The Annals of Applied Probability
2016, Vol. 26, No. 3, 1837–1887
DOI: 10.1214/15-AAP1135

GAUSSIAN FLUCTUATIONS FOR LINEAR SPECTRAL STATISTICS OF LARGE RANDOM COVARIANCE MATRICES

By Jamal Najim[1] and Jianfeng Yao[2]

Université Paris Est and CNRS and Télécom Paristech

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

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

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 of class C^3, that is, has three continuous derivatives. The main improvements with respect to Bai and Silverstein's CLT [*Ann. Probab.* **32** (2004) 553–605] are twofold: First, we consider general entries with finite fourth moment, but whose fourth cumulant is nonnull, that is, whose fourth moment may differ from the moment of a (real or complex) Gaussian random variable. As a consequence, extra terms proportional to

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

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. Second, we relax the analyticity assumption over f by representing the linear statistics with the help of Helffer–Sjöstrand's formula.

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.

Received October 2014; revised July 2015.
[1]Supported by ANR Grant ANR-12-MONU-003 and Labex Bézout.
[2]Supported by program "Futur et ruptures" of Fondation Télécom.
AMS 2000 subject classifications. Primary 15A52; secondary 15A18, 60F15.
Key words and phrases. Large random matrix, linear statistics of the eigenvalues, central limit theorem.

CONTENTS

1. Introduction. Empirical random covariance matrices, whose probabilistic study may be traced back to Wishart [56] in the late twenties, play an important role in applied mathematics. After Marčenko and Pastur's seminal contribution [41] in 1967, the large dimensional setting (where the dimension of the observations is of the same order as the size of the sample) has drawn a growing interest, and important theoretical contributions [7, 34, 50] found many applications in multivariate statistics, electrical engineering, mathematical finance, etc.; cf. [4, 17, 39, 42]. The aim of this paper is to describe the fluctuations for linear spectral statistics of large empirical random covariance matrices. It will complete the picture already provided by Bai and Silverstein [7] and will hopefully provide a generic result of interest for practitioners.

The model. 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 uniformly bounded in N. The entries $(X_{ij}^n; i \le N, j \le n, n \ge 1)$ 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 $R_n^{1/2} X_{\cdot 1}^n$, with the population covariance matrix R_n. In the asymptotic regime where

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

(a condition that will be simply referred as $N, n \to \infty$ in the sequel), we study the fluctuations of linear spectral statistics of the form:

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

where $\mathrm{tr}(A)$ 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 [3], Girko (see [21, 22] and the references therein), Jonsson [35], Khorunzhiy et al. [38], Johansson [33], Sinai and Soshnikov [52, 53], Cabanal-Duvillard [14], Guionnet [24], Bai and Silverstein [7], Anderson and Zeitouni [2], Pan et al. [44, 45], Chatterjee [16], Lytova and Pastur [40], Bai et al. [6], Shcherbina [49], etc. There are also contributions for heavy-tailed entries (see, e.g., Benaych-Georges et al. [10]).

In their 2004 article [7], 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 neighborhood of the asymptotic spectrum of $\Sigma_n\Sigma_n^*$.

Such a result proved to be highly useful in probability theory, statistics and various other fields.

The purpose of this article is to establish a CLT for linear spectral statistics (1.3) for general entries X_{ij}^n with finite fourth moment and for nonanalytic functions f, sufficiently regular, hence to relax both assumptions (1) and (2) in [7].

It is well known since the paper by Khorunzhiy et al. [38] 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 [45], 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. [27, 37], CLTs have been established for specific linear statistics of interest in information theory, with general entries and (possibly noncentered) covariance random matrices with a variance profile. In Bao et al. [9], the CLT is established for the white model (where R_n is equal to the identity matrix) with general entries with finite fourth moment, featuring terms in the covariance proportional to the square of the second nonabsolute moment and to the fourth cumulant.

In Lytova and Pastur [40] and Shcherbina [49], both assumptions have been relaxed for the white model. In [40], it has been proved that mild integrability conditions over the Fourier transform of f was enough to establish the CLT. In Bai et al. [6], fluctuations for the white model are addressed as well, for nonanalytic functions f. Following Shcherbina's ideas, Guédon et al. [23] establish a CLT for linear statistics of large covariance matrices with vectors with log-concave distribution. Following Lytova and Pastur, Yao [57] relaxes the analyticity assumption in [7] by using interpolation techniques and Fourier transforms. We follow here a different approach, inspired from Bordenave [12].

Non-Gaussian entries. The presence of matrix R_n yields interesting phenomena at the CLT level when considering entries with non-Gaussian fourth moment: terms proportional to the fourth cumulant and to $|\mathbb{E}(X_{11}^n)^2|^2$ appear in the asymptotic variance (described in Section 2.3); however, their convergence is not granted under usual assumptions (roughly, under the convergence of R_n's spectrum), mainly because these extra terms also depend on the eigenvectors of R_n. As a consequence, such terms may not converge unless some very strong structural assumption over R_n (such as R_n diagonal) is made. This lack of convergence has consequences on the description of the fluctuations.

Denote by $L_n(f)$ the (approximately) centered version of the linear statistics (1.3), to be properly defined below. Instead of expressing the CLT in the usual way, that is ($\stackrel{\mathcal{D}}{\to}$ stands for the convergence in distribution)

$$L_n(f) \xrightarrow[N,n\to\infty]{\mathcal{D}} \mathcal{N}(\mathcal{B}_\infty^f, \Theta_\infty^f), \tag{1.4}$$

for some well-defined parameters $\mathcal{B}_\infty^f, \Theta_\infty^f$, we prove 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 that there exists a family of Gaussian random variables $\mathcal{N}(\mathcal{B}_n^f, \Theta_n^f)$, such that

$$d_{\mathcal{LP}}(L_n(f), \mathcal{N}(\mathcal{B}_n^f, \Theta_n^f)) \xrightarrow[N,n\to\infty]{} 0, \tag{1.5}$$

where $d_{\mathcal{LP}}$ denotes the Lévy–Prohorov distance (and in particular metrizes the convergence of laws). Details are provided in Section 2.5 and the fluctuation results are stated in Theorem 1 [for the resolvent $f(\lambda) = (\lambda - z)^{-1}$] and Theorem 2 (for f of class C^3, the space of functions with third continuous derivative).

From a technical point of view, the analysis of the extra term proportionnal to the fourth cumulant requires to cope with quadratic forms of the resolvent (counterpart of isotropic Marčenko–Pastur law). We provide the needed results in Section 5.

Expressing the CLT as in (1.5) makes it possible to avoid any cumbersome assumption related to the joint convergence of R_n's eigenvectors and eigenvalues; the technical price to pay however is the need to get various uniform (in N, n) controls over the sequence $\mathcal{N}(\mathcal{B}_n, \Theta_n)$. This is achieved by introducing a matrix meta-model in Section 2.6. The case where matrix R_n is diagonal is simpler and the fluctuations express in the usual way (1.4); it is handled in Section 3.4. Remarks on the white case ($R_n = I_N$) are also provided in Sections 3.5 and 4.2.

This framework may also prove to be useful for other interesting models such as large dimensional information-plus-noise type matrices [18, 28] and more generally mixed models combining large dimensional deterministic and random matrices.

Nonanalytic functions. In Section 3, we establish the CLT for the trace of the resolvent

$$\operatorname{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1}.$$

In order to transfer the CLT from the resolvent to the linear statistics of the eigenvalues $\operatorname{tr} f(\Sigma_n\Sigma_n^*)$, we will use (Dynkin–)Helffer–Sjöstrand's representation formula[3] for a function f of class C^{k+1} and with compact support [20, 32]. Denote by $\Phi_k(f) : \mathbb{C}^+ \to \mathbb{C}$ the function

$$\Phi_k(f)(x + \mathbf{i}y) = \sum_{\ell=0}^{k} \frac{(\mathbf{i}y)^\ell}{\ell!} f^{(\ell)}(x)\chi(y), \tag{1.6}$$

where $\chi : \mathbb{R} \to \mathbb{R}^+$ is smooth, compactly supported, with value 1 in a neighborhood of 0. Function $\Phi_k(f)$ coincides with f on the real line and is an appropriate extension of f to the complex plane. Let $\overline{\partial} = \partial_x + \mathbf{i}\partial_y$, then Helffer–Sjöstrand's formula writes

$$\operatorname{tr} f(\Sigma_n\Sigma_n^*) = \frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(f)(z) \operatorname{tr}(\Sigma_n\Sigma_n^* - zI_N)^{-1} \ell_2(dz), \tag{1.7}$$

where ℓ_2 stands for the Lebesgue measure over $\mathbb{C}^+$. An elementary proof of formula (1.7) can be found in [13], Chapter 5. Closest to our work are the papers by Pizzo, O'Rourke, Renfrew and Soshnikov [43, 47] where the fluctuations of the entries of regular functions of Wigner and large covariance matrices are studied; see also the paper by Bao et al. [8] where a CLT for partial linear eigenvalue statistics is established for Wigner matrices.

We believe that formula (1.7) provides a very streamlined way to handle nonanalytic functions and in fact enables us to state the fluctuations for the linear statistics for functions of class C^3, a lower regularity requirement than in [6, 40, 57]; in Shcherbina's article [49], the requirements over the functions are lower and expressed in terms of Sobolev norms $\|f\|_{3/2+\varepsilon} < \infty$, a condition that is fulfilled if f is C^2 (with bounded derivatives in L^2).

[3] In [31], Notes of Chapter 8, it is written "*This formula is due to Dynkin but was popularized by Helffer and Sjöstrand in the context of spectral theory, leading many authors to call it the Helffer–Sjöstrand formula.*"

Bias in the CLT and asymptotic expansion for the linear spectral statistics. Beside the fluctuations, a substantial part of this article is devoted to the study of the bias that we describe hereafter. In order to center the linear spectral statistics $\operatorname{tr} f(\Sigma_n \Sigma_n^*)$, we consider the (first-order) expansion of $\frac{1}{N}\mathbb{E}\operatorname{tr} f(\Sigma_n \Sigma_n^*)$

$$\frac{1}{N}\mathbb{E}\operatorname{tr} f(\Sigma_n \Sigma_n^*) = \mathcal{E}_{0,n}(f) + \mathcal{O}\left(\frac{1}{N}\right),$$

where $\mathcal{E}_{0,n}(f)$ is $\mathcal{O}(1)$ and does not depend on the distribution of the entries of X_n, and define $L_n(f)$ as

$$L_n(f) = \operatorname{tr} f(\Sigma_n \Sigma_n^*) - N\mathcal{E}_{0,n}(f).$$

A precise description of $L_n(f)$ is provided in Section 2.4. In order to fully characterize the fluctuations of $L_n(f)$, we must study the second-order expansion of $\frac{1}{N}\mathbb{E}\operatorname{tr} f(\Sigma_n \Sigma_n^*)$,

$$\frac{1}{N}\mathbb{E}\operatorname{tr} f(\Sigma_n \Sigma_n^*) = \mathcal{E}_{0,n}(f) + \frac{\mathcal{E}_{1,n}(f)}{N} + o\left(\frac{1}{N}\right),$$

which will naturally yield the bias of $L_n(f)$, as $\mathbb{E}L_n(f) = \mathcal{E}_{1,n}(f) + o(1)$. Asymptotic expansions for various matrix ensembles have already been studied; see, for instance, Pastur et al. [1], Bai and Silverstein [7], Haagerup and Thorbjørnsen [25, 26], Schultz [48], Capitaine and Donati-Martin [15], Vallet et al. [55], Hachem et al. [30], etc.

The asymptotic bias is expressed in Theorem 1 for the resolvent. In order to lift asymptotic expansions from the resolvent to smooth functions, we combine ideas from Haagerup and Thorbjørnsen [25] and Loubaton et al. [30, 55] together with some Gaussian interpolation and the use of Helffer–Sjöstrand's formula. For smooth functions, the statement is given in Theorem 3. Somehow surprisingly, the condition over function f is stronger for the asymptotic expansion to hold than for the CLT as function f needs to be of class C^{18} (cf. Remark 4.4).

2. General background—variance and bias formulas.

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

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

Assumption A-1. The random variables $(X_{ij}^n; 1 \le i \le N(n), 1 \le j \le n, n \ge 1)$ are independent and identically distributed. They satisfy

$$\mathbb{E}X_{ij}^n = 0, \qquad \mathbb{E}|X_{ij}^n|^2 = 1 \quad \text{and} \quad \mathbb{E}|X_{ij}^n|^4 < \infty.$$

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 $N, n \to \infty$:

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

In particular, we will have

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

2.2. *Resolvent, canonical equation and deterministic equivalents.* Denote by $Q_n(z)$ (resp., $\widetilde{Q}_n$) the resolvent of matrix $\Sigma_n \Sigma_n^*$ (resp., of $\Sigma_n^* \Sigma_n$):

$$Q_n(z) = (\Sigma_n \Sigma_n^* - zI_N)^{-1}, \qquad \widetilde{Q}_n(z) = (\Sigma_n^* \Sigma_n - zI_n)^{-1}, \tag{2.1}$$

and by $f_n(z)$ and $\tilde{f}_n(z)$ their normalized traces which are the Stieltjes transforms of the empirical distribution of $\Sigma_n \Sigma_n^*$'s and $\Sigma_n^* \Sigma_n$'s eigenvalues:

$$f_n(z) = \frac{1}{N} \operatorname{tr} Q_n(z), \qquad \tilde{f}_n(z) = \frac{1}{n} \operatorname{tr} \widetilde{Q}_n(z). \tag{2.2}$$

The following canonical equation[4] admits a unique solution t_n in the class of Stieltjes transforms of probability measures (see, e.g., [7]):

$$t_n(z) = \frac{1}{N} \operatorname{tr}(-zI_N + (1-c_n)R_n - zc_n t_n(z) R_n)^{-1}, \qquad z \in \mathbb{C} \setminus \mathbb{R}^+. \tag{2.3}$$

The function t_n being introduced, we can define the following $N \times N$ matrix:

$$T_n(z) = (-zI_N + (1-c_n)R_n - zc_n t_n(z) R_n)^{-1}. \tag{2.4}$$

Matrix $T_n(z)$ can be thought of as a *deterministic equivalent* of the resolvent $Q_n(z)$ in the sense that it approximates the resolvent in various senses. For instance,

$$\frac{1}{N} \operatorname{tr} T_n(z) - \frac{1}{N} \operatorname{tr} Q_n(z) \xrightarrow[N,n\to\infty]{} 0, \qquad z \in \mathbb{C} \setminus \mathbb{R}^+$$

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

$$u_n^* Q_n(z) v_n - u_n^* T_n(z) v_n \xrightarrow[N,n\to\infty]{} 0, \tag{2.5}$$

[4]We borrow the name "canonical equation" from V. L. Girko who established in [21, 22] canonical equations associated to various models of large random matrices.

where (u_n) and (v_n) are deterministic $N\times 1$ vectors with uniformly bounded Euclidean 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.

If $R_n = I_N$, then t_n is simply the Stieltjes transform of Marčenko–Pastur's distribution [41] with parameter c_n.

2.3. *Entries with nonnull fourth cumulant and the limiting covariance for the trace of the resolvent.* As in [7], we first study 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:

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

If the entries are real or complex standard Gaussian, then $\mathcal{V}=1$ or 0 and $\kappa = 0$. 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:

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

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 (RHS) of (2.6), a second one will be proportional to $|\mathcal{V}|^2$, and a third one to κ. In order to describe these terms, let

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

The quantity $\tilde{t}_n(z)$ is the deterministic equivalent associated to $n^{-1}\operatorname{tr}(\Sigma_n^*\Sigma_n - zI_n)^{-1}$. Denote by R_n^T the transpose matrix of R_n (notice that since R_n is Hermitian, $R_n^T = \overline{R}_n$ and we shall use this latter notation) and by T_n^T, the transpose matrix[5] of T_n:

$$T_n^T(z) = (-zI_N + (1-c_n)\overline{R}_n - zc_n t_n(z)\overline{R}_n)^{-1}; \tag{2.8}$$

notice that the definition of $t_n(z)$ in (2.3) does not change if R_n is replaced by $\overline{R}_n$ since the spectrum of both matrices R_n and $\overline{R}_n$ is the same. We can

[5] Beware that T_n^T is not the entry-wise conjugate of T_n, due to the presence of z.

now describe the limiting covariance of the trace of the resolvent

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

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

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

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

$$\begin{aligned}
\Theta_{2,n}(z_1,z_2) &\stackrel{\triangle}{=} \frac{z_1^2 z_2^2 \tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{n} \\
&\qquad \times \sum_{i=1}^N (R_n^{1/2} T_n^2(z_1) R_n^{1/2})_{ii} (R_n^{1/2} T_n^2(z_2) R_n^{1/2})_{ii},
\end{aligned} \tag{2.12}$$

with

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

For alternative formulas for $\Theta_{0,n}$ and $\Theta_{2,n}$, see Remarks 3.2 and 3.3.

At first sight, these formulas (established in Section 5) 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 [7]. 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

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

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 $\overline{R}_n^{1/2}$. 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, that is, $\mathcal{V}=1$, then it can be easily proved that $\Theta_{0,n}=\Theta_{1,n}$ hence the factor 2 in [7] 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 (e.g., if R_n is diagonal) and any assumption which enforces the convergence of such terms (as, e.g., in [45], Theorem 1.4) implicitly implies an asymptotic joint behavior 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.

2.4. *Representation of the linear statistics and limiting bias.* Recall that $t_n(z)$ is the Stieltjes transform of a probability measure $\mathcal{F}_n$:

$$t_n(z)=\int_{\mathcal{S}_n}\frac{\mathcal{F}_n(d\lambda)}{\lambda-z} \tag{2.16}$$

with support $\mathcal{S}_n$ included in a compact set. The purpose of this article is to describe the fluctuations of the linear statistics

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

as $N,n\to\infty$.

For a smooth enough function f of class C^{k+1} with bounded support, one can rely on Helffer–Sjöstrand's formula and write

$$\begin{aligned} L_n(f)&=\operatorname{tr} f(\Sigma_n\Sigma_n^*)-N\int f(\lambda)\mathcal{F}_n(d\lambda)\\ &=\frac{1}{\pi}\operatorname{Re}\int_{\mathbb{C}^+}\overline{\partial}\Phi_k(f)(z)\{\operatorname{tr} Q_n(z)-Nt_n(z)\}\ell_2(dz), \end{aligned} \tag{2.18}$$

where $\Phi_k(f)$ is defined in (1.6) and the last equality follows from the fact that

$$\int f(\lambda)\mathcal{F}_n(d\lambda)=\frac{1}{\pi}\operatorname{Re}\int_{\mathbb{C}^+}\overline{\partial}\Phi_k(f)(z)t_n(z)\ell_2(dz).$$

Based on (2.18), we shall first study the fluctuations of

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

for $z \in \mathbb{C}^+$. The first difference in the RHS 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

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

where

$$\begin{aligned}&\mathcal{B}_{1,n}(z) \\ &\quad \stackrel{\triangle}{=} -z^3 \tilde{t}_n^3 \left(\frac{1}{n} \operatorname{tr} R_n^{1/2} T_n^2(z) R_n^{1/2} \overline{R}_n^{1/2} T_n^T(z) \overline{R}_n^{1/2}\right) \\ &\qquad \Bigg/ \Bigg(\left(1 - z^2 \tilde{t}_n^2 \frac{1}{n} \operatorname{Tr} R_n^2 T_n^2\right) \\ &\qquad \times \left(1 - |\mathcal{V}|^2 z^2 \tilde{t}_n^2 \frac{1}{n} \operatorname{Tr} R_n^{1/2} T_n(z) R_n^{1/2} \overline{R}_n^{1/2} T_n^T(z) \overline{R}_n^{1/2}\right)\Bigg),\end{aligned} \tag{2.20}$$

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

The previous discussion on the terms $\Theta_{1,n}$ and $\Theta_{2,n}$ also applies to the terms $\mathcal{B}_{1,n}$ and $\mathcal{B}_{2,n}$ (whose expressions are established in Section 5) which are likely not to converge for similar reasons.

2.5. *Gaussian processes and the central limit theorem.* A priori, the mean $\mathcal{B}_n$ and covariance Θ_n of $(\operatorname{tr} Q_n - Nt_n)$ 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 $(G_n(z), z \in \mathcal{C})_{n\in\mathbb{N}}$ of tight Gaussian processes with mean and covariance

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

We then express the fluctuations of the centralized trace as

$$d_{\mathcal{LP}}((\operatorname{tr} Q_n(z) - Nt_n(z)), G_n(z)) \underset{N,n\to\infty}{\longrightarrow} 0,$$

with $d_{\mathcal{LP}}$ the Lévy–Prohorov distance between P and Q probability measures over borel sets of $\mathbb{R}, \mathbb{R}^d, \mathbb{C}$ or $\mathbb{C}^d$:

$$d_{\mathcal{LP}}(P, Q) = \inf\{\varepsilon > 0, P(A) \le Q(A^\varepsilon) + \varepsilon \text{ for all Borel sets } A\}, \tag{2.22}$$

where A^ε is an ε-blow up of A (cf. [19], Section 11.3, for more details). If X is a random variable and $\mathcal{L}(X)$ its distribution, denote (with a slight abuse of notation) by $d_{\mathcal{LP}}(X, Y) \stackrel{\triangle}{=} d_{\mathcal{LP}}(\mathcal{L}(X), \mathcal{L}(Y))$.

Similarly, we will express the fluctuations of $L_n(f)$ as

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

where $\mathcal{N}_n(f)$ is a well-identified Gaussian random variable.

2.6. *A meta-model argument.* As we need to cope with a sequence of Gaussian processes (G_n) instead of a single one, it will be necessary to establish various properties uniform in n, N such as:

1. the tightness of the sequence (G_n) (cf. Section 5.2);
2. a uniform bound over the variances of $(\operatorname{Tr} G_n(z))$ (cf. Section 6.2), needed to extend the CLT to nonanalytic functionals;
3. a uniform bound over the biases of $(\operatorname{Tr} G_n(z))$ (cf. Section 7.1.1), needed to compute the bias for nonanalytic functionals.

A direct approach based on the mere definition of process G_n's parameters seems difficult, mainly due to the definitions of Θ_n and $\mathcal{B}_n$ which rely on quantities (t_n and $\tilde{t}_n$) defined as solutions of fixed-point equations. Since the previous properties will be established for the processes $(\operatorname{Tr} Q_n - N t_n)$ anyway, the idea is to transfer them to G_n by means of the following matrix meta-model.

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

$$R_n(M) = \begin{pmatrix} R_n & 0 & \cdots \\ 0 & \ddots & 0 \\ \cdots & 0 & R_n \end{pmatrix}. \tag{2.23}$$

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

$$\Sigma_n(M) = \frac{1}{\sqrt{Mn}} R_n(M)^{1/2} X_n(M), \tag{2.24}$$

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 Assumption A-1. The interest of introducing 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 (3.1), 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 structure of $R_n(M)$, a straightforward computation yields (N, n fixed)

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

Similarly, denote by $\mathcal{B}_{n,M}$ and $\Theta_{n,M}$ 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

$$\forall M \geq 1, \qquad \mathcal{B}_{n,M} = \mathcal{B}_n \quad \text{and} \quad \Theta_{n,M} = \Theta_n.$$

An interesting feature of this meta-model lies in the fact that all the quantities associated to $\Sigma_n(M)\Sigma_n(M)^*$ converge as $M \to \infty$ to the deterministic equivalents t_n, $\tilde{t}_n$, etc. As a consequence, one can easily transfer all the estimates obtained for

$$(\operatorname{Tr}(\Sigma_n(M)\Sigma_n(M)^* - zI_{NM})^{-1} - NMt_n)$$

to the process (G_n).

3. Statement of the CLT for the trace of the resolvent.

3.1. *Further notation.* 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, e.g., [7]):

$$T_n(z) = -\frac{1}{z}(I_N + \tilde{t}_n(z)R_n)^{-1} \quad \text{and} \quad \tilde{t}_n(z) = -\frac{1}{z(1 + (1/n)\operatorname{tr} R_n T_n(z))}. \tag{3.1}$$

Recall the definition of $\mathcal{F}_n$ in (2.16) and let similarly $\widetilde{\mathcal{F}}_n$ be the probability distribution associated to $\tilde{t}_n$. The central object of study is the signed measure

$$N(F^{\Sigma_n \Sigma_n^*} - \mathcal{F}_n) = n(F^{\Sigma_n^* \Sigma_n} - \widetilde{\mathcal{F}}_n),$$

and its Stieltjes transform

$$M_n(z) = N(f_n(z) - t_n(z)) = n(\tilde{f}_n(z) - \tilde{t}_n(z)). \tag{3.2}$$

Denote by $o_P(1)$ any random variable which converges to zero in probability as $N, n \to \infty$.

3.2. *Truncation.* In this section, we closely follow Bai and Silverstein [7]. We recall the framework developed there and introduce some additional notation.

Consider a sequence of positive numbers (δ_n) which satisfies

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

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

$$\mathbb{P}(\Sigma_n \Sigma_n^* \neq \widehat{\Sigma}_n \widehat{\Sigma}_n^*) \underset{N,n\to\infty}{\longrightarrow} 0 \tag{3.3}$$

from which we deduce

$$\operatorname{tr}(\Sigma_n \Sigma_n^* - zI_N)^{-1} - \operatorname{tr}(\widehat{\Sigma}_n \widehat{\Sigma}_n^* - zI_N)^{-1} \xrightarrow[N,n\to\infty]{\mathcal{P}} 0, \tag{3.4}$$

where $\rightarrow^{\mathcal{P}}$ stands for the convergence in probability. Define $\widetilde{\Sigma}_n = n^{-1/2} R_n^{1/2} \widetilde{X}_n$ where $\widetilde{X}_n$ is a $N \times n$ matrix having (i,j)th entry $(\widehat{X}_{ij} - \mathbb{E}\widehat{X}_{ij})/\sigma_n$, where $\sigma_n^2 = \mathbb{E}|\widehat{X}_{ij} - \mathbb{E}\widehat{X}_{ij}|^2$. Using the fact that $\lambda (\in \mathbb{R}) \mapsto \frac{1}{\lambda - z}$ is Lipschitz with Lipschitz constant $|z|^{-2}$, we obtain

$$\mathbb{E}|\operatorname{tr}(\widehat{\Sigma}_n \widehat{\Sigma}_n^* - zI_N)^{-1} - \operatorname{tr}(\widetilde{\Sigma}_n \widetilde{\Sigma}_n^* - zI_N)^{-1}| \leq \frac{1}{|z|^2} \sum_{i=1}^{N} \mathbb{E}|\tilde{\lambda}_i - \hat{\lambda}_i| \xrightarrow[N,n\to\infty]{(a)} 0,$$

where $\tilde{\lambda}_i = \lambda_i(\widetilde{\Sigma}_n \widetilde{\Sigma}_n^*)$, $\hat{\lambda}_i = \lambda_i(\widehat{\Sigma}_n \widehat{\Sigma}_n^*)$ and (a) follows from similar arguments as in [5], Section 9.7.1. Hence,

$$\operatorname{tr}(\widehat{\Sigma}_n \widehat{\Sigma}_n^* - zI_N)^{-1} - \operatorname{tr}(\widetilde{\Sigma}_n \widetilde{\Sigma}_n^* - zI_N)^{-1} \xrightarrow[N,n\to\infty]{\mathcal{P}} 0. \tag{3.5}$$

Combining (3.4) and (3.5), we obtain

$$\operatorname{tr}(\Sigma_n \Sigma_n^* - zI_N)^{-1} - \operatorname{tr}(\widetilde{\Sigma}_n \widetilde{\Sigma}_n^* - zI_N)^{-1} \xrightarrow[N,n\to\infty]{\mathcal{P}} 0.$$

Moreover, the moments are asymptotically not affected by these different steps:

$$\max(|\mathbb{E}\widetilde{X}_{ij}^2 - \mathbb{E}X_{ij}^2|; (\mathbb{E}|\widetilde{X}_{ij}|^2 - 1); (\mathbb{E}|\widetilde{X}_{ij}|^4 - \mathbb{E}|X_{ij}|^4)) \underset{N,n\to\infty}{\longrightarrow} 0. \tag{3.6}$$

Note in particular that the fourth cumulant of $\widetilde{X}_{ij}$ 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 (we shall simply write X_{ij} and all related quantities with X_{ij}'s truncated, centralized, renormalized with no superscript any more).

3.3. *The central limit theorem for the resolvent.* We extend below Bai and Silverstein's master lemma [7], Lemma 1.1. Let A be such that

$$A > \boldsymbol{\lambda}_R^+ (1 + \sqrt{\boldsymbol{\ell}^+})^2.$$

Denote by D, D^+ and D_ε the domains

$$\begin{aligned} D &= [0, A] + \mathbf{i}[0, 1], \\ D^+ &= [0, A] + \mathbf{i}(0, 1], \\ D_\varepsilon &= [0, A] + \mathbf{i}[\varepsilon, 1] \qquad (\varepsilon > 0). \end{aligned} \tag{3.7}$$

THEOREM 1. *Assume that Assumption A-1 and Assumption A-2 hold true, then:*

1. *The process $\{M_n(\cdot)\}$ as defined in (3.2) forms a tight sequence on D_ε, more precisely,*

$$\sup_{z_1, z_2 \in D_\varepsilon, n \ge 1} \frac{\mathbb{E}|M_n(z_1) - M_n(z_2)|^2}{|z_1 - z_2|^2} < \infty.$$

2. *There exists a sequence $(G_n(z), z \in D^+)$ of two-dimensional Gaussian processes with mean*

$$\mathbb{E}G_n(z) = |\mathcal{V}|^2 \mathcal{B}_{1,n}(z) + \kappa \mathcal{B}_{2,n}(z), \tag{3.8}$$

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

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

and

$$\operatorname{cov}(G_n(z_1), \overline{G_n(z_2)}) = \operatorname{cov}(G_n(z_1), G_n(\overline{z_2})),$$

with $z_1, z_2 \in D^+ \cup \overline{D^+}$ with $\overline{D^+} = \{\bar{z}, z \in D^+\}$ and where $\Theta_{0,n}$, $\Theta_{1,n}$ and $\Theta_{2,n}$ are defined in (2.9), (2.10)–(2.12). Moreover, $(G_n(z), z \in D_\varepsilon)$ is tight.

3. *For any continuous functional F from $C(D_\varepsilon; \mathbb{C})$ to $\mathbb{C}$,*

$$\mathbb{E}F(M_n) - \mathbb{E}F(G_n) \underset{N,n\to\infty}{\longrightarrow} 0.$$

REMARK 3.1. 1. The tightness of the process $\{M_n\}$ immediately follows from Bai and Silverstein's lemma as this result has been proved in [7], Lemma 1.1, under Assumption A-1 with no extra conditions on the moments of the entries.

2. Differences between Theorem 1 and [7], 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 [7], Lemma 1.1.

3. Since the extra terms may not converge, we need to consider a sequence of Gaussian processes instead of a single Gaussian process as in [7], 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 $\{M_n\}$ to $\{G_n\}$ (see, e.g., Section 5.2.1).

5. Following Bai and Silverstein [7], it is relatively straightforward with the help of Cauchy's formula to describe the fluctuations of $L_n(f)$ for f analytic with Theorem 1 at hand. We skip this step since we will directly extend the CLT to nonanalytic functions f in Section 4.

REMARK 3.2. A closer look to Bai and Silverstein's proof [7], Section 2, page 578, yields the following alternative expression for the term $\Theta_{0,n}$:

$$\Theta_{0,n}(z_1,z_2)=\frac{\partial}{\partial z_2}\left\{\frac{\partial \mathcal{A}_{0,n}(z_1,z_2)}{\partial z_1}\frac{1}{1-\mathcal{A}_{0,n}(z_1,z_2)}\right\}, \tag{3.9}$$

with

$$\mathcal{A}_{0,n}(z_1,z_2)=\frac{z_1z_2}{n}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\operatorname{tr}\{R_nT_n(z_1)R_nT_n(z_2)\}. \tag{3.10}$$

Such an expression will be helpful in Section 6.2. As an interesting consequence: In the case where R_n and X_n have real entries [in particular $\mathcal{V}=\mathbb{E}(X_{ij})^2=1$], then $\mathcal{A}_{0,n}=\mathcal{A}_n$ and $\Theta_{0,n}=\Theta_{1,n}$.

REMARK 3.3. A closer look to the proof below [see, e.g., (5.21)] yields the following formula for $\Theta_{2,n}$ which will be of help in the sequel:

$$\Theta_{2,n}(z_1,z_2)=\frac{1}{n}\sum_{i=1}^{N}\frac{\partial}{\partial z_1}[z_1T_n(z_1)]_{ii}\frac{\partial}{\partial z_2}[z_2T_n(z_2)]_{ii}. \tag{3.11}$$

The proof of Theorem 1 is postponed to Section 5.

The end of the section is devoted to various specializations of Theorem 1 in the case where matrix R_n is diagonal. In this case, the results are simpler to express and comparisons can easily be made with related works.

3.4. *Covariance and bias in the special case of diagonal matrices* (R_n). This case partially falls into the framework developed in Pan and Zhou [45] (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)], $T_n=T_n^T$ and simplifications occur for the following terms:

$$\mathcal{A}_n(z_1,z_2)=\frac{z_1z_2}{n}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\operatorname{tr}R_nT_n(z_1)R_nT_n(z_2),$$

$$\Theta_{2,n}(z_1,z_2) = \frac{z_1^2 z_2^2 \tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{n} \operatorname{tr}(R_n^2 T_n^2(z_1) T_n^2(z_2)),$$

$$\mathcal{B}_{1,n}(z) = -z^3 \tilde{t}_n^3 \frac{(1/n)\operatorname{tr} R_n^2 T_n^3}{(1 - z^2\tilde{t}_n^2(1/n)\operatorname{Tr} R_n^2 T_n^2)(1 - |\mathcal{V}|^2 z^2 \tilde{t}_n^2 (1/n)\operatorname{Tr} R_n^2 T_n^2)},$$

$$\mathcal{B}_{2,n}(z) = -z^3 \tilde{t}_n^3 \frac{(1/n)\operatorname{tr} R_n^2 T_n^3}{1 - z^2 \tilde{t}_n^2 (1/n)\operatorname{tr} R_n^2 T_n^2}.$$

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):

$$\mathcal{A}_n(z_1,z_2) \underset{N,n\to\infty}{\longrightarrow} \mathcal{A}(z_1,z_2) = c\tilde{t}(z_1)\tilde{t}(z_2) \int \frac{\lambda^2 F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z_1))(1+\lambda\tilde{t}(z_2))},$$

$$\Theta_{1,n}(z_1,z_2) \underset{N,n\to\infty}{\longrightarrow} \Theta_1(z_1,z_2) = \frac{\partial}{\partial z_2}\left\{ \frac{\partial \mathcal{A}(z_1,z_2)}{\partial z_1} \frac{1}{1 - |\mathcal{V}|^2 \mathcal{A}(z_1,z_2)} \right\},$$

$$\Theta_{2,n}(z_1,z_2) \underset{N,n\to\infty}{\longrightarrow} \Theta_2(z_1,z_2) = c\tilde{t}'(z_1)\tilde{t}'(z_2) \int \frac{\lambda^2 F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z_1))^2(1+\lambda\tilde{t}(z_2))^2},$$

$$\mathcal{B}_{1,n}(z) \underset{N,n\to\infty}{\longrightarrow} \mathcal{B}_1(z) = -\frac{cz^3\tilde{t}^3(z)}{(1-\mathcal{A}(z,z))(1-|\mathcal{V}|^2\mathcal{A}(z,z))} \int \frac{\lambda^2 F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z))^3},$$

$$\mathcal{B}_{2,n}(z) \underset{N,n\to\infty}{\longrightarrow} \mathcal{B}_2(z) = -\frac{cz^3\tilde{t}^3(z)}{1-\mathcal{A}(z,z)} \int \frac{\lambda^2 F^{\mathbf{R}}(d\lambda)}{(1+\lambda\tilde{t}(z))^3},$$

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 3.1. *Assume that Assumptions A-1 and A-2 hold true. Assume moreover that R_n is diagonal and that the convergence assumption (2.14) holds true. Then $M_n(\cdot)$ converges weakly on D_ε [defined in (3.7)] to a two-dimensional Gaussian process $N(\cdot)$ satisfying*

$$\mathbb{E}N(z) = \mathcal{B}(z) \qquad \text{where } \mathcal{B} = |\mathcal{V}|^2\mathcal{B}_1 + \kappa\mathcal{B}_2, z \in D_\varepsilon$$

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

$$\operatorname{cov}(N(z_1), N_(z_2)) = \Theta(z_1,z_2)$$
$$\text{where } \Theta = \Theta_0 + |\mathcal{V}|^2\Theta_1 + \kappa\Theta_2, z_1, z_2 \in D_\varepsilon \cup \overline{D_\varepsilon}$$

and Θ_0 defined in (2.15) and Θ_1, Θ_2 defined above.

3.5. *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

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

and

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

This formula is in accordance with [9], formula (2.2) (use [9], equation (3.4), to equate both). If needed, one can then use the explicit expression of the Stieltjes transform of Marčenko–Pastur distribution (cf. also Proposition 4.2 below).

4. Statement of the CLT for nonanalytic functionals. In order to lift the CLT from the trace of the resolvent to a smooth function f, the key ingredient is Helffer–Sjöstrand's formula (1.7). Let

$$\begin{aligned}L_n(f) &\overset{(a)}{=} \operatorname{Tr} f(\Sigma_n\Sigma_n^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda)\\ &= (\operatorname{Tr} f(\Sigma_n\Sigma_n^*) - \mathbb{E}\operatorname{Tr} f(\Sigma_n\Sigma_n^*))\\ &\quad + \left(\mathbb{E}\operatorname{Tr} f(\Sigma_n\Sigma_n^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda)\right)\\ &\overset{\triangle}{=} L_n^1(f) + L_n^2(f),\end{aligned} \tag{4.1}$$

where $\mathcal{F}_n$ in (a) is defined in (2.16). We describe the fluctuations of $L_n^1(f)$ for nonanalytic functions f in Section 4.1 and study the bias $L_n^2(f)$ in Section 4.3.

4.1. *Fluctuations for the linear spectral statistics.* Denote by $C_c^\infty(\mathbb{R}^d)$ [resp., $C_c^m(\mathbb{R}^d)$] the set of infinitely differentiable (resp., C^m) functions from $\mathbb{R}^d$ to $\mathbb{R}$ with compact support; by $C_c^{mp}(\mathbb{R}^2)$ the set of functions from $\mathbb{R}^2$ to $\mathbb{R}$ m times differentiable with respect to the first coordinate and p times with respect to the second one. As usual, if the subscript c is removed in the sets above, then the corresponding functions may no longer have a compact support.

Theorem 2. *Assume that A-1 and A-2 hold true. Let $f_1,\dots,f_k$ be in $C_c^3(\mathbb{R})$. Consider the centered Gaussian random vector $Z_n^1(\mathbf{f}) \stackrel{\triangle}{=} (Z_n^1(f_1),\dots, Z_n^1(f_k))$ with covariance*

$$\begin{aligned} &\operatorname{cov}(Z_n^1(f), Z_n^1(g)) \\ &\quad = \frac{1}{2\pi^2}\operatorname{Re}\int_{(\mathbb{C}^+)^2} \overline{\partial}\Phi_2(f)(z_1)\overline{\overline{\partial}\Phi_2(g)(z_2)}\Theta_n(z_1,\bar z_2)\ell_2(dz_1)\ell_2(dz_2) \\ &\qquad + \frac{1}{2\pi^2}\operatorname{Re}\int_{(\mathbb{C}^+)^2} \overline{\partial}\Phi_2(f)(z_1)\overline{\partial}\Phi_2(g)(z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2), \end{aligned} \tag{4.2}$$

for $f,g\in\{f_1,\dots,f_k\}$, where $\Phi_2(f)$ and $\Phi_2(g)$ are defined as in (1.6), and where Θ_n is defined in (2.9); let

$$L_n^1(\mathbf{f}) = (L_n^1(f_1),\dots,L_n^1(f_k)) \qquad \textit{with } L_n^1(f) = \operatorname{tr} f(\Sigma_n\Sigma_n^*) - \mathbb{E}\operatorname{tr} f(\Sigma_n\Sigma_n^*).$$

Then the sequence of $\mathbb{R}^k$-valued random vectors $Z_n^1(\mathbf{f})$ is tight and the following convergence holds true:

$$d_{\mathcal{LP}}(L_n^1(\mathbf{f}), Z_n^1(\mathbf{f})) \underset{N,n\to\infty}{\longrightarrow} 0, \tag{4.3}$$

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

$$\mathbb{E}F(L_n^1(\mathbf{f})) - \mathbb{E}F(Z_n^1(\mathbf{f})) \underset{N,n\to\infty}{\longrightarrow} 0. \tag{4.4}$$

The proof of Theorem 2 is postponed to Section 6.

We provide hereafter some information on the covariance operator.

Let $N_1,N_2\in\mathbb{N}$ and $f\in C_c^{N_1+1,N_2+1}(\mathbb{R}^2)$; denote by $z_1 = x+\mathbf{i}u$, $z_2 = y+\mathbf{i}v$ and let $\Phi_{N_1,N_2}(f)$ be defined as

$$\Phi_{N_1,N_2}(f)(z_1,z_2) = \sum_{\substack{n_1=0:N_1\\ n_2=0:N_2}} \frac{\partial^{n_1+n_2}}{\partial x^{n_1}\partial y^{n_2}} f(x,y)\frac{(\mathbf{i}u)^{n_1}}{n_1!}\frac{(\mathbf{i}v)^{n_2}}{n_2!}\chi(u)\chi(v), \tag{4.5}$$

where $\chi:\mathbb{R}\to\mathbb{R}^+$ is smooth, compactly supported with value 1 in a neighborhood of the origin. Denote by $\overline{\partial}_1 = \partial_x + \mathbf{i}\partial_u$ and $\overline{\partial}_2 = \partial_y+\mathbf{i}\partial_v$.

Proposition 4.1. *For every $f\in C_c^{3,3}(\mathbb{R}^2)$, denote by*

$$\begin{aligned} \Upsilon(f) &= \frac{1}{2\pi^2}\operatorname{Re}\int_{(\mathbb{C}^+)^2} \overline{\partial}_2\overline{\partial}_1\Phi_{2,2}(f)(z_1,z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2) \\ &\quad + \frac{1}{2\pi^2}\operatorname{Re}\int_{(\mathbb{C}^+)^2} \overline{\partial}_2\overline{\partial}_1\Phi_{2,2}(f)(z_1,\overline{z_2})\Theta_n(z_1,\overline{z_2})\ell_2(dz_1)\ell_2(dz_2). \end{aligned}$$

Then $\Upsilon(f)$ is a distribution (in the sense of L. Schwartz) on $C_c^{3,3}(\mathbb{R}^2)$. Moreover, Υ admits the following boundary value representation:

$$\begin{aligned}\Upsilon(f) = -\frac{1}{4\pi^2}\lim_{\varepsilon\searrow 0}\int_{\mathbb{R}^2} f(x,y)\{&\Theta_n(x+\mathbf{i}\varepsilon, y+\mathbf{i}\varepsilon) + \Theta_n(x-\mathbf{i}\varepsilon, y-\mathbf{i}\varepsilon) \\ &- \Theta_n(x-\mathbf{i}\varepsilon, y+\mathbf{i}\varepsilon) - \Theta_n(x+\mathbf{i}\varepsilon, y-\mathbf{i}\varepsilon)\}\,dx\,dy.\end{aligned} \tag{4.6}$$

Notice that for every $f,g\in C_c^3(\mathbb{R})$ then $f\otimes g\in C_c^{3,3}(\mathbb{R}^2)$ [where $(f\otimes g)(x,y) = f(x)g(y)$] and

$$\Upsilon(f\otimes g) = \operatorname{cov}(Z_n^1(f), Z_n^1(g)).$$

The proof of Proposition 4.1 is postponed to Section 6.3.

Remark 4.1. By relying on Tillmann's results [54], one may prove that the support of Υ (as a distribution) is included in $\mathcal{S}_n\times\mathcal{S}_n$. We provide a more direct approach in a slightly simpler case in Section 4.2.

4.2. *More covariance formulas.* We provide here more explicit formulas for the variance than those given in Theorem 2 and Proposition 4.1; we also verify that these formulas are in agreement with other formulas available in the literature.

Recall that by [51], Theorem 1.1, the limit $\lim_{\varepsilon\searrow 0}\tilde{t}_n(x+\mathbf{i}\varepsilon)$ denoted by $\tilde{t}_n(x)$ exists for all $x\in\mathbb{R}$, $x\neq 0$; the same holds true for t_n.

Proposition 4.2. *Assume that Assumptions A-1 and A-2 hold true and let $f,g\in C_c^3(\mathbb{R})$; assume moreover for simplicity that $\mathcal{V}=\mathbb{E}X_{ij}^2$ is either equal to 0 or 1 and that R_n has real entries. Then the covariance of $(Z_n(f), Z_n(g))$ in Theorem 2 writes*

$$\begin{aligned}\operatorname{cov}(Z_n^1(f), Z_n^1(g)) &= \frac{1+|\mathcal{V}|^2}{2\pi^2}\int_{\mathcal{S}_n^2} f'(x)g'(y)\ln\left|\frac{\tilde{t}_n(x)-\overline{\tilde{t}_n(y)}}{\tilde{t}_n(x)-\tilde{t}_n(y)}\right|dx\,dy \\ &\quad+\frac{\kappa}{\pi^2 n}\sum_{i=1}^N\left(\int_{\mathcal{S}_n} f'(x)\operatorname{Im}(xT_n(x))_{ii}\,dx\right)\left(\int_{\mathcal{S}_n} g'(y)\operatorname{Im}(yT_n(y))_{ii}\,dy\right).\end{aligned} \tag{4.7}$$

The proof for Proposition 4.2 is postponed to Section 6.4.

Remark 4.2. Notice that the first term in the RHS matches with the expression provided in [7], equation (1.17) (see also [5], equation (9.8.8)).

Remark 4.3. Concerning the cumulant term, we shall compare it with the explicit formula provided in [40] (see also [46]) in the case where $R_n = I_N$.

Recall that in the context of Marčenko–Pastur's theorem where $R_n = I_N$, we have $\mathcal{S}_n = [\lambda^-, \lambda^+]$ where $\lambda^- = (1-\sqrt{c_n})^2$, $\lambda^+ = (1+\sqrt{c}_n)^2$ and $(T_n(x))_{ii} = t_n(x)$. We will prove hereafter that

$$\begin{aligned}
&\frac{\kappa c_n}{\pi^2}\left(\int_{\lambda^-}^{\lambda^+} f'(x)\,\mathrm{Im}\{x t_n(x)\}\,dx\right)\left(\int_{\lambda^-}^{\lambda^+} g'(y)\,\mathrm{Im}\{y t_n(y)\}\,dy\right)\\
&\qquad = \frac{\kappa}{4c_n\pi^2}\left(\int_{\lambda^-}^{\lambda^+} f(x)\frac{x-(1+c_n)}{\sqrt{(\lambda^+-x)(x-\lambda^-)}}\,dx\right)\\
&\qquad\quad \times\left(\int_{\lambda^-}^{\lambda^+} g(y)\frac{y-(1+c_n)}{\sqrt{(\lambda^+-y)(y-\lambda^-)}}\,dy\right).
\end{aligned}\tag{4.8}$$

Notice that the LHS of the equation above is the cumulant term as provided in (4.7) if $R_n = I_N$ while the RHS is the cumulant term as provided[6] in [40].

In the case where $R_n = I_N$, the Stieltjes transform of Marčenko–Pastur's distribution has an explicit form given by (see, e.g., [46], Chapter 7)

$$t_n(z) = \frac{1}{2c_n z}\{\sqrt{(z-(1+c_n))^2-4c_n} - (z-(1-c_n))\},$$

where the branch of the square root is fixed by its asymptotics: $z-(1+c)+o(1)$ as $z\to\infty$. In particular, if $x\in[\lambda^-,\lambda^+]$ then

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

Hence,

$$\mathrm{Im}\{x t_n(x)\} = \frac{\sqrt{(\lambda^+-x)(x-\lambda^-)}}{2c_n}.$$

It remains to perform an integration by parts to get

$$\begin{aligned}
\int_{\lambda^-}^{\lambda^+} f'(x)\,\mathrm{Im}\{x t_n(x)\}\,dx &= -\int_{\lambda^-}^{\lambda^+} f'(x)\frac{\sqrt{(\lambda^+-x)(x-\lambda^-)}}{2c_n}\,dx\\
&= \frac{1}{2c_n}\int_{\lambda^-}^{\lambda^+} f(x)\frac{(1+c_n)-x}{\sqrt{(\lambda^+-x)(x-\lambda^-)}}\,dx
\end{aligned}$$

which yields (4.8).

As a corollary of Proposition 4.2, we obtain the following extension of Theorem 2.

[6] Denote by the superscript $^{\mathrm{LP}}$ the quantities in [40] and use the correspondence $c^{\mathrm{LP}} \leftrightarrow 1/c$, $a^{\mathrm{LP}} \leftrightarrow c$ and $\kappa_4^{\mathrm{LP}} \leftrightarrow (a^{\mathrm{LP}})^4\kappa = c^2\kappa$ to check that the RHS of (4.8) equates the formula provided in [40].

Recall that $\mathcal{S}_n$ is the support of the probability measure $\mathcal{F}_n$. Due to Assumption A-2, it is clear that

$$\mathcal{S}_n \subset \mathcal{S}_\infty \stackrel{\triangle}{=} [0, \boldsymbol{\lambda}_R^+(1+\sqrt{\boldsymbol{\ell}^+})^2], \tag{4.9}$$

uniformly in n. Denote by $\mathbf{h} \in C_c^\infty(\mathbb{R})$ a function whose value is 1 on a η-neighborhood $\mathcal{S}_\infty^\eta$ of $\mathcal{S}_\infty$.

COROLLARY 4.3. *Assume that Assumptions A-1 and A-2 hold true and let $f_\ell \in C^3(\mathbb{R})$ with $1 \le \ell \le k$; assume moreover that $\mathcal{V} = \mathbb{E}X_{ij}^2$ is either equal to 0 or 1 and that R_n has real entries. Let $\mathbf{h} \in C_c^\infty(\mathbb{R})$ be as above. Then (4.3)–(4.4) remain true with $L_n^1(\mathbf{f})$ replaced by*

$$L_n^{1,\mathbf{h}}(\mathbf{f}) = (\operatorname{tr} f_\ell(\Sigma_n\Sigma_n^*) - \mathbb{E}\operatorname{tr}(f_\ell\mathbf{h})(\Sigma_n\Sigma_n^*); 1 \le \ell \le k)$$

and with the Gaussian random vector $Z_n^1(\mathbf{fh})$ as in Theorem 2.

The proof of Corollary 4.3 is postponed to Section 6.5.

4.3. *First-order expansions for the bias in the case of nonanalytic functionals.*

THEOREM 3. *Assume Assumptions A-1 and A-2 hold true and let $f \in C_c^{18}(\mathbb{R})$. Denote by*

$$Z_n^2(f) = \frac{1}{\pi}\operatorname{Re}\int_{\mathbb{C}^+} \overline{\partial}\Phi_{17}(f)(z)\mathcal{B}_n(z)\ell_2(dz), \tag{4.10}$$

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

$$\mathbb{E}\operatorname{Tr}(f)(\Sigma_n\Sigma_n^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda) - Z_n^2(f) \underset{N,n\to\infty}{\longrightarrow} 0.$$

The proof of Theorem 3 is postponed to Section 7.

REMARK 4.4 (Why eighteen?). A quick sketch of the proof of Theorem 3 provides some hints. Let f have a bounded support. By Gaussian interpolation (whose cost is $f \in C^8$), we only need to prove

$$\mathbb{E}\operatorname{Tr} f(\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda) \to 0,$$

where $\Sigma_n^{\mathbb{C}}$ is the counterpart of Σ_n with $\mathcal{N}_{\mathbb{C}}(0,1)$ i.i.d. entries. The proof of the latter is based on Helffer–Sjöstrand's formula

$$\mathbb{E}\operatorname{Tr} f(\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda) = \frac{1}{\pi}\operatorname{Re}\int_{\mathbb{C}^+} \overline{\partial}\Phi_k(f)\{\operatorname{Tr}\mathbb{E}Q_n^{\mathbb{C}} - Nt_n\}\,d\ell_2,$$

where $Q_n^{\mathbb{C}} = (\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^* - zI_N)^{-1}$, and on the following estimate, stated in Proposition 7.2:

$$|\mathbb{E}\operatorname{Tr}(\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^* - zI_N)^{-1} - Nt_n(z)| \le \frac{1}{n}P_{12}(|z|)P_{17}(|\operatorname{Im}(z)|^{-1}), \tag{4.11}$$

where P_k denotes a polynomial with degree k and positive coefficients. In view of Proposition 6.2, f needs to be of class C^{18}. If one can improve estimate (4.11) and decrease the powers of $|\operatorname{Im}(z)|^{-1}$, then one will automatically lower the regularity assumption over f. Notice that in the case of the Gaussian unitary ensemble, counterpart of (4.11) features $|\operatorname{Im}(z)|^{-7}$ on its RHS (cf. [25], Lemma 6.1), hence the needed regularity is $f \in C^8$ in this case.

PROPOSITION 4.4. *Let $Z_n^2(f)$ be defined as in (4.10), then Z_n^2 is a distribution (in the sense of L. Schwartz) on $C_c^{18}(\mathbb{R})$ and*

$$Z_n^2(f) = \frac{-\mathbf{i}}{2\pi}\lim_{\varepsilon\searrow 0}\int_{\mathbb{R}} f(x)\{\mathcal{B}_n(x+\mathbf{i}\varepsilon) - \mathcal{B}_n(x-\mathbf{i}\varepsilon)\}\,dx. \tag{4.12}$$

Moreover, the singular points of $\mathcal{B}_n(z)$ are included in $\mathcal{S}_n$ and so is the support of Z_n^2 (as a distribution). In particular, one can extend Z_n^2 to $C^{18}(\mathbb{R})$ by

$$\check{Z}_n^2(f) = Z_n^2(f\mathbf{h}), \qquad f \in C^{18}(\mathbb{R}),$$

where $\check{Z}_n^2$ is the extension to $C^{18}(\mathbb{R})$ and $\mathbf{h} \in C_c^\infty(\mathbb{R})$ has value 1 on $\mathcal{S}_n$.

The proof of Proposition 4.4 is postponed to Section 7.2.

COROLLARY 4.5. *Assume Assumptions A-1 and A-2 hold true. Let $f \in C^{18}(\mathbb{R})$ and $\mathbf{h} \in C_c^\infty(\mathbb{R})$ be a function whose value is 1 on a neighborhood of $\mathcal{S}_\infty$, then the following convergence holds true:*

$$\mathbb{E}\operatorname{Tr}(f\mathbf{h})(\Sigma_n\Sigma_n^*) - N\int f(\lambda)\mathcal{F}_n(d\lambda) - \check{Z}_n^2(f) \xrightarrow[N,n\to\infty]{} 0.$$

The proof is straighforward and is therefore omitted.

5. Proof of Theorem 1 (CLT for the trace of the resolvent). Recall that $M_n(z) = \operatorname{tr} Q_n(z) - Nt_n(z)$. It will be convenient to decompose $M_n(z)$ as

$$M_n(z) = M_n^1(z) + M_n^2(z) \qquad \text{where } \begin{cases} M_n^1(z) = \operatorname{tr} Q_n(z) - \operatorname{tr}\mathbb{E}Q_n(z), \\ M_n^2(z) = N(\mathbb{E}f_n(z) - t_n(z)). \end{cases} \tag{5.1}$$

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 $\mathcal{G}_j$, the σ-field generated by $\xi_1, \dots, \xi_j$:

$$\mathbb{E}_j = \mathbb{E}(\cdot|\mathcal{G}_j). \tag{5.2}$$

By convention, $\mathbb{E}_0 = \mathbb{E}$. We split Theorem 1 into intermediate results. Recall the definitions of D_ε, D^+ and D in (3.7). Let

$$\Gamma = D^+ \cup \overline{D^+} \qquad \text{where } \overline{D^+} = \{\bar{z}, z \in D^+\}.$$

PROPOSITION 5.1. *Assume that Assumptions A-1 and A-2 hold true; let $z_1, z_2 \in \Gamma$, then*

$$M_n^1(z) = \sum_{j=1}^n Z_j^n(z) + o_P(1),$$

where the Z_j^n's are martingale increments with respect to the σ-field $\mathcal{G}_i$ and

$$\sum_{j=1}^n \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) - \Theta_n(z_1, z_2) \xrightarrow[N,n\to\infty]{\mathcal{P}} 0, \tag{5.3}$$

$$\sum_{j=1}^n \mathbb{E}_{j-1} Z_j^n(z_1) \overline{Z_j^n(z_2)} - \Theta_n(z_1, \overline{z_2}) \xrightarrow[N,n\to\infty]{\mathcal{P}} 0, \tag{5.4}$$

where Θ_n is defined in (2.9). Moreover,

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

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

PROPOSITION 5.2. *There exists a sequence $(G_n(z), z \in \Gamma)$ of two-dimensional Gaussian processes with mean $\mathbb{E}G_n(z) = \mathcal{B}_n(z)$ and covariance*

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

Moreover, $(G_n(z), z \in D_\varepsilon)$ is tight.

5.1. *Proof for Proposition 5.1.* The fact that (M_n) is a tight sequence has already been established in [7] (regardless of the assumption $\kappa = 0$ and $|\mathcal{V}| = 0/1$). In order to proceed, we shall rely heavily on the proof of [7], Lemma 1.1, which is the crux of Bai and Silverstein's paper. In Section 5.1.1, we recall the main steps of Bai and Silverstein's computations of the variance/covariance. In Sections 5.1.2 and 5.1.3, we compute the extra terms in the limiting variance. In Section 5.1.4, we compute the limiting bias. In Section 5.3, we finally conclude the proof of Theorem 1 and address various subtleties which appear due to the existence of a sequence of Gaussian limiting processes.

In the sequel, we shall drop subscript n and write Q and R instead of Q_n and R_n. Denote by $Q_j(z)$ the resolvent of matrix $\Sigma\Sigma^* - \xi_j\xi_j^*$, that is,

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

The following quantities will be needed:

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

5.1.1. *Preliminary variance computations.* We briefly review in this section the main steps related to the computation of the variance/covariance as presented in [7]. These standard steps will finally lead to equation (5.8) which will be the starting point of the computations associated to the $|\mathcal{V}|^2$- and κ-terms of the variance.

Let $z \in \Gamma$:

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

where $\mathbb{E}_j$ is introduced in (5.2). Denote by

$$Z_j^n(z) = -\mathbb{E}_j\left(\bar{\beta}_j(z)\delta_j(z) - \bar{\beta}_j^2(z)\varepsilon_j(z)\frac{1}{n}\operatorname{tr} RQ_j^2(z)\right) = -\mathbb{E}_j\frac{d}{dz}(\bar{\beta}_j(z)\varepsilon_j(z)).$$

Hence,

$$M_n^1(z) = N(f_n(z) - \mathbb{E}f_n(z)) = \sum_{j=1}^{n} Z_j^n(z) + o_P(1).$$

The RHS 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 [11], Theorem 35.12, and the variations below in Lemmas 5.5 and 5.6).

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 $\mathbb{E}_{j-1} Z_j^n(z_1)\overline{Z_j^n(z_2)} = \mathbb{E}_{j-1} Z_j^n(z_1) \times Z_j^n(\overline{z_2})$. Since the set Γ is stable by complex conjugation, it is sufficient to study the limiting behavior of

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

in order to prove (5.3) and (5.4). Now,

$$\begin{aligned} &\sum_{j=1}^{n} \mathbb{E}_{j-1} Z_j^n(z_1) Z_j^n(z_2) \\ &\quad = \frac{\partial^2}{\partial z_1 \partial z_2} \left\{ \sum_{j=1}^{n} \mathbb{E}_{j-1} [\mathbb{E}_j(\bar{\beta}_j(z_1)\varepsilon_j(z_1)) \mathbb{E}_j(\bar{\beta}_j(z_2)\varepsilon_j(z_2))] \right\}. \end{aligned} \tag{5.5}$$

Following the same arguments as in [7], page 571, one can prove that it is sufficient to study the convergence in probability of

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

Moreover,

$$\begin{aligned} &\sum_{j=1}^{n} \mathbb{E}_{j-1} [\mathbb{E}_j(\bar{\beta}_j(z_1)\varepsilon_j(z_1)) \mathbb{E}_j(\bar{\beta}_j(z_2)\varepsilon_j(z_2))] \\ &\quad = \sum_{j=1}^{n} b_n(z_1) b_n(z_2) \mathbb{E}_{j-1} [\mathbb{E}_j \varepsilon_j(z_1) \mathbb{E}_j \varepsilon_j(z_2)] + o_P(1) \\ &\quad = \sum_{j=1}^{n} z_1 \tilde{t}_n(z_1) z_2 \tilde{t}_n(z_2) \mathbb{E}_{j-1} [\mathbb{E}_j \varepsilon_j(z_1) \mathbb{E}_j \varepsilon_j(z_2)] + o_P(1). \end{aligned} \tag{5.6}$$

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

$$\sum_{j=1}^{n} \mathbb{E}_{j-1} (\mathbb{E}_j \varepsilon_j(z_1) \mathbb{E}_j \varepsilon_j(z_2)), \qquad z_1, z_2 \in \Gamma. \tag{5.7}$$

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

$$\begin{aligned}
&\sum_{j=1}^n \mathbb{E}_{j-1}(\mathbb{E}_j\varepsilon_j(z_1)\mathbb{E}_j\varepsilon_j(z_2))\\
&\qquad = \frac{1}{n^2}\sum_{j=1}^n \mathrm{tr}(R^{1/2}\mathbb{E}_jQ_j(z_1)R\mathbb{E}_jQ_j(z_2)R^{1/2})\\
&\qquad\quad + \frac{|\mathcal{V}|^2}{n^2}\sum_{j=1}^n \mathrm{tr}(R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})^T)\\
&\qquad\quad + \frac{\kappa}{n^2}\sum_{j=1}^n\sum_{i=1}^N (R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2})_{ii}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})_{ii}.
\end{aligned}\tag{5.8}$$

The limiting behavior of the first term of the RHS has been completely described in [7] where it has been shown that

$$\begin{aligned}
&\frac{\partial^2}{\partial z_1\partial z_2}\left\{z_1z_2\tilde{t}_n(z_1)\tilde{t}_n(z_2)\frac{1}{n^2}\sum_{j=1}^n \mathrm{tr}(R^{1/2}\mathbb{E}_jQ_j(z_1)R\mathbb{E}_jQ_j(z_2)R^{1/2})\right\}\\
&\qquad = \Theta_{0,n}(z_1,z_2) + o_P(1),
\end{aligned}\tag{5.9}$$

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

We shall focus on the second and third terms.

5.1.2. *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 whether R is replaced by $\overline{R}$ in (2.3) and (3.1) since t_n and $\tilde{t}_n$ only depend on the spectrum of R (which is the same as the spectrum of $\overline{R}$). Notice also that $(R^{1/2})^T = \overline{R}^{1/2}$, hence

$$(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})^T = \overline{R}^{1/2}\mathbb{E}_jQ_j^T(z_2)\overline{R}^{1/2}.$$

Recall the definition of $T_n^T(z)$ given by (2.8). Taking into account the fact that for a deterministic matrix A,

$$\mathbb{E}\xi_j^TA\xi_j = \frac{\mathcal{V}}{n}\mathrm{tr}(\overline{R}^{1/2}AR^{1/2}) \quad \text{and} \quad \mathbb{E}\xi_j^*A\bar{\xi}_j = \frac{\overline{\mathcal{V}}}{n}\mathrm{tr}(R^{1/2}A\overline{R}^{1/2}),\tag{5.10}$$

and following closely [7], Section 2, it is a matter of bookkeeping to establish that

$$\frac{|\mathcal{V}|^2z_1z_2}{n^2}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\sum_{j=1}^n \mathrm{tr}(R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})^T)$$

$$=|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)\times\frac{1}{n}\sum_{j=1}^{n}\frac{1}{1-((j-1)/n)|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}+o_P(1) \tag{5.11}$$

$$=\int_0^{|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}\frac{dz}{1-z}+o_P(1),$$

where

$$\mathcal{A}_n(z_1,z_2)=\frac{z_1z_2}{n}\tilde{t}_n(z_1)\tilde{t}_n(z_2)\,\mathrm{tr}\{R^{1/2}T_n(z_1)R^{1/2}\overline{R}^{1/2}T_n^T(z_2)\overline{R}^{1/2}\}.$$

Finally,

$$\begin{aligned}\frac{\partial^2}{\partial z_1\partial z_2}(5.11)&=|\mathcal{V}|^2\Theta_{1,n}(z_1,z_2)+o_P(1)\\&=|\mathcal{V}|^2\frac{\partial}{\partial z_2}\left\{\frac{\partial\mathcal{A}_n(z_1,z_2)/\partial z_1}{1-|\mathcal{V}|^2\mathcal{A}_n(z_1,z_2)}\right\}+o_P(1).\end{aligned} \tag{5.12}$$

5.1.3. *The cumulant term in the variance.* We now handle the term proportional to κ in (5.8):

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{N}(R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2})_{ii}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})_{ii}. \tag{5.13}$$

The objective is to prove that $\mathbb{E}_jQ_j(z)$ 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 [30] for large covariance matrices based on a noncentered matrix model with separable variance profile.

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

$$\boldsymbol{\delta}_z=\mathrm{dist}(z,\mathbb{R}^+). \tag{5.14}$$

PROPOSITION 5.3. *Assume that Assumptions A-1 and A-2 hold true and let u_n be a deterministic $N\times 1$ vector, then*

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

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

Proof of Proposition 5.3 is an easy adaptation of [30], Proposition 2.7; see also the proof of Proposition 6.4 below. It is therefore omitted.

PROPOSITION 5.4. *Assume that Assumptions A-1 and A-2 hold true, then the following convergence holds true:*

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

The proof below has been suggested by a referee whom we thank; it substantially simplifies the initial one.

PROOF OF PROPOSITION 5.4. We first transform the sum to be calculated:

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{N}(R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2})_{ii}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})_{ii}. \tag{5.15}$$

Using Proposition 5.3 enables us to replace the conditional expectation $\mathbb{E}_i$ by the true expectation in every term $(R^{1/2}\mathbb{E}_jQ_j(z)R^{1/2})_{ii}$. Now using the fact that

$$Q=Q_j-\frac{Q_j\xi_j\xi_j^*Q_j}{1+\xi_j^*Q_j\xi_j}, \tag{5.16}$$

one can replace $\mathbb{E}Q_j$ by $\mathbb{E}Q$. We now prove the following estimate:

$$|\mathbb{E}u^*Q(z)v-u^*T(z)v|\le\frac{C}{\sqrt{n}\,\mathrm{Im}^k(z)}\|u\|\|v\|, \tag{5.17}$$

where neither K nor k depend on N,n. Clearly, Proposition 5.4 follows from (5.17).

Using (5.16) and the associated fact that $(Q(z)\xi_j)_i=\beta_j(z)(Q_j(z)\xi_j)_i$, we get

$$\begin{aligned}\mathbb{E}u^*Q(z)\Sigma\Sigma^*T(z)v&=\sum_j\mathbb{E}\beta_j(u^*Q_j(z)\xi_j\xi_j^*T(z)v)\\&\overset{(a)}{=}-z\tilde{t}_n(z)\sum_j\mathbb{E}(u^*Q_j(z)\xi_j\xi_j^*T(z)v)+O\left(\frac{1}{\sqrt{n}\,\mathrm{Im}^k(z)}\right)\\&=-z\frac{\tilde{t}_n(z)}{n}\sum_j\mathbb{E}(u^*Q_j(z)RT(z)v)+O\left(\frac{1}{\sqrt{n}\,\mathrm{Im}^k(z)}\right)\\&\overset{(b)}{=}-z\tilde{t}_n(z)\mathbb{E}u^*Q(z)RT(z)v+O\left(\frac{\|u\|\|v\|}{\sqrt{n}\,\mathrm{Im}^k(z)}\right),\end{aligned} \tag{5.18}$$

where we used that $\mathbb{E}|\beta_j(z) + z\tilde{t}_n(z)|^2 \le Kn^{-1}|\operatorname{Im}^{-k}(z)|$ (see, e.g., [7]) to prove (a) and we used (5.16) to replace Q_j by Q in (b).

On the other hand,

$$
\begin{aligned}
\mathbb{E}u^*Q(z)\Sigma\Sigma^*T(z)v &= \mathbb{E}u^*Q(z)(\Sigma\Sigma^* - zI_N + zI_N)T(z)v \\
&= u^*T(z)v + z\mathbb{E}u^*Q(z)T(z)v.
\end{aligned}
\tag{5.19}
$$

Taking into account that by Definitions (2.4) and (2.7)

$$
T(z) = (-z\tilde{t}_n(z)c_nR - zI_N)^{-1},
$$

we get

$$
\begin{aligned}
&u^*T(z)v - u^*\mathbb{E}Q(z)v \\
&\quad = u^*T(z)v - \mathbb{E}u^*Q(z)(-z\tilde{t}_n(z)c_nR - zI_N)T(z)v \\
&\quad = u^*T(z)v + z\tilde{t}_n(z)c_n\mathbb{E}u^*Q(z)RT(z)v + z\mathbb{E}u^*Q(z)T(z)v \\
&\quad \overset{(a)}{=} \mathbb{E}u^*Q(z)\Sigma\Sigma^*T(z)v + z\tilde{t}_n(z)c_n\mathbb{E}u^*Q(z)RT(z)v \\
&\quad \overset{(b)}{=} O\left(\frac{\|u\|\|v\|}{\sqrt{n}\operatorname{Im}^k(z)}\right),
\end{aligned}
$$

where (a) follows from (5.19) and (b) from (5.18). This completes the proof of (5.17), hence the proof of Proposition 5.4. $\square$

Combining the result in Proposition 5.4 together with (5.6) and (5.8), we have proved so far that

$$
\begin{aligned}
&\frac{\partial^2}{\partial z_1 \partial z_2}\Bigg\{\frac{z_1z_2\tilde{t}_n(z_1)\tilde{t}_n(z_2)}{n^2} \\
&\qquad \times \sum_{j=1}^{n}\sum_{i=1}^{N}(R^{1/2}\mathbb{E}_jQ_j(z_1)R^{1/2})_{ii}(R^{1/2}\mathbb{E}_jQ_j(z_2)R^{1/2})_{ii}\Bigg\} \\
&\quad = \frac{1}{n}\sum_{i=1}^{N}\frac{\partial^2}{\partial z_1\partial z_2}\{z_1z_2\tilde{t}_n(z_1)\tilde{t}_n(z_2)(R_n^{1/2}T_n(z_1)R_n^{1/2})_{ii}(R_n^{1/2}T_n(z_2)R_n^{1/2})_{ii}\} \\
&\qquad + o_P(1).
\end{aligned}
\tag{5.20}
$$

Taking into account (3.1) and the matrix identity $U(I+VU)^{-1}V = 1 - (I+UV)^{-1}$, we obtain

$$
(5.20) = \frac{1}{n}\sum_{i=1}^{N}\frac{\partial^2}{\partial z_1\partial z_2}(I_N - (I_N + \tilde{t}_n(z_1)R_n)^{-1})_{ii}(I_N - (I_N + \tilde{t}_n(z_2)R_n)^{-1})_{ii}
$$

$$+ o_P(1)$$

$$\text{(5.21)} = \frac{1}{n}\sum_{i=1}^{N} \frac{\partial}{\partial z_1}[z_1 T_n(z_1)]_{ii} \frac{\partial}{\partial z_2}[z_2 T_n(z_2)]_{ii} + o_P(1)$$

$$= \frac{z_1^2 z_2^2 \tilde{t}_n'(z_1)\tilde{t}_n'(z_2)}{n} \sum_{i=1}^{N} (R_n^{1/2} T_n^2(z_1) R_n^{1/2})_{ii} (R_n^{1/2} T_n^2(z_2) R_n^{1/2})_{ii} + o_P(1)$$

$$= \Theta_{2,n}(z_1, z_2) + o_P(1),$$

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

Now gathering (5.9), (5.12) and (5.21), 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 5.1.

5.1.4. *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 immediately 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 (3.1) yields

$$-z - \frac{1}{\tilde{t}_n(z)} + \frac{1}{n} \operatorname{tr} R_n (I_N + \tilde{t}_n(z) R_n)^{-1} = 0. \tag{5.22}$$

Following Bai and Silverstein [7], Section 4, we introduce the quantity $A_n(z)$ defined as

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

hence

$$-\frac{A_n(z)}{\mathbb{E}\tilde{f}_n(z)} = -z - \frac{1}{\mathbb{E}\tilde{f}_n(z)} + \frac{1}{n}\operatorname{tr} R_n (I_N + \mathbb{E}\tilde{f}_n(z) R_n)^{-1}. \tag{5.23}$$

Subtracting (5.22) to (5.23) finally yields

$$\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z)$$

$$
\begin{aligned}
&= -A_n(z)\tilde{t}_n(z) \\
&\quad \times \left[1 - \frac{\tilde{t}_n(z)\mathbb{E}\tilde{f}_n(z)}{n} \operatorname{tr} R_n^2 (I_N + \mathbb{E}\tilde{f}_n(z)R_n)^{-1}(I_N + \tilde{t}_n(z)R_n)^{-1}\right]^{-1},
\end{aligned}
$$

which is the counterpart of [7], equation (4.12). The same arguments as in [7] now yield

$$
\begin{aligned}
&n(\mathbb{E}\tilde{f}_n(z) - \tilde{t}_n(z)) \\
&\qquad = -nA_n(z)\tilde{t}_n(z)\left[1 - \frac{\tilde{t}_n^2(z)}{n} \operatorname{tr} R_n^2 (I_N + \tilde{t}_n(z)R_n)^{-2}\right]^{-1} + o(1).
\end{aligned} \tag{5.24}
$$

It remains to study the behavior of $nA_n(z)$. Following [7], equation (4.10), we obtain

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

Applying (2.6) to the right term to the RHS of the previous equation (recall that $R^T = \overline{R}$), we obtain

$$
\begin{aligned}
nA_n(z) &= -|\mathcal{V}|^2 \frac{b_n^2}{n}\mathbb{E} \operatorname{tr} R_n^{1/2} Q_1 (\mathbb{E}\tilde{f}_n R_n + I_N)^{-1} R_n^{1/2} \overline{R}_n^{1/2} Q_1^T \overline{R}_n^{1/2} \\
&\qquad - \kappa \frac{b_n^2}{n} \sum_{i=1}^N (R_n^{1/2} Q_1 R_n^{1/2})_{ii} (R_n^{1/2} Q_1 (\mathbb{E}\tilde{f}_n R_n + I_N)^{-1} R_n^{1/2})_{ii} + o(1).
\end{aligned} \tag{5.25}
$$

The first term of the RHS has been fully analyzed in [7] in the case where R_n and X_n are real matrices. We can adapt these computations to the general case and get the following identity[7]:

$$
\begin{aligned}
&-|\mathcal{V}|^2 \frac{b_n^2}{n}\mathbb{E} \operatorname{tr} R_n^{1/2} Q_1 (\mathbb{E}\tilde{f}_n R_n + I_N)^{-1} R_n^{1/2} \overline{R}_n^{1/2} Q_1^T \overline{R}_n^{1/2} \\
&\qquad = |\mathcal{V}|^2 \frac{\frac{z^3 \tilde{t}_n^2}{n} \operatorname{tr} R_n^{1/2} T_n^2(z) R_n^{1/2} \overline{R}_n^{1/2} T_n^T(z) \overline{R}_n^{1/2}}{1 - \frac{|\mathcal{V}|^2 z^2 \tilde{t}_n^2}{n} \operatorname{tr} R_n^{1/2} T_n(z) R_n^{1/2} \overline{R}_n^{1/2} T_n^T(z) \overline{R}_n^{1/2}} + o(1),
\end{aligned} \tag{5.26}
$$

---

[7] Details can be found in the previous version of this article, arxiv:1309.3728v3.

where $T_n^T(z)$ is defined in (2.8). The term proportional to the cumulant in (5.25) can be analyzed as in Section 5.1.3, and one can prove that

$$\begin{aligned} & -\kappa \frac{b_n^2}{n} \sum_{i=1}^N (R_n^{1/2} Q_1 R_n^{1/2})_{ii} (R_n^{1/2} Q_1 (\mathbb{E}\tilde{f}_n R_n + I_N)^{-1} R_n^{1/2})_{ii} \\ & \qquad = -\kappa \frac{z^2 \tilde{t}_n^2}{n} \sum_{i=1}^N (R_n^{1/2} T_n R_n^{1/2})_{ii} (R_n^{1/2} T_n (\tilde{t}_n R_n + I_N)^{-1} R_n^{1/2})_{ii} + o(1). \end{aligned} \tag{5.27}$$

We now plug (5.26) and (5.27) into (5.24) to conclude

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

The proof of Proposition 5.1 is completed.

5.2. *Proof of Proposition 5.2.* Recall the meta-model introduced in Section 2.6.

5.2.1. *The Gaussian process G_n.* Let

$$M_{n,M}(z) = \operatorname{tr}(\Sigma_n(M)\Sigma_n(M)^* - z I_{NM})^{-1} - MN t_n(z).$$

Applying Proposition 5.1 to the matrix model $\Sigma_n(M)\Sigma_n(M)^*$ yields

$$\forall z \in \Gamma, \qquad M_{n,M}^1(z) = \sum_{j=1}^{nM} Z_j^M(z) + o_P(1),$$

where the Z_j^M's are martingale increments and

$$\begin{aligned} \sum_{j=1}^{nM} \mathbb{E}_{j-1} Z_j^M(z_1) Z_j^M(z_2) & \xrightarrow[N,n \ \text{fixed}, M\to\infty]{\mathcal{P}} \Theta_n(z_1, z_2), \\ M_{n,M}^2(z) & \xrightarrow[N,n \ \text{fixed}, M\to\infty]{} \mathcal{B}_n(z). \end{aligned}$$

Notice that there is a genuine limit in the previous convergence. Applying the central limit theorem for martingales [11], Theorem 35.12, plus the tightness argument for $(M_{n,M}(z), z \in \Gamma)$ provided by Proposition 5.1 immediately yields the fact that $M_{n,M}$ converges in distribution to a Gaussian process $(G_n(z), z \in \Gamma)$ with mean $\mathcal{B}_n(z)$ and covariance function $\Theta_n(z_1, z_2)$.

5.2.2. *Tightness of the sequence of Gaussian processes* (G_n). In order to prove that the sequence of Gaussian processes (G_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 (2.23). 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 5.1, the family $\{M_{n,M}; M \geq 1\}_{N,n\to\infty}$ is tight, hence relatively compact in distribution. As the distribution $\mathcal{L}(G_n)$ of the Gaussian process G_n is the limit (in M) of the distribution $\mathcal{L}(M_{n,M})$ of $M_{n,M}$, $\mathcal{L}(G_n)$ belongs to the closure of $\{\mathcal{L}(M_{n,M})\}$, which is compact. Finally, $\{\mathcal{L}(G_n)\}$ is included in a compact set, hence is relatively compact. In particular, the family of Gaussian processes (G_n) is tight.

5.3. *Proof of Theorem 1.* The two propositions below are minor variations of known results. They will be helpful to conclude the proof of Theorem 1.

LEMMA 5.5 (CLT for martingales I). *Suppose that for each n $Y_{n1}, Y_{n2}, \dots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{G}_{n,j}\}$ having second moments. Assume moreover that (Θ_n^2) is a sequence of nonnegative real numbers, uniformly bounded. If*

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

and for each $\varepsilon > 0$,

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

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

$$\mathbb{E}f\left(\sum_{j=1}^{r_n} Y_{nj}\right) - \mathbb{E}f(Z_n) \xrightarrow[n\to\infty]{} 0, \tag{5.28}$$

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

Hereafter is the multidimensional and complex extension of Lemma 5.5 we shall rely on in the sequel.

LEMMA 5.6 (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 $\{\mathcal{G}_{n,j}; 1 \le j \le r_n\}$ having second moments. Write*

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

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

$$\sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^k \overline{Y}_{nj}^d | \mathcal{G}_{n,j-1}) - \Theta_n(k,\ell) \xrightarrow[n\to\infty]{\mathcal{P}} 0, \tag{5.29}$$

$$\sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^k Y_{nj}^\ell | \mathcal{G}_{n,j-1}) - \widetilde{\Theta}_n(k,\ell) \xrightarrow[n\to\infty]{\mathcal{P}} 0, \tag{5.30}$$

and for each $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathbb{E}(|Y_{nj}|^2 1_{|Y_{nj}|>\varepsilon}) \xrightarrow[n\to\infty]{} 0, \tag{5.31}$$

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

$$\mathbb{E} f\left(\sum_{j=1}^{r_n} Y_{nj}\right) - \mathbb{E} f(Z_n) \xrightarrow[n\to\infty]{} 0, \tag{5.32}$$

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

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

Lemmas 5.5 and 5.6 are variations around the central limit theorem for martingales (see Billingsley [11], Theorem 35.12) which enables us to prove (in the real case)

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

and Lévy theorem for the weak convergence criterion via characteristic functions (see Kallenberg [36], Theorems 5.3 and 5.5) which yields (5.32) from the above convergence. Details of the proof are omitted.

LEMMA 5.7 (Tightness and weak convergence). *Let K be a compact set in $\mathbb{C}$; let $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be random elements in $C(K, \mathbb{C})$. Assume that for all $d \ge 1$, for all $z_1, \dots, z_d \in K$, for all $f \in C(\mathbb{C}^d, \mathbb{C})$ we have*

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

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 5.7 can be proved as [36], Lemma 16.2; the proof is therefore omitted.

We are now in position to conclude.

In order to apply Lemma 5.6, 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 5.2.

Proposition 5.1 together with Lemma 5.6 (notice that condition (5.31) can be proved as in [7]) yield the fact that for every $z_1, \dots, z_d \in \Gamma$ and for every bounded continuous function $f : \Gamma^d \to \mathbb{C}$

$$\mathbb{E} f(M_n(z_1), \dots, M_n(z_d)) - \mathbb{E} f(G_n(z_1), \dots, G_n(z_d)) \xrightarrow[N,n\to\infty]{} 0,$$

where G_n is well defined by Proposition 5.2. Now the tightness of M_n and G_n together with Lemma 5.7 yield the last statement of Theorem 1.

6. Proof of Theorem 2 (fluctuations for nonanalytic functionals).

In this section, we will assume that the random variables (X_{ij}^n) are truncated, centered and normalized, following Section 3.2.

6.1. *Useful properties.* Recall that $\mathcal{S}_n \subset \mathcal{S}_\infty \stackrel{\triangle}{=} [0, \boldsymbol{\lambda}_R^+(1+\sqrt{\boldsymbol{\ell}^+})^2]$ uniformly in n. Denote by $\mathbf{h} \in C_c^\infty(\mathbb{R})$ a function whose value is 1 on a η-neighborhood $\mathcal{S}_\infty^\eta$ of $\mathcal{S}_\infty$.

PROPOSITION 6.1. *1. Assume that Assumptions A-1 and A-2 hold true; let the random variables (X_{ij}^n) be truncated as in Section 3.2, function $\mathbf{h}$ be defined as above and $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then*

$$\operatorname{tr} f(\Sigma_n \Sigma_n^*) - \operatorname{tr}(f\mathbf{h})(\Sigma_n\Sigma_n^*) \xrightarrow[N,n\to\infty]{\text{a.s.}} 0.$$

2. Let $\mathbf{h}_n$ be a smooth function on $\mathbb{R}$ with compact support and whose value is 1 on a η-neighborhood $\mathcal{S}_n^\eta$ of $\mathcal{S}_n$; then

$$\int_{\mathbb{R}} f(\lambda)\mathcal{F}_n(d\lambda) = \int_{\mathbb{R}} (f\mathbf{h}_n)(\lambda)\mathcal{F}_n(d\lambda).$$

The proof of Proposition 6.1 is straightforward and is based on the fact that almost surely

$$\limsup_{N,n\to\infty} \|\Sigma_n\Sigma_n^*\| < \boldsymbol{\lambda}_R^+(1+\sqrt{\boldsymbol{\ell}^+})^2 + \eta,$$

a fact that can be found in [5] for instance. Details are left to the reader.

The following proposition underlines how a sufficient regularity of function f compensates the singularity in $\operatorname{Im}(z)^{-1}$ near the real axis.

PROPOSITION 6.2. *Let μ, ν be two probability measures on $\mathbb{R}$ and g_μ and g_ν their associated Stieltjes transforms. Assume that*

$$|g_\mu(z) - g_\nu(z)| \le \frac{|h(z)|}{\mathrm{Im}(z)^k}, \qquad z \in \mathbb{C}^+,$$

where h is a continuous function over $\mathrm{cl}(\mathbb{C}^+)$, the closure of $\mathbb{C}^+$.

Let $f : \mathbb{R} \to \mathbb{R}$ be a function of order C^{k+1} with bounded support; recall the definition of $\Phi_k(f)$ in (1.6) and denote by

$$\|f\|_{k+1} = \sup_{0 \le \ell \le k+1} \|f^{(\ell)}\|_\infty \qquad \text{where } \|g\|_\infty = \sup_{x \in \mathbb{R}} |g(x)|.$$

Then

$$\begin{aligned}\left|\int f\,d\mu - \int f\,d\nu\right| &\le \frac{1}{\pi}\left|\int_{\mathbb{C}^+} \overline{\partial}\Phi_k(f)(z)\{g_\mu(z) - g_\nu(z)\}\ell_2(dz)\right| \\ &\le K\|f\|_{k+1}\int_{\mathrm{supp}(f)\times\mathrm{supp}(\chi)} |h(z)|\ell_2(dz) \qquad (6.1)\\ &\le K'\|f\|_{k+1}.\end{aligned}$$

PROOF. Write

$$\begin{aligned}\overline{\partial}\Phi_k(f)(x + \mathbf{i}y) &= \partial_x\Phi_k(f)(x + \mathbf{i}y) + \mathbf{i}\partial_y\Phi_k(f)(x + \mathbf{i}y) \\ &= \frac{(\mathbf{i}y)^k f^{(k+1)}(x)}{k!}\chi(y) + \mathbf{i}\sum_{\ell=0}^{k}\frac{(\mathbf{i}y)^\ell f^{(\ell)}(x)}{\ell!}\chi'(y).\end{aligned}$$

From this and the fact that χ is equal to 1 for y small enough, we deduce that

$$\overline{\partial}\Phi_k(f)(x + \mathbf{i}y) = \frac{(\mathbf{i}y)^k f^{(k+1)}(x)}{k!}$$

near the real axis. Hence, $|\overline{\partial}\Phi_k(f)(x + \mathbf{i}y)| \le 1_{\mathrm{supp}(f)\times\mathrm{supp}(\chi)}(x, y)K\|f\|_{k+1}y^k$ near the real axis, which yields (6.1). $\square$

6.2. *Proof of Theorem 2.* Recall the definition of the sets D, D^+ and D_ε given in (3.7) and the fact that constant $A > \boldsymbol{\lambda}_R^+(1 + \sqrt{\boldsymbol{\ell}^+})^2$.

LEMMA 6.3. *Let $(\varphi_n(z), z \in D^+ \cup \overline{D^+})_{n\in\mathbb{N}}$ and $(\psi_n(z), z \in D^+ \cup \overline{D^+})_{n\in\mathbb{N}}$ be centered complex-valued continuous random processes and such that $\varphi(\bar{z}) = \overline{\varphi(z)}$ and $\psi(\bar{z}) = \overline{\psi(z)}$. Assume that:*

(i) *The following convergence in distribution holds true: for all $d \ge 1$ and $(z_1, \dots, z_d) \in D^+$,*

$$d_{\mathcal{LP}}((\varphi_n(z_1), \dots, \varphi_n(z_d)), (\psi_n(z_1), \dots, \psi_n(z_d))) \xrightarrow[n\to\infty]{} 0.$$

(ii) *For all* $\varepsilon > 0$, $\varphi_n(z)$ *and* $\psi_n(z)$ *are tight on* D_ε.

(iii) *The process* $(\psi_n(z))$ *is Gaussian with covariance matrix* $\kappa_n(z_1, z_2)$, $(z_1, z_2 \in D^+ \cup \overline{D^+})$.

(iv) *The following estimates hold true:*

$$\forall n \in \mathbb{N}, \forall z \in D^+, \qquad \operatorname{var} \varphi_n(z) \le \frac{1}{\operatorname{Im}(z)^{2k}} \quad \textit{and} \quad \operatorname{var} \psi_n(z) \le \frac{1}{\operatorname{Im}(z)^{2k}}.$$

(v) *Let functions* $g_\ell : \mathbb{R} \to \mathbb{R}$ $(1 \le \ell \le L)$ *be* C^{k+1} *and have compact support.*

Then

$$d_{\mathcal{LP}}\left(\frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(\mathbf{g})(z)\varphi_n(z)\ell_2(dz), \frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(\mathbf{g})(z)\psi_n(z)\ell_2(dz)\right) \xrightarrow[n\to\infty]{} 0,$$

where

$$\overline{\partial}\Phi_k(g_j)(z) = (\partial_x + \mathbf{i}\partial_y) \sum_{\ell=0}^{k} \frac{(\mathbf{i}y)^\ell}{\ell!} g_j^{(\ell)}(x)\chi(y) \quad \textit{and}$$

$$\overline{\partial}\Phi_k(\mathbf{g}) = (\overline{\partial}\Phi_k(g_j); 1 \le j \le L)$$

with χ *being smooth, compactly supported with value 1 in a neighborhood of 0. Moreover,*

$$\frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(\mathbf{g})(z)\psi_n(z)\ell_2(dz)$$

is centered Gaussian with covariance matrix

$$\begin{aligned}
&\operatorname{cov}\left(\frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(g_k)(z)\psi_n(z)\ell_2(dz), \frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial}\Phi_k(g_\ell)(z)\psi_n(z)\ell_2(dz)\right) \\
&= \frac{1}{2\pi^2} \operatorname{Re} \int_{(\mathbb{C}^+)^2} \overline{\partial}\Phi_k(g_k)(z_1)\overline{\overline{\partial}\Phi_k(g_\ell)(z_2)}\kappa_n(z_1, \bar{z}_2)\ell_2(dz_1)\ell_2(dz_2) \\
&\quad + \frac{1}{2\pi^2} \operatorname{Re} \int_{(\mathbb{C}^+)^2} \overline{\partial}\Phi_k(g_k)(z_2)\overline{\partial}\Phi_k(g_\ell)(z_2)\kappa_n(z_1, z_2)\ell_2(dz_1)\ell_2(dz_2),
\end{aligned} \tag{6.2}$$

for $1 \le k, \ell \le L$.

The proof of Lemma 6.3 is provided in Appendix A.1.

The strategy to prove Theorem 2 closely follows this lemma. Denote by

$$\varphi_n(z) = \operatorname{tr} Q_n(z) - \mathbb{E} \operatorname{tr} Q_n(z) \quad \text{and} \quad \psi_n(z) = G_n(z) - \mathbb{E}G_n(z),$$

the process G_n being defined in Theorem 1, then conditions (i), (ii) and (iii) are immediate consequences of Theorem 1. In order to check condition (iv), we establish the following proposition.

PROPOSITION 6.4. *Assume that Assumptions A-1 and A-2 hold true, then:*

(i) *(Bordenave [12], Hachem et al. [29], Lemma 6.3, Shcherbina [49]). For all $z \in \mathbb{C}^+$,*

$$\operatorname{var} \operatorname{tr} Q_n(z) \leq \frac{C}{\operatorname{Im}(z)^4}.$$

(ii) *For all $z \in \mathbb{C}^+$,*

$$\operatorname{var} G_n(z) \leq \frac{C}{\operatorname{Im}(z)^4},$$

where C is a constant that may depend polynomially on $|z|$.

The first part of the proposition is classical and its proof is omitted (for the details, see footnote 7). Proof of Proposition 6.4(ii) is postponed to Appendix A.2.

Taking into account the estimates established in Proposition 6.4 immediately yields the first part of Theorem 2 in the case where functions (g_ℓ) have a bounded support and satisfy (v) with $k = 2$, that is, are C^3. It remains to prove the equivalence between (4.3) and (4.4), but this immediately follows from the following.

PROPOSITION 6.5. *Let (X_n) and (Y_n) be $\mathbb{C}^d$-valued random variables and assume that both sequences are tight, then the following are equivalent:*

(i) *the following convergence holds true: $d_{\mathcal{LP}}(X_n, Y_n) \underset{n\to\infty}{\longrightarrow} 0$;*

(ii) *for every continuous bounded function $f : \mathbb{C}^d \to \mathbb{C}$, $\mathbb{E}f(X_n) - \mathbb{E}f(Y_n) \underset{n\to\infty}{\longrightarrow} 0$.*

Proposition 6.5 can be proved easily by contradiction using the fact that $d_{\mathcal{LP}}$ meterizes the convergence of laws; its proof is hence omitted.

6.3. *Proof of Proposition 4.1.* Let $f \in C_c^\infty(\mathbb{R}^2)$. A simple but lengthy computation yields the fact that

$$\begin{aligned} &\overline{\partial}_2 \overline{\partial}_1 \Phi_{N_1,N_2}(f)(x + \mathbf{i}u, y + \mathbf{i}v) \\ &\quad = \frac{\partial^{N_1+N_2+2}}{\partial x^{N_1+1} \partial y^{N_2+1}} f(x,y) \times \frac{(\mathbf{i}u)^{N_1}}{N_1!} \frac{(\mathbf{i}v)^{N_2}}{N_2!} \end{aligned} \tag{6.3}$$

for u, v small enough. Let now $N_1 = N_2 = 2$. Since $|\Theta_n(z_1, z_2)| \leq K|z_1 z_2|^{-2}$ for any $z_1, z_2 \in \mathbb{C}^+$ and z_1, z_2 in a compact set (use Cauchy–Schwarz and apply Proposition 6.4), $\Upsilon(f)$ is well defined. Let K be a compact set in $\mathbb{R}^2$

and let $f \in C_c^\infty(\mathbb{R}^2)$ with support included in K, then one can easily prove that

$$|\Upsilon(f)| \le C_K \|f\|_{3,3} \qquad \text{with } \|f\|_{3,3} = \sup_{\substack{\ell,p\le 3\\(x,y)\in K}} \|\partial_x^\ell \partial_y^p f(x,y)\|_\infty.$$

This in particular implies that Υ is a distribution on $C_c^\infty(\mathbb{R}^2)$, of finite order $(3,3)$, and hence uniquely extends as a distribution on $C_c^{3,3}(\mathbb{R}^2)$.

Moreover, $\Upsilon(f)$ can be written as

$$\begin{aligned}\Upsilon(f) = {} & \lim_{\varepsilon\downarrow 0} \frac{1}{2\pi^2} \operatorname{Re} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(z_1,z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2)\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{2\pi^2} \operatorname{Re} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(z_1,\overline{z_2})\Theta_n(z_1,\overline{z_2})\ell_2(dz_1)\ell_2(dz_2),\end{aligned}$$

where $\mathbb{C}_\varepsilon^+ = \{z \in \mathbb{C}, \operatorname{Im}(z) \ge \varepsilon\}$. Taking into account the facts that

$$\overline{\overline{\partial}_2 \overline{\partial}_1 \Phi_{n_1,n_2}(f)(z_1,z_2)} = \overline{\partial}_2 \overline{\partial}_1 \Phi_{n_1,n_2}(f)(\overline{z_1},\overline{z_2}) \quad \text{and} \quad \overline{\Theta_n(z_1,z_2)} = \Theta_n(\overline{z_1},\overline{z_2}),$$

we can expand $\Upsilon(f)$ as

$$\begin{aligned}\Upsilon(f) = {} & \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(z_1,z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2)\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(\overline{z_1},\overline{z_2})\Theta_n(\overline{z_1},\overline{z_2})\ell_2(dz_1)\ell_2(dz_2)\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(z_1,\overline{z_2})\Theta_n(z_1,\overline{z_2})\ell_2(dz_1)\ell_2(dz_2)\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{(\mathbb{C}_\varepsilon^+)^2} \overline{\partial}_2 \overline{\partial}_1 \Phi_{2,2}(f)(\overline{z_1},z_2)\Theta_n(\overline{z_1},z_2)\ell_2(dz_1)\ell_2(dz_2).\end{aligned}$$

We now apply twice Green's formula to each integral and obtain

$$\begin{aligned}\Upsilon(f) = {} & -\lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \Phi_{2,2}(f)(x_1+\mathbf{i}\varepsilon, x_2+\mathbf{i}\varepsilon)\Theta_n(x_1+\mathbf{i}\varepsilon, x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\\ & - \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \Phi_{2,2}(f)(x_1-\mathbf{i}\varepsilon, x_2-\mathbf{i}\varepsilon)\Theta_n(x_1-\mathbf{i}\varepsilon, x_2-\mathbf{i}\varepsilon)\,dx_1\,dx_2\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \Phi_{2,2}(f)(x_1+\mathbf{i}\varepsilon, x_2-\mathbf{i}\varepsilon)\Theta_n(x_1+\mathbf{i}\varepsilon, x_2-\mathbf{i}\varepsilon)\,dx_1\,dx_2\\ & + \lim_{\varepsilon\downarrow 0} \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \Phi_{2,2}(f)(x_1-\mathbf{i}\varepsilon, x_2+\mathbf{i}\varepsilon)\Theta_n(x_1-\mathbf{i}\varepsilon, x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2.\end{aligned}$$

Notice that the sign changes in the two last integrals follow from the contour orientations in Green's formula. We now prove

$$\begin{aligned}&\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^2}\Phi_{2,2}(f)(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\\&\qquad=\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^2}f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2.\end{aligned}\tag{6.4}$$

The three other integrals can be handled similarly, and this will achieve the boundary value representation (4.6) for $\Upsilon(f)$.

Using the mere definition of $\Phi_{N_1,N_2}(f)$ [cf. (4.5)] and Green's formula, we get

$$\begin{aligned}&\int_{(\mathbb{C}_\varepsilon^+)^2}\overline{\partial}_2\overline{\partial}_1\Phi_{1,0}(f)(z_1,z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2)\\&\qquad=-\int_{\mathbb{R}^2}\Phi_{1,0}(f)(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\\&\qquad=-\int_{\mathbb{R}^2}f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\\&\qquad\quad-\mathbf{i}\varepsilon\int_{\mathbb{R}^2}\partial_x f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2.\end{aligned}$$

We extract the first term of the RHS from the equation above. Taking into account (6.3) and the fact that $|\Theta_n(z_1,z_2)|\le|z_1z_2|^{-2}$ for z_1,z_2 in a compact set of $\mathbb{C}\setminus\mathbb{R}$, we obtain

$$\limsup_{\varepsilon\downarrow 0}\left|\varepsilon^3\int_{\mathbb{R}^2}f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\right|<\infty.$$

By applying the same argument to the quantity

$$\int_{(\mathbb{C}_\varepsilon^+)^2}\overline{\partial}_2\overline{\partial}_1\Phi_{4-\ell,0}(f)(z_1,z_2)\Theta_n(z_1,z_2)\ell_2(dz_1)\ell_2(dz_2)$$

for $\ell=2$ then $\ell=1$ and $\ell=0$, we can similarly prove that

$$\limsup_{\varepsilon\downarrow 0}\left|\varepsilon^\ell\int_{\mathbb{R}^2}f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\right|<\infty\qquad\text{for }\ell=2,1,0.$$

We finally obtain

$$\limsup_{\varepsilon\downarrow 0}\left|\int_{\mathbb{R}^2}f(x_1,x_2)\Theta_n(x_1+\mathbf{i}\varepsilon,x_2+\mathbf{i}\varepsilon)\,dx_1\,dx_2\right|<\infty.$$

Expanding $\Phi_{2,2}(f)$ into (6.4) and using the above estimate immediately yields (6.4).

The proof of Proposition 4.1 is complete.

6.4. *Proof of Proposition 4.2.* The covariance writes (in short)

$$\begin{aligned}
&\operatorname{cov}(Z_n^1(f), Z_n^1(g)) \\
&\qquad = -\frac{1}{4\pi^2} \lim_{\varepsilon\downarrow 0} \sum_{\pm_1,\pm_2} (\pm_1\pm_2) \int f(x)g(y)\Theta_n(x \pm_1 \mathbf{i}\varepsilon, y \pm_2 \mathbf{i}\varepsilon)\, dx\, dy,
\end{aligned}$$

where $\pm_1, \pm_2 \in \{+,-\}$ and $\pm_1\pm_2$ is the sign resulting from the product $\pm_1 1$ by $\pm_2 1$. Unfolding $\Theta_n = \Theta_{0,n} + |\mathcal{V}|^2\Theta_{1,n} + \kappa\Theta_{2,n}$, we have three terms to compute. According to the assumptions of Proposition 4.2, either $|\mathcal{V}|^2$ equals 1 or 0. In the latter case, the term corresponding to $\Theta_{1,n}$ vanishes; if $|\mathcal{V}|^2 = 1$, then the quantities $\mathcal{A}_n$ and $\mathcal{A}_{0,n}$ [resp., defined in (2.13) and (3.10)] are equal, and so are $\Theta_{0,n}$ and $\Theta_{1,n}$. We first establish

$$\begin{aligned}
&-\frac{1}{4\pi^2} \lim_{\varepsilon\downarrow 0} \sum_{\pm_1,\pm_2} (\pm_1\pm_2) \int f(x)g(y)\Theta_{0,n}(x \pm_1 \mathbf{i}\varepsilon, y \pm_2 \mathbf{i}\varepsilon)\, dx\, dy \\
&\qquad = \frac{1}{2\pi^2} \int_{\mathcal{S}_n^2} f'(x)g'(y) \ln\left|\frac{\tilde{t}_n(x) - \overline{\tilde{t}_n(y)}}{\tilde{t}_n(x) - \tilde{t}_n(y)}\right| dx\, dy.
\end{aligned} \tag{6.5}$$

The proof relies on formula (3.9) and the following expression of $\mathcal{A}_{0,n}$:

$$1 - \mathcal{A}_{0,n}(z_1, z_2) = \frac{(z_1 - z_2)\tilde{t}_n(z_1)\tilde{t}_n(z_2)}{\tilde{t}_n(z_1) - \tilde{t}_n(z_2)} \tag{6.6}$$

which can be obtained using (3.1). Using (3.9) and performing a double integration by parts yields

$$\begin{aligned}
&\int f(x)g(y)\Theta_{0,n}(x + \mathbf{i}\varepsilon, y + \mathbf{i}\varepsilon)\, dx\, dy \\
&\qquad = \int f'(x)g'(y) \ln|1 - \mathcal{A}_{0,n}(x + \mathbf{i}\varepsilon, y + \mathbf{i}\varepsilon)|\, dx\, dy \\
&\qquad\quad + \mathbf{i} \int f'(x)g'(y) \operatorname{Arg}(1 - \mathcal{A}_{0,n}(x + \mathbf{i}\varepsilon, y + \mathbf{i}\varepsilon))\, dx\, dy.
\end{aligned}$$

Following [7], Section 5, we need only to consider the logarithm term and show its convergence since the Arg term will eventually disappear (functions f and g being real, the covariance will be real as well). Using (6.6), we obtain

$$\begin{aligned}
&\int f'(x)g'(y) \ln|1 - \mathcal{A}_{0,n}(x + \mathbf{i}\varepsilon, y + \mathbf{i}\varepsilon)|\, dx\, dy \\
&\qquad = \int f'(x)g'(y) \ln\left|\frac{(x - y)\tilde{t}_n(x + \mathbf{i}\varepsilon)\tilde{t}_n(y + \mathbf{i}\varepsilon)}{\tilde{t}_n(x + \mathbf{i}\varepsilon) - \tilde{t}_n(y + \mathbf{i}\varepsilon)}\right| dx\, dy
\end{aligned}$$

and the sum writes

$$\sum_{\pm_1,\pm_2}(\pm_1\pm_2)\int f(x)g(y)\Theta_n(x\pm_1\mathbf{i}\varepsilon, y\pm_2\mathbf{i}\varepsilon)\,dx\,dy$$

$$= 2\int f'(x)g'(y)\ln\left\{\left|\frac{(x-y)\tilde{t}_n(x+\mathbf{i}\varepsilon)\tilde{t}_n(y+\mathbf{i}\varepsilon)}{\tilde{t}_n(x+\mathbf{i}\varepsilon)-\tilde{t}_n(y+\mathbf{i}\varepsilon)}\right|\right.$$

$$\left.\times\left|\frac{\tilde{t}_n(x+\mathbf{i}\varepsilon)-\tilde{t}_n(y-\mathbf{i}\varepsilon)}{(x-y+2\mathbf{i}\varepsilon)\tilde{t}_n(x+\mathbf{i}\varepsilon)\tilde{t}_n(y-\mathbf{i}\varepsilon)}\right|\right\}dx\,dy$$

$$\stackrel{(a)}{=} 2\int f'(x)g'(y)\left\{\ln\left|\frac{x-y}{x-y+2\mathbf{i}\varepsilon}\right| + \ln\left|\frac{\tilde{t}_n(x+\mathbf{i}\varepsilon)-\tilde{t}_n(y-\mathbf{i}\varepsilon)}{\tilde{t}_n(x+\mathbf{i}\varepsilon)-\tilde{t}_n(y+\mathbf{i}\varepsilon)}\right|\right\}dx\,dy,$$

where (a) follows from the fact that $\tilde{t}_n(\bar{z})=\overline{\tilde{t}_n(z)}$ and $|z|=|\bar{z}|$. It is straightforward to prove that the first integral of the RHS vanishes as $\varepsilon\to 0$. Using similar arguments as in [7], Section 5, one can prove that

$$\sum_{\pm_1,\pm_2}(\pm_1\pm_2)\int f(x)g(y)\Theta_n(x\pm_1\mathbf{i}\varepsilon, y\pm_2\mathbf{i}\varepsilon)\,dx\,dy$$

$$= 2\int f'(x)g'(y)\ln\left|\frac{\tilde{t}_n(x)-\overline{\tilde{t}_n(y)}}{\tilde{t}_n(x)-\tilde{t}_n(y)}\right|dx\,dy,$$

which is the desired result. We now establish

$$\begin{aligned} &-\frac{\kappa}{4\pi^2}\lim_{\varepsilon\downarrow 0}\sum_{\pm_1,\pm_2}(\pm_1\pm_2)\int f(x)g(y)\Theta_{2,n}(x\pm_1\mathbf{i}\varepsilon, y\pm_2\mathbf{i}\varepsilon)\,dx\,dy \\ &\quad=\frac{\kappa}{\pi^2 n}\sum_{i=1}^{N}\left(\int_{\mathcal{S}_n} f'(x)\,\mathrm{Im}(xT_n(x))_{ii}\,dx\right)\left(\int_{\mathcal{S}_n} g'(y)\,\mathrm{Im}(yT_n(y))_{ii}\,dy\right). \end{aligned}\tag{6.7}$$

Due to formula (3.11), we only need to prove

$$\begin{aligned} &\frac{\mathbf{i}}{2\pi}\lim_{\varepsilon\downarrow 0}\sum_{\pm\in\{+,-\}}\pm\int f(x)\frac{\partial}{\partial x}[(x\pm\mathbf{i}\varepsilon)T_n(x\pm\mathbf{i}\varepsilon)]_{ii}\,dx \\ &\quad=\frac{1}{\pi}\int_{\mathcal{S}_n} f'(x)\,\mathrm{Im}(xT_n(x))_{ii}\,dx. \end{aligned}\tag{6.8}$$

Performing an integration by parts and taking into account the fact that $T_n(\bar{z})=\overline{T_n(z)}$ yields

$$\frac{\mathbf{i}}{2\pi}\lim_{\varepsilon\downarrow 0}\sum_{\pm\in\{+,-\}}\pm\int f(x)\frac{\partial}{\partial x}[(x\pm\mathbf{i}\varepsilon)T_n(x\pm\mathbf{i}\varepsilon)]_{ii}\,dx$$

$$=-\frac{\mathbf{i}}{2\pi}\lim_{\varepsilon\downarrow 0}\int f'(x)2\mathbf{i}\,\mathrm{Im}[(x+\mathbf{i}\varepsilon)T_n(x+\mathbf{i}\varepsilon)]_{ii}\,dx$$

$$\overset{(a)}{=} \frac{1}{\pi} \int_{\mathcal{S}_n} f'(x) \operatorname{Im}(xT_n(x))_{ii} \, dx,$$

where step (a) follows from the fact that

$$\inf_{\substack{1 \le i \le N, \\ z \in (0,A] \times (0,B]}} |(1 + \tilde{t}_n(z)\lambda_i)| > 0, \tag{6.9}$$

where the λ_i's stand for R_n's eigenvalues. In fact, assume that (6.9) holds true, then using the spectral decomposition of R_n, the pointwise convergence of $\tilde{t}_n(z)$ to $\tilde{t}_n(x)$ as $\mathbb{C}^+ \ni z \to x \in \mathbb{R}$ (see, e.g., [51]) and formula (3.1), then one obtains the pointwise convergence

$$\operatorname{Im}[(x + \mathbf{i}\varepsilon)T_n(x + \mathbf{i}\varepsilon)]_{ii} \xrightarrow[\varepsilon \to 0]{} \operatorname{Im}[xT_n(x)]_{ii}$$

for $x > 0$. Since $\operatorname{Im}(\tilde{t}(x)) = 0$ outside $\mathcal{S}_n$, so is $\operatorname{Im}[xT_n(x)]_{ii}$. Finally, (6.9) provides a uniform bound for $\operatorname{Im}[(x + \mathbf{i}\varepsilon)T_n(x + \mathbf{i}\varepsilon)]_{ii}$ and (a) follows from the dominated convergence theorem. It remains to prove (6.9). Assume that the infimum is zero, then there exists $\lambda^* \in \{\lambda_1, \dots, \lambda_N\}$ with $\lambda^* \neq 0$ and a sequence (z_ℓ) such that $\tilde{t}_n(z_\ell) \to -\frac{1}{\lambda^*}$ and $z_\ell \to x^* \in \mathbb{R}$. Formula (3.1) yields

$$\forall z \in \mathbb{C}^+, \qquad \tilde{t}_n(z) = \frac{1}{-z + \frac{1}{n} \sum_{i=1}^N \frac{\lambda_i}{1 + \tilde{t}_n(z)\lambda_i}}$$

$$\Leftrightarrow \quad \frac{1}{n} \sum_{i=1}^N \frac{\lambda_i}{1 + \tilde{t}_n(z)\lambda_i} = \frac{1}{\tilde{t}_n(z)} + z.$$

Taking $z = z_\ell$ yields a contradiction since the LHS goes to infinity while the RHS remains bounded. Necessarily, (6.9) holds true and (6.7) is proved.

The proof of Proposition 4.2 is complete by gathering (6.5), (6.7) and using the fact that $\Theta_{0,n} = \Theta_{1,n}$.

6.5. *Proof of Corollary 4.3.* In order to establish the fluctuations in the case where functions (f_ℓ) are C^3 in a neighborhood of $\mathcal{S}_\infty$ but may not have a bounded support, we proceed as following: Write

$$\begin{aligned} &\operatorname{tr} f_\ell(\Sigma_n \Sigma_n^*) - \mathbb{E} \operatorname{tr}(f_\ell \mathbf{h})(\Sigma_n \Sigma_n^*) \\ &\quad = \underbrace{\operatorname{tr} f_\ell(\Sigma_n \Sigma_n^*) - \operatorname{tr}(f_\ell \mathbf{h})(\Sigma_n \Sigma_n^*)}_{\Gamma_\ell^1} + \underbrace{\operatorname{tr}(f_\ell \mathbf{h})(\Sigma_n \Sigma_n^*) - \mathbb{E} \operatorname{tr}(f_\ell \mathbf{h})(\Sigma_n \Sigma_n^*)}_{\Gamma_\ell^2}. \end{aligned}$$

By Proposition 6.1, the vector (Γ_ℓ^1) almost surely converges to zero while the fluctuations for vector (Γ_ℓ^2) are described by Theorem 2 with covariance given by Proposition 4.2, where functions f_k and f_ℓ must be replaced by $(f_k \mathbf{h})$ and $(f_\ell \mathbf{h})$. The variance formula provided in this proposition shows that $\operatorname{cov}(Z_1^n(f_\ell \mathbf{h}), Z_1^n(f_{\ell'} \mathbf{h}))$ does not depend on function $\mathbf{h}$ as long as $\mathbf{h}$ has value 1 on $\mathcal{S}_n$.

7. Proof of Theorem 3 (bias for nonanalytic functionals).

7.1. *Proof of Theorem 3.* Denote by $X_n^{\mathbb{C}}$ a $N \times n$ matrix whose entries are independent standard complex circular Gaussian r.v. [i.e., $X_{ij}^{\mathbb{C}} = U + iV$ where U, V are independent $\mathcal{N}(0, 2^{-1})$ random variables]; denote accordingly $\Sigma_n^{\mathbb{C}} = n^{-1/2} R^{1/2} X_n^{\mathbb{C}}$, $\xi_j^{\mathbb{C}} = (\Sigma_n^{\mathbb{C}})_{\cdot j}$ and

$$Q_n^{\mathbb{C}}(z) = (-zI_N + \Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^*)^{-1}.$$

We split the bias into two terms:

$$\begin{aligned}
\mathbb{E} \operatorname{Tr} f(\Sigma_n \Sigma_n^*) &- N \int f(\lambda) \mathcal{F}_n(d\lambda) \\
&= \mathbb{E} \operatorname{Tr} f(\Sigma_n \Sigma_n^*) - \mathbb{E} \operatorname{Tr} f(\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^*) \\
&\quad + \mathbb{E} \operatorname{Tr} f(\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^*) - N \int f(\lambda) \mathcal{F}_n(d\lambda) \\
&\stackrel{\triangle}{=} T_1 + T_2.
\end{aligned}$$

We will prove the following. Provided that function f is of class C^8 with bounded support, then

$$\begin{aligned}
&\mathbb{E} \operatorname{Tr} f(\Sigma_n \Sigma_n^*) - \mathbb{E} \operatorname{Tr} f(\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^*) - \frac{1}{\pi} \operatorname{Re} \int_{\mathbb{C}^+} \overline{\partial} \Phi_7(f)(z) \mathcal{B}_n(z) \ell_2(dz) \\
&\qquad \underset{N,n\to\infty}{\longrightarrow} 0.
\end{aligned} \tag{7.1}$$

Provided that function f is of class C^{18} with bounded support, then

$$\mathbb{E} \operatorname{Tr} f(\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^*) - N \int f(\lambda) \mathcal{F}_n(d\lambda) \underset{N,n\to\infty}{\longrightarrow} 0. \tag{7.2}$$

As one can check, it is much more demanding in terms of assumptions to prove (7.2) than (7.1). Convergence in (7.2) should be compared to the results in Haagerup and Thørbjornsen [25] (counterpart in the GUE case), Schultz [48] (GOE), Capitaine and Donati-Martin [15], Loubaton et al. [55] ("signal plus noise" model), etc.

7.1.1. *Proof of (7.1).* The heart of the proof lies in Helffer–Sjöstrand's formula, in Theorem 1 (bias part) and in a dominated convergence argument. By Theorem 1,

$$\mathbb{E} \operatorname{Tr}(\Sigma_n \Sigma_n^* - zI_N)^{-1} - N t_n(z) - \mathcal{B}_n(z) \underset{N,n\to\infty}{\longrightarrow} 0.$$

The same argument yields

$$\mathbb{E} \operatorname{Tr}(\Sigma_n^{\mathbb{C}} (\Sigma_n^{\mathbb{C}})^* - zI_N)^{-1} - N t_n(z) \underset{N,n\to\infty}{\longrightarrow} 0,$$

because in the later case $\mathcal{V} = \kappa = 0$, hence the bias is zero for the matrix model $\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^*$. Subtracting yields

$$\mathbb{E}\,\mathrm{Tr}\,Q_n(z) - \mathbb{E}\,\mathrm{Tr}\,Q_n^{\mathbb{C}}(z) - \mathcal{B}_n(z) \underset{N,n\to\infty}{\longrightarrow} 0.$$

The following proposition will be of help.

Proposition 7.1. *Assume that Assumptions A-1 and A-2 hold true, then*

$$|\mathbb{E}\,\mathrm{Tr}\,Q(z) - \mathbb{E}\,\mathrm{Tr}\,Q^{\mathbb{C}}(z)| \le K\frac{|z|^3}{\mathrm{Im}(z)^7}, \tag{7.3}$$

where K is independent from N, n, z.

The proof is based on classical rank-one perturbation arguments and is omitted (details can be found in Section 5.1.3 of the previous version of this article—see footnote 7).

In order to transfer this bound to $\mathcal{B}_n(z)$, we invoke a meta-model argument (cf. Section 2.6): Consider matrix $\Sigma_n(M)$ and its counterpart $\Sigma_n^{\mathbb{C}}(M)$ as defined in (2.24) and recall that in this case, we have a genuine limit:

$$\begin{aligned}\mathbb{E}\,\mathrm{Tr}(\Sigma_n(M)\Sigma_n^*(M) - zI_{NM})^{-1} - \mathbb{E}\,\mathrm{Tr}(\Sigma_n^{\mathbb{C}}(M)(\Sigma_n^{\mathbb{C}}(M))^* - zI_{NM})^{-1}\\ \underset{\substack{M\to\infty\\ N,n \text{ fixed}}}{\longrightarrow} \mathcal{B}_n(z).\end{aligned}$$

Since the estimate (7.3) remains true for all $M \ge 1$, we obtain

$$\begin{aligned}|\mathcal{B}_n(z)| &= \lim_{M\to\infty} |\mathbb{E}\,\mathrm{Tr}(\Sigma_n(M)\Sigma_n(M)^* - zI_{MN})^{-1}\\ &\qquad - \mathbb{E}\,\mathrm{Tr}(\Sigma^{\mathbb{C}}(M)_n(\Sigma_n^{\mathbb{C}}(M))^* - zI_{NM})^{-1}|\\ &\le K\frac{|z|^3}{\mathrm{Im}(z)^7}.\end{aligned} \tag{7.4}$$

Write

$$\begin{aligned}&\mathbb{E}\,\mathrm{Tr}\,f(\Sigma_n\Sigma_n^*) - \mathbb{E}\,\mathrm{Tr}\,f(\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^*) - \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+}\overline{\partial}\Phi(f)(z)\mathcal{B}_n(z)\ell_2(dz)\\ &\quad = \frac{1}{\pi}\mathrm{Re}\int_{\mathbb{C}^+}\overline{\partial}\Phi(f)(z)\{\mathbb{E}\,\mathrm{Tr}\,Q_n(z) - \mathbb{E}\,\mathrm{Tr}\,Q_n^{\mathbb{C}}(z) - \mathcal{B}_n(z)\}\ell_2(dz).\end{aligned} \tag{7.5}$$

In view of (7.5), we need a dominated convergence argument in order to prove (7.1); such an argument follows from Proposition 6.2, (7.3) and (7.4) as long as f is of class C^8 with large but bounded support. This completes the proof of (7.1).

7.1.2. *Proof of (7.2).* The gist of the proof lies in the following proposition.

PROPOSITION 7.2. *Denote by $P_\ell(X)$ a polynomial in X with degree ℓ and positive coefficients, then*

$$|\mathbb{E}\operatorname{Tr}(\Sigma_n^{\mathbb{C}}(\Sigma_n^{\mathbb{C}})^* - zI_N)^{-1} - Nt_n(z)| \le \frac{1}{n} P_{12}(|z|) P_{17}(|\operatorname{Im}(z)|^{-1}).$$

The proof of Proposition 7.2 builds upon techniques borrowed from [25, 55] and is omitted. Details can be found in the previous version of this article; see footnote 7 and [57].

Using Helffer–Sjöstrand's formula, Proposition 7.2 together with Proposition 6.2 immediately yield (7.2) for any f of class C^{18} with large but bounded support.

7.2. *Proof of Proposition 4.4.* One can easily prove that Z_n^2 is a distribution on $C_c^{18}(\mathbb{R})$ following the lines of proof of Proposition 4.1. Similarly, one can establish the boundary value representation (4.12). It remains to prove that the singular points of $\mathcal{B}_n(z)$ are included in $\mathcal{S}_n$. Following the definitions of $\mathcal{B}_{1,n}$ and $\mathcal{B}_{2,n}$ [cf. (2.20) and (2.21)], we simply need to prove that the quantities

$$\left(1 - z^2 \tilde{t}_n^2 \frac{1}{n} \operatorname{Tr} R_n^2 T_n^2\right) \quad \text{and} \quad \left(1 - |\mathcal{V}|^2 z^2 \tilde{t}_n^2 \frac{1}{n} \operatorname{Tr} R_n^{1/2} T_n(z) R_n^{1/2} \overline{R}_n^{1/2} T_n^T(z) \overline{R}_n^{1/2}\right)$$

are invertible for $z \notin \mathcal{S}_n$. We focus on the first one. Assume first that $z \in \mathbb{C} \setminus \mathbb{R}$. Using the inequality $|\operatorname{tr}(AB)| \le (\operatorname{tr}(AA^*)\operatorname{tr}(BB^*))^{1/2}$ yields

$$\left| z^2 \tilde{t}_n^2(z) \frac{1}{n} \operatorname{Tr} R_n^2 T_n^2(z) \right| \le \frac{|z|^2 |\tilde{t}_n(z)|^2}{n} \operatorname{tr} R_n T_n(z) R_n T_n^*(z).$$

Since $T_n^*(z) = T_n(\bar{z})$, we can assume without loss of generality that $z_1, z_2 \in \mathbb{C}^+$:

$$\begin{aligned} \left| 1 - z^2 \tilde{t}_n^2(z) \frac{1}{n} \operatorname{Tr} R_n^2 T_n^2(z) \right| &\ge 1 - \frac{|z|^2 |\tilde{t}_n(z)|^2}{n} \operatorname{tr} R_n T_n(z) R_n T_n^*(z) \\ &= |\tilde{t}_n(z)|^2 \frac{\operatorname{Im}(z)}{\operatorname{Im}(\tilde{t}_n(z))}, \end{aligned} \tag{7.6}$$

where the last identity can be found in the previous version of this article [equation (A.15)]; see footnote 7. In order to extend the previous estimate to $z \in \mathbb{R} \setminus \mathcal{S}_n$, let $z = x + \mathbf{i}y$ with $x \in \mathbb{R} \setminus \mathcal{S}_n$; then a direct computation yields

$$\frac{\operatorname{Im}(\tilde{t}_n(z))}{\operatorname{Im}(z)} = \int \frac{\widetilde{F}_n(d\lambda)}{|\lambda - z|^2} \xrightarrow[y \searrow 0]{} \int \frac{\widetilde{F}_n(d\lambda)}{|\lambda - x|^2} \ne 0.$$

Therefore, by continuity $(z) \mapsto 1 - z^2 \tilde{t}_n^2(z) \frac{1}{n} \operatorname{Tr} R_n^2 T_n^2(z)$ does not vanish on $\mathbb{C} \setminus \mathcal{S}_n$ and $\mathcal{B}_{1,n}$ is analytic on this set. We can similarly prove that $\mathcal{B}_{2,n}$ is also analytic on the same set. Consider now a function $f \in C_c^{18}(\mathbb{R})$ whose support is disjoint from $\mathcal{S}_n$, then it is straightforward to check that $Z_n^2(f) = 0$ and the proof of the proposition is completed.

APPENDIX: REMAINING PROOFS

A.1. Proof of Lemma 6.3. By Proposition 6.2,

$$\begin{aligned}\mathbb{E}\left|\int_D \overline{\partial}\Phi(g)(z)\varphi_n(z)\ell_2(dz)\right| &\leq \int_D |\overline{\partial}\Phi(g)(z)|\mathbb{E}|\varphi_n(z)|\ell_2(dz)\\ &\leq \|g\|_{k+1,\infty}\int_D \operatorname{Im}(z)^k\{\operatorname{var}\varphi_n(z)\}^{1/2}\ell_2(dz)\\ &< \infty,\end{aligned}$$

by (iii) and (iv). Hence, $\frac{1}{\pi}\operatorname{Re}\int_D \overline{\partial}\Phi(g)(z)\varphi_n(z)\ell_2(dz)$ is a well-defined a.s. finite random variable. This estimate, uniform in n, readily implies the tightness of

$$\left(\frac{1}{\pi}\operatorname{Re}\int_D \overline{\partial}\Phi(g)(z)\varphi_n(z)\ell_2(dz); n \in \mathbb{N}\right).$$

Notice that the integrals with ψ_n instead of φ_n are similarly well defined and tight.

By conditions (i) and (ii), we obtain

$$\begin{aligned}&d_{\mathcal{LP}}\left(\frac{1}{\pi}\operatorname{Re}\int_{D_\varepsilon}\overline{\partial}\Phi(\mathbf{g})(z)\varphi_n(z)\ell_2(dz), \frac{1}{\pi}\operatorname{Re}\int_{D_\varepsilon}\overline{\partial}\Phi(\mathbf{g})(z)\psi_n(z)\ell_2(dz)\right)\\ &\quad\xrightarrow[N,n\to\infty]{} 0\end{aligned} \tag{A.1}$$

(apply Lemma 5.7).

Let $\mathbf{g} = (g_\ell; 1 \leq \ell \leq L)$ and $f : \mathbb{C}^L \to \mathbb{C}$ be bounded and continuous. Consider the following notation:

$$\xi_n = \frac{1}{\pi}\operatorname{Re}\int_D \overline{\partial}\Phi(\mathbf{g})(z)\varphi_n(z)\ell_2(dz), \qquad \xi_n^\varepsilon = \frac{1}{\pi}\operatorname{Re}\int_{D_\varepsilon}\overline{\partial}\Phi(\mathbf{g})(z)\varphi_n(z)\ell_2(dz),$$

$$\eta_n = \frac{1}{\pi}\operatorname{Re}\int_D \overline{\partial}\Phi(\mathbf{g})(z)\psi_n(z)\ell_2(dz), \qquad \eta_n^\varepsilon = \frac{1}{\pi}\operatorname{Re}\int_{D_\varepsilon}\overline{\partial}\Phi(\mathbf{g})(z)\psi_n(z)\ell_2(dz).$$

We have

$$\begin{aligned}&|\mathbb{E}f(\xi_n) - \mathbb{E}f(\eta_n)|\\ &\quad\leq |\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)| + |\mathbb{E}f(\xi_n^\varepsilon) - \mathbb{E}f(\eta_n^\varepsilon)| + |\mathbb{E}f(\eta_n^\varepsilon) - \mathbb{E}f(\eta_n)|.\end{aligned} \tag{A.2}$$

Given $\rho > 0$, we first prove that for all $n \geq 1$,

$$|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)| \leq (4\|f\|_\infty + 1)\rho \tag{A.3}$$

for ε small enough.

We have

$$\mathbb{P}\{|\xi_n - \xi_n^\varepsilon| > \delta\} \leq \frac{1}{\delta}\left(\int_{[0,A]+\mathbf{i}[0,\varepsilon]} |\overline{\partial}\Phi(\mathbf{g})(z)|\mathbb{E}|\varphi_n(z)|\ell_2(dz)\right) \tag{A.4}$$

which can be made arbitrarily small if ε is small enough, independently from n. Now,

$$\begin{aligned}|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)| \leq{}& |\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)|1_{\{|\xi_n - \xi_n^\varepsilon| > \eta\}} \\ &+ |\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)|1_{\{|\xi_n - \xi_n^\varepsilon| \leq \eta, |\xi_n| \vee |\xi_n^\varepsilon| > K\}} \\ &+ |\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi_n^\varepsilon)|1_{\{|\xi_n - \xi_n^\varepsilon| \leq \eta, |\xi_n| \vee |\xi_n^\varepsilon| \leq K\}}.\end{aligned}$$

First, invoke the tightness of $|\xi_n| \vee |\xi_n^\varepsilon|$ and choose K large enough so that the second term of the RHS is lower than $2\|f\|_\infty\rho$; then choose $\eta > 0$ small enough so that f being absolutely continuous over $\{z \in \mathbb{C}^+, |z| \leq K\}$, the third term of the RHS is lower that ρ; finally for such K and η, take advantage of (A.4) and choose ε small enough so that the first term of the RHS is lower than $2\|f\|_\infty\rho$. Equation (A.3) is proved.

One can similarly prove that $|\mathbb{E}f(\eta_n) - \mathbb{E}f(\eta_n^\varepsilon)| \leq (4\|f\|_\infty + 1)\rho$ for $\varepsilon > 0$ small enough. Such ε being fixed, it remains to control the second term of the RHS of (A.2), but this immediately follows from (A.1).

In order to prove that η_n is multivariate Gaussian with prescribed covariance (6.2), we first consider η_n^ε. Approximating the integral in η_n^ε by Riemann sums and using the fact that weak limits of Gaussian vectors are Gaussian immediately yields that η_n^ε is a Gaussian vector with covariance matrix

$$\begin{aligned}&[\mathrm{cov}(\eta_n^\varepsilon)]_{k\ell} \\ &\quad = \frac{1}{\pi^2}\mathbb{E}\left\{\mathrm{Re}\int_{D^\varepsilon} \overline{\partial}\Phi(g_k)(z)\psi_n(z)\ell_2(dz)\,\mathrm{Re}\int_{D^\varepsilon} \overline{\partial}\Phi(g_\ell)(z)\psi_n(z)\ell_2(dz)\right\}\end{aligned}$$

for $1 \leq k, \ell \leq L$. Using the elementary identity

$$\mathrm{Re}(z)\,\mathrm{Re}(z') = \frac{\mathrm{Re}(z\overline{z'}) + \mathrm{Re}(zz')}{2},$$

we obtain

$$\begin{aligned}&[\mathrm{cov}(\eta_n^\varepsilon)]_{k\ell} \\ &\quad = \frac{1}{2\pi^2}\,\mathrm{Re}\int_{(D^\varepsilon)^2} \overline{\partial}\Phi(g_k)(z_1)\overline{\overline{\partial}\Phi(g_\ell)(z_2)}\mathbb{E}\psi_n(z_1)\overline{\psi_n(z_2)}\ell_2(dz_1)\ell_2(dz_2) \\ &\qquad + \frac{1}{2\pi^2}\,\mathrm{Re}\int_{(D^\varepsilon)^2} \overline{\partial}\Phi(g_k)(z_1)\overline{\partial}\Phi(g_\ell)(z_2)\mathbb{E}\psi_n(z_1)\psi_n(z_2)\ell_2(dz_1)\ell_2(dz_2).\end{aligned}$$

Using the fact that $\overline{\psi_n(z_2)} = \psi_n(\overline{z_2})$ yields

$$
\begin{aligned}
&[\mathrm{cov}(\eta_n^\varepsilon)]_{k\ell} \\
&\quad = \frac{1}{2\pi^2} \operatorname{Re} \int_{(D^\varepsilon)^2} \overline{\partial}\Phi(g_k)(z_1) \overline{\overline{\partial}\Phi(g_\ell)(z_2)} \kappa_n(z_1, \overline{z_2}) \ell_2(dz_1)\ell_2(dz_2) \\
&\qquad + \frac{1}{2\pi^2} \operatorname{Re} \int_{(D^\varepsilon)^2} \overline{\partial}\Phi(g_k)(z_1) \overline{\partial}\Phi(g_\ell)(z_2) \kappa_n(z_1, z_2) \ell_2(dz_1)\ell_2(dz_2).
\end{aligned}
$$

In order to lift the Gaussianity from η_n^ε to η_n and to extend the covariance formula from the one above to formula (6.2), we rely on the approximation theorem [36], Theorem 4.28, and on assumptions (iv) and (v) on the variance estimates and on the regularity of functions g_k, g_ℓ in Lemma 6.3.

The proof of Lemma 6.3 is complete.

A.2. Proof of Proposition 6.4(ii). We rely on a meta-model argument (cf. Section 2.6). Denote by

$$M_{n,M}^1(z) = \mathrm{Tr}(\Sigma_n(M)\Sigma_n^*(M) - zI_N)^{-1} - \mathbb{E}\,\mathrm{Tr}(\Sigma_n(M)\Sigma_n^*(M) - zI_N)^{-1},$$

then by Proposition 6.4(i), we get

$$\mathrm{var}\{\mathrm{tr}(\Sigma_n(M)\Sigma_n^*(M) - zI_N)^{-1}\} \le \frac{C}{\mathrm{Im}(z)^4},$$

moreover $M_{n,M}^1(z)$ converges in distribution to $\psi_n(z)$ as $M \to \infty$, N and n being fixed (see, e.g., the details in Section 5.2). Consider the continuous bounded function $h_K(x) = |x|^2 \wedge K$, then

$$\mathbb{E} h_K(\psi_n(z)) = \lim_{M\to\infty} \mathbb{E} h_K(M_{n,M}^1(z)) \le \limsup_{M\to\infty} \mathbb{E}|M_{n,M}^1(z)|^2 \le \frac{C}{\mathrm{Im}(z)^4}.$$

Now letting $K \to \infty$ yields the desired bound by monotone convergence theorem.

Acknowledgements. We are particularly indebted to Charles Bordenave who drew our attention to Helffer–Sjöstrand's formula and related variance estimates, which substantially shorten the initial proof of fluctuations for nonanalytic functions; we thank Reinhold Meise for his help to understand Tillmann's article; we also thank Djalil Chafaï, Walid Hachem and Philippe Loubaton for fruitful discussions; finally, we thank the referee whose remarks have substantially simplified the proof of Proposition 5.4.

REFERENCES

[1] Albeverio, S., Pastur, L. and Shcherbina, M. (2001). On the $1/n$ expansion for some unitary invariant ensembles of random matrices. *Comm. Math. Phys.* **224** 271–305. MR1869000

[2] Anderson, G. W. and Zeitouni, O. (2006). A CLT for a band matrix model. *Probab. Theory Related Fields* **134** 283–338. MR2222385

[3] Arharov, L. V. (1971). Limit theorems for the characteristic roots of a sample covariance matrix. *Dokl. Akad. Nauk SSSR* **199** 994–997. MR0309171

[4] Bai, Z., Chen, Y. and Liang, Y.-C., eds. (2009). *Random Matrix Theory and Its Applications: Multivariate Statistics and Wireless Communications. Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore* **18**. World Scientific, Hackensack, NJ. MR2603191

[5] Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[6] Bai, Z., Wang, X. and Zhou, W. (2010). Functional CLT for sample covariance matrices. *Bernoulli* **16** 1086–1113. MR2759170

[7] Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

[8] Bao, Z., Pan, G. and Zhou, W. (2013). Central limit theorem for partial linear eigenvalue statistics of Wigner matrices. *J. Stat. Phys.* **150** 88–129. MR3018879

[9] Bao, Z., Pan, G. and Zhou, W. (2013). On the MIMO channel capacity for the general channels. Preprint.

[10] Benaych-Georges, F., Guionnet, A. and Male, C. (2014). Central limit theorems for linear statistics of heavy tailed random matrices. *Comm. Math. Phys.* **329** 641–686. MR3210147

[11] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

[12] Bordenave, C. (2013). Personal communication.

[13] Bordenave, C. (2013). A short course on random matrices (preliminary draft). Available at http://www.math.univ-toulouse.fr/~bordenave/coursRMT.pdf.

[14] Cabanal-Duvillard, T. (2001). Fluctuations de la loi empirique de grandes matrices aléatoires. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 373–402. MR1831988

[15] Capitaine, M. and Donati-Martin, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. MR2317545

[16] Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121

[17] Couillet, R. and Debbah, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge. MR2884783

[18] Dozier, R. B. and Silverstein, J. W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123

[19] Dudley, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358

[20] Dyn'kin, E. M. (1972). An operator calculus based on the Cauchy–Green formula. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **30** 33–39. MR0328640

[21] Girko, V. L. (2001). *Theory of Stochastic Canonical Equations. Vol. I. Mathematics and Its Applications* **535**. Kluwer Academic, Dordrecht. MR1887675

[22] Girko, V. L. (2001). *Theory of Stochastic Canonical Equations. Vol. II. Mathematics and Its Applications* **535**. Kluwer Academic, Dordrecht. MR1887676

[23] Guédon, O., Lytova, A., Pajor, A. and Pastur, L. (2014). The central limit theorem for linear eigenvalue statistics of the sum of independent random matrices of rank one. In *Spectral Theory and Differential Equations. Amer. Math. Soc. Transl. Ser. 2* **233** 145–164. Amer. Math. Soc., Providence, RI. MR3307778

[24] Guionnet, A. (2002). Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 341–384. MR1899457

[25] Haagerup, U. and Thorbjørnsen, S. (2005). A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)* **162** 711–775. MR2183281

[26] Haagerup, U. and Thorbjørnsen, S. (2012). Asymptotic expansions for the Gaussian unitary ensemble. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **15** 1250003, 41. MR2922846

[27] Hachem, W., Kharouf, M., Najim, J. and Silverstein, J. W. (2012). A CLT for information-theoretic statistics of non-centered Gram random matrices. *Random Matrices Theory Appl.* **1** 1150010, 50. MR2934716

[28] Hachem, W., Loubaton, P. and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235

[29] Hachem, W., Loubaton, P. and Najim, J. (2008). A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.* **18** 2071–2130. MR2473651

[30] Hachem, W., Loubaton, P., Najim, J. and Vallet, P. (2013). On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 36–63. MR3060147

[31] Helffer, B. (2013). *Spectral Theory and Its Applications. Cambridge Studies in Advanced Mathematics* **139**. Cambridge Univ. Press, Cambridge. MR3027462

[32] Helffer, B. and Sjöstrand, J. (1989). Équation de Schrödinger avec champ magnétique et équation de Harper. In *Schrödinger Operators (Sønderborg, 1988). Lecture Notes in Physics* **345** 118–197. Springer, Berlin. MR1037319

[33] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. MR1487983

[34] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[35] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38. MR0650926

[36] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[37] Kammoun, A., Kharouf, M., Hachem, W. and Najim, J. (2009). A central limit theorem for the SINR at the LMMSE estimator output for large-dimensional signals. *IEEE Trans. Inform. Theory* **55** 5048–5063. MR2596959

[38] Khorunzhy, A. M., Khoruzhenko, B. A. and Pastur, L. A. (1996). Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37** 5033–5060. MR1411619

[39] Laloux, L., Cizeau, P., Bouchaud, J.-P. and Potters, M. (1999). Noise dressing of financial correlation matrices. *Phys. Rev. Lett.* **83** 1467.

[40] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434

[41] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72** 507–536. MR0208649

[42] Münnix, M. C., Schäfer, R. and Guhr, T. (2014). A random matrix approach to credit risk. *PLoS ONE* **9** e98030.

[43] O'Rourke, S., Renfrew, D. and Soshnikov, A. (2013). On fluctuations of matrix entries of regular functions of Wigner matrices with non-identically distributed entries. *J. Theoret. Probab.* **26** 750–780. MR3090549

[44] Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 655–677. MR3189088

[45] Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244

[46] Pastur, L. and Shcherbina, M. (2011). *Eigenvalue Distribution of Large Random Matrices*. *Mathematical Surveys and Monographs* **171**. Amer. Math. Soc., Providence, RI. MR2808038

[47] Pizzo, A., Renfrew, D. and Soshnikov, A. (2012). Fluctuations of matrix entries of regular functions of Wigner matrices. *J. Stat. Phys.* **146** 550–591. MR2880032

[48] Schultz, H. (2005). Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* **131** 261–309. MR2117954

[49] Shcherbina, M. (2011). Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Zh. Mat. Fiz. Anal. Geom.* **7** 176–192, 197, 199. MR2829615

[50] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

[51] Silverstein, J. W. and Choi, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541

[52] Sinai, Ya. and Soshnikov, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)* **29** 1–24. MR1620151

[53] Soshnikov, A. (2000). The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.* **28** 1353–1370. MR1797877

[54] Tillmann, H.-G. (1953). Randverteilungen analytischer Funktionen und Distributionen. *Math. Z.* **59** 61–83. MR0057345

[55] Vallet, P., Loubaton, P. and Mestre, X. (2012). Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Trans. Inform. Theory* **58** 1043–1068. MR2918009

[56] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

[57] Yao, J. (2013). Estimation et fluctuations de fonctionnelles de grandes matrices aléatoires. Ph.D. thesis, Télécom ParisTech.

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

Télécom Paristech
46 rue Barrault
75634 Paris cedex
France
E-mail: jianfeng.yao.sh@gmail.com