Gaussian fluctuations for linear spectral statistics of large random covariance matrices

Jamal Najim

GAUSSIAN FLUCTUATIONS FOR LINEAR SPECTRAL STATISTICS OF LARGE RANDOM COVARIANCE MATRICES

JAMAL NAJIM

ABSTRACT. Consider a $N \times n$ matrix $\Sigma_n = \frac{1}{\sqrt{n}} R_n^{1/2} X_n$, where R_n is a nonnegative definite Hermitian matrix and X_n is a random matrix with i.i.d. real or complex standardized entries. The fluctuations of the linear statistics of the eigenvalues:

$$\text{Trace } f(\Sigma_n \Sigma_n^*) = \sum_{i=1}^{N} f(\lambda_i), \quad (\lambda_i) \text{ eigenvalues of } \Sigma_n \Sigma_n^*,$$

are shown to be gaussian, in the regime where both dimensions of matrix Σ_n go to infinity at the same pace and in the case where f is an analytic function. The main improvement with respect to Bai and Silverstein's CLT [3] lies in the fact that we consider general entries with finite fourth moment, but whose fourth cumulant is non-null, i.e. whose fourth moment may differ from the moment of a (real or complex) Gaussian random variable. As a consequence, extra terms proportional to

$$|\mathcal{V}|^2 = |\mathbb{E}(X_{11}^n)^2|^2 \quad \text{and} \quad \kappa = \mathbb{E}\left|X_{ij}^n\right|^4 - |\mathcal{V}|^2 - 2$$

appear in the limiting variance and in the limiting bias, which not only depend on the spectrum of matrix R_n but also on its eigenvectors.

The CLT is expressed in terms of vanishing Lévy-Prohorov distance between the linear statistics' distribution and a Gaussian probability distribution, the mean and the variance of which depend upon N and n and may not converge.

AMS 2000 subject classification: Primary 15A52, Secondary 15A18, 60F15.
Key words and phrases: large random matrix, fluctuations, linear statistics of the eigenvalues, central limit theorem.

1. INTRODUCTION

Consider a $N \times n$ random matrix $\Sigma_n = (\xi_{ij}^n)$ given by:

$$\Sigma_n = \frac{1}{\sqrt{n}} R_n^{1/2} X_n \ , \tag{1.1}$$

where $N = N(n)$ and R_n is a $N \times N$ nonnegative definite hermitian matrix with spectral norm uniformily bounded in N. The entries $(X_{ij}^n \ ; i \leq N, j \leq n, n \geq 1)$ of matrices (X_n) are real or complex, independent and identically distributed (i.i.d.) with mean 0 and variance 1. Matrix $\Sigma_n \Sigma_n^*$ models a sample covariance matrix, formed from n samples of the random vector $R_n^{1/2} X_{\cdot 1}$, with the population covariance matrix R_n.

Date: September 13, 2013.

Since the seminal work of Marčenko and Pastur [24] in 1967, the study of the spectrum of large covariance matrices of the type $X_n X_n^*$ under the asymptotic regime where:

$$N, n \to \infty \quad \text{and} \quad 0 < \liminf \frac{N}{n} \leq \limsup \frac{N}{n} < \infty \,, \tag{1.2}$$

(a condition that will be simply referred as $N, n \to \infty$ in the sequel) has drawn a considerable interest.

In this article, we study the fluctuations of linear spectral statistics of the form:

$$\operatorname{tr} f(\Sigma_n \Sigma_n^*) = \sum_{i=1}^{N} f(\lambda_i) \,, \qquad \text{as} \quad N, n \to \infty \tag{1.3}$$

where $\operatorname{tr}(A)$ refers to the trace of A and the λ_i's are the eigenvalues of $\Sigma_n \Sigma_n^*$. This subject has a rich history with contributions by Arharov [2], Girko (see [12, 13] and the references therein), Jonsson [19], Khorunzhiy et al. [22], Johansson [18], Sinai and Soshnikov [27, 28], Cabanal-Duvillard [8], Guionnet [14], Bai and Silverstein [3], Anderson and Zeitouni [1], Pan and Zhou [25], Chatterjee [9], Lytova and Pastur [23], Bai et al. [5], etc. There are also more recent contributions for heavytailed entries (see for instance Benaych-Georges et al. [6]).

In their '04 article [3], Bai and Silverstein established a CLT for the linear spectral statistics (1.3) as the dimensions N and n grow to infinity at the same pace ($N/n \to c \in (0, \infty)$) and under two important assumptions:

(1) The entries (X_{ij}^n) are centered with unit variance and a finite fourth moment equal to the fourth moment of a (real or complex) gaussian standard variable.
(2) Function f in (1.3) is analytic in a neighbourhood of the asymptotic spectrum of $\Sigma_n \Sigma_n^*$.

Such a highly cited result proved to be useful in probability theory, statistics and various other fields; and as a consequence, many attempts have been done to relax both assumptions.

It is well known since the paper by Khorunzhiy et al. [22] that if the fourth moment of the entries differs from the fourth moment of a Gaussian random variable, then a term appears in the variance of the trace of the resolvent, which is proportional to the fourth cumulant of the entries. This term does not appear if Assumption (1) holds true, because in this case, the fourth cumulant is zero.

In Pan and Zhou [25], Assumption (1) has been relaxed under an additional assumption on matrix R_n, which somehow enforces structural conditions on R_n (in particular, these conditions are satisfied if matrix R_n is diagonal). In Hachem et al. [21, 15], CLTs have been established for specific linear statistics of interest in information theory, with general entries and (possibly non-centered) covariance random matrices with a variance profile. In Lytova and Pastur [23], both assumptions have been relaxed for the "white" model, when R_n is equal to the identity matrix. In this case, it has been proved that a mild integrability condition over the Fourier transform of f was enough to establish the CLT. In Bai, Wang and Zhou [5], fluctuations for the white model are addressed as well; Assumption (2) is relaxed and functions with continuous fourth order derivatives are considered. An important feature of these works [15, 21, 22, 23, 25] is the presence of a term proportional to the fourth cumulant in the variance.

The purpose of this article is to establish a CLT for linear spectral statistics (1.3) for general entries X_{ij}^n with fourth moment finite, and hence to fully relax Assumption (1) in

[3], while keeping f analytic . In order to address this question, we take full advantage of Bai and Silverstein's CLT; we closely follow and extend their computations.

Interesting phenomena appear when considering entries with non-Gaussian fourth moment. As was previously mentioned, a term proportional to the fourth cumulant appears in the variance of the linear statistics; another term, proportional to $|\mathbb{E}(X_{11}^n)^2|^2$, also appears (as already noticed in [15]). The presence of these terms whose convergence is not granted under usual assumptions (see for instance the condition (2.14) hereafter) will modify the nature of the fluctuations of the linear statistics.

Denote by $L_n(f)$ the (approximately) centered version of the linear statistics (1.3), to be properly defined below. In Bai and Silverstein's setup (see also Pan and Zhou [25]), the CLT is expressed in the usual way, i.e. ($\xrightarrow{\mathcal{D}}$ stands for the convergence in distribution):

$$L_n(f) \xrightarrow[N,n\to\infty]{\mathcal{D}} \mathcal{N}(\mathcal{B}_\infty, \Theta_\infty) \ ,$$

for some well-defined parameters $\mathcal{B}_\infty, \Theta_\infty$ depending on function f and on the limiting spectrum of R_n and $\Sigma_n\Sigma_n^*$.

If Assumption (1) is relaxed, such a convergence may not hold. We prove in this case that the distribution of the linear statistics $L_n(f)$ becomes close to a family of Gaussian distributions, whose parameters (mean and variance) may not converge. More precisely, we establish in the sequel that there exists a family of Gaussian random variables $\mathcal{N}(\mathcal{B}_n, \Theta_n)$, such that

$$d_{\mathcal{L}P}\left(L_n(f), \mathcal{N}(\mathcal{B}_n, \Theta_n)\right) \xrightarrow[N,n\to\infty]{} 0 \ , \tag{1.4}$$

where $d_{\mathcal{L}P}$ denotes the Lévy-Prohorov distance (and in particular metrizes the convergence of laws).

The difference here with respect to [3, 25] lies in the fact that the Gaussian parameters $\mathcal{B}_n$ and Θ_n not only depend on the limiting spectrum of R_n and $\Sigma_n\Sigma_n^*$ but also on the behaviour of the eigenvectors[1] of R_n; a fact that is due to the presence of terms proportional to both $|\mathbb{E}(X_{11}^n)^2|^2$ and the fourth cumulant. As a consequence, these parameters may not converge unless some very strong structural assumption over R_n (such as R_n diagonal) is made, which would ensure a joint convergence of R_n's spectrum and eigenvectors. Notice that this absence of convergence is not observed in [23] because in the case where R_n is equal to the identity I_N, there are no issues with its eigenvectors.

Expressing the CLT as in (1.4) makes it possible to circumvent this joint convergence issue and related cumbersome assumptions. This framework may also prove to be useful for other interesting models such as large dimensional information-plus-noise type matrices [10, 16] and more generally mixed models combining deterministic and random large dimensional matrices.

Outline of the article. The general framework of the paper together with the main formulas (canonical equations, deterministic equivalents, variance and bias formulas) are presented in Section 2. The central limit theorem (Theorem 3.1) is stated in Section 3. Section 4 is devoted to the proof of Theorem 3.1; in particular, the cumulant term in the variance is identified in subsection 4.3.3 and both the existence and the tightness of the Gaussian limiting process are established in subsection 4.4.

[1] In a slightly different context, such a phenomenon has been noticed in Hachem et al. [17].

Acknowledgement. The author would like to thank Djalil Chafaï, Walid Hachem and Jianfeng Yao for fruitful discussions.

2. Background and presentation of the results

2.1. **Resolvent, canonical equation and deterministic equivalents.** Denote by $Q_n(z)$ the resolvent of matrix $\Sigma_n\Sigma_n^*$:

$$Q_n(z) = \left(\Sigma_n\Sigma_n^* - zI_N\right)^{-1} \ , \tag{2.1}$$

and by $f_n(z)$ its normalized trace which is the Stieltjes transform of the empirical distribution of $\Sigma_n\Sigma_n^*$'s eigenvalues:

$$f_n(z) = \frac{1}{N}\mathrm{tr}\, Q_n(z) \ . \tag{2.2}$$

The following canonical equation[2] admits a unique solution t_n in the class of Stieltjes transforms of probability measures (see for instance [3]):

$$t_n(z) = \frac{1}{N}\mathrm{tr}\, \left(-zI_N + (1-c_n)R_n - zc_nt_n(z)R_n\right)^{-1} \ , \quad z \in \mathbb{C}\setminus\mathbb{R}^+ \ , \tag{2.3}$$

where c_n stands for the ratio N/n. The function t_n being introduced, we can define the following $N \times N$ matrix

$$T_n(z) = \left(-zI_N + (1-c_n)R_n - zc_nt_n(z)R_n\right)^{-1} \ . \tag{2.4}$$

Matrix $T_n(z)$ can be thought of as a *deterministic equivalent* of the resolvent $Q_n(z)$ in the sense that it is a deterministic quantity, easily computable, which does not depend on the distribution of the entries and which approximates the resolvent in various senses. For instance,

$$\frac{1}{N}\mathrm{tr}\, T_n(z) - \frac{1}{N}\mathrm{tr}\, Q_n(z) \xrightarrow[N,n\to\infty]{} 0$$

(in probability or almost surely). Otherwise stated, $t_n(z) = N^{-1}\mathrm{tr}\, T_n(z)$ is the deterministic equivalent of $f_n(z)$. As we shall see later in this paper,

$$u_n^*Q_nv_n - u_n^*T_nv_n \xrightarrow[N,n\to\infty]{} 0 \tag{2.5}$$

where (u_n) and (v_n) are deterministic $N\times 1$ vectors with uniformily bounded euclidian norms in N. As a consequence of (2.5), not only T_n conveys information on the limiting spectrum of the resolvent Q_n but also on the eigenvectors of Q_n.

2.2. **Entries with non-null fourth cumulant and the limiting covariance for the trace of the resolvent.** In [3], an important preliminary step to establish the CLT for linear statistics is to compute the CLT for the trace of the resolvent. Let $\mathcal{V}$ be the second moment of the random variable X_{ij} and κ its fourth cumulant:

$$\mathcal{V} = \mathbb{E}(X_{ij}^n)^2 \quad \text{and} \quad \kappa = \mathbb{E}\left|X_{ij}^n\right|^4 - |\mathcal{V}|^2 - 2 \ .$$

[2] We borrow the name "canonical equation" from V.L. Girko who established in [12, 13] canonical equations associated to various models of large random matrices.

If the entries are real or complex standard Gaussian, then $\kappa = 0$ and $\mathcal{V} = 0$ or 1. Otherwise the fourth cumulant is a priori no longer equal to zero. This induces extra-terms in the computation of the limiting variance, mainly due to the following $(\mathcal{V}, \kappa)$-dependent identity:

$$\mathbb{E}(X_{\cdot 1}^* A X_{\cdot 1} - \operatorname{tr} A)(X_{\cdot 1}^* B X_{\cdot 1} - \operatorname{tr} B) = \operatorname{tr} AB + |\mathcal{V}|^2 \operatorname{tr} AB^T + \kappa \sum_{i=1}^N A_{ii} B_{ii} \ , \tag{2.6}$$

where $X_{\cdot 1}$ stands for the first column (of dimension $N \times 1$) of matrix X_n and where A, B are deterministic $N \times N$ matrices. As a consequence, there will be three terms in the limiting covariance of the quantity (1.3); one will raise from the first term of the right hand side (r.h.s.) of (2.6), a second one will be proportional to $|\mathcal{V}|^2$, and a third one will be proportional to κ. As will be described in the sequel, the two last terms behave differently than the first one. In order to describe these terms, we first need to introduce more notations. Let

$$\tilde{t}_n(z) \quad = \quad -\frac{1 - c_n}{z} + c_n t_n(z) \ . \tag{2.7}$$

The quantity $\tilde{t}_n(z)$ is the deterministic equivalent associated to $n^{-1}\operatorname{tr}(\Sigma_n^* \Sigma_n - z I_n)^{-1}$. Denote by $\bar{R}_n$ the (entry-wise) conjugate matrix of R_n, and by $\bar{T}_n$, the matrix[3]:

$$\bar{T}_n(z) = \left(-z I_N + (1 - c_n)\bar{R}_n - z c_n t_n(z) \bar{R}_n\right)^{-1} \ ; \tag{2.8}$$

notice that the definition of $t_n(z)$ in (2.3) does not change if R_n is replaced by $\bar{R}_n$ since the spectrum of both matrices R_n and $\bar{R}_n$ is the same. We can now describe the limiting covariance of the trace of the resolvent, which is a key step in Bai and Silverstein's approach:

$$\begin{aligned} \operatorname{cov}\left(\operatorname{tr} Q_n(z_1), \operatorname{tr} Q_n(z_2)\right) &= \Theta_{0,n}(z_1, z_2) + |\mathcal{V}|^2 \Theta_{1,n}(z_1, z_2) + \kappa \Theta_{2,n}(z_1, z_2) + o(1) \\ &\triangleq \Theta_n(z_1, z_2) + o(1) \ , \end{aligned} \tag{2.9}$$

where $o(1)$ is a term that converges to zero as $N, n \to \infty$ and

$$\Theta_{0,n}(z_1, z_2) \quad \triangleq \quad \left\{ \frac{\tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{(\tilde{t}_n(z_1) - \tilde{t}_n(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right\} \tag{2.10}$$

$$\Theta_{1,n}(z_1, z_2) \quad \triangleq \quad \frac{\partial}{\partial z_2} \left\{ \frac{\partial \mathcal{A}_n(z_1, z_2)}{\partial z_1} \frac{1}{1 - |\mathcal{V}|^2 \mathcal{A}_n(z_1, z_2)} \right\} \tag{2.11}$$

$$\Theta_{2,n}(z_1, z_2) \quad \triangleq \quad \frac{z_1^2 z_2^2 \tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{n} \sum_{i=1}^N \left(R_n^{1/2} T_n^2(z_1) R_n^{1/2} \right)_{ii} \left(R_n^{1/2} T_n^2(z_2) R_n^{1/2} \right)_{ii} \tag{2.12}$$

with

$$\mathcal{A}_n(z_1, z_2) = \frac{z_1 z_2}{n} \tilde{t}_n(z_1)\tilde{t}_n(z_2) \operatorname{tr} \left\{ R_n^{1/2} T_n(z_1) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z_2) \bar{R}_n^{1/2} \right\} \ . \tag{2.13}$$

At first sight, these formulas (established in Section 4) may seem complicated; however, much information can be inferred from them.

The term $\Theta_{0,n}$. This term is familiar as it already appears in Bai and Silverstein's CLT [3]. Notice that the quantities $\tilde{t}_n$ and $\tilde{t}_n'$ only depend on the spectrum of matrix R_n. Hence, under the additional assumption that:

$$c_n \xrightarrow[N,n\to\infty]{} c \in (0, \infty) \quad \text{and} \quad F^{R_n} \xrightarrow[N,n\to\infty]{\mathcal{L}} F^{\mathbf{R}} \ , \tag{2.14}$$

[3]Beware that $\bar{T}_n$ is **not** the entry-wise conjugate of T_n, due to the presence of z.

where F^{R_n} denotes the empirical distribution of R_n's eigenvalues and $F^{\mathbf{R}}$ is a probability measure, it can easily be proved that

$$\Theta_{0,n}(z_1,z_2) \xrightarrow[N,n\to\infty]{} \Theta_0(z_1,z_2) = \left\{ \frac{\tilde{t}'(z_1)\tilde{t}'(z_2)}{(\tilde{t}(z_1)-\tilde{t}(z_2))^2} - \frac{1}{(z_1-z_2)^2} \right\} , \tag{2.15}$$

where $\tilde{t}, \tilde{t}'$ are the limits of $\tilde{t}_n, \tilde{t}'_n$ under (2.14).

The term $\Theta_{1,n}$. The interesting phenomenon lies in the fact that this term involves products of matrices $R_n^{1/2}$ and its conjugate $\bar{R}_n^{1/2}$. These matrices have the same spectrum but conjugate eigenvectors. If R_n is not real, the convergence of $\Theta_{1,n}$ is not granted, even under (2.14). If however R_n and X_n's entries are real, i.e. $\mathcal{V}=1$, then it can be easily proved that $\Theta_{0,n}=\Theta_{1,n}$ hence the factor 2 in [3] between the complex and the real covariance.

The term $\Theta_{2,n}$. This term involves quantities of the type $(R_n^{1/2}T_nR_n^{1/2})_{ii}$ which not only depend on the spectrum of matrix R_n but also on its eigenvectors. As a consequence, the convergence of such terms does not follow from an assumption such as (2.14), except in some particular cases (for instance if R_n is diagonal) and any assumption which enforces the convergence of such terms (as for instance in [25, Theorem 1.4]) implicitely implies an asymptotic joint behaviour between R_n's eigenvectors and eigenvalues. We shall adopt a different point of view here and will not assume the convergence of these quantities.

2.3. **Representation of the linear statistics and limiting bias.** Recall that $t_n(z)$ is the Stieltjes transform of a probability measure F_n:

$$t_n(z) = \int \frac{F_n(d\lambda)}{\lambda - z} . \tag{2.16}$$

The purpose of this article is to describe the fluctuations of the linear statistics

$$L_n(f) = \sum_{i=1}^{N} f(\lambda_i) - N \int f(\lambda)\, F_n(d\lambda) \tag{2.17}$$

as $N,n\to\infty$, in the case where the function f is analytic in a neighborhood of the limiting support (to be properly defined) of the spectrum of $\Sigma_n\Sigma_n^*$ and $\mathcal{C}$ is a contour surrounding this limiting spectrum. The function f being analytic, one can rely on the following version of Cauchy representation formula:

$$\begin{aligned} L_n(f) &= \operatorname{tr} f(\Sigma_n\Sigma_n^*) - N\int f(\lambda)F_n(d\lambda) \\ &= -\frac{1}{2i\pi}\oint_C f(z)\,\{\operatorname{tr} Q_n(z) - Nt_n(z)\}\,dz . \end{aligned} \tag{2.18}$$

where the last equality follows from the fact that

$$\int f(\lambda)F_n(d\lambda) = -\frac{1}{2\mathbf{i}\pi}\oint f(z)t_n(z)dz ,$$

an immediate consequence of (2.16), Cauchy's representation formula and Fubini's theorem. Based on (2.18), we shall first study the fluctuations of:

$$\begin{aligned} &\operatorname{tr} Q_n(z) - Nt_n(z) \\ &= \operatorname{tr} Q_n(z) - \mathbb{E}\operatorname{tr} Q_n(z) + \mathbb{E}\operatorname{tr} Q_n(z) - Nt_n(z) \end{aligned}$$

for $z \in \mathcal{C}$. The first difference in the r.h.s. will yield the fluctuations with a covariance $\Theta_n(z_1, z_2)$ described in (2.9) while the second difference, deterministic, will yield the bias:

$$\begin{aligned} \mathbb{E}\mathrm{tr}\, Q_n(z) - N t_n(z) &= |\mathcal{V}|^2 \mathcal{B}_{1,n}(z) + \kappa \mathcal{B}_{2,n}(z) + o(1) \\ &\triangleq \mathcal{B}_n(z) + o(1) \end{aligned} \tag{2.19}$$

where

$$\mathcal{B}_{1,n}(z) \triangleq -z^3 \tilde{t}_n^3 \frac{\frac{1}{n}\mathrm{tr}\, R_n^{1/2} T_n^2(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2}}{\left(1 - z^2 \tilde{t}_n^2 \frac{1}{n}\mathrm{Tr}\, R_n^2 T_n^2\right)\left(1 - |\mathcal{V}|^2 z^2 \tilde{t}_n^2 \frac{1}{n}\mathrm{Tr}\, R_n^{1/2} T_n(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2}\right)} \tag{2.20}$$

$$\mathcal{B}_{2,n}(z) \triangleq -z^3 \tilde{t}_n^3 \frac{\frac{1}{n}\sum_{i=1}^N \left(R_n^{1/2} T_n R_n^{1/2}\right)_{ii} \left(R_n^{1/2} T_n^2 R_n^{1/2}\right)_{ii}}{1 - z^2 \tilde{t}_n^2 \frac{1}{n}\mathrm{tr}\, R_n^2 T_n^2} \tag{2.21}$$

The discussions on the terms $\Theta_{1,n}$ and $\Theta_{2,n}$ also apply to the terms $\mathcal{B}_{1,n}$ and $\mathcal{B}_{2,n}$ (whose expressions are established in Section 4) which are likely not to converge for similar reasons.

2.4. **Gaussian processes and the central limit theorem.** A priori, the mean $\mathcal{B}_n$ and covariance Θ_n of $(\mathrm{tr}\, Q_n - N t_n)$ do not converge. Hence, we shall express the Gaussian fluctuations of the linear statistics (2.17) in the following way: we first prove the existence of a family $(N_n(z), z \in \mathcal{C})_{n\in\mathbb{N}}$ of tight Gaussian processes with mean and covariance:

$$\begin{aligned} \mathbb{E}N_n(z) &= \mathcal{B}_n(z)\ , \\ \mathrm{cov}(N_n(z_1), N_n(z_2)) &= \Theta_n(z_1, z_2)\ . \end{aligned}$$

If X is a $\mathbb{R}$ or $\mathbb{R}^d$-valued random variable, denote by $\mathcal{L}(X)$ its distribution. If P and Q are probability measures over $\mathbb{R}^d$, the Lévy-Prohorov distance between P and Q is defined as:

$$d_{\mathcal{L}P}(P, Q) = \inf\left\{\varepsilon > 0,\ P(A) \leq Q(A^\varepsilon) + \varepsilon \quad \text{for all Borel sets } A \subset \mathbb{R}^d\right\}\ , \tag{2.22}$$

where A^ε is an ε-blow up of A (cf. [11, Section 11.3] for more details). If X and Y are random variables, we denote (with a slight abuse of notation) by $d_{\mathcal{L}P}(X, Y) \triangleq d_{\mathcal{L}P}(\mathcal{L}(X), \mathcal{L}(Y))$.

The fluctuations of the centralized trace will be described as:

$$d_{\mathcal{L}P}\left(\left(\mathrm{tr}\, Q_n(z) - N t_n(z)\right), N_n(z)\right) \xrightarrow[N,n\to\infty]{} 0\ .$$

A multi-dimensional version of the previous convergence together with a tightness argument will yield:

$$d_{\mathcal{L}P}\left(L_n(f), -\frac{1}{2\mathbf{i}\pi}\oint_{\mathcal{C}} f(z) N_n(z)\, dz\right) \xrightarrow[N,n\to\infty]{} 0\ ,$$

where $-\frac{1}{2\mathbf{i}\pi}\oint_{\mathcal{C}} f(z) N_n(z)\, dz$ is a Gaussian random variable with well-identified parameters.

3. Statement of the CLT

3.1. **Assumptions and further notations.** Recall the asymptotic regime where $N, n \to \infty$, cf. (1.2), and denote by

$$c_n = \frac{N}{n}\ , \qquad \ell^- \triangleq \liminf \frac{N}{n} \qquad \text{and} \qquad \ell^+ \triangleq \limsup \frac{N}{n}\ .$$

Assumption A-1. *The random variables* $(X_{ij}^n \ ; \ 1 \leq i \leq N(n), 1 \leq j \leq n\,, n \geq 1)$ *are independent and identically distributed. They satisfy* $\mathbb{E}X_{ij}^n = 0$, $\mathbb{E}|X_{ij}^n|^2 = 1$ *and* $\mathbb{E}|X_{ij}^n|^4 < \infty$.

Assumption A-2. *Consider a sequence* (R_n) *of deterministic, nonnegative definite hermitian* $N \times N$ *matrices, with* $N = N(n)$. *The sequence* $(R_n, n \geq 1)$ *is bounded for the spectral norm as* $N \to \infty$*:*

$$\sup_{n \geq 1} \|R_n\| < \infty \ .$$

In particular, we will have:

$$0 \quad \leq \quad \boldsymbol{\lambda}_R^- \quad \stackrel{\triangle}{=} \quad \liminf_{N,n\to\infty} \|R_n\| \quad \leq \quad \boldsymbol{\lambda}_R^+ \quad \stackrel{\triangle}{=} \quad \limsup_{N,n\to\infty} \|R_n\| \quad < \quad \infty \ .$$

Assumption A-3. *Let* f *be a function on* $\mathbb{R}$, *analytic on an open interval containing*

$$\left[\boldsymbol{\lambda}_R^- \left(1 - \sqrt{\boldsymbol{\ell}^{\pm}}\right)^2, \boldsymbol{\lambda}_R^+ \left(1 + \sqrt{\boldsymbol{\ell}^{+}}\right)^2\right] \ , \tag{3.1}$$

where $\left(1 - \sqrt{\boldsymbol{\ell}^{\pm}}\right)^2 = \min\left(\left(1 - \sqrt{\boldsymbol{\ell}^{-}}\right)^2 \ ; \ \left(1 - \sqrt{\boldsymbol{\ell}^{+}}\right)^2\right)$.

If A is a $N \times N$ matrix with real eigenvalues, denote by F^A the empirical distribution of the eigenvalues $(\delta_i(A), i = 1 : N)$ of A, that is:

$$F^A(dx) = \frac{1}{n} \sum_{i=1}^{N} \delta_{\lambda_i(A)}(dx) \ .$$

Recall the definitions of Q_n, t_n, T_n and $\tilde{t}_n$ (cf. (2.1), (2.3), (2.4) and (2.7)). The following relations hold true (see for instance [3]):

$$T_n(z) \quad = \quad -\frac{1}{z}\left(I_N + \tilde{t}_n(z) R_n\right)^{-1} \quad \text{and} \quad \tilde{t}_n(z) \quad = \quad -\frac{1}{z\left(1 + \frac{1}{n}\mathrm{tr}\, R_n T_n(z)\right)} \tag{3.2}$$

Recall that F_n is the probability distribution whose Stieltjes transform is $t_n(z)$ and let $\tilde{F}_n$ be the probability distribution associated to $\tilde{t}_n(z)$. The central object of study is the signed measure:

$$G_n = N\left(F^{\Sigma_n \Sigma_n^*} - F_n\right) = n\left(F^{\Sigma_n^* \Sigma_n} - \tilde{F}_n\right) \ . \tag{3.3}$$

3.2. The Central Limit Theorem.

Theorem 3.1. *Assume that (A-1) and (A-2) hold true, and let* $f_1, \cdots, f_k$ *satisfy (A-3). Consider the random vector*

$$\begin{aligned} L_n(\mathbf{f}) \quad &\stackrel{\triangle}{=} \quad (L_n(f_1), \cdots, L_n(f_k)) \\ &= \quad \left(\int f_1(x) d\, G_n(x), \cdots, \int f_k(x) d\, G_n(x)\right) \end{aligned}$$

and the Gaussian random vector $Z_n(\mathbf{f}) \stackrel{\triangle}{=} (Z_n(f_1), \cdots, Z_n(f_k))$ *with means*

$$\mathbb{E}Z_n(f) \quad = \quad \frac{1}{2\mathbf{i}\pi} \oint f(z) \mathcal{B}_n(z) dz \quad \textit{where} \quad \mathcal{B}_n \ = \ |\mathcal{V}|^2 \mathcal{B}_{1,n} + \kappa\, \mathcal{B}_{2,n} \tag{3.4}$$

these quantities being defined in (2.19), (2.20) *and* (2.21), *and covariance*

$$\operatorname{cov}\left(Z_n(f), Z_n(g)\right) \;=\; -\frac{1}{4\pi^2}\oint\oint f(z_1)g(z_2)\Theta_n(z_1,z_2)dz_1dz_2 \tag{3.5}$$

where $\Theta_n = \Theta_{0,n} + |\mathcal{V}|^2\Theta_{1,n} + \kappa\Theta_{2,n}$, *these quantities being defined in* (2.9), (2.10)-(2.12). *The contours in* (3.4) *and* (3.5) *(two in* (3.5) *which are assumed to be nonoverlapping) are closed, taken in the positive direction in the complex plane and enclosing the interval* (3.1).

Then, the sequence of $\mathbb{R}^k$*-valued random vectors* $Z_n(\mathbf{f})$ *is tight and the following convergence holds true:*

$$d_{\mathcal{LP}}\left(L_n(\mathbf{f}), Z_n(\mathbf{f})\right) \xrightarrow[N,n\to\infty]{} 0\ ,$$

or equivalently for every continuous bounded function $h : \mathbb{R}^k \to \mathbb{C}$,

$$\mathbb{E}\,h(L_n(\mathbf{f})) - \mathbb{E}\,h(Z_n(\mathbf{f})) \xrightarrow[N,n\to\infty]{} 0\ .$$

The proof of Theorem 3.1 heavily relies on Lemma 4.1 in Section 4, which extends [3, Lemma 1.1].

3.3. **The special case of diagonal matrices** (R_n)**.** This case partially falls into the framework developed in Pan and Zhou [25] (note that the case $\mathcal{V} \neq 0$ and 1 is not handled there). Matrix R_n being nonnegative definite hermitian, its entries are real positive if R_n is assumed to be diagonal. In this case, matrix T_n is diagonal as well (cf. (2.4)), $T_n = \bar{T}_n$ and simplifications occur for the following terms:

$$\begin{aligned}
\mathcal{A}_n &= \frac{z_1z_2}{n}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\operatorname{tr} R_nT_n(z_1)R_nT_n(z_2)\ ,\\
\Theta_{2,n}(z_1,z_2) &= \frac{z_1^2z_2^2\tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{n}\operatorname{tr}\left(R_n^2T_n^2(z_1)T_n^2(z_2)\right)\ ,\\
\mathcal{B}_{1,n}(z) &\triangleq -z^3\tilde{t}_n^3\frac{\frac{1}{n}\operatorname{tr} R_n^2T_n^3(z)}{\left(1-z^2\tilde{t}_n^2\,\frac{1}{n}\operatorname{Tr} R_n^2T_n^2\right)\left(1-|\mathcal{V}|^2z^2\tilde{t}_n^2\,\frac{1}{n}\operatorname{Tr} R_n^2T_n^2(z)\right)}\ ,\\
\mathcal{B}_{2,n}(z) &\triangleq -z^3\tilde{t}_n^3\frac{\frac{1}{n}\operatorname{tr} R_n^2T_n^3}{1-z^2\tilde{t}_n^2\,\frac{1}{n}\operatorname{tr} R_n^2T_n^2}\ .
\end{aligned}$$

As one may notice, all the terms in the variance and the bias now only depend on the spectrum of R_n. Hence, the following convergence holds true under the extra assumption (2.14):

$$\begin{aligned}
\mathcal{A}_n(z_1,z_2) &\xrightarrow[N,n\to\infty]{} \mathcal{A}(z_1,z_2) = c\,\tilde{t}(z_1)\tilde{t}(z_2)\int\frac{\lambda^2F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z_1))(1+\lambda\tilde{t}(z_2))}\ ,\\
\Theta_{1,n}(z_1,z_2) &\xrightarrow[N,n\to\infty]{} \Theta_1(z_1,z_2) = \frac{\partial}{\partial z_2}\left\{\frac{\partial\mathcal{A}(z_1,z_2)}{\partial z_1}\frac{1}{1-|\mathcal{V}|^2\mathcal{A}(z_1,z_2)}\right\}\ ,\\
\Theta_{2,n}(z_1,z_2) &\xrightarrow[N,n\to\infty]{} \Theta_2(z_1,z_2) = c\,\tilde{t}'(z_1)\tilde{t}'(z_2)\int\frac{\lambda^2F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z_1))^2(1+\lambda\tilde{t}(z_2))^2}\ ,\\
\mathcal{B}_{1,n}(z) &\xrightarrow[N,n\to\infty]{} \mathcal{B}_1(z) = -\frac{cz^3\tilde{t}^3(z)}{(1-\mathcal{A}(z,z))(1-|\mathcal{V}|^2\mathcal{A}(z,z))}\int\frac{\lambda^2F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z))^3}\ ,\\
\mathcal{B}_{2,n}(z) &\xrightarrow[N,n\to\infty]{} \mathcal{B}_2(z) = -\frac{cz^3\tilde{t}^3(z)}{1-\mathcal{A}(z,z)}\int\frac{\lambda^2F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z))^3}\ .
\end{aligned}$$

where $\tilde{t}, \tilde{t}'$ are the limits of $\tilde{t}_n, \tilde{t}'_n$ under (2.14). This can be packaged into the following result:

Corollary 3.2. *Assume that (A-1) and (A-2) hold true. Assume moreover that R_n is diagonal, that the convergence assumption (2.14) holds true and let $f_1, \cdots, f_k$ satisfy (A-3). Consider the Gaussian random vector $Z(\mathbf{f}) = (Z(f_1), \cdots, Z(f_k))$ with means*

$$\mathbb{E}Z(f) = \frac{1}{2\mathbf{i}\pi} \oint f(z)\mathcal{B}(z)dz \qquad \text{where} \quad \mathcal{B} \;=\; |\mathcal{V}|^2\mathcal{B}_1 + \kappa\,\mathcal{B}_2$$

and $\mathcal{B}_1$ and $\mathcal{B}_2$ are defined above and covariance

$$\operatorname{cov}\left(Z(f), Z(g)\right) = -\frac{1}{4\pi^2} \oint\oint f(z_1)g(z_2)\Theta(z_1, z_2)dz_1dz_2 \quad \text{where} \quad \Theta \;=\; \Theta_0 + |\mathcal{V}|^2\Theta_1 + \kappa\Theta_2$$

and Θ_0 defined in (2.15) and Θ_1, Θ_2 defined above. The contours in (3.4) and (3.5) (two in (3.5) which are assumed to be nonoverlapping) are closed, taken in the positive direction in the complex plane and enclosing the interval (3.1). Then,

$$L_n(\mathbf{f}) \xrightarrow[N,n\to\infty]{\mathcal{D}} Z(\mathbf{f}) \ .$$

3.4. **Additional computations in the case where R_n is the identity.** In this section, we assume that $R_n = I_N$.

The term proportional to $|\mathcal{V}^2|$. In this case, the quantity $\mathcal{A}(z_1, z_2)$ takes the simplified form

$$\mathcal{A}(z_1, z_2) = \frac{c\,\tilde{t}_1\tilde{t}_2}{(1+\tilde{t}_1)(1+\tilde{t}_2)} \ .$$

where we denote $\tilde{t}_i = \tilde{t}(z_i),\ i = 1, 2$. Straightforward computations yield:

$$\frac{\partial}{\partial z_i}\mathcal{A}(z_1, z_2) = \frac{\tilde{t}'_i}{(1+\tilde{t}_i)\tilde{t}_i}\mathcal{A}(z_1, z_2) \ , \ i = 1, 2 \ .$$

and

$$\Theta_1(z_1, z_2) \;=\; \frac{c\,\tilde{t}'_1\tilde{t}'_2}{(1+\tilde{t}_1)^2(1+\tilde{t}_2)^2\left(1 - |\mathcal{V}|^2\mathcal{A}(z_1, z_2)\right)^2} \;=\; \frac{c\,\tilde{t}'_1\tilde{t}'_2}{\left((1+\tilde{t}_1)(1+\tilde{t}_2) - |\mathcal{V}|^2 c\,\tilde{t}_1\tilde{t}_2\right)^2} \ .$$

If needed, one can then use the explicit expression of the Stieltjes transform of Marčenko-Pastur distribution (see below).

The cumulant term. In the particular case where $R_n = I_N$, Lytova and Pastur [23] (see also [26]) provided an explicit formula for the cumulant term of the covariance based on the Stieltjes transform of Marčenko-Pastur distribution. We recover this formula hereafter and prove that

$$-\frac{\kappa}{4\pi^2} \oint\oint \varphi(z_1)\varphi(z_2)\Theta_2(z_1, z_2)dz_1dz_2 = \frac{\kappa}{4c\pi^2}\left(\int_{\lambda^-}^{\lambda^+} \varphi(\lambda)\frac{\lambda - (1+c)}{\sqrt{(\lambda^+ - \lambda)(\lambda - \lambda^-)}}d\lambda\right)^2 \ , \tag{3.6}$$

which is in accordance[4] with Lytova and Pastur's result [23, Eq. (4.65)].

[4]Denote by the superscript $^{\text{LP}}$ the quantities in [23] and use the correspondance $c^{\text{LP}} \leftrightarrow 1/c$, $a^{\text{LP}} \leftrightarrow c$ and $\kappa_4^{\text{LP}} \leftrightarrow (a^{\text{LP}})^4\kappa = c^2\kappa$ to check that the r.h.s. of (3.6) equates the formula provided in [23].

Specifying Θ_2 in the case $R_n = I_N$, we obtain:

$$-\frac{\kappa}{4\pi^2}\oint\oint \varphi(z_1)\varphi(z_2)\Theta_2(z_1,z_2)dz_1dz_2 \tag{3.7}$$
$$= -\frac{\kappa c}{4\pi^2}\left(\oint \varphi(z)\frac{\tilde{t}'(z)}{(1+\tilde{t}(z))^2}dz\right)^2 = -\frac{\kappa c}{4\pi^2}\left(\oint \varphi'(z)\frac{1}{1+\tilde{t}(z)}dz\right)^2 = -\frac{\kappa c}{4\pi^2}\left(\oint \varphi'(z)zt(z)dz\right)^2 .$$

Now, in this case, $t(z)$ is the Stieltjes transform of Marčenko-Pastur distribution and has an explicit form (see for instance [26, Chapter 7]):

$$t(z) = \frac{1}{2cz}\left\{\sqrt{(z-(1+c))^2-4c} - (z-(1-c))\right\}$$

where the branch of the square root is fixed by its asymptotics: $z-(1+c)+o(1)$ as $z\to\infty$. We have:

$$(3.7) = -\frac{\kappa c}{4\pi^2}\left(\oint \varphi(z)(zt(z))'dz\right)^2 = -\frac{\kappa c}{4\pi^2}\left(\oint \varphi(z)\frac{z-(1+c)}{2c\sqrt{(z-(1+c))^2-4c}}dz\right)^2$$

It remains to deform the contour into the cuts $[\lambda^-,\lambda^+]$ where $\lambda^- = (1-\sqrt{c})^2$ and $\lambda^+ = (1+\sqrt{c})^2$ and to use the relations

$$\sqrt{(z-(1+c))^2-4c}\big|_{z=\lambda\pm\mathbf{i}0} = \pm\mathbf{i}\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}$$

for $\lambda\in[(1-\sqrt{c})^2,(1+\sqrt{c})^2]$ to obtain

$$(3.7) = -\frac{\kappa c}{4\pi^2}\left(\oint \varphi(z)\frac{z-(1+c)}{2c\sqrt{(z-(1+c))^2-4c}}dz\right)^2 = -\frac{\kappa c}{4\pi^2}\left(\int_{\lambda^-}^{\lambda^+} \varphi(\lambda)\frac{\lambda-(1+c)}{\mathbf{i}c\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}}d\lambda\right)^2$$
$$= \frac{\kappa}{4c\pi^2}\left(\int_{\lambda^-}^{\lambda^+} \varphi(\lambda)\frac{\lambda-(1+c)}{\sqrt{(\lambda^+-\lambda)(\lambda-\lambda^-)}}d\lambda\right)^2 ,$$

which yields (3.6).

4. Proof of Theorem 3.1

Denote by $\xrightarrow{\mathcal{P}}$ convergence in probability and by $o_P(1)$ any random variable which converges to zero in probability.

4.1. **Truncation.** In this section, we closely follow Bai and Silverstein [3]. We recall the framework developed there and introduce some additional notations.

4.1.1. *Truncation of random variables.* Consider a sequence of positive numbers (δ_n) which satisfies:

$$\delta_n\to 0, \quad \delta_n n^{1/4}\to\infty \quad \text{and} \quad \delta_n^{-4}\int_{\{|X_{11}|\geq\delta_n\sqrt{N}\}}|X_{11}|^4\to 0$$

as $N,n\to\infty$. Let $\widehat{\Sigma}_n = n^{-1/2}R_n^{1/2}\widehat{X}_n$ where $\widehat{X}_n$ is a $N\times n$ matrix having (i,j)th entry $X_{ij}1_{\{|X_{ij}|<\delta_n\sqrt{N}\}}$. This truncation step yields:

$$\mathbb{P}\left(\Sigma_n\Sigma_n^*\neq\widehat{\Sigma}_n\widehat{\Sigma}_n^*\right)\xrightarrow[N,n\to\infty]{} 0 . \tag{4.1}$$

Define $\widehat{\Sigma}_n = n^{-1/2} R_n^{1/2} \widehat{X}_n$ where $\widehat{X}_n$ is a $N \times n$ matrix having (i,j)th entry $(\widehat{X}_{ij} - \mathbb{E}\widehat{X}_{ij})/\sigma_n$, where $\sigma_n^2 = \mathbb{E}|\widehat{X}_{ij} - \mathbb{E}\widehat{X}_{ij}|^2$. Let $\widehat{G}_n = N(F^{\widehat{\Sigma}_n \widehat{\Sigma}_n^*} - F_n)$ and $\widetilde{G}_n = N(F^{\widetilde{\Sigma}_n \widetilde{\Sigma}_n^*} - F_n)$, then whenever f fulfills (A-3), the following holds true (cf. [3]):

$$\forall \eta > 0, \quad \mathbb{P}\left\{ \left| \int f d\widehat{G}_n - \int f d\widetilde{G}_n \right| > \eta \right\} \xrightarrow[N,n\to\infty]{} 0 \, . \tag{4.2}$$

Combining (4.1) and (4.2), we obtain $\int f dG_n - \int f d\widetilde{G}_n \to 0$ in probability. Moreover, the moments are asymptotically not affected by these different steps:

$$\max\left(\left| \mathbb{E}\widetilde{X}_{ij}^2 - \mathbb{E}X_{ij}^2 \right| \; ; \left(\mathbb{E}|\widetilde{X}_{ij}|^2 - 1 \right) \; ; \left(\mathbb{E}|\widetilde{X}_{ij}|^4 - \mathbb{E}|X_{ij}|^4 \right) \right) \xrightarrow[N,n\to\infty]{} 0 \, . \tag{4.3}$$

Note in particular that the fourth cumulant of $\tilde{X}_{ij}$ converges to that of X_{ij}. Hence, it is sufficient to consider variables truncated at $\delta_n \sqrt{n}$, centralized and renormalized. This will be assumed in the sequel with no superscript (we shall simply write X_{ij} and all related quantities with X_{ij}'s truncated, centralized, renormalized with no superscript any more).

4.1.2. *Truncation of process.* Let x_r be any number greater than the right endpoint of interval (3.1). Let x_ℓ be any negative number if the left endpoint of (3.1) is zero. Otherwise, let $x_\ell \in (0, \boldsymbol{\lambda}_R^-(1-\sqrt{\ell^\pm})^2)$. Let $y_0 > 0$ and consider:

$$\mathcal{C} = \{x_\ell + \mathbf{i}y : y \in [0, y_0]\} \cup \{x + \mathbf{i}y_0 : x \in [x_\ell, x_r]\} \cup \{x_r + \mathbf{i}y : y \in [0, y_0]\} \tag{4.4}$$

We also denote by $\overline{\mathcal{C}} = \{z : \bar{z} \in \mathcal{C}\}$ and by $\Gamma = \mathcal{C} \cup \overline{\mathcal{C}}$.

We now introduce the Stieltjes transform of G_n. Let $M_n(z)$ be defined as

$$M_n(z) = \operatorname{tr} Q_n(z) - N t_n(z) = \operatorname{tr} (\Sigma_n^* \Sigma_n - z I_n)^{-1} - n \tilde{t}_n(z) \tag{4.5}$$

and $M_n(\bar{z}) = \overline{M_n(z)}$.

Let (ε_n) be a real sequence decreasing to zero and satisfying for some $\alpha \in (0,1)$, $\varepsilon_n \geq n^{-\alpha}$. Let

$$\begin{aligned} \mathcal{C}_\ell &= \begin{cases} \{x_\ell + \mathbf{i}v \; : \; v \in [n^{-1}\varepsilon_n, v_0]\} & \text{if } x_\ell \geq 0 \\ \{x_\ell + \mathbf{i}v \; : \; v \in [0, v_0]\} & \text{if } x_\ell < 0 \end{cases}, \\ \mathcal{C}_r &= \{x_r + \mathbf{i}v \; : \; [n^{-1}\varepsilon_n, v_0]\} \, . \end{aligned}$$

Denote by $\mathcal{C}_n = \mathcal{C}_\ell \cup \{x + \mathbf{i}y_0 : x \in [x_\ell, x_r]\} \cup \mathcal{C}_r$. We can now define $\widehat{M}_n(\cdot)$, a truncated version of $M_n(\cdot)$:

$$\widehat{M}_n(z) = \begin{cases} M_n(z), & \text{for } z \in \mathcal{C}_n, \\ M_n(x_r + \mathbf{i}n^{-1}\varepsilon_n), & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n], \\ & \text{and if } x_\ell > 0, \\ M_n(x_\ell + in^{-1}\varepsilon_n), & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n], \end{cases} \tag{4.6}$$

If f satisfies (A-3), it has been proved in [3] that

$$\begin{aligned} &\left| \oint_\Gamma f(z) \left(M_n(z) - \widehat{M}_n(z) \right) dz \right| \\ &\qquad \leq 4 \sup_{z \in \Gamma} |f(z)| \varepsilon_n \Big(\left| \max(\lambda_{\max}(R_n)(1+\sqrt{c_n})^2, \lambda_{\max}(\Sigma_n \Sigma_n^*)) - x_r \right|^{-1} \\ &\qquad\qquad + \left| \min(\lambda_{\min}(R_n) 1_{(0,1)}(c_n)(1-\sqrt{c_n})^2, \lambda_{\min}(\Sigma_n \Sigma_n^*)) - x_\ell \right|^{-1} \Big) \end{aligned} \tag{4.7}$$

which almost surely converges to zero as $N, n \to \infty$.

It is therefore sufficient to study $\widehat{M}_n$ instead of M_n; we hence replace M_n by $\widehat{M}_n$ with no loss in the sequel.

4.2. **Extension of Bai and Silverstein's master lemma.** We state below the counterpart of [3, Lemma 1.1].

Lemma 4.1. *Assume that (A-1) and (A-2) hold true , then*

(1) *The process $\{\widehat{M}_n(\cdot)\}$ forms a tight sequence on $\mathcal{C}\cup\overline{\mathcal{C}}$, more precisely:*

$$\sup_{z_1,z_2\in\mathcal{C},n\geq 1}\frac{\mathbb{E}\left|\widehat{M}_n(z_1)-\widehat{M}_n(z_2)\right|^2}{|z_1-z_2|^2}<\infty$$

(2) *There exists a tight sequence $(N_n(z),\ z\in\mathcal{C}\cup\overline{\mathcal{C}})$ of two-dimensional Gaussian processes with mean*

$$\mathbb{E}N_n(z) \;=\; |\mathcal{V}|^2\mathcal{B}_{1,n}(z)+\kappa\mathcal{B}_{2,n}(z) \tag{4.8}$$

where $\mathcal{B}_{1,n}(z)$ and $\mathcal{B}_{2,n}(z)$ are defined in (2.20) and (2.21), and covariance:

$$\begin{aligned}\operatorname{cov}\left(N_n(z_1),N_n(z_2)\right) &= \mathbb{E}\left(N_n(z_1)-\mathbb{E}N_n(z_1)\right)\left(N_n(z_2)-\mathbb{E}N_n(z_2)\right)\\ &= \Theta_{0,n}(z_1,z_2)+|\mathcal{V}|^2\,\Theta_{1,n}(z_1,z_2)+\kappa\Theta_{2,n}(z_1,z_2)\end{aligned}$$

where $\Theta_{0,n}$, $\Theta_{1,n}$ and $\Theta_{2,n}$ are defined in (2.9), (2.10)-(2.12).

(3) *For any functional F from $C\left(\mathcal{C}\cup\overline{\mathcal{C}};\mathbb{C}\right)$ to $\mathbb{C}$, then*

$$\mathbb{E}F(\widehat{M}_n)-\mathbb{E}F\left(N_n\right)\xrightarrow[N,n\to\infty]{}0$$

Remark 4.1. (1) The tightness of the process $\{\widehat{M}_n\}$ immediatly follows from Bai and Silverstein's lemma as this result has been proved in [3, Lemma 1.1] under Assumption (A-1) with no extra conditions on the moments of the entries.
(2) Differences between Lemma 4.1 and [3, Lemma 1.1] appear in the bias and in the covariance where there are respectively two terms instead of one and three terms instead of one in [3, Lemma 1.1].
(3) Since the extra terms do not converge, we need to consider a sequence of Gaussian processes instead of a single Gaussian process as in [3, Lemma 1.1].
(4) In order to prove that the sequence of Gaussian processes is tight, we introduce a meta-matrix model to transfer the tightness of $\{\hat{M}_n\}$ to $\{N_n\}$ (see for instance Section 4.4.1).

It will be convenient to decompose $M_n(z)$ as:

$$M_n(z)=M_n^1(z)+M_n^2(z)\quad\text{where}\quad\left\{\begin{array}{rcl}M_n^1(z) &=& \operatorname{tr}Q_n(z)-\operatorname{tr}\mathbb{E}Q_n(z)\\ M_n^2(z) &=& N\left(\mathbb{E}f_n(z)-t_n(z)\right)\end{array}\right. \tag{4.9}$$

We shall naturally extend this decomposition to $\widehat{M}_n(z)=\widehat{M}_n^1(z)+\widehat{M}_n^2(z)$.

Denote by ξ_j the $N\times 1$ vector $\xi_j=\Sigma_{\cdot j}=\frac{1}{\sqrt{n}}R^{1/2}X_{\cdot j}$ and by $\mathbb{E}_j$ the conditional expectation with respect to the σ-field generated by $\xi_1,\cdots,\xi_j$; by convention, $\mathbb{E}_0=\mathbb{E}$.

We split Lemma 4.1 into intermediate results.

Proposition 4.2. *Assume that (A-1) and (A-2) hold true; let $z_1, z_2 \in \Gamma = \mathcal{C} \cup \bar{\mathcal{C}}$, then:*

$$\widehat{M}_n^1(z_1) = \sum_{j=1}^{n} Z_j^n(z_1) + o_P(1) \ ,$$

where the Z_j^n's are martingale increments and

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) - \Theta_n(z_1, z_2) \quad \xrightarrow[N,n\to\infty]{\mathcal{P}} \quad 0 \ , \tag{4.10}$$

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) \overline{Z_j^n(z_2)} - \Theta_n(z_1, \overline{z_2}) \quad \xrightarrow[N,n\to\infty]{\mathcal{P}} \quad 0 \ , \tag{4.11}$$

where Θ_n is defined in (2.9) and

$$\widehat{M}_n^2(z_1) - \mathcal{B}_n(z_1) \xrightarrow[N,n\to\infty]{} 0 \ ,$$

where $\mathcal{B}_n$ is defined in (2.19).

Proposition 4.3. *There exists a tight sequence $(N_n(z),\ z \in \Gamma)$ of two-dimensional Gaussian processes with mean $\mathbb{E}N_n(z) = \mathcal{B}_n(z)$ and covariance*

$$\begin{aligned} \operatorname{cov}\left(N_n(z_1), N_n(z_2)\right) &= \mathbb{E}\left(N_n(z_1) - \mathbb{E}N_n(z_1)\right)\left(N_n(z_2) - \mathbb{E}N_n(z_2)\right) \\ &= \Theta_n(z_1, z_2) \ . \end{aligned}$$

The proofs of both propositions follow hereafter.

4.3. **Proof for Proposition 4.2.** The fact that $(\widehat{M}_n)$ is a tight sequence has already been established in [3] (regardless of the assumption $\kappa = 0$ and $|\mathcal{V}| = 0/1$). In order to proceed, we shall heavily rely on the proof of [3, Lemma 1.1] which is the crux of Bai and Silverstein's paper. In Section 4.3.1 we review the main steps of Bai and Silverstein's computations of the variance/covariance. In Section 4.3.3, we compute the limiting variance. In Section 4.3.4, we compute the limiting bias (some details are postponed to Appendix A). In Section 4.5, we finally conclude the proof of Lemma 4.1 and address various subtleties which appear due to the existence of a sequence of Gaussian limiting processes.

Recall that $Q_n(z) = (-zI_N + \Sigma_n\Sigma_n^*)^{-1}$, that $\xi_j = \Sigma_{\cdot j} = \frac{1}{\sqrt{n}} R^{1/2} X_{\cdot j}$ and denote by $Q_j(z)$ the resolvent of matrix $\Sigma\Sigma^* - \xi_j\xi_j^*$, i.e.

$$Q_j(z) = \left(-zI + \Sigma\Sigma^* - \xi_j\xi_j^*\right)^{-1} \ .$$

The following quantities will be needed:

$$\begin{aligned} f_n(z) &= \frac{1}{N}\mathrm{tr}\, Q_n(z)\ , \\ \beta_j(z) &= \frac{1}{1+\xi_j^* Q_j(z)\xi_j}\ , \\ \bar\beta_j(z) &= \frac{1}{1+\frac{1}{n}\mathrm{tr}\, R_n Q_j(z)}\ , \\ b_n(z) &= \frac{1}{1+\frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n Q_1(z)}\ , \\ \varepsilon_j(z) &= \xi_j^* Q_j(z)\xi_j - \frac{1}{n}\mathrm{tr}\, R_n Q_j(z)\ , \\ \delta_j(z) &= \xi_j^* Q_j^2(z)\xi_j - \frac{1}{N}\mathrm{tr}\, R_n Q_j^2(z) \ = \ \frac{d}{dz}\varepsilon_j(z)\ . \end{aligned}$$

We shall frequently drop subscript n and write Q and R instead of Q_n and R_n in the sequel.

4.3.1. *Preliminary variance computations.* We briefly review in this section the main steps related to the computation of the variance/covariance as presented in [3]. These standard steps will finally lead to Eq. (4.15) which will be the starting point of the computations associated to the $|\mathcal{V}|^2$- and κ-terms of the variance.

Let $z \in \mathcal{C}_n$.

$$\begin{aligned} N\left(f_n(z) - \mathbb{E}f_n(z)\right) &= -\sum_{j=1}^n \left(\mathbb{E}_j - \mathbb{E}_{j-1}\right)\beta_j(z)\xi_j^* Q_j^2(z)\xi_j \\ &= -\sum_{j=1}^n \mathbb{E}_j\left(\bar\beta_j(z)\delta_j(z) - \bar\beta_j^2(z)\varepsilon_j(z)\frac{1}{n}\mathrm{tr}\, R\, Q_j^2\right) + o_P(1)\ . \end{aligned}$$

Denote by

$$Z_j^n(z) = -\mathbb{E}_j\left(\bar\beta_j(z)\delta_j(z) - \bar\beta_j^2(z)\varepsilon_j(z)\frac{1}{n}\mathrm{tr}\, R\, Q_j^2(z)\right) = -\mathbb{E}_j\frac{d}{dz}\left(\bar\beta_j(z)\varepsilon_j(z)\right)\ .$$

Hence,

$$\forall z \in \mathcal{C}_n\ , \quad N\left(f_n(z) - \mathbb{E}f_n(z)\right) = \sum_{j=1}^n Z_j^n(z) + o_P(1)$$

or equivalently

$$\forall z \in \mathcal{C}\ , \quad \widehat{M}_n(z) = \sum_{j=1}^n Z_j^n(z) + o_P(1)\ .$$

The r.h.s. appears as a sum of martingale increments. Such a decomposition is important since it will enable us to rely on powerful CLTs for martingales (see [7, Theorem 35.12], and the variations below in Lemmas 4.7 and 4.8). These CLTs rely on the study of the terms:

$$\sum_{j=1}^n \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) \quad \text{and} \quad \sum_{j=1}^n \mathbb{E}_{j-1} Z_j^n(z_1)\overline{Z_j^n(z_2)}\ .$$

Notice that since $\overline{Z_j^n(z)} = Z_j^n(\bar{z})$, we have $\mathbb{E}_{j-1} Z_j^n(z_1)\overline{Z_j^n(z_2)} = \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(\overline{z_2})$ with $\overline{z_2} \in \Gamma$ for $z_2 \in \Gamma$; it is thus sufficient to study the limiting behavior of:

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) \ , \quad z_1, z_2 \in \Gamma$$

in order to prove (4.10) and (4.11). Now,

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) = \frac{\partial^2}{\partial z_1 \partial z_2} \left\{ \sum_{j=1}^{n} \mathbb{E}_{j-1} \left[\mathbb{E}_j \left(\bar{\beta}_j(z_1) \varepsilon_j(z_1) \right) \mathbb{E}_j \left(\bar{\beta}_j(z_2) \varepsilon_j(z_2) \right) \right] \right\} . \tag{4.12}$$

Following the same arguments as in [3, pp. 571], one can prove that it is sufficient to study the convergence in probability of

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} \left[\mathbb{E}_j \left(\bar{\beta}_j(z_1) \varepsilon_j(z_1) \right) \mathbb{E}_j \left(\bar{\beta}_j(z_2) \varepsilon_j(z_2) \right) \right] .$$

Moreover,

$$\begin{aligned} &\sum_{j=1}^{n} \mathbb{E}_{j-1} \left[\mathbb{E}_j \left(\bar{\beta}_j(z_1) \varepsilon_j(z_1) \right) \mathbb{E}_j \left(\bar{\beta}_j(z_2) \varepsilon_j(z_2) \right) \right] \\ &\quad = \sum_{j=1}^{n} b_n(z_1) b_n(z_2) \mathbb{E}_{j-1} \left[\mathbb{E}_j \varepsilon_j(z_1) \mathbb{E}_j \varepsilon_j(z_2) \right] + o_P(1) \ , \\ &\quad = \sum_{j=1}^{n} z_1 \tilde{t}_n(z_1) z_2 \tilde{t}_n(z_2) \mathbb{E}_{j-1} \left[\mathbb{E}_j \varepsilon_j(z_1) \mathbb{E}_j \varepsilon_j(z_2) \right] + o_P(1) \ . \end{aligned} \tag{4.13}$$

Hence, it is finally sufficient to study the limiting behaviour (in terms of convergence in probability) of the quantity:

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} \left(\mathbb{E}_j \, \varepsilon_j(z_1) \mathbb{E}_j \, \varepsilon_j(z_2) \right) \ , \quad z_1, z_2 \in \Gamma \ . \tag{4.14}$$

Denote by A^T the transpose matrix of A. Applying (2.6) yields :

$$\begin{aligned} \sum_{j=1}^{n} \mathbb{E}_{j-1} \left(\mathbb{E}_j \, \varepsilon_j(z_1) \mathbb{E}_j \, \varepsilon_j(z_2) \right) &= \frac{1}{n^2} \sum_{j=1}^{n} \mathrm{tr} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R \mathbb{E}_j Q_j(z_2) R^{1/2} \right) \\ &\quad + \frac{|\mathcal{V}|^2}{n^2} \sum_{j=1}^{n} \mathrm{tr} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2} \right)^T \right) \\ &\quad + \frac{\kappa}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2} \right)_{ii} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2} \right)_{ii} . \end{aligned} \tag{4.15}$$

The limiting behaviour of the first term of the r.h.s. has been completely described in [3] where it has been shown that:

$$\frac{\partial^2}{\partial z_1 \partial z_2}\left\{ z_1 z_2 \tilde{t}_n(z_1)\tilde{t}_n(z_2)\frac{1}{n^2}\sum_{j=1}^{n} \mathrm{tr}\,\left(R^{1/2}\mathbb{E}_j Q_j(z_1) R \mathbb{E}_j Q_j(z_2) R^{1/2}\right)\right\} = \Theta_{0,n}(z_1,z_2)+o_P(1)\,, \tag{4.16}$$

with $\Theta_{0,n}(z_1,z_2)$ is defined in (2.10).

We shall focus on the second and third terms.

4.3.2. *The term proportional to $|\mathcal{V}|^2$ in the variance.* Notice first that the value of t_n and $\tilde{t}_n$ is the same wether R is replaced by $\bar{R}$ in (2.3) and (3.2) since t_n and $\tilde{t}_n$ only depend on the spectrum of R (which is the same as the spectrum of $\bar{R}$). Notice also that $(R^{1/2})^T = \bar{R}^{1/2}$, hence:

$$\left(R^{1/2}\mathbb{E}_j Q_j(z_2) R^{1/2}\right)^T = \bar{R}^{1/2}\mathbb{E}_j Q_j^T(z_2)\bar{R}^{1/2}\ .$$

Recall the definition of $\bar{T}_n(z)$ given by (2.8). Taking into account the fact that for a deterministic matrix A,

$$\mathbb{E}\xi_j^T A\,\xi_j = \frac{\mathcal{V}}{n}\,\mathrm{tr}\,(\bar{R}^{1/2}AR^{1/2}) \quad \text{and} \quad \mathbb{E}\xi_j^* A\,\bar{\xi}_j = \frac{\bar{\mathcal{V}}}{n}\,\mathrm{tr}\,(R^{1/2}A\bar{R}^{1/2})\,, \tag{4.17}$$

and following closely [3, Section 2], it is a matter of bookkeeping[5] to establish that:

$$\begin{aligned}
&\frac{|\mathcal{V}|^2 z_1 z_2}{n^2}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\sum_{j=1}^{n}\mathrm{tr}\,\left(R^{1/2}\mathbb{E}_j Q_j(z_1)R^{1/2}\left(R^{1/2}\mathbb{E}_j Q_j(z_2)R^{1/2}\right)^T\right) \qquad (4.18)\\
&\quad = |\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)\times\frac{1}{n}\sum_{j=1}^{n}\frac{1}{1-\left(\frac{j-1}{n}\right)|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}+o_P(1)\\
&\quad = \int_0^{|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}\frac{dz}{1-z}+o_P(1)
\end{aligned}$$

where

$$\mathcal{A}_n(z_1,z_2) = \frac{z_1 z_2}{n}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\mathrm{tr}\,\left\{R^{1/2}T_n(z_1)R^{1/2}\bar{R}^{1/2}\bar{T}_n(z_2)\bar{R}^{1/2}\right\}\ .$$

Finally,

$$\frac{\partial^2}{\partial z_1\partial z_2}(4.18) \;=\; |\mathcal{V}|^2\Theta_{1,n}(z_1,z_2)+o_P(1) \;=\; |\mathcal{V}|^2\frac{\partial}{\partial z_2}\left\{\frac{\partial\mathcal{A}_n(z_1,z_2)/\partial z_1}{1-|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}\right\}+o_P(1)\,. \tag{4.19}$$

4.3.3. *The cumulant term in the variance.* We now handle the term proportional to κ in (4.15):

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{N}\left(R^{1/2}\mathbb{E}_j Q_j(z_1)R^{1/2}\right)_{ii}\left(R^{1/2}\mathbb{E}_j Q_j(z_2)R^{1/2}\right)_{ii}\,. \tag{4.20}$$

The objective is to prove that $\mathbb{E}_j Q_j(z)$ can be replaced by $T_n(z)$ in the formula above, which boils down to prove a convergence of quadratic forms of the type (2.5). Such a convergence has already been established in [17] for large covariance matrices based on a non-centered matrix model with separable variance profile.

[5]Similar computations for the term proportional to $|\mathcal{V}|^2$ in the bias are outlined in Appendix A.

By interpolating between the quantity (4.20) and its counterpart when the entries are complex i.i.d. standard Gaussian, we will be able to rely on the results in [17] by using the unitary invariance of a Gaussian matrix (see Proposition 4.5 and Eq. (4.26) below).

Let $\boldsymbol{\delta}_z$ be the distance between the point $z \in \mathbb{C}$ and the real nonnegative axis $\mathbb{R}^+$:

$$\boldsymbol{\delta}_z \;=\; \mathrm{dist}(z, \mathbb{R}^+) \; . \tag{4.21}$$

Proposition 4.4. *Assume that (A-1) and (A-2) hold true and let u_n be a deterministic $N \times 1$ vector, then:*

$$\mathbb{E}\,|u_n^* Q(z) u_n - u_n^*\,\mathbb{E}Q(z)\,u_n|^2 \leq \frac{1}{n}\Phi(|z|)\Psi\left(\frac{1}{\boldsymbol{\delta}_z}\right)\|u_n\|^2 \; ,$$

where Φ and Ψ are fixed polynomials with coefficients independent from N, n, z and (u_n).

Proof of Proposition 4.4 is an easy adaptation[6] of [17, Prop. 2.7].

Denote by $\tilde{X}_n$ a $N \times n$ matrix whose entries are independent standard complex circular Gaussian r.v. (i.e. $\tilde{X}_{ij} = U + iV$ where U, V are independent $\mathcal{N}(0, 2^{-1})$ random variables); denote accordingly $\tilde{\Sigma}_n = n^{-1/2} R^{1/2} \tilde{X}_n$, $\tilde{\xi}_j = \left(\tilde{\Sigma}_n\right) \cdot j$ and $\tilde{Q}_n(z) = (-zI_N + \tilde{\Sigma}_n \tilde{\Sigma}_n^*)^{-1}$. We now drop subscripts N and n.

Proposition 4.5. *Assume that (A-1) and (A-2) hold true and let u_n be a deterministic $N \times 1$ vector, then:*

$$\left|u_n^* \mathbb{E}\, Q(z) u_n - u_n^*\,\mathbb{E}\tilde{Q}(z)\,u_n\right| \leq \frac{1}{\sqrt{n}}\Phi(|z|)\Psi\left(\frac{1}{\boldsymbol{\delta}_z}\right)\|u_n\|^2 \; ,$$

where Φ and Ψ are fixed polynomials with coefficients independent from N, n, z and (u_n).

Proof. Consider the resolvent

$$Q^{(i)}(z) = \left(\sum_{\ell=1}^{i} \tilde{\xi}_i \tilde{\xi}_i^* + \sum_{\ell=i+1}^{n} \xi_i \xi_i^* - zI_N\right)^{-1}$$

defined for $1 \leq i \leq n-1$. Denote by $Q^{(0)} = Q$ and by $Q^{(n)} = \tilde{Q}$ and write

$$u^* \mathbb{E}(Q - \tilde{Q})u = \sum_{i=1}^{n} u^* \mathbb{E}(Q^{(i-1)} - Q^{(i)})u \; . \tag{4.22}$$

We shall evaluate the difference $u^*\mathbb{E}(Q^{(0)} - Q^{(1)})u$, the other ones being handled similarly. Denote by $\check{Q}(z) = (\sum_{i=2}^{n} \xi_i \xi_i^* - zI_N)^{-1}$, then:

$$Q^{(0)} = \check{Q} - \frac{\check{Q}\xi_1 \xi_1^* \check{Q}}{1 + \xi_1^* \check{Q} \xi_1} \quad \text{and} \quad Q^{(1)} = \check{Q} - \frac{\check{Q}\tilde{\xi}_1 \tilde{\xi}_1^* \check{Q}}{1 + \tilde{\xi}_1^* \check{Q} \tilde{\xi}_1} \; .$$

[6] Notice in particular all the cancellations that appear when adapting the proof of [17, Prop. 2.7], due to the fact that Σ_n is centered here; notice also the fact that R not being diagonal has virtually no impact.

Dropping the subscript 1 to lighten the notations, we get:

$$\begin{aligned} u^*\mathbb{E}\left(Q^{(0)}-Q^{(1)}\right)u &= u^*\mathbb{E}\left(\frac{\check{Q}\tilde{\xi}\tilde{\xi}^*\check{Q}}{1+\tilde{\xi}^*\check{Q}\tilde{\xi}}-\frac{\check{Q}\xi\xi^*\check{Q}}{1+\xi^*\check{Q}\xi}\right)u \\ &= u^*\mathbb{E}\left(\frac{\check{Q}\tilde{\xi}\tilde{\xi}^*\check{Q}}{1+\tilde{\xi}^*\check{Q}\tilde{\xi}}-\frac{\check{Q}\tilde{\xi}\tilde{\xi}^*\check{Q}}{1+\frac{1}{n}\mathrm{tr}\,R\check{Q}}\right)u \\ &\quad +u^*\mathbb{E}\left(\frac{\check{Q}\tilde{\xi}\tilde{\xi}^*\check{Q}}{1+\frac{1}{n}\mathrm{tr}\,R\check{Q}}-\frac{\check{Q}\xi\xi^*\check{Q}}{1+\frac{1}{n}\mathrm{tr}\,R\check{Q}}\right)u \\ &\quad +u^*\mathbb{E}\left(\frac{\check{Q}\xi\xi^*\check{Q}}{1+\frac{1}{n}\mathrm{tr}\,R\check{Q}}-\frac{\check{Q}\xi\xi^*\check{Q}}{1+\xi^*\check{Q}\xi}\right)u\ . \end{aligned}$$

The second term in the r.h.s. above is zero (simply compute the conditional expectation with respect to $\check{Q}$), the first and third term are of a similar nature; we therefore only estimate the third one denoted by Δ_3 below.

$$\begin{aligned} |\Delta_3| &= \left|u^*\mathbb{E}\left(\frac{\check{Q}\xi\xi^*\check{Q}}{1+\frac{1}{n}\mathrm{tr}\,R\check{Q}}-\frac{\check{Q}\xi\xi^*\check{Q}}{1+\xi^*\check{Q}\xi}\right)u\right| \\ &= \left|\mathbb{E}\frac{\xi^*\check{Q}uu^*\check{Q}\xi}{(1+\xi^*\check{Q}\xi)(1+\frac{1}{n}\mathrm{tr}\,R\check{Q})}\left(\xi^*\check{Q}\xi-\frac{1}{n}\mathrm{tr}\,R\check{Q}\right)\right| \\ &\le \frac{|z|^2}{|\mathrm{Im}(z)|^2}\left\{\mathbb{E}\left|\xi^*\check{Q}\xi-\frac{1}{n}\mathrm{tr}\,R\check{Q}\right|^2\mathbb{E}\left|\xi^*\check{Q}uu^*\check{Q}\xi\right|^2\right\}^{1/2}\ , \end{aligned}$$

where the last inequality follows from Cauchy-Schwarz inequality plus the fact that both $\left(-z(1+\xi^*\check{Q}\xi)\right)^{-1}$ and $\left(-z(1+n^{-1}\mathrm{tr}\,R\check{Q})\right)^{-1}$ are Stieltjes transforms and hence upper-bounded in module by $|\mathrm{Im}(z)|^{-1}$. A control for the first expectation in the above inequality directly follows from classical estimates (see for instance [4, Lemma B.26]):

$$\mathbb{E}\left|\xi^*\check{Q}\xi-\frac{1}{n}\mathrm{tr}\,R\check{Q}\right|^2\le\frac{K}{n^2}\mathbb{E}|X_{11}|^4\mathbb{E}\left(\mathrm{tr}\,R\check{Q}R\check{Q}^*\right)\le\frac{K}{n}\frac{\|R\|^2}{|\mathrm{Im}(z)|^2}c_n\mathbb{E}|X_{11}|^4\ , \tag{4.23}$$

where K is a constant whose value may change from line to line but which remains independent from N,n. The second expectation can be handled in the following way:

$$\begin{aligned} \mathbb{E}\left|\xi^*\check{Q}uu^*\check{Q}\xi\right|^2 &= \mathbb{E}\left|\xi^*\check{Q}uu^*\check{Q}\xi-\frac{1}{n}\mathrm{tr}\,R\check{Q}uu^*\check{Q}+\frac{1}{n}\mathrm{tr}\,R\check{Q}uu^*\check{Q}\right|^2 \\ &\le 2\mathbb{E}\left|\xi^*\check{Q}uu^*\check{Q}\xi-\frac{1}{n}\mathrm{tr}\,R\check{Q}uu^*\check{Q}\right|^2+2\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\,R\check{Q}uu^*\check{Q}\right|^2 \\ &\le \frac{K}{n^2}\mathbb{E}|X_{11}|^4\mathbb{E}\mathrm{tr}\,(R^{1/2}\check{Q}uu^*\check{Q}R^{1/2})(R^{1/2}\check{Q}uu^*\check{Q}R^{1/2})^* \\ &\quad +\frac{2}{n^2}\mathbb{E}\left|u^*\check{Q}R\check{Q}u\right|^2 \\ &\le \frac{K}{n^2}\frac{\|R\|^2\|u\|^4}{|\mathrm{Im}(z)|^4}\ . \end{aligned} \tag{4.24}$$

It now remains to gather (4.23) and (4.24) to get:

$$\left| u^* \mathbb{E}\left(Q^{(0)} - Q^{(1)}\right) u \right| \leq \frac{1}{n\sqrt{n}} \Phi(|z|) \Psi\left(\frac{1}{\boldsymbol{\delta}_z}\right) \|u\|^2.$$

Finally, the result follows by upper-bounding each term of the sum in (4.22).

□

Corollary 4.6. *Assume that (A-1) and (A-2) hold true, then the following convergence holds true:*

$$\frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2}\right)_{ii} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2}\right)_{ii} - \frac{1}{n} \sum_{i=1}^{N} \left(R^{1/2} T(z_1) R^{1/2}\right)_{ii} \left(R^{1/2} T(z_2) R^{1/2}\right)_{ii} \xrightarrow[n,N\to\infty]{\mathcal{P}} 0 \ .$$

Proof. We first transform the sum to be calculated:

$$\frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2}\right)_{ii} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2}\right)_{ii} \ . \tag{4.25}$$

Using Proposition 4.4 enables us to replace the conditional expectation $\mathbb{E}_i$ by the true expectation in every term $\left(R^{1/2} \mathbb{E}_j Q_j(z) R^{1/2}\right)_{ii}$. Now using the fact that

$$Q = Q_j - \frac{Q_j \xi_j \xi_j^* Q_j}{1 + \xi_j^* Q_j \xi_j}$$

and computations similar to those made in Proposition 4.5, one can replace $\mathbb{E}Q_j(z)$ by $\mathbb{E}Q(z)$. Finally, by Proposition 4.5, $\mathbb{E}Q(z)$ can be replaced by $\mathbb{E}\tilde{Q}(z)$. We are led to study the sum:

$$\frac{1}{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}\tilde{Q}(z_1) R^{1/2}\right)_{ii} \left(R^{1/2} \mathbb{E}\tilde{Q}(z_2) R^{1/2}\right)_{ii} \ .$$

Denote by $R_n = U_n \Delta U_n^*$ the spectral decomposition of covariance matrix R_n. Since matrix U_n is unitary, then $Y_n = U_n^* \tilde{X}_n$ has i.i.d. standard complex Gaussian entries and the resolvent writes:

$$\begin{aligned} \tilde{Q}(z) &= \left(R_n^{1/2} \tilde{X}_n \tilde{X}_n^* R_n^{1/2} - z I_N\right)^{-1} = U_n \left(\Delta^{1/2} Y_n Y_n^* \Delta^{1/2} - z I_N\right)^{-1} U_n^* \\ &\stackrel{\triangle}{=} U_n Q_\Delta(z) U_n^* \ . \end{aligned} \tag{4.26}$$

Denote by $T_\Delta(z)$ the matrix

$$T_\Delta(z) = \left(-z I_N + (1 - c_n)\Delta - z c_n t_n(z) \Delta\right)^{-1} \ ,$$

where $t_n(z)$ is defined in (2.3); notice that the definition of $t_n(z)$ only depends on the spectrum of R_n (or equivalently Δ); notice also that

$$T_n(z) = U_n \, T_\Delta(z) \, U_n^* \ . \tag{4.27}$$

It has been proved in [17, Theorem 1.1] that for every deterministic $N \times 1$ vector v_n:

$$\mathbb{E} \left| v_n^* \left(Q_\Delta(z) - T_\Delta(z)\right) v_n \right|^2 \leq \frac{1}{n} \Phi_2(|z|) \Psi_2\left(\frac{1}{\boldsymbol{\delta}_z}\right) \|v_n\|^4 \ .$$

Hence,

$$|v_n^* \mathbb{E} Q_\Delta(z) v_n - v_n^* T_\Delta(z) v_n| \leq \left(\mathbb{E} \left| v_n^* \left(Q_\Delta(z) - T_\Delta(z) \right) v_n \right|^2 \right)^{1/2}$$
$$\leq \frac{\|v_n\|^2}{\sqrt{n}} \sqrt{\Phi_2(|z|) \Psi_2\left(\boldsymbol{\delta}_z^{-1}\right)} \leq \frac{\|v_n\|^2}{\sqrt{n}} \left(\frac{1 + \Phi_2(|z|)}{2} \right) \left(\frac{1 + \Psi_2\left(\boldsymbol{\delta}_z^{-1}\right)}{2} \right)$$

In particular, let e_i be the i$^{\text{th}}$ coordinate vector, then

$$\left| \left(R^{1/2} \mathbb{E} Q(z) R^{1/2} \right)_{ii} - \left(R^{1/2} T(z) R^{1/2} \right)_{ii} \right|$$
$$= \left| \left(R^{1/2} U \mathbb{E} Q_\Delta(z) U^* R^{1/2} \right)_{ii} - \left(R^{1/2} U T_\Delta(z) U^* R^{1/2} \right)_{ii} \right|$$
$$\leq \frac{\|R^{1/2} U^* e_i\|^2}{\sqrt{n}} \left(\frac{1 + \Phi_2(|z|)}{2} \right) \left(\frac{1 + \Psi_2\left(\boldsymbol{\delta}_z^{-1}\right)}{2} \right) ,$$

which completes the proof. □

Combining the result in Corollary 4.6 together with (4.13) and (4.15), we have proved so far that:

$$\frac{\partial^2}{\partial z_1 \partial z_2} \left\{ \frac{z_1 z_2 \tilde{t}_n(z_1) \tilde{t}_n(z_2)}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2} \right)_{ii} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2} \right)_{ii} \right\}$$
$$= \frac{1}{n} \sum_{i=1}^{N} \frac{\partial^2}{\partial z_1 \partial z_2} \left\{ z_1 z_2 \tilde{t}_n(z_1) \tilde{t}_n(z_2) \left(R_n^{1/2} T_n(z_1) R_n^{1/2} \right)_{ii} \left(R_n^{1/2} T_n(z_2) R_n^{1/2} \right)_{ii} \right\} + o_P(1) .$$

Taking into account (3.2) and the matrix identity $U(I + VU)^{-1} V = 1 - (I + UV)^{-1}$, we obtain:

$$\frac{\partial^2}{\partial z_1 \partial z_2} \left\{ \frac{z_1 z_2 \tilde{t}_n(z_1) \tilde{t}_n(z_2)}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{N} \left(R^{1/2} \mathbb{E}_j Q_j(z_1) R^{1/2} \right)_{ii} \left(R^{1/2} \mathbb{E}_j Q_j(z_2) R^{1/2} \right)_{ii} \right\}$$
$$= \frac{1}{n} \sum_{i=1}^{N} \frac{\partial^2}{\partial z_1 \partial z_2} \left(I_N - (I_N + \tilde{t}_n(z_1) R_n)^{-1} \right)_{ii} \left(I_N - (I_N + \tilde{t}_n(z_2) R_n)^{-1} \right)_{ii} + o_P(1) ,$$
$$= \frac{z_1^2 z_2^2 \tilde{t}_n'(z_1) \tilde{t}_n'(z_2)}{n} \sum_{i=1}^{N} \left(R_n^{1/2} T_n^2(z_1) R_n^{1/2} \right)_{ii} \left(R_n^{1/2} T_n^2(z_2) R_n^{1/2} \right)_{ii} + o_P(1) ,$$
$$= \Theta_{2,n}(z_1, z_2) + o_P(1) , \tag{4.28}$$

where $\Theta_{2,n}$ is given by formula (2.12).

Now gathering (4.16), (4.19) and (4.28), we have established so far:

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) = \Theta_n(z_1, z_2) + o_P(1)$$

which is the first part of Proposition 4.2.

4.3.4. *Computations for the bias.* In this section, we are interested in the computation of $N(\mathbb{E}f_n(z) - t_n(z))$. As

$$\tilde{f}_n(z) = -\frac{(1-c_n)}{z} + c_n f_n(z) \quad \text{and} \quad \tilde{t}_n(z) = -\frac{(1-c_n)}{z} + c_n t_n(z) \ ,$$

we immediatly obtain $N(\mathbb{E}f_n(z) - t_n(z)) = n(\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z))$. Combining (2.7) and (3.2) yields:

$$-z - \frac{1}{\tilde{t}_n(z)} + \frac{1}{n}\mathrm{tr}\, R_n \left(I_N + \tilde{t}_n(z)R_n\right)^{-1} = 0 \ . \tag{4.29}$$

Following Bai and Silverstein [3, Section 4], we introduce the quantity $A_n(z)$ defined as:

$$\begin{aligned} A_n(z) &= z\mathbb{E}\tilde{f}_n(z) + 1 + \frac{1}{n}\mathrm{tr}\, \left(I_N + \mathbb{E}\tilde{f}_n(z)R_n\right)^{-1} - c_n \\ &= z\mathbb{E}\tilde{f}_n(z) + 1 + \frac{1}{n}\mathrm{tr}\, \left(I_N + \mathbb{E}\tilde{f}_n(z)R_n\right)^{-1} - \frac{1}{n}\mathrm{tr}\, I_N^{-1} \\ &= -\mathbb{E}\tilde{f}_n(z)\left(-z - \frac{1}{\mathbb{E}\tilde{f}_n(z)} + \frac{1}{n}\mathrm{tr}\, R_n(I_N + \mathbb{E}\tilde{f}_n(z)R_n)^{-1}\right) \ , \end{aligned}$$

hence

$$-\frac{A_n(z)}{\mathbb{E}\tilde{f}_n(z)} = -z - \frac{1}{\mathbb{E}\tilde{f}_n(z)} + \frac{1}{n}\mathrm{tr}\, R_n(I_N + \mathbb{E}\tilde{f}_n(z)R_n)^{-1} \ . \tag{4.30}$$

Substracting (4.29) to (4.30) finally yields:

$$\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z) = -A_n(z)\tilde{t}_n(z)\left[1 - \frac{\tilde{t}_n(z)\mathbb{E}\tilde{f}_n(z)}{n}\mathrm{tr}\, R_n^2 \left(I_N + \mathbb{E}\tilde{f}_n(z)R_n\right)^{-1}\left(I_N + \tilde{t}_n(z)R_n\right)^{-1}\right]^{-1} \ ,$$

which is the counterpart of [3, Eq. (4.12)]. The same arguments as in [3] now yields:

$$n\left(\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z)\right) = -nA_n(z)\tilde{t}_n(z)\left[1 - \frac{\tilde{t}_n^2(z)}{n}\mathrm{tr}\, R_n^2 \left(I_N + \tilde{t}_n(z)R_n\right)^{-2}\right]^{-1} + o(1) \ . \tag{4.31}$$

It remains to study the behaviour of $nA_n(z)$. Following [3, Eq. (4.10)], we obtain:

$$\begin{aligned} nA_n(z) = \\ &\frac{b_n^2}{n}\mathbb{E}\mathrm{tr}\, Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N\right)^{-1} R_n Q_1 R_n - b_n^2\, n\, \mathbb{E}\Bigg[\left(\xi_1^* Q_1 \xi_1 - \frac{1}{n}\mathrm{tr}\, Q_1 R_n\right) \\ &\qquad \times \left(\xi_1^* Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N\right)^{-1}\xi_1 - \frac{1}{n}\mathrm{tr}\, Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N\right)^{-1} R_n\right)\Bigg] + o(1) \ . \end{aligned}$$

Applying (2.6) to the right term to the r.h.s. of the previous equation (recall that $R^T = \bar{R}$), we obtain:

$$\begin{aligned} nA_n(z) = &-|\mathcal{V}|^2 \frac{b_n^2}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2} Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N\right)^{-1} R_n^{1/2}\bar{R}_n^{1/2} Q_1^T \bar{R}_n^{1/2} \\ &- \kappa\frac{b_n^2}{n}\sum_{i=1}^N \left(R_n^{1/2} Q_1 R_n^{1/2}\right)_{ii} \left(R_n^{1/2} Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N\right)^{-1} R_n^{1/2}\right)_{ii} + o(1) \ . \end{aligned} \tag{4.32}$$

The first term of the r.h.s. has been fully analyzed in [3] in the case where R_n and X_n are real matrices. We adapt these computations to the general case and outline in Appendix A

the proof of the identity:

$$- |\mathcal{V}|^2 \frac{b_n^2}{n} \mathbb{E}\mathrm{tr}\, R_n^{1/2} Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N \right)^{-1} R_n^{1/2} \bar{R}_n^{1/2} Q_1^T \bar{R}_n^{1/2} = |\mathcal{V}|^2 \frac{\frac{z^3 \tilde{t}_n^2}{n} \mathrm{tr}\, R_n^{1/2} T_n^2(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2}}{1 - \frac{|\mathcal{V}|^2 z^2 \tilde{t}_n^2}{n} \mathrm{tr}\, R_n^{1/2} T_n(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2}} + o(1) , \quad (4.33)$$

where $\bar{T}_n(z)$ is defined in (2.8). The term proportional to the cumulant in (4.32) can be analyzed as in Section 4.3.3, and one can prove that:

$$- \kappa \frac{b_n^2}{n} \sum_{i=1}^N \left(R_n^{1/2} Q_1 R_n^{1/2} \right)_{ii} \left(R_n^{1/2} Q_1 \left(\mathbb{E}\tilde{f}_n R_n + I_N \right)^{-1} R_n^{1/2} \right)_{ii} = -\kappa \frac{z^2 \tilde{t}_n^2}{n} \sum_{i=1}^N \left(R_n^{1/2} T_n R_n^{1/2} \right)_{ii} \left(R_n^{1/2} T_n \left(\tilde{t}_n R_n + I_N \right)^{-1} R_n^{1/2} \right)_{ii} + o(1) . \quad (4.34)$$

We now plug (4.33) and (4.34) into (4.31) to conclude.

$$\begin{aligned} n \left(\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z) \right) &= -|\mathcal{V}|^2 \frac{z^3 \tilde{t}_n^3}{n} \frac{\mathrm{tr}\, R_n^{1/2} T_n^2(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2}}{\left(1 - \frac{|\mathcal{V}|^2 z^2 \tilde{t}_n^2}{n} \mathrm{tr}\, R_n^{1/2} T_n(z) R_n^{1/2} \bar{R}_n^{1/2} \bar{T}_n(z) \bar{R}_n^{1/2} \right) \left(1 - \frac{\tilde{t}_n^2}{n} \mathrm{tr}\, R_n^2 T_n^2 \right)} \\ &\quad - \kappa \frac{z^3 \tilde{t}_n^3}{n} \sum_{i=1}^N \frac{\left(R_n^{1/2} T_n R_n^{1/2} \right)_{ii} \left(R_n^{1/2} T_n^2 R_n^{1/2} \right)_{ii}}{1 - \frac{z^2 \tilde{t}_n^2}{n} \mathrm{tr}\, R_n^2 T_n^2} + o(1) . \end{aligned}$$

Proof of Proposition 4.2 is completed.

4.4. **Proof of Proposition 4.3.**

4.4.1. *The Gaussian process N_n.* In order to get some insight on properties related to deterministic equivalents, it is sometimes useful to consider matrix models which actually converge toward these deterministic equivalents. For instance, instead of having $f_n(z) - t_n(z) \to 0$, one may construct a matrix model for which the normalized resolvent would satisfy

$$f_n^M(z) \xrightarrow[M \to \infty]{} t_n(z)$$

for some extra parameter M. We proceed along these lines hereafter.

Let N, n and R_n be fixed and consider the $NM \times NM$ matrix

$$R_n(M) = \begin{pmatrix} R_n & 0 & \cdots \\ & \ddots & \\ \cdots & 0 & R_n \end{pmatrix} . \quad (4.35)$$

Matrix $R_n(M)$ is a block matrix with $N \times N$ diagonal blocks equal to R_n, and zero blocks elsewhere; for all $M \geq 1$ the spectral norm of $R_n(M)$ is equal to the spectral norm of R_n (which is fixed). In particular the sequence $(R_n(M); M \geq 1)$ with N, n fixed satisfies assumption (A-2) with $(R_n(M); M \geq 1)$ instead of (R_n). Consider now the random matrix model:

$$\Sigma_n(M) = \frac{1}{\sqrt{Mn}} R_n(M)^{1/2} X_n(M) \quad (4.36)$$

where $X_n(M)$ is a $MN \times Mn$ matrix with i.i.d. random entries with the same distribution as the X_{ij}'s and satisfying (A-1). The interest of introducing the random matrix $\Sigma_n(M)$

lies in the fact that matrices $\Sigma_n(M)\Sigma_n(M)^*$ and $\Sigma_n\Sigma_n^*$ have loosely speaking the same deterministic equivalents. Denote by t_n, T_n and $\tilde{t}_n$ the deterministic equivalents of $\Sigma_n\Sigma_n^*$ as defined in (2.3), (2.4) and (3.2), and by $t_n(M)$, $T_n(M)$ and $\tilde{t}_n(M)$ their counterparts for the model $\Sigma_n(M)\Sigma_n(M)^*$. Taking advantage of the block nature of $R_n(M)$, a straightforward computation yields (N, n fixed):

$$\forall M \geq 1, \quad t_n(M) = t_n, \quad \tilde{t}_n(M) = \tilde{t}_n \quad \text{and} \quad T_n(M) = \begin{pmatrix} T_n & 0 & \cdots \\ & \ddots & \\ \cdots & 0 & T_n \end{pmatrix} .$$

Similarly, denote by $\mathcal{B}_{n,M}(,z)$ and $\Theta_{n,M}(z_1,z_2)$ the quantities given by formulas (2.19) and (2.9) when replacing N, t_n, T_n and $\tilde{t}_n$ by NM, $t_n(M)$, $T_n(M)$ and $\tilde{t}_n(M)$. Straightforward computation yields:

$$\forall M \geq 1 \ , \quad \mathcal{B}_{n,M}(z) = \mathcal{B}_n(z) \quad \text{and} \quad \Theta_{n,M}(z_1,z_2) = \Theta_n(z_1,z_2) \ .$$

Denote by $\widehat{M}_{n,M}(z)$ the truncated process[7] associated to

$$M_{n,M}(z) = \operatorname{tr}\left(\Sigma_n(M)\Sigma_n(M)^* - zI_{NM}\right)^{-1} - MNt_n(z) \ .$$

Applying Proposition 4.2 to the matrix model $\Sigma_n(M)\Sigma_n(M)^*$ yields: For $z \in \Gamma$:

$$\widehat{M}^1_{n,M}(z) = \sum_{j=1}^{nM} Z_j^M(z) + o_P(1) \ ,$$

where the Z_j^M's are martingale increments and

$$\sum_{j=1}^{nM} \mathbb{E}_{j-1} Z_j^M(z_1) Z_j^M(z_2) \quad \xrightarrow[N,n \text{ fixed }, \ M\to\infty]{\mathcal{P}} \quad \Theta_n(z_1,z_2) \ ,$$

$$\widehat{M}^2_{n,M}(z) \quad \xrightarrow[N,n \text{ fixed}, M\to\infty]{} \quad \mathcal{B}_n(z) \ .$$

Notice that there is a genuine limit in the previous convergence. Applying the central limit theorem for martingales [7, Theorem 35.12] plus the tightness argument for $(\widehat{M}_{n,M}(z), z \in \Gamma)$ provided by Proposition 4.2 immediatly yields the fact that $\widehat{M}_{n,M}$ converges in distribution to a Gaussian process $(N_n(z), z \in \Gamma)$ with mean $\mathcal{B}_n(z)$ and covariance function $\Theta_n(z_1,z_2)$.

4.4.2. *Tightness of the sequence of Gaussian processes* (N_n). In order to prove that the sequence of Gaussian processes (N_n) is tight, we shall prove, according to Prohorov's theorem, that it is relatively compact in distribution. Consider the set of matrices:

$$\{(R_n(M), M \geq 1) \, ; \ R_n \text{ is a } N \times n \text{ matrix}, \ N = N(n); n \geq 1\}$$

where $R_n(M)$ is defined in (4.35). Since $\|R_n(M)\| = \|R_n\|$ for all $M \geq 1$, we have

$$\sup_{M\geq 1, N, n\to\infty} \|R_n(M)\| = \sup_{N,n\to\infty} \|R_n\| < \infty$$

by Assumption (A-2). Hence, by Proposition 4.2, the family $\{\widehat{M}_{n,M}; M \geq 1\}_{N,n\to\infty}$ is tight, hence relatively compact in distribution. As the distribution $\mathcal{L}(N_n)$ of the Gaussian process

[7] In order to fully define the truncated process, one may specify the thereshold $\varepsilon_n(M)$ as $\varepsilon_n(1) = \varepsilon_n$, $\varepsilon_n(M) \to_M 0$ and $\varepsilon_n(M) \geq (NM)^{-\alpha}$ - the contour $\mathcal{C}_n(M)$ where both processes coincide can be defined accordingly.

N_n is the limit (in M) of the distribution $\mathcal{L}(\widehat{M}_{n,M})$ of $\widehat{M}_{n,M}$, $\mathcal{L}(N_n)$ belongs to the closure of $\{\mathcal{L}(\widehat{M}_{n,M})\}$, which is compact. Finally, $\{\mathcal{L}(N_n)\}$ is included in a compact set, hence is relatively compact. In particular, the family of Gaussian processes (N_n) is tight.

4.5. **Proof of Lemma 4.1.** The two propositions below are minor variations of known results. They will be helpful to conclude the proof of Lemma 4.1.

Lemma 4.7 (CLT for martingales I)**.** *Suppose that for each n $Y_{n1}, Y_{n2}, \cdots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{n,j}\}$ having second moments. Assume moreover that (Θ_n^2) is a sequence of nonnegative real numbers, uniformly bounded. If*

$$\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 \mid \mathcal{F}_{n,j-1}\right) - \Theta_n^2 \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ ,$$

and for each $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 1_{|Y_{nj}|>\varepsilon}\right) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ ,$$

then, for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}f\left(\sum_{j=1}^{r_n} Y_{nj}\right) - \mathbb{E}f(Z_n) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ , \tag{4.37}$$

where Z_n is a centered Gaussian random variable with variance Θ_n^2.

Hereafter is the multidimensional and complex extension of Lemma 4.7 we shall rely on in the sequel:

Lemma 4.8 (CLT for martingales II)**.** *Suppose that for each n $(Y_{nj}; 1 \le j \le r_n)$ is a $\mathbb{C}^d$-valued martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{n,j}; 1 \le j \le r_n\}$ having second moments. Write:*

$$Y_{nj}^T = (Y_{nj}^1, \cdots, Y_{nj}^d) \ .$$

Assume moreover that $(\Theta_n(k,\ell))_n$ and $(\tilde{\Theta}_n(k,\ell))_n$ are uniformly bounded sequences of complex numbers, for $1 \le k, \ell \le d$. If

$$\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^k \bar{Y}_{nj}^d \mid \mathcal{F}_{n,j-1}\right) - \Theta_n(k,\ell) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ , \tag{4.38}$$

$$\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^k Y_{nj}^\ell \mid \mathcal{F}_{n,j-1}\right) - \tilde{\Theta}_n(k,\ell) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ , \tag{4.39}$$

and for each $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathbb{E}\left(|Y_{nj}|^2 1_{|Y_{nj}|>\varepsilon}\right) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ , \tag{4.40}$$

then, for every bounded continuous function $f : \mathbb{C}^d \to \mathbb{R}$

$$\mathbb{E}f\left(\sum_{j=1}^{r_n} Y_{nj}\right) - \mathbb{E}f(Z_n) \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ , \tag{4.41}$$

where Z_n is a $\mathbb{C}^d$-valued centered Gaussian random vector with parameters

$$\mathbb{E} Z_n Z_n^* = (\Theta_n(k,\ell))_{k,\ell} \qquad \text{and} \qquad \mathbb{E} Z_n Z_n^T = (\tilde{\Theta}_n(k,\ell))_{k,\ell} \ .$$

Lemmas 4.7 and 4.8 are variations around the Central Limit Theorem for martingales (see Billingsley [7, Theorem 35.12]) which enables us to prove (in the real case):

$$\forall t \in \mathbb{R}\,, \quad \mathbb{E} e^{\mathbf{i}t \sum_{j=1}^{r_n} Y_{nj}} - e^{-\frac{t^2 \Theta_n^2}{2}} \to 0$$

and Lévy theorem for the weak convergence criterion via characteristic functions (see Kallenberg [20, Theorem 5.3 and Theorem 5.5]) which yields (4.41) from the above convergence. Details of the proof are omitted.

Lemma 4.9 (Tightness and weak convergence)**.** *Let K be a compact set in $\mathbb{C}$; let $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ be random elements in $C(K,\mathbb{C})$. Assume that for all $d \geq 1$, for all $z_1, \cdots, z_d \in K$, for all $f \in C(\mathbb{C}^d, \mathbb{C})$ we have:*

$$\mathbb{E} f(X_n(z_1), \cdots, X_n(z_d)) - \mathbb{E} f(Y_n(z_1), \cdots, Y_n(z_d)) \xrightarrow[n\to\infty]{} 0 \ .$$

Assume moreover that (X_n) and (Y_n) are tight, then for every continuous and bounded functional $F : C(K,\mathbb{C}) \to \mathbb{C}$, we have:

$$\mathbb{E} F(X_n) - \mathbb{E} F(Y_n) \xrightarrow[n\to\infty]{} 0 \ .$$

Lemma 4.9 can be proved as [20, Lemma 16.2]; the proof is therefore omitted.

We are now in position to conclude.

In order to apply Lemma 4.8, it remains to check that Θ_n as defined in (2.9) is uniformly bounded for $z_1, z_2 \in \Gamma$ fixed but this is an easy byproduct of Proposition 4.3.

Proposition 4.2 together with Lemma 4.8 (notice that condition (4.40) can be proved as in [3]) yield the fact that for every $z_1, \cdots, z_d \in \mathcal{C} \cup \bar{\mathcal{C}}$ and for every bounded continuous function $f : \left(\mathcal{C} \cup \bar{\mathcal{C}}\right)^d \to \mathbb{C}$:

$$\mathbb{E} f(\widehat{M}_n(z_1), \cdots, \widehat{M}_n(z_d)) - \mathbb{E} f(N_n(z_1), \cdots, N_n(z_d)) \xrightarrow[N,n\to\infty]{} 0 \ ,$$

where N_n is well-defined by Proposition 4.3. Now the tightness of $\widehat{M}_n$ and N_n together with Lemma 4.9 yield the last statement of Lemma 4.1.

4.6. **Proof of Theorem 3.1.** Theorem 3.1 is now almost completely proved. It remains to prove that $L_n(\mathbf{f})$ and $Z_n(\mathbf{f})$ being tight, the following equivalence holds true:

$$d_{\mathcal{L}P}\left(L_n(\mathbf{f}), Z_n(\mathbf{f})\right) \xrightarrow[N,n\to\infty]{} 0 \quad \Longleftrightarrow \quad \mathbb{E}\, h(L_n(\mathbf{f})) - \mathbb{E}\, h(Z_n(\mathbf{f})) \xrightarrow[N,n\to\infty]{} 0$$

but this can be proved easily by contradiction using the fact that $d_{\mathcal{L}P}$ metrizes the convergence of laws.

Appendix A. Proof of Proposition 4.2: remaining computations for the bias

In this section, we outline the proof of identity (4.33) which we recall below:

$$-|\mathcal{V}|^2\frac{b_n^2}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}Q_1\left(\mathbb{E}\tilde{f}_nR_n+I_N\right)^{-1}R_n^{1/2}\bar{R}_n^{1/2}Q_1^T\bar{R}_n^{1/2} = |\mathcal{V}|^2\frac{\frac{z^3\tilde{t}_n^2}{n}\mathrm{tr}\, R_n^{1/2}T_n^2(z)R_n^{1/2}\bar{R}_n^{1/2}\bar{T}_n(z)\bar{R}_n^{1/2}}{1-\frac{|\mathcal{V}|^2z^2\tilde{t}_n^2}{n}\mathrm{tr}\, R_n^{1/2}T_n(z)R_n^{1/2}\bar{R}_n^{1/2}\bar{T}_n(z)\bar{R}_n^{1/2}}+o(1)\ . \quad (A.1)$$

The proof closely follows computations in [3, Section 4] and is essentially a matter of bookkeeping; in particular, all the estimates established there remain valid in the context where R_n and X_n are not real. We shall focus here on the algebraic identities.

We first replace Q_1 by Q and approximate Q by (cf. [3, Eq. 4.13]):

$$Q(z)=-(zI_N-b_n(z)R_n)^{-1}+b_n(z)A(z)+B(z)+C(z) \quad (A.2)$$

where

$$A(z)=\sum_{j=1}^n(zI_N-b_n(z)R_n)^{-1}(\xi_j\xi_j^*-n^{-1}R_n)Q_j(z)\ .$$

The terms $B(z)$ and $C(z)$ will not contribute in the sequel. Denote by

$$\begin{aligned} M &= (\mathbb{E}\tilde{f}_nR_n+I_N)^{-1}R_n^{1/2}\bar{R}_n^{1/2},\\ \mathcal{T} &= \frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}Q_1\left(\mathbb{E}\tilde{f}_nR_n+I_N\right)^{-1}R_n^{1/2}\bar{R}_n^{1/2}Q_1^T\bar{R}_n^{1/2}\ . \end{aligned}$$

We have:

$$\begin{aligned} \mathcal{T} &= \frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}Q_1MQ_1^T\bar{R}_n^{1/2}\\ &= \frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}QMQ^T\bar{R}_n^{1/2}+o(1)\\ &= -\frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}MQ^TR_n^{1/2}+\frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}A(z)MQ^T\bar{R}^{1/2}+o(1)\\ &\triangleq \mathcal{T}_1+\mathcal{T}_2+o(1) \end{aligned} \quad (A.3)$$

In order to compute $\mathcal{T}_1$, we approximate Q^T in the same way as in (A.2); we take into account the fact that for some deterministic matrix Γ, $\mathbb{E}\,\mathrm{tr}\,(\Gamma A)=0$; we also use the approximation $b_n(z)=-z\tilde{t}_n(z)+o(1)$ and equation (3.2). The computation of $\mathcal{T}_1$ then easily follows:

$$\begin{aligned} \mathcal{T}_1 &= -\frac{1}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}MQ^T\bar{R}_n^{1/2}\\ &= \frac{1}{n}\mathrm{tr}\, R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}M(zI_N-b_n(z)\bar{R}_n)^{-1}\bar{R}_n^{1/2}+o(1)\\ &= -\frac{z}{n}\mathrm{tr}\, R_n^{1/2}T_n^2(z)R_n^{1/2}\bar{R}_n^{1/2}\bar{T}_n(z)\bar{R}_n^{1/2}+o(1)\ . \end{aligned}$$

We now focus on the term

$$\begin{aligned}\mathcal{T}_2 &= \frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}A(z)MQ^T\bar{R}^{1/2}\\ &= \frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}\sum_{j=1}^n(zI_N-b_n(z)R_n)^{-1}(\xi_j\xi_j^*-n^{-1}R_n)Q_j(z)MQ^T\bar{R}^{1/2}\\ &= \frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}\sum_{j=1}^n(zI_N-b_n(z)R_n)^{-1}\left\{\xi_j\xi_j^*Q_j(z)M(Q^T-Q_j^T)+D(z)+E(z)\right\}\bar{R}^{1/2}\end{aligned}$$

where

$$\begin{aligned}D(z) &= \xi_j\xi_j^*Q_jMQ_j^T-n^{-1}R_nMQ_jMQ_j^T\\ E(z) &= n^{-1}R_nM(Q_j^T-Q^T)\end{aligned}$$

will not contribute. Using the rank-one perturbation identity for $Q^T-Q_j^T$, we obtain:

$$\begin{aligned}\mathcal{T}_2 &= \frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}\sum_{j=1}^n(zI_N-b_n(z)R_n)^{-1}\xi_j\xi_j^*Q_j(z)M(Q^T-Q_j^T)\bar{R}^{1/2}+o(1)\\ &= -\frac{b_n(z)}{n}\mathbb{E}\mathrm{tr}\, R_n^{1/2}\sum_{i=1}^n(zI_N-b_n(z)R_n)^{-1}\xi_j\xi_j^*Q_j(z)M\frac{Q_j^T\bar{\xi}_j\bar{\xi}_j^*Q_j^T}{1+\bar{\xi}_j^*Q_j^T\bar{\xi}_j}\bar{R}^{1/2}+o(1)\\ &= -\frac{b_n(z)}{n}\sum_{j=1}^n\mathbb{E}\frac{1}{1+\bar{\xi}_j^*Q_j^T\bar{\xi}_j}\left(\bar{\xi}_j^*Q_j^T\bar{R}^{1/2}R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}\xi_j\right)\left(\xi_j^*Q_j(z)MQ_j^T\bar{\xi}_j\right)+o(1)\,.\end{aligned}$$

In order to pursue the computation of $\mathcal{T}_2$, we shall perform the following approximations: The quantity $(1+\bar{\xi}_j^*Q_j^T\bar{\xi}_j)^{-1}$ can be replaced by b_n and the two remaining quadratic forms in the expectation can be decorrelated. Now, using formulas (4.17), we obtain:

$$\begin{aligned}\mathcal{T}_2 &= -\frac{b_n^2(z)}{n}\sum_{j=1}^n\mathbb{E}\left(\bar{\xi}_j^*Q_j^T\bar{R}^{1/2}R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}\xi_j\right)\mathbb{E}\left(\xi_j^*Q_j(z)MQ_j^T\bar{\xi}_j\right)+o(1)\\ &= -\frac{|\mathcal{V}|^2b_n^2(z)}{n}\sum_{j=1}^n\mathbb{E}\mathrm{tr}\,\left(\bar{R}^{1/2}Q_j^T\bar{R}^{1/2}R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}R^{1/2}\right)\mathbb{E}\mathrm{tr}\,\left(R^{1/2}Q_j(z)MQ_j^T\bar{R}^{1/2}\right)+o(1)\end{aligned}$$

We can now replace Q_j by Q with no loss and use equation (A.2) to obtain:

$$\begin{aligned}\mathcal{T}_2 &= -\frac{|\mathcal{V}|^2b_n^2(z)}{n^2}\mathbb{E}\mathrm{tr}\,\left(\bar{R}^{1/2}Q^T(z)\bar{R}^{1/2}R_n^{1/2}(zI_N-b_n(z)R_n)^{-1}R^{1/2}\right)\\ &\qquad\times\mathbb{E}\mathrm{tr}\,\left(R^{1/2}\left(-(zI_N+b_n(z)R_n)^{-1}+b_n(z)A(z)\right)MQ^T\bar{R}^{1/2}\right)+o(1)\\ &= |\mathcal{V}|^2b_n^2(z)\frac{1}{n}\mathrm{tr}\, R^{1/2}T(z)R^{1/2}\bar{R}^{1/2}\bar{T}(z)\bar{R}^{1/2}\left(\mathcal{T}_1+\mathcal{T}_2\right)+o(1)\end{aligned}\qquad(A.4)$$

Denote by

$$\mathcal{T}_3=\frac{1}{n}\mathrm{tr}\, R^{1/2}T(z)R^{1/2}\bar{R}^{1/2}\bar{T}(z)\bar{R}^{1/2}\,.$$

We now extract $\mathcal{T}_2$ from (A.4) and plug it into (A.3). We finally obtain:

$$\mathcal{T} = \mathcal{T}_1+|\mathcal{V}|^2b_n^2(z)\frac{\mathcal{T}_1\mathcal{T}_3}{1-|\mathcal{V}|^2b_n^2(z)\mathcal{T}_3}+o(1) = \frac{\mathcal{T}_1}{1-|\mathcal{V}|^2b_n^2(z)\mathcal{T}_3}+o(1)\,.$$

Multiplying $\mathcal{T}$ by $-|\mathcal{V}|^2b_n^2(z)=-|\mathcal{V}|^2z^2\tilde{t}_n^2(z)$ finally yields (A.1).

References

[1] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

[2] L. V. Arharov. Limit theorems for the characteristic roots of a sample covariance matrix. *Dokl. Akad. Nauk SSSR*, 199:994–997, 1971.

[3] Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

[4] Z.D. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[5] Z.D. Bai, X. Wang, and W. Zhou. Functional CLT for sample covariance matrices. *Bernoulli*, 16(4):1086–1113, 2010.

[6] F. Benaych-George, A. Guionnet, and C. Male. Central limit theorems for linear statistics of heavy tailed random matrices. *preprint*, 2013.

[7] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[8] T. Cabanal-Duvillard. Fluctuations de la loi empirique de grandes matrices aléatoires. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(3):373–402, 2001.

[9] S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.

[10] R. Brent Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007.

[11] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[12] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[13] V. L. Girko. *Theory of stochastic canonical equations. Vol. II*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[14] A. Guionnet. Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):341–384, 2002.

[15] W. Hachem, M. Kharouf, J. Najim, and J. W. Silverstein. A CLT for information-theoretic statistics of non-centered Gram random matrices. *Random Matrices Theory Appl.*, 1(2):1150010, 50, 2012.

[16] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3):875–930, 2007.

[17] W. Hachem, P. Loubaton, J. Najim, and P. Vallet. On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1):36–63, 2013.

[18] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[19] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[20] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[21] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A central limit theorem for the sinr at the lmmse estimator output for large-dimensional signals. *Information Theory, IEEE Transactions on*, 55(11):5048–5063, nov. 2009.

[22] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur. Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.*, 37(10):5033–5060, 1996.

[23] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

[24] V. A. Marcenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[25] G. M. Pan and W. Zhou. Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.*, 18(3):1232–1270, 2008.

[26] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[27] Ya. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)*, 29(1):1–24, 1998.

[28] A. Soshnikov. The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.*, 28(3):1353–1370, 2000.

Jamal Najim,
Institut Gaspard Monge LabInfo, UMR 8049
Université Paris Est Marne-la-Vallée
5, Boulevard Descartes,
Champs sur Marne, 77454 Marne-la-Vallée Cedex 2, France
e-mail: najim@univ-mlv.fr