to the matrix measure $(2\pi)^{-1}S(e^{\imath\theta})d\theta$ on $\mathbb{T}$. Assumption 1.3, which is reminiscent of the notion of *regular measures* found in this literature [35, 36], will play a major role in our analysis that will be presented in Sec. 3.2.

We now consider the proof of Theorem 1.1, starting with a well-known linearization trick. Write

$$H = \begin{bmatrix} \Omega^{-L} & Y^* \\ Y & z \end{bmatrix} \in \mathbb{C}^{(N+n)\times(N+n)}.$$

By using the well-known formula for the inverse of a partitioned matrix that involves the Schur complements (see [20, §0.7.3]), it follows that $\|(Y\Omega^L Y^* - z)^{-1}\| \le \|H^{-1}\|$. Hence

$$s_{N+n-1}(H) \le s_{n-1}(Y\Omega^L Y^* - z), \tag{3.2}$$

and the problem then is to control $s_{N+n-1}(H)$. A similar problem, considered in [40], was that of the smallest singular value of a symmetric random matrix with iid elements above the diagonal, see also [25]. Even though our matrix is quite different from theirs, we borrow many of their ideas as well of their predecessors, such as [31].

First, it can be shown that for $C > 0$ large enough, $\mathbb{P}[\|Y\| \ge C]$ is exponentially small (Lemma 3.7) by using some standard Gaussian calculations. Then, as in these articles, the control over $s_{N+n-1}(H)$ will be provided on the event $[\|Y\| \le C]$. More specifically, we will prove that, for some constant $c > 0$,

$$\mathbb{P}[[s_{N+n-1}(H) \le N^{-3/2}t] \cap [\|Y\| \le C]] \le \varepsilon t + \exp(-c\varepsilon^2 n).$$

The smallest singular value of H can be obtained from the variational characterization

$$s_{N+n-1}(H) = \inf_{u \in \mathbb{S}^{N+n-1}} \|Hu\|.$$

A well-established method to control the smallest singular value of a random matrix is to study the action of this matrix on the so-called *compressible* and *incompressible* vectors [23, 31]. Let $\theta, \rho \in (0,1)$ be fixed. A vector in $\mathbb{S}^{m-1}$ is called θ-*sparse* if it does not have more than $\lfloor \theta m \rfloor$ non-zero elements. Given $\theta, \rho \in (0,1)$, we define the set of (θ, ρ)-compressible vectors as

$$\operatorname{comp}(\theta, \rho) = \mathbb{S}^{m-1} \cap \bigcup_{\substack{\mathcal{I} \subset [m] \\ |\mathcal{I}| = \lfloor \theta m \rfloor}} \mathcal{V}_\rho^{\mathbb{C}^m}(\mathbb{S}_{\mathcal{I}}^{m-1}).$$

In other words, this is the set of all unit vectors at a distance less or equal to ρ from the set of the θ-sparse unit vectors. The set $\operatorname{incomp}(\theta, \rho)$ of (θ, ρ)-incompressible vectors is the complementary set $\mathbb{S}^{m-1} \setminus \operatorname{comp}(\theta, \rho)$.

Getting back to the variational characterization, and writing $u = [v^\mathsf{T}\ w^\mathsf{T}]^\mathsf{T}$ with $v \in \mathbb{C}^n$, we (roughly) define the set

$$\mathcal{S} = \{u \in \mathbb{S}^{N+n-1} : \|v\| \le \text{ a constant or } v/\|v\| \text{ is compressible}\},$$

and we write

$$s_{N+n-1}(H) = \inf_{u \in \mathcal{S}} \|Hu\| \wedge \inf_{u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}} \|Hu\|. \tag{3.3}$$

For the first infimum, we focus on the component v of the vector u because v impacts the first n columns of H which are nearly independent. Relying on the decomposition $y_k = \tilde{y}_k + \check{y}_k$, we first show that $\mathbb{P}[\|Hu\| \le c]$ is exponentially small when $v/\|v\|$ is a sparse vector, and then we complete the analysis by an ε-net argument.

The second infimum requires other arguments. Let h_k be the kth column of H, and so $H_k = [h_0 \cdots h_{k-1}\ h_{k+1} \cdots h_{N+n-1}] \in \mathbb{C}^{(N+n)\times(N+n-1)}$. Following a by now well-known idea of [31, Lemma 3.5], the infimum over the incompressible vectors can be controlled by managing the distances $\operatorname{dist}(h_k, H_k)$ between h_k and $\operatorname{span}(H_k)$. Given $k \in [n]$, let $\Omega_k = \operatorname{diag}(\omega^\ell)_{\ell \ne k} \in \mathbb{C}^{(n-1)\times(n-1)}$. Let

$$G_k = \begin{bmatrix} \Omega_k^{-L} & Y_k^* \\ Y_k & z \end{bmatrix} \in \mathbb{C}^{(N+n-1)\times(N+n-1)},$$

and partition G_k^{-1} as

$$G_k^{-1} = \begin{bmatrix} E_k & F_k \\ P_k & D_k \end{bmatrix}, \quad E_k \in \mathbb{C}^{(n-1)\times(n-1)},\ D_k \in \mathbb{C}^{N\times N}.$$

Then, after some algebra, we get the following equation (similar to what was obtained in [40]):

$$\operatorname{dist}(h_k, H_k) = \frac{|\omega^{-kL} - y_k^* D_k y_k|}{\sqrt{1 + \|y_k^* P_k\|^2 + \|y_k^* D_k\|^2}}.$$

Let us assume temporarily that the $\check{y}_k$ are equal to zero. Restricting ourselves to the indices k for which $s_{N-1}(S_k) \sim 1$ (which is allowed), we can show that for these indices

$$\sqrt{1 + \|y_k^* P_k\|^2 + \|y_k^* D_k\|^2} \sim \|S_k^{1/2} D_k S_k^{1/2}\|_{\mathrm{HS}}. \tag{3.4}$$

Consequently, the control of $\operatorname{dist}(h_k, H_k)$ can be reduced to the control of the probability (with respect to the law of y_k) that, $y_k^* D_k y_k / \|S_k^{1/2} D_k S_k^{1/2}\|_{\mathrm{HS}}$ lies in a small ball. Due to the Gaussian nature of y_k, this task is easy and leads to a rate of $N^{-1}t$ in the statement of Theorem 1.1.

Now consider the term that involves $\check{y}_k$. Even though $\mathbb{E}\|\check{y}_k\|^2$ is small, its interdependence with (P_k, D_k) prevents us from obtaining an approximation similar to (3.4). Its presence in fact is responsible for the $N^{-3/2}t$ term (instead of $N^{-1}t$) in the statement of Theorem 1.1.

The proof of Theorem 1.2 is laid out on a similar canvas. Let $k \in [\lfloor N^\beta \rfloor, \lfloor N/2 \rfloor]$. In Lemma 3.12, we show that the smallest singular value of $H_{\cdot,[N+n-k]}$, is a lower bound for $s_{N-k-1}(Y\Omega^L Y^* - z)$. As in the proof of Theorem 1.1, this value can be characterized through the action of $H_{\cdot,[N+n-k]}$ on the compressible and incompressible vectors. The former can be handled exactly as for H. The latter can also be reduced to a distance problem, and indeed this term is easier to tackle than earlier, thanks to the rectangular nature of $H_{\cdot,[N+n-k]}$.

3.2. *Statistical analysis of the process* (y_k)

Recall the decomposition $y_k = \tilde{y}_k + \check{y}_k$, with $\check{y}_k = \mathbb{E}[y_k \mid Y_k]$. We now derive approximations for $n\mathbb{E}[y_k y_k^*]$ and $n\mathbb{E}[\tilde{y}_k \tilde{y}_k^*]$. Let $\mathcal{R}$ denote the $Nn \times Nn$ block-Toeplitz matrix

$$\mathcal{R} = \mathbb{E} \operatorname{vec} X (\operatorname{vec} X)^* = \frac{1}{n} \begin{bmatrix} R_0 & R_{-1} & & R_{-n+1} \\ R_1 & \ddots & \ddots & \\ & \ddots & \ddots & R_{-1} \\ R_{n-1} & & R_1 & R_0 \end{bmatrix} = \frac{1}{n}[R_{k-\ell}]_{k,\ell=0}^{n-1}. \tag{3.5}$$

We also write the Fourier matrix F as

$$\mathsf{F} = \begin{bmatrix} \mathsf{f}_0 \\ \vdots \\ \mathsf{f}_{n-1} \end{bmatrix},$$

with $\mathsf{f}_k \in \mathbb{C}^{1\times n}$ being the kth row of F, and we define the function

$$\mathsf{a}(e^{\imath\theta}) = \begin{bmatrix} 1 & e^{-\imath\theta} & \cdots & e^{-\imath(n-1)\theta} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^n.$$

From Eq. (1.2), we have

$$\mathcal{R} = \frac{1}{2\pi n}\int_0^{2\pi} (\mathsf{a}(e^{\imath\theta}) \otimes I_N) S(e^{\imath\theta})(\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\, d\theta.$$

Observe that $\mathcal{R}$ is invertible, since the spectral density $S(e^{\imath\theta})$ is nontrivial, as is well known. Indeed, the equation $\mathcal{R}u = 0$ with $u = [u_0^{\mathsf{T}}, \ldots, u_{n-1}^{\mathsf{T}}]^{\mathsf{T}}$ being a non-zero vector with $u_\ell \in \mathbb{C}^N$ would lead to the identity $\int_0^{2\pi} p(e^{\imath\theta})^* S(e^{\imath\theta}) p(e^{\imath\theta})\, d\theta = 0$ where $p(z) = \sum_{\ell=0}^{n-1} z^\ell u_\ell$ is a non-zero polynomial. We also have that $\|n\mathcal{R}\| \le \boldsymbol{M}$ from Assumption 1.1(ii) by a similar argument.

Lemma 3.1. *Let Assumption* 1.1 *hold true. Then,*

$$\max_{k\in[n]} \|n\mathbb{E}[y_k y_k^*] - S^{(N)}(e^{2\imath\pi k/n})\| \xrightarrow[N\to\infty]{0}.$$

Proof. Let $k \in [n]$. Since $\mathsf{F} = \mathsf{F}^{\mathsf{T}}$, we have $y_k = X\mathsf{F}e_k = X\mathsf{f}_k^{\mathsf{T}} = (\mathsf{f}_k \otimes I_N)\operatorname{vec} X$. Hence,

$$\begin{aligned} n\mathbb{E}y_k y_k^* &= n(\mathsf{f}_k \otimes I_N)\mathcal{R}(\mathsf{f}_k^* \otimes I_N) \\ &= \frac{1}{2\pi}\int_0^{2\pi} ((\mathsf{f}_k\mathsf{a}(e^{\imath\theta})) \otimes I_N) S(e^{\imath\theta})((\mathsf{a}(e^{\imath\theta})^*\mathsf{f}_k^*) \otimes I_N)\, d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} |\mathsf{f}_k\mathsf{a}(e^{\imath\theta})|^2 S(e^{\imath\theta})\, d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} F_{n-1}(\theta - 2\pi k/n) S(e^{\imath\theta})\, d\theta, \end{aligned} \tag{3.6}$$

where F_{n-1} is the Fejér kernel, defined as

$$F_{n-1}(\theta) = \frac{1}{n}\left|\sum_{\ell=0}^{n-1} e^{\imath\ell\theta}\right|^2 = \frac{1}{n}\frac{\sin(n\theta/2)^2}{\sin(\theta/2)^2}. \tag{3.7}$$

This kernel satisfies

$$\frac{1}{2\pi}\int_0^{2\pi} F_{n-1}(\theta)\, d\theta = 1, \quad \text{and} \quad F_{n-1}(\theta) \le n\left(\frac{\pi^2}{n^2\theta^2} \wedge \frac{\pi^2}{4}\right) \quad \text{for } \theta \in [-\pi, \pi],$$

see, e.g. [35, p. 136]. For an arbitrarily small number $\eta > 0$, we know by Assumption 1.1(i) that there exists $\delta > 0$ independent of k and N such that $\|S(e^{\imath\theta}) - S(e^{2\imath\pi k/n})\| \le \eta$ if $|\theta - 2\pi k/n| \le \delta$. Splitting the integral that appears on the right side of (3.6) as $\int_{|\theta-2\pi k/n|\le\delta} + \int_{|\theta-2\pi k/n|>\delta}$ and using the properties of the Fejér kernel provided above, along with Assumption 1.1(ii), we obtain the result of the lemma after some routine derivations. □

The handling of $n\mathbb{E}[\tilde{y}_k\tilde{y}_k^*]$ is more involved.

Proposition 3.1. *Let Assumptions* 1.1 *and* 1.3 *hold true. Then,*

$$\forall\delta > 0, \max_{k\in[n]} \|n\mathbb{E}[\tilde{y}_k\tilde{y}_k^*] - S^{(N)}(e^{2\imath\pi k/n})\| \mathbb{1}_{S^{(N)}(e^{2\imath\pi k/n})\ge\delta I_N} \xrightarrow[N\to\infty]{} 0.$$

Proof of Proposition 3.1.

The first step is to provide a single letter expression for $\mathbb{E}[\tilde{y}_k\tilde{y}_k^*]$. This expression is reminiscent of the formula to calculate the partial covariance of real-valued random variables.

Lemma 3.2. *For any given* $k \in [n], \mathbb{E}[\tilde{y}_k\tilde{y}_k^*] = ((\mathsf{f}_k \otimes I)\mathcal{R}^{-1}(\mathsf{f}_k^* \otimes I))^{-1}$.

Proof. The covariance matrix $\mathbb{E}[\tilde{y}_k\tilde{y}_k^*]$ coincides with the conditional covariance matrix $\mathrm{cov}(y_k \,|\, Y_k)$ of y_k with respect to $\sigma(Y_k)$. Let Z_0 and Z_1 be two random square integrable vectors with arbitrary dimensions. Writing

$$\mathbb{E}\begin{bmatrix} Z_0 \\ Z_1 \end{bmatrix} [Z_0^* \quad Z_1^*] = \begin{bmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{bmatrix}$$

with the block dimensions at the right-hand side being compatible with the dimensions of Z_0 and Z_1, it is well known that $\mathrm{cov}(Z_0 \mid Z_1) = \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}$ when Σ_{11} is invertible. Observe that

$$\mathbb{E}\,\mathrm{vec}\,Y(\mathrm{vec}\,Y)^* = \mathbb{E}\,\mathrm{vec}(X\mathsf{F})(\mathrm{vec}(X\mathsf{F}))^* = (\mathsf{F} \otimes I_N)\mathcal{R}(\mathsf{F}^* \otimes I_N).$$

Let $\mathsf{F}_k \in \mathbb{C}^{(n-1)\times n}$ be the matrix that remains after taking the row f_k out of F, and let $\mathcal{R}^{1/2}$ be the Hermitian square root of $\mathcal{R}$. Then

$$\begin{aligned}\mathbb{E}[\tilde{y}_k\tilde{y}_k^*] &= (\mathsf{f}_k \otimes I_N)\mathcal{R}^{1/2}(I_{(n-1)N} - \mathcal{R}^{1/2}(\mathsf{F}_k^* \otimes I_N)((\mathsf{F}_k \otimes I_N) \\ &\quad \times \mathcal{R}(\mathsf{F}_k^* \otimes I_N))^{-1}(\mathsf{F}_k \otimes I_N)\mathcal{R}^{1/2})\mathcal{R}^{1/2}(\mathsf{f}_k^* \otimes I_N) \\ &= (\mathsf{f}_k \otimes I_N)\mathcal{R}^{1/2}\Pi^{\perp}_{\mathcal{R}^{1/2}(\mathsf{F}_k^*\otimes I_N)}\mathcal{R}^{1/2}(\mathsf{f}_k^* \otimes I_N).\end{aligned}$$

Let $\mathcal{C}$ be a positively oriented circle with center zero and a radius small enough so that $\mathcal{R}$ has no eigenvalue in the closed disk delineated by $\mathcal{C}$. This implies that the positive definite matrix $(\mathsf{F}_k \otimes I)\mathcal{R}(\mathsf{F}_k^* \otimes I)$ does not have eigenvalues in this closed disk either. With this choice, the projection $\Pi^{\perp}_{\mathcal{R}^{1/2}(\mathsf{F}_k^*\otimes I_N)}$ can be expressed as a contour integral, which leads us to write

$$\begin{aligned}\mathbb{E}[\tilde{y}_k\tilde{y}_k^*] &= \frac{-1}{2\imath\pi}\oint_{\mathcal{C}}(\mathsf{f}_k\otimes I_N)\mathcal{R}^{1/2}(\mathcal{R}^{1/2}(\mathsf{F}_k^*\otimes I_N)(\mathsf{F}_k\otimes I_N)\mathcal{R}^{1/2}-z)^{-1}\\ &\quad\times\mathcal{R}^{1/2}(\mathsf{f}_k^*\otimes I_N)\,dz\\ &= \frac{-1}{2\imath\pi}\oint_{\mathcal{C}}(\mathsf{f}_k\otimes I_N)\mathcal{R}^{1/2}(\mathcal{R}-\mathcal{R}^{1/2}(\mathsf{f}_k^*\otimes I_N)(\mathsf{f}_k\otimes I_N)\mathcal{R}^{1/2}-z)^{-1}\\ &\quad\times\mathcal{R}^{1/2}(\mathsf{f}_k^*\otimes I_N)\,dz.\end{aligned}$$

Let $\mathcal{Q}(z)=(\mathcal{R}-z)^{-1}$ when $z\in\mathbb{C}$ is not an eigenvalue of $\mathcal{R}$, and let

$$\Sigma(z)=(\mathsf{f}_k\otimes I_N)\mathcal{R}^{1/2}\mathcal{Q}(z)\mathcal{R}^{1/2}(\mathsf{f}_k^*\otimes I_N)=I_N+z(\mathsf{f}_k\otimes I_N)\mathcal{Q}(z)(\mathsf{f}_k^*\otimes I_N).$$

Now, using the Sherman–Morrison–Woodbury formula, we get

$$\begin{aligned}\mathbb{E}[\tilde{y}_k\tilde{y}_k^*] &= \frac{-1}{2\imath\pi}\oint_{\mathcal{C}}(\Sigma(z)+\Sigma(z)(I_N-\Sigma(z))^{-1}\Sigma(z))dz\\ &= \frac{1}{2\imath\pi}\oint_{\mathcal{C}}z^{-1}((\mathsf{f}_k\otimes I_N)\mathcal{Q}(z)(\mathsf{f}_k^*\otimes I_N))^{-1}dz\\ &= ((\mathsf{f}_k\otimes I_N)\mathcal{R}^{-1}(\mathsf{f}_k^*\otimes I_N))^{-1},\end{aligned}$$

where the last equality is obtained by expressing $\mathcal{Q}(z)$ in terms of the spectral decomposition of $\mathcal{R}$ and by using that $(2\imath\pi)^{-1}\oint_{\mathcal{C}}z^{-1}(\lambda-z)^{-1}dz=\lambda^{-1}$ when λ is outside the closed disk enclosed by $\mathcal{C}$. □

We now reinterpret the expression provided by this lemma in the light of the matrix orthogonal polynomial theory. Let us quickly review some of the basic results of this theory. For the proofs of these results, the reader may consult [12].

For each N, consider the $N\times N$ matrix-valued measure ϱ_N defined on $\mathbb{T}$ as

$$d\varrho_N(\theta)=\frac{1}{2\pi}S^{(N)}(e^{\imath\theta})d\theta.$$

Given two $\mathbb{C}^{N\times N}$ matrix-valued polynomials F,G, we define the $N\times N$ matrix sesquilinear function $\langle\!\langle F,G\rangle\!\rangle_{\varrho_N}$ with respect to ϱ_N as

$$\langle\!\langle F,G\rangle\!\rangle_{\varrho_N}=\int_0^{2\pi}F(e^{\imath\theta})^*d\varrho_N(\theta)G(e^{\imath\theta}).$$

We now define the sequence $(\Phi_\ell^{\varrho_N})_{\ell=0,1,2,\ldots}$ of matrix orthogonal polynomials with respect to $\langle\!\langle\cdot,\cdot\rangle\!\rangle_{\varrho_N}$. The following conditions, which are analogous to a Gram–Schmidt orthogonalization, are enough to define this sequence:

- $\Phi_\ell^{\varrho_N}(z)$ is a $\mathbb{C}^{N\times N}$-valued monic matrix polynomial of degree ℓ, and is thus written as $\Phi_\ell^{\varrho_N}(z)=z^\ell I_N+$ lower order coefficients.
- For each $\ell\geq 1$, the relation $\langle\!\langle\Phi_\ell^{\varrho_N},z^k\rangle\!\rangle_{\varrho_N}=0$ holds for $k=0,1,\ldots,\ell-1$.

Since $S^{(N)}$ is nontrivial, the matrix $\langle\!\langle \Phi_\ell^{\varrho_N}, \Phi_\ell^{\varrho_N} \rangle\!\rangle_{\varrho_N}$ is positive definite for each $\ell \in \mathbb{N}$. Thus, one can define the sequence $(\varphi_\ell^{\varrho_N})_{\ell\in\mathbb{N}}$ of the normalized versions of the polynomials $\Phi_\ell^{\varrho_N}$ as

$$\varphi_\ell^{\varrho_N}(z) = \Phi_\ell^{\varrho_N}(z)\kappa_\ell^{\varrho_N},$$

where the matrices $(\kappa_\ell^{\varrho_N})_{\ell\in\mathbb{N}}$ are chosen in such a way that $\langle\!\langle \varphi_\ell^{\varrho_N}, \varphi_k^{\varrho_N} \rangle\!\rangle_{\varrho_N} = \mathbb{1}_{k=\ell} I_N$. This identity determines the matrices $\kappa_\ell^{\varrho_N}$ up to a right multiplication by a unitary matrix, the convenient choice of which is specified in [12, § 3.2] and is not relevant here.

The *Christoffel–Darboux* (*CD*) kernel of order ℓ for the measure ϱ_N is defined for $z, u \in \mathbb{C}$ as

$$K_\ell^{\varrho_N}(z,u) = \sum_{k=0}^{\ell} \varphi_k^{\varrho_N}(z)\varphi_k^{\varrho_N}(u)^*. \tag{3.8}$$

This function is a reproducing kernel, in the sense that for each matrix polynomial $P(z)$ with degree less or equal to ℓ, the following equation holds and defines the kernel $K_\ell^{\varrho_N}$:

$$\int_0^{2\pi} K_\ell^{\varrho_N}(z, e^{\imath\theta}) d\varrho_N(\theta) P(e^{\imath\theta}) = P(z).$$

The CD kernel $K_\ell^{\varrho_N}(e^{\imath\theta}, e^{\imath\theta})$ is invertible, and it satisfies the following variational formula: for each $\theta \in [0, 2\pi)$ and for each $N \times N$ matrix polynomial $P(z)$ of degree at most ℓ, with

$$P(e^{\imath\theta}) = I_N,$$

it holds that

$$\langle\!\langle P, P \rangle\!\rangle_{\varrho_N} \geq K_\ell^{\varrho_N}(e^{\imath\theta}, e^{\imath\theta})^{-1}, \tag{3.9}$$

with equality if and only if $P(z) = P_\ell^{\varrho_N,\theta}(z)$, with

$$P_\ell^{\varrho_N,\theta}(z) = K_\ell^{\varrho_N}(z, e^{\imath\theta}) K_\ell^{\varrho_N}(e^{\imath\theta}, e^{\imath\theta})^{-1}. \tag{3.10}$$

Getting back to our problem, it turns out that our covariance matrix $\mathbb{E}[\tilde{y}_k \tilde{y}_k^*]$ can be very simply expressed in terms of a CD kernel:

Lemma 3.3. $\mathbb{E}[\tilde{y}_k \tilde{y}_k^*] = K_{n-1}^{\varrho_N}(e^{2\imath\pi k/n}, e^{2\imath\pi k/n})^{-1}$.

Proof. With Lemma 3.2 at hand, this result is an instance of a well-known result, see, e.g. [21] for the scalar measure case. We reproduce its proof for completeness. Let $P(z) = \sum_{k=0}^{n-1} z^k P_k$ be an arbitrary $N \times N$ matrix polynomial of degree at most $n-1$. Stack the coefficients of P in the matrix $\mathsf{P} = [P_0^\mathsf{T} \quad \cdots \quad P_{n-1}^T]^\mathsf{T}$, so that $P(e^{\imath\theta}) = (\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\mathsf{P}$. We have

$$\begin{aligned}
&\frac{1}{2\pi n}\int_0^{2\pi} (\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\mathcal{R}^{-1}(\mathsf{a}(e^{\imath\psi}) \otimes I_N)S(e^{\imath\psi})P(e^{\imath\psi})\, d\psi \\
&\quad = (\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\mathcal{R}^{-1}\left(\frac{1}{2\pi n}\int_0^{2\pi}(\mathsf{a}(e^{\imath\psi}) \otimes I_N)S(e^{\imath\psi})(\mathsf{a}(e^{\imath\psi})^* \otimes I_N)\, d\psi\right)\mathsf{P} \\
&\quad = (\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\mathcal{R}^{-1}\mathcal{R}\mathsf{P} \\
&\quad = P(e^{\imath\theta}).
\end{aligned}$$

From the uniqueness of the CD kernel as a reproducing kernel, we thus obtain that

$$n^{-1}(\mathsf{a}(e^{\imath\theta})^* \otimes I_N)\mathcal{R}^{-1}(\mathsf{a}(e^{\imath\psi}) \otimes I_N) = K_{n-1}^{\varrho_N}(e^{\imath\theta}, e^{\imath\psi}).$$

The result follows upon observing that $\mathsf{f}_k = n^{-1/2}\mathsf{a}(e^{2\pi\imath k/n})^*$. □

Now, Proposition 3.1 will be obtained by studying the asymptotics of the CD kernel, a subject with a rich history in the literature of orthogonal polynomials. These asymptotics are well understood in the case where the underlying measure has a so-called regularity property à la Stahl and Totik [36] (see [35] for an extensive treatment of these ideas). We adapt the approach detailed in [35, §2.15 and §2.16], to the matrix measure case. This will consist of two main steps:

- We show that the polynomial $P_{n-1}^{\varrho_N;\theta}$, for which the variational inequality (3.9) for $\ell = n-1$ is an equality, satisfies

$$\max_{\theta,\psi\in[0,2\pi)} \|P_{n-1}^{\varrho_N;\theta}(e^{\imath\psi})\| \le e^{n\varepsilon_N} \quad \text{with } 0 \le \varepsilon_N \to 0. \tag{3.11}$$

  This is where Assumption 1.3 comes into play.
- With the help of the variational characterization of the CD kernel, and by making use of the previous result coupled with a matrix version of the so-called Nevai's trial polynomial technique [24, 35], we show that

$$\begin{aligned}
&\forall\varepsilon,\delta > 0, \exists N_0 \in \mathbb{N}, \forall N > N_0, \forall\theta_0 \in [0,2\pi), S^{(N)}(e^{\imath\theta_0}) \ge \delta I_N \\
&\quad \Rightarrow nK_{n-1}^{\varrho_N}(e^{\imath\theta_0}, e^{\imath\theta_0})^{-1} \ge (1-\varepsilon)S^{(N)}(e^{\imath\theta_0}).
\end{aligned} \tag{3.12}$$

In the following derivations, the sequence $\varepsilon_N \to 0$ can change from one display to another.

For a given $N \in \mathbb{N}\backslash\{0\}$, consider the scalar measure ζ_N on $\mathbb{T}$ defined as

$$d\zeta_N(\theta) = \frac{1}{2\pi}s_{N-1}(S^{(N)}(e^{\imath\theta}))\, d\theta.$$

Consider the sesquilinear form on the scalar polynomials

$$\langle f, g\rangle_{\zeta_N} = \int_0^{2\pi} f(e^{\imath\theta})^* g(e^{\imath\theta})\, d\zeta_N(e^{\imath\theta}),$$

and denote as $(b_\ell^{\zeta_N})_{\ell\in\mathbb{N}}$ the sequence of monic orthogonal polynomials with respect to $\langle\cdot,\cdot\rangle_{\zeta_N}$, which are the analogues of the $\Phi_\ell^{\varrho_N}$ above. Let $\tau_\ell^{\zeta_N} = \langle b_\ell^{\zeta_N}, b_\ell^{\zeta_N}\rangle_{\zeta_N}$, and consider the orthonormal polynomials $(\beta_\ell^{\zeta_N})_{\ell\in\mathbb{N}}$ defined as $\beta_\ell^{\zeta_N}(z) = b_\ell^{\zeta_N}(z)/\sqrt{\tau_\ell^{\zeta_N}}$ (these are the analogues of the $\varphi_\ell^{\varrho_N}$). To establish (3.11), we first show that

$$\max_{\ell\in[n]}\max_{\theta\in[0,2\pi)} |\beta_\ell^{\zeta_N}(e^{\imath\theta})| \leq e^{n\varepsilon_N} \quad \text{with } \varepsilon_N \geq 0,\ \varepsilon_N \to 0. \tag{3.13}$$

Write

$$\sigma_N^2 = \tau_0^{\zeta_N} = \int_0^{2\pi} d\zeta_N(e^{\imath\theta}) = \frac{1}{2\pi}\int_0^{2\pi} s_{N-1}(S^{(N)}(e^{\imath\theta}))d\theta.$$

The sequence $(\tau_\ell^{\zeta_N})_\ell$ satisfies the recursion $\tau_{\ell+1}^{\zeta_N} = (1-|\alpha_\ell|^2)\tau_\ell^{\zeta_N}$, thus, $\tau_{\ell+1}^{\zeta_N} = \sigma_N^2\prod_{k=0}^{\ell}(1-|\alpha_k|^2)$, where $(\alpha_\ell)_{\ell=0}^\infty$ is the sequence of the so-called Verblunsky's coefficients associated to the measure ζ_N [35, §1.7 and §1.8]. Moreover, by the celebrated Szegö's theorem [35, Theorem 1.8.6], the non-increasing sequence $(\tau_\ell^{\zeta_N})_\ell$ converges as $\ell\to\infty$ towards $\exp((2\pi)^{-1}\int_0^{2\pi}\log s_{N-1}(S^{(N)}(e^{\imath\theta}))\, d\theta)$.

From this, we first deduce with the help of Assumption 1.1(ii) that there is a constant $C > 0$ such that

$$C \geq \sigma_N^2 \geq \tau_{n-1}^{\zeta_N} \geq \exp\left(\frac{1}{2\pi}\int_0^{2\pi}\log s_{N-1}(S^{(N)}(e^{\imath\theta}))d\theta\right).$$

Furthermore, by inspecting the proof of [35, Theorem 2.15.3], in particular, Inequality (2.15.21) of that proof, and by using in addition Inequalities (2.15.13) then (2.15.15) of [35], we get that for $|z| > 1$,

$$\begin{aligned}
\max_{\ell\in[n]}\frac{|b_\ell^{\zeta_N}(z)|}{|z|^\ell} &\leq \exp\left(\sum_{k=0}^{n-1}|\alpha_k|\right) \\
&\leq \exp\left(\frac{1}{n}\sqrt{\frac{1}{n}\sum_{k=0}^{n-1}|\alpha_k|^2}\right) \\
&\leq \exp\left(\frac{1}{n}\sqrt{-\frac{1}{n}\log\prod_{k=0}^{n-1}(1-|\alpha_k|^2)}\right) \\
&= \exp\left(n\sqrt{\frac{1}{n}(-\log\tau_n^{\zeta_N} + \log\sigma_N^2)}\right).
\end{aligned}$$

Thus, since

$$\frac{1}{n}\log C \geq \frac{1}{n}\log \sigma_N^2 \geq \frac{1}{n}\log \tau_n^{\zeta_N} \geq \frac{1}{2\pi n}\int_0^{2\pi} \log s_{N-1}(S^{(N)}(e^{\imath\theta}))d\theta,$$

we obtain from Assumption 1.3 that there exists a non-negative sequence $\varepsilon_N \to_N 0$ such that

$$\max_{\ell\in[n]} \frac{|b_\ell^{\zeta_N}(z)|}{|z|^\ell} \leq \exp(n\varepsilon_N).$$

By the maximum principle, for any $\theta \in [0, 2\pi)$, any $\ell \in [n]$ and any $r > 1$,

$$\begin{aligned}|\beta_\ell^{\zeta_N}(e^{\imath\theta})| &= (\tau_\ell^{\zeta_N})^{-1/2}|b_\ell^{\zeta_N}(e^{\imath\theta})| \leq (\tau_n^{\zeta_N})^{-1/2} \sup_{\psi\in[0,2\pi)} |b_\ell^{\zeta_N}(re^{\imath\psi})| \\ &\leq r^\ell(\tau_n^{\zeta_N})^{-1/2}\exp(n\varepsilon_N),\end{aligned}$$

and by Szegö's theorem and Assumption 1.3 again, we obtain (3.13) by choosing r as close to one as desired.

Proceeding, we now consider the matrix measure $d\zeta_N(e^{\imath\theta}) \otimes I_N = (2\pi)^{-1}s_{N-1}(S^{(N)}(e^{\imath\theta}))I_N d\theta$, equipped with the sesquilinear function

$$\langle\langle F, G\rangle\rangle_{\zeta_N\otimes I_N} = \frac{1}{2\pi}\int_0^{2\pi} s_{N-1}(S^{(N)}(e^{\imath\theta}))F^*(e^{\imath\theta})G(e^{\imath\theta})d\theta.$$

It is clear that the ℓth normalized orthogonal polynomial for $\langle\langle\cdot,\cdot\rangle\rangle_{\zeta_N\otimes I_N}$ is $\beta_\ell^{\zeta_N}(z)I_N$, and the associated ℓth CD kernel is

$$K_\ell^{\zeta_N\otimes I_N}(z,u) = \sum_{k=0}^{\ell} \beta_k^{\zeta_N}(z)(\beta_k^{\zeta_N}(u))^* I_N.$$

Obviously, for each Borel set $A \subset [0, 2\pi)$, $\zeta_N(A)I_N \leq \varrho_N(A)$ in the semi-definite ordering. Therefore, by the variational characterization of the CD kernels, we have

$$\begin{aligned}K_{n-1}^{\zeta_N\otimes I_N}(e^{\imath\theta}, e^{\imath\theta}) &\geq \langle\langle P_{n-1}^{\varrho_N,\theta}, P_{n-1}^{\varrho_N,\theta}\rangle\rangle_{\zeta_N\otimes I_N}^{-1} \geq \langle\langle P_{n-1}^{\varrho_N,\theta}, P_{n-1}^{\varrho_N,\theta}\rangle\rangle_{\varrho_N}^{-1} \\ &= K_{n-1}^{\varrho_N}(e^{\imath\theta}, e^{\imath\theta})\end{aligned}$$

for all $\theta \in [0, 2\pi)$. Recalling the definition (3.8) of the kernel $K_\ell^{\varrho_N}(z,u)$, we obtain from Inequality (3.13) that

$$\max_{\ell\in[n]}\max_{\theta\in[0,2\pi)} \|\varphi_\ell^{\varrho_N}(e^{\imath\theta})\| \leq e^{n\varepsilon_N} \quad \text{with} \quad 0 \leq \varepsilon_N \to 0.$$

Also notice that

$$K_{n-1}^{\varrho_N}(e^{\imath\theta}, e^{\imath\theta}) \geq \varphi_0^{\varrho_N}(e^{\imath\theta})\varphi_0^{\varrho_N}(e^{\imath\theta})^* = \left(\int_0^{2\pi} d\varrho_N(\theta)\right)^{-1} = (R_0^{(N)})^{-1}.$$

Thus, using Assumption 1.1(ii) and recalling Definition (3.10) of the polynomials $P_\ell^{\varrho_N,\theta}$, we obtain the bound (3.11).

We now prove (3.12). Let $\delta > 0$ and $\theta_0 \in (0, 2\pi)$ be such that $S^{(N)}(e^{\imath\theta_0}) \geq \delta I_N$. Consider the matrix measure

$$d\varsigma_N(\theta) = \frac{1}{2\pi} S^{(N)}(e^{\imath\theta_0})\, d\theta.$$

The CD kernels for this measure are constant and satisfy the identity $\ell K_{\ell-1}^{\varsigma_N}(e^{\imath\theta}, e^{\imath\theta})^{-1} = S^{(N)}(e^{\imath\theta_0})$ for each integer $\ell > 0$, as can be checked by the direct calculation of the orthogonal polynomials for $d\varsigma_N(\theta)$, or by the application of Lemmas 3.2 and 3.3, with the blocks of the matrix $\mathcal{R}$ being replaced with $R_k = \mathbb{1}_{k=0} S^{(N)}(e^{\imath\theta_0})$.

Let $\eta > 0$ be an arbitrarily small number. By Assumption 1.1(i), we can choose $h > 0$ small enough so that for each $N > 0$, $\boldsymbol{w}(S^{(N)}, h) < \eta\delta$. Let $\theta \in [\theta_0 - h, \theta_0 + h]$. Then, for each vector $u \in \mathbb{C}^N$ with $\|u\| = 1$, we have

$$u^* S^{(N)}(e^{\imath\theta}) u = u^* S^{(N)}(e^{\imath\theta_0}) u \left(1 + \frac{u^*(S^{(N)}(e^{\imath\theta}) - S^{(N)}(e^{\imath\theta_0}))u}{u^* S^{(N)}(e^{\imath\theta_0}) u} \right),$$

and hence,

$$\forall \theta \in [\theta_0 - h, \theta_0 + h],\ S^{(N)}(e^{\imath\theta_0}) \leq (1 + 2\eta) S^{(N)}(e^{\imath\theta}).$$

The degree one polynomial $q(z) = 0.5(ze^{-\imath\theta_0} + 1)$ can be easily shown to satisfy $q(e^{\imath\theta_0}) = \sup_\theta |q(e^{\imath\theta})| = 1$, and $\sup_{\theta-\theta_0 \in [-\pi,\pi), |\theta-\theta_0| \geq h} |q(e^{\imath\theta})| = \cos(h/2) < 1$. Let $\tilde{\eta} > 0$ be arbitrarily small, and let $m = n - 1 + \lfloor n\tilde{\eta} \rfloor$. Consider the "trial polynomial" $Q_m(e^{\imath\theta}) = P_{n-1}^{\varrho_N;\theta_0}(e^{\imath\theta}) q(e^{\imath\theta})^{\lfloor n\tilde{\eta} \rfloor}$ with degree m. This polynomial has the following features:

- $Q_m(e^{\imath\theta_0}) = I_N$.
- By Assertion (3.11) that we just proved, for $\theta - \theta_0 \in [-\pi, \pi)$, $|\theta - \theta_0| \geq h$, for all large N,

$$\|Q_n(e^{\imath\theta}\| \leq |\cos(h/2)|^{n\tilde{\eta}} \|P_{n-1}^{\varrho_N;\theta_0}(e^{\imath\theta})\| \leq |\cos(h/2)|^{n\tilde{\eta}/2}.$$

By the variational characterization of the CD kernels, we have

$$\begin{aligned}
\frac{1}{m+1} S^{(N)}(e^{\imath\theta_0}) &= (K_m^{\varsigma_N}(e^{\imath\theta_0}, e^{\imath\theta_0}))^{-1} \\
&= \langle\!\langle P_m^{\varsigma_N,\theta_0}, P_m^{\varsigma_N,\theta_0} \rangle\!\rangle_{\varsigma_N} \\
&\leq \langle\!\langle Q_m, Q_m \rangle\!\rangle_{\varsigma_N} \\
&\leq (1 + 2\eta) \int_{\theta: |\theta-\theta_0| \leq h} Q_m(e^{\imath\theta}) d\varrho_N(\theta) Q_m(e^{\imath\theta})^* \\
&\quad + \int_{\theta: |\theta-\theta_0| > h} Q_m(e^{\imath\theta}) d\varsigma_N(\theta) Q_m(e^{\imath\theta})^*
\end{aligned}$$

$$\leq (1+2\eta)\langle\langle P_{n-1}^{\varrho_N;\theta_0}, P_{n-1}^{\varrho_N,\theta_0}\rangle\rangle_{\varrho_N} + |\cos(h/2)|^{n\tilde{\eta}}\boldsymbol{M}$$
$$= (1+2\eta)K_{n-1}^{\varrho_N}(e^{\imath\theta_0}, e^{\imath\theta_0})^{-1} + |\cos(h/2)|^{n\tilde{\eta}}\boldsymbol{M}.$$

Since η and $\tilde{\eta}$ are arbitrary, Assertion 3.12 holds true.

We now have all the elements to complete the proof of Proposition 3.3. Let $k \in [n]$ be such that $S^{(N)}(e^{2\imath\pi k/n}) \geq \delta I_N$. Thanks to Lemma 3.3, we can replace $K_{n-1}^{\varrho_N}(e^{\imath\theta_0}, e^{\imath\theta_0})^{-1}$ for $\theta_0 = 2\pi k/n$ with $\mathbb{E}\tilde{y}_k\tilde{y}_k^*$ in Assertion 3.12. This provides a lower bound on $\mathbb{E}\tilde{y}_k\tilde{y}_k^*$. To obtain an upper bound on this matrix, we recall that $\tilde{y}_k$ and $\check{y}_k$ are independent, resulting in $\mathbb{E}\tilde{y}_k\tilde{y}_k^* \leq \mathbb{E}y_k y_k^*$, and use Lemma 3.1.

This completes the proof of Proposition 3.1. We note for completeness that we could have established the upper bound by using a Nevai's polynomial trial technique as well.

3.3. *Technical results needed in the proofs of Theorems* 1.1 *and* 1.2

The following lemma can be proved similarly to, e.g. [41, Corollary 4.2.13].

Lemma 3.4 (Covering number). *Let $\mathcal{U} \subset \mathbb{C}^m$ be a k-dimensional subspace. Given $\rho > 0$, there exists a ρ-net of the Euclidean unit-ball of $\mathcal{U}$ with a cardinality bounded above by $(3/\rho)^{2k}$.*

Lemma 3.5. *Let $a_0, \ldots, a_{m-1}$ be non-negative real numbers such that there exist $0 < c \leq C$ for which $c \leq m^{-1}\sum a_k$, and $\max a_k \leq C$. Given $x \in (0, c]$, the set $\mathcal{I}(x) \subset [m]$ defined as*

$$\mathcal{I}(x) = \{k \in [m] : a_k > x\}$$

satisfies

$$|\mathcal{I}| \geq \frac{c-x}{C-x}m.$$

Proof. We have $C|\mathcal{I}| \geq \sum_{k\in\mathcal{I}} a_k \geq mc - \sum_{k\in\mathcal{I}^c} a_k \geq mc - (m - |\mathcal{I}|)x$, hence the result. □

For any random vector $\xi = [\xi_0, \ldots, \xi_{m-1}]^\mathsf{T} \in \mathbb{C}^m$, *Lévy's anti-concentration* function is defined for $t > 0$ as

$$\mathcal{L}(\xi, t) = \sup_{d\in\mathbb{C}^m} \mathbb{P}[\|\xi - d\| \leq t]$$

(when the probability $\mathbb{P}$ above is taken with respect to some random vector x, we denote the associated Lévy's anti-concentration function as $\mathcal{L}_x(\xi, t)$). Assuming that the elements ξ_ℓ of ξ are independent, letting $k \in [m]$, and defining $\xi^{(k)} = [\xi_0, \ldots, \xi_k]^\mathsf{T}$, the following restriction result is well-known [31, Lemma 2.1].

$$\mathcal{L}(\xi, t) \leq \mathcal{L}(\xi^{(k)}, t).$$

We shall need the following rather standard results on Gaussian vectors.

Lemma 3.6 (Norm and anti-concentration results for Gaussian vectors). *Let* $\xi = [\xi_0, \ldots, \xi_{N-1}]^\mathsf{T} \sim \mathcal{N}_\mathbb{C}(0, I_N)$, *and let* Σ *be an* $N \times N$ *covariance matrix. Then:*

(1) *For* $t > 0$, *it holds that* $\mathbb{P}[\|\Sigma^{1/2}\xi\| \geq \sqrt{2Nt}] \leq \exp(-(t/\|\Sigma\| - 1)N)$.
Assume that $s_{m-1}(\Sigma) \geq \alpha$ *for some* $m \in [N]$ *and some* $\alpha > 0$. *Then*:
(2) *There exists a constant* $c_{3.6,2} > 0$ *such that* $\mathbb{P}[\|\Sigma^{1/2}\xi\| \leq \sqrt{\alpha m/2}] \leq \exp(-c_{3.6,2}m)$.
(3) *There exists a constant* $C_{3.6,3} > 0$ *such that for each deterministic non-zero matrix* $M \in \mathbb{C}^{N\times N}$ *and each deterministic vector* $a \in \mathbb{C}^N$,

$$\mathcal{L}((\xi + a)^* M(\xi + a)/\|M\|_{HS}, t) \leq C_{3.6,3}t$$

(4) *There exists a constant* $C_{3.6,4} = C_{3.6,4}(\alpha) > 0$ *such that* $\mathcal{L}(\Sigma^{1/2}\xi, \sqrt{m}t) \leq (C_{3.6,4}t)^m$.
(5) *For each non-zero deterministic matrix* $M \in \mathbb{C}^{N\times N}$,

$$\mathbb{P}[\|M\xi\|^2 \geq t\|M\|_{HS}^2] \leq \exp(1 - t/2).$$

This lemma is proved in the appendix.

The following results specifically concern our matrix model. The first one will be needed to bound the spectral norm of our matrix Y.

Lemma 3.7 (Spectral norm of $\boldsymbol{Y}$). *Let Assumption* 1.1(ii) *hold true. Then, there are two constants* $c_{3.7}, c'_{3.7} > 0$ *such that for each* $t \geq c'_{3.7}$,

$$\mathbb{P}[\|Y\| \geq t] \leq \exp(-c_{3.7}Nt^2).$$

Proof. Since $Y = X\mathsf{F}$, it is enough to prove the lemma for X. By a standard ε-net argument (see [37, Lemma 2.3.2]), we know that

$$\mathbb{P}[\|X\| \geq t] \leq \sum_{u\in\boldsymbol{N}} \mathbb{P}[\|Xu\| \geq t/2],$$

where $\boldsymbol{N}$ is a 1/2–net of $\mathbb{S}^{n-1}$. Let $u \in \boldsymbol{N}$. Observing that $Xu \sim \mathcal{N}_\mathbb{C}(0, (u^\mathsf{T} \otimes I_N)\mathcal{R}(\bar{u} \otimes I_N))$, and recalling that $n\|\mathcal{R}\| \leq \boldsymbol{M}$, Lemma 3.6(1) shows that

$$\mathbb{P}[\|Xu\| \geq t/2] \leq \exp\left(-\left(\frac{t^2 n}{8\boldsymbol{M}N}\right) - 1\right) N).$$

By Lemma 3.4 and the union bound, we thus have that

$$\mathbb{P}[\|X\| \geq t] \leq \exp\left(-\left(\frac{t^2 n}{8\boldsymbol{M}N} - 1\right) N + (2\log 6)n\right).$$

Choosing t large enough and using (1.1), we get the result. □

We shall need to use a discrete frequency counterpart of Assumption 1.2:

Lemma 3.8. *Let Assumptions* 1.1(i) *and* 1.2 *hold true. Given* $N \in \mathbb{N}\backslash\{0\}$ *and* $\alpha > 0$, *let* $\mathcal{C}_{\mathrm{good}}(\alpha)$ *be the subset of* $[n]$ *defined as*

$$\mathcal{C}_{\mathrm{good}}(\alpha) = \{k \in [n] : S(e^{2\imath\pi k/n}) \geq \alpha I_N\}.$$

Then,

$$\forall\, \kappa \in (0,1), \quad \exists\, \alpha > 0, \quad |\mathcal{C}_{\mathrm{good}}(\alpha)| \geq (1-\kappa)n \quad \textit{for all large } N.$$

Proof. Let us identify a set $\mathcal{I} \subset [n]$ with the discretization of the unit-circle $\mathbb{T}$ given as $\{\exp(2\imath\pi k/n) : k \in \mathcal{I}\}$, and let us also denote any of these two sets as $\mathcal{I}$. Given a small real number $h > 0$, let $\mathcal{V}_h^{\mathbb{T}}(\mathcal{I})$ be the closed h-neighborhood of $\mathcal{I}$ on $\mathbb{T}$ equipped with the curvilinear distance. We shall show that

$$|\mathcal{I}| \leq \frac{n}{2\pi}\,\mathrm{Leb}(\mathcal{V}_h^{\mathbb{T}}(\mathcal{I})) + \frac{\pi}{h}. \tag{3.14}$$

We first observe that $\mathcal{V}_h^{\mathbb{T}}(\mathcal{I})$ consists of a finite number of disjoint closed arcs, the length of each arc being greater or equal to $2h$. Given $\varphi_0 \in [0, 2\pi)$ and $\varphi_1 \in [\varphi_0 + 2h, \varphi_0 + 2h + 2\pi)$, let $\{e^{\imath\varphi} : \varphi_0 \leq \varphi \leq \varphi_1\}$ be one such arc. Then, there are two integers $k_{\min} \leq k_{\max}$ in $\mathcal{I}$ such that $2k_{\min}\pi/n = \varphi_0 + h$ and $2k_{\max}\pi/n = \varphi_1 - h$, and all the elements of $\mathcal{I}$ in this arc belong to the set $\{k_{\min}, \ldots, k_{\max}\}$. Since there are at most $k_{\max} - k_{\min} + 1 = n(\varphi_1 - \varphi_0)/(2\pi) - nh/\pi + 1$ of these elements, and furthermore, since $\mathcal{V}_h^{\mathbb{T}}(\mathcal{I})$ consists of at most $\lfloor \pi/h \rfloor$ arcs, we obtain the inequality (3.14).

Fix an arbitrary $\kappa \in (0,1)$. By Assumption 1.2, there exists $\delta > 0$ such that the set $B = \{z \in \mathbb{T} : s_{N-1}(S(z)) \leq \delta\}$ satisfies $\mathrm{Leb}(B) \leq \kappa\pi$. Let $\alpha = \delta/2$, and assume that the set $\mathcal{C}_{\mathrm{bad}}(\alpha) = [n]\backslash\mathcal{C}_{\mathrm{good}}(\alpha)$ is non-empty (otherwise the result of the lemma is true). Relying on Assumption 1.1(i), let $h > 0$ be such that $\boldsymbol{w}(S, h) \leq \alpha$ for all N. Let $k \in \mathcal{C}_{\mathrm{bad}}(\alpha)$. By the Wielandt–Hoffmann theorem and the triangle inequality, $s_{N-1}(S(e^{\imath\theta})) \leq \delta$ for each θ such that $|\theta - 2\pi k/n| \leq h$. In other words, $\mathcal{V}_h^{\mathbb{T}}(\mathcal{C}_{\mathrm{bad}}(\alpha)) \subset B$. By inequality (3.14), we thus obtain that

$$|\mathcal{C}_{\mathrm{bad}}(\alpha)| \leq \frac{n}{2}\kappa + \frac{\pi}{h},$$

which implies that $|\mathcal{C}_{\mathrm{bad}}(\alpha)| \leq \kappa n$ for all large n. □

We now enter the proofs of Theorems 1.1 and 1.2. Given $C > 0$, we denote as $\mathcal{E}_{\mathrm{op}}(C)$ the event

$$\mathcal{E}_{\mathrm{op}}(C) = [\|Y\| \leq C].$$

In the remainder of this section, the constants will be referred to by the letters c, a, or C, possibly with primes or numerical indices that refer generally to the statements (lemmas, propositions, ...) where these constants appear for the first time. These constants do not depend on N. The typical statements where they appear are of the type

$$\mathbb{P}[[\cdots \le a] \cap \mathcal{E}_{\text{op}}(C)] \le C'a + \exp(-cN),$$

or others in the same vein. Often, such inequalities hold true for all N large enough. This detail will not always be mentioned.

3.4. *Proof of Theorem* 1.1

As mentioned in Sec. 3.1, our starting point for proving Theorem 1.1 is the variational characterization (3.3).

3.4.1. *Compressible vectors*

In the following statement and in the proof, the vectors $u \in \mathbb{S}^{N+n-1}$ will always take the form $u = [v^\mathsf{T}\ w^\mathsf{T}]^\mathsf{T}$ with $v \in \mathbb{C}^n$.

Proposition 3.2 (Compressible vectors). *There exist constants* $\theta_{3.2}$, $\rho_{3.2}$, $C_{3.2} \in (0,1)$, *and* $a_{3.2}$, $c_{3.2} > 0$ *such that, for the set*

$$\mathcal{S} := \{u \in \mathbb{S}^{N+n-1} : \|v\| \le C_{3.2} \text{ or } v/\|v\| \in \text{comp}(\theta_{3.2}, \rho_{3.2})\}$$

(*here* $v/\|v\| = 0$ *if* $v = 0$), *we have*

$$\mathbb{P}[[\inf_{u\in\mathcal{S}} \|Hu\| \le a_{3.2}] \cap \mathcal{E}_{\text{op}}(C)] \le \exp(-c_{3.2}n).$$

Proof. For each $u \in \mathbb{S}^{N+n-1}$ the inequality $\|Hu\| \le a$ for $a > 0$ implies

$$\|Yv + zw\| \le a, \tag{3.15a}$$

$$\|\Omega^{-L}v + Y^*w\| \le a. \tag{3.15b}$$

Take $a = |z|/2$. On the event $\mathcal{E}_{\text{op}}(C)$, we obtain from (3.15a) that

$$|z|/2 \ge |z|\|w\| - \|Y\|\|v\| \ge |z|(1 - \|v\|) - C\|v\|,$$

since $\sqrt{1-x^2} \ge 1 - x$ on $[0,1]$. Hence

$$\|v\| \ge C_{3.2} \triangleq \frac{|z|/2}{|z| + C},$$

which trivially implies that

$$\mathbb{P}[[\inf_{u\in\mathbb{S}^{N+n-1}:\|v\|<C_{3.2}} \|Hu\| \le |z|/2] \cap \mathcal{E}_{\text{op}}(C)] = 0.$$

Let $\mathcal{I} \subset [n]$ be such that $|\mathcal{I}| = \lfloor \theta_{3.2}n \rfloor$, where $\theta_{3.2}$ will be chosen later. Let $V_\mathcal{I} \subset \mathbb{C}^n$ be the set of vectors v such that $\|v\| \in [C_{3.2}, 1]$ and $v/\|v\| \in \mathbb{S}^{n-1}_\mathcal{I}$. Let $\boldsymbol{v}$ be a deterministic vector in $V_\mathcal{I}$, and define the random vector $\boldsymbol{w} = -z^{-1}Y\boldsymbol{v}$. Writing

$\boldsymbol{u} = [\boldsymbol{v}^\top\ \boldsymbol{w}^\top]^\top$, we shall show that there exists a constant $c > 0$ such that for all $t > 0$ small enough,

$$\mathbb{P}[[\|H\boldsymbol{u}\| \le t] \cap \mathcal{E}_{\rm op}(C)] \le (ct)^n. \tag{3.16}$$

On the event $\mathcal{E}_{\rm op}(C)$, we have

$$\begin{aligned}\frac{C}{|z|} \ge \|\boldsymbol{w}\| \ge \frac{\|Y^*\boldsymbol{w}\|}{\|Y\|} &= \frac{\|Y^*\boldsymbol{w} + \Omega^{-L}\boldsymbol{v} - \Omega^{-L}\boldsymbol{v}\|}{\|Y\|} \\ &\ge \frac{\|\boldsymbol{v}\| - \|Y^*\boldsymbol{w} + \Omega^{-L}\boldsymbol{v}\|}{\|Y\|} \\ &\ge \frac{C_{3.2} - \|Y^*\boldsymbol{w} + \Omega^{-L}\boldsymbol{v}\|}{C}.\end{aligned}$$

By the choice of $\boldsymbol{w}$, we have $[\|H\boldsymbol{u}\| \le t] = [\|\Omega^{-L}\boldsymbol{v} + Y^*\boldsymbol{w}\| \le t]$. Assume that $t \le C_{3.2}/2$, and define the interval $J = [C_{3.2}/(2C), C/|z|]]$. Then,

$$[\|H\boldsymbol{u}\| \le t] \cap \mathcal{E}_{\rm op}(C)] \subset [\|\boldsymbol{w}\| \in [(C_{3.2} - t)/C, C/|z|]] \subset [\|\boldsymbol{w}\| \in J].$$

If $J = \emptyset$, then the inequality (3.16) holds trivially. Otherwise, we write

$$\mathbb{P}[[\|H\boldsymbol{u}\| \le t] \cap \mathcal{E}_{\rm op}(C)] \le \mathbb{P}[[\|(Y_{\cdot,\mathcal{I}^c})^*\boldsymbol{w}\| \le t] \cap [\|\boldsymbol{w}\| \in J]],$$

and for each $\ell \in \mathcal{I}^c$, we write

$$y_\ell = \tilde{\boldsymbol{y}}_\ell + \check{\boldsymbol{y}}_\ell \quad \text{where } \check{\boldsymbol{y}}_\ell = \mathbb{E}[y_\ell \,|\, Y_{\cdot,\mathcal{I}}].$$

Similarly, we write

$$Y_{\cdot,\mathcal{I}^c} = \tilde{Y}_{\cdot,\mathcal{I}^c} + \check{Y}_{\cdot,\mathcal{I}^c} \quad \text{where } \check{Y}_{\cdot,\mathcal{I}^c} = \mathbb{E}[Y_{\cdot,\mathcal{I}^c}|Y_{\cdot,\mathcal{I}}].$$

Since $\boldsymbol{w}$ is $\sigma(Y_{\cdot,\mathcal{I}})$ measurable, the variables $\widetilde{Y}_{\cdot,\mathcal{I}^c}$ and $\boldsymbol{w}$ are independent, and we get from the last inequality that

$$\mathbb{P}[[\|H\boldsymbol{u}\| \le t] \cap \mathcal{E}_{\rm op}(C)] \le \sup_{w:\|w\|\in J} \mathcal{L}_{\widetilde{Y}_{\cdot,\mathcal{I}^c}}((\widetilde{Y}_{\cdot,\mathcal{I}^c})^* w, t).$$

For each deterministic vector w such that $\|w\| \in J$, we thus need to consider the probability law of the Gaussian vector $(\widetilde{Y}_{\cdot,\mathcal{I}^c})^* w$ by studying its covariance matrix $\Sigma = n\mathbb{E}(\widetilde{Y}_{\cdot,\mathcal{I}^c})^* w w^* \widetilde{Y}_{\cdot,\mathcal{I}^c}$. We first bound the spectral norm of Σ. Writing $(Y_{\cdot,\mathcal{I}^c})^* w = (I_{|\mathcal{I}^c|} \otimes w^\top)\overline{\rm vec}(\widetilde{Y}_{\cdot,\mathcal{I}^c})$, we have

$$\Sigma = n(I_{|\mathcal{I}^c|} \otimes w^\top)\mathbb{E}\overline{\rm vec}(\widetilde{Y}_{\cdot,\mathcal{I}^c})\,{\rm vec}(\widetilde{Y}_{\cdot,\mathcal{I}^c})^\top(I_{|\mathcal{I}^c|} \otimes \bar{w}),$$

thus,

$$\begin{aligned}\Sigma^\top &= n(I_{|\mathcal{I}^c|} \otimes w^*)\mathbb{E}\,{\rm vec}(\widetilde{Y}_{\cdot,\mathcal{I}^c})\,{\rm vec}(\widetilde{Y}_{\cdot,\mathcal{I}^c})^*(I_{|\mathcal{I}^c|} \otimes w) \\ &\le n(I_{|\mathcal{I}^c|} \otimes w^*)\mathbb{E}\,{\rm vec}(Y_{\cdot,\mathcal{I}^c})\,{\rm vec}(Y_{\cdot,\mathcal{I}^c})^*(I_{|\mathcal{I}^c|} \otimes w)\end{aligned}$$

in the semidefinite ordering. Using that $Y = X\mathsf{F}$, we have $\mathrm{vec}(Y_{\cdot,\mathcal{I}^c}) = ((\mathsf{F}_{\cdot,\mathcal{I}^c})^\mathsf{T} \otimes I_N)\,\mathrm{vec}(X)$, thus,

$$\Sigma^\mathsf{T} \le n((\mathsf{F}_{\cdot,\mathcal{I}^c})^\mathsf{T} \otimes w^*)\mathcal{R}(\overline{\mathsf{F}_{\cdot,\mathcal{I}^c}} \otimes w),$$

which shows that

$$\|\Sigma\| \le \boldsymbol{M}\|w\|^2 \le C_{\max} = \boldsymbol{M}C^2/|z|^2.$$

Recall that $y_\ell = \tilde{y}_\ell + \check{y}_\ell$, and observe that $\sigma(Y_{\cdot,\mathcal{I}}) \subset \sigma(Y_\ell)$ for each $\ell \in \mathcal{I}^c$. Therefore, for an arbitrary deterministic vector $u \in \mathbb{C}^N$, we have by Jensen's inequality

$$\begin{aligned} \mathbb{E}[|u^*\check{\boldsymbol{y}}_\ell|^2] = \mathbb{E}[|\mathbb{E}[u^* y_\ell|Y_{\cdot,\mathcal{I}}]|^2] &= \mathbb{E}[|\mathbb{E}[\mathbb{E}[u^* y_\ell \,|\, Y_\ell]|Y_{\cdot,\mathcal{I}}]|^2] \\ &\le \mathbb{E}[|\mathbb{E}[u^* y_\ell|Y_\ell]|^2] \\ &= \mathbb{E}[|u^*\check{y}_\ell|^2] \end{aligned}$$

for such ℓ. Since

$$\mathbb{E}y_\ell y_\ell^* = \mathbb{E}\tilde{y}_\ell\tilde{y}_\ell^* + \mathbb{E}\check{y}_\ell\check{y}_\ell^* = \mathbb{E}\tilde{\boldsymbol{y}}_\ell\tilde{\boldsymbol{y}}_\ell^* + \mathbb{E}\check{\boldsymbol{y}}_\ell\check{\boldsymbol{y}}_\ell^*,$$

we get from the previous inequality that

$$\mathbb{E}\tilde{y}_\ell\tilde{y}_\ell^* \le \mathbb{E}\tilde{\boldsymbol{y}}_\ell\tilde{\boldsymbol{y}}_\ell^*.$$

Choosing $\theta_{3.2} \le 1/4$, we have that $|\mathcal{I}^c| \ge 3n/4$. By Lemma 3.8, there exists $\alpha > 0$ such that $|\mathcal{I}^c \cap \mathcal{C}_{\mathrm{good}}(\alpha)| \ge n/2$. With this, we have

$$\mathrm{tr}\,\Sigma = nw^*\mathbb{E}\widetilde{Y}_{\cdot,\mathcal{I}^c}(\widetilde{Y}_{\cdot,\mathcal{I}^c})^* w = n\sum_{k\in\mathcal{I}^c} w^*\mathbb{E}\tilde{\boldsymbol{y}}_k\tilde{\boldsymbol{y}}_k^* w \ge n\sum_{k\in\mathcal{I}^c\cap\mathcal{C}_{\mathrm{good}}} w^*\mathbb{E}\tilde{y}_k\tilde{y}_k^* w.$$

By Proposition 3.1, we thus obtain that for all large N,

$$\mathrm{tr}\,\Sigma \ge \frac{\alpha}{2}|\mathcal{I}^c \cap \mathcal{C}_{\mathrm{good}}|\|w\|^2 \ge n\frac{\alpha}{4}\frac{C_{3.2}^2}{4C^2}.$$

Write $C_{\min} = \alpha C_{3.2}^2 n/(16C^2|\mathcal{I}^c|)$, and let k be the largest integer in $[|\mathcal{I}^c|]$ such that $s_{k-1}(\Sigma) \ge C_{\min}/2$. Then, we obtain from Lemma 3.5 that

$$k \ge \frac{C_{\min}}{2C_{\max} - C_{\min}}|\mathcal{I}^c| \ge \frac{C_{\min}}{2C_{\max} - C_{\min}}\frac{3n}{4}.$$

With this at hand, we can deduce Inequality (3.16) from Lemma 3.6(4).

Now, still fixing $\mathcal{I}$ and t as above, set $\rho_{3.2}$ as

$$\rho_{3.2} = \frac{t}{3 + C|z|^{-1}(2C+1)}.$$

Set $a_{3.2} = (|z|/2) \wedge \rho_{3.2}$, and let $u = [v^\mathsf{T}\ w^\mathsf{T}]^\mathsf{T} \in \mathbb{S}^{N+n-1}$ be such that $v \in \mathcal{V}^{\mathbb{C}^n}_{\rho_{3.2}}(V_\mathcal{I})$ and $\|Hu\| \le a_{3.2}$. Let $\mathcal{K}_{\rho_{3.2}}$ be a $\rho_{3.2}$-net of $V_\mathcal{I}$. By Lemma 3.4, we can choose $\mathcal{K}_{\rho_{3.2}}$ in such a way that $|\mathcal{K}_{\rho_{3.2}}| \le (3/\rho_{3.2})^{2\theta_{3.2}n}$. By the triangle inequality, there is $\boldsymbol{v} \in \mathcal{K}_{\rho_{3.2}}$ such that $\|v - \boldsymbol{v}\| \le 2\rho_{3.2}$. Let $\boldsymbol{w} = -z^{-1}Y\boldsymbol{v}$, and write $\boldsymbol{u} = [\boldsymbol{v}^\mathsf{T}\ \boldsymbol{w}^\mathsf{T}]^\mathsf{T}$. Since $\|Yv + zw\| \le \rho_{3.2}$ by (3.15a), we have

$$\|Y(v - \boldsymbol{v}) + z(w - \boldsymbol{w})\| \le \rho_{3.2}.$$

Thus, on the event $\mathcal{E}_{\mathrm{op}}(C)$,

$$\|w - \boldsymbol{w}\| \le |z|^{-1}(\rho_{3.2} + 2\rho_{3.2}\|Y\|) \le |z|^{-1}(\rho_{3.2} + 2\rho_{3.2}C).$$

From the inequality $\|\Omega^{-L}v + Y^*w\| \le \rho_{3.2}$ (see (3.15b)), we get

$$\|\Omega^{-L}\boldsymbol{v} + Y^*\boldsymbol{w} + \Omega^{-L}(v - \boldsymbol{v}) + Y^*(w - \boldsymbol{w})\| \le \rho_{3.2}.$$

Thus,

$$\|\Omega^{-L}\boldsymbol{v} + Y^*\boldsymbol{w}\| \le (3 + C|z|^{-1}(1 + 2C))\rho_{3.2} = t.$$

This implies that $\|H\boldsymbol{u}\| \le t$. As a consequence,

$$\begin{aligned}&[\exists u \in \mathbb{S}^{N+n-1} : v \in \mathcal{V}_{\rho_{3.2}}(V_\mathcal{I}),\ \|Hu\| \le a_{3.2}] \subset [\exists \boldsymbol{v} \in \mathcal{K}_{\rho_{3.2}} :\\ &\quad \|H[\boldsymbol{v}^\mathsf{T}, -z^{-1}(Y\boldsymbol{v})^\mathsf{T}]^\mathsf{T}\| \le t].\end{aligned}$$

Applying the union bound and using Inequality (3.16), we therefore get that

$$\mathbb{P}\left[\left[\inf_{u\in\mathbb{S}^{N+n-1}:\|v\|\ge C_{3.2},v\in\mathcal{V}_{\rho_{3.2}}(V_\mathcal{I})} \|Hu\| \le a_{3.2}\right] \cap \mathcal{E}_{\mathrm{op}}(C)\right] \le (3/\rho_{3.2})^{2\theta_{3.2}n}(ct)^n.$$

Now, considering all the sets $\mathcal{I} \subset [n]$ such that $|\mathcal{I}| = \lfloor\theta_{3.2}n\rfloor$, and using the bound $\binom{n}{m} \le (en/m)^m$ along with the union bound again, we get that

$$\begin{aligned}&\mathbb{P}\left[\left[\inf_{u\in\mathbb{S}^{N+n-1}:\|v\|\ge C_{3.2},v\in\mathrm{comp}(\theta_{3.2},\rho_{3.2})} \|Hu\| \le a_{3.2}\right] \cap \mathcal{E}_{\mathrm{op}}(C)\right]\\ &\quad \le (e/\theta_{3.2})^{\theta_{3.2}n}(3/\rho_{3.2})^{2\theta_{3.2}n}(ct)^n.\end{aligned}$$

The proof is completed by choosing $\theta_{3.2}$ small enough. □

3.4.2. *Incompressible vectors*

We now consider the action of H on the vectors $u \in \mathbb{S}^{N+n-1}$ that belong to the complement of the set $\mathcal{S}$ of Proposition 3.2 in the unit-sphere.

Proposition 3.3 (Incompressible vectors for the smallest singular value). *There exists a constant $c_{3.3} > 0$ such that for $\varepsilon > 0$ arbitrarily small,*

$$\mathbb{P}\left[\left[\inf_{u\in\mathbb{S}^{N+n-1}\setminus\mathcal{S}} \|Hu\| \le n^{-3/2}t\right] \cap \mathcal{E}_{\mathrm{op}}(C)\right] \le \varepsilon t + \exp(-c_{3.3}\varepsilon^2 N).$$

The remainder of this subsection is devoted to the proof of this proposition. As in the proof of Proposition 3.2, a vector $u \in \mathbb{S}^{N+n-1}$ will be written $u = [v^\mathsf{T}\ w^\mathsf{T}]^\mathsf{T}$ with $v = [v_0, \ldots, v_{n-1}]^\mathsf{T} \in \mathbb{C}^n$. When $u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}$, the vector $\tilde{v} = v/\|v\| = [\tilde{v}_0, \ldots, \tilde{v}_{n-1}]^\mathsf{T}$ belongs now to $\operatorname{incomp}(\theta_{3.2}, \rho_{3.2})$. Associate with any such vector u, the set

$$\mathcal{J}_u = \{k \in [n] : |v_k| \geq \rho_{3.2} C_{3.2}/\sqrt{2n}\}.$$

Then, $|\mathcal{J}_u| > \theta_{3.2} n$. Indeed, the set $\widetilde{\mathcal{J}}_v = \{k \in [n] : |\tilde{v}_k| \geq \rho_{3.2}/\sqrt{2n}\}$ satisfies $|\widetilde{\mathcal{J}}_v| > \theta_{3.2} n$. To see this, denote as $\Pi_{\widetilde{\mathcal{J}}_v}$ the orthogonal projector on the vectors supported by $\widetilde{\mathcal{J}}_v$, and check that $\|\tilde{v} - \Pi_{\widetilde{\mathcal{J}}_v}(\tilde{v})/\|\Pi_{\widetilde{\mathcal{J}}_v}(\tilde{v})\|\| \leq \rho_{3.2}$. If $|\widetilde{\mathcal{J}}_v| \leq \theta_{3.2} n$, we get a contradiction to the fact that $\tilde{v} \in \operatorname{incomp}(\theta_{3.2}, \rho_{3.2})$. It remains to check that if $u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}$, then $\widetilde{\mathcal{J}}_v \subset \mathcal{J}_u$.

Now, choose $\alpha_{3.3} > 0$ small enough so that the set $\mathcal{C}_{\text{good}}(\alpha)$ from Lemma 3.8 satisfies $|\mathcal{C}_{\text{good}}(\alpha_{3.3})| \geq (1 - \theta_{3.2}/2)n$. Observe that with this choice, $|\mathcal{J}_u \cap \mathcal{C}_{\text{good}}(\alpha_{3.3})| \geq \theta_{3.2} n/2$.

We now use the canvas of the proof of [31, Lemma 3.5] to relate the infimum over the incompressible vectors with the distance between column subspaces of H. Specifically, for each $u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}$, we have

$$\begin{aligned} \|Hu\| &\geq \max_{k \in \mathcal{J}_u \cap \mathcal{C}_{\text{good}}(\alpha_{3.3})} |v_k| \operatorname{dist}(h_k, H_k) \\ &\geq \frac{\rho_{3.2} C_{3.2}}{\sqrt{2n}} \max_{k \in \mathcal{J}_u \cap \mathcal{C}_{\text{good}}(\alpha_{3.3})} \operatorname{dist}(h_k, H_k). \end{aligned}$$

Thus,

$$\begin{aligned} &\left[\inf_{u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}} \|Hu\| \leq \frac{\rho_{3.2} C_{3.2} t}{\sqrt{2} n^{3/2}} \right] \\ &\quad \subset \left[\inf_{u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}} \max_{k \in \mathcal{J}_u \cap \mathcal{C}_{\text{good}}(\alpha_{3.3})} \operatorname{dist}(h_k, H_k) \leq \frac{t}{n} \right]. \end{aligned}$$

Denoting as $\mathcal{E}$ the event at the right-hand side of this inclusion, we have from what precedes that

$$\mathbb{1}_{\mathcal{E}} \leq \frac{2}{\theta_{3.2} n} \sum_{k \in \mathcal{C}_{\text{good}}(\alpha_{3.3})} \mathbb{1}_{[\operatorname{dist}(h_k, H_k) \leq t/n]},$$

which implies that

$$\begin{aligned} &\mathbb{P}\left[\left[\inf_{u \in \mathbb{S}^{N+n-1} \setminus \mathcal{S}} \|Hu\| \leq \frac{\rho_{3.2} C_{3.2} t}{\sqrt{2} n^{3/2}} \right] \cap \mathcal{E}_{\text{op}}(C) \right] \\ &\quad \leq \frac{2}{\theta_{3.2} n} \sum_{k \in \mathcal{C}_{\text{good}}(\alpha_{3.3})} \mathbb{P}[[\operatorname{dist}(h_k, H_k) \leq t/n] \cap \mathcal{E}_{\text{op}}(C)]. \end{aligned} \tag{3.17}$$

A workable formula for $\mathrm{dist}(h_k, H_k)$ is provided by the following lemma.

Lemma 3.9. *Let $M \in \mathbb{C}^{m\times m}$, and partition this matrix as*

$$M = [m_0 \quad M_{-0}] = \begin{bmatrix} m_{00} & m_{01} \\ m_{10} & M_{11} \end{bmatrix},$$

where m_0 is the first column of M, M_{-0} is the submatrix that remains after extracting m_0, and m_{00} is the first element of the vector m_0. Assume that M_{11} is invertible. Then,

$$\mathrm{dist}(m_0, M_{-0}) = \frac{|m_{00} - m_{01}M_{11}^{-1}m_{10}|}{\sqrt{1 + \|m_{01}M_{11}^{-1}\|^2}}.$$

Proof. Write $\mathrm{dist}(m_0, M_{-0})^2 = m_0^*(I - M_{-0}(M_{-0}^*M_{-0})^{-1}M_{-0}^*)m_0$, and observe that $M_{-0}^*M_{-0} = M_{11}^*M_{11} + m_{01}^*m_{01}$. Letting $v = M_{11}^{-*}m_{01}^*$, we obtain by the Sherman–Morrison–Woodbury identity that

$$(M_{-0}^*M_{-0})^{-1} = M_{11}^{-1}\left(I - \frac{vv^*}{1+\|v\|^2}\right)M_{11}^{-*},$$

thus,

$$I - M_{-0}(M_{-0}^*M_{-0})^{-1}M_{-0}^* = I - \begin{bmatrix} m_{01} \\ M_{11} \end{bmatrix} M_{11}^{-1}\left(I - \frac{vv^*}{1+\|v\|^2}\right)M_{11}^{-*}[m_{01}^* \quad M_{11}^*]$$

$$= \frac{1}{1+\|v\|^2}\begin{bmatrix} 1 & -v^* \\ -v & vv^* \end{bmatrix} = \frac{1}{1+\|v\|^2}\begin{bmatrix} 1 \\ -v \end{bmatrix}[1 \quad -v^*].$$

This leads to

$$\mathrm{dist}(m_0, M_{-0})^2 = \frac{1}{1+\|v\|^2}\left|[1 \quad -v^*]\begin{bmatrix} m_{00} \\ m_{10} \end{bmatrix}\right|^2,$$

which is the required result. □

For convenience, let is recall the matrices G_k and G_k^{-1} from Sec. 3.1:

$$G_k = \begin{bmatrix} \Omega_k^{-L} & Y_k^* \\ Y_k & z \end{bmatrix}, \quad G_k^{-1} = \begin{bmatrix} E_k & F_k \\ P_k & D_k \end{bmatrix}.$$

Applying the above lemma on the matrix H with the column k being used instead of Column zero, we get that

$$\mathrm{dist}(h_k, H_k) = \frac{|\omega^{-kL} - y_k^*D_ky_k|}{\sqrt{1 + \|[0 \quad y_k^*]G_k^{-1}\|^2}} = \frac{|\omega^{-kL} - y_k^*D_ky_k|}{\sqrt{1 + \|y_k^*P_k\|^2 + \|y_k^*D_k\|^2}}. \tag{3.18}$$

We shall use the notation

$$\mathrm{Num}_k = |\omega^{-kL} - y_k^*D_ky_k|, \quad \text{and} \quad \mathrm{Den}_k = \sqrt{1 + \|y_k^*P_k\|^2 + \|y_k^*D_k\|^2}.$$

Lemma 3.10. *The following facts hold true*:

(1) *Assume that $k \in \mathcal{C}_{\rm good}(\alpha_{3.3})$. Then, there exist $c_{3.10}, C_{3.10} > 0$ such that*

$$\mathbb{P}[[\|[0 \quad y_k^*]G_k^{-1}\| \le C_{3.10}] \cap \mathcal{E}_{\rm op}(C)] \le \exp(-c_{3.10}N).$$

(2) *On $\mathcal{E}_{\rm op}(C)$, for each $k \in [n]$, we have $\|P_k\|_{HS} \le C\|D_k\|_{HS}$.*

Proof. (1) Consider

$$\|y_k\| = \|[0 \quad y_k^*]G_k^{-1}G_k\| \le \|[0 \quad y_k^*]G_k^{-1}\|\|G_k\|.$$

On $\mathcal{E}_{\rm op}(C)$, we have $\|G_k\| \le C + |z| + 1$. Moreover, since $k \in \mathcal{C}_{\rm good}(\alpha_{3.3})$, and since $\|n\mathbb{E}y_k y_k^* - S(e^{2\imath\pi k/n})\|$ is small by Lemma 3.1, there exist two constants $c', C' > 0$ such that

$$\mathbb{P}[\|y_k\| \le C'] \le \exp(-c'N)$$

by Lemma 3.6(2). The proof of Item (1) is complete.

(2) We show that for each constant non-zero vector $u \in \mathbb{C}^N$,

$$\|u^* P_k\| \le C\|u^* D_k\|.$$

The result follows then from the identity $\|M\|_{\rm HS}^2 = \sum_k \|e_k^* M\|^2$, valid for each matrix M. The vectors $v^* = u^* P_k$ and $w^* = u^* D_k$ satisfy

$$[v^* \quad w^*] = [0 \quad u^*]G_k^{-1},$$

and hence,

$$[0 \quad u^*] = [v^* \quad w^*]\begin{bmatrix}\Omega_k^{-L} & Y_k^* \\ Y_k & z\end{bmatrix}.$$

In particular, $v^*\Omega_k^{-L} + w^* Y_k = 0$, and this shows that $\|v\| \le C\|w\|$ on $\mathcal{E}_{\rm op}(C)$. ☐

Lemma 3.10 will be used to control the value of Den_k. Fix an arbitrary small real number $\varepsilon > 0$, and define the event

$$\mathcal{E}_{\rm Den}(\varepsilon) = [\mathrm{Den}_k^2 \ge \varepsilon\|D_k\|_{\rm HS}^2].$$

Lemma 3.11 (Denominator not too large). *Assume that $k \in \mathcal{C}_{\rm good}(\alpha_{3.3})$. Then, there exists $c_{3.11} > 0$ such that for all large n, $\mathbb{P}[\mathcal{E}_{Den}(\varepsilon) \cap \mathcal{E}_{\rm op}(C)] \le \exp(-c_{3.11}\varepsilon n)$.*

Proof. Defining the event $\mathcal{E} = [\mathrm{Den}_k^2 \ge (C_{3.10}^{-2} + 1)\|[0 \quad y_k^*]G_k^{-1}\|^2]$, we have

$$\mathbb{P}[\mathcal{E} \cap \mathcal{E}_{\rm op}(C)] = \mathbb{P}[[C_{3.10}^2 \ge \|[0 \quad y_k^*]G_k^{-1}\|^2] \cap \mathcal{E}_{\rm op}(C)] \le \exp(-c_{3.10}n)$$

by Lemma 3.10(1). Writing $\mathcal{E}'(\varepsilon) = [\|[0 \quad y_k^* G_k^{-1}\|^2 \geq \varepsilon \|D_k\|_{\mathrm{HS}}^2]$, we have

$$\begin{aligned}\mathbb{P}[\mathcal{E}'(\varepsilon) \cap \mathcal{E}_{\mathrm{op}}(C)] &= \mathbb{P}[[\|y_k^* P_k\|^2 + \|y_k^* D_k\|^2 \geq \varepsilon \|D_k\|_{\mathrm{HS}}^2] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\leq \mathbb{P}[\|y_k^* P_k\|^2 \geq \varepsilon \|P_k\|_{\mathrm{HS}}^2/(2C^2)] + \mathbb{P}[\|y_k^* D_k\|^2 \\ &\geq \varepsilon \|D_k\|_{\mathrm{HS}}^2/2],\end{aligned}$$

where we used Lemma 3.10(2). in the first inequality.

Note that

$$\|y_k^* D_k\|^2 \leq 2\|\tilde{y}_k^* D_k\|^2 + 2\|\check{y}_k^* D_k\|^2 \quad \text{and} \quad \|\check{y}_k^* D_k\| \leq \|\check{y}_k\|^2 \|D_k\|_{\mathrm{HS}}.$$

Hence there exists $c_N \to \infty$ such that

$$\begin{aligned}\mathbb{P}[\|y_k^* D_k\|^2 \geq \varepsilon \|D_k\|_{\mathrm{HS}}^2/2] &\leq \mathbb{P}[\|\tilde{y}_k^* D_k\|^2 \geq \varepsilon \|D_k\|_{\mathrm{HS}}^2/8] + \mathbb{P}[\|\check{y}_k^*\|^2 \geq \varepsilon/8] \\ &\leq \mathbb{P}[\|\tilde{y}_k^* D_k\|^2 \geq \varepsilon \|S_k^{1/2} D_k\|_{\mathrm{HS}}^2/(8\boldsymbol{M})] + e^{-c_N n} \\ &\leq e^{1-\varepsilon n/(16\boldsymbol{M})} + e^{-c_N n},\end{aligned}$$

where we used Lemma 3.6(1) along with Proposition 3.1 in the second inequality, and Lemma 3.6(5) in the third one. We have a similar bound for $\mathbb{P}[\|y_k^* P_k\|^2 \geq \varepsilon \|P_k\|_{\mathrm{HS}}^2/(2C^2)]$. Consequently, there exists $c > 0$ such that $\mathbb{P}[\mathcal{E}'(\varepsilon) \cap \mathcal{E}_{\mathrm{op}}(C)] \leq e^{-c\varepsilon n}$ for all large N.

Observing that $\mathcal{E}_{\mathrm{Den}}(\varepsilon) \cap \mathcal{E}^{\mathrm{c}} \subset \mathcal{E}'(\varepsilon/(C_{3.10}^{-2} + 1))$, we obtain as a consequence of these inequalities that there exists $c_{3.11} > 0$ such that for all large N,

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{\mathrm{Den}}(\varepsilon) \cap \mathcal{E}_{\mathrm{op}}(C)] &\leq \mathbb{P}[\mathcal{E}_{\mathrm{Den}}(\varepsilon) \cap \mathcal{E}^{\mathrm{c}} \cap \mathcal{E}_{\mathrm{op}}(C)] + \mathbb{P}[\mathcal{E} \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\leq \mathbb{P}[\mathcal{E}'(\varepsilon/(C_{3.10}^{-2} + 1)) \cap \mathcal{E}_{\mathrm{op}}(C)] + \mathbb{P}[\mathcal{E} \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\leq \exp(-c_{3.11}\varepsilon n).\end{aligned}$$

The proof of the lemma is now complete. □

Now, we get back to the expression in (3.17). As in Sec. 3.1, we use the shorthand notation $S_k = S(e^{2\imath\pi k/n})$. Given $k \in \mathcal{C}_{\mathrm{good}}(\alpha_{3.3})$, we have

$$\begin{aligned}&\mathbb{P}[[\mathrm{dist}(h_k, H_k) \leq t/n] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad = \mathbb{P}[[\mathrm{Num}_k/\mathrm{Den}_k \leq t/n] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad \leq \mathbb{P}\left[\left[\frac{\mathrm{Num}_k}{\mathrm{Den}_k} \leq t/n\right] \cap \mathcal{E}_{\mathrm{Den}}(\varepsilon)^{\mathrm{c}} \cap \mathcal{E}_{\mathrm{op}}(C)\right] + \mathbb{P}[\mathcal{E}_{\mathrm{Den}}(\varepsilon) \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad \leq \mathbb{P}\left[\frac{n\mathrm{Num}_k}{\|D_k\|_{\mathrm{HS}}} \leq t\sqrt{\varepsilon}\right] + \exp(-c_{3.11}\varepsilon n) \\ &\quad \leq \mathbb{P}\left[\frac{n\mathrm{Num}_k}{\|S_k^{1/2} D_k S_k^{1/2}\|_{\mathrm{HS}}} \leq t\sqrt{\varepsilon}/\alpha_{3.3}\right] + \exp(-c_{3.11}\varepsilon n).\end{aligned}$$

Writing $\text{Num}_k = |\omega^{-kL} + (\tilde{y}_k + \check{y}_k)^* D_k(\tilde{y}_k + \check{y}_k)|$, we obtain from Lemma 3.6(3) and Proposition 3.1 that for some constant $c > 0$,

$$\begin{aligned}\mathbb{P}[[\text{dist}(h_k, H_k) \le t/n] \cap \mathcal{E}_{\text{op}}(C)] &\le \mathbb{E}_{Y_k} \mathcal{L}_{\tilde{y}_k}\left(\frac{(\tilde{y}_k + \check{y}_k)^* D_k(\tilde{y}_k + \check{y}_k)}{\|S_k^{1/2} D_k S_k^{1/2}\|_{\text{HS}}}, \frac{t\sqrt{\varepsilon}}{\alpha_{3.3}}\right) \\ &\quad + \exp(-c_{3.11}\varepsilon n) \\ &\le ct\sqrt{\varepsilon} + \exp(-c_{3.11}\varepsilon n).\end{aligned}$$

Proposition 3.3 is obtained by combining the last inequality with Inequality (3.17).

Theorem 1.1: *End of proof.*

The proof is now completed by combining Propositions 3.2, 3.3 and Lemma 3.7.

3.5. *Proof of Theorem* 1.2

Throughout the proof, we fix $k \in [\lfloor N^\beta \rfloor, \lfloor N/2 \rfloor]$. None of the constants that will appear in the course of the proof will depend on k.

The following lemma refines Inequality (3.2).

Lemma 3.12. *For any* $k \in [N]$, *we have*

$$s_{N+n-k-1}(H_{\cdot,[N+n-k]}) \le s_{N-k-1}(Y\Omega^L Y^* - z).$$

Proof. Write any vector $u \in \mathbb{C}^{N+n-k}$ as $u = [v^\mathsf{T}\ w^\mathsf{T}]^\mathsf{T}$ with $w \in \mathbb{C}^{N-k}$. Writing

$$H_{\cdot,[N+n-k]}u = \begin{bmatrix} \Omega^{-L} & (Y_{[N-k],\cdot})^* \\ Y & z(I_N)_{\cdot,[N-k]} \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} \Omega^{-L}v + (Y_{[N-k],\cdot})^* w \\ Yv + z(I_N)_{\cdot,[N-k]}w \end{bmatrix},$$

we have

$$\begin{aligned} s_{N+n-k-1}(H_{\cdot,[N+n-k]}) &= \min_{u \in \mathbb{S}^{N+n-k-1}} \|H_{\cdot,[N+n-k]}u\| \\ &\le \min_{u \in \mathbb{S}^{N+n-k-1} : v = -\Omega^L (Y_{[N-k],\cdot})^* w} \|H_{\cdot,[N+n-k]}u\| \\ &\le \min_{w \in \mathbb{S}^{N-k-1}} \|(zI_N - Y\Omega^L Y^*)_{\cdot,[N-k]}w\|. \end{aligned}$$

On the other hand, using the variational characterization of the singular values of a matrix, see [20, Theorem 4.3.15], we can write

$$\min_{w \in \mathbb{S}^{N-k-1}} \|(Y\Omega^L Y^* - zI_N)_{\cdot,[N-k]}w\| \le s_{N-k-1}(Y\Omega^L Y^* - zI_N),$$

and hence the result follows. ☐

Similar to what we did for controlling the smallest singular value of H, we use the characterization $s_{N+n-k-1}(H_{\cdot,[N+n-k]}) = \min_{u \in \mathbb{S}^{N+n-k-1}} \|H_{\cdot,[N+n-k]}u\|$, and we partition the set $\mathbb{S}^{N+n-k-1}$ into sets of compressible and incompressible vectors.

So, write any vector $u \in \mathbb{S}^{N+n-k-1}$ as $u = [v^\top \ w^\top]^\top$ with $w \in \mathbb{C}^{N-k}$. Let the set $\mathcal{S} \subset \mathbb{S}^{N+n-k-1}$ be as in the statement of Proposition 3.2. We of course have

$$s_{N+n-k-1}(H_{\cdot,[N+n-k]}) = \inf_{u \in \mathcal{S}} \|H_{\cdot,[N+n-k]}u\| \wedge \inf_{u \in \mathbb{S}^{N+n-k-1} \setminus \mathcal{S}} \|H_{\cdot,[N+n-k]}u\|. \tag{3.19}$$

It can be readily checked that Proposition 3.2 remains true, once we change the values of the constants that appear in its statement as needed. So, finally, the contribution of the term $\inf_{u \in \mathcal{S}} \|H_{\cdot,[N+n-k]}u\|$ has been estimated, and we are left with the incompressible vectors. Regarding these, we have the following proposition.

Proposition 3.4 (Incompressible vectors for the small singular values). *There exist constants $a_{3.4} > 0$ and $c_{3.4} > 0$ such that for all N larger than an integer that is independent of $k \in [\lfloor N^\beta \rfloor, \lfloor N/2 \rfloor]$,*

$$\mathbb{P}\left[\left[\inf_{u \in \mathbb{S}^{N+n-k-1} \setminus \mathcal{S}} \|H_{\cdot,[N+n-k]}u\| \le a_{3.4}\sqrt{k}/N\right] \cap \mathcal{E}_{\text{op}}(C)\right] \le \exp(-c_{3.4}k).$$

Remember the definition of the set $\mathcal{C}_{\text{good}}(\alpha) \subset [n]$ in Lemma 3.8. To prove this proposition, we begin by mimicking the argument that lead to Inequality (3.17) above. Specifically, there exist positive constants c, c' and α such that

$$\mathbb{P}\left[\left[\inf_{u \in \mathbb{S}^{N+n-k-1} \setminus \mathcal{S}} \|H_{\cdot,[N+n-k]}u\| \le \frac{ct}{n^{1/2}}\right] \cap \mathcal{E}_{\text{op}}(C)\right] \le \frac{c'}{n} \sum_{\ell \in \mathcal{C}_{\text{good}}(\alpha)} \mathbb{P}[[\text{dist}(h_\ell, H_{\cdot,[N+n-k]\setminus\{\ell\}}) \le t] \cap \mathcal{E}_{\text{op}}(C)], \tag{3.20}$$

and $|\mathcal{C}_{\text{good}}(\alpha)|$ is of order N.

This inequality calls for the following remark.

Remark 3.1. By checking the structure of the vector h_ℓ and the matrix $H_{\cdot,[N+n-k]\setminus\{\ell\}}$ (see below), one can intuitively infer that with high probability, $\text{dist}(h_\ell, H_{\cdot,[N+n-k]\setminus\{\ell\}})^2$ is of order $\text{codim}(H_{\cdot,[N+n-k]\setminus\{\ell\}})/N = (k-1)/N$. Taking $t = \sqrt{k/N}$ and recalling the characterization (3.19), we get that $s_{N+n-k-1} \gtrsim \sqrt{k}/N$, which is what is predicted by Theorem 1.2. However, one naturally expects that these singular values grow linearly in k (as k/N) which would make the result of this theorem sub-optimal. Actually, this sub-optimality is due in the first place to Inequality (3.20), which is too conservative for obtaining optimal bounds on the small singular values that we are dealing with here.

In [32], a tighter inequality is used to control the smallest singular value of a rectangular matrix. Obtaining a corresponding inequality appears to be a quite involved task in the present context.

Proof of Proposition 3.4. We need to bound the summands at the right-hand side of Inequality (3.20). To this end, assume for notational simplicity that $\ell = 0$

(assuming without loss of generality that $0 \in \mathcal{C}_{\rm good}(\alpha)$). In this case,

$$h_0 = \begin{bmatrix} 1 \\ 0_{n-1} \\ y_0 \end{bmatrix},$$

and

$$H_{\cdot,[N+n-k]\setminus\{0\}} = \begin{bmatrix} 0_{n-1}^{\mathsf T} & (y_0)^*_{[N-k]} \\ (\Omega^{-L})_{[1:n],[1:n]} & (Y_{[N-k],[1:n]})^* \\ Y_{\cdot,[1:n]} & \begin{bmatrix} zI_{N-k} \\ 0_{k\times(N-k)} \end{bmatrix} \end{bmatrix} \in \mathbb{C}^{(N+n)\times(N+n-k-1)}.$$

Write

$$\underline{h}_0 = \begin{bmatrix} 0_{n-1} \\ y_0 \end{bmatrix} \in \mathbb{C}^{N+n-1},$$

and let $\underline{H}_0$ be the matrix obtained by taking out the row 0 from $H_{\cdot,[N+n-k]\setminus\{0\}}$, i.e.

$$\underline{H}_0 = \begin{bmatrix} (\Omega^{-L})_{[1:n],[1:n]} & (Y_{[N-k],[1:n]})^* \\ Y_{\cdot,[1:n]} & \begin{bmatrix} zI_{N-k} \\ 0_{k\times(N-k)} \end{bmatrix} \end{bmatrix} \in \mathbb{C}^{(N+n-1)\times(N+n-k-1)}.$$

Note that $\underline{h}_0$ and $\underline{H}_0$ are "almost" independent. However, this is not so for h_0 and $H_{\cdot,[N+n-k]\setminus\{0\}}$. Let $a \in \mathbb{C}^{N+n-k-1}$ be the vector such that $H_{\cdot,[N+n-k]\setminus\{0\}}a = \Pi_{H_{\cdot,[N+n-k]\setminus\{0\}}} h_0$. Then,

$$\begin{aligned} \operatorname{dist}(h_0, H_{\cdot,[N+n-k]\setminus\{0\}})^2 &= \|h_0 - H_{\cdot,[N+n-k]\setminus\{0\}}a\|^2 \geq \|\underline{h}_0 - \underline{H}_0 a\|^2 \\ &\geq \operatorname{dist}(\underline{h}_0, \underline{H}_0)^2. \end{aligned}$$

Consider again the decomposition $y_0 = \tilde{y}_0 + \check{y}_0$ with $\check{y}_0 = \mathbb{E}[y_0 | Y_{\cdot,[1:n]}]$, and write

$$\tilde{h}_0 = \begin{bmatrix} 0_{n-1} \\ \tilde{y}_0 \end{bmatrix}, \quad \text{and} \quad \check{h}_0 = \begin{bmatrix} 0_{n-1} \\ \check{y}_0 \end{bmatrix} \quad \in \mathbb{C}^{N+n-1}.$$

Then, writing $G_0 = [\check{h}_0 \quad \underline{H}_0]$, we have

$$\operatorname{dist}(\underline{h}_0, \underline{H}_0) = \operatorname{dist}(\tilde{h}_0 + \check{h}_0, \underline{H}_0) \geq \operatorname{dist}(\tilde{h}_0 + \check{h}_0, G_0) = \operatorname{dist}(\tilde{h}_0, G_0).$$

Observe that $\tilde{h}_0$ and G_k are independent, and that $N + n - k - 1 \leq \operatorname{rank}(G_0) \leq N+n-k$ with probability one. Let $r = N+n-1-\operatorname{rank}(G_0)$, and let $A = [V^{\mathsf T}\, W^{\mathsf T}]^{\mathsf T} \in \mathbb{C}^{(N+n-1)\times r}$ be an isometry matrix such that $AA^* = \Pi^{\perp}_{G_0}$. Here the partitioning of

A is such that $V \in \mathbb{C}^{(n-1)\times r}$, and $W \in \mathbb{C}^{N\times r}$. Note that $r \in \{k-1, k\}$ w.p. 1. Since $AA^* = \Pi^{\perp}_{G_0}$, it holds that $\underline{H}_0^* A = 0$, which can be elaborated as

$$\begin{aligned} 0 &= (\Omega^L)_{[1:n],[1:n]} V + (Y_{\cdot,[1:n]})^* W \\ 0 &= Y_{[N-k],[1:n]} U + [\bar{z} I_{N-k} \quad 0_{(N-k)\times k}] W. \end{aligned} \tag{3.21}$$

Assume that $\|W\|^2_{\mathrm{HS}} < ar$ with $a = /(1 + C^2)$. By Eq. (3.21), on the event $\mathcal{E}_{\mathrm{op}}(C)$, we have $\|V\|^2_{\mathrm{HS}} \le C^2 ar$, which contradicts the identity $\|V\|^2_{\mathrm{HS}} + \|W\|^2_{\mathrm{HS}} = r$. Thus,

$$\mathbb{P}[[\|W\|^2_{\mathrm{HS}} < ar] \cap \mathcal{E}_{\mathrm{op}}(C)] = 0.$$

Using this result, we now provide a control over $\mathrm{dist}(\tilde{h}_0, G_0) = \|W^* \tilde{y}_0\|$. It is clear that $\tilde{y}_0$ and W are independent. On the event $[\|W\|^2_{\mathrm{HS}} \ge ar]$, we know by Lemma 3.5 that

$$s_{\lfloor \frac{a}{2-a} r \rfloor - 1}(W)^2 \ge a/2.$$

Moreover, since $0 \in \mathcal{C}_{\mathrm{good}}(\alpha)$, Proposition 3.1, with the help of Assumptions 1.1 and 1.3, shows that $n\mathbb{E}\tilde{y}_0\tilde{y}_0^* \ge (\alpha/2) I_N$ for all N large. Consequently, on the event $[\|W\|^2_{\mathrm{HS}} \ge ar]$, the conditional distribution of $W^*\tilde{y}_k$ with respect to the sigma-field $\sigma(W)$ is Gaussian with a covariance matrix $\Sigma \in \mathcal{H}^r_+$ such that $s_{\lfloor \frac{a}{2-a} r \rfloor - 1}(\Sigma) \ge a\alpha/(4n)$ for all large N. Put $t^2 = a\alpha \lfloor \frac{a}{2-a} r \rfloor/(8n) \sim r/n$. By what precedes and by Lemma 3.6(2), we have

$$\begin{aligned} &\mathbb{P}[[\mathrm{dist}(h_0, H_{\cdot,[N+n-k]\setminus\{0\}})^2 \le t^2] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad \le \mathbb{P}[[\mathrm{dist}(\tilde{h}_0, G_0)^2 \le t^2] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad \le \mathbb{P}[[\mathrm{dist}(\tilde{h}_0, G_0)^2 \le t^2] \cap [\|W\|^2_{\mathrm{HS}} \ge ar]] + \mathbb{P}[[\|W\|^2_{\mathrm{HS}} < ar] \cap \mathcal{E}_{\mathrm{op}}(C)] \\ &\quad \le \exp\left(-c_{3.6,2} \left\lfloor \frac{a}{2-a} r \right\rfloor \right). \end{aligned}$$

The proof of Proposition 3.4 is completed by using Inequality (3.20). □

Proof of Theorem 1.2. To complete the proof, we need to just combine Lemma 3.12 with the characterization (3.19), apply Propositions 3.2 and 3.4, respectively, to the two terms in this characterization, and use Lemma 3.7 to bound $\mathbb{P}[\mathcal{E}_{\mathrm{op}}(C)]$. □

4. Proofs Regarding the Singular Value Distribution of $\widehat{R}_L^{(N)} - z$

4.1. *Proof of Theorem* 1.3

We recall that $\mathcal{M}^m_+ = \{M \in \mathbb{C}^{m\times m},\ \Im M > 0\}$.

Lemma 4.1. *Given an integer $m > 0$, let $M \in \mathcal{M}_+^m$. Then M is invertible, and $\|M^{-1}\| \leq \|(\Im M)^{-1}\|$. Moreover, $-M^{-1} \in \mathcal{M}_+^m$.*

Proof. For each non-zero vector u, we have

$$\|u\|\|Mu\| \geq |u^* Mu| = |u^* \Re M u + \imath u^* \Im M u| \geq u^* \Im M u > 0,$$

and hence M is invertible. For an arbitrary non-zero vector v, there is $u \neq 0$ such that $v = Mu$, and we have from the former display that

$$\|M^{-1}v\|\|v\| \geq (M^{-1}v)^*(\Im M)M^{-1}v \geq \|(\Im M)^{-1}\|^{-1}\|M^{-1}v\|^2,$$

and hence the second result follows. We finally have

$$\Im(M^{-1}) = M^{-1}\frac{M^* - M}{2\imath}M^{-*} = -M^{-1}(\Im M)M^{-*} < 0. \qquad \square$$

The following result is well known. We provide its proof for completeness.

Lemma 4.2. *A probability measure $\check{\nu}$ is symmetric if and only if its Stieltjes transform $g_{\check{\nu}}$, seen as an analytic function on $\mathbb{C}\backslash\mathbb{R}$, satisfies $g_{\check{\nu}}(-\eta) = -g_{\check{\nu}}(\eta)$.*

Proof. The necessity is obvious from the definition of the Stieltjes transform and from the fact that $\check{\nu}(d\lambda) = \check{\nu}(-d\lambda)$. To prove the sufficiency, we use the Perron inversion formula, which says that for any function $\varphi \in C_c(\mathbb{R})$,

$$\int_{\mathbb{R}} \varphi(x)\check{\nu}(dx) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \Im g_{\check{\nu}}(x + \imath\varepsilon)dx.$$

By a simple variable change at the right-hand side, and by using the equalities $g_{\check{\nu}}(-\eta) = -g_{\check{\nu}}(\eta)$ and $g_{\check{\nu}}(\bar{\eta}) = \bar{g}_{\check{\nu}}(\eta)$, we obtain that $\int \varphi(x)\check{\nu}(dx) = \int \varphi(-x)\check{\nu}(dx)$, and hence the result follows. $\square$

The operator $\mathscr{T}$ introduced before Theorem 1.3 has the following properties.

Lemma 4.3. *Suppose $M \in \mathbb{C}^{2N\times 2N}$ and $S \in \mathcal{H}_+^N$. Then,*

$$\|\mathscr{T}((I_2 \otimes S)M)\| \leq \|M\| \operatorname{tr} S/n \leq (N/n)\|S\|\|M\|.$$

If $\Im M \geq \rho I_{2N}$ for some $\rho > 0$, then

$$\Im \mathscr{T}((I_2 \otimes S)M) \geq \rho n^{-1}(\operatorname{tr} S)I_2.$$

Proof. To obtain the first result, we write

$$\begin{aligned}\mathscr{T}((I_2 \otimes S)M) &= \mathscr{T}((I_2 \otimes S^{1/2})M(I_2 \otimes S^{1/2})) \\ &= \frac{1}{n}\sum_{\ell \in [N]} \begin{bmatrix} e_\ell^* S^{1/2} & \\ & e_\ell^* S^{1/2} \end{bmatrix} M \begin{bmatrix} S^{1/2}e_\ell & \\ & S^{1/2}e_\ell \end{bmatrix}.\end{aligned}$$

Hence,

$$\|\mathscr{T}((I_2\otimes S)M)\| \le n^{-1}\|M\| \sum_{\ell\in[N]} \|S^{1/2}e_\ell\|^2 = \|M\| \operatorname{tr} S/n.$$

To obtain the second inequality, just observe that $\Im\mathscr{T}((I_2\otimes S)M) = \mathscr{T}((I_2\otimes S)\Im M)$ and follow the same derivation as above. □

Proof of Theorem 1.3. Given a function $M(\eta) \in \mathfrak{S}^{2N}$, let

$$A(M(\eta), e^{\imath\theta}) = \mathscr{T}((I_2 \otimes \Sigma(e^{\imath\theta}))M(\eta)) + U_L(e^{\imath\theta}).$$

We shall show that $A(M(\eta), e^{\imath\theta})^{-1}$ exists, is holomorphic as a function of η, is continuous as a function of θ, and satisfies $\Im A(M(\eta), e^{\imath\theta})^{-1} \le 0$.

First let $\rho > 0$ be such that $\Im M(\eta) \ge \rho I$. By Lemma 4.3, $\Im A(M(\eta), e^{\imath\theta}) \ge \rho n^{-1} \operatorname{tr}\Sigma(e^{\imath\theta}) I_2$. Thus, if $\Sigma(e^{\imath\theta}) \ne 0$, then $A(M(\eta), e^{\imath\theta})^{-1}$ exists, is holomorphic in η, and $\Im A(M(\eta), e^{\imath\theta})^{-1} < 0$ by Lemma 4.1.

Otherwise, $A(M(\eta), e^{\imath\theta}) = U_L(e^{\imath\theta})$, which is trivially invertible, and $\Im A(M(\eta), e^{\imath\theta})^{-1} = 0$. In summary, $A(M(\eta), e^{\imath\theta})^{-1}$ is holomorphic in η and satisfies $\Im A(M(\eta), e^{\imath\theta})^{-1} \le 0$, for each $\theta \in [0, 2\pi)$. Moreover, the continuity of $A(M(\eta), e^{\imath\cdot})^{-1}$ follows from the continuity of $A(M(\eta), e^{\imath\cdot})$.

From these properties of A, it follows that

$$B(M(\eta)) = (2\pi)^{-1} \int_0^{2\pi} A(M(\eta), e^{\imath\theta})^{-1} \otimes \Sigma(e^{\imath\theta}) d\theta$$

exists, is holomorphic as a function of η, and satisfies $\Im B(M(\eta)) \le 0$. Since $\Im\left[\begin{smallmatrix}\eta & z\\ \bar{z} & \eta\end{smallmatrix}\right] = \Im\eta I_2$, we get by Lemma 4.1 that the function $\mathcal{F}_{\Sigma,z}(M(\eta),\eta)$ is holomorphic in $\eta \in \mathbb{C}_+$, and takes values in $\mathcal{M}_+^{2N}$. Furthermore, since $M \in \mathfrak{S}^{2N}$, it holds that $\|M(\eta)\| \le 1/(\Im\eta)$, and it is easy to show that $\lim_{t\to\infty} \imath t \mathcal{F}_{\Sigma,z}(M, \imath t) = -I_{2N}$. In summary, $\mathcal{F}_{\Sigma,z}(M(\eta),\eta) \in \mathfrak{S}^{2N}$ as a function of η when $M \in \mathfrak{S}^{2N}$.

Let us now establish the uniqueness of the solution of Eq. (1.3) in the class $\mathfrak{S}^{2N}$. Assume that $P(z,\cdot)$ and $P'(z,\cdot)$ are two such solutions. Then,

$$\begin{aligned}\mathcal{F}_{\Sigma,z}(P,\eta) - \mathcal{F}_{\Sigma,z}(P',\eta) &= \mathcal{F}_{\Sigma,z}(P,\eta)(B(P') - B(P))\mathcal{F}_{\Sigma,z}(P',\eta)\\ &= \mathcal{F}_{\Sigma,z}(P,\eta) \times \left\{\frac{1}{2\pi}\int_0^{2\pi} \left(A(P', e^{\imath\theta})^{-1}\mathscr{T}((I_2\otimes\Sigma(e^{\imath\theta}))\right.\right.\\ &\qquad \left.\left. \times (P - P'))A(P, e^{\imath\theta})^{-1}\right) \otimes \Sigma(e^{\imath\theta}) d\theta\right\} \times \mathcal{F}_{\Sigma,z}(P',\eta).\end{aligned}$$

Define the domain

$$\mathcal{D} = \left\{\eta \in \mathbb{C}_+ : \Im\eta > 4\|\Sigma\|_\infty \left(\frac{N}{n} \vee \sqrt{\frac{N}{n}}\right)\right\}.$$

Let $\eta \in \mathcal{D}$. Using the Inequality $\|P\|, \|P'\| \le 1/\Im\eta$, along with Lemma 4.3, it can be checked that $\|A(P, e^{\imath\theta})^{-1}\|, \|A(P', e^{\imath\theta})^{-1}\| \le 2$ when $\eta \in \mathcal{D}$. Using Lemma 4.3

again, we have

$$\|P - P'\| = \|\mathcal{F}_{\Sigma,z}(P,\eta) - \mathcal{F}_{\Sigma,z}(P',\eta)\| \le \frac{4}{(\Im\eta)^2}\frac{N}{n}\|\Sigma\|_\infty^2 \|P - P'\|$$
$$\le \frac{1}{2}\|P - P'\|, \tag{4.1}$$

which shows that $P(z,\eta) = P'(z,\eta)$ for $\eta \in \mathcal{D}$, and hence, for $\eta \in \mathbb{C}_+$.

To show the existence of the solution, set $P_0(z,\eta) = -\eta^{-1}I_{2N} \in \mathfrak{S}^{2N}$, and consider the iterations

$$P_{k+1}(z,\eta) = \mathcal{F}_{\Sigma,z}(P_k(z,\eta),\eta).$$

Then, $P_k(z,\cdot) \in \mathfrak{S}^{2N}$ for each k, and furthermore, the sequence $P_k(z,\eta)$ converges on $\mathcal{D}$ to a function $P_\infty(z,\eta)$ which satisfies $P_\infty(z,\eta) = \mathcal{F}_{\Sigma,z}(P_\infty(z,\eta),\eta)$ by Banach's fixed point theorem. Furthermore, given arbitrary vectors $a, b \in \mathbb{C}^{2N}$, the sequence of holomorphic functions $(a^* P_k(z,\cdot)b)_k$ on $\mathcal{D}$ is a *normal family*, thus, their limit $a^* P_\infty(z,\cdot)b$ is holomorphic on $\mathcal{D}$ by the normal family theorem. Since a and b are arbitrary, $P_\infty(z,\cdot)$ is a holomorphic matrix function on $\mathcal{D}$ that satisfies the properties of a matrix Stieltjes transform stated in Proposition 1.2. This shows that $P_\infty(z,\cdot)$ is the unique solution of Eq. (1.3) in $\mathfrak{S}^{2N}$.

It remains to establish the last result of Theorem 1.3. Extending the domain of $\mathcal{F}_{\Sigma,z}(P(z,\eta),\eta)$ in the parameter η to $\mathbb{C}\backslash\mathbb{R}$, we show that P is the solution of the equation $P = \mathcal{F}_{\Sigma,z}(P,\eta)$ if and only if the matrix function $P' = [\begin{smallmatrix} -P_{00} & P_{01} \\ P_{10} & -P_{11}\end{smallmatrix}]$ is the solution of the equation $P' = \mathcal{F}_\Sigma(P',-\eta)$. This can be demonstrated by a direct calculation: writing

$$T(e^{\imath\theta},z,\eta) = \mathscr{T}((I_2 \otimes \Sigma(e^{\imath\theta}))P(z,\eta)) = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11}\end{bmatrix},$$

we have

$$(T(e^{\imath\theta},z,\eta) + U_L(e^{\imath\theta}))^{-1} = \frac{1}{\Delta}\begin{bmatrix} t_{11} & -(t_{01} + e^{-\imath L\theta}) \\ -(t_{10} + e^{\imath L\theta}) & t_{00}\end{bmatrix},$$

with $\Delta = t_{00}t_{11} - (t_{01} + e^{-\imath L\theta})(t_{10} + e^{\imath L\theta})$. Thus,

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11}\end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2\pi}\int \frac{t_{11}}{\Delta}\Sigma - \eta & -\frac{1}{2\pi}\int \frac{(t_{01} + e^{-\imath L\theta})}{\Delta}\Sigma - z \\ -\frac{1}{2\pi}\int \frac{(t_{10} + e^{\imath L\theta})}{\Delta}\Sigma - \bar{z} & \frac{1}{2\pi}\int \frac{t_{00}}{\Delta}\Sigma - \eta \end{bmatrix}^{-1}.$$

Recalling the formula for the inverse of a partitioned matrix (see [20, §0.7.3])

$$\begin{bmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{bmatrix}^{-1} = \begin{bmatrix} (M_{00} - M_{01}M_{11}^{-1}M_{10})^{-1} & -M_{00}^{-1}M_{01}(M_{11} - M_{10}M_{00}^{-1}M_{01})^{-1} \\ -(M_{11} - M_{10}M_{00}^{-1}M_{01})^{-1}M_{10}M_{00}^{-1} & (M_{11} - M_{10}M_{00}^{-1}M_{01})^{-1} \end{bmatrix}$$

we get the required result by a direct checking.

Using this result in conjunction with Lemma 4.2, we obtain that for each deterministic vector $u \in \mathbb{C}^N$, the scalar measures $(u^*\Lambda_{ii}u)(dt)$ are symmetric for $i = 0, 1$. This shows at once that the measures Λ_{ii} are symmetric. The proof of Theorem 1.3 is now complete. □

4.2. *Proof of Theorem* 1.4

Recall that $\boldsymbol{M}$ is the bound provided by Assumption 1.1(ii). From now on, we shall write $\boldsymbol{\gamma}_{\sup} = \sup_N N/n$.

We first establish the tightness of the sequence $(\check{\boldsymbol{\nu}}_{z,N})_N$. As is well known [14], this is equivalent to showing that $-\imath t g_{\check{\boldsymbol{\nu}}_{z,N}}(\imath t) \to 1$ as $t \to \infty$ uniformly in N.

Given $M \in \mathcal{M}_+^{2N}$, let

$$A(M, e^{\imath\theta}) \triangleq \mathscr{T}((I_2 \otimes S(e^{\imath\theta}))M) + U_L(e^{\imath\theta}).$$

By Lemma 4.3,

$$\|\mathscr{T}((I_2 \otimes S(e^{\imath\theta}))G(z, \imath t))\| \leq \boldsymbol{\gamma}_{\sup}\boldsymbol{M}/t.$$

Thus, for $t \geq 2\boldsymbol{\gamma}_{\sup}\boldsymbol{M}$, $\|A(G(z, \imath t), e^{\imath\theta})^{-1}\| \leq 2$. Since

$$\begin{aligned} G(z, \imath t) &= \mathcal{F}_{S,z}(G(z, \imath t), \imath t) \\ &= \left(\frac{1}{2\pi}\int A(G(z, \imath t), e^{\imath\theta})^{-1} \otimes S(e^{\imath\theta})d\theta - \begin{bmatrix} \imath t & z \\ \bar{z} & \imath t \end{bmatrix} \otimes I_N\right)^{-1}, \end{aligned}$$

it is clear that

$$\frac{-\imath t}{2N} \operatorname{tr} G(z, \imath t) \xrightarrow[t\to\infty]{} 1$$

uniformly in N, thus, the sequence $(\check{\boldsymbol{\nu}}_{z,N})_N$ is tight.

The remainder of the proof is devoted towards establishing the convergence (1.4). Recalling the expression $\widehat{R}_L = XJ^LX^*$ provided at the beginning of Sec. 3.1, we have

$$Q(z,\eta) = (\boldsymbol{H}(\widehat{R}_L - z) - \eta)^{-1} = \begin{bmatrix} -\eta & XJ^LX^* - z \\ XJ^{-L}X^* - \bar{z} & -\eta \end{bmatrix}^{-1} = \begin{bmatrix} Q_{00} & Q_{01} \\ Q_{10} & Q_{11} \end{bmatrix}.$$

We begin by bounding the variance of $(2N)^{-1} \operatorname{tr} DQ$ at the left-hand side of (1.4).

Proposition 4.1. *Under Assumption* 1.1(ii), *for each deterministic matrix* $B \in \mathbb{C}^{N\times N}$ *and each* $u, v \in \{0,1\}$,

$$\mathbb{V}\mathrm{ar}(\operatorname{tr} BQ_{uv}(z,\eta)) \le 8\boldsymbol{\gamma}_{\sup}\boldsymbol{M}^2\|B\|^2/(\Im\eta)^4.$$

The proof of this proposition will be based on the well-known Poincaré–Nash (PN) inequality [10, 26], which is also a particular case of the Brascamp–Lieb inequality. Let $\boldsymbol{v} = [v_0, \ldots, v_{m-1}]^\mathsf{T}$ be a complex Gaussian random vector with $\mathbb{E}\boldsymbol{v} = 0$, $\mathbb{E}\boldsymbol{v}\boldsymbol{v}^\mathsf{T} = 0$, and $\mathbb{E}[\boldsymbol{v}\boldsymbol{v}^*] = \Sigma$. Let $\varphi = \varphi(v_0, \ldots, v_{m-1}, \bar{v}_0, \ldots, \bar{v}_{m-1})$ be a C^1 complex function which is polynomially bounded together with its derivatives. Then, writing

$$\nabla_{\boldsymbol{v}}\varphi = [\partial\varphi/\partial v_0, \ldots, \partial\varphi/\partial v_{m-1}]^\mathsf{T} \quad \text{and} \quad \nabla_{\bar{\boldsymbol{v}}}\varphi = [\partial\varphi/\partial \bar{v}_0, \ldots, \partial\varphi/\partial \bar{v}_{m-1}]^\mathsf{T},$$

the *PN inequality* is

$$\mathbb{V}\mathrm{ar}(\varphi(\boldsymbol{v})) \le \mathbb{E}[\nabla_{\boldsymbol{v}}\varphi(\boldsymbol{v})^T \ \Sigma \ \overline{\nabla_{\boldsymbol{v}}\varphi(\boldsymbol{v})}] + \mathbb{E}[(\nabla_{\bar{\boldsymbol{v}}}\varphi(\boldsymbol{v}))^* \ \Sigma \ \nabla_{\bar{\boldsymbol{v}}}\varphi(\boldsymbol{v})]. \tag{4.2}$$

If we rewrite $Q(z,\eta)$ as $Q(z,\eta) = Q^X$ to emphasize the dependence of the resolvent on the matrix X, then, given a matrix $\Delta \in \mathbb{C}^{N\times n}$, the resolvent identity implies that

$$\begin{aligned} Q^{X+\Delta} - Q^X &= -Q^{X+\Delta}((Q^{X+\Delta})^{-1} - (Q^X)^{-1})Q^X \\ &= -Q^{X+\Delta} \\ &\quad \times \begin{bmatrix} & (X+\Delta)J^L(X+\Delta)^* - XJ^LX^* \\ (X+\Delta)J^{-L}(X+\Delta)^* - XJ^{-L}X^* & \end{bmatrix} Q^X. \end{aligned}$$

Using this equation, we can obtain the expression of $\partial a^*Q_{uv}b/\partial\bar{x}_{ij}$, where $a, b \in \mathbb{C}^N$, $u, v \in \{0,1\}$, $i \in [N]$, and $j \in [n]$. Indeed, taking $\Delta = e_{N,i}e_{n,j}^*$ in the former expression, we get after a simple derivation that

$$\frac{\partial a^*Q_{uv}b}{\partial\bar{x}_{ij}} = -[a^*Q_{u1}XJ^{-L}]_j[Q_{0v}b]_i - [a^*Q_{u0}XJ^L]_j[Q_{1v}b]_i. \tag{4.3}$$

Proof of Proposition 4.1. We apply Inequality (4.2) by, respectively, replacing $\boldsymbol{v}$ and φ with $\operatorname{vec} X$ and $\operatorname{tr} BQ_{uv}$ (seen as a function of X).

Given $k, i \in [N]$, and $j \in [n]$, we have

$$\frac{\partial [e_k^* BQ_{uv} e_k]}{\partial \bar{x}_{ij}} = -[BQ_{u1} X J^{-L}]_{kj} [Q_{0v}]_{ik} - [BQ_{u0} X J^L]_{kj} [Q_{1v}]_{ik},$$

and thus,

$$\frac{\partial \operatorname{tr} BQ_{uv}}{\partial \bar{x}_{ij}} = -[Q_{0v} BQ_{u1} X J^{-L}]_{ij} - [Q_{1v} BQ_{u0} X J^L]_{ij}.$$

Let us focus on the second term at the right-hand side of Inequality (4.2). Observing that $\mathbb{E}[x_{i_1 j_1} \bar{x}_{i_2 j_2}] = n^{-1} [R_{j_1 - j_2}]_{i_1, i_2}$, and recalling the expression of the block-Toeplitz matrix $\mathcal{R}$ given by Eq. (3.5), we have

$$\begin{aligned}
&\sum_{i_1, i_2 \in [N]} \sum_{j_1, j_2 \in [n]} \mathbb{E}\left[\overline{\frac{\partial \operatorname{tr} BQ_{uv}}{\partial \bar{x}_{i_1 j_1}}} \mathbb{E}[x_{i_1 j_1} \bar{x}_{i_2 j_2}] \frac{\partial \operatorname{tr} BQ_{uv}}{\partial \bar{x}_{i_2 j_2}} \right] \\
&\quad \le 2\mathbb{E} \operatorname{vec}(Q_{0v} BQ_{u1} X J^{-L})^* \mathcal{R} \operatorname{vec}(Q_{0v} BQ_{u1} X J^{-L}) \\
&\qquad + 2\mathbb{E} \operatorname{vec}(Q_{1v} BQ_{u0} X J^{L})^* \mathcal{R} \operatorname{vec}(Q_{1v} BQ_{u0} X J^{L}) \\
&\quad \le \frac{2\boldsymbol{M}}{n} (\mathbb{E}\|Q_{0v} BQ_{u1} X J^{-L}\|_{\mathrm{HS}}^2 + \mathbb{E}\|Q_{1v} BQ_{u0} X J^{L}\|_{\mathrm{HS}}^2) \\
&\quad \le \frac{4\boldsymbol{M}\|B\|^2}{(\Im \eta)^4 n} \mathbb{E}\|X\|_{\mathrm{HS}}^2 \\
&\quad \le \frac{4\gamma_{\sup} \boldsymbol{M}^2 \|B\|^2}{(\Im \eta)^4}.
\end{aligned}$$

The first term at the right-hand side of Inequality (4.2) is treated similarly, leading to the bound given in the statement, and the proof is complete. ☐

In order to establish the convergence (1.4), using Proposition 4.1 and the Borel–Cantelli lemma, it will be enough to show that

$$\forall \eta \in \mathbb{C}_+, \quad \frac{1}{2N} \operatorname{tr} D^{(N)} (\mathbb{E} Q^{(N)}(z, \eta) - G^{(N)}(z, \eta)) \xrightarrow[N \to \infty]{} 0. \tag{4.4}$$

Following the general canvas of [28], we approximate our process $\boldsymbol{x}^{(N)}$ by a Moving Average process with a finite memory. We shall construct from this MA process a resolvent that will be more easily manageable than $Q^{(N)}$.

With reference to Theorem 1.3, first consider a discrete analogue of the integral within the expression of $\mathcal{F}_{\Sigma,z}$. A straightforward adaptation of its proof yields the following proposition and we omit the details.

Proposition 4.2. *Let $\Sigma : \mathbb{T} \to \mathcal{H}_+^N$ be a continuous function, and let $z \in \mathbb{C}$. Then, the conclusions of Theorem* 1.3 *remain true if the function $\mathcal{F}_{\Sigma,z}$ there is replaced* *with*

$$\check{\mathcal{F}}_{\Sigma,z}(M(\eta),\eta) = \Bigg(\frac{1}{n}\sum_{\ell\in[n]}\Bigg(\mathscr{T}((I_2\otimes\Sigma(e^{2\imath\pi\ell/n}))M(\eta)) + \begin{bmatrix} & e^{-2\imath\pi\ell L/n} \\ e^{-2\imath\pi\ell L/n} & \end{bmatrix}\Bigg)^{-1} \otimes \Sigma(e^{2\imath\pi\ell/n}) - \begin{bmatrix}\eta & z\\ \bar{z} & \eta\end{bmatrix}\otimes I_N\Bigg)^{-1}.$$

Now, given an integer constant $K>0$, let us define the function $\widehat{S}^{(N,K)}$ on $\mathbb{T}$ as

$$\widehat{S}^{(N,K)}(e^{\imath\theta}) = \frac{1}{2\pi}\int_0^{2\pi} F_K(e^{\imath(\theta-\psi)})S^{(N)}(e^{\imath\psi})\,d\psi,$$

where F_K is the Fejér kernel (see Eq. (3.7)). This function has the following properties:

(1) By the non-negativity of the Fejér kernel, $\widehat{S}^{(N,K)}$ is a spectral density.
(2) By replacing $F_K(e^{\imath(\theta-\psi)})$ with the first expression of this kernel provided by (3.7), and by developing the integrand above, we obtain that $\widehat{S}^{(N,K)}$ is a Laurent trigonometric polynomial of the form $\widehat{S}^{(N,K)}(e^{\imath\theta}) = \sum_{\ell=-K}^{K} e^{\imath\ell\theta}\widetilde{R}_\ell^{(N,K)}$.
(3) With Assumptions 1.1(i) and 1.1(ii), we have

$$\sup_N \|\widehat{S}^{(N,K)} - S^{(N)}\|_\infty^{\mathbb{T}} \xrightarrow[K\to\infty]{} 0. \tag{4.5}$$

Relation (4.5) can be established by splitting the integral that defines $\widehat{S}^{(N,K)}(e^{\imath\theta})$ into two pieces as $\int_0^{2\pi} = \int_{\psi:|\theta-\psi|\le\delta} + \int_{\psi:|\theta-\psi|>\delta}$ for a properly chosen $\delta>0$, and by using the properties of the Fejér kernel provided after Eq. (3.7).

Consider the implicit equation

$$\widehat{G}^{(N,K)}(z,\eta) = \check{\mathcal{F}}_{\widehat{S}^{(N,K)},z}(\widehat{G}^{(N,K)}(z,\eta),\eta) \quad (\text{in } \mathfrak{S}^{2N}).$$

From Proposition 4.2, the solution $\widehat{G}^{(N,K)}(z,\cdot)$ exists and is unique. The following three propositions will be proved in Sec. 4.3.

Proposition 4.3. *For each η such that $\Im\eta \ge C_{4.3}$ with $C_{4.3} = C_{4.3}(\boldsymbol{\gamma}_{\sup}, \boldsymbol{M}) > 0$,*

$$\limsup_N \|\widehat{G}^{(N,K)}(z,\eta) - G^{(N)}(z,\eta)\| \xrightarrow[K\to\infty]{} 0.$$

Let $(\hat{x}_k^{(N,K)})_{k\in\mathbb{Z}}$ be a $\mathbb{C}^N$-valued stationary centered Gaussian process with the spectral density $n^{-1}\widehat{S}^{(N,K)}$. Define

$$\widehat{X}^{(N,K)} = [\hat{x}_0^{(N,K)} \quad \cdots \quad \hat{x}_{n-1}^{(N,K)}] \text{ and}$$
$$\widehat{Q}^{(N,K)}(z,\eta) = (\boldsymbol{H}(\widehat{X}^{(N,K)}J^L(\widehat{X}^{(N,K)})^* - z) - \eta I)^{-1}.$$

We then have the following proposition.

Proposition 4.4. *Fix $K > 0$. Then, for an arbitrary deterministic matrix $D^{(N)} \in \mathbb{C}^{2N\times 2N}$ with $\|D^{(N)}\| = 1$, we have*

$$\frac{1}{N}|\operatorname{tr} D^{(N)}(\mathbb{E}\widehat{Q}^{(N,K)}(z,\eta) - \widehat{G}^{(N,K)}(z,\eta))| \leq \frac{CK}{\sqrt{N}},$$

where $C > 0$ depends only on $\eta, \boldsymbol{M}$, and $\boldsymbol{\gamma}_{\sup}$.

We note here that the bound provided in the statement of this proposition is not optimal but is good enough for our purpose.

Proposition 4.5. *With $D^{(N)}$ as in the previous proposition,*

$$\limsup_N \frac{1}{N}|\operatorname{tr} D^{(N)}(\mathbb{E}Q^{(N)}(z,\eta) - \mathbb{E}\widehat{Q}^{(N,K)}(z,\eta))| \xrightarrow[K\to\infty]{} 0.$$

Theorem 1.4: end of the proof.

We write

$$\begin{aligned}\frac{1}{N}\operatorname{tr} D^{(N)}(\mathbb{E}Q^{(N)} - G^{(N)}) &= \frac{1}{N}\operatorname{tr} D^{(N)}(\mathbb{E}Q^{(N)} - \mathbb{E}\widehat{Q}^{N,K}) + \frac{1}{N}\operatorname{tr} D^{(N)} \\ &\quad \times (\mathbb{E}\widehat{Q}^{N,K} - \widehat{G}^{(N,K)}) + \frac{1}{N}\operatorname{tr} D^{(N)}(\widehat{G}^{(N,K)} - G^{(N)}) \\ &\triangleq \chi_1(N,K) + \chi_2(N,K) + \chi_3(N,K).\end{aligned}$$

Fix an arbitrarily small $\varepsilon > 0$. Let $K_0, N_0 > 0$ be such that, by Propositions 4.5 and 4.3,

$$|\chi_1(N,K_0)|, |\chi_3(N,K_0)| \leq \varepsilon \quad \text{for all } N \geq N_0 \text{ and } \Im\eta > C_{4.3}.$$

By Proposition 4.4, $\chi_2(N,K_0) \to_{N\to\infty} 0$. Thus, $(2N)^{-1}\operatorname{tr} D^{(N)}(\mathbb{E}Q^{(N)} - G^{(N)}) \to_N 0$ first for $\Im\eta > C_{4.3}$, and hence for each $\eta \in \mathbb{C}_+$ by analyticity. Thus (4.4) is established. This completes the proof of Theorem 1.4.

4.3. *Remaining proofs for Sec. 4.2*

Proof of Proposition 4.3. Let $\bar{S}^{(N,K)}(e^{\imath\theta}) = \widehat{S}^{(N,K)}(e^{2\imath\pi k/n})$ and $\bar{U}_L(e^{\imath\theta}) = U_L(e^{2\imath\pi k/n})$ for $\theta \in [2\pi k/n, 2\pi(k+1)/n)$ be the respective stepwise continuous versions of the functions $\widehat{S}^{(N,K)}$ and U_L with step size $2\pi/n$. Within this proof, we re-denote the function $\mathcal{F}_{\Sigma,z}$ defined in the statement of Theorem 1.3 as $\mathcal{F}_{\Sigma,z,U_L}$ to stress the dependence on U_L. With this notation, it is obvious that $\breve{\mathcal{F}}_{\widehat{S}^{(N,K)},z} = \mathcal{F}_{\bar{S}^{(N,K)},z,\bar{U}_L}$.

In the rest of the proof, we often drop the superscripts $^{(N)}$ and $^{(N,K)}$ and the subscript $_L$ for brevity. Given $M \in \mathfrak{S}^{2N}$, put $A_{\boldsymbol{S},\boldsymbol{U}}(M,e^{\imath\theta}) \triangleq \mathscr{T}((I_2 \otimes \boldsymbol{S}(e^{\imath\theta}))M) + \boldsymbol{U}(e^{\imath\theta})$, where $(\boldsymbol{S},\boldsymbol{U}) = (S,U)$ or $(\bar{S},\bar{U})$. Write $B_{\boldsymbol{S},\boldsymbol{U}}(M) = (2\pi)^{-1}\int_0^{2\pi} A_{\boldsymbol{S},\boldsymbol{U}}(M,e^{\imath\theta})^{-1} \otimes \boldsymbol{S}(e^{\imath\theta})d\theta$.

We also assume that K is large enough so that

$$\sup_N \|\bar{S}\|_\infty^{\mathbb{T}} \le 2\boldsymbol{M}.$$

By dropping the unnecessary parameters from the notations, we write

$$\begin{aligned}
G - \widehat{G} &= \mathcal{F}_{S,U}(G) - \mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}) = \mathcal{F}_{S,U}(G)(B_{\bar{S},\bar{U}}(\widehat{G}) - B_{S,U}(G))\mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}) \\
&= \mathcal{F}_{S,U}(G)\left(\frac{1}{2\pi}\int_0^{2\pi}(A_{\bar{S},\bar{U}}(\widehat{G}, e^{\imath\theta})^{-1} \otimes \bar{S}(e^{\imath\theta}) \right. \\
&\quad \left. - A_{S,U}(G, e^{\imath\theta})^{-1} \otimes S(e^{\imath\theta}))d\theta\right)\mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}) \\
&= \mathcal{F}_{S,U}(G)\left(\frac{1}{2\pi}\int (A_{\bar{S},\bar{U}}(\widehat{G})^{-1} \otimes \bar{S} - A_{S,U}(\widehat{G})^{-1} \otimes S \right. \\
&\quad \left. + (A_{S,U}(\widehat{G})^{-1} - A_{S,U}(G)^{-1}) \otimes S)d\theta\right)\mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}) \\
&= \mathcal{F}_{S,U}(G)\left(\frac{1}{2\pi}\int (A_{\bar{S},\bar{U}}(\widehat{G})^{-1} \otimes (\bar{S} - S) + (A_{\bar{S},\bar{U}}(\widehat{G})^{-1} - A_{S,U}(\widehat{G})^{-1}) \otimes S \right. \\
&\quad \left. + (A_{S,U}(\widehat{G})^{-1} - A_{S,U}(G)^{-1}) \otimes S)d\theta\right)\mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}) \\
&\triangleq \mathcal{F}_{S,U}(G)\left(\frac{1}{2\pi}\int (\chi_1 + \chi_2 + \chi_3)d\theta\right)\mathcal{F}_{\bar{S},\bar{U}}(\widehat{G}).
\end{aligned}$$

By Lemma 4.3, we have that $\|\mathscr{T}((I_2 \otimes \boldsymbol{S}(e^{\imath\theta}))\boldsymbol{G})\| \le 2\boldsymbol{\gamma}_{\sup}\boldsymbol{M}/\Im\eta$ for any of the possibilities for $\boldsymbol{S}$ and for $\boldsymbol{G} = G, \widehat{G}$. Thus, for

$$2\boldsymbol{\gamma}_{\sup}\boldsymbol{M}/\Im\eta \le 1/2,$$

we have $\|A_{\boldsymbol{S}}(\boldsymbol{G}, e^{\imath\theta})^{-1}\| \le 2$. Therefore,

$$\|\chi_1\| \le 2\|\bar{S} - S\|_\infty^{\mathbb{T}}.$$

Moreover,

$$\begin{aligned}
\chi_2 &= (A_{\bar{S},\bar{U}}(\widehat{G})^{-1} - A_{S,U}(\widehat{G})^{-1}) \otimes S \\
&= (A_{\bar{S},\bar{U}}(\widehat{G})^{-1}(\mathscr{T}((I_2 \otimes (\bar{S} - S))\widehat{G}) + U - \bar{U})A_{S,U}(\widehat{G})^{-1}) \otimes S
\end{aligned}$$

satisfies

$$\|\chi_2\| \le 4\boldsymbol{M}\boldsymbol{\gamma}_{\sup}(\Im\eta)^{-1}\|\bar{S} - S\|_\infty^{\mathbb{T}} + 4\boldsymbol{M}\|U - \bar{U}\|$$

for the same values of η. By mimicking the calculation that lead to Inequality (4.1), we obtain

$$\|\chi_3\| \le \frac{4}{(\Im\eta)^2}\boldsymbol{\gamma}_{\sup}\boldsymbol{M}^2\|G - \widehat{G}\|.$$

Using the inequality $\|\mathcal{F}_{\boldsymbol{S}}(\boldsymbol{G})\| \leq 1/(\Im\eta)$, we thus arrive at

$$\left(1 - \frac{4\boldsymbol{\gamma}_{\sup}\boldsymbol{M}^2}{(\Im\eta)^4}\right)\|G^{(N)} - \widehat{G}^{(N,K)}\| \leq \frac{2}{(\Im\eta)^2}\left(1 + \frac{4\boldsymbol{\gamma}_{\sup}\boldsymbol{M}}{\Im\eta}\right)\|\bar{S}^{(N,K)} - S^{(N)}\|_\infty^{\mathbb{T}} + \frac{4\boldsymbol{M}}{(\Im\eta)^2}\|\bar{U}_L - U_L\|_\infty^{\mathbb{T}}.$$

Thus, if $\Im\eta > (8\boldsymbol{\gamma}_{\sup}\boldsymbol{M}^2)^{1/4} \vee (4\boldsymbol{\gamma}_{\sup}\boldsymbol{M})$, then

$$\|G^{(N)} - \widehat{G}^{(N,K)}\| \leq C\|\bar{S}^{(N,K)} - S^{(N)}\|_\infty^{\mathbb{T}} + C'\|\bar{U}_L - U_L\|_\infty^{\mathbb{T}}$$

for some constants $C, C' > 0$. Now the proof can be completed by using the convergence (4.5). □

Proof of Proposition 4.4. From now on, C is a positive constant that depends on η, $\boldsymbol{M}$, and $\boldsymbol{\gamma}_{\sup}$ at most, and can change from line to line. Recalling Properties 1 and 2 of the density $\widehat{S}^{(N,K)}$ that were stated in Sec. 4.2, our first step is to apply the well-known operator version of the Fejér–Riesz theorem (see [29, Sec. 6.6]) to $\widehat{S}^{(N,K)}$. This implies that for each (N, K), there exists an $N \times N$ matrix trigonometric polynomial

$$P^{(N,K)}(e^{\imath\theta}) = \sum_{\ell=0}^{K} e^{\imath\ell\theta} B_\ell^{(N,K)}$$

such that $\widehat{S}^{(N,K)}(e^{\imath\theta}) = P^{(N,K)}(e^{\imath\theta})P^{(N,K)}(e^{\imath\theta})^*$.

Letting $\boldsymbol{\xi}^{(N)} = (\boldsymbol{\xi}_k^{(N)})_{k\in\mathbb{Z}}$ be an i.i.d. process with $\boldsymbol{\xi}_k^{(N)} \sim \mathcal{N}_{\mathbb{C}}(0, I_N)$, the process $(\hat{x}_k^{(N,K)})_{k\in\mathbb{Z}}$ that we used to construct the resolvent $\widehat{Q}^{(N,K)}$ can be defined as

$$\hat{x}_k^{(N,K)} = \frac{1}{\sqrt{n}}\sum_{\ell=0}^{K} B_\ell^{(N,K)}\boldsymbol{\xi}_{k-\ell}^{(N)}.$$

Define the finite sequence of random vectors $(\tilde{x}_k^{(N,K)})_{k\in[n]}$ as

$$\tilde{x}_k^{(N,K)} = \frac{1}{\sqrt{n}}\sum_{\ell=0}^{K} B_\ell^{(N,K)}\boldsymbol{\xi}_{(k-\ell) \bmod n}^{(N)}$$

(thus, $\tilde{x}_k^{(N,K)}$ is the analogue of $\hat{x}_k^{(N,K)}$ obtained through a circular convolution). Define

$$\widetilde{X}^{(N,K)} = [\tilde{x}_0^{(N,K)} \quad \cdots \quad \tilde{x}_{n-1}^{(N,K)}] \in \mathbb{C}^{N\times n},$$

and

$$\widetilde{Q}^{(N,K)}(z, \eta) = (\boldsymbol{H}(\widetilde{X}^{(N,K)}J^L(\widetilde{X}^{(N,K)})^* - z) - \eta I)^{-1}.$$

Observing that $\operatorname{rank}(\widehat{X} - \widetilde{X}) \leq K$, we get

$$\operatorname{rank}(\boldsymbol{H}(\widehat{X} J^L \widehat{X}^* - z) - \boldsymbol{H}(\widetilde{X} J^L \widetilde{X}^* - z)) \leq 4K.$$

Thus, for each matrix $D \in \mathbb{C}^{2N \times 2N}$, the inequality

$$|\operatorname{tr} D(\widehat{Q}(z, \eta) - \widetilde{Q}(z, \eta))| \leq \frac{4K\|D\|}{\Im \eta} \tag{4.6}$$

holds (see [34, Lemma 2.6]). We can thus work with $\widetilde{Q}$ in place of $\widehat{Q}$ for establishing Proposition 4.4.

For $k \in [n]$, let

$$w_k^{(N,K)} = \frac{1}{\sqrt{n}} \sum_{\ell=0}^{n-1} e^{2\imath\pi k\ell/n} \tilde{x}_\ell^{(N,K)}$$

be the discrete Fourier transform of the finite sequence $(\tilde{x}_0, \ldots, \tilde{x}_{n-1})$, and define the matrix

$$W^{(N,K)} = [w_0^{(N,K)} \quad \cdots \quad w_{n-1}^{(N,K)}] = \widetilde{X}^{(N,K)} \mathsf{F} \in \mathbb{C}^{N \times n},$$

where F is the Fourier matrix defined in (3.1). Since the $\tilde{x}_\ell$ are built through a circular convolution, the vectors w_k are independent, and $w_k \sim \mathcal{N}_{\mathbb{C}}(0, n^{-1}\widehat{S}(e^{2\imath\pi k/n}))$. Recalling that $J = \mathsf{F}\Omega\mathsf{F}^*$ with $\Omega = \operatorname{diag}(\omega^\ell)_{\ell=0}^{n-1}$ and $\omega = \exp(-2\imath\pi/n)$, we can write

$$\widetilde{Q}(z, \eta) = (\boldsymbol{H}(\widetilde{X} J^L \widetilde{X}^* - z) - \eta I)^{-1} = (\boldsymbol{H}(W\Omega^L W^* - z) - \eta I)^{-1}.$$

The remainder of the proof will be devoted towards showing that

$$\|\mathbb{E}\widetilde{Q}^{(N,K)}(z, \eta) - \widehat{G}^{(N,K)}(z, \eta)\| \leq CN^{-1/2}, \tag{4.7}$$

taking advantage of the independence of the columns of W. This bound, used in conjunction with the bound (4.6), immediately leads to the result of Proposition 4.4.

The proof of Inequality (4.7) relies on the NP inequality that we used above, as well as on the well-known Integration by Parts (IP) formula for Gaussian functionals [17, 22]. Recalling the definition of the vector $\boldsymbol{v}$ and the functional φ after Proposition 4.1, the IP formula reads as

$$\mathbb{E} v_k \varphi(\boldsymbol{v}) = \sum_{\ell=0}^{n-1} [\Sigma]_{k\ell} \mathbb{E}\left[\frac{\partial \varphi(\boldsymbol{v})}{\partial \bar{v}_\ell}\right].$$

Write $\widetilde{Q}(z, \eta) = \left[\begin{smallmatrix} \widetilde{Q}_{00} & \widetilde{Q}_{01} \\ \widetilde{Q}_{10} & \widetilde{Q}_{11} \end{smallmatrix}\right]$, and $W = [w_{ij}]_{i \in [N], j \in [n]}$. By reproducing verbatim the derivation that we made to obtain the Identity (4.3), we have

$$\frac{\partial a^* \widetilde{Q}_{uv} b}{\partial \bar{w}_{ij}} = -[a^* \widetilde{Q}_{u1} W \Omega^{-L}]_j [\widetilde{Q}_{0v} b]_i - [a^* \widetilde{Q}_{u0} W \Omega^L]_j [\widetilde{Q}_{1v} b]_i.$$

In the subsequent derivations, we write $\widehat{S}_k = \widehat{S}(e^{2\imath\pi k/n})$. Let a, b be two constant vectors in $\mathbb{C}^N$. Write $b = [b_0, \ldots, b_{N-1}]^\mathsf{T}$, and let $\alpha_{uv}(\ell) = \mathbb{E}[[a^* \widetilde{Q}_{uv} W]_\ell [W^* b]_\ell]$. By the IP formula, we have

$$\begin{aligned}\alpha_{uv}(\ell) &= \sum_{i\in[N]} \mathbb{E}[[a^*\widetilde{Q}_{uv}]_i w_{i\ell}[W^*b]_\ell] \\ &= \frac{1}{n}\sum_{i,m\in[N]} [\widehat{S}_\ell]_{im}\mathbb{E}\frac{\partial([a^*\widetilde{Q}_{uv}]_i[W^*b]_\ell)}{\partial \bar{w}_{m\ell}} \\ &= \frac{1}{n}\sum_{i,m}[\widehat{S}_\ell]_{im}\mathbb{E}[-[a^*\widetilde{Q}_{u1}W\Omega^{-L}]_\ell[\widetilde{Q}_{0v}]_{mi}[W^*b]_\ell \\ &\quad - [a^*\widetilde{Q}_{u0}W\Omega^L]_\ell[\widetilde{Q}_{1v}]_{mi}[W^*b]_\ell + [a^*\widetilde{Q}_{uv}]_i b_m] \\ &= -\mathbb{E}\left[[a^*\widetilde{Q}_{u1}W]_\ell[W^*b]_\ell\omega^{-\ell L}\frac{\operatorname{tr}\widetilde{Q}_{0v}\widehat{S}_\ell}{n}\right] \\ &\quad - \mathbb{E}\left[[a^*\widetilde{Q}_{u0}W]_\ell[W^*b]_\ell\omega^{\ell L}\frac{\operatorname{tr}\widetilde{Q}_{1v}\widehat{S}_\ell}{n}\right] + \frac{\mathbb{E}[a^*\widetilde{Q}_{uv}\widehat{S}_\ell b]}{n}.\end{aligned}$$

Write $\widetilde{qs}_{uv}(\ell) = \mathbb{E}\operatorname{tr}\widetilde{Q}_{uv}\widehat{S}_\ell/n$. We shall isolate the terms $\widetilde{qs}_{0v}(\ell)$ and $\widetilde{qs}_{1v}(\ell)$ in the last display by resorting to the following lemma. □

Lemma 4.4. *It holds that* $\mathbb{V}\mathrm{ar}\,\widetilde{qs}_{uv}(\ell) \le C/n^2$ *and* $\mathbb{V}\mathrm{ar}[a^*\widetilde{Q}_{uv}W]_\ell[W^*b]_\ell \le C\|a\|^2\|b\|^2/n$.

Proof. The first bound is obtained by repeating almost word for word the proof of Proposition 4.1. To obtain the second bound, we also use the NP inequality again. We start by writing

$$\begin{aligned}\frac{\partial[a^*\widetilde{Q}_{uv}W]_\ell[W^*b]_\ell}{\partial\bar{w}_{ij}} &= \sum_k \frac{\partial[a^*\widetilde{Q}_{uv}]_k w_{k\ell}[W^*b]_\ell}{\partial\bar{w}_{ij}} \\ &= [a^*\widetilde{Q}_{uv}W]_\ell b_i \mathbb{1}_{j=\ell} - [a^*\widetilde{Q}_{u1}W\Omega^{-L}]_j[Q_{0v}W]_{i\ell}[W^*b]_\ell \\ &\quad - [a^*\widetilde{Q}_{u0}W\Omega^L]_j[Q_{1v}W]_{i\ell}[W^*b]_\ell.\end{aligned}$$

We focus on the second term at the right-hand side of Inequality (4.2), treating separately the three terms at the right-hand side of the last display. Starting with the first term, we get

$$\begin{aligned}&\frac{1}{n}\sum_{i_1,i_2\in[N]}\mathbb{E}\overline{[a^*\widetilde{Q}_{uv}W]_\ell b_{i_1}}[\widehat{S}_\ell]_{i_1,i_2}[a^*\widetilde{Q}_{uv}W]_\ell b_{i_2} \\ &= \frac{1}{n}b^*\widehat{S}_\ell b\,\mathbb{E}|[a^*\widetilde{Q}_{uv}W]_\ell|^2 \le \|a\|^2\|b\|^2 C/n.\end{aligned}$$

Turning to the second of these terms, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i_1,i_2\in[N]}\sum_{j\in[n]}\mathbb{E}|[W^*b]_\ell|^2\,|[a^*\widetilde{Q}_{u1}W\Omega^{-L}]_j|^2\overline{[Q_{0v}W]_{i_1\ell}}[\widehat{S}_j]_{i_1,i_2}[Q_{0v}W]_{i_2\ell}\\
&\quad\le\frac{\boldsymbol{M}}{n}\mathbb{E}|[W^*b]_\ell|^2\,\|a^*\widetilde{Q}_{u1}W\|^2\|[Q_{0v}W]_{\cdot,\ell}\|^2\\
&\quad\le\|a\|^2\|b\|^2C/n^2,
\end{aligned}$$

where the last inequality can be obtained by applying, e.g. Lemma 3.7 along with standard inequalities. The third term can be handled similarly. □

Thanks to these bounds and to Cauchy–Schwarz inequality, we obtain the identity

$$\alpha_{uv}(\ell)=-\alpha_{u0}(\ell)\omega^{\ell L}\widetilde{q}s_{1v}(\ell)-\alpha_{u1}(\ell)\omega^{-\ell L}\widetilde{q}s_{0v}(\ell)+n^{-1}\mathbb{E}[a^*\widetilde{Q}_{uv}\widehat{S}_\ell b]+\varepsilon,$$

with $\|\varepsilon\|\le C\|a\|\|b\|n^{-3/2}$. This leads to the system of equations

$$\begin{bmatrix}1+\omega^{\ell L}\widetilde{q}s_{10}(\ell) & \omega^{-\ell L}\widetilde{q}s_{00}(\ell)\\ \omega^{\ell L}\widetilde{q}s_{11}(\ell) & 1+\omega^{-\ell L}\widetilde{q}s_{01}(\ell)\end{bmatrix}\begin{bmatrix}\alpha_{u0}(\ell)\\ \alpha_{u1}(\ell)\end{bmatrix}=\frac{1}{n}\begin{bmatrix}\mathbb{E}[a^*\widetilde{Q}_{u0}\widehat{S}_\ell b]\\ \mathbb{E}[a^*\widetilde{Q}_{u1}\widehat{S}_\ell b]\end{bmatrix}+\boldsymbol{\varepsilon},$$

with $\|\boldsymbol{\varepsilon}\|\le C\|a\|\|b\|n^{-3/2}$. The matrix at the left-hand side of this expression, that we denote as T_ℓ, is written as

$$T_\ell=I_2+\mathscr{T}\left((I_2\otimes\widehat{S}_\ell)\left(\widetilde{Q}^{\mathsf{T}}\begin{bmatrix} & \omega^{-\ell L}\\ \omega^{\ell L} & \end{bmatrix}\right)\right).$$

Assume that K is large enough so that $\|\widehat{S}_\ell\|\le 2\boldsymbol{M}$ for all $\ell\in[n]$, and take $\Im\eta\ge 4\boldsymbol{\gamma}_{\sup}\boldsymbol{M}$. Then by Lemma 4.3 we get that $\|T_\ell-I_2\|\le 1/2$. Thus, the determinant

$$d(\ell)=(1+\omega^{\ell L}\widetilde{q}s_{10}(\ell))(1+\omega^{-\ell L}\widetilde{q}s_{01}(\ell))-\widetilde{q}s_{00}(\ell)\widetilde{q}s_{11}(\ell)$$

of T_ℓ is such that $|d(\ell)|$ is bounded away from zero uniformly in N and ℓ. Solving our system, and reusing henceforth the notations ε and $\boldsymbol{\varepsilon}$ at will, we get that

$$\begin{aligned}
\begin{bmatrix}\alpha_{u0}(\ell)\\ \alpha_{u1}(\ell)\end{bmatrix}&=\frac{1}{n}\frac{1}{d(\ell)}\begin{bmatrix}(1+\omega^{-\ell L}\widetilde{q}s_{01}(\ell)) & -\omega^{-\ell L}\widetilde{q}s_{00}(\ell)\\ -\omega^{\ell L}\widetilde{q}s_{11}(\ell) & (1+\omega^{\ell L}\widetilde{q}s_{10}(\ell))\end{bmatrix}\\
&\quad\times\begin{bmatrix}\mathbb{E}[a^*\widetilde{Q}_{u0}\widehat{S}_\ell b]\\ \mathbb{E}[a^*\widetilde{Q}_{u1}\widehat{S}_\ell b]\end{bmatrix}+\boldsymbol{\varepsilon},
\end{aligned}\tag{4.8}$$

with $\|\boldsymbol{\varepsilon}\|\le C\|a\|\|b\|n^{-3/2}$.

Now, keeping in mind the identity $\widetilde{Q}\big(\boldsymbol{H}(W\Omega^L W^*) - \left[\begin{smallmatrix} \eta & z \\ \bar{z} & \eta \end{smallmatrix}\right] \otimes I_N\big) = I$, our purpose is to find an approximant of the matrix

$$\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*) = \begin{bmatrix} \widetilde{Q}_{01} W\Omega^{-L}W^* & \widetilde{Q}_{00} W\Omega^L W^* \\ \widetilde{Q}_{11} W\Omega^{-L}W^* & \widetilde{Q}_{10} W\Omega^L W^* \end{bmatrix}.$$

To that end, we write

$$\mathbb{E}[a^*\widetilde{Q}_{u0} W\Omega^L W^* b] = \sum_{\ell\in[n]} \omega^{\ell L}\alpha_{u0}(\ell), \quad \text{and}$$
$$\mathbb{E}[a^*\widetilde{Q}_{u1} W\Omega^{-L} W^* b] = \sum_{\ell\in[n]} \omega^{-\ell L}\alpha_{u1}(\ell),$$

and we use Eq. (4.8) to obtain

$$\begin{aligned} &[\mathbb{E}a^*\widetilde{Q}_{u1}W\Omega^{-L}W^*b \quad \mathbb{E}a^*\widetilde{Q}_{u0}W\Omega^L W^*b] = a^*\mathbb{E}[\widetilde{Q}_{u0} \quad \widetilde{Q}_{u1}] \\ &\times \frac{1}{n}\sum_{\ell\in[n]} \frac{1}{d(\ell)} \begin{bmatrix} -\widetilde{q}s_{11}(\ell) & \omega^{\ell L} + \widetilde{q}s_{01}(\ell) \\ \omega^{-\ell L} + \widetilde{q}s_{10}(\ell) & -\widetilde{q}s_{00}(\ell) \end{bmatrix} \otimes (\widehat{S}_\ell b) + \boldsymbol{\varepsilon} \end{aligned} \tag{4.9}$$

with $\|\boldsymbol{\varepsilon}\| \le C\|a\|\|b\|n^{-1/2}$ (note that we lost a factor of n^{-1} because of the summation $\sum_{\ell\in[n]}$). Let $U_{L,\ell} \triangleq U_L(e^{2\imath\pi\ell/n}) = \left[\begin{smallmatrix} & \omega^{\ell L} \\ \omega^{-\ell L} & \end{smallmatrix}\right]$, and define the matrix

$$\begin{aligned} C(z,\eta) &\triangleq \frac{1}{n}\sum_{\ell\in[n]} \frac{1}{d(\ell)} \begin{bmatrix} -\widetilde{q}s_{11}(\ell) & \omega^{\ell L} + \widetilde{q}s_{01}(\ell) \\ \omega^{-\ell L} + \widetilde{q}s_{10}(\ell) & -\widetilde{q}s_{00}(\ell) \end{bmatrix} \otimes \widehat{S}_\ell \\ &= \frac{1}{n}\sum_{\ell\in[n]} \left(\begin{bmatrix} \widetilde{q}s_{00}(\ell) & \widetilde{q}s_{01}(\ell) \\ \widetilde{q}s_{10}(\ell) & \widetilde{q}s_{11}(\ell) \end{bmatrix} + U_{L,\ell} \right)^{-1} \otimes \widehat{S}_\ell \\ &= \frac{1}{n}\sum_{\ell\in[n]} (\mathscr{T}((I_2 \otimes \widehat{S}_\ell)\mathbb{E}\widetilde{Q}) + U_{L,\ell})^{-1} \otimes \widehat{S}_\ell. \end{aligned}$$

Then, Eq. (4.9) can be rewritten as

$$\mathbb{E}a^*\widetilde{Q}_{u1}W\Omega^{-L}W^*b = a^*\mathbb{E}[\widetilde{Q}_{u0} \quad \widetilde{Q}_{u1}]C\begin{bmatrix} b \\ 0 \end{bmatrix} + \varepsilon,$$
$$\mathbb{E}a^*\widetilde{Q}_{u0}W\Omega^{L}W^*b = a^*\mathbb{E}[\widetilde{Q}_{u0} \quad \widetilde{Q}_{u1}]C\begin{bmatrix} 0 \\ b \end{bmatrix} + \varepsilon',$$

with $|\varepsilon|, |\varepsilon'| \le C\|a\|\|b\|n^{-1/2}$. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two constant vectors in $\mathbb{C}^{2N}$. Recalling the expression of $\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*)$ above, the last display can be written compactly as

$$|\boldsymbol{a}(\mathbb{E}\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*) - \mathbb{E}\widetilde{Q}C)\boldsymbol{b}| \leq C\|\boldsymbol{a}\|\|\boldsymbol{b}\|n^{-1/2},$$

equivalently,

$$\|\mathbb{E}\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*) - \mathbb{E}\widetilde{Q}C\| \leq Cn^{-1/2}.$$

Since $\mathbb{E}\widetilde{Q} \in \mathfrak{S}^{2N}$, it is easy to prove, mostly by mimicking the first part of the proof of Theorem 1.3, that the matrix function

$$R(z,\eta) = \left(C(z,\eta) - \begin{bmatrix} \eta & z \\ \bar{z} & \eta \end{bmatrix} \otimes I_N \right)^{-1}$$

is well defined for $\eta \in \mathbb{C}\backslash\mathbb{R}$, and $R(z,\cdot) \in \mathfrak{S}^{2N}$. In particular, $\|R(z,\eta)\| \leq 1/\Im\eta$ for $\eta \in \mathbb{C}_+$. We therefore have

$$\begin{aligned}\|\mathbb{E}\widetilde{Q} - R\| &= \|\mathbb{E}\widetilde{Q}(R^{-1} - \widetilde{Q}^{-1})R\| = \|(\mathbb{E}\widetilde{Q}C - \mathbb{E}\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*))R\| \\ &\leq \|(\mathbb{E}\widetilde{Q}C - \mathbb{E}\widetilde{Q}\boldsymbol{H}(W\Omega^L W^*)\|\|R\| \\ &\leq Cn^{-1/2}.\end{aligned}$$

To complete the proof of Proposition 4.4, it remains to control the norm $\|R-\widehat{G}\|$. Remembering that $\widehat{G}$ is defined through the implicit equation in Proposition 4.3, we use the contraction property of $\breve{\mathcal{F}}_{\widehat{S},z}(\cdot,\eta)$ to gain this control. By mimicking the derivation that led to Inequality (4.1), we obtain that if K is large enough so that $\|\widehat{S}\|_\infty^{\mathbb{T}} \leq 2\boldsymbol{M}$, and if $\Im\eta$ is large enough, then

$$\|\breve{\mathcal{F}}_{\widehat{S},z}(M,\eta) - \breve{\mathcal{F}}_{\widehat{S},z}(M',\eta)\| \leq \frac{1}{2}\|M - M'\|.$$

Note that $R = \breve{\mathcal{F}}_{\widehat{S},z}(\mathbb{E}\widetilde{Q},\eta)$. Therefore, if $\Im\eta$ is large enough, we have

$$\begin{aligned}\|R - \widehat{G}\| &= \|\breve{\mathcal{F}}_{\widehat{S},z}(\mathbb{E}\widetilde{Q},\eta) - \breve{\mathcal{F}}_{\widehat{S},z}(R,\eta) + \breve{\mathcal{F}}_{\widehat{S},z}(R,\eta) - \breve{\mathcal{F}}_{\widehat{S},z}(\widehat{G},\eta)\| \\ &\leq \frac{1}{2}\|\mathbb{E}\widetilde{Q} - R\| + \frac{1}{2}\|R - \widehat{G}\|,\end{aligned}$$

leading to

$$\|R - \widehat{G}\| \leq \|\mathbb{E}\widetilde{Q} - R\| \leq Cn^{-1/2}.$$

Finally, $\|\mathbb{E}\widetilde{Q} - \widehat{G}\| \leq \|\mathbb{E}\widetilde{Q} - R\| + \|R - \widehat{G}\| \leq Cn^{-1/2}$. The proof of Proposition 4.4 can now be completed by combining this bound with the inequality (4.6).

Proof of Proposition 4.5. We can assume that the processes $(x_k^{(N)})_k$ and $(\hat{x}_k^{(N,K)})_k$ that constitute the columns of the matrices $X^{(N)}$ and $\widehat{X}^{(N,K)}$ respectively, are generated by applying the filters with Fourier transforms $n^{-1/2}S^{(N)}(e^{\imath\theta})^{1/2}$ and

$n^{-1/2}\widehat{S}^{(N,K)}(e^{\imath\theta})^{1/2}$, respectively, to the same i.i.d. process $\boldsymbol{\xi}^{(N)} = (\boldsymbol{\xi}_k^{(N)})_{k\in\mathbb{Z}}$ with $\boldsymbol{\xi}_k^{(N)} \sim \mathcal{N}_{\mathbb{C}}(0, I_N)$. This being the case, we have

$$\mathbb{E}\|x_k^{(N)} - \hat{x}_k^{(N,K)}\|^2 = \frac{1}{n}\frac{1}{2\pi}\int_0^{2\pi} \|S^{(N)}(e^{\imath\theta})^{1/2} - \widehat{S}^{(N,K)}(e^{\imath\theta})^{1/2}\|_{\mathrm{HS}}^2 d\theta.$$

Hence, by (4.5), we get that

$$\sup_N \mathbb{E}\|x_k^{(N)} - \hat{x}_k^{(N,K)}\|^2 \xrightarrow[K\to\infty]{} 0.$$

For any two square matrices M_1 and M_2 of same order, by Cauchy–Schwarz inequality, $|\operatorname{tr} M_1M_2| \le \|M_1\|_{\mathrm{HS}}\|M_2\|_{\mathrm{HS}}$. Thus, by the resolvent identity,

$$\begin{aligned}\frac{1}{N^2}|\operatorname{tr} D(\mathbb{E}Q(z,\eta) - \mathbb{E}\widehat{Q}(z,\eta))|^2 &\le \frac{1}{N^2}\mathbb{E}|\operatorname{tr} DQ(z,\eta)(\boldsymbol{H}(\widehat{X}J^L\widehat{X}^*)\\ &\quad - \boldsymbol{H}(XJ^LX^*))\widehat{Q}(z,\eta)|^2\\ &\le \frac{1}{(\Im\eta)^4}\frac{1}{N}\mathbb{E}\|\boldsymbol{H}(\widehat{X}J^L\widehat{X}^*) - \boldsymbol{H}(XJ^LX^*)\|_{\mathrm{HS}}^2.\end{aligned}$$

Writing $\mathbb{E}\|XJ^LX^* - \widehat{X}J^L\widehat{X}^*\|_{\mathrm{HS}}^2 = \mathbb{E}\|(X-\widehat{X})J^LX^* + \widehat{X}J^L(X-\widehat{X})^*\|_{\mathrm{HS}}^2$, it is enough to bound $\mathbb{E}\|(X-\widehat{X})J^LX^*\|_{\mathrm{HS}}^2$. Given a constant $\kappa > 0$, we have

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\|(X-\widehat{X})J^LX^*\|_{\mathrm{HS}}^2\\ &\quad\le \frac{1}{n}\mathbb{E}\|(X-\widehat{X})\|_{\mathrm{HS}}^2\|X\|^2\\ &\quad\le \frac{\kappa^2}{n}\mathbb{E}\|(X-\widehat{X})\|_{\mathrm{HS}}^2 + \frac{1}{n}\mathbb{E}\|(X-\widehat{X})\|_{\mathrm{HS}}^2\|X\|^2\mathbb{1}_{\|X\|>\kappa}\\ &\quad\le \kappa^2\mathbb{E}\|x_k - \hat{x}_k\|^2 + \frac{1}{n}(\mathbb{E}(\|X-\widehat{X}\|_{\mathrm{HS}}^4)^{1/2}(\mathbb{E}\|X\|^4\mathbb{1}_{\|X\|>\kappa})^{1/2}.\end{aligned}$$

With the help of Lemma 3.7, the second term in the last expression can be made as small as desired, independently of N, when N is large enough, by choosing κ large enough. The first term converges to zero as $K \to \infty$ as shown above. □

4.4. *Corollary 1.2: Sketch of the proof*

We have here $G(z,\eta) = \mathcal{F}_{I_N,z}(G(z,\eta),\eta)$ for $\eta \in \mathbb{C}_+$. We also know from the proof of Theorem 1.3 that if we start with $P_0(z,\eta) = -\eta^{-1}I_{2N}$, then the iterates $P_{k+1}(z,\eta) = \mathcal{F}_{I,z}(P_k(z,\eta),\eta)$ converge to $G(z,\eta)$ uniformly on the compacts of $\mathbb{C}_+$ in the parameter η. Writing $\mathscr{T}(P_k) = \begin{bmatrix} p_{00,k} & p_{01,k}\\ p_{10,k} & p_{11,k}\end{bmatrix}$ where the argument (z,η) is omitted, by developing the expression of $\mathcal{F}_{I_N,z}(P_k,\eta)$, we get that

$$P_{k+1} = \begin{bmatrix} \frac{1}{2\pi}\int \frac{p_{11,k}}{\Delta_k(e^{\imath\theta})}d\theta - \eta & -\frac{1}{2\pi}\int \frac{(p_{01,k}+e^{-\imath\theta})}{\Delta_k(e^{\imath\theta})}d\theta - z \\ -\frac{1}{2\pi}\int \frac{(p_{10,k}+e^{\imath\theta})}{\Delta_k(e^{\imath\theta})}d\theta - \bar{z} & \frac{1}{2\pi}\int \frac{p_{00,k}}{\Delta_k(e^{\imath\theta})}d\theta - \eta \end{bmatrix}^{-1} \otimes I_N,$$

where $\Delta_k(e^{\imath\theta}) = p_{00,k}p_{11,k} - (p_{01,k}+e^{-\imath\theta})(p_{10,k}+e^{\imath\theta})$. Setting from now on $\eta = \imath t$ with $t > 0$, it is obvious that $p_{00,0} = p_{11,0} = \imath h_0$ for $h_0 = 1/t > 0$, and $p_{01,0} = \bar{p}_{10,0}$ ($= 0$ here). Assuming that $p_{00,k} = p_{11,k} = \imath h_k$ for some $h_k > 0$ and $p_{01,k} = \bar{p}_{10,k}$, it is not difficult to show by developing the last display that the same properties hold for P_{k+1}. Passing to the limit, we obtain that $g_{00}(z,\imath t) = g_{11}(z,\imath t) = \imath h$ for some $h > 0$, and $g_{01}(z,\imath t) = \bar{g}_{10}(z,\imath t)$, where we wrote $\mathscr{T}(G) = \left[\begin{smallmatrix} g_{00} & g_{01} \\ g_{10} & g_{11}\end{smallmatrix}\right]$. With this at hand, the equation $G(z,\imath t) = \mathcal{F}_{I,z}(G(z,\imath t),\imath t)$ becomes

$$\begin{aligned} G &= \left(\frac{1}{2\pi}\int_0^{2\pi} \frac{1}{\widetilde{\Delta}(e^{\imath\theta})}\begin{bmatrix} -\imath h & g_{01}+e^{-\imath\theta} \\ \bar{g}_{01}+e^{\imath\theta} & -\imath h\end{bmatrix} d\theta - \begin{bmatrix} \imath t & z \\ \bar{z} & \imath t\end{bmatrix}\right)^{-1} \otimes I_N \\ &= \frac{n}{N}\begin{bmatrix} \imath h & g_{01} \\ \bar{g}_{01} & \imath h\end{bmatrix} \otimes I_N, \end{aligned}$$

where $\widetilde{\Delta}(e^{\imath\theta}) = h^2 + |g_{01}+e^{-\imath\theta}|^2$. After some calculation, this equation can be equivalently restated in the form of the following system of two equations:

$$\begin{aligned} \frac{N}{n} &= \frac{1}{2\pi}\int_0^{2\pi} \frac{h^2 + |g_{01}|^2 + g_{01}e^{\imath\theta}}{\widetilde{\Delta}(e^{\imath\theta})}d\theta + th - \bar{z}g_{01}, \\ 0 &= \frac{1}{2\pi}\int_0^{2\pi} \frac{he^{-\imath\theta}}{\widetilde{\Delta}(e^{\imath\theta})}d\theta - zh - tg_{01}. \end{aligned}$$

This coincides with the system of [7, Eqs. (33a) and (33b)].

Acknowledgments

The authors would like to thank the referees for their detailed and insightful comments. This work has been partially funded by the Indo-French Center for Applicable Mathematics, project High-Dimensional Random Matrix Models with Applications. The work of WH has been partially supported by the French ANR grant HIDITSA (ANR-17-CE40-0003).

Appendix A. Proof of Lemma 3.6

(1) Using Markov's inequality, an obvious integration with respect to the exponential distribution, and using the inequality $-\log(1-x) \le 2x$ for $x \in [0,1/2]$, we

have

$$\begin{aligned}\mathbb{P}[\|\Sigma^{1/2}\xi\| \geq \sqrt{2Nt}] &= \mathbb{P}\left[\sum_{\ell=0}^{N-1} s_\ell(\Sigma)|\xi_\ell|^2 \geq 2Nt\right] \\ &= \mathbb{P}\left[\exp\frac{\sum_{\ell=0}^{N-1} s_\ell(\Sigma)|\xi_\ell|^2}{2\|\Sigma\|} \geq \exp\frac{Nt}{\|\Sigma\|}\right] \\ &\leq e^{-Nt/\|\Sigma\|}\mathbb{E}\left[\exp\frac{\sum_{\ell=0}^{N-1} s_\ell(\Sigma)|\xi_\ell|^2}{2\|\Sigma\|}\right] \\ &= e^{-\frac{Nt}{\|\Sigma\|}}\prod_{\ell=0}^{N-1}(1 - s_\ell(\Sigma)/(2\|\Sigma\|))^{-1} \\ &\leq \exp\left(-\frac{Nt}{\|\Sigma\|} + \sum_{\ell=0}^{N-1}\frac{s_\ell(\Sigma)}{\|\Sigma\|}\right) \\ &\leq e^{-(t/\|\Sigma\|-1)N}.\end{aligned}$$

(2) We have $\|\Sigma^{1/2}\xi\|^2 \stackrel{\mathcal{L}}{=} \sum_{\ell=0}^{N-1} s_\ell(\Sigma)|\xi_\ell|^2 \geq \alpha\sum_{\ell=0}^{m-1}|\xi_\ell|^2$. Thus, by a calculation similar to above

$$\mathbb{P}[\|\Sigma^{1/2}\xi\| \leq \sqrt{\alpha m/2}] \leq \mathbb{P}\left[\sum_{\ell=0}^{m-1}|\xi_\ell|^2 \leq m/2\right] \leq \exp(-c_{3.6,2}m).$$

(3) We obviously have that

$$(\xi + a)^*M(\xi + a) = (\xi + a)^*\Re M(\xi + a) + \imath(\xi + a)^*\Im M(\xi + a),$$

and both $(\xi + a)^*\Re M(\xi + a)$ and $(\xi + a)^*\Im M(\xi + a)$ are real. Furthermore,

$$\|M\|_{\mathrm{HS}}^2 = \|\Re M\|_{\mathrm{HS}}^2 + \|\Im M\|_{\mathrm{HS}}^2.$$

Let us assume that $\|\Re M\|_{\mathrm{HS}} \geq \|M\|_{\mathrm{HS}}/\sqrt{2}$, otherwise, we replace $\Re M$ with $\Im M$. From these facts, we deduce that

$$\mathcal{L}((\xi + a)^*M(\xi + a)/\|M\|_{\mathrm{HS}}, t) \leq \mathcal{L}((\xi + a)^*\Re M(\xi + a)/\|\Re M\|_{\mathrm{HS}}, \sqrt{2}t).$$

By a spectral factorization of the Hermitian matrix $\Re M$,

$$(\xi + a)^*\Re M(\xi + a)/\|\Re M\|_{\mathrm{HS}} \stackrel{\mathcal{L}}{=} \sum_{\ell\in[N]}\beta_\ell|\xi_\ell + u_\ell|^2,$$

where the u_ℓ are deterministic complex numbers, and the β_ℓ are deterministic reals that satisfy $\sum\beta_\ell^2 = 1$, since they are the eigenvalues of $\|\Re M\|_{\mathrm{HS}}^{-1}\Re M$. The random variable $|\xi_\ell + u_\ell|^2$ is non-central chi-squared with two degrees of freedom and has the density

$$f_\ell(x) = \exp(-(x + |u_\ell|^2))\boldsymbol{I}_0(2|u_\ell|\sqrt{x})$$

on $\mathbb{R}_+$, where $\boldsymbol{I}_0$ is the modified Bessel function. Since these densities are bounded by one [1], we can get the result from [33, Theorem 1.2].

(4) By the restriction property of Lévy's anti-concentration function

$$\mathcal{L}(\Sigma^{1/2}\xi, \sqrt{m}t) \leq \sup_{(d_0,\ldots,d_{m-1})\in\mathbb{C}^m} \mathbb{P}\left[\sum_{\ell=0}^{m-1} |s_\ell(\Sigma)^{1/2}\xi_\ell - d_\ell|^2 \leq mt^2\right].$$

We furthermore use the following well-known *tensorization* result (see [31, Lemma 2.2]): Suppose $\{w_0, \ldots, w_{m-1}\}$ is a collection of independent non-negative random variables such that there is a constant $c > 0$ such that for each $t \geq 0$,

$$\mathbb{P}[w_\ell \leq t] \leq ct.$$

Then there exists a constant $c' > 0$ so that

$$\mathbb{P}\left[\sum_{\ell=0}^{m-1} w_\ell^2 \leq mt^2\right] \leq (c't)^m.$$

In the present case, for each $\ell \in [m]$ and each $d_\ell \in \mathbb{C}$, $w_\ell^2 = [|s_\ell(\Sigma)^{1/2}\xi_\ell - d_\ell|]^2$ is a non-central chi-squared random variable with two degrees of freedom and has a density bounded by a constant that depends only on α, as can be checked from the previous item. Thus, the tensorization argument applies and the result follows.

(5) By a singular value decomposition of M, we obtain that

$$\mathbb{P}[\|M\xi\|^2 \geq t\|M\|_{\mathrm{HS}}^2] = \mathbb{P}\left[\sum_{\ell=0}^{N-1} \sigma_\ell^2 |\xi_\ell|^2 \geq t\right],$$

where $\sum \sigma_\ell^2 = 1$. Writing

$$\mathbb{P}\left[\sum_\ell \sigma_\ell^2 |\xi_\ell|^2 \geq t\right] = \mathbb{P}\left[\exp\left(\sum_\ell \sigma_\ell^2 |\xi_\ell|^2/2\right) \geq \exp(t/2)\right]$$

and following the arguments given for Item (1), we easily get the result.

Proof of Lemma 3.6 is now complete.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, Vol. 55 (Dover Publications, 1964).

[2] M. Bhattacharjee and A. Bose, Large sample behaviour of high dimensional autocovariance matrices, *Ann. Statist.* **44**(2) (2016) 598–628.

[3] M. Bhattacharjee and A. Bose, Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application, *Ann. Statist.* **47**(6) (2019) 3470–3503.

[4] V. Bolotnikov, On a general moment problem on the half axis, *Linear Algebra Appl.* **255** (1997) 57–112.

[5] Ch. Bordenave and D. Chafaï, Around the circular law, *Probab. Surv.* **9** (2012) 1–89.
[6] A. Bose and M. Bhattacharjee, *Large Covariance and Autocovariance Matrices* (Chapman & Hall/CRC, Boca Raton, London, New York, 2018).
[7] A. Bose and W. Hachem, Smallest singular value and limit eigenvalue distribution of a class of non-hermitian random matrices with statistical application, *J. Multi. Anal.* **178** (2020) 104623.
[8] A. Böttcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices* (SIAM, Philadelphia, PA, 2005).
[9] D. R. Brillinger, *Time Series: Data Analysis and Theory* (SIAM, 2001).
[10] S. Chatterjee and A. Bose, A new method for bounding rates of convergence of empirical spectral distributions, *J. Theoret. Probab.* **17**(4) (2004) 1003–1019.
[11] N. Cook, W. Hachem, J. Najim and D. Renfrew, Non-Hermitian random matrices with a variance profile (I): Deterministic equivalents and limiting ESDs, *Electron. J. Probab.* **23** (2018) 110.
[12] D. Damanik, A. Pushnitski and B. Simon, The analytic theory of matrix orthogonal polynomials, *Surv. Approx. Theory* **4** (2008) 1–85.
[13] J. Feinberg and A. Zee, Non-Hermitian random matrix theory: Method of Hermitian reduction, *Nucl. Phys. B* **504**(3) (1997) 579–608.
[14] J. S. Geronimo and T. P. Hill, Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform, *J. Approx. Theory* **121**(1) (2003) 54–60.
[15] F. Gesztesy and E. Tsekanovskii, On matrix-valued Herglotz functions, *Math. Nachr.* **218** (2000) 61–138.
[16] V. L. Girko, The circular law, *Teor. Veroyatnost. Primenen.* **29**(4) (1984) 669–679. English translation: *Theory Probab. Appl.* **29**(4) (1984) 694–706.
[17] J. Glimm and A. Jaffe, *Quantum Physics. A Functional Integral Point of View*, 2nd edn. (Springer-Verlag, New York, 1987).
[18] F. Götze and A. Tikhomirov, The circular law for random matrices, *Ann. Probab.* **38**(4) (2010) 1444–1491.
[19] W. Hachem, Ph. Loubaton and J. Najim, Deterministic equivalents for certain functionals of large random matrices, *Ann. Appl. Probab.* **17**(3) (2007) 875–930.
[20] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1990). Corrected reprint of the 1985 original.
[21] T. Kailath, A. Vieira and M. Morf, Inverses of Toeplitz operators, innovations, and orthogonal polynomials, *SIAM Rev.* **20**(1) (1978) 106–119.
[22] A. M. Khorunzhy and L. A. Pastur, Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness, *Comm. Math. Phys.* **153**(3) (1993) 605–646.
[23] A. E. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, *Adv. Math.* **195**(2) (2005) 491–523.
[24] P. G. Nevai, Orthogonal polynomials, *Mem. Amer. Math. Soc.* **18**(213) (1979) v+185.
[25] H. H. Nguyen, On the least singular value of random symmetric matrices, *Electron. J. Probab.* **17** (2012) 1–19.
[26] L. A. Pastur, A simple approach to the global regime of Gaussian ensembles of random matrices, *Ukraïn. Mat. Zh.* **57**(6) (2005) 790–817.
[27] L. A. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, Mathematical Surveys and Monographs, Vol. 171 (American Mathematical Society, Providence, RI, 2011).

[28] C. Peligrad and M. Peligrad, The limiting spectral distribution in terms of spectral density, *Random Matrices Theory Appl.* **5**(1) (2016) 1650003, Erratum **5**(1) (2016) 1792001-1.

[29] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory*, Oxford Mathematical Monographs (Oxford University Press, New York; Clarendon Press, Oxford, 1985).

[30] Y. A. Rozanov, *Stationary Random Processes* (Holden-Day, Inc., San Francisco, California-London-Amsterdam, 1967). Translated from the Russian by A. Feinstein.

[31] M. Rudelson and R. Vershynin, The Littlewood–Offord problem and invertibility of random matrices, *Adv. Math.* **218**(2) (2008) 600–633.

[32] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, *Comm. Pure Appl. Math.* **62**(12) (2009) 1707–1739.

[33] M. Rudelson and R. Vershynin, Small ball probabilities for linear images of high-dimensional distributions, *Int. Math. Res. Not. IMRN* **2015**(19) (2015) 9594–9617.

[34] J. W. Silverstein and Z. D. Bai, On the empirical distribution of eigenvalues of a class of large-dimensional random matrices, *J. Multiv. Anal.* **54**(2) (1995) 175–192.

[35] B. Simon, *Szegő's Theorem and Its Descendants: Spectral Theory for L^2 Perturbations of Orthogonal Polynomials*, M. B. Porter Lectures (Princeton University Press, Princeton, NJ, 2011).

[36] H. Stahl and V. Totik, *General Orthogonal Polynomials*, Encyclopedia of Mathematics and its Applications, Vol. 43 (Cambridge University Press, Cambridge, 1992).

[37] T. Tao, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, Vol. 132 (American Mathematical Society, Providence, RI, 2012).

[38] T. Tao and V. Vu, Random matrices: Universality of ESDs and the circular law, *Ann. Probab.* **38**(5) (2010) 2023–2065. With an appendix by Manjunath Krishnapur.

[39] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators* (Princeton University Press, Princeton, NJ, 2005).

[40] R. Vershynin, Invertibility of symmetric random matrices, *Random Struct. Algorithms* **44**(2) (2014) 135–182.

[41] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 2018).

[42] J. Yao and W. Yuan, On eigenvalue distributions of large auto-covariance matrices, preprint (2020), arXiv:2011.09165.