Asymptotic freeness for rectangular random matrices and large deviations for sample covariance matrices with sub-Gaussian tails

Benjamin Groux[*]

Abstract

We establish a large deviation principle for the empirical spectral measure of a sample covariance matrix with sub-Gaussian entries, which extends Bordenave and Caputo's result for Wigner matrices having the same type of entries [7]. To this aim, we need to establish an asymptotic freeness result for rectangular free convolution, more precisely, we give a bound in the subordination formula for information-plus-noise matrices.

1 Introduction

Throughout this paper, $\mathcal{P}(E)$ will denote the set of probability measures on a space E, $\mathcal{M}_{n,p}(\mathbb{R})$ (resp. $\mathcal{M}_{n,p}(\mathbb{C})$) the set of $n \times p$ real (resp. complex) matrices, $\mathcal{H}_n(\mathbb{C})$ the set of $n \times n$ Hermitian matrices, A^t (resp. A^*) the transpose (resp. transconjugate) of a matrix A, and $\mathrm{Tr}(A)$ its trace. Besides, for a random variable X, $\mathring{X}$ denotes the centred variable $X - \mathbb{E}(X)$. Finally, for two real numbers x, y, we denote by $x \wedge y$ the minimum of x and y.

1.1 Large deviation results in random matrix theory

Let us first recall some basic facts in random matrix theory (RMT). A central object in RMT is the *empirical spectral measure* of a matrix $A \in \mathcal{H}_n(\mathbb{C})$, namely the probability measure on $\mathbb{R}$ defined by

---
[*]Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 45 avenue des États-Unis, 78035 Versailles Cedex, France. E-mail: benjamin.groux@uvsq.fr

$$\mu_A = \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(A)} \,,$$

where $\lambda_1(A), \ldots, \lambda_n(A)$ denote the eigenvalues of A.

It is well known (cf. [21]) that if X is a *Wigner matrix*, i.e. $X \in \mathcal{H}_n(\mathbb{C})$ and the families of centred independent and identically distributed (i.i.d.) random variables $(X_{j,j})_{1 \leq j \leq n}$, $(X_{j,k})_{1 \leq j < k \leq n}$ are independent, and if the variance $\mathrm{Var}(X_{1,2}) = \mathbb{E}\,|X_{1,2} - \mathbb{E}(X_{1,2})|^2$ equals 1, then almost surely, the spectral measure $\mu_{X/\sqrt{n}}$ converges weakly towards the *semicircular distribution* μ_{sc}, i.e. for any bounded continuous $f : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n \to +\infty} \int_{\mathbb{R}} f \, d\mu_{X/\sqrt{n}} = \int_{\mathbb{R}} f \, d\mu_{\mathrm{sc}} \,.$$

The semicircular distribution μ_{sc} is the probability measure on $\mathbb{R}$ defined by

$$d\mu_{\mathrm{sc}}(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \, \mathbf{1}_{[-2,2]}(x) \, dx \,.$$

In the case of a *sample covariance matrix*, i.e. a matrix XX^* with $X \in \mathcal{M}_{n,p}(\mathbb{C})$ having centred i.i.d. entries, if $\mathrm{Var}(X_{1,1}) = 1$, then almost surely, the spectral measure $\mu_{XX^*/p}$ converges weakly towards the *Marcenko-Pastur distribution* $\mu_{\mathrm{MP},c}$ with ratio c as $n, p \to +\infty$ with $\frac{n}{p} \to c \in (0, +\infty)$ (cf. [18]). This probability measure on $\mathbb{R}$ is defined by

$$d\mu_{\mathrm{MP},c}(x) = \max\left(1 - \frac{1}{c}, 0\right) \delta_0 + \frac{\sqrt{(b_c - x)(x - a_c)}}{2\pi x c} \, \mathbf{1}_{[a_c, b_c]}(x) \, dx$$

with $a_c = (1 - \sqrt{c})^2$ and $b_c = (1 + \sqrt{c})^2$.

For these two models in which the empirical spectral measure converges, we can investigate the speed of convergence and more particularly large deviation principles.

We recall from [9] that a sequence of random variables $(Z_n)_{n \geq 1}$ with values in a topological space $(E, \mathcal{O})$ with σ-Borel field $\mathcal{B}$ satisfies the *large deviation principle* (LDP) with speed v and rate function I in the topology $\mathcal{O}$ if

- $I : E \to [0, +\infty]$ is a lower semi-continuous function, i.e. the level set $\{x \in E \mid I(x) \leq t\}$ is closed for every $t \geq 0$,
- $v : \mathbb{N} \to (0, +\infty)$ admits a limit equal to $+\infty$,
- for all $B \in \mathcal{B}$,

$$-\inf_{x \in \mathrm{Int}(B)} I(x) \leq \liminf_{n \to +\infty} \frac{1}{v(n)} \log \mathbb{P}(Z_n \in B)$$

$$\leq \limsup_{n \to +\infty} \frac{1}{v(n)} \log \mathbb{P}(Z_n \in B) \leq -\inf_{x \in \mathrm{Clo}(B)} I(x)$$

where $\mathrm{Int}(B)$ and $\mathrm{Clo}(B)$ denote resp. the interior and the closure of B.

We also recall that the rate function I is said to be *good* if the level set $\{x \in E \mid I(x) \leq t\}$ is compact for every $t \geq 0$.

In [4], Ben Arous and Guionnet proved that if X is in the GUE, i.e. X is a Wigner matrix and $X_{1,1}$ (resp. $X_{1,2}$) has law $\mathcal{N}(0,1)$ (resp. $\mathcal{N}_2\left(0, \frac{1}{2}I_2\right)$), then $\mu_{X/\sqrt{n}}$ satisfies a LDP in $\mathcal{P}(\mathbb{R})$ with speed n^2 and rate function

$$I(\mu) = \frac{1}{2} \int x^2 \, d\mu(x) - \iint \log|x - y| \, d\mu(x) d\mu(y) - \frac{3}{4} \,.$$

This result was extended to LUE matrices, i.e. sample covariance matrices XX^* where X has standard Gaussian entries, by Hiai and Petz [16]. Note that in fact, these two LDPs

do not concern only Gaussian matrices but also more general unitarily invariant models. They strongly rely on the fact that for the considered models, the joint distribution of the eigenvalues has an explicit form, which is also the case in [12].

In [7], Bordenave and Caputo managed to obtain a LDP for Wigner matrices in another case, where the distribution of the $X_{j,k}$'s has sub-Gaussian tails. This is remarkable because here the joint distribution of the eigenvalues is unknown. Let us recall their result.

Definition 1.1. *For $\alpha > 0$ and $a \in (0, +\infty]$, we denote by $\mathcal{S}_\alpha(a)$ the class of complex random variables Z such that*

$$\lim_{t \to +\infty} -t^{-\alpha} \log \mathbb{P}(|Z| \geq t) = a \tag{1.1}$$

and such that $|Z|$ and $Z/|Z|$ are independent for large values of $|Z|$, i.e. there exist $t_0 > 0$ and a probability measure ϑ_a on the unit circle $\mathbb{S}^1$ such that for all $t \geq t_0$ and all measurable sets $U \subset \mathbb{S}^1$, we have

$$\mathbb{P}(Z/|Z| \in U \ \cap \ |Z| \geq t) = \vartheta_a(U) \, \mathbb{P}(|Z| \geq t) \,.$$

In particular, a real random variable Z belongs to $\mathcal{S}_\alpha(a)$ if it satisfies (1.1) and there exist $t_0 > 0$ and a probability measure ϑ_a on $\{-1, 1\}$ such that for all $t \geq t_0$ and all $U \subset \{-1, 1\}$, we have

$$\mathbb{P}(|Z| \geq t \ \cap \ \mathrm{sign}(Z) \in U) = \vartheta_a(U) \, \mathbb{P}(|Z| \geq t) \,. \tag{1.2}$$

Note that the first hypothesis implies that a random variable in $\mathcal{S}_\alpha(a)$ has finite moments of all orders. Besides, the second assumption is a technical one, which is used only when proving lower bounds, see (3.16).

Theorem 1.2 (see [7, Theorem 1.1]). *Let X be a Wigner matrix. We assume that $\mathrm{Var}(X_{1,2}) = 1$, $X_{1,2} \in \mathcal{S}_\alpha(a)$, and $X_{1,1} \in \mathcal{S}_\alpha(b)$ for some $\alpha \in (0, 2)$ and $a, b \in (0, +\infty]$. Then the spectral measure $\mu_{X/\sqrt{n}}$ satisfies the LDP with speed $n^{1+\alpha/2}$ and good rate function*

$$J(\mu) = \begin{cases} \Phi(\nu) & \text{if there exists } \nu \in \mathcal{P}(\mathbb{R}) \text{ such that } \mu = \mu_{\mathrm{sc}} \boxplus \nu \\ +\infty & \text{otherwise} \end{cases}$$

where $\Phi : \mathcal{P}(\mathbb{R}) \to [0, +\infty]$ is a good rate function (see [7] for further details) and $\boxplus$ denotes free convolution (see Section 1.2).

Let us make a few remarks about this result. Roughly speaking, after random matrix considerations, the proof of Theorem 1.2 consists in proving a LDP for some sparse random graphs associated to the Wigner matrix X. Therefore, the rate function Φ expresses as the supremum of functions of probability measures on graphs and it can not be computed in general. However, in some particular cases, it is possible to compute $\Phi(\nu)$. For example, if ν is a symmetric distribution on $\mathbb{R}$, $b < \infty$ and the support of ϑ_b is $\{-1, 1\}$, then we have

$$\Phi(\nu) = \left(\frac{a}{2} \wedge b\right) m_\alpha(\nu) \,,$$

where $m_\alpha(\nu)$ denotes the α-th moment of ν.

Theorem 1.7 below will extend Theorem 1.2 to sample covariance matrices XX^* with $X_{1,1} \in \mathcal{S}_\alpha(a)$ for some $\alpha \in (0, 2)$, $a \in (0, +\infty]$. Note that to simplify, we will assume that X is a <u>real</u> random matrix.

Let us mention here that LDPs for the top eigenvalue of Wigner matrices have also been obtained in Ben Arous and Guionnet's setting, see [1, p. 81], and for the model introduced by Bordenave and Caputo in [2]. Furthermore, LDPs for the extreme eigenvalues of sample covariance matrices have been proved in a general setting in [13].

1.2   Deformed matrix models

After understanding the behaviour of the spectral measure of Wigner matrices or sample covariance matrices, the question of deformations of these models has been investigated. Several types of deformations have been studied, the main ones being matrices of the type $X + A$ with $A \in \mathcal{H}_n(\mathbb{C})$ (additive deformation), $\Sigma^{1/2} X X^* \Sigma^{1/2}$ with $\Sigma \in \mathcal{H}_n(\mathbb{C})$ positive definite (multiplicative deformation) or $(X + A)(X + A)^*$ with $A \in \mathcal{M}_{n,p}(\mathbb{C})$ (information-plus-noise model).

A tool to study the spectral measure of a deformation is free probability, and more particularly free convolutions. Let us recall their definitions.

Theorem 1.3 (see [17, Proposition 4.3.9]). *Let A, B be two independent $n \times n$ Hermitian random matrices such that*

- *either A or B is unitarily invariant, i.e. for $M = A$ or B, for any unitary $U \in \mathcal{M}_n(\mathbb{C})$, UMU^* has the same law as M,*

- *μ_A and μ_B converge weakly in probability to some distributions μ_1 and μ_2 on $\mathbb{R}$ as $n \to +\infty$.*

Then, as $n \to +\infty$, the spectral measure μ_{A+B} converges weakly in probability to a deterministic distribution depending only on μ_1 and μ_2. This distribution is called the free (additive) convolution *of μ_1 and μ_2, and is denoted by $\mu_1 \boxplus \mu_2$.*

A similar result also exists for the singular values of the sum of two rectangular matrices and it is due to Benaych-Georges. The *empirical singular value distribution* of a matrix $A \in \mathcal{M}_{n,p}(\mathbb{C})$ is the probability measure on $\mathbb{R}_+$ defined by

$$\nu_A = \frac{1}{n \wedge p} \sum_{k=1}^{n \wedge p} \delta_{\sigma_k(A)} \,,$$

where $\sigma_1(A), \ldots, \sigma_{n \wedge p}(A)$ denote the singular values of A, i.e. the square roots of the eigenvalues of the positive matrix AA^* (resp. A^*A) if $n \leq p$ (resp. $n \geq p$).

Theorem 1.4 (see [5, Theorem 3.13]). *Let A, B be two independent $n \times p$ random matrices such that*

- *either A or B is bi-unitarily invariant, i.e. for $M = A$ or B, for any unitary matrices $U \in \mathcal{M}_n(\mathbb{C})$ and $V \in \mathcal{M}_p(\mathbb{C})$, UMV has the same law as M,*

- *ν_A and ν_B converge weakly in probability to some distributions μ_1 and μ_2 on $\mathbb{R}_+$ as $n, p \to +\infty$ with $\frac{n}{p} \to c \in [0, +\infty)$.*

Then, as $n \to +\infty$, the singular value distribution ν_{A+B} converges weakly in probability to a deterministic distribution depending only on μ_1, μ_2, and c. This distribution is called the rectangular free convolution *with ratio c of μ_1 and μ_2, and is denoted by $\mu_1 \boxplus_c \mu_2$.*

Free convolutions can be characterized in terms of another key object in RMT, namely Stieltjes transform. For a probability measure μ on $\mathbb{R}$, we call the *Stieltjes transform* of μ the function $G_\mu : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C}$ defined by

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - x} \, d\mu(x)$$

for all $z \in \mathbb{C} \setminus \mathbb{R}$. The following properties are obvious:

$$|G_\mu(z)| \leq \frac{1}{|\operatorname{Im} z|}$$

and

$$|G_\mu(z) - G_\mu(z')| \leq \frac{|z - z'|}{|\operatorname{Im} z|.|\operatorname{Im} z'|} .$$

We will use them implicitly in this paper.

Note that the notion of Stieltjes transform is related to the resolvent one, since for a matrix $A \in \mathcal{H}_n(\mathbb{C})$, we have $G_{\mu_A}(z) = \frac{1}{n} \operatorname{Tr}((zI_n - A)^{-1})$. Useful properties of resolvents we will use in this paper are gathered in Appendix B.2.

Stieltjes transform allows to express subordination relations for free convolutions. To state these relations, we need some additional notations. For $\mu \in \mathcal{P}(\mathbb{R})$, we denote by μ^2 the distribution of X^2 when X has law μ. Similarly, for $\mu \in \mathcal{P}(\mathbb{R}_+)$, we denote by $\sqrt{\mu}$ the symmetrization of the distribution ν of $\sqrt{X}$ when X has law μ, i.e. the symmetric distribution on $\mathbb{R}$ defined by $\sqrt{\mu}(B) = \frac{\nu(B)+\nu(-B)}{2}$ for all Borel set B. We have the following subordination formulas, the first one is due to Biane (cf. [6]) and the second one is obtained from Dozier and Silverstein's work [10] and from the paper [5] by Benaych-Georges.

Proposition 1.5. • *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $\nu = \mu \boxplus \mu_{sc}$. We have*

$$G_\nu(z) = G_\mu\left(z - G_\nu(z)\right) . \tag{1.3}$$

• *Let $\mu \in \mathcal{P}(\mathbb{R}_+)$, $c \geq 0$, and $\nu = \left(\sqrt{\mu} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$. We have*

$$\frac{G_\nu(z)}{1 - cG_\nu(z)} = G_\mu\left(z(1 - cG_\nu(z))^2 - (1 - c)(1 - cG_\nu(z))\right) . \tag{1.4}$$

In Theorem 1.6 below, we are interested in the information-plus-noise model and we control the distance between the spectral measure and the corresponding rectangular free convolution, by bounding the difference between the two sides in (1.4) evaluated at the average Stieltjes transform.

1.3 Main results

Note that in the rest of the paper, we will only consider <u>real</u> matrices for ease but our results may generalize to complex matrices adapting the proofs. The main difficulty in the complex case is to adapt the general integration by parts formula (2.21) which is used several times in this paper, which would lead to heavier computations.

Let us define, for $s, t > 0$, the distance $d_{s,t}$ on $\mathcal{P}(\mathbb{R})$ by

$$d_{s,t}(\mu, \nu) = \sup_{z \in V_{s,t}} |G_\mu(z) - G_\nu(z)| , \tag{1.5}$$

where

$$V_{s,t} = \left\{ z \in \mathbb{C} \mid \operatorname{Im} z > s, \left|\frac{\operatorname{Re} z}{\operatorname{Im} z}\right| < t \right\} . \tag{1.6}$$

As the distance d defined in [7], $d_{s,t}$ metrizes weak convergence. Let us mention that for all $\mu, \nu \in \mathcal{P}(\mathbb{R})$, we have

$$d_{s,t}(\mu, \nu) \leq \min\left(d_{\mathrm{KS}}(\mu, \nu), W_1(\mu, \nu)\right) , \tag{1.7}$$

where d_{KS} and W_1 are respectively the Kolmogorov-Smirnov and L^1-Wasserstein distances on $\mathcal{P}(\mathbb{R})$. Some key inequalities for the distance between two empirical spectral measures are summarized in Appendix B.3.

Our first main result is the following.

Theorem 1.6. *We assume that $c_n = \frac{n}{p}$ is bounded below and above by two constants in $(0, +\infty)$. Let $c \geq 0$. There exist $s, t > 0$ and a constant $c_{s,t} > 0$ such that for any random matrix $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ with i.i.d. entries satisfying $\mathrm{Var}(Y_{1,1}) = 1$ and $\mathbb{E}(Y_{1,1}^4) < +\infty$, for any deterministic matrix $M \in \mathcal{M}_{n,p}(\mathbb{R})$, and for all n large enough, we have*

$$
d_{s,t}\left(\mathbb{E}\,\mu_{(Y/\sqrt{p}+M)(Y/\sqrt{p}+M)^t}, \left(\sqrt{\mu_{MM^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2 \right)
$$
$$
\leq \; c_{s,t}\left(\mathbb{E}\,|\mathring{Y}_{1,1}|^3 + \mathbb{E}(\mathring{Y}_{1,1}^{\,4}) \right)\left(\frac{1}{\sqrt{n}} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n} \right)
$$
$$
+ \; c_{s,t}\left(|c_n - c| + \frac{1}{n} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n^{5/4}} \right),
$$

where $\mathring{Y}$ is the matrix whose entries are given by $\mathring{Y}_{j,k} = Y_{j,k} - \mathbb{E}(Y_{j,k})$.

This result allows to understand the influence of the deformation in the information-plus-noise model. First, we can observe a decorrelation between the classical term $\frac{1}{\sqrt{n}}$ and the Frobenius norm of the deformation divided by a better power of n, namely $\frac{\mathrm{Tr}(MM^t)^{1/2}}{n}$. It is important for us to get this precise estimate since in Section 3, we apply Theorem 1.6 to a matrix M whose Frobenius norm is not bounded but of order $\sqrt{n}\log n$.

Besides, it is interesting to compare Theorem 1.6 to the Wigner case (cf. [7, Theorem 2.6]). Bordenave and Caputo investigated additive deformations and obtained that in this model, the distance between the spectral measure and the corresponding free additive convolution is bounded by $\frac{1}{\sqrt{n}}$. This bound is uniform in the deformation M and it depends on the initial matrix through its moments only. In the case of sample covariance matrices, it would have been surprising if we had obtained a better bound. Table 1 below permits to compare Bordenave and Caputo's results with ours in the Gaussian and the general cases. We recall that LOE matrices are sample covariance matrices XX^t where X has standard real Gaussian entries.

In addition to this, let us mention that in [8], the authors were interested in the case of Wigner matrices whose entries have a symmetric distribution satisfying a Poincaré inequality, which leads to better bounds than [7].

	Gaussian	Non-Gaussian
Wigner matrix	Deformed GUE matrix $\dfrac{1}{n}$	Deformed Wigner matrix $\dfrac{1}{\sqrt{n}}$
Covariance matrix	Deformed LOE matrix $\dfrac{1}{n} + \dfrac{\mathrm{Tr}(MM^t)^{1/2}}{n^{5/4}}$	Info-plus-noise matrix $\dfrac{1}{\sqrt{n}} + \dfrac{\mathrm{Tr}(MM^t)^{1/2}}{n}$

Table 1: Bound in the subordination relation (1.3) or (1.4) for different matrix models.

Theorem 1.6 above will be used in the proof of our second main result.

Theorem 1.7. *Let $X \in \mathcal{M}_{n,p}(\mathbb{R})$ be a random matrix such that $c_n = \frac{n}{p} \to c \in (0, +\infty)$. We assume that $\mathrm{Var}(X_{1,1}) = 1$ and that there exist $\alpha \in (0, 2)$ and $a \in (0, +\infty]$ such that $X_{1,1} \in \mathcal{S}_\alpha(a)$.*
Then, the empirical spectral measure $\mu_{XX^t/p}$ satisfies the LDP with speed $n^{1+\alpha/2}$ in $\mathcal{P}(\mathbb{R}_+)$, governed by the good rate function J' defined by

$$
J'(\mu) = \begin{cases} \dfrac{a}{c^{\alpha/2}} m_{\alpha/2}(\nu) & \text{if there exists } \nu \in \mathcal{P}(\mathbb{R}_+) \text{ s.t. } \mu = \left(\sqrt{\nu} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2 \\ & \qquad \text{and } \nu(\{0\}) \geq \max\left(0, 1 - \frac{1}{c}\right) \\ +\infty & \text{otherwise} \end{cases}
$$

where $m_p(\mu) = \int_{\mathbb{R}} |x|^p \, d\mu(x)$ denotes the p-th moment of a distribution μ.

It is very similar to Bordenave and Caputo's result (see Theorem 1.2), the main difference being the explicit expression of the rate function in all cases. This is due to the fact that here, we can achieve large deviations explicitly because all entries of our matrix X have the same distribution, which is not the case for the Wigner matrices considered in [7]. Consequently, a large deviation $\mu_{XX^t/p} \sim \left(\sqrt{\nu} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ can be reached explicitly by considering the event $\mu_{\mathrm{diag}(X)\,\mathrm{diag}(X)^t/p} \sim \nu$, where $\mathrm{diag}(X)$ is the matrix X from which we have removed off-diagonal entries, and large deviations can be proved directly using Schatten's inequality, see Section 3.2. In the Wigner case, this is not possible to do so because, roughly speaking, there are two ways to achieve large deviations, using diagonal entries or off-diagonal entries.

Note also that contrary to J, the rate function J' does not depend on the support of the law ϑ_a appearing in Definition 1.1.

Finally, we mention that our proof does not use sparse random graphs as in Bordenave and Caputo's original proof [7] but it would be possible to do so (see [14, Section 4.6]).

The rest of the paper is organized as follows. In Section 2, we prove the bound for rectangular free convolution stated in Theorem 1.6. In Section 3, we prove the large deviation principle in Theorem 1.7. In Appendix A, we state and prove concentration results used in Sections 2 and 3. Finally, in Appendix B, we summarize miscellaneous inequalities and identities used throughout the paper.

2 Asymptotic freeness

This section is devoted to the proof of Theorem 1.6. This theorem is in fact a consequence of the following, as we will see in Section 2.1.

Theorem 2.1 (Bound in subordination formula (1.4))**.** *We assume that $c_n = \frac{n}{p}$ is bounded below and above by two constants in $(0, +\infty)$. Let $c \geq 0$. There exist $s, t > 0$ and a function f, bounded on the domain $V_{s,t}$ defined by (1.6), such that for any random matrix $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ with i.i.d. entries satisfying $\mathrm{Var}(Y_{1,1}) = 1$ and $\mathbb{E}(Y_{1,1}^4) < +\infty$, for any deterministic matrix $M \in \mathcal{M}_{n,p}(\mathbb{R})$, for all n large enough, and for all $z \in V_{s,t}$, we have*

$$
\begin{aligned}
\Big|\underline{g}(z) - (1 - c\underline{g}(z))G_{\mu_{MM^t}}&\left(z(1 - c\underline{g}(z))^2 - (1-c)(1 - c\underline{g}(z))\right)\Big| \\
&\leq f(z)\left(\mathbb{E}\,|\mathring{Y}_{1,1}|^3 + \mathbb{E}(\mathring{Y}_{1,1}^4)\right)\left(\frac{1}{\sqrt{n}} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n}\right) \\
&\quad + f(z)\left(|c_n - c| + \frac{1}{n} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n^{5/4}}\right),
\end{aligned}
$$

where $g(z) = G_{\mu_{(Y/\sqrt{p}+M)(Y/\sqrt{p}+M)^t}}(z)$ and $\underline{g}(z) = \mathbb{E}(g(z))$.

The proof of Theorem 2.1 follows the same lines as Bordenave and Caputo's one for the bound in subordination formula (1.3) for free additive convolution (see [7, Theorem A.1]). It consists in two main steps: the Gaussian case and the general case, which is deduced from the Gaussian case by interpolation. The starting point is also a resolvent identity, see Proposition B.2 (ii).

However, contrary to the Wigner case, when considering the difference between the resolvents of two covariance matrices AA^* and $(A + B)(A + B)^*$, not only does B appear in the difference but also A. Thus, if we want to show that the difference between the resolvents is small when B is small, we must be more careful because some new phenomena appear.

Let us mention that in the Gaussian case, we have a better bound (see Proposition 2.3).

In the proof, we define

$$
X = \frac{Y}{\sqrt{p}} + M.
$$

We also consider $s > 2$, $t > 0$, and along the proof, s can increase and t can decrease. Moreover, f will denote a bounded function on $V_{s,t}$, which can also change from one line to another. In particular, for all $z \in V_{s,t}$ and $x < y$, since we have

$$\frac{|z|^x}{|\operatorname{Im} z|^y} = \frac{1}{|\operatorname{Im} z|^{y-x}} \left(\left(\frac{\operatorname{Re} z}{\operatorname{Im} z}\right)^2 + 1 \right)^{x/2} < \frac{1}{s^{y-x}}(t^2 + 1)^{x/2},$$

we will write

$$\frac{|z|^x}{|\operatorname{Im} z|^y} \leq f(z)$$

as soon as $x < y$.

Before starting the proofs, let us state a lemma we will use in the different steps. $B_{\mathbb{C}}(z, \delta)$ denotes here the ball with centre $z \in \mathbb{C}$ and radius $\delta > 0$ for the usual distance in $\mathbb{C}$.

Lemma 2.2. *For $\mu \in \mathcal{P}(\mathbb{R})$ and $z \in \mathbb{C}$, we define*

$$\phi_{z,\mu} : (h, \gamma) \mapsto (1 - \gamma h) G_\mu(z(1 - \gamma h)^2 - (1 - \gamma)(1 - \gamma h)).$$

There exist $s, t > 0$ and $l_{s,t}, l'_{s,t} \in (0, 1)$ such that

- *for all $\mu \in \mathcal{P}(\mathbb{R})$, $z \in V_{s,t}$, and $\gamma \geq 0$, the function $\phi_{z,\mu}(., \gamma)$ is Lipschitz on $B_{\mathbb{C}}\left(0, \frac{1}{s}\right)$ with constant $l_{s,t}$,*
- *for all $\mu \in \mathcal{P}(\mathbb{R})$, $z \in V_{s,t}$, and $h \in B_{\mathbb{C}}\left(0, \frac{1}{s}\right)$, the function $\phi_{z,\mu}(h, .)$ is Lipschitz on $[0, +\infty)$ with constant $l'_{s,t}$.*

The proof of this lemma consists in simple computations and is left to the reader. Let us mention however that it relies on the inequality

$$|\operatorname{Im} \eta| \geq \operatorname{Im} z \left(\frac{(\sigma - 1)(\sigma^2 - 2)}{\sigma^2(\sigma + 1)} - \frac{2t(\sigma + 1)}{\sigma^2} \right) - \frac{|1 - \gamma|}{\sigma} \tag{2.1}$$

where $\eta = z(1 - \gamma h)^2 - (1 - \gamma)(1 - \gamma h)$ and $\sigma = \frac{s}{\gamma}$. This inequality is an easy consequence of the development of $\operatorname{Im} \eta$. We will use it once again later.

Furthermore, note that choosing a larger s and a smaller t, $l_{s,t}$ and $l'_{s,t}$ can be as close to 0 as wanted.

2.1 Proof of Theorem 1.6

First, let us deduce Theorem 1.6 from Theorem 2.1.

Proof. We define $\nu = \left(\sqrt{\mu_{MM^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ and we consider the function $\phi_{z,\mu_{MM^t}}$ defined in Lemma 2.2. Subordination formula (1.4) can be rewritten $\phi_{z,\mu_{MM^t}}(G_\nu(z), c) = G_\nu(z)$ for all $z \in \mathbb{C} \setminus \mathbb{R}$. Consequently, using Lemma 2.2, there exist $s, t > 0$ and $l_{s,t} \in (0, 1)$ such that for all $z \in V_{s,t}$,

$$
\begin{aligned}
\left| \underline{g}(z) - G_\nu(z) \right| &\leq \left| \underline{g}(z) - \phi_{z,\mu_{MM^t}}(\underline{g}(z), c) \right| + \left| \phi_{z,\mu_{MM^t}}(\underline{g}(z), c) - \phi_{z,\mu_{MM^t}}(G_\nu(z), c) \right| \\
&\leq \left| \underline{g}(z) - \phi_{z,\mu_{MM^t}}(\underline{g}(z), c) \right| + l_{s,t} \left| \underline{g}(z) - G_\nu(z) \right|
\end{aligned}
$$

thus

$$\left| \underline{g}(z) - G_\nu(z) \right| \leq \frac{1}{1 - l_{s,t}} \left| \underline{g}(z) - \phi_{z,\mu_{MM^t}}(\underline{g}(z), c) \right|.$$

From Theorem 2.1 in which we majorize f by a constant depending on s, t and from the definition (1.5) of $d_{s,t}$, we finally get

$$
d_{s,t}\left(\mathbb{E}\, \mu_{XX^t}, \nu\right) \leq c_{s,t} \left(\mathbb{E}\, |\mathring{Y}_{1,1}|^3 + \mathbb{E}(\mathring{Y}_{1,1}{}^4) \right) \left(\frac{1}{\sqrt{n}} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n} \right)
$$
$$
+ c_{s,t} \left(|c_n - c| + \frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right). \quad \square
$$

2.2 The Gaussian case

In this subsection, we assume that $Y_{1,1}$ is a standard Gaussian. We will prove the following bound.

Proposition 2.3. *We use the same notations as in Theorem 2.1. There exist $s, t > 0$ and a function f, bounded on $V_{s,t}$, such that for any random matrix $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ with i.i.d. standard Gaussian entries, for any deterministic matrix $M \in \mathcal{M}_{n,p}(\mathbb{R})$, for all n large enough, and for all $z \in V_{s,t}$, we have*

$$\left| \underline{g} - (1 - c\underline{g})G_{\mu_{MM^t}}(z(1 - c\underline{g})^2 - (1 - c)(1 - c\underline{g})) \right| \leq f(z)\left(|c_n - c| + \frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right).$$

To prove Proposition 2.3, we will follow and improve some computations by Dumont *et al.*, see [11, Appendix II]. In the proof, we will simply denote $g(z)$ and $\underline{g}(z)$ by g and $\underline{g}$ respectively.

Lemma 2.4 (adaptation from [20, Formula (122)]). *Let $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ be a random matrix with i.i.d. standard Gaussian entries, let $M \in \mathcal{M}_{n,p}(\mathbb{R})$ be a deterministic matrix, and let $z \in \mathbb{C} \setminus \mathbb{R}$. For all integer n, we have*

$$\underline{g} - \frac{1}{n}\operatorname{Tr}(R) = \frac{1}{n}\operatorname{Tr}(\Delta R) - \frac{c_n}{n^2}\operatorname{Tr}(\Delta)\operatorname{Tr}(\mathbb{E}(S)R)$$

$$+ \frac{1}{n}\operatorname{Tr}(\Delta'R) - \frac{c_n}{n^2}\operatorname{Tr}(\Delta')\operatorname{Tr}(\mathbb{E}(S)R) \quad (2.2)$$

where

$$S = (zI_n - XX^t)^{-1},$$

$$R = \left((z(1 - c_n\underline{g}) - 1 + c_n)I_n - \frac{1}{1 - c_n\underline{g}}MM^t \right)^{-1}, \quad (2.3)$$

$$\Delta = \frac{1}{p(1 - c_n\underline{g})}\mathbb{E}\left(\overbrace{\operatorname{Tr}(SXM^t)}^{\circ}\mathring{S} \right) + \frac{c_n z}{1 - c_n\underline{g}}\mathbb{E}(\mathring{g}S)$$

$$+ \frac{c_n}{p(1 - c_n\underline{g})^2}\mathbb{E}\left(\mathring{g}\overbrace{\operatorname{Tr}(SXM^t)}^{\circ} \right)\mathbb{E}(S), \quad (2.4)$$

and

$$\Delta' = \frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(SXM^t S) + \frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(zS^2 - S)$$

$$+ \frac{1}{p^2(1 - c_n\underline{g})^2}\mathbb{E}(\operatorname{Tr}(S^2 XM^t))\mathbb{E}(S). \quad (2.5)$$

In this lemma, we compare $\underline{g}$ to $\frac{1}{n}\operatorname{Tr}(R)$ because, using the notations in Lemma 2.2, we have $\frac{1}{n}\operatorname{Tr}(R) = \phi_{z,\mu_{MM^t}}(\underline{g}, c_n)$, so $\frac{1}{n}\operatorname{Tr}(R)$ is close to $\phi_{z,\mu_{MM^t}}(\underline{g}, c)$ by Lemma 2.2. That is interesting if we have in mind our goal, which is Proposition 2.3.

Note that the proof of Lemma 2.4 follows the same lines as [20, Formula (122)] since it relies on the Gaussian integration by parts formula (2.20) and it uses the differentiation formula (B.1), so we do not give it here.

However, we can observe an important difference between Formula (122) in [20] and Lemma 2.4, namely the terms in Δ'. In fact, the background here is not exactly the same as in [20]. Indeed, Vallet *et al.* consider complex Gaussian entries with independent real and imaginary parts having the same distribution in the matrix Y, whereas we consider

real Gaussian entries. Consequently, some simplifications do not occur any longer and a new term appears. Behind this phenomenon is the quantity $\zeta = K_{1,1} + 2iK_{1,2} - K_{2,2}$, where K denotes the covariance matrix of the Gaussian vector $(\operatorname{Re} Y_{1,1}, \operatorname{Im} Y_{1,1})$. This quantity is equal to 0 in [20] and to 1 here, that is why we have an additional term.

In the next lemma, we bound the different terms appearing in (2.2). For this, we will use the concentration bounds (A.1) and (A.3) for the terms in Δ and standard inequalities on traces and resolvents (see Propositions B.1 and B.2) for the terms in Δ'. Our computations will partially follow those in [20].

Lemma 2.5. *There exist $s, t > 0$ and a function f, bounded on $V_{s,t}$, such that for all Y, M, n, and z as in Proposition 2.3, we have*

$$\left| \underline{g} - \frac{1}{n} \operatorname{Tr} R \right| \le f(z) \left(\frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right) .$$

This lemma shows that $\frac{1}{n} \operatorname{Tr}(R)$ is a deterministic equivalent to the Stieltjes transform $g(z) = \frac{1}{n} \operatorname{Tr}(S)$ as soon as $\frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}}$ tends to 0 as $n \to +\infty$, i.e. when the perturbation M is not too large.

We can compare this result with the bound obtained in [20, Proposition 6]. Two main differences must be highlighted. First, as we mentioned above, the model is not exactly the same. Indeed, we consider real Gaussian entries and not complex Gaussian entries with independent real and imaginary parts, which produces an additional term in Δ'. However, the terms in Δ are present in both cases, so we can compare the bounds for these terms. In [20], the authors assume that $\|M\|$ is uniformly bounded in n and get the bound $\frac{f(z)}{n^2}$. Here, for the terms in Δ, we will get the bound

$$f(z) \left(\frac{1}{n^2} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{17/8}} \right) .$$

Moreover, if we use the bound (A.2) instead of (A.3) in the proof, and if we observe that $\operatorname{Tr}(MM^t)^{1/2} \le \sqrt{n}\|M\|$, then we get the bound $\frac{f(z)}{n^2}(1 + \|M\|)$, which is the same as in [20] when $\|M\|$ is uniformly bounded in n. Consequently, our bound has two advantages: it is slightly better than the bound in [20] and it applies without any assumption on M.

Proof of Lemma 2.5. First of all, let us remark that

$$\left| \frac{1}{1 - c_n \underline{g}} \right| \le f(z)$$

since $|\underline{g}| \le \frac{1}{|\operatorname{Im} z|}$. Besides, we have

$$\|R\| \le f(z) \tag{2.6}$$

because on the one hand, $\frac{R}{1 - c_n \underline{g}}$ is a resolvent evaluated at $\eta = z(1 - c_n \underline{g})^2 - (1 - c_n)(1 - c_n \underline{g})$ so its operator norm is less than $\frac{1}{|\operatorname{Im} \eta|}$, and on the other hand, we have the inequalities $|1 - c_n \underline{g}| \le 1 + \frac{c_n}{|\operatorname{Im} z|}$ and (2.1) (we apply the latter with $\gamma = c_n$ and $h = \underline{g}$).

By Proposition B.1 (ii), it follows that

$$\left| \frac{1}{n} \operatorname{Tr}(\mathbb{E}(S)R) \right| \le \frac{f(z)}{|\operatorname{Im} z|}$$

or just

$$\left| \frac{1}{n} \operatorname{Tr}(\mathbb{E}(S)R) \right| \le f(z) . \tag{2.7}$$

Note that more precise bounds can be obtained, see [20, Appendix E].

We will now see how to majorize the four terms in (2.2).

Let us recall that Δ is defined by

$$\Delta = \frac{1}{p(1-c_n\underline{g})} \mathbb{E}\left(\overbrace{\mathrm{Tr}(SXM^t)}^{\circ}\mathring{S}\right) + \frac{c_n z}{1-c_n\underline{g}} \mathbb{E}(\mathring{g}S) + \frac{c_n}{p(1-c_n\underline{g})^2} \mathbb{E}\left(\mathring{g}\overbrace{\mathrm{Tr}(SXM^t)}^{\circ}\right) \mathbb{E}(S)$$

and observe that $\mathrm{Tr}(SXM^t) = \mathrm{Tr}(X^t SM)$.

We start bounding the term $\left|\frac{c_n}{n^2} \mathrm{Tr}(\Delta)\mathrm{Tr}(\mathbb{E}(S)R)\right|$ in (2.2). First, using the Cauchy-Schwarz inequality and the concentration bounds (A.1), (A.3), we get

$$\left| \frac{1}{n}\mathrm{Tr}\left(\frac{1}{p(1-c_n\underline{g})} \mathbb{E}\left(\overbrace{\mathrm{Tr}(SXM^t)}^{\circ}\mathring{S}\right)\right)\right|$$

$$= \left| \frac{1}{n}\mathrm{Tr}\left(\frac{1}{p(1-c_n\underline{g})} \mathbb{E}\left[(\mathrm{Tr}(SMX^t) - \mathbb{E}\,\mathrm{Tr}(SMX^t))(S - \mathbb{E}\,S)\right]\right)\right|$$

$$= \left| \frac{c_n}{1-c_n\underline{g}} \mathbb{E}\left[\frac{1}{n}(\mathrm{Tr}(X^t SM) - \mathbb{E}\,\mathrm{Tr}(X^t SM))\frac{1}{n}(\mathrm{Tr}(S) - \mathbb{E}\,\mathrm{Tr}(S))\right]\right|$$

$$\leq \frac{c_n}{|1-c_n\underline{g}|} \mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}(X^t SM)\right)^{1/2} \mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}(S)\right)^{1/2}$$

$$\leq \frac{c_n}{|1-c_n\underline{g}|} \left(\frac{9c_n v(z)}{n^{9/4}}\mathrm{Tr}(MM^t)\right)^{1/2}\left(\frac{4c_n u(z)}{n^2}\right)^{1/2}$$

$$\leq \frac{f(z)}{n^{17/8}}\mathrm{Tr}(MM^t)^{1/2}, \tag{2.8}$$

where $u(z)$ and $v(z)$ are defined in Proposition A.1.

Next, using the identity $\underline{g} = \frac{1}{n}\mathrm{Tr}(S)$ and (A.1), we have

$$\left|\frac{1}{n}\mathrm{Tr}\left(\frac{c_n z}{1-c_n\underline{g}}\mathbb{E}(\mathring{g}S)\right)\right| = \left|\frac{c_n z}{1-c_n\underline{g}}\mathrm{Var}(\underline{g})\right| \leq \frac{c_n|z|}{|1-c_n\underline{g}|}\frac{4c_n u(z)}{n^2} \leq \frac{f(z)}{n^2} \tag{2.9}$$

where, for the last inequality, we used the definition of $u(z)$ to get

$$\frac{|z|u(z)}{|1-c_n\underline{g}|} \leq f(z).$$

The same arguments also allow to show that

$$\left|\frac{1}{n}\mathrm{Tr}\left(\frac{c_n}{p(1-c_n\underline{g})^2}\mathbb{E}\left(\mathring{g}\overbrace{\mathrm{Tr}(SXM^t)}^{\circ}\right)\mathbb{E}(S)\right)\right| \leq \frac{f(z)}{n^{17/8}}\mathrm{Tr}(MM^t)^{1/2}. \tag{2.10}$$

Combining inequalities from (2.7) to (2.10) gives

$$\left|\frac{c_n}{n^2}\mathrm{Tr}(\Delta)\mathrm{Tr}(\mathbb{E}(S)R)\right| \leq f(z)\left(\frac{1}{n^2} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n^{17/8}}\right). \tag{2.11}$$

Computations are similar for the term $\frac{1}{n}\mathrm{Tr}(\Delta R)$, using the additional inequalities (2.6) and $\mathrm{Tr}(RR^*)^{1/2} \leq \sqrt{n}\|R\|$ (see Proposition B.1 (iv)). For instance, we have

$$\left|\frac{1}{n}\mathrm{Tr}\,\mathbb{E}\left(\overbrace{\mathrm{Tr}(SXM^t)}^{\circ}\mathring{S}R\right)\right|$$

$$= \left|\frac{1}{n}\mathrm{Tr}\left(\frac{1}{p(1-c_n\underline{g})}\mathbb{E}\left[(\mathrm{Tr}(SXM^t) - \mathbb{E}\,\mathrm{Tr}(SXM^t))(S - \mathbb{E}\,S)R\right]\right)\right|$$

$$\leq \quad \frac{c_n}{|1 - c_n\underline{g}|} \operatorname{Var}\left(\frac{1}{n}\operatorname{Tr}(X^t SM)\right)^{1/2} \operatorname{Var}\left(\frac{1}{n}\operatorname{Tr}(SR)\right)^{1/2}$$

$$\leq \quad \frac{c_n}{|1 - c_n\underline{g}|} \left(\frac{9c_n v(z)}{n^{9/4}}\operatorname{Tr}(MM^t)\right)^{1/2} \left(\frac{4c_n u(z)}{n^{5/2}}\|R\|\operatorname{Tr}(RR^*)^{1/2}\right)^{1/2}$$

$$\leq \quad \frac{c_n}{|1 - c_n\underline{g}|} \left(\frac{9c_n v(z)}{n^{9/4}}\operatorname{Tr}(MM^t)\right)^{1/2} \left(\frac{4c_n u(z)}{n^2}f(z)^2\right)^{1/2}$$

$$\leq \quad \frac{f(z)}{n^{17/8}}\operatorname{Tr}(MM^t)^{1/2}\,.$$

Combining with

$$\left|\frac{1}{n}\operatorname{Tr}\left(\frac{c_n z}{1 - c_n\underline{g}}\mathbb{E}(\mathring{g}SR)\right)\right| \leq \frac{f(z)}{n^2}$$

and

$$\left|\frac{1}{n}\operatorname{Tr}\left(c_n\,\mathbb{E}\left(\mathring{g}\overbrace{\operatorname{Tr}(SXM^t)}^{\circ}\right)\mathbb{E}(S)R\right)\right| \leq \frac{f(z)}{n^{17/8}}\operatorname{Tr}(MM^t)^{1/2}\,,$$

which have a similar proof, we thus have

$$\left|\frac{1}{n}\operatorname{Tr}(\Delta R)\right| \leq f(z)\left(\frac{1}{n^2} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{17/8}}\right)\,. \tag{2.12}$$

We have bounded the terms in Lemma 2.4 in which Δ appears thanks to the concentration bounds proved in Appendix A. We will now consider the terms in which Δ' appears, in other words the terms not present in [20]. To this, we will only use inequalities on traces and resolvents (see Propositions B.1 and B.2). Let us recall that Δ' is defined by

$$\Delta' = \frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(SXM^t S) + \frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(zS^2 - S) + \frac{1}{p^2(1 - c_n\underline{g})^2}\mathbb{E}(\operatorname{Tr}(S^2 XM^t))\mathbb{E}(S)\,.$$

We bound $\operatorname{Tr}(\Delta')$ using inequalities (i)–(iv) in Proposition B.1 and Proposition B.2 (i), (iv). We have

$$\begin{aligned}
&\left|\operatorname{Tr}(SXM^t S)\right| \\
&\leq \quad \operatorname{Tr}(SXX^t S^*)^{1/2}\operatorname{Tr}(M^t SS^* M)^{1/2} \\
&\leq \quad \left(\frac{n}{|\operatorname{Im}\overline{z}|}\left(\frac{|z|}{|\operatorname{Im} z|} + 1\right)\right)^{1/2}\left(\frac{\sqrt{p}}{|\operatorname{Im} z|.|\operatorname{Im}\overline{z}|}\|M\|\operatorname{Tr}(MM^t)^{1/2}\right)^{1/2} \\
&\leq \quad n^{3/4}f(z)\operatorname{Tr}(MM^t)^{1/2} \tag{2.13}
\end{aligned}$$

so

$$\left|\frac{1}{n}\operatorname{Tr}\left(\frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(SXM^t S)\right)\right| \leq \frac{f(z)}{n^{5/4}}\operatorname{Tr}(MM^t)^{1/2}\,. \tag{2.14}$$

In addition, we have

$$\begin{aligned}
\left|\frac{1}{n}\operatorname{Tr}\left(\frac{1}{p(1 - c_n\underline{g})}\mathbb{E}(zS^2 - S)\right)\right| &\leq \quad \frac{1}{np|1 - c_n\underline{g}|}n\left(\frac{|z|}{|\operatorname{Im} z|^2} + \frac{1}{|\operatorname{Im} z|}\right) \\
&\leq \quad \frac{f(z)}{n} \tag{2.15}
\end{aligned}$$

and using (2.13) again, we find

$$\left|\frac{1}{n}\operatorname{Tr}\left(\frac{1}{p^2(1 - c_n\underline{g})^2}\mathbb{E}(\operatorname{Tr}(S^2 XM^t))\mathbb{E}(S)\right)\right| \leq \frac{f(z)}{n^{5/4}}\operatorname{Tr}(MM^t)^{1/2}\,. \tag{2.16}$$

Consequently, the combination of (2.7), (2.14), (2.15), and (2.16) gives

$$\left| \frac{c_n}{n^2} \operatorname{Tr}(\Delta') \operatorname{Tr}(\mathbb{E}(S)R) \right| \leq f(z) \left(\frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right) . \tag{2.17}$$

By very similar calculations, we get

$$\left| \frac{1}{n} \operatorname{Tr}(\Delta' R) \right| \leq f(z) \left(\frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right) . \tag{2.18}$$

Finally, combining relation (2.2) with inequalities (2.11), (2.12), (2.17), and (2.18), we get

$$\left| \underline{g} - \frac{1}{n} \operatorname{Tr}(R) \right| \leq f(z) \left(\frac{1}{n} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n^{5/4}} \right) . \qquad \Box$$

Finally, the Gaussian case (Proposition 2.3) follows from Lemma 2.5 and the second part of Lemma 2.2, since we have

$$\frac{1}{n} \operatorname{Tr}(R) = (1 - c_n \underline{g}) G_{\mu_{MM^t}} \left(z(1 - c_n \underline{g})^2 - (1 - c_n)(1 - c_n \underline{g}) \right).$$

2.3 The general case

We now only assume that $\operatorname{Var}(Y_{1,1}) = 1$ and that $\mathbb{E}(Y_{1,1}^4) < +\infty$. Let $\widehat{Y} \in \mathcal{M}_{n,p}(\mathbb{R})$ be an independent random matrix such that the $\widehat{Y}_{j,k}$'s are i.i.d. standard Gaussians, we define $\widehat{X} = \frac{\widehat{Y}}{\sqrt{p}} + M$.

Furthermore, for all $u \in [0,1]$, inspired by Pastur and Shcherbina [19, Theorem 18.3.1], we define $Y(u) = \sqrt{u}Y + \sqrt{1-u}\widehat{Y}$, $X(u) = \frac{Y(u)}{\sqrt{p}} + M$, and $S(u) = (zI_n - X(u)X(u)^t)^{-1}$.

Since $Y(0) = \widehat{Y}$ and $Y(1) = Y$, this provides an interpolation between the Gaussian case and the general case. The idea is that along this interpolation, the average Stieltjes transform does not change a lot, which will allow us to bring back the general case to the Gaussian case.

Proposition 2.6. *There exist $s, t > 0$ and a function f, bounded on $V_{s,t}$, such that for any random matrix $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ with i.i.d. entries satisfying $\operatorname{Var}(Y_{1,1}) = 1$, $\mathbb{E}(Y_{1,1}^4) < +\infty$, and $\mathbb{E}(Y_{1,1}) = 0$, for any deterministic matrix $M \in \mathcal{M}_{n,p}(\mathbb{R})$, for all n large enough, and for all $z \in V_{s,t}$, we have*

$$\left| \mathbb{E} G_{\mu_{XX^t}}(z) - \mathbb{E} G_{\mu_{\widehat{X}\widehat{X}^t}}(z) \right| \leq f(z) \left(\mathbb{E} |Y_{1,1}|^3 + \mathbb{E}(Y_{1,1}^4) \right) \left(\frac{1}{\sqrt{n}} + \frac{\operatorname{Tr}(MM^t)^{1/2}}{n} \right) .$$

Proof. The proof consists in four main steps. After developing $\mathbb{E} G_{\mu_{XX^t}}(z) - \mathbb{E} G_{\mu_{\widehat{X}\widehat{X}^t}}(z)$ thanks to the interpolation, we use integration by parts formulas (see Lemma 2.7). Then, we respectively focus on bounds for the main terms and for the rests in these integrations by parts.

> **First step: Development of** $\mathbb{E} G_{\mu_{XX^t}}(z) - \mathbb{E} G_{\mu_{\widehat{X}\widehat{X}^t}}(z)$**.**

Let $u \in [0,1]$ and $h \in [-u, 1-u]$. Proposition B.2 (ii), applied to $A = X(u)$ and $B = \frac{1}{\sqrt{p}}(Y(u+h) - Y(u))$ gives

$$S(u+h) - S(u) = S(u+h) \left(X(u) \left(\frac{Y(u+h) - Y(u)}{\sqrt{p}} \right)^t + \frac{Y(u+h) - Y(u)}{\sqrt{p}} X(u)^t \right.$$

$$\left. + \frac{Y(u+h) - Y(u)}{\sqrt{p}} \left(\frac{Y(u+h) - Y(u)}{\sqrt{p}} \right)^t \right) S(u).$$

Dividing by h and taking $h \to 0$, we get for all $u \in [0,1]$,

$$
\begin{aligned}
S'(u) &= S(u)\left(X(u)\frac{Y'(u)^t}{\sqrt{p}} + \frac{Y'(u)}{\sqrt{p}}X(u)^t\right)S(u) \\
&= \frac{1}{\sqrt{p}}S(u)\left(\left(\frac{\sqrt{u}Y}{\sqrt{p}} + \frac{\sqrt{1-u}\widehat{Y}}{\sqrt{p}} + M\right)\left(\frac{Y^t}{2\sqrt{u}} - \frac{\widehat{Y}^t}{2\sqrt{1-u}}\right) \right. \\
&\qquad \left. + \left(\frac{Y}{2\sqrt{u}} - \frac{\widehat{Y}}{2\sqrt{1-u}}\right)\left(\frac{\sqrt{u}Y^t}{\sqrt{p}} + \frac{\sqrt{1-u}\widehat{Y}^t}{\sqrt{p}} + M^t\right)\right)S(u).
\end{aligned}
$$

Thus we can rewrite

$$
\begin{aligned}
&G_{\mu_{XX^t}}(z) - G_{\mu_{\widehat{X}\widehat{X}^t}}(z) \\
&= \frac{1}{n}\operatorname{Tr}S(1) - \frac{1}{n}\operatorname{Tr}S(0) \\
&= \frac{1}{n}\int_0^1 \operatorname{Tr}S'(u)\,du \\
&= \frac{1}{2n\sqrt{p}}\int_0^1 \operatorname{Tr}\left[S(u)^2\left[2\frac{YY^t}{\sqrt{p}} - 2\frac{\widehat{Y}\widehat{Y}^t}{\sqrt{p}} + \left(\sqrt{\frac{1-u}{u}} - \sqrt{\frac{u}{1-u}}\right)\frac{Y\widehat{Y}^t}{\sqrt{p}} \right.\right. \\
&\qquad \left.\left. + \left(\sqrt{\frac{1-u}{u}} - \sqrt{\frac{u}{1-u}}\right)\frac{\widehat{Y}Y^t}{\sqrt{p}} + \frac{MY^t}{\sqrt{u}} - \frac{M\widehat{Y}^t}{\sqrt{1-u}} + \frac{YM^t}{\sqrt{u}} - \frac{\widehat{Y}M^t}{\sqrt{1-u}}\right]\right]\,du.
\end{aligned}
$$

Denoting by

$$
\begin{aligned}
\text{(I)} &= \operatorname{Tr}S(u)^2\left[\frac{YY^t}{\sqrt{p}} - \sqrt{\frac{u}{1-u}}\frac{Y\widehat{Y}^t}{\sqrt{p}}\right] \\
&= \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{p}}\left[S(u)_{j,k}^2 Y_{k,l}Y_{j,l} - \sqrt{\frac{u}{1-u}}S(u)_{j,k}^2 Y_{k,l}\widehat{Y}_{j,l}\right],
\end{aligned}
$$

$$
\text{(II)} = \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{p}}\left[S(u)_{j,k}^2 Y_{k,l}Y_{j,l} - \sqrt{\frac{u}{1-u}}S(u)_{j,k}^2 \widehat{Y}_{k,l}Y_{j,l}\right],
$$

$$
\text{(III)} = \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{p}}\left[\sqrt{\frac{1-u}{u}}S(u)_{j,k}^2 Y_{k,l}\widehat{Y}_{j,l} - S(u)_{j,k}^2 \widehat{Y}_{k,l}\widehat{Y}_{j,l}\right],
$$

$$
\text{(IV)} = \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{p}}\left[\sqrt{\frac{1-u}{u}}S(u)_{j,k}^2 \widehat{Y}_{k,l}Y_{j,l} - S(u)_{j,k}^2 \widehat{Y}_{k,l}\widehat{Y}_{j,l}\right],
$$

$$
\text{(V)} = \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{u}}S(u)_{j,k}^2 M_{k,l}Y_{j,l} - \frac{1}{\sqrt{1-u}}S(u)_{j,k}^2 M_{k,l}\widehat{Y}_{j,l},
$$

and

$$
\text{(VI)} = \sum_{\substack{1\le j,k\le n \\ 1\le l\le p}} \frac{1}{\sqrt{u}}S(u)_{j,k}^2 Y_{k,l}M_{j,l} - \frac{1}{\sqrt{1-u}}S(u)_{j,k}^2 \widehat{Y}_{k,l}M_{j,l},
$$

where $S(u)_{j,k}^2$ must be read $(S(u)^2)_{j,k}$, we finally rewrite

$$
\mathbb{E}\,G_{\mu_{XX^t}}(z) - \mathbb{E}\,G_{\mu_{\widehat{X}\widehat{X}^t}}(z) = \frac{1}{2n\sqrt{p}}\int_0^1 \mathbb{E}[\text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)} + \text{(VI)}]\,du. \quad (2.19)
$$

Second step: Integrations by parts.

Let us recall the formulas we will use below, in particular the Stein characterization of a normal distribution (2.20).

Lemma 2.7 (see [19, Formulas (2.1.39) and (18.1.19)]). *(i) Let a function $F \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ and ξ a random variable with distribution $\mathcal{N}(0, \sigma^2)$. If $\mathbb{E}\,|F'(\xi)| < +\infty$, then*

$$\mathbb{E}\,(\xi F(\xi)) = \sigma^2\,\mathbb{E}\,(F'(\xi))\,. \tag{2.20}$$

(ii) More generally, let p be an integer, a function $F \in \mathcal{C}^{p+1}(\mathbb{R}, \mathbb{R})$, and a real random variable ξ. If $\mathbb{E}\,|\xi|^{p+2} < +\infty$ and the derivatives $F', \ldots, F^{(p+1)}$ are bounded on $\mathbb{R}$, then

$$\mathbb{E}\,(\xi F(\xi)) = \sum_{j=0}^{p} \frac{\kappa_{j+1}}{j!}\,\mathbb{E}(F^{(j)}(\xi)) + \varepsilon_p \tag{2.21}$$

where the κ_{j+1}'s are the cumulants of the distribution of ξ and

$$|\varepsilon_p| \leq C_p\,\mathbb{E}\,|\xi|^{p+2}.\|F^{(p+1)}\|_\infty\,, \qquad C_p \leq \frac{1 + (3 + 2p)^{p+2}}{(p+1)!}\,.$$

We will apply the Gaussian (2.20) or the general (2.21) integration by parts formula for all j, k, l in order to decompose

$$\mathbb{E}[(\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV}) + (\mathrm{V}) + (\mathrm{VI})]$$

as a sum of terms that we can bound.

Note a first crucial point here. As we want to apply Theorem 1.6 to the matrices Y and C in Section 3 in order to obtain (3.10), it will not be sufficient to use the integration by parts formula up to order 2, that is why we will be interested in terms of order 3 in this formula.

From now, $D_{a,b}$ denotes the derivation with respect to $Y_{a,b}$.

Let $u \in [0, 1]$, $j, k \in [\![1, n]\!]$, and $l \in [\![1, p]\!]$. We denote by F_1 and G_1 the functions defined by $F_1(Y_{j,l}) = Y_{k,l} S(u)_{j,k}^2$ and $G_1(\widehat{Y}_{j,l}) = Y_{k,l} S(u)_{j,k}^2$. We have

$$F_1'(Y_{j,l}) = \frac{2\sqrt{u}}{\sqrt{p}} Y_{k,l} S(u)_{j,k}.D_{j,l} S(u)_{j,k} + \delta_{j,k} S(u)_{k,k}^2\,,$$

$$F_1''(Y_{j,l}) = \frac{2u}{p} Y_{k,l}\left((D_{j,l}S(u)_{j,k})^2 + S(u)_{j,k}.D_{j,l}^2 S(u)_{j,k}\right) + \frac{4\sqrt{u}}{\sqrt{p}}\delta_{j,k} S(u)_{k,k}.D_{k,l} S(u)_{k,k}\,,$$

$$F_1^{(3)}(Y_{j,l}) = \frac{2u^{3/2}}{p^{3/2}} Y_{k,l}\left(3 D_{j,l}S(u)_{j,k}.D_{j,l}^2 S(u)_{j,k} + S(u)_{j,k}.D_{j,l}^3 S(u)_{j,k}\right)$$
$$+ \frac{6u}{p}\delta_{j,k}\left((D_{k,l}S(u)_{k,k})^2 + S(u)_{k,k}.D_{k,l}^2 S(u)_{k,k}\right)\,,$$

and

$$G_1'(\widehat{Y}_{j,l}) = \frac{2\sqrt{1-u}}{\sqrt{p}} Y_{k,l} S(u)_{j,k}.D_{j,l} S(u)_{j,k}\,.$$

Applied conditionally to the variables $\{Y_{a,b},\ 1 \leq a \leq n,\ 1 \leq b \leq p\} \cup \{\widehat{Y}_{a,b},\ (a,b) \neq (j,l)\}$, (2.20) gives

$$\widehat{\mathbb{E}}_{j,l}(S(u)_{j,k}^2 Y_{k,l} \widehat{Y}_{j,l}) = \mathrm{Var}(\widehat{Y}_{j,l})\widehat{\mathbb{E}}_{j,l}(G_1'(\widehat{Y}_{j,l}))\,,$$

where $\widehat{\mathbb{E}}_{j,l}$ denotes the associated conditional expectation. Similarly, from (2.21), we have

$$\mathbb{E}_{j,l}(S(u)_{j,k}^2 Y_{k,l} Y_{j,l}) = \mathrm{Var}(Y_{j,l})\,\mathbb{E}_{j,l}(F_1'(Y_{j,l})) + \frac{\kappa_3(Y_{j,l})}{2}\,\mathbb{E}_{j,l}(F_1''(Y_{j,l})) + \varepsilon_{1,j,k,l}\,,$$

where $\mathbb{E}_{j,l}$ denotes the expectation conditionally to the variables $\{Y_{a,b},\ (a,b) \neq (j,l)\} \cup \{\widehat{Y}_{a,b},\ 1 \leq a \leq n,\ 1 \leq b \leq p\}$.

Taking the expectation, we thus have

$$\mathbb{E}\left[S(u)_{j,k}^2 Y_{k,l} Y_{j,l} - \sqrt{\frac{u}{1-u}} S(u)_{j,k}^2 Y_{k,l} \widehat{Y}_{j,l}\right]$$

$$= \mathrm{Var}(Y_{j,l})\,\mathbb{E}(F_1'(Y_{j,l})) + \frac{\kappa_3(Y_{j,l})}{2}\,\mathbb{E}(F_1''(Y_{j,l})) + \mathbb{E}(\varepsilon_{1,j,k,l}) - \sqrt{\frac{u}{1-u}}\,\mathrm{Var}(\widehat{Y}_{j,l})\,\mathbb{E}(G_1'(\widehat{Y}_{j,l}))$$

$$= \delta_{j,k}\,\mathbb{E}(S(u)_{k,k}^2) + \frac{\kappa_3(Y_{j,l})}{2}\,\mathbb{E}(F_1''(Y_{j,l})) + \mathbb{E}(\varepsilon_{1,j,k,l})$$

with, for all j, k, l,

$$|\varepsilon_{1,j,k,l}| \leq \frac{1+7^4}{6}\,\mathbb{E}(Y_{1,1}^4).\|F_1^{(3)}\|_\infty\,.$$

Dividing by $\sqrt{p}$ and summing on j, k, l, we thus have

$$\mathbb{E}(\mathrm{I}) = \underbrace{\sqrt{p}\,\mathbb{E}(\mathrm{Tr}\,S(u)^2)}_{(\mathrm{I}.1)} + \underbrace{\frac{\kappa_3(Y_{1,1})u}{p^{3/2}}\sum_{j,k,l}\mathbb{E}(Y_{k,l}(D_{j,l}S(u)_{j,k})^2)}_{(\mathrm{I}.2)}$$

$$+ \underbrace{\frac{\kappa_3(Y_{1,1})u}{p^{3/2}}\sum_{j,k,l}\mathbb{E}(Y_{k,l}S(u)_{j,k}.D_{j,l}^2 S(u)_{j,k})}_{(\mathrm{I}.3)}$$

$$+ \underbrace{\frac{2\kappa_3(Y_{1,1})\sqrt{u}}{p}\sum_{k,l}\mathbb{E}(S(u)_{k,k}.D_{k,l}S(u)_{k,k})}_{(\mathrm{I}.4)} + \frac{1}{\sqrt{p}}\sum_{j,k,l}\mathbb{E}(\varepsilon_{1,j,k,l})\,. \qquad (2.22)$$

Since $S(u)^t = S(u)$, we also have

$$\mathbb{E}(\mathrm{II}) = (\mathrm{I}.1) + (\mathrm{I}.2) + (\mathrm{I}.3) + (\mathrm{I}.4) + \frac{1}{\sqrt{p}}\sum_{j,k,l}\mathbb{E}(\varepsilon_{2,j,k,l}) \qquad (2.23)$$

with

$$|\varepsilon_{2,j,k,l}| \leq \frac{1+7^4}{6}\,\mathbb{E}(Y_{1,1}^4).\|F_1^{(3)}\|_\infty\,.$$

Similarly, considering $F_3(Y_{k,l}) = \widehat{Y}_{j,l}S(u)_{j,k}^2$, we get

$$\mathbb{E}(\mathrm{III}) = \underbrace{-\sqrt{p}\,\mathbb{E}(\mathrm{Tr}\,S(u)^2)}_{(\mathrm{III}.1)} + \underbrace{\frac{\kappa_3(Y_{1,1})\sqrt{u(1-u)}}{p^{3/2}}\sum_{j,k,l}\mathbb{E}(\widehat{Y}_{j,l}(D_{k,l}S(u)_{j,k})^2)}_{(\mathrm{III}.2)}$$

$$+ \underbrace{\frac{\kappa_3(Y_{1,1})\sqrt{u(1-u)}}{p^{3/2}}\sum_{j,k,l}\mathbb{E}(\widehat{Y}_{j,l}S(u)_{j,k}.D_{k,l}^2 S(u)_{j,k})}_{(\mathrm{III}.3)} + \frac{1}{\sqrt{p}}\sum_{j,k,l}\mathbb{E}(\varepsilon_{3,j,k,l})\,, \qquad (2.24)$$

$$\mathbb{E}(\mathrm{IV}) = (\mathrm{III}.1) + (\mathrm{III}.2) + (\mathrm{III}.3) + \frac{1}{\sqrt{p}}\sum_{j,k,l}\mathbb{E}(\varepsilon_{4,j,k,l}) \qquad (2.25)$$

with

$$|\varepsilon_{3,j,k,l}| \leq \frac{1+7^4}{6}\,\mathbb{E}(Y_{1,1}^4).\|F_3^{(3)}\|_\infty \quad \text{and} \quad |\varepsilon_{4,j,k,l}| \leq \frac{1+7^4}{6}\,\mathbb{E}(Y_{1,1}^4).\|F_3^{(3)}\|_\infty\,,$$

and considering $F_5(Y_{j,l}) = M_{k,l}S(u)_{j,k}^2$, we get

$$\mathbb{E}(\mathrm{V}) = \underbrace{\frac{\kappa_3(Y_{1,1})\sqrt{u}}{p} \sum_{j,k,l} \mathbb{E}(M_{k,l}(D_{j,l}S(u)_{j,k})^2)}_{(\mathrm{V}.1)}$$

$$+ \underbrace{\frac{\kappa_3(Y_{1,1})\sqrt{u}}{p} \sum_{j,k,l} \mathbb{E}(M_{k,l}S(u)_{j,k}.D_{j,l}^2 S(u)_{j,k})}_{(\mathrm{V}.2)} + \sum_{j,k,l} \mathbb{E}(\varepsilon_{5,j,k,l}), \quad (2.26)$$

$$\mathbb{E}(\mathrm{VI}) = (\mathrm{V}.1) + (\mathrm{V}.2) + \sum_{j,k,l} \mathbb{E}(\varepsilon_{6,j,k,l}) \qquad (2.27)$$

with

$$|\varepsilon_{5,j,k,l}| \le \frac{1+7^4}{6} \mathbb{E}(Y_{1,1}^4).\|F_5^{(3)}\|_\infty \quad \text{and} \quad |\varepsilon_{6,j,k,l}| \le \frac{1+7^4}{6} \mathbb{E}(Y_{1,1}^4).\|F_5^{(3)}\|_\infty .$$

We have thus rewritten (2.19) as

$$\mathbb{E}\, G_{\mu_{XX^t}}(z) - \mathbb{E}\, G_{\mu_{\widetilde{X}\widetilde{X}^t}}(z)$$

$$= \frac{1}{n\sqrt{p}} \int_0^1 \mathbb{E}[(\mathrm{I}.2) + (\mathrm{I}.3) + (\mathrm{I}.4) + (\mathrm{III}.2) + (\mathrm{III}.3) + (\mathrm{V}.1) + (\mathrm{V}.2)]\, du$$

$$+ \frac{1}{2n\sqrt{p}} \int_0^1 \mathbb{E}\left[\sum_{j,k,l}\left(\frac{1}{\sqrt{p}}\sum_{i=1}^4 \varepsilon_{i,j,k,l} + \sum_{i=5}^6 \varepsilon_{i,j,k,l}\right)\right]\, du .$$

Third step: Bounds for the main terms.

We will develop the different terms in this expression with the differentiation formulas in Proposition B.2 (vii), and bound them thanks to inequalities on traces and resolvents (see Propositions B.1 and B.2 again).

Because some computations are very similar, we will be interested in the terms (I.2), (I.3), and (I.4) only.

Note that in order to simplify the notations, from now, we will denote S and X for $S(u)$ and $X(u)$.

Let us start with the term (I.2). Using (B.1), we have

$$\sum_{j,k,l} \mathbb{E}(Y_{k,l}(D_{j,l}S(u)_{j,k})^2)$$

$$= \sum_{j,k,l} \mathbb{E}\left[Y_{k,l}(SX)_{j,l}^2 S_{j,k}^2 + 2Y_{k,l}(SX)_{j,l}S_{j,k}S_{j,j}(SX)_{k,l} + Y_{k,l}S_{j,j}^2(SX)_{k,l}^2\right]$$

$$= \mathbb{E}\left[\mathrm{Tr}(Y(X^tS)^{\circ 2}S^{\circ 2}) + 2\,\mathrm{Tr}((Y\circ SX)X^tS\,\mathrm{diag}(S)S) + \sum_j S_{j,j}^2 \cdot \sum_{k,l} Y_{k,l}(SX)_{k,l}^2\right]$$

where $\circ$ is the Hadamard product (see Appendix B.1), $S^{\circ 2}$ stands for $S \circ S$, and $\mathrm{diag}(S)$ is the matrix S from which we have removed off-diagonal entries. Note that it is crucial here to rewrite precisely the terms with the Hadamard product and then to bound the traces rather than bound directly the entries. Indeed, it allows us to get better powers of n in the bound, which is crucial if we have in mind the large deviations in Section 3.

Using Propositions B.1, B.2, and the Cauchy-Schwarz inequality in $\mathbb{C}^{np}$, and denoting by y a square root of z, we have

$$|\mathrm{Tr}(Y(X^tS)^{\circ 2}S^{\circ 2})| \le \sqrt{p}\|X^tS\|^2.\|S\|^2\,\mathrm{Tr}(YY^t)^{1/2}$$

$$\le \frac{\sqrt{p}}{|\mathrm{Im}\, y|^2|\mathrm{Im}\, z|^2}\,\mathrm{Tr}(YY^t)^{1/2} ,$$

$$| \operatorname{Tr}((Y \circ SX)X^t S \operatorname{diag}(S)S)| \leq \sqrt{p}\|X^t S\|.\| \operatorname{diag}(S)\|.\|S\| \operatorname{Tr}((Y \circ SX)(Y \circ SX)^*)^{1/2}$$
$$\leq \frac{\sqrt{p}}{|\operatorname{Im} y|^2|\operatorname{Im} z|^2} \operatorname{Tr}(YY^t)^{1/2},$$

and

$$\left| \sum_j S_{j,j}^2 . \sum_{k,l} Y_{k,l}(SX)_{k,l}^2 \right| \leq \frac{n}{|\operatorname{Im} z|^2|\operatorname{Im} y|} \sum_{k,l} |Y_{k,l}(SX)_{k,l}|$$
$$\leq \frac{n}{|\operatorname{Im} z|^2|\operatorname{Im} y|} \left(\sum_{k,l} Y_{k,l}^2 \right)^{1/2} \left(\sum_{k,l} |(SX)_{k,l}|^2 \right)^{1/2}$$
$$= \frac{n \operatorname{Tr}(YY^t)^{1/2}}{|\operatorname{Im} z|^2|\operatorname{Im} y|} \operatorname{Tr}(SXX^t S^*)^{1/2}$$
$$\leq \frac{n \operatorname{Tr}(YY^t)^{1/2}}{|\operatorname{Im} z|^2|\operatorname{Im} y|} \left(\frac{n|z|}{|\operatorname{Im} z|^2} + \frac{n}{|\operatorname{Im} z|} \right)^{1/2}.$$

Using also the bound (B.7), there exists a function f, bounded on $V_{s,t}$, independent from Y, M, and n, such that for all $z \in V_{s,t}$, we have

$$|(\mathrm{I.2})| \leq |\kappa_3(Y_{1,1})|f(z)\,\mathbb{E}(\operatorname{Tr}(YY^t)^{1/2}).$$

But for a centred random variable, the third cumulant equals the third moment, so this inequality can be rewritten

$$|(\mathrm{I.2})| \leq \mathbb{E}\,|Y_{1,1}|^3 f(z)\,\mathbb{E}(\operatorname{Tr}(YY^t)^{1/2}).$$

We adopt the same strategy for the term (I.3). Using (B.5), we get

$$\sum_{j,k,l} \mathbb{E}(Y_{k,l}S(u)_{j,k}.D_{j,l}^2 S(u)_{j,k})$$
$$= \sum_{j,k,l} \mathbb{E}[Y_{k,l}S_{j,k}.2(S_{j,j}S_{j,k} + (SX)_{j,l}^2 S_{j,k} + S_{j,j}(X^t SX)_{l,l}S_{j,k} + 2S_{j,j}(SX)_{j,l}(SX)_{k,l})]$$
$$= 2\,\mathbb{E}\left[\sum_{j,l} S_{j,j}(S^{\circ 2}Y)_{j,l} + \operatorname{Tr}(Y(X^t S)^{\circ 2}S^{\circ 2}) \right.$$
$$\left. + \sum_{j,l} S_{j,j}(S^{\circ 2}Y)_{j,l}(X^t SX)_{l,l} + 2\operatorname{Tr}((Y \circ SX)X^t S \operatorname{diag}(S)S) \right]$$

so, using the previous bounds, and also

$$\left| \sum_{j,l} S_{j,j}(S^{\circ 2}Y)_{j,l} \right| \leq \frac{1}{|\operatorname{Im} z|}\sqrt{np} \left(\sum_{j,l} |(S^{\circ 2}Y)_{j,l}|^2 \right)^{1/2}$$
$$= \frac{\sqrt{np}}{|\operatorname{Im} z|} \operatorname{Tr}(S^{\circ 2}YY^t(S^{\circ 2})^*)^{1/2}$$
$$\leq \frac{\sqrt{np}}{|\operatorname{Im} z|} \left(\sqrt{n}\frac{\|Y\|}{|\operatorname{Im} z|^4} \operatorname{Tr}(YY^t)^{1/2} \right)^{1/2}$$
$$\leq \frac{n^{3/4}p^{1/2}}{|\operatorname{Im} z|^3} \operatorname{Tr}(YY^t)^{1/2}$$

and

$$\left| \sum_{j,l} S_{j,j} (S^{\circ 2} Y)_{j,l} (X^t S X)_{l,l} \right| \leq \frac{1}{|\operatorname{Im} z|} \left(1 + \left| \frac{z}{\operatorname{Im} z} \right| \right) \sqrt{np} \left(\sum_{j,l} |(S^{\circ 2} Y)_{j,l}|^2 \right)^{1/2}$$

$$\leq \frac{n^{3/4} p^{1/2}}{|\operatorname{Im} z|^3} \left(1 + \left| \frac{z}{\operatorname{Im} z} \right| \right) \operatorname{Tr}(YY^t)^{1/2},$$

the same arguments as above lead to

$$|(\mathrm{I}.3)| \leq \frac{\mathbb{E} |Y_{1,1}|^3}{n^{1/4}} f(z) \, \mathbb{E}(\operatorname{Tr}(YY^t)^{1/2}).$$

Besides, we have

$$\sum_{k,l} \mathbb{E}(S(u)_{k,k} . D_{k,l} S(u)_{k,k}) = \sum_{k,l} \mathbb{E}[S_{k,k} . 2 S_{k,k} (SX)_{k,l}]$$

and

$$\left| \sum_{k,l} S_{k,k}^2 (SX)_{k,l} \right| \leq \frac{1}{|\operatorname{Im} z|^2} \sqrt{np} \left(\sum_{k,l} |(SX)_{k,l}|^2 \right)^{1/2}$$

$$= \frac{\sqrt{np}}{|\operatorname{Im} z|^2} \operatorname{Tr}(SXX^t S^*)^{1/2} \leq \frac{\sqrt{np}}{|\operatorname{Im} z|^2} \left(\frac{n|z|}{|\operatorname{Im} z|^2} + \frac{n}{|\operatorname{Im} z|} \right)^{1/2}$$

thus we get

$$|(\mathrm{I}.4)| \leq \mathbb{E} |Y_{1,1}|^3 f(z) \sqrt{n}.$$

We finally have

$$|(\mathrm{I}.2) + (\mathrm{I}.3) + (\mathrm{I}.4)| \leq \mathbb{E} |Y_{1,1}|^3 f(z) \left(\mathbb{E}(\operatorname{Tr}(YY^t)^{1/2}) + \sqrt{n} \right). \qquad (2.28)$$

Very similar computations allow to show that

$$|(\mathrm{III}.2) + (\mathrm{III}.3)| \leq \mathbb{E} |Y_{1,1}|^3 f(z) \, \mathbb{E}(\operatorname{Tr}(\widehat{Y}\widehat{Y}^t)^{1/2})$$

and

$$|(\mathrm{V}.1) + (\mathrm{V}.2)| \leq \mathbb{E} |Y_{1,1}|^3 f(z) \sqrt{n} \operatorname{Tr}(MM^t)^{1/2}. \qquad (2.29)$$

If we remember that $Y_{1,1}$ and $\widehat{Y}_{1,1}$ have mean zero and variance 1, we have $\mathbb{E}(\operatorname{Tr}(YY^t))^{1/2} \leq \sqrt{np}$ and $\mathbb{E}(\operatorname{Tr}(\widehat{Y}\widehat{Y}^t))^{1/2} \leq \sqrt{np}$ by Jensen's inequality. Finally, we can write

$$|(\mathrm{I}.2) + (\mathrm{I}.3) + (\mathrm{I}.4) + (\mathrm{III}.2) + (\mathrm{III}.3) + (\mathrm{V}.1) + (\mathrm{V}.2)|$$

$$\leq \mathbb{E} |Y_{1,1}|^3 f(z) \left(n + \sqrt{n} \operatorname{Tr}(MM^t)^{1/2} \right). \qquad (2.30)$$

Fourth step: Bounds for the rests.

The only thing to be left is to bound the rests appeared in the integration by parts formulas. We recall that for all $j, k \in [\![1, n]\!]$, $l \in [\![1, p]\!]$, we have

$$|\varepsilon_{1,j,k,l}| \leq \frac{1 + 7^4}{6} \mathbb{E}(Y_{1,1}^4) . \|F_1^{(3)}\|_\infty.$$

Using the expression of $F_1^{(3)}(Y_{j,l})$, differentiation formulas (B.1), (B.5), (B.6), and inequalities (iv)–(vi) in Proposition B.2, there exists a function f, independent from Y, M, n, j, k, l, bounded on $V_{s,t}$, such that

$$|\varepsilon_{1,j,k,l}| \leq f(z)\,\mathbb{E}(Y_{1,1}^4)\left(\frac{u^{3/2}}{p^{3/2}}|Y_{k,l}| + \frac{u}{p}\delta_{j,k}\right).$$

So, using the Cauchy-Schwarz inequality in $\mathbb{R}^{np}$, we have

$$\begin{aligned}
\frac{1}{\sqrt{p}}\left|\sum_{j,k,l}\mathbb{E}(\varepsilon_{1,j,k,l})\right| &\leq f(z)\,\mathbb{E}(Y_{1,1}^4)\left(\frac{1}{p}\sum_{k,l}\mathbb{E}\,|Y_{k,l}| + \sqrt{n}\right) \\
&\leq f(z)\,\mathbb{E}(Y_{1,1}^4)\left(\mathbb{E}(\mathrm{Tr}(YY^t)^{1/2}) + \sqrt{n}\right) \\
&\leq f(z)\,\mathbb{E}(Y_{1,1}^4)n.
\end{aligned} \tag{2.31}$$

The same bound holds for $\frac{1}{\sqrt{p}}\left|\sum_{j,k,l}\mathbb{E}(\varepsilon_{2,j,k,l})\right|$. Similarly, we get

$$\frac{1}{\sqrt{p}}\left|\sum_{j,k,l}\mathbb{E}(\varepsilon_{3,j,k,l}) + \mathbb{E}(\varepsilon_{4,j,k,l})\right| \leq f(z)\,\mathbb{E}(Y_{1,1}^4)n \tag{2.32}$$

and

$$\left|\sum_{j,k,l}\varepsilon_{5,j,k,l} + \varepsilon_{6,j,k,l}\right| \leq f(z)\,\mathbb{E}(Y_{1,1}^4)\sqrt{n}\,\mathrm{Tr}(MM^t)^{1/2}. \tag{2.33}$$

Finally, combining relations from (2.19) to (2.33), we get

$$\left|\mathbb{E}\,G_{\mu_{XX^t}}(z) - \mathbb{E}\,G_{\mu_{\widehat{X}\widehat{X}^t}}(z)\right| \leq f(z)\left(\mathbb{E}\,|Y_{1,1}|^3 + \mathbb{E}(Y_{1,1}^4)\right)\left(\frac{1}{\sqrt{n}} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n}\right). \tag{2.34}$$
$$\square$$

We can now conclude the proof of the general case and obtain Theorem 2.1. In fact, in Proposition 2.6, we assumed that $\mathbb{E}(Y_{1,1}) = 0$, so we only have to remove this assumption.

Proof. We recall that $\mathring{X} = X - \mathbb{E}(X)$ by definition. We also define $\underline{g}(z) = \mathbb{E}\,G_{\mu_{XX^t}}(z)$, $g_\circ(z) = \mathbb{E}\,G_{\mu_{\mathring{X}\mathring{X}^t}}(z)$, and $\widehat{g}(z) = \mathbb{E}\,G_{\mu_{\widehat{X}\widehat{X}^t}}(z)$. Using the notations in Lemma 2.2, we have

$$\begin{aligned}
&\left|\underline{g}(z) - (1 - c\underline{g}(z))G_{\mu_{MM^t}}(z(1 - c\underline{g}(z))^2 - (1 - c)(1 - c\underline{g}(z)))\right| \\
&\leq \left|\underline{g}(z) - g_\circ(z)\right| + |g_\circ(z) - \widehat{g}(z)| + \left|\widehat{g}(z) - \phi_{z,\mu_{MM^t}}(\widehat{g}(z), c)\right| \\
&\quad + \left|\phi_{z,\mu_{MM^t}}(\widehat{g}(z), c) - \phi_{z,\mu_{MM^t}}(g_\circ(z), c)\right| \\
&\quad + \left|\phi_{z,\mu_{MM^t}}(g_\circ(z), c) - \phi_{z,\mu_{MM^t}}(\underline{g}(z), c)\right| \\
&\leq (1 + l_{s,t})\left|\underline{g}(z) - g_\circ(z)\right| + (1 + l_{s,t})|g_\circ(z) - \widehat{g}(z)| + \left|\widehat{g}(z) - \phi_{z,\mu_{MM^t}}(\widehat{g}(z), c)\right|
\end{aligned}$$

for s large enough and t small enough by Lemma 2.2.

Since the matrix $X - \mathring{X} = \mathbb{E}(X)$ has rank at most 1, using the relations (1.5), (1.7), and (B.9), we have

$$\left|G_{\mu_{XX^t}}(z) - G_{\mu_{\mathring{X}\mathring{X}^t}}(z)\right| \leq d_{s,t}\left(\mu_{XX^t}, \mu_{\mathring{X}\mathring{X}^t}\right) \leq d_{\mathrm{KS}}\left(\mu_{XX^t}, \mu_{\mathring{X}\mathring{X}^t}\right) \leq \frac{1}{n}.$$

This inequality, Proposition 2.3 (the Gaussian case) applied to $\widehat{Y}$, and Proposition 2.6 (the centred case) applied to $\mathring{Y}$ permit to get finally

$$\left| \underline{g}(z) - (1 - c\underline{g}(z))G_{\mu_{MM^t}}(z(1 - c\underline{g}(z))^2 - (1 - c)(1 - c\underline{g}(z))) \right|$$

$$\leq (1 + l_{s,t}).\frac{1}{n} + f(z)\left(\mathbb{E}\,|\mathring{Y}_{1,1}|^3 + \mathbb{E}(\mathring{Y}_{1,1}^4) \right)\left(\frac{1}{\sqrt{n}} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n} \right)$$

$$+ f(z)\left(|c_n - c| + \frac{1}{n} + \frac{\mathrm{Tr}(MM^t)^{1/2}}{n^{5/4}} \right). \quad \square$$

3 Large deviations

This section is devoted to the proof of Theorem 1.7. In this section, $X \in \mathcal{M}_{n,p}(\mathbb{R})$ is a random matrix such that $c_n = \frac{n}{p} \to c \in (0, +\infty)$. Moreover, we assume that $\mathrm{Var}(X_{1,1}) = 1$ and that there exist $\alpha \in (0, 2)$ and $a \in (0, +\infty]$ such that $X_{1,1} \in \mathcal{S}_\alpha(a)$ (see Definition 1.1).

We define

$$\varepsilon(n) = \frac{1}{\log n}$$

and we decompose the matrix X as

$$\frac{X}{\sqrt{p}} = A + B + C + D\,, \tag{3.1}$$

where A, B, C, D are the matrices defined by

$$A_{j,k} = \frac{X_{j,k}}{\sqrt{p}}\mathbf{1}_{|X_{j,k}|<(\log n)^{2/\alpha}} \qquad B_{j,k} = \frac{X_{j,k}}{\sqrt{p}}\mathbf{1}_{(\log n)^{2/\alpha}\leq|X_{j,k}|\leq\varepsilon(n)\sqrt{p}}$$

$$C_{j,k} = \frac{X_{j,k}}{\sqrt{p}}\mathbf{1}_{\varepsilon(n)\sqrt{p}<|X_{j,k}|\leq\varepsilon(n)^{-1}\sqrt{p}} \qquad D_{j,k} = \frac{X_{j,k}}{\sqrt{p}}\mathbf{1}_{\varepsilon(n)^{-1}\sqrt{p}<|X_{j,k}|}\,.$$

This decomposition, introduced in [7], must be understood as follows. First, the matrix A is the bulk of the matrix $X/\sqrt{p}$ because of the tail assumption (1.1), see for instance Lemma 3.4 (ii) below. Besides, we keep in C the terms of order $\sqrt{p}$ in X, which produce large deviations as we briefly explained at the end of Introduction (note that we could have chosen an other sequence $\varepsilon(n)$ going slowly to 0).

3.1 Exponential equivalences

The goal of this subsection is to prove the following.

Proposition 3.1. *There exist $s, t > 0$ such that the random distributions $\mu_{XX^t/p}$ and $\left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ are $d_{s,t}$-exponentially equivalent at scale $n^{1+\alpha/2}$ as $n \to +\infty$, i.e. for all $\delta > 0$, we have*

$$\lim_{n\to+\infty} \frac{1}{n^{1+\alpha/2}}\log\mathbb{P}\left(d_{s,t}\left(\mu_{XX^t/p}, \left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2 \right) \geq \delta \right) = -\infty\,.$$

The strategy to prove Proposition 3.1 is similar to the one in [7]. First, we explain why B and D do not contribute to large deviations in Lemmas 3.2 and 3.3 and then, we show that the measures $\mu_{(A+C)(A+C)^t}$ and $\left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ are exponentially equivalent thanks to a conditioning and a coupling argument in which several tools are needed, such as the concentration property (A.15) and the asymptotic freeness result stated in Theorem 1.6. From now on, we consider $s > 2$ and $t > 0$.

First, the contribution of D is negligible.

Lemma 3.2. $\mu_{XX^t/p}$ *and* $\mu_{(A+B+C)(A+B+C)^t}$ *are exponentially equivalent.*

The proof is very similar to what is done in [7], the only difference being the use of (B.9) instead of (B.8). Therefore, it will not be repeated here.

The contribution of B can also be neglected.

Lemma 3.3. $\mu_{XX^t/p}$ *and* $\mu_{(A+C)(A+C)^t}$ *are exponentially equivalent.*

Proof. From Lemma 3.2, the triangle inequality, Lemma 1.2.15 in [9], and the inequalities $d_{s,t} \leq W_1 \leq W_2$, it is sufficient to prove that for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(W_2\left(\mu_{(A+B+C)(A+B+C)^t}, \mu_{(A+C)(A+C)^t} \right) \geq \delta \right) = -\infty \,.$$

From (B.11), which is the analogue of the Hoffman-Wielandt inequality (B.10) for covariance matrices, it is sufficient to check that for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{2}{n^2} \operatorname{Tr}((A+B+C)(A+B+C)^t \right.$$
$$\left. + (A+C)(A+C)^t) \operatorname{Tr}(BB^t) \geq \delta \right) = -\infty \,.$$

Let $\delta > 0$. We have

$$\operatorname{Tr}((A+C)(A+C)^t) \leq \operatorname{Tr}((A+B+C)(A+B+C)^t) \leq \operatorname{Tr}\left(\frac{1}{p} XX^t \right)$$

using the decomposition (3.1). Thus,

$$\mathbb{P}\left(\frac{2}{n^2} \operatorname{Tr}((A+B+C)(A+B+C)^t + (A+C)(A+C)^t) \operatorname{Tr}(BB^t) \geq \delta \right)$$
$$\leq \quad \mathbb{P}\left(\frac{4}{n^2 p} \operatorname{Tr}(XX^t) \operatorname{Tr}(BB^t) \geq \delta \right)$$
$$\leq \quad \mathbb{P}\left(\frac{1}{np} \operatorname{Tr}(XX^t) \geq \mathbb{E}(X_{1,1}^2) + \delta \right) + \mathbb{P}\left(\frac{4}{n} \operatorname{Tr}(BB^t) \geq \frac{\delta}{\mathbb{E}(X_{1,1}^2) + \delta} \right) . \quad (3.2)$$

On the one hand, since $\operatorname{Tr}(XX^t)$ is the sum of np i.i.d. random variables, from Cramér's theorem in $\mathbb{R}$ (see [9, Theorem 2.2.3]), we have

$$\lim_{n \to +\infty} \frac{1}{np} \log \mathbb{P}\left(\frac{1}{np} \operatorname{Tr}(XX^t) \geq \mathbb{E}(X_{1,1}^2) + \delta \right) = -\sup_{\theta \in \mathbb{R}} \left(\theta(\mathbb{E}(X_{1,1}^2) + \delta) - \log \mathbb{E}(e^{\theta X_{1,1}}) \right) < 0$$

so, since $\alpha < 2$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{1}{np} \operatorname{Tr}(XX^t) \geq \mathbb{E}(X_{1,1}^2) + \delta \right) = -\infty \,. \quad (3.3)$$

On the other hand, since $\frac{n}{p} \to c$, the same arguments as in [7] lead to

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{4}{n} \operatorname{Tr}(BB^t) \geq \frac{\delta}{\mathbb{E}(X_{1,1}^2) + \delta} \right) = -\infty \,. \quad (3.4)$$

Finally, combining (3.2), (3.3), (3.4), and Lemma 1.2.15 in [9], we get the exponential equivalence of $\mu_{XX^t/p}$ and $\mu_{(A+C)(A+C)^t}$. $\qquad \square$

Before proving Proposition 3.1, we need some additional properties.

Lemma 3.4. *(i) We have*

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(\frac{1}{n} \operatorname{Tr}(CC^t) > (\log n)^2 \right) = -\infty \,.$$

(ii) Defining $I = \{(j,k) \mid |X_{j,k}| \geq (\log n)^{2/\alpha}\}$, for all $\delta > 0$, we have

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(|I| \geq \delta n^{1+\alpha/2} \right) = -\infty \,.$$

(iii) We denote by P_n the distribution of $X_{1,1}$ conditionally to $\{|X_{1,1}| < (\log n)^{2/\alpha}\}$. Let Z_n be a random variable with distribution P_n. There exists $\zeta > 0$ such that

$$\sup_{n \in \mathbb{N}} \max \left(\mathbb{E}(Z_n^2), (\mathbb{E}(Z_n^2))^2, \mathbb{E}(Z_n^4) \right) \leq \zeta \,.$$

Furthermore, the variance of Z_n, denoted by σ_n^2, tends to $\operatorname{Var}(X_{1,1}) = 1$ as $n \to +\infty$ and more precisely, there exists $\eta > 0$ such that

$$|\sigma_n^2 - 1| \leq \eta e^{-a(\log n)^2/4} \,.$$

Proof. The proofs of (i) and (ii) exactly follow the proof of Lemma 2.4 in [7]. Therefore, we will only prove (iii).

Let Z_n be a random variable with distribution P_n defined as above. We have

$$\mathbb{E}(Z_n^2) = \mathbb{E}(X_{1,1}^2 \mid |X_{1,1}| < (\log n)^{2/\alpha}) = \frac{\mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)}{\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha})} \,.$$

But thanks to hypothesis (1.1), $X_{1,1}^2$ is integrable, so by the dominated convergence theorem, $\mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)$ tends to $\mathbb{E}(X_{1,1}^2)$ as $n \to +\infty$. Besides, $\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha})$ tends to 1, so $\mathbb{E}(Z_n^2)$ tends to $\mathbb{E}(X_{1,1}^2)$ as $n \to +\infty$.

The same arguments show that $\mathbb{E}(Z_n^4)$ tends to $\mathbb{E}(X_{1,1}^4)$ as $n \to +\infty$. We can deduce that there exists a real number ζ such that

$$\sup_{n \in \mathbb{N}} \max \left(\mathbb{E}(Z_n^2), (\mathbb{E}(Z_n^2))^2, \mathbb{E}(Z_n^4) \right) \leq \zeta \,.$$

Moreover, we have

$$\sigma_n^2 = \operatorname{Var}(X_{1,1} \mid |X_{1,1}| < (\log n)^{2/\alpha})$$

$$= \frac{\mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)}{\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha})} - \left(\frac{\mathbb{E}\left(X_{1,1} \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)}{\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha})} \right)^2 \,.$$

Using similar arguments, we prove that σ_n^2 tends to $\operatorname{Var}(X_{1,1}) = 1$ as $n \to +\infty$. More precisely, we can write

$$
\begin{aligned}
\sigma_n^2 - 1 &= \frac{\mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)} - \left(\frac{\mathbb{E}\left(X_{1,1} \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)} \right)^2 \\[4pt]
&\quad - \mathbb{E}(X_{1,1}^2) + (\mathbb{E}(X_{1,1}))^2 \\[4pt]
&= \frac{\mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right) - \mathbb{E}(X_{1,1}^2) \, \mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)} \\[4pt]
&\quad + \frac{\mathbb{E}(X_{1,1})^2 \, \mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2 - \left(\mathbb{E}\left(X_{1,1} \, \mathbf{1}_{|X_{1,1}| < (\log n)^{2/\alpha}} \right) \right)^2}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2} \\[4pt]
&= \frac{\mathbb{E}(X_{1,1}^2) \, \mathbb{P}\left(|X_{1,1}| \geq (\log n)^{2/\alpha} \right) - \mathbb{E}\left(X_{1,1}^2 \, \mathbf{1}_{|X_{1,1}| \geq (\log n)^{2/\alpha}} \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)} \\[4pt]
&\quad + \frac{\mathbb{E}(X_{1,1})^2 \left(\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2 - 1 \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2} \\[4pt]
&\quad + \frac{2 \, \mathbb{E}(X_{1,1}) \, \mathbb{E}\left(X_{1,1} \, \mathbf{1}_{|X_{1,1}| \geq (\log n)^{2/\alpha}} \right)}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2} - \frac{\mathbb{E}\left(X_{1,1} \, \mathbf{1}_{|X_{1,1}| \geq (\log n)^{2/\alpha}} \right)^2}{\mathbb{P}\left(|X_{1,1}| < (\log n)^{2/\alpha} \right)^2} \quad (3.5)
\end{aligned}
$$

where $\left(\mathbb{E}\left(X_{1,1}\, \mathbf{1}_{|X_{1,1}|<(\log n)^{2/\alpha}} \right) \right)^2 = \left(\mathbb{E}(X_{1,1}) - \mathbb{E}\left(X_{1,1}\, \mathbf{1}_{|X_{1,1}|\geq(\log n)^{2/\alpha}} \right) \right)^2$ was used to get the last equality.

From hypothesis (1.1), for n large enough, we have

$$\mathbb{P}(|X_{1,1}| \geq (\log n)^{2/\alpha}) \leq e^{-\frac{a}{2}(\log n)^2}.$$

Besides,

$$\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha})^2 - 1 = -\mathbb{P}(|X_{1,1}| \geq (\log n)^{2/\alpha}) \left(\mathbb{P}(|X_{1,1}| < (\log n)^{2/\alpha}) + 1 \right)$$

and by the Cauchy-Schwarz inequality, we have

$$\left| \mathbb{E}\left(X_{1,1}\, \mathbf{1}_{|X_{1,1}|\geq(\log n)^{2/\alpha}} \right) \right| \leq \mathbb{E}(X_{1,1}^2)^{1/2}\, \mathbb{P}(|X_{1,1}| \geq (\log n)^{2/\alpha})^{1/2}$$

and

$$\left| \mathbb{E}\left(X_{1,1}^2\, \mathbf{1}_{|X_{1,1}|\geq(\log n)^{2/\alpha}} \right) \right| \leq \mathbb{E}(X_{1,1}^4)^{1/2}\, \mathbb{P}(|X_{1,1}| \geq (\log n)^{2/\alpha})^{1/2}.$$

Going back to (3.5), we have for n large enough

$$|\sigma_n^2 - 1| \leq 2\,\mathbb{E}(X_{1,1}^2)e^{-\frac{a}{2}(\log n)^2} + 2\,\mathbb{E}(X_{1,1}^4)^{1/2}e^{-\frac{a}{4}(\log n)^2} + 4\,\mathbb{E}(X_{1,1})^2 e^{-\frac{a}{2}(\log n)^2}$$
$$+ 4|\mathbb{E}(X_{1,1})|\,\mathbb{E}(X_{1,1}^2)^{1/2}e^{-\frac{a}{4}(\log n)^2} + 2\,\mathbb{E}(X_{1,1}^2)e^{-\frac{a}{2}(\log n)^2}.$$

Because the moments of $X_{1,1}$ are finite, we can deduce that there exists a real number η such that

$$|\sigma_n^2 - 1| \leq \eta e^{-a(\log n)^2/4}. \qquad \square$$

We can now prove Proposition 3.1.

Proof of Proposition 3.1. The proof relies on a conditioning with respect to the entries of X which are not in A and on a coupling argument to remove the dependency between A and C.

We use here the same notations as [7]. We denote by $\mathcal{F}$ the σ-algebra

$$\mathcal{F} = \sigma \left\{ X_{j,k}\, \mathbf{1}_{|X_{j,k}|\geq(\log n)^{2/\alpha}} \right\},$$

by $\mathbb{P}_{\mathcal{F}}$ and $\mathbb{E}_{\mathcal{F}}$ the probability and the expectation conditionally to $\mathcal{F}$, and we denote by E and F the events

$$E = \left\{ \frac{1}{n}\,\mathrm{Tr}(CC^t) \leq (\log n)^2 \right\}$$

and

$$F = \left\{ |I| < n^{1+\alpha/2} \right\},$$

with $I = \{(j,k) \mid |X_{j,k}| \geq (\log n)^{2/\alpha}\}$. Thus, the matrix C is $\mathcal{F}$-measurable and the events E and F belong to $\mathcal{F}$. Moreover, from Lemma 3.4 (i)–(ii), we have

$$\lim_{n\to+\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(E^c) = -\infty \quad \text{and} \quad \lim_{n\to+\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(F^c) = -\infty. \qquad (3.6)$$

The idea to prove Proposition 3.1 is to get the successive exponential equivalences

$$\mu_{XX^t/p} \approx \mu_{(A+C)(A+C)^t} \approx \mathbb{E}_{\mathcal{F}}\, \mu_{(A+C)(A+C)^t} \approx \mathbb{E}_{\mathcal{F}}\, \mu_{(A'+C)(A'+C)^t}$$
$$= \mathbb{E}_{\mathcal{F}}\, \mu_{\left(\frac{Y}{\sqrt{p}}+C\right)\left(\frac{Y}{\sqrt{p}}+C\right)^t} \approx \mathbb{E}_{\mathcal{F}}\, \mu_{\left(\frac{Y}{\sigma_n\sqrt{p}}+C\right)\left(\frac{Y}{\sigma_n\sqrt{p}}+C\right)^t} \approx \left(\sqrt{\mu_{CC^t}}\, \boxplus_c\, \sqrt{\mu_{\mathrm{MP},c}} \right)^2,$$

where A' will be defined below. Let us recall that the first one has been obtained in Lemma 3.3.

• Conditionally to $\mathcal{F}$, $\sqrt{p}A$ is a random matrix with independent entries bounded by $(\log n)^{2/\alpha}$. From the concentration result (A.15) applied to $Y = \sqrt{p}A$, $M = C$, $\kappa = (\log n)^{2/\alpha}$ and using that $\alpha < 2$, we get for all $\delta > 0$ and n large enough,

$$\mathbf{1}_E\, \mathbb{P}_{\mathcal{F}} \left(d_{s,t} \left(\mu_{(A+C)(A+C)^t}, \mathbb{E}_{\mathcal{F}}\, \mu_{(A+C)(A+C)^t} \right) \geq \delta \right) \leq \frac{\beta(\log n)^{2/\alpha}}{\delta^{3/2}} \exp\left(-\frac{n^2\delta^5}{\beta(\log n)^{8/\alpha}} \right)$$

hence

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}\left(E \cap \left\{ d_{s,t}\left(\mu_{(A+C)(A+C)^t}, \mathbb{E}_{\mathcal{F}}\, \mu_{(A+C)(A+C)^t} \right) \geq \delta \right\} \right) = -\infty\,. \quad (3.7)$$

• We will now use a coupling argument. We consider an independent random matrix Y whose entries are i.i.d. with distribution P_n defined in Lemma 3.4, and we denote by A' the matrix defined by

$$A'_{j,k} = \mathbf{1}_{(j,k)\notin I}\, A_{j,k} + \mathbf{1}_{(j,k)\in I}\, \frac{Y_{j,k}}{\sqrt{p}}\,.$$

Consequently, $\sqrt{p}A'$ and Y have the same distribution and are independent from $\mathcal{F}$. In particular, we will use later that for all bounded continuous f, we have $\mathbb{E}_{\mathcal{F}}(f(Y)) = \mathbb{E}(f(Y))$.

From the inequalities (B.11) and $d_{s,t} \leq W_2$, we have

$$d_{s,t}\big(\mu_{(A+C)(A+C)^t}, \mu_{(A'+C)(A'+C)^t}\big)^4$$
$$\leq \frac{2}{n^2} \operatorname{Tr}((A-A')(A-A')^t) \operatorname{Tr}((A+C)(A+C)^t + (A'+C)(A'+C)^t)$$
$$= \frac{2}{n^2} \left(\sum_{j,k} \mathbf{1}_{(j,k)\in I}\, \frac{Y_{j,k}^2}{p} \right) \left(\sum_{j,k} (A_{j,k}+C_{j,k})^2 + (A'_{j,k}+C_{j,k})^2 \right)$$
$$= \frac{2}{n^2 p} \left(\sum_{(j,k)\in I} Y_{j,k}^2 \right) \left(\sum_{j,k} A_{j,k}^2 + 2C_{j,k}^2 + (A'_{j,k})^2 \right)$$
$$= \frac{2}{n^2 p} \left(\sum_{(j,k)\in I} Y_{j,k}^2 \right) \left(\sum_{j,k} A_{j,k}^2 + 2\operatorname{Tr}(CC^t) + \sum_{j,k} A_{j,k}^2 + \sum_{(j,k)\in I} \frac{Y_{j,k}^2}{p} \right)$$
$$\leq \frac{2}{n^2 p} \left(2 \left(\sum_{(j,k)\in I} Y_{j,k}^2 \right) \left(np\frac{(\log n)^{4/\alpha}}{p} + \operatorname{Tr}(CC^t) \right) + \frac{1}{p} \left(\sum_{(j,k)\in I} Y_{j,k}^2 \right)^2 \right).$$

With definition (1.5) of $d_{s,t}$ and conditional Jensen's inequality for the concave function $x \mapsto x^{1/4}$, we thus have

$$\mathbf{1}_E\, \mathbf{1}_F\, d_{s,t}\left(\mathbb{E}_{\mathcal{F}}\, \mu_{(A+C)(A+C)^t}, \mathbb{E}_{\mathcal{F}}\, \mu_{(A'+C)(A'+C)^t} \right)$$
$$\leq \mathbf{1}_E\, \mathbf{1}_F\, \mathbb{E}_{\mathcal{F}}\, d_{s,t}\left(\mu_{(A+C)(A+C)^t}, \mu_{(A'+C)(A'+C)^t} \right)$$
$$\leq \left[\frac{2\,\mathbf{1}_E\, \mathbf{1}_F}{n^2 p} \left(2(n(\log n)^{4/\alpha} + \operatorname{Tr}(CC^t))\, \mathbb{E}_{\mathcal{F}} \left(\sum_{(j,k)\in I} Y_{j,k}^2 \right) + \frac{1}{p} \mathbb{E}_{\mathcal{F}} \left(\sum_{\substack{(j,k)\in I \\ (l,m)\in I}} Y_{j,k}^2 Y_{l,m}^2 \right) \right) \right]^{1/4}$$

because $\mathbf{1}_E$, $\mathbf{1}_F$, and $\operatorname{Tr}(CC^t)$ are $\mathcal{F}$-measurable. Since the events $\{(j,k) \in I\}$ are $\mathcal{F}$-measurable and Y is independent from $\mathcal{F}$, we have

$$\mathbb{E}_{\mathcal{F}} \left(\sum_{j,k} \mathbf{1}_{(j,k) \in I} \, Y_{j,k}^2 \right) = \sum_{j,k} \mathbf{1}_{(j,k) \in I} \, \mathbb{E}(Y_{j,k}^2) = |I| . \, \mathbb{E}(Y_{1,1}^2) \le \zeta |I|$$

from Lemma 3.4 (iii), and similarly

$$\mathbb{E}_{\mathcal{F}} \left(\sum_{j,k,l,m} \mathbf{1}_{(j,k) \in I} \, \mathbf{1}_{(l,m) \in I} \, Y_{j,k}^2 Y_{l,m}^2 \right) \le \zeta |I|^2 .$$

So we have

$$\mathbf{1}_E \, \mathbf{1}_F \, d_{s,t} \left(\mathbb{E}_{\mathcal{F}} \, \mu_{(A+C)(A+C)^t}, \mathbb{E}_{\mathcal{F}} \, \mu_{(A'+C)(A'+C)^t} \right)$$
$$\le \quad \left[\frac{2\zeta \, \mathbf{1}_E \, \mathbf{1}_F}{n^2 p} \left(2(n(\log n)^{4/\alpha} + \mathrm{Tr}(CC^t))|I| + \frac{1}{p}|I|^2 \right) \right]^{1/4}$$
$$\le \quad \left[\frac{2\zeta c_n}{n^3} \left(2 \left(n(\log n)^{4/\alpha} + n(\log n)^2 \right) n^{1+\alpha/2} + c_n n^{1+\alpha} \right) \right]^{1/4}$$
$$\le \quad \left[\frac{2\zeta c_n}{n^3} . 3n(\log n)^{4/\alpha} n^{1+\alpha/2} \right]^{1/4}$$
$$= \quad (6\zeta c_n)^{1/4} \frac{(\log n)^{1/\alpha}}{n^{1/4-\alpha/8}}$$

for n large enough (we used here the fact that $\frac{4}{\alpha} > 2$). It follows that for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P} \left(E \cap F \right.$$
$$\left. \cap \left\{ d_{s,t} \left(\mathbb{E}_{\mathcal{F}} \, \mu_{(A+C)(A+C)^t}, \mathbb{E}_{\mathcal{F}} \, \mu_{(A'+C)(A'+C)^t} \right) \ge \delta \right\} \right) = -\infty . \quad (3.8)$$

We recall here that $\mathbb{E}_{\mathcal{F}} \, \mu_{(A'+C)(A'+C)^t} = \mathbb{E}_{\mathcal{F}} \, \mu_{\left(\frac{Y}{\sqrt{p}} + C \right) \left(\frac{Y}{\sqrt{p}} + C \right)^t} .$

• In addition, from (B.11), we have

$$d_{s,t} \left(\mu_{\left(\frac{Y}{\sqrt{p}} + C \right) \left(\frac{Y}{\sqrt{p}} + C \right)^t}, \mu_{\left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t} \right)^4$$
$$\le \quad \frac{2}{n^2} \mathrm{Tr} \left(\left(1 - \frac{1}{\sigma_n} \right)^2 \frac{YY^t}{p} \right)$$
$$\times \mathrm{Tr} \left(\left(\frac{Y}{\sqrt{p}} + C \right) \left(\frac{Y}{\sqrt{p}} + C \right)^t + \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t \right)$$
$$\le \quad \frac{2}{n^2 p} \left(1 - \frac{1}{\sigma_n} \right)^2 \mathrm{Tr}(YY^t) \left(\sum_{j,k} \left(\frac{Y_{j,k}}{\sqrt{p}} + C_{j,k} \right)^2 + \left(\frac{Y_{j,k}}{\sigma_n \sqrt{p}} + C_{j,k} \right)^2 \right)$$
$$\le \quad \frac{2}{n^2 p} \left(1 - \frac{1}{\sigma_n} \right)^2 \mathrm{Tr}(YY^t) \left(4 \, \mathrm{Tr}(CC^t) + \frac{2}{p} \left(1 + \frac{1}{\sigma_n^2} \right) \mathrm{Tr}(YY^t) \right)$$

so, using conditional Jensen's inequality and doing as above, we get

$$\mathbf{1}_E \, \mathbb{E}_{\mathcal{F}} \, d_{s,t} \left(\mu_{\left(\frac{Y}{\sqrt{p}} + C \right) \left(\frac{Y}{\sqrt{p}} + C \right)^t}, \mu_{\left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t} \right)$$
$$\le \quad \left[\frac{2}{n^2 p} \left(1 - \frac{1}{\sigma_n} \right)^2 \left(4n(\log n)^2 \, \mathbb{E}_{\mathcal{F}} \, \mathrm{Tr}(YY^t) + \frac{2}{p} \left(1 + \frac{1}{\sigma_n^2} \right) \mathbb{E}_{\mathcal{F}}(\mathrm{Tr}(YY^t)^2) \right) \right]^{1/4}$$

$$\leq \quad \left[\frac{2}{n^2 p} \left(\frac{\sigma_n^2 - 1}{\sigma_n(\sigma_n + 1)} \right)^2 \left(4n(\log n)^2 . np\zeta + \frac{2}{p} \left(1 + \frac{1}{\sigma_n^2} \right) . n^2 p^2 \zeta \right) \right]^{1/4}$$

$$= \quad \left(\frac{\sigma_n^2 - 1}{\sigma_n(\sigma_n + 1)} \right)^{1/2} \left(8\zeta(\log n)^2 + 4\zeta \left(1 + \frac{1}{\sigma_n^2} \right) \right)^{1/4} .$$

By Lemma 3.4 (iii), we deduce from it that for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P} \left(E \cap \right.$$
$$\left. \left\{ d_{s,t} \left(\mathbb{E}_{\mathcal{F}} \mu_{\left(\frac{Y}{\sqrt{p}} + C \right) \left(\frac{Y}{\sqrt{p}} + C \right)^t}, \mathbb{E}_{\mathcal{F}} \mu_{\left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t} \right) \geq \delta \right\} \right) = -\infty. \quad (3.9)$$

• To finish, we define $\sigma_n^2 = \mathrm{Var}(Y_{1,1})$ as in Lemma 3.4 (iii). Since C is $\mathcal{F}$-measurable, Y is independent from $\mathcal{F}$, and $\frac{1}{\sigma_n^4} \mathbb{E}_{\mathcal{F}}(Y_{1,1}^4) \leq 2\zeta < +\infty$ for n large enough, we can apply Theorem 1.6 to Y/σ_n and C, conditionally to $\mathcal{F}$. Therefore, for n large enough, s large enough, and t small enough, we have

$$\mathbf{1}_E \, d_{s,t} \left(\mathbb{E}_{\mathcal{F}} \mu_{\left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t}, \left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}} \right)^2 \right)$$

$$\leq \quad c_{s,t} \, \mathbf{1}_E \left(\frac{1}{\sigma_n^3} \mathbb{E} |\mathring{Y}_{1,1}|^3 + \frac{1}{\sigma_n^4} \mathbb{E}(\mathring{Y}_{1,1}^4) \right) \left(\frac{1}{\sqrt{n}} + \frac{\mathrm{Tr}(CC^t)^{1/2}}{n} \right)$$

$$+ c_{s,t} \, \mathbf{1}_E \left(|c_n - c| + \frac{1}{n} + \frac{\mathrm{Tr}(CC^t)^{1/2}}{n^{5/4}} \right)$$

$$\leq \quad c_{s,t} \left(\frac{8(\log n)^{6/\alpha}}{\sigma_n^3} + \frac{16(\log n)^{8/\alpha}}{\sigma_n^4} \right) \left(\frac{1}{\sqrt{n}} + \frac{\log n}{\sqrt{n}} \right)$$

$$+ c_{s,t} \left(|c_n - c| + \frac{1}{n} + \frac{\log n}{n^{3/4}} \right)$$

using that for all $j \in [\![1, n]\!]$ and $k \in [\![1, p]\!]$, we have $|\mathring{Y}_{j,k}| = |Y_{j,k} - \mathbb{E}_{\mathcal{F}}(Y_{j,k})| \leq 2(\log n)^{2/\alpha}$. Therefore, for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P} \left(E \cap \right.$$
$$\left. \left\{ d_{s,t} \left(\mathbb{E}_{\mathcal{F}} \mu_{\left(\frac{Y}{\sigma_n \sqrt{p}} + C \right) \left(\frac{Y}{\sigma_n \sqrt{p}} + C \right)^t}, \left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}} \right)^2 \right) \geq \delta \right\} \right) = -\infty. \quad (3.10)$$

• To conclude, combining equalities from (3.6) to (3.9), Lemma 3.3, and Lemma 1.2.15 in [9], for s large enough and t small enough, we have for all $\delta > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P} \left(d_{s,t} \left(\mu_{XX^t/p}, \left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}} \right)^2 \right) \geq \delta \right) = -\infty. \qquad \square$$

3.2 Large deviations for $\mu_{C'}$

In the previous subsection, we proved that $\mu_{XX^t/p}$ and $\left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}} \right)^2$ are exponentially equivalent. Consequently, to obtain the large deviations of $\mu_{XX^t/p}$ (Theorem 1.7), it is sufficient to study the large deviations of μ_{CC^t} and to apply the contraction principle (see [9, Theorem 4.2.1]). To this, in this subsection, we will study the large deviations of

$$C' = \left(\begin{array}{c|c} 0 & C \\ \hline C^t & 0 \end{array} \right)$$

and prove the following, from which we will deduce the large deviations of μ_{CC^t} thanks to identity (3.13) and conclude in the next subsection.

Proposition 3.5. *The measure $\mu_{C'}$ satisfies the LDP with speed $n^{1+\alpha/2}$ in $\mathcal{P}(\mathbb{R})$, for weak topology and good rate function Φ' defined by*

$$\Phi'(\mu) = \begin{cases} \frac{a}{2} \frac{c+1}{c^{1+\alpha/2}} m_\alpha(\mu) & \text{if } \mu \text{ is symmetric and } \mu(\{0\}) \geq \frac{|1-c|}{1+c} \\ +\infty & \text{otherwise} \end{cases}, \qquad (3.11)$$

where $m_\alpha(\mu) = \int_{\mathbb{R}} |x|^\alpha \, d\mu(x)$ denotes the α-th moment of μ.

Note that Φ' is a good rate function because it is well known that for all $m \geq 0$ and $p > 0$, the set

$$K_{p,m} = \left\{ \mu \in \mathcal{P}(\mathbb{R}) \; \middle| \; \int_{\mathbb{R}} |x|^p \, d\mu(x) \leq m \right\} \qquad (3.12)$$

is compact for weak topology. Moreover, the domain of Φ' can be explained thanks to Lemma 3.6 (i).

Lemma 3.6. *Let $M \in \mathcal{M}_{n,p}(\mathbb{R})$ and*

$$M' = \left(\begin{array}{c|c} 0 & M \\ \hline M^t & 0 \end{array} \right).$$

(i) *The distribution $\mu_{M'}$ is symmetric and $\mu_{M'}(\{0\}) \geq \frac{|1-c_n|}{1+c_n}$.*

(ii) *We have*

$$\mu_{M'}^2 = \frac{2c_n}{c_n + 1} \mu_{MM^t} + \frac{1-c_n}{1+c_n} \delta_0. \qquad (3.13)$$

(iii) *If M is diagonal, in the sense that only the entries $M_{j,j}$, $1 \leq j \leq n \wedge p$, can be non-zero, then*

$$\mu_{M'} = \frac{1}{n+p} \sum_{j=1}^{n \wedge p} \left(\delta_{M_{j,j}} + \delta_{-M_{j,j}} \right) + \frac{|1-c_n|}{1+c_n} \delta_0. \qquad (3.14)$$

The proof of this lemma relies on the identity

$$(M')^2 = \left(\begin{array}{c|c} MM^t & 0 \\ \hline 0 & M^t M \end{array} \right).$$

Since it does not present any difficulty, it is left to the reader.

We also need a second lemma, which consists in two estimates for the distribution of $X_{1,1}$. These estimates come from the particular form of this distribution, see hypotheses (1.1) and (1.2).

Lemma 3.7. *(i) There exists a sequence $(\eta_n)_{n \in \mathbb{N}}$ converging to 0 such that for all $x \geq \varepsilon(n)$, we have*

$$\mathbb{P}(|X_{1,1}| \geq x\sqrt{p}) \leq e^{-(a-\eta_n)x^\alpha p^{\alpha/2}}. \qquad (3.15)$$

(ii) *We denote by S_a the support of the distribution ϑ_a defined by (1.2). There exists a sequence $(a_n)_{n \in \mathbb{N}}$ converging to a such that for all $x \in \mathbb{R}$ satisfying $|x| \geq \varepsilon(n)$ and $\operatorname{sign}(x) \in S_a$, for all $\gamma > 0$, and for all n large enough, we have*

$$\mathbb{P}\left(\frac{X_{1,1}}{\sqrt{p}} \in (x - \gamma, x + \gamma) \right) \geq e^{-a_n |x|^\alpha p^{\alpha/2}}. \qquad (3.16)$$

The computations to get these inequalities are explained in [7, p. 26] and are left to the reader.

We will now prove Proposition 3.5. Let us mention that Schatten's inequality (B.12) will be crucial in the proof since it will allow to link the α-th moment of the spectral measure $\mu_{C'}$ and the entries of C'.

From now on, we denote by $B_{s,t}(\mu, \delta)$ the ball with centre $\mu \in \mathcal{P}(\mathbb{R})$ and radius $\delta > 0$ for the distance $d_{s,t}$.

Proof of Proposition 3.5. Since the set of symmetric probability measures on $\mathbb{R}$ is closed for weak topology, it is enough to prove the LDP on this set, see [9, Lemma 4.1.5].

Upper bound. Let μ be a symmetric probability measure on $\mathbb{R}$. Since the function m_α is lower semi-continuous, there exists a continuous function h such that $h(0) = 0$ and

$$\mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \leq \mathbb{P}(m_\alpha(\mu_{C'}) \geq m_\alpha(\mu) - h(\delta))$$

for all δ small enough. Moreover, by Schatten's inequality (B.12) and the fact that $\sum_{i=1}^k a_i^r \leq \left(\sum_{i=1}^k a_i\right)^r$ for all $r \geq 1$, $a_1, \ldots, a_k \geq 0$, we have

$$m_\alpha(\mu_{C'}) \leq \frac{1}{n+p} \sum_{j=1}^{n+p} \left(\sum_{k=1}^{n+p} |C'_{j,k}|^2\right)^{\alpha/2} \leq \frac{1}{n+p} \sum_{j=1}^{n+p} \sum_{k=1}^{n+p} |C'_{j,k}|^\alpha.$$

Consequently,

$$
\begin{aligned}
&\mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \\
&\leq\ \mathbb{P}\left(\frac{1}{n+p} \sum_{j,k} |C'_{j,k}|^\alpha \geq m_\alpha(\mu) - h(\delta)\right) \\
&=\ \mathbb{P}\left(\frac{2}{(c_n+1)p^{1+\alpha/2}} \sum_{j,k} |X_{j,k}|^\alpha \mathbf{1}_{\varepsilon(n)\sqrt{p} < |X_{j,k}| \leq \varepsilon(n)^{-1}\sqrt{p}} \geq m_\alpha(\mu) - h(\delta)\right) \\
&\leq\ e^{-\frac{a_1}{2}(c_n+1)p^{1+\alpha/2}(m_\alpha(\mu) - h(\delta))} \left(\mathbb{E}\left(e^{a_1|X_{1,1}|^\alpha \mathbf{1}_{\varepsilon(n)\sqrt{p} < |X_{j,k}| \leq \varepsilon(n)^{-1}\sqrt{p}}}\right)\right)^{np}
\end{aligned}
$$

for all $a_1 \in (0, a)$ by Chernoff's inequality. Besides, from hypothesis (1.1), there exists $a_2 \in (a_1, a)$ such that for all x large enough, $\mathbb{P}(|X_{1,1}| \geq x) \leq \exp(-a_2 x^\alpha)$. Let us also recall the following integration by parts formula: for all $\nu \in \mathcal{P}(\mathbb{R})$ and $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$,

$$\int_a^b f(x) \, d\nu(x) = f(a)\nu([a, +\infty)) - f(b)\nu([b, +\infty)) + \int_a^b f'(x)\nu([x, +\infty)) \, dx.$$

Denoting by $\mathbb{P}_{|X_{1,1}|}$ the law of $|X_{1,1}|$, we thus have for n large enough,

$$
\begin{aligned}
&\mathbb{E}\exp\left(a_1 |X_{1,1}|^\alpha \mathbf{1}_{\varepsilon(n)\sqrt{p} < |X_{1,1}| \leq \varepsilon(n)^{-1}\sqrt{p}}\right) \\
&\leq\ 1 + \int_{\varepsilon(n)\sqrt{p}}^{\varepsilon(n)^{-1}\sqrt{p}} e^{a_1 x^\alpha} \, d\mathbb{P}_{|X_{1,1}|}(x) \\
&\leq\ 1 + e^{a_1\varepsilon(n)^\alpha p^{\alpha/2} - a_2\varepsilon(n)^\alpha p^{\alpha/2}} + \int_{\varepsilon(n)\sqrt{p}}^{\varepsilon(n)^{-1}\sqrt{p}} a_1 \alpha x^{\alpha-1} e^{a_1 x^\alpha - a_2 x^\alpha} \, dx \\
&\leq\ 1 + e^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}} - \frac{a_1}{a_1 - a_2} e^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}} \\
&=\ 1 + \frac{a_2}{a_2 - a_1} e^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}} \\
&\leq\ \exp\left(\frac{a_2}{a_2 - a_1} e^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}}\right)
\end{aligned}
$$

hence

$$
\begin{aligned}
&\mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \\
&\leq \exp\left(-\frac{a_1}{2}(c_n+1)p^{1+\alpha/2}(m_\alpha(\mu) - h(\delta)) + \frac{a_2}{a_2 - a_1} np e^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}}\right).
\end{aligned}
$$

So, for all δ small enough and all $a_1 \in (0, a)$,

$$\limsup_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \leq -\frac{a_1}{2} \frac{c+1}{c^{1+\alpha/2}} (m_\alpha(\mu) - h(\delta))$$

and finally

$$\limsup_{\delta \to 0} \limsup_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \leq -\frac{a}{2} \frac{c+1}{c^{1+\alpha/2}} m_\alpha(\mu) . \qquad (3.17)$$

In the case of a μ satisfying $\mu(\{0\}) < \frac{|1-c|}{1+c}$, we have a better result. Indeed, inspired by [16], we can observe that for all ε small enough, there exists $R > 0$ such that

$$\mu([-R, R]) < \frac{|1-c|}{1+c} - \varepsilon .$$

Therefore,

$$\left\{ \mu' \in \mathcal{P}(\mathbb{R}) \mid \mu'([-R, R]) < \frac{|1-c|}{1+c} - \varepsilon \right\}$$

is a neighbourhood of μ in which, for n large enough, almost surely, $\mu_{C'}$ is not. So we have

$$\lim_{\delta \to 0} \lim_{n \to +\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) = -\infty . \qquad (3.18)$$

We have obtained the upper bound of the LDP.

Lower bound. Let $\mu \in \mathcal{P}(\mathbb{R})$ be a symmetric measure such that $\mu(\{0\}) \geq \frac{|1-c|}{1+c}$. There exists $\tilde{\mu} \in \mathcal{P}(\mathbb{R}_+)$ such that

$$\mu = \frac{|1-c|}{1+c} \delta_0 + \frac{1 \wedge c}{1+c} (\tilde{\mu} + (-\operatorname{Id})_\sharp \tilde{\mu}) ,$$

where $(-\operatorname{Id})_\sharp \tilde{\mu}$ denotes the push-forward of $\tilde{\mu}$ by $-\operatorname{Id}$.

We denote by $x_1, \ldots, x_{n \wedge p}$ the quantiles of $\tilde{\mu}$ of orders $\frac{1}{1+n \wedge p}, \ldots, \frac{n \wedge p}{1+n \wedge p}$, we also define $n_0 = \min\{j \in [\![1, n \wedge p]\!] \mid x_j \geq \varepsilon(n)\}$, and

$$M' = \left(\begin{array}{c|c} 0 & M \\ \hline M^t & 0 \end{array} \right)$$

with $M \in \mathcal{M}_{n,p}(\mathbb{R})$ defined by $M_{j,j} = x_j$ for all $j \in [\![n_0, n \wedge p]\!]$ and $M_{j,k} = 0$ otherwise.

From (3.14), we have

$$\mu_{M'} = \frac{1}{n+p} \sum_{j=1}^{n \wedge p} \left(\delta_{M_{j,j}} + \delta_{-M_{j,j}} \right) + \frac{|1-c_n|}{1+c_n} \delta_0 .$$

Besides,

$$m_\alpha(\mu) = \frac{2(1 \wedge c)}{1+c} \int_0^{+\infty} |x|^\alpha \, d\tilde{\mu}(x) \geq \frac{2(1 \wedge c)}{(1+c)(1+n \wedge p)} \sum_{j=1}^{n \wedge p} |M_{j,j}|^\alpha . \qquad (3.19)$$

Let us also remark that by construction, $d_{s,t}(\mu, \mu_{M'})$ tends to 0 as $n \to +\infty$.

Let $\delta > 0$. For n large enough, we thus have

$$d_{s,t}(\mu, \mu_{M'}) < \frac{\delta}{2} . \qquad (3.20)$$

Using respectively (3.20), the Hoffman-Wielandt inequality (B.10) with the inequality $d_{s,t} \leq W_2$, the independence of the $X_{j,k}$'s, the inequalities (3.15) and (3.16), the fact that $1 \leq n_0 \leq n \wedge p$, and (3.19), we get for n large enough,

$$\mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta))$$

$$\geq \mathbb{P}\left(\mu_{C'} \in B_{s,t}\left(\mu_{M'}, \frac{\delta}{2}\right)\right)$$

$$\geq \mathbb{P}\left(\frac{1}{n+p}\operatorname{Tr}((C'-M')^2) \leq \frac{\delta^2}{4}\right)$$

$$= \mathbb{P}\left(\sum_{j,k}(C_{j,k}-M_{j,k})^2 \leq \frac{\delta^2(n+p)}{8}\right)$$

$$\geq \mathbb{P}\left(\forall j \in [\![n_0, n \wedge p]\!],\ (C_{j,j}-M_{j,j})^2 \leq \frac{\delta^2(n+p)}{8(n \wedge p)}\right.$$

$$\left. \cap\ \forall(j,k)\ \text{different},\ C_{j,k}=0\right)$$

$$\geq \mathbb{P}\left(\forall j \in [\![n_0, n \wedge p]\!],\ \frac{X_{j,j}}{\sqrt{p}} \in \left(M_{j,j}-\sqrt{\frac{\delta^2(n+p)}{8(n \wedge p)}}, M_{j,j}+\sqrt{\frac{\delta^2(n+p)}{8(n \wedge p)}}\right)\right.$$

$$\left. \cap\ \forall(j,k)\ \text{different},\ |X_{j,k}| < \varepsilon(n)\sqrt{p}\right)$$

$$\geq \prod_{j=n_0}^{n \wedge p} e^{-a_n|M_{j,j}|^\alpha p^{\alpha/2}}\left(1-e^{-(a-\eta_n)\varepsilon(n)^\alpha p^{\alpha/2}}\right)^{np-(n \wedge p-n_0+1)}$$

$$\geq \frac{1}{2}\exp\left(-a_n \sum_{j=n_0}^{n \wedge p}|M_{j,j}|^\alpha p^{\alpha/2}\right)$$

$$\geq \frac{1}{2}\exp\left(-\frac{a_n}{2}p^{1+\alpha/2}\frac{(1+c)\left(\frac{1}{p}+c \wedge 1\right)}{1 \wedge c}m_\alpha(\mu)\right)$$

Note that we can apply (3.15) and (3.16), even if it means to swap $\tilde{\mu}$ and $(-\operatorname{Id})_\sharp\tilde{\mu}$ in order to apply (3.16). We finally get

$$\liminf_{n \to +\infty} \frac{1}{n^{1+\alpha/2}}\log\mathbb{P}(\mu_{C'} \in B_{s,t}(\mu, \delta)) \geq -\frac{a}{2}\frac{c+1}{c^{1+\alpha/2}}m_\alpha(\mu) \tag{3.21}$$

for all $\delta > 0$.

This is the lower bound of the LDP.

Exponential tightness. Let $A > 0$ and $m = \frac{2Ac^{1+\alpha/2}}{a(1+c)}$. We recall that the set $K_{\alpha,m}$ defined by (3.12) is compact. Moreover, using the computations done for the upper bound, we have

$$\mathbb{P}(\mu_{C'} \notin K_{\alpha,m}) = \mathbb{P}(m_\alpha(\mu_{C'}) > m)$$

$$\leq \exp\left(-\frac{a_1}{2}(c_n+1)p^{1+\alpha/2}m + \frac{a_2}{a_2-a_1}npe^{-(a_2-a_1)\varepsilon(n)^\alpha p^{\alpha/2}}\right)$$

for all $a_1 \in (0, a)$ and some $a_2 \in (a_1, a)$. It follows that

$$\limsup_{n \to +\infty} \frac{1}{n^{1+\alpha/2}}\log\mathbb{P}(\mu_{C'} \notin K_{\alpha,m}) \leq -\frac{a}{2}\frac{c+1}{c^{1+\alpha/2}}m = -A. \tag{3.22}$$

The combination of (3.17), (3.18), (3.21), and (3.22) is the desired LDP. $\qquad \square$

3.3 Conclusion

To conclude this section, we show how to deduce the LDP for $\mu_{XX^t/p}$ (Theorem 1.7) from the LDP for $\mu_{C'}$ (Proposition 3.5).

Proposition 3.8. *The measure μ_{CC^t} satisfies the LDP with speed $n^{1+\alpha/2}$ in $\mathcal{P}(\mathbb{R}_+)$, for weak topology and the good rate function Ψ' defined by*

$$\Psi'(\nu) = \begin{cases} \frac{a}{c^{\alpha/2}} m_{\alpha/2}(\nu) & \text{if } \nu(\{0\}) \geq \max\left(0, 1 - \frac{1}{c}\right) \\ +\infty & \text{otherwise} \end{cases} \tag{3.23}$$

Proof. We define

$$T_n : \mu \mapsto \frac{1}{2}\left(1 + \frac{1}{c_n}\right)\mu^2 + \frac{1}{2}\left(1 - \frac{1}{c_n}\right)\delta_0$$

and

$$T : \mu \mapsto \frac{1}{2}\left(1 + \frac{1}{c}\right)\mu^2 + \frac{1}{2}\left(1 - \frac{1}{c}\right)\delta_0,$$

so that $\mu_{CC^t} = T_n(\mu_{C'})$, see (3.13).

Besides, we have

$$\lim_{n \to +\infty} d_{s,t}(\mu_{CC^t}, T(\mu_{C'})) = 0. \tag{3.24}$$

Indeed, let $n \in \mathbb{N}$ and z in the domain $V_{s,t}$ defined by (1.6). We have

$$\begin{aligned}
\left|G_{\mu_{CC^t}}(z) - G_{T(\mu_{C'})}(z)\right| &= \left|\int_{\mathbb{R}} \frac{1}{z-x} d(T_n(\mu_{C'}))(x) - \int_{\mathbb{R}} \frac{1}{z-x} d(T(\mu_{C'}))(x)\right| \\
&= \left|\frac{1}{2}\left(\frac{1}{c_n} - \frac{1}{c}\right)\int_{\mathbb{R}} \frac{1}{z-x} d(\mu_{C'}^2)(x) + \frac{1}{2}\left(\frac{1}{c} - \frac{1}{c_n}\right)\frac{1}{z}\right| \\
&\leq \frac{1}{|\operatorname{Im} z|}\left|\frac{1}{c_n} - \frac{1}{c}\right|,
\end{aligned}$$

so, taking the upper bound on $z \in V_{s,t}$ and the limit as $n \to +\infty$, we get (3.24).

The contraction principle applied to the function T, see [9, Theorem 4.2.1], will allow us to conclude. Indeed, T takes its values in $\mathcal{P}(\mathbb{R}_+)$ and is continuous for weak topology. This strategy will make appear the good rate function Ψ' defined for all $\nu \in \mathcal{P}(\mathbb{R}_+)$ by

$$\Psi'(\nu) = \inf\{\Phi'(\mu), \ \mu \in \mathcal{P}(\mathbb{R}) \text{ s.t. } \nu = T(\mu)\}.$$

For all $\mu \in \mathcal{P}(\mathbb{R})$, we have $(T(\mu))(\{0\}) = \frac{1}{2}\left(1 + \frac{1}{c}\right)\mu(\{0\}) + \frac{1}{2}\left(1 - \frac{1}{c}\right)$, hence

$$\mu(\{0\}) \geq \frac{|1-c|}{1+c} \quad \Leftrightarrow \quad (T(\mu))(\{0\}) \geq \frac{1}{2}\frac{|1-c| + (c-1)}{c} = \max\left(0, 1 - \frac{1}{c}\right).$$

Therefore, for all $\nu \in \mathcal{P}(\mathbb{R}_+)$ such that $\nu(\{0\}) \geq \max\left(0, 1 - \frac{1}{c}\right)$, there exists a symmetric $\mu \in \mathcal{P}(\mathbb{R})$ satisfying $\mu(\{0\}) \geq \frac{|1-c|}{1+c}$ and $\nu = T(\mu)$. We have in this case

$$\Phi'(\mu) = \frac{a}{2}\frac{c+1}{c^{1+\alpha/2}}\int_{\mathbb{R}} |x|^\alpha \, d\mu(x) = \frac{a}{c^{\alpha/2}}\int_{\mathbb{R}} |x|^{\alpha/2} \, d(T(\mu))(x) = \frac{a}{c^{\alpha/2}} m_{\alpha/2}(\nu)$$

hence $\Psi'(\nu) = \frac{a}{c^{\alpha/2}} m_{\alpha/2}(\nu)$. In the case where $\nu(\{0\}) < \max\left(0, 1 - \frac{1}{c}\right)$, we can not find a symmetric $\mu \in \mathcal{P}(\mathbb{R})$ satisfying $\mu(\{0\}) \geq \frac{|1-c|}{1+c}$ and $\nu = T(\mu)$, so $\Psi'(\nu) = +\infty$. Thus, we have computed $\Psi'(\nu)$ for every $\nu \in \mathcal{P}(\mathbb{R}_+)$.

Lower bound. Let $\mu \in \mathcal{P}(\mathbb{R}_+)$ and $\delta > 0$. From (3.24), for n large enough, we have $d_{s,t}(\mu_{CC^t}, T(\mu_{C'})) \leq \frac{\delta}{2}$, hence

$$\mathbb{P}(\mu_{CC^t} \in B_{s,t}(\mu, \delta)) \geq \mathbb{P}\left(T(\mu_{C'}) \in B_{s,t}\left(\mu, \frac{\delta}{2}\right)\right).$$

By Proposition 3.5 and the contraction principle, we thus have

$$\liminf_{n\to+\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{CC^t} \in B_{s,t}(\mu,\delta)) \geq - \inf_{\nu \in B_{s,t}(\mu,\delta/2)} \Psi'(\nu) \geq -\Psi'(\mu). \qquad (3.25)$$

Upper bound. Let F be a closed subset of $\mathcal{P}(\mathbb{R}_+)$ and $\delta > 0$. From (3.24), for n large enough, we have $d_{s,t}(\mu_{CC^t}, T(\mu_{C'})) \leq \delta$, so

$$\mathbb{P}(\mu_{CC^t} \in F) \leq \mathbb{P}(T(\mu_{C'}) \in F^\delta),$$

where F^δ denotes the δ-neighbourhood of F for the distance $d_{s,t}$, namely

$$F^\delta = \{\nu \in \mathcal{P}(\mathbb{R}_+) \mid \exists \mu \in F, \, d_{s,t}(\mu,\nu) \leq \delta\}.$$

Applying the contraction principle again, we thus have

$$\limsup_{n\to+\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{CC^t} \in F) \leq - \inf_{\nu \in F^\delta} \Psi'(\nu).$$

This is true for all $\delta > 0$ so, taking the limit as $\delta \to 0$, we get (see [9, Lemma 4.1.6(a)])

$$\limsup_{n\to+\infty} \frac{1}{n^{1+\alpha/2}} \log \mathbb{P}(\mu_{CC^t} \in F) \leq - \inf_{\nu \in F} \Psi'(\nu). \qquad (3.26)$$

Combining (3.25) and (3.26), we can conclude that μ_{CC^t} satisfies the announced LDP. $\qquad \square$

Because rectangular free convolution is continuous for weak topology, see [5, Theorem 3.12], the function $\mu \mapsto \left(\sqrt{\mu} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ is so, therefore, by Proposition 3.8 and the contraction principle, $\left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ satisfies the LDP with speed $n^{1+\alpha/2}$ on $\mathcal{P}(\mathbb{R}_+)$ governed by the good rate function

$$J'(\mu) = \begin{cases} \Psi'(\nu) & \text{if there exists } \nu \in \mathcal{P}(\mathbb{R}_+) \text{ such that } \mu = \left(\sqrt{\nu} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2 \\ +\infty & \text{otherwise} \end{cases}.$$

Thanks to the exponential equivalence between $\mu_{XX^t/p}$ and $\left(\sqrt{\mu_{CC^t}} \boxplus_c \sqrt{\mu_{\mathrm{MP},c}}\right)^2$ obtained in Proposition 3.1, we can conclude that $\mu_{XX^t/p}$ satisfies the same LDP, see [9, Theorem 4.2.13], which ends the proof of Theorem 1.7.

A Concentration bounds for the information-plus-noise model

A.1 Concentration for some functions of the resolvent

In Section 2, in order to prove Lemma 2.5, we needed the following concentration estimates.

Proposition A.1 (adaptation from [20, Lemma 8])**.** *Let $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ be a random matrix with i.i.d. entries, let $M \in \mathcal{M}_{n,p}(\mathbb{R})$ be a deterministic matrix and let $z \in \mathbb{C} \setminus \mathbb{R}$.*

We define $X = \frac{Y}{\sqrt{p}} + M$, $S = (zI_n - XX^t)^{-1}$ the resolvent of XX^t, $c_n = \frac{n}{p}$, and $\sigma^2 = \mathrm{Var}(Y_{1,1})$.

We assume that the distribution of $Y_{1,1}$ has mean zero and satisfy the following Poincaré inequality:

$$\forall f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}) \text{ s.t. } \mathbb{E}(f'(Y_{1,1})^2) < +\infty, \quad \mathrm{Var}(f(Y_{1,1})) \leq \sigma^2 \, \mathbb{E}(f'(Y_{1,1})^2).$$

Then, for all deterministic matrices $U \in \mathcal{M}_n(\mathbb{R})$ and $V \in \mathcal{M}_{n,p}(\mathbb{R})$, and for all integers n, p, we have

$$\mathrm{Var}\left(\frac{1}{n} \mathrm{Tr}(SU)\right) \leq \frac{4\sigma^2 c_n}{n^{5/2}} u(z) \|U\| (\mathrm{Tr}(UU^t))^{1/2} \qquad (A.1)$$

and

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}(X^t S V)\right) \leq \frac{9\sigma^2 c_n}{n^2} v(z)$$
$$\times \max\left(\|V\|\frac{(\mathrm{Tr}(VV^t))^{1/2}}{n^{1/2}}, \|V\|^{3/2}\frac{(\mathrm{Tr}(VV^t))^{1/4}}{n^{1/4}}, \|V\|^{5/4}\frac{(\mathrm{Tr}(VV^t))^{3/8}}{n^{3/8}}\right), \quad \text{(A.2)}$$

where

$$u(z) = \frac{|z|}{|\mathrm{Im}\,z|^4} + \frac{1}{|\mathrm{Im}\,z|^3}$$

and

$$v(z) = \max\left(\frac{1}{|\mathrm{Im}\,z|^2}, \frac{|z|}{|\mathrm{Im}\,z|^3} + \frac{1}{|\mathrm{Im}\,z|^2}, \left(\frac{|z|}{|\mathrm{Im}\,z|^2} + \frac{1}{|\mathrm{Im}\,z|}\right)^2\right).$$

Remark A.2.
- In the proof of Lemma 2.5, we apply (A.1) to $U = I_n$ and $U = R$, and we apply (A.2) to $V = M$.
- Since $\|V\| \leq \mathrm{Tr}(VV^t)^{1/2}$, (A.2) implies

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}(X^t S V)\right) \leq \frac{9\sigma^2 c_n}{n^{9/4}} v(z)\,\mathrm{Tr}(VV^t). \quad \text{(A.3)}$$

  Having in mind the large deviations in Section 3, we want to get a bound in $\mathrm{Tr}(MM^t)$ in Lemma 2.5, that is why we use (A.3) instead of (A.2) in its proof.
- We get here slightly better bounds than [20]. Indeed, we can recover their results from ours since $\mathrm{Tr}(AA^t) \leq n\|A\|^2$ (see Proposition B.1 (iv)). This improvement is due to the fact that we used the inequality $|\mathrm{Tr}(AC)| \leq \sqrt{n}\|A\|(\mathrm{Tr}(CC^*))^{1/2}$ (see Proposition B.1 (iii)) instead of $|\mathrm{Tr}(AC)| \leq n\|A\|.\|C\|$.
- If the distribution of $Y_{1,1}$ satisfies the Poincaré inequality with a constant C instead of σ^2, then σ^2 must be replaced by C in the bounds (A.1) and (A.2).
- In the case of complex matrices Y, M, U, V, the bounds are very similar and only the constants change.

Proof. Using the sub-additivity property of variance, the Poincaré inequality, and the differentiation formula (B.1), we get

$$\mathrm{Var}\left(\frac{1}{n}\mathrm{Tr}(SU)\right) = \mathrm{Var}\left(\sum_{1\leq j,k\leq n}\frac{1}{n}S_{j,k}U_{k,j}\right)$$
$$\leq \sigma^2\,\mathbb{E}\left[\sum_{a,b}\left(\frac{1}{n}\sum_{j,k}\frac{1}{\sqrt{p}}D_{a,b}S_{j,k}.U_{k,j}\right)^2\right]$$
$$= \sigma^2\,\mathbb{E}\left[\sum_{a,b}\frac{1}{n^2 p}\left(\sum_{j,k}(SX)_{j,b}S_{a,k}U_{k,j} + S_{j,a}(X^t S)_{b,k}U_{k,j}\right)^2\right]$$
$$= \frac{\sigma^2}{n^2 p}\,\mathbb{E}\left[\sum_{a,b}((SUSX)_{a,b} + (X^t SUS)_{b,a})^2\right]$$
$$= \frac{\sigma^2}{n^2 p}\,\mathbb{E}\left[\sum_{a,b}(SUSX)_{a,b}^2 + 2(SUSX)_{a,b}(X^t SUS)_{b,a} + (X^t SUS)_{b,a}^2\right]$$
$$= \frac{\sigma^2}{n^2 p}\,\mathbb{E}\left[\mathrm{Tr}(SUSX(SUSX)^t) + 2\,\mathrm{Tr}(SUSXX^t SUS) + \mathrm{Tr}(X^t SUS(X^t SUS)^t)\right].$$
$$\text{(A.4)}$$

Using the resolvent identity $SXX^t = XX^tS = zS - I_n$, the inequality $|\operatorname{Tr}(AB)| \leq \sqrt{n} \operatorname{Tr}(AA^*)^{1/2}\|B\|$ (see Proposition B.1 (iii)), and $\|S\| \leq \frac{1}{|\operatorname{Im} z|}$ (see Proposition B.2 (iv)), we get

$$\left|\operatorname{Tr}(SUSX(SUSX)^t)\right| = \left|\operatorname{Tr}(U(zS - I_n)SU^tS^2)\right|$$

$$\leq \sqrt{n}\|U\| \operatorname{Tr}(UU^t)^{1/2} \left(\frac{|z|}{|\operatorname{Im} z|^4} + \frac{1}{|\operatorname{Im} z|^3}\right) \tag{A.5}$$

and very similarly,

$$\left|\operatorname{Tr}(SUSXX^tSUS)\right| \leq \sqrt{n}\|U\| \operatorname{Tr}(UU^t)^{1/2} \left(\frac{|z|}{|\operatorname{Im} z|^4} + \frac{1}{|\operatorname{Im} z|^3}\right) \tag{A.6}$$

and

$$\left|\operatorname{Tr}(X^tSUS(X^tSUS)^t)\right| \leq \sqrt{n}\|U\| \operatorname{Tr}(UU^t)^{1/2} \left(\frac{|z|}{|\operatorname{Im} z|^4} + \frac{1}{|\operatorname{Im} z|^3}\right). \tag{A.7}$$

Combining (A.4), (A.5), (A.6) and (A.7), we conclude that

$$\operatorname{Var}\left(\frac{1}{n}\operatorname{Tr}(SU)\right) \leq \frac{4\sigma^2 c_n}{n^{5/2}} \left(\frac{1}{|\operatorname{Im} z|^3} + \frac{|z|}{|\operatorname{Im} z|^4}\right) \|U\| \operatorname{Tr}(UU^t)^{1/2}.$$

Let us now prove the second inequality. By the same arguments as above, we get

$$\operatorname{Var}\left(\frac{1}{n}\operatorname{Tr}(X^tSV)\right) = \operatorname{Var}\left(\sum_{\substack{1 \leq k,l \leq n \\ 1 \leq j \leq p}} \frac{1}{n}X_{k,j}S_{k,l}V_{l,j}\right)$$

$$\leq \sigma^2 \mathbb{E}\left[\sum_{a,b} \left(\frac{1}{n}\sum_{j,k,l} D_{a,b}X_{k,j}.S_{k,l}V_{l,j} + X_{k,j}.\frac{1}{\sqrt{p}}D_{a,b}S_{k,l}.V_{l,j}\right)^2\right]$$

$$= \frac{\sigma^2}{n^2 p} \mathbb{E}\left[\sum_{a,b} \left((SV)_{a,b} + (SVX^tSX)_{a,b} + (X^tSVX^tS)_{b,a}\right)^2\right]$$

$$= \frac{\sigma^2}{n^2 p} \mathbb{E}\left[\operatorname{Tr}(SV(SV)^t) + \operatorname{Tr}(SVX^tSX(SVX^tSX)^t) + \operatorname{Tr}(X^tSVX^tS(X^tSVX^tS)^t)\right.$$

$$\left. +2\operatorname{Tr}(SVX^tSX(SV)^t) + 2\operatorname{Tr}(SVX^tSVX^tS) + 2\operatorname{Tr}(SVX^tSXX^tSVX^tS)\right]. \tag{A.8}$$

We will now bound these terms, always using the resolvent identities $SXX^t = XX^tS = zS - I_n$, $XX^tS^* = S^*XX^t = \bar{z}S^* - I_n$, inequalities (i)–(iii) in Proposition B.1, and the bound $\|S\| \leq \frac{1}{|\operatorname{Im} z|}$. We get for example

$$\left|\operatorname{Tr}(SV(SV)^t)\right| = \left|\operatorname{Tr}(SVV^tS)\right| \leq \sqrt{n}\|V\| \operatorname{Tr}(VV^t)^{1/2}\frac{1}{|\operatorname{Im} z|^2} \tag{A.9}$$

and

$$\left|\operatorname{Tr}(SVX^tSX(SVX^tSX)^t)\right|$$
$$= \left|\operatorname{Tr}(V^tS^2VX^t(zS - I_n)SX)\right|$$
$$\leq \operatorname{Tr}(V^tS^2VV^t(S^*)^2V)^{1/2} \operatorname{Tr}(X^t(zS - I_n)SXX^tS^*(\bar{z}S^* - I_n)X)^{1/2}$$
$$= \operatorname{Tr}(V^tS^2VV^t(S^*)^2V)^{1/2} \operatorname{Tr}(S^*(\bar{z}S^* - I_n)XX^t(zS - I_n)(zS - I_n))^{1/2}$$
$$\leq \left(\sqrt{n}\|V\|^3\|S\|^4 \operatorname{Tr}(VV^t)^{1/2}\right)^{1/2} \operatorname{Tr}((\bar{z}S^* - I_n)^2(zS - I_n)^2)^{1/2}$$
$$\leq n^{1/4}\frac{1}{|\operatorname{Im} z|^2}\|V\|^{3/2} \operatorname{Tr}(VV^t)^{1/4}.\sqrt{n}\left(\frac{|z|}{|\operatorname{Im} z|} + 1\right)^2$$
$$= n^{3/4}\|V\|^{3/2} \operatorname{Tr}(VV^t)^{1/4} \left(\frac{|z|}{|\operatorname{Im} z|^2} + \frac{1}{|\operatorname{Im} z|}\right)^2. \tag{A.10}$$

Very similarly, we get

$$\left|\mathrm{Tr}(X^t S V X^t S (X^t S V X^t S)^t)\right| \leq n^{3/4} \|V\|^{3/2} \mathrm{Tr}(VV^t)^{1/4} \left(\frac{|z|}{|\mathrm{Im}\, z|^2} + \frac{1}{|\mathrm{Im}\, z|} \right)^2, \quad \text{(A.11)}$$

$$\left|\mathrm{Tr}(S V X^t S X V^t S)\right| \leq n^{5/8} \|V\|^{5/4} \mathrm{Tr}(VV^t)^{3/8} \left(\frac{|z|}{|\mathrm{Im}\, z|^3} + \frac{1}{|\mathrm{Im}\, z|^2} \right), \quad \text{(A.12)}$$

$$\left|\mathrm{Tr}(S V X^t S V X^t S)\right| \leq \sqrt{n} \|V\| \mathrm{Tr}(VV^t)^{1/2} \left(\frac{|z|}{|\mathrm{Im}\, z|^3} + \frac{1}{|\mathrm{Im}\, z|^2} \right), \quad \text{(A.13)}$$

and finally

$$\left|\mathrm{Tr}(S V X^t S X X^t S V X^t S)\right| \leq \sqrt{n} \|V\| \mathrm{Tr}(VV^t)^{1/2} \left(\frac{|z|}{|\mathrm{Im}\, z|^2} + \frac{1}{|\mathrm{Im}\, z|} \right)^2. \quad \text{(A.14)}$$

Combining inequalities from (A.8) to (A.14), we finally obtain the announced bound. □

Remark A.3. In the proof above, it is possible to improve some majorizations using the inequality $\|SX\| \leq \frac{1}{|\mathrm{Im}\, y|}$ where y is a square root of z (see Proposition B.2 (v)). For instance, it allows to get

$$|\mathrm{Tr}(X^t S V X^t S (X^t S V X^t S)^t)| \leq \sqrt{n} \,\mathrm{Tr}(VV^t)^{1/2} \|V\| \cdot \frac{1}{|\mathrm{Im}\, y|^4}$$

instead of (A.11). However, in (A.10), which is the other dominant term in (A.8), we can not improve the power of n by this strategy.

A.2 Concentration of the empirical spectral measure

In Section 3, in order to prove Proposition 3.1, we needed the following concentration bound.

Proposition A.4 (adaptation from [7, Theorem 2.5]). *Let* $\kappa > 1$, $Y \in \mathcal{M}_{n,p}(\mathbb{R})$ *a random matrix with i.i.d. entries bounded by* κ, $M \in \mathcal{M}_{n,p}(\mathbb{R})$ *a deterministic matrix such that* $\frac{1}{n} \mathrm{Tr}(MM^t) \leq \kappa^2$, *and* $s, t > 0$. *We assume that* $c_n = \frac{n}{p} \to c \in (0, +\infty)$ *as* $n \to +\infty$.

There exists $\beta > 0$ *such that for all* s *large enough,* t *small enough,* n *large enough, and* $\delta \in \left[\left(\frac{\beta \kappa^2}{n} \right)^{2/5}, 1 \right]$, *we have*

$$\mathbb{P}\left(d_{s,t}\left(\mu_{(Y/\sqrt{p}+M)(Y/\sqrt{p}+M)^t}, \mathbb{E}\,\mu_{(Y/\sqrt{p}+M)(Y/\sqrt{p}+M)^t} \right) \geq \delta \right) \leq \frac{\beta \kappa}{\delta^{3/2}} \exp\left(-\frac{n^2 \delta^5}{\beta \kappa^4} \right). \quad \text{(A.15)}$$

Remark A.5. • Here κ is a constant but we are interested in the dependence on κ in the bound since we apply (A.15) to a κ depending on n in Section 3.
  • This result remains true if Y and M are complex matrices and the entries of Y have independent real and imaginary parts.

Proof. We will apply [15, Theorem 1.3(b)] to the $(n+p) \times (n+p)$ matrix

$$\mathbf{X}_A = \left(\begin{array}{c|c} 0 & \frac{Y}{\sqrt{p}} + M \\ \hline \left(\frac{Y}{\sqrt{p}} + M \right)^t & 0 \end{array} \right).$$

The matrix M is not present in [15] but it is possible to do so because

$$\frac{1}{n+p} \mathrm{Tr}(\mathbf{X}_A^2) \leq \frac{2}{n} \sum_{j,k} \left(\frac{Y_{j,k}}{\sqrt{p}} + M_{j,k} \right)^2 \leq \frac{4}{np} \sum_{j,k} Y_{j,k}^2 + \frac{4}{n} \mathrm{Tr}(MM^t)$$

so, thanks to the hypotheses on Y and M, we have $\frac{1}{n+p}\operatorname{Tr}(\mathbf{X}_A^2) \leq 8\kappa^2$. Therefore, the argument in [15, p. 132] does not change and we can apply [15, Theorem 1.3(b)] adding the matrix M. Consequently, there exists $\beta > 0$ such that for all n large enough and $\delta \in \left[\left(\frac{\beta\kappa^2}{n}\right)^{2/5}, 1\right]$, we have

$$\mathbb{P}\left(\sup_f \left|\int f\, d\mu_{\mathbf{X}_A} - \mathbb{E}\int f\, d\mu_{\mathbf{X}_A}\right| \geq \delta\right) \leq \frac{\beta\kappa}{\delta^{3/2}} e^{-n^2\delta^5/\beta\kappa^4}$$

where the supremum is taken over all bounded Lipschitz functions f such that

$$\sup_{x\in\mathbb{R}} |f(x)| + \sup_{x\neq y}\left|\frac{f(x) - f(y)}{x - y}\right| \leq 1. \tag{A.16}$$

Moreover, using (B.7), we can check that when s is large enough and t is small enough, for every $z \in V_{s,t}$, the function $f : x \mapsto \frac{1}{z-x^2}$ is Lipschitz, bounded, and satisfies (A.16). Noting in addition that

$$\int \frac{1}{z-x^2}\, d\mu_{\mathbf{X}_A}(x) = \frac{1}{n+p}\left(2n\int \frac{1}{z-x}\, d\mu_{(Y/\sqrt{p}+M)(Y/\sqrt{p}+M)^t}(x) + (n-p).\frac{1}{z}\right)$$

and using the definition (1.5) of $d_{s,t}$, we find (A.15) even if it means to change β. $\qquad\square$

B Technical tools

In this appendix, we summarize miscellaneous results used throughout the paper.

B.1 Traces and matricial norms inequalities

For a matrix $A \in \mathcal{M}_{n,p}(\mathbb{C})$, we denote by $\|A\|$ its operator norm associated to Euclidean norms and

$$\|A\|_\infty = \max_{1\leq j\leq n, 1\leq k\leq p} |A_{j,k}|.$$

If B is an other matrix in $\mathcal{M}_{n,p}(\mathbb{C})$, we denote by $A \circ B$ the Hadamard product of A and B, i.e. the matrix defined by $(A \circ B)_{j,k} = A_{j,k}B_{j,k}$. Finally, $\operatorname{diag}(A)$ denotes the matrix whose entries are given by $A_{j,k}\delta_{j,k}$, where δ is the Kronecker delta.

Proposition B.1. *Let* $A, B \in \mathcal{M}_{n,p}(\mathbb{C})$, $C \in \mathcal{M}_{p,n}(\mathbb{C})$, $D \in \mathcal{M}_n(\mathbb{C})$, $E \in \mathcal{M}_{p,q}(\mathbb{C})$. *We have the following.*

(i) $|\operatorname{Tr}(AC)| \leq (\operatorname{Tr}(AA^*))^{1/2}(\operatorname{Tr}(CC^*))^{1/2}$,

(ii) $|\operatorname{Tr}(AC)| \leq n\|A\|.\|C\|$,

(iii) $|\operatorname{Tr}(AC)| \leq \sqrt{n}\|A\|(\operatorname{Tr}(CC^*))^{1/2}$,

(iv) $\|A\| \leq (\operatorname{Tr}(AA^*))^{1/2} \leq \sqrt{n}\|A\|$,

(v) $(\operatorname{Tr}(AA^*))^{1/2} \leq \sqrt{np}\|A\|_\infty$,

(vi) $\|\operatorname{diag}(D)\| = \|\operatorname{diag}(D)\|_\infty \leq \|D\|_\infty$,

(vii) $\|A \circ B\|_\infty \leq \|A\|_\infty\|B\|_\infty$,

(viii) $\|A \circ B\| \leq \|A\|.\|B\|$,

(ix) $\operatorname{Tr}((A \circ B)(A \circ B)^*) \leq \operatorname{Tr}(AA^*)\|B\|_\infty^2$.

Most of these points are classical or easy to check. Note that the combination (iii) of (i) and (ii) will be crucial for us and that a proof of (viii) requires the use of the matrices

$$A' = \begin{pmatrix} A_{1,1} & & A_{1,2} & & A_{1,p} & \\ & \ddots & & \ddots & \cdots & \ddots \\ & A_{n,1} & & A_{n,2} & & A_{n,p} \end{pmatrix}$$

and

$$
B' = \begin{pmatrix}
B_{1,1} & & & & \\
\vdots & & & & \\
B_{n,1} & & & & \\
& B_{1,2} & & & \\
& \vdots & & & \\
& B_{n,2} & & & \\
& & \ddots & & \\
& & & B_{1,p} & \\
& & & \vdots & \\
& & & B_{n,p} &
\end{pmatrix}.
$$

B.2 Properties of resolvents

Let $A \in \mathcal{H}_n(\mathbb{C})$ and $z \in \mathbb{C} \setminus \mathbb{R}$. The *resolvent of A at z* is the matrix $R(A) = (zI_n - A)^{-1}$. For $A \in \mathcal{M}_{n,p}(\mathbb{C})$, we denote by $S(A)$ the resolvent $R(AA^*)$, or just S if no confusion can arise.

Proposition B.2. *Let $A, B \in \mathcal{M}_{n,p}(\mathbb{C})$ and $z \in \mathbb{C} \setminus \mathbb{R}$. We have the following.*

(i) $SAA^* = AA^*S = zS - I_n$,

(ii) $S(A + B) - S(A) = S(A + B)(AB^* + BA^* + BB^*)S(A)$,

(iii) $G_{\mu_{AA^*}}(z) = \frac{1}{n} \operatorname{Tr} S$,

(iv) $\|S\|_\infty \le \|S\| \le \frac{1}{|\operatorname{Im} z|}$,

(v) $\|SA\|_\infty \le \|SA\| \le \frac{1}{|\operatorname{Im} y|}$, *where y is a square root of z,*

(vi) $\|A^*SA\|_\infty \le \|A^*SA\| \le 1 + \left|\frac{z}{\operatorname{Im} z}\right|$.

(vii) *We denote by $D_{a,b}$ the derivation w.r.t. $\operatorname{Re} A_{a,b}$ and by δ the Kronecker delta. For all $a, j, k \in [\![1, n]\!]$ and $b, l, m \in [\![1, p]\!]$, we have*

$$
D_{a,b}S_{j,k} = (SA)_{j,b}S_{a,k} + S_{j,a}(A^*S)_{b,k}, \tag{B.1}
$$

$$
D_{a,b}(SA)_{j,l} = (SA)_{j,b}(SA)_{a,l} + S_{j,a}(A^*SA)_{b,l} + \delta_{b,l}S_{j,a}, \tag{B.2}
$$

$$
D_{a,b}(A^*S)_{l,k} = (A^*SA)_{l,b}S_{a,k} + (A^*S)_{l,a}(A^*S)_{b,k} + \delta_{b,l}S_{a,k}, \tag{B.3}
$$

$$
\begin{aligned}
D_{a,b}(A^*SA)_{l,m} = {} & (A^*SA)_{l,b}(SA)_{a,m} + (A^*S)_{l,a}(A^*SA)_{b,m} \\
& + \delta_{b,m}(A^*S)_{l,a} + \delta_{b,l}(SA)_{a,m},
\end{aligned} \tag{B.4}
$$

$$
\begin{aligned}
D_{a,b}^2 S_{j,k} = 2[& S_{j,a}S_{a,k} + (SA)_{j,b}(SA)_{a,b}S_{a,k} + S_{j,a}(A^*SA)_{b,b}S_{a,k} \\
& + S_{j,a}(A^*S)_{b,a}(A^*S)_{b,k} + (SA)_{j,b}S_{a,a}(A^*S)_{b,k}],
\end{aligned} \tag{B.5}
$$

$$
\begin{aligned}
D_{a,b}^3 S_{j,k} = {} & 6[(SA)_{j,b}S_{a,a}S_{a,k} + S_{j,a}(A^*S)_{b,a}S_{a,k} + S_{j,a}(SA)_{a,b}S_{a,k} \\
& + S_{j,a}S_{a,a}(A^*S)_{b,k} + (SA)_{j,b}(SA)_{a,b}^2 S_{a,k} + S_{j,a}(A^*S)_{b,a}^2(A^*S)_{b,k} \\
& + (SA)_{j,b}(SA)_{a,b}S_{a,a}(A^*S)_{b,k} + (SA)_{j,b}S_{a,a}(A^*S)_{b,a}(A^*S)_{b,k} \\
& + S_{j,a}(A^*SA)_{b,b}(SA)_{a,b}S_{a,k} + (SA)_{j,b}S_{a,a}(A^*SA)_{b,b}S_{a,k} \\
& + S_{j,a}(A^*S)_{b,a}(A^*SA)_{b,b}S_{a,k} + S_{j,a}(A^*SA)_{b,b}S_{a,a}(A^*S)_{b,k}].
\end{aligned} \tag{B.6}
$$

Rectangular asymptotic freeness and large deviations of random matrices

Most of these relations are classical or obtained by simple computations. Note however that (v) and (vi) respectively follow from the identities

$$
\left(\begin{array}{c|c} \overline{y}I_n & -A \\ \hline -A^* & \overline{y}I_p \end{array} \right)^{-1} = \left(\begin{array}{c|c} \overline{y}S^* & A(\overline{y}^2 I_p - A^*A)^{-1} \\ \hline (SA)^* & \overline{y}(\overline{y}^2 I_p - A^*A)^{-1} \end{array} \right)
$$

and $A^*(zI_n - AA^*)^{-1}A = -I_p + z(zI_p - A^*A)^{-1}$.

Note also that if y is a square root of z and z belongs to the domain $V_{s,t}$ defined by (1.6), then we can easily prove that

$$
|\operatorname{Im} y|^2 = \frac{\operatorname{Im} z}{2}\left(\sqrt{\frac{(\operatorname{Re} z)^2}{(\operatorname{Im} z)^2} + 1} - \frac{\operatorname{Re} z}{\operatorname{Im} z} \right) > \frac{s}{2}(\sqrt{t^2+1} - t) > 0. \tag{B.7}
$$

B.3 Inequalities for empirical spectral measures

Proposition B.3 (Rank inequality, see [7, Lemma B.1]). *Let $A, B \in \mathcal{H}_n(\mathbb{C})$. We have*

$$
d_{\mathrm{KS}}(\mu_A, \mu_B) \le \frac{1}{n}\operatorname{rank}(A - B). \tag{B.8}
$$

Proposition B.4 (Rank inequality for covariance matrices, see [3, Theorem A.44]). *Let $A, B \in \mathcal{M}_{n,p}(\mathbb{C})$. We have*

$$
d_{\mathrm{KS}}(\mu_{AA^*}, \mu_{BB^*}) \le \frac{1}{n}\operatorname{rank}(A - B). \tag{B.9}
$$

Proposition B.5 (Hoffman-Wielandt inequality, see [7, Lemma B.2]). *Let $A, B \in \mathcal{H}_n(\mathbb{C})$. We have*

$$
W_2^2(\mu_A, \mu_B) \le \frac{1}{n}\operatorname{Tr}((A - B)^2), \tag{B.10}
$$

where W_2 denotes the L^2-Wasserstein distance on $\mathcal{P}(\mathbb{R})$.

Proposition B.6 (see [3, Corollary A.42]). *Let $A, B \in \mathcal{M}_{n,p}(\mathbb{C})$. We have*

$$
W_2^4(\mu_{AA^*}, \mu_{BB^*}) \le \frac{2}{n^2}\operatorname{Tr}(AA^* + BB^*)\operatorname{Tr}((A - B)(A - B)^*). \tag{B.11}
$$

Proposition B.7 (Schatten's inequality, see [22, Theorem 3.32]). *Let $A \in \mathcal{H}_n(\mathbb{C})$ and $p \in (0, 2]$. We have*

$$
\int_{\mathbb{R}} |x|^p \, d\mu_A(x) \le \frac{1}{n}\sum_{k=1}^n \left(\sum_{j=1}^n |A_{k,j}|^2 \right)^{p/2}. \tag{B.12}
$$

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. MR-2760897

[2] F. Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. arXiv:1502.07983v1, 2015. MR-3492936

[3] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010. MR-2567175

[4] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997. MR-1465640

[5] F. Benaych-Georges. Rectangular random matrices, related convolution. *Probab. Theory Related Fields*, 144(3–4):471–515, 2009. MR-2496440

[6] P. Biane. On the free convolution with a semi-circular distribution. *Indiana Univ. Math. J.*, 46(3):705–718, 1997. MR-1488333

[7] C. Bordenave and P. Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.*, 42(6):2454–2496, 2014. MR-3265172

[8] M. Capitaine, C. Donati-Martin, D. Féral, and M. Février. Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab.*, 16(64):1750–1792, 2011. MR-2835253

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998. MR-1619036

[10] R. B. Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007. MR-2322123

[11] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach. *IEEE Trans. Inform. Theory*, 56(3):1048–1069, 2010. MR-2723661

[12] P. Eichelsbacher, J. Sommerauer, and M. Stolz. Large deviations for disordered bosons and multiple orthogonal polynomial ensembles. *J. Math. Phys.*, 52(7):073510, 16, 2011. MR-2849043

[13] A. Fey, R. van der Hofstad, and M. J. Klok. Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Adv. in Appl. Probab.*, 40(4):1048–1071, 2008. MR-2488532

[14] B. Groux. Grandes déviations de matrices aléatoires et Équation de Fokker-Planck libre. *PhD thesis*, Université Paris-Saclay, 2016.

[15] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000. MR-1781846

[16] F. Hiai and D. Petz. Eigenvalue density of the Wishart matrix and large deviations. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 1(4):633–646, 1998. MR-1665279

[17] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000. MR-1746976

[18] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72(114):507–536, 1967. MR-0208649

[19] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011. MR-2808038

[20] P. Vallet, P. Loubaton, and X. Mestre. Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case. *IEEE Trans. Inform. Theory*, 58(2):1043–1068, 2012. MR-2918009

[21] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958. MR-0095527

[22] X. Zhan. *Matrix inequalities*, volume 1790 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. MR-1927396