Concentration of the spectral measure for large matrices

A. GUIONNET[*], 0. ZEITOUNI [†]

2/1/2000

Abstract

We derive concentration inequalities for functions of the empirical measure of eigenvalues for large, random, self adjoint matrices, with not necessarily Gaussian entries. The results presented apply in particular to non-Gaussian Wigner and Wishart matrices. We also provide concentration bounds for non commutative functionals of random matrices.

1 Introduction and statement of results

Consider a random $N \times N$ Hermitian matrix $\mathbf{X}$ with i.i.d. complex entries (except for the symmetry constraint) satisfying a moment condition. It is well known since Wigner [28] that the spectral measure of $N^{-1/2}\mathbf{X}$ converges to the semicircle law. This observation has been generalized to a large class of matrices, e.g. sample covariance matrices of the form $\mathbf{XRX}^*$ where $\mathbf{R}$ is a deterministic diagonal matrix ([19]), band matrices (see [5, 16, 20]), etc. For the Wigner case, this convergence has been supplemented by Central Limit Theorems, see [15] for the case of Gaussian entries and [17], [22] for the general case.

Our goal in this paper is to study deviations beyond the central limit theorem regime. For Gaussian entries, full large deviation principles have been derived in [4] (for the Wigner case) and [14] (for Wishart matrices, that is sample covariance matrices with $\mathbf{R} = I$). These papers are based on the explicit form of the joint law of the eigenvalues, and this technique does not seem to extend to either non-Gaussian entries or even to general sample covariance matrices. Still in the Gaussian case, [12] derived upper bounds of large deviations type for the spectral measure of band matrices, including certain sample covariance matrices. Her derivation is based on the tools in [9] who considered more general non-commutative functionals of Wigner matrices.

We propose here a different approach to the estimation of deviations from the typical behaviour, which is based on the concentration inequalities developed in recent years by M. Talagrand ([23]-[25]). It is interesting to note that the crucial part of our argument, Lemma 1.2, uses computations which appear, in slightly different form, in the theory of free probability. Unlike the approach of [9] and [12], there is no hope to obtain by this technique complementary lower bounds, and the constants determining the rates are clearly not optimal. On the other hand, the Gaussian assumption can be dispensed with, and the derivation

[*]DMA, Ecole Normale Superieure, 45 rue d'Ulm, 75005 Paris, France

[†]Department of Electrical Engineering, Technion, Haifa 32000, Israel. This work was partially supported by a grant from the Israel Science Foundation, administered by the Israel Academy of Sciences

in general is much easier. It is also worthwhile to note that the speed of convergence of the probabilities of large deviations that we obtain is the same as in the Gaussian Wigner case. In particular, in certain cases it yields all the tightness needed for a CLT statement.

In the rest of this section, we define precisely the model of random matrices we consider, state concentration bounds for these models, and show how these imply exponential decay of Wasserstein distance between the empirical spectral measure $\hat{\mu}^N$ and its mean. We also present our main technical lemma, Lemma 1.2, which allows us to transfer properties of a function $f : \mathbb{R} \mapsto \mathbb{R}$ to properties of the function $\int f d\hat{\mu}^N$ viewed as a function of the entries of the random matrix. In Section 1.2, we combine our basic concentration estimates with known results on the convergence of $\hat{\mu}^N$ in some classical random matrix models to deduce exponential convergence to an N-independent limit. Section 1.3 contains concentration results for a class of non-commutative functionals of an ensemble of independent random matrices. Finally, Section 2 is devoted to the proofs of our results.

1.1 Concentration for large inhomogeneous random matrices

Let $\mathcal{M}^{sa}_{N\times N}(\mathbb{C})$ be the set of complex entries $N \times N$ self-adjoint matrices. Let f be a real valued function on $\mathbb{R}$. f can as well be seen as a function from $\mathcal{M}^{sa}_{N\times N}(\mathbb{C})$ into $\mathcal{M}^{sa}_{N\times N}(\mathbb{C})$ if we set, for $M \in \mathcal{M}^{sa}_{N\times N}(\mathbb{C})$ so that $M = UDU^*$ for a diagonal real matrix D and a unitary matrix U,

$$f(M) = Uf(D)U^*$$

where $f(D)$ is the diagonal matrix with entries $(f(D_{11}), .., f(D_{NN}))$ and U^* denotes the (complex) adjoint of U.

Let tr denote the trace on $\mathcal{M}^{sa}_{N\times N}(\mathbb{C})$ given by $\mathrm{tr}(M) = \sum_{i=1}^N M_{ii}$, and set $\mathrm{tr}_N(M) := N^{-1}\mathrm{tr}(M)$. We shall first consider the concentration of the real valued random variable $\mathrm{tr}_N(f(\mathbf{X}_A))$ for inhomogeneous random matrices given by

$$\mathbf{X}_A = ((\mathbf{X}_A)_{ij})_{1\le i,j\le N}, \quad \mathbf{X}_A = \mathbf{X}_A^*, \quad (\mathbf{X}_A)_{ij} = \frac{1}{\sqrt{N}} A_{ij}\omega_{ij}$$

with

$$\omega := (\omega^R + i\omega^I) = (\omega_{ij})_{1\le i,j\le N} = (\omega^R_{ij} + \sqrt{-1}\omega^I_{ij})_{1\le i,j\le N}, \omega_{ij} = \bar{\omega}_{ji}, \qquad A = (A_{ij})_{1\le i,j\le N}, A_{ij} = \bar{A}_{ji},$$

$\{\omega_{ij}, 1 \le i \le j \le N\}$ are independent complex random variables with laws $\{P_{ij}, 1 \le i \le j \le N\}$, P_{ij} being a probability measure on $\mathbb{C}$ with

$$P_{ij}(\omega_{ij} \in \cdot) = \int 1\!\!1_{u+iv\in\cdot} P^R_{ij}(du) P^I_{ij}(dv),$$

and A is a non-random complex matrix with entries $\{A_{ij}, 1 \le i \le j \le N\}$ uniformly bounded by, say, a. When needed, we shall write $\mathbf{X}_A = \mathbf{X}_A(\omega)$. We let $\Omega_N = \{(\omega^R_{ij}, \omega^I_{ij})\}_{1\le i\le j\le N}$, and denote by $\mathbb{P}^N$ the law $\mathbb{P}^N = \otimes_{1\le i\le j\le N}(P^R_{ij} \otimes P^I_{ij})$ on Ω_N, with $P^I_{ii} = \delta_0$, and set $\mathbb{E}^N$ to be the corresponding expectation.

We shall also use the following notations and definitions.

For a compact set K, denote by $|K|$ its diameter, that is the maximal distance between two points of K. For a Lipschitz function $f : I\!\!R^k \mapsto I\!\!R$, we define the Lipschitz constant $|f|_{\mathcal{L}}$ by

$$|f|_{\mathcal{L}} = \sup_{x,y} \frac{|f(x) - f(y)|}{||x - y||},$$

where here and throughout, $||\cdot||$ denotes the Euclidean norm on $I\!\!R^k$.

We say that a measure ν on $I\!\!R$ satisfies the logarithmic Sobolev inequality with (not necessarily optimal) constant c if, for any differentiable function f,

$$\int f^2 \log \frac{f^2}{\int f^2 d\nu} d\nu \leq 2c \int |f'|^2 d\nu.$$

Recall that a measure ν satisfying the logarithmic Sobolev inequality possesses sub-Gaussian tails ([18]). Recall also that the Gaussian law [11], any probability measure ν absolutely continuous with respect to Lebesgue measure satisfying the Bobkov and Götze [6] condition (including $\nu(dx) = Z^{-1}e^{-|x|^\alpha}dx$ for $\alpha \geq 2$), as well as any distribution absolutely continuous with respect to them possessing a bounded above and below density, satisfies the logarithmic Sobolev inequality [18, Section 7.1], [13, Propriété 4.5].

Our main result is

Theorem 1.1 *a)Assume that the $(P_{ij}, i \leq j, i,j \in I\!\!N)$ are uniformly compactly supported, that is that there exists a compact set $K \subset \mathbb{C}$ so that for any $1 \leq i \leq j \leq N$, $P_{ij}(K^c) = 0$. Assume f is convex and Lipschitz. Then, for any $\delta > \delta_0(N) := 8|K|\sqrt{\pi}a|f|_{\mathcal{L}}/N > 0$,*

$$I\!\!P^N \left(\omega : |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - I\!\!E^N[\mathrm{tr}_N(f(\mathbf{X}_A))]| \geq \delta\right) \leq 4e^{-\frac{1}{16|K|^2a^2|f|_{\mathcal{L}}^2}N^2(\delta-\delta_0(N))^2}.$$

b) If the $(P_{ij}^R, P_{ij}^I, 1 \leq i \leq j \leq N)$ satisfy the logarithmic Sobolev inequality with uniform constant c, then for any Lipschitz function f, for any $\delta > 0$,

$$I\!\!P^N \left(\omega : |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - I\!\!E^N[\mathrm{tr}_N(f(\mathbf{X}_A))]| \geq \delta\right) \leq 2e^{-\frac{1}{8ca^2|f|_{\mathcal{L}}^2}N^2\delta^2}.$$

Remark: Note that the concentration results stated in Theorem 1.1 above hold also in the case that $P_{ij}^I = \delta_0$ for all i,j, i.e. for inhomogeneous *real* Wigner matrices. Studying separately this case roughly improves the constants in the exponents by a factor 2. This comment applies to the other theorems in this paper.

The above theorem is based on Talagrand's results [24], and Ledoux [18]. To obtain it, we have to see $\mathrm{tr}_N f(\mathbf{X}_A)$ as a function of the entries $\{\omega_{ij}^R, \omega_{ij}^I, 1 \leq i \leq j \leq N\}$ and prove the following lemma, which is at the heart of our results.

Lemma 1.2 *a) If f is a real valued convex function on $I\!\!R$, it holds that $(\omega^R, \omega^I) \mapsto \mathrm{tr}(f(\mathbf{X}_A(\omega)))$ is convex. b) If f is a Lipschitz function on $I\!\!R$, $(\omega^R, \omega^I) \mapsto \mathrm{tr}(f(\mathbf{X}_A(\omega)))$ is a Lipschitz function on $I\!\!R^{N^2}$ with Lipschitz constant bounded by $2a|f|_{\mathcal{L}}$. If f is differentiable, we more precisely have*

$$\sum_{1\leq i\leq j\leq N} (\partial_{\omega_{ij}^R} \mathrm{tr}(f(\mathbf{X}_A)))^2 + \sum_{1\leq i<j\leq N} (\partial_{\omega_{ij}^I} \mathrm{tr}(f(\mathbf{X}_A)))^2 \leq 4a^2|f|_{\mathcal{L}}^2.$$

In Theorem 1.1, we proved that for any compactly supported family of measures $(P_{i,j}, i \leq j)$, concentration was holding for convex functions, and obtained a sub-Gaussian speed for this concentration. Next, we push the argument to obtain concentration for non convex functions, up to a loss in the accuracy of the speed. Towards this end, let

$$||f||_{\mathcal{L}} = ||f||_\infty + |f|_{\mathcal{L}}, \quad \mathcal{F}_{\text{lip}} = \{f : ||f||_{\mathcal{L}} \leq 1\},$$

and for $\mathcal{K} \subset I\!R$ compact, let

$$\mathcal{F}_{\text{lip},\mathcal{K}} := \{f : \text{supp}(f) \subset \mathcal{K}, \ f \in \mathcal{F}_{\text{lip}}\}.$$

Theorem 1.3 *Let $(P_{i,j}, i \leq j)$ be compactly supported probability measures on $K \subset\subset I\!R$. Fix $\delta_1(N) = 8|K|\sqrt{\pi}a/N$, $S = \max_{z \in K} |z|^2$ and $M = \sqrt{8Sa^2}$.*
a) Let $\mathcal{K} \in I\!R$ be compact. Then, for any $\delta \geq 4\sqrt{|\mathcal{K}|\delta_1(N)}$,

$$I\!P^N \left(\omega : \sup_{f \in \mathcal{F}_{\text{lip},\mathcal{K}}} |\text{tr}_N(f(\mathbf{X}_A)) - I\!E^N[\text{tr}_N(f(\mathbf{X}_A))]| \geq \delta \right) \leq \frac{32|\mathcal{K}|}{\delta} \exp\left(-N^2 \frac{1}{16|K|^2 a^2} [\frac{\delta^2}{16|\mathcal{K}|} - \delta_1(N)]^2 \right). \tag{1.1}$$

b) For any $\delta > (128(M + \sqrt{\delta})\delta_1(N))^{2/5}$,

$$I\!P_A^N \left(\omega : \sup_{f \in \mathcal{F}_{\text{lip}}} |\text{tr}_N(f(\mathbf{X}_A)) - I\!E^N[\text{tr}_N(f(\mathbf{X}_A))]| \geq \delta \right)$$

$$\leq \frac{128(M + \sqrt{\delta})}{\delta^{3/2}} \exp\left(-N^2 \frac{1}{16|K|^2 a^2} [\frac{\delta^{5/2}}{128(M + \sqrt{\delta})} - \delta_1(N)]^2 \right).$$

The results above have direct implication on concentration for the empirical measure of eigenvalues with respect to the Wasserstein distance, given for any probability measures (μ, ν) on $I\!R$ by

$$d(\mu, \nu) = \sup_{f \in \mathcal{F}_{\text{lip}}} |\int f d\mu - \int f d\nu|.$$

Let $\hat{\mu}^N$ denote the empirical measure of the eigenvalues of $\mathbf{X}_A$, that is $\text{tr}_N(f(\mathbf{X}_A)) = \int f(x)\hat{\mu}^N(dx)$ for every bounded Borel f. We have the following immediate corollary.

Corollary 1.4 *a) With the assumptions and notations of Theorem 1.3 b), we have for $\delta > (128(M + \sqrt{\delta})\delta_1(N))^{2/5}$,*

$$I\!P^N(\omega : d(\hat{\mu}^N, \ I\!E^N\hat{\mu}^N) > \delta) \leq \frac{128(M + \sqrt{\delta})}{\delta^{3/2}} \exp\left(-N^2 \frac{1}{16|K|^2 a^2} [\frac{\delta^{5/2}}{128(M + \sqrt{\delta})} - \delta_1(N)]^2 \right).$$

b) Assume that $(P_{ij}, 1 \leq i, j \leq N)$ satisfy the logarithmic Sobolev inequality with uniform constant c. Then, there exist positive universal constants C_1 and C_2 so that for any $\delta > 0$,

$$I\!P^N \left(\omega : d(\hat{\mu}^N, \ I\!E^N(\hat{\mu}^N)) \geq \delta \right) \leq \frac{C_1}{\delta^{3/2}} e^{-\frac{C_2}{ca^2} N^2 \delta^5}.$$

Next, we also get the following extension.

Theorem 1.5 *Assume that $(P_{ij}, 1 \le i,j \le N)$ satisfy the logarithmic Sobolev inequality with uniform constant c. Then, for any probability measure μ on $\mathbb{R}$, for any $\delta > 0$,*

$$\mathbb{P}^N\left(\omega : d(\hat{\mu}^N, \mu) \ge \mathbb{E}^N\left(d(\hat{\mu}^N, \mu)\right) + \delta\right) \le e^{-\frac{N^2}{8ca^2}\delta^2}.$$

1.2 Concentration of the spectral measure for classical matrices

In the previous section, we obtained concentration of the marginals of the spectral measure for inhomogeneous Wigner's matrices for which we do not know a priori that $\mathbb{E}[\mathrm{tr}_N(f(\mathbf{X}_A))]$ converges. Here, we shall restrict ourselves to models for which such a result is already known including Wigner's matrices (with non necessarily Gaussian entries), band matrices, diluted matrices,Wishart's ensembles including inhomogeneous Wishart matrices. In such models, the previous result will allow us to deduce from already known L^1 convergence, an exponential speed for this convergence. Hereafter, we shall state results under the natural normalization

$$\int x dP_{ij}(x) = 0, \qquad \int |x|^2 dP_{ij}(x) = 1. \tag{1.2}$$

Wigner's matrices

First, let us concentrate on Wigner's matrices. In this case, we have $A_{ij} = 1$ for all i,j, and we write $\mathbf{X}_A = \mathbf{X}_1$. Then, it is well known that if the P_{ij}'s satisfy the moments conditions (see [2] for instance)

$$\sup_{ij} \int |x|^4 dP_{ij}(x) < \infty, \tag{1.3}$$

one has

$$\lim_{N\to\infty} \mathbb{E}^N(\mathrm{tr}_N(f)) = \int f(x) d\sigma(x)$$

with σ the semicircle law

$$\sigma(dx) = \frac{1}{2\pi}\sqrt{4 - x^2}1_{|x|\le 2} dx. \tag{1.4}$$

Here, we shall restrict ourselves according to the previous section (see the assumptions of Theorem 1.1) to entries with all moments finite. Further, Bai (see [2] for real entries and [3, Theorem 3.6] for the complex case) proved that if F_N is the distribution function $F_N(x) = \mathbb{E}^N(\mathrm{tr}_N(\mathbb{1}_{\mathbf{X}_1 \le x}))$, and F is the distribution function of the semicircle law (1.4), then there exists a finite constant C so that for any $N \in \mathbb{N}$,

$$\sup_{x\in\mathbb{R}} |F_N(x) - F(x)| \le C N^{-\frac{1}{4}}.$$

As a direct consequence, it is not hard to verify that if f is differentiable with $f' \in L^1(\mathbb{R})$,

$$\delta_N(f) := |\ \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_1))] - \int f(x) d\sigma(x)| \le C|f'|_{L^1(\mathbb{R})} N^{-\frac{1}{4}}.$$

Recall that it is known since Wigner [28] that for any $k \in \mathbb{N}$, $\mathbb{E}^N[\mathrm{tr}_N((\mathbf{X}_1^{2k}))]$ can be bounded independently of N. For general Lipschitz functions, one can use an approximating scheme to see that for any $\epsilon > 0$, there exists a finite constant $C(\epsilon) > 0$ such that

$$\delta_N(f) \le C(\epsilon) ||f||_{\mathcal{L}} N^{-\frac{1}{4}+\epsilon}.$$

We can therefore deduce from Theorem 1.1 the

Corollary 1.6 *a) Under the assumptions of Theorem 1.1 a) and (1.2), for any Lipschitz function f, for any $\delta > \delta_2(N) := 8|K|\sqrt{\pi}a|f|_{\mathcal{L}}/N + \delta_N(f) > 0$,*

$$\mathbb{P}^N\left(\omega : |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - \int f(x)d\sigma(x)| \geq \delta\right) \leq 4e^{-\frac{1}{16|K|^2a^2|f|_{\mathcal{L}}^2}N^2(\delta-\delta_2(N))^2}.$$

b) Under the assumptions of Theorem 1.1 b) and (1.2), for any Lipschitz function f, for any $\delta > \delta_N(f)$,

$$\mathbb{P}^N\left(\omega : |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - \int f(x)d\sigma(x)| \geq \delta\right) \leq 2e^{-\frac{1}{8ca^2|f|_{\mathcal{L}}^2}N^2(\delta-\delta_N(f))^2}.$$

Of course, Corollary 1.4 and Theorem 1.5 also provide exponential convergence of Wasserstein's distance between the spectral measure of Wigner's matrices and the semicircle law. We emphasize that they apply both for real and complex Wigner matrices.

Now we turn to more interesting consequences of the possible inhomogeneity in the matrices studied in the last section and first to the most natural

Band matrices Let us recall that band matrices are matrices of type $\mathbf{X}_A$ with $A_{i,j}$ null for $|i-j|$ big enough. More generally, one can consider the sometimes called generalized deformed Wigner ensemble described by a sequence $\mathbf{X}_{A^N} \in \mathcal{M}^{sa}_{N\times N}(\mathbb{C})$, $N \in \mathbb{N}$, so that there exists a non random bounded continuous function ϕ so that

$$\lim_{N\to\infty} \sup_{u\in[0,1]} \int_0^1 du|\sum_{i,j} A^N_{ij} 1\!\!1_{t\in[i/N,i+1/N[} 1\!\!1_{u\in[j/N,j+1/N[} - \phi(t,u)| = 0.$$

We denote by $\mathbf{X}_\phi$ the matrix $\mathbf{X}_A$ for such a choice of A. It was observed since [7, 10, 16, 20, 12] under various hypotheses that the empirical measure of $\mathbf{X}_\phi$ converges when the $(P_{ij}, 1 \leq i \leq j \leq N)$ satisfy (1.2) and have some uniformly bounded moments. Let us recall the result from [16, section 8.3] for real symmetric $\mathbf{X}_\phi$ (see [5, 20, 12] for Hermitian matrices) which as far as we know have the weakest hypotheses. Assuming that $\sup_{i,j} P_{i,j}[|x|^3]$ is finite, they found that if k_ϕ is the unique analytic function on $[0,1]\times\mathbb{C}\backslash\mathbb{R}$, so that

$$k_\phi(x,z) = (z - K_\phi(k_\phi(.,z))(x))^{-1}$$

where K_ϕ is the operator in $L^2([0,1])$ with kernel ϕ, for any $z \in \mathbb{C}\setminus\mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{E}^N[\mathrm{tr}_N\left((z-\mathbf{X}_\phi)^{-1}\right)] = m_\phi(z) := \int_0^1 k_\phi(x,z)dx. \tag{1.5}$$

Let σ_ϕ be the unique probability measure with Stieljes transform m_ϕ. Let us recall how the convergence of $\mathbb{E}^N(\mathrm{tr}_N(f(\mathbf{X}_\phi)))$ for Lipschitz functions f can be deduced from the convergence of the Stieljes transform. The main observation is that if m is the Stieljes transform of a probability measure μ, $m(z) = \int(z-x)^{-1}d\mu(x)$, $z \in \mathbb{C}\backslash\mathbb{R}$, then, with $z = u+iv$ and $\Im(z) = v$ denoting the imaginary part of z,

$$\pi^{-1}\Im(m(z)) = \int_{-\infty}^{+\infty} \frac{v}{\pi\left((x-u)^2+v^2\right)}\mu(dx) = \frac{dP_v * \mu}{du}(u) \tag{1.6}$$

with P_v the Cauchy law with parameter $v \neq 0$. Thus, for $f \in \mathcal{F}_{\text{lip}} \cap L^1(\mathbb{R})$,

$$
\begin{aligned}
&| \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_\phi))] - \int f d\sigma_\phi| \\
&\leq | \int \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_\phi + u))]P_v(du) - \int f dP_v * \sigma_\phi| + 4||f||_{\mathcal{L}} \int (|u| \wedge 1) P_v(du) \\
&= \pi^{-1} | \int f(u) \Im \left(\mathbb{E}^N[\mathrm{tr}_N((u + iv - \mathbf{X}_\phi)^{-1})] - m_\phi(u + iv)\right) du| + 4||f||_{\mathcal{L}} \int (|u| \wedge 1) P_v(du) \quad (1.7)
\end{aligned}
$$

It is now clear from (1.5) that the first term in the right hand side of (1.7) goes to zero as N goes to infinity by dominated convergence for $v \neq 0$ and then that the second is as small as one wishes when v goes to zero. Hence, using the a-priori moment bounds on $\mathbf{X}_\phi$, for any $f \in \mathcal{F}_{\text{lip}}$,

$$
\lim_{N \to \infty} \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_\phi))] = \int f d\sigma_\phi.
$$

Further, this convergence can be seen to be uniform over $\mathcal{F}_{\text{lip}}$ by considering more carefully the first integral in the right hand side of (1.7).

As a conclusion, for band-matrices, not only Theorems 1.1, 1.3, 1.5 and Corollary 1.4 hold but actually one has also the convergence of $\mathbb{E}^N[\mathrm{tr}_N f(\mathbf{X}_A)]$ (for $f \in \mathcal{F}_{\text{lip}}$) and of $\mathbb{E}^N \hat{\mu}^N$.

In the same spirit, one can consider

Diluted random matrices Diluted random matrices appear in diluted models of spin glasses. They are described as matrices $\mathbf{X}_A$ with A now random with $\{0,1\}$-valued entries. Assume for simplicity that $\{A_{ij}, 1 \leq i \leq j \leq N\}$ are i.i.d with law $Q = p\delta_0 + (1-p)\delta_1$ (p eventually depending on N) and denote $Q^N = \otimes_{1 \leq i \leq j \leq N} Q$ with E^N the corresponding expectation. Convergence of the empirical measure of $\mathbf{X}_A$ for such random matrices A is discussed in [16], p.193-197. We here want to stress the following concentration result

Corollary 1.7 *Assume that the $(P_{ij}, i \leq j, i, j \in \mathbb{N})$ are uniformly compactly supported, that is that there exists a compact set $K \subset \mathbb{C}$ so that for any $1 \leq i \leq j \leq N$, $P_{ij}(K^c) = 0$. Set $L = \max\{\sup_{x \in K} x^2, |K|^2\}$. Assume f is convex and Lipschitz. Then, for any $\delta > \delta_3(N) := 4(|K| + \sup_{x \in K} |x|)\sqrt{\pi}|f|_{\mathcal{L}}/N > 0$,*

$$
Q^N \otimes \mathbb{P}^N \left(\omega, \mathbf{A} : |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - E^N \ \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_A))]| \geq \delta\right) \leq 8e^{-\frac{1}{64L|f|_{\mathcal{L}}^2} N^2(\delta - \delta_3(N))^2}.
$$

Concentration for the Wasserstein distance can be derived as in Corollary 1.4.

We now turn to the well known

Wishart's matrices Let us recall that if $\mathbf{Y}$ is a $N \times M$ matrix, $N \leq M$, with independent entries $\omega_{kl} = (\omega_{kl}^R + i\omega_{kl}^I)$ of law P_{kl}, $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^*$ is a so-called Wishart's matrix. We let throughout $\mathbb{P}^{N,M} = \otimes_{1 \leq k \leq N, 1 \leq l \leq M} P_{kl}$. For the sake of completeness, let us consider inhomogeneous Wishart matrices given, for a diagonal real matrix $\mathbf{R} = (\lambda_1, ..\lambda_M)$ with $\lambda_i \geq 0$ by $\mathbf{Z} = \mathbf{Y}\mathbf{R}\mathbf{Y}^*$. To deduce the concentration of the spectral measure for such matrices from Theorem 1.1, note that if we consider

$$\begin{aligned} A_{ij} &= 0 \quad \text{for } 1 \le i \le N, 1 \le j \le N \\ A_{ij} &= 0 \quad \text{for } M+1 \le i \le M+N, M+1 \le j \le N+M \\ A_{ij} &= \sqrt{\lambda_{i-M}} \quad \text{for } N+1 \le i \le M+N, 1 \le j \le N \\ A_{ij} &= \sqrt{\lambda_{j-M}} \quad \text{for } 1 \le i \le N, N+1 \le j \le N+M \end{aligned}$$

then $A = A^*$ and if we consider $\mathbf{X}_A \in \mathcal{M}^{sa}_{(N+M)\times(N+M)}(\mathbb{C})$ constructed as in the previous section, $\mathbf{X}_A$ can be written as

$$\begin{pmatrix} 0 & \mathbf{Y}\mathbf{R}^{\frac{1}{2}} \\ \mathbf{R}^{\frac{1}{2}}\mathbf{Y}^* & 0 \end{pmatrix}.$$

Now, it is straightforward to see that $(\mathbf{X}_A)^2$ is equal to

$$\begin{pmatrix} \mathbf{YRY}^* & 0 \\ 0 & \mathbf{R}^{\frac{1}{2}}\mathbf{Y}^*\mathbf{Y}\mathbf{R}^{\frac{1}{2}} \end{pmatrix}$$

In particular, for any measurable function f,

$$\mathrm{tr}(f((\mathbf{X}_A)^2)) = 2\mathrm{tr}(f(\mathbf{YRY}^*)) + (M-N)f(0).$$

It is therefore a direct consequence of Theorem 1.1 that

Corollary 1.8 *With $(Y_{kl}, 1 \le k \le N, 1 \le l \le M)$ independent random variables, and with $\mathbb{P}^{N,M}$ as above, we let $\mathbf{R}$ be a non negative diagonal matrix with finite spectral radius ρ. Set $\mathbf{Z} = \mathbf{YRY}^*$. Then,*
a) If the $(P_{kl}, 1 \le k \le N, 1 \le l \le M)$ are supported in a compact set K, for any increasing convex Lipschitz function f so that $g(x) = f(x^2)$ has finite Lipschitz norm $|g|_{\mathcal{L}} \equiv |||f|||_{\mathcal{L}}$, for any $\delta > \delta_0(N+M) := 4|K|\sqrt{\pi\rho}|||f|||_{\mathcal{L}}/(N+M)$,

$$\mathbb{P}^{N,M}\left(|\mathrm{tr}_N(f(\mathbf{Z})) - \mathbb{E}^{N,M}(\mathrm{tr}_N(f(\mathbf{Z})))| > \delta\frac{M+N}{N}\right) \le 4e^{-\frac{1}{4|K|^2\rho|||f|||^2_{\mathcal{L}}}(\delta-\delta_0(N+M))^2(N+M)^2}.$$

b) If the $(P_{kl}, 1 \le k \le N, 1 \le l \le M)$ satisfy the logarithmic Sobolev inequality with uniformly bounded constant c, the above result holds for any Lipschitz functions f: for any Lipschitz function f, for any $\delta > 0$,

$$\mathbb{P}^{N,M}\left(|\mathrm{tr}_N f(\mathbf{Z})) - \mathbb{E}^{N,M}(\mathrm{tr}_N f(\mathbf{Z})))| > \delta\frac{M+N}{N}\right) \le 2e^{-\frac{1}{2c\rho|||f|||^2_{\mathcal{L}}}\delta^2(N+M)^2}.$$

The proof is in fact straightforward since, with the above remarks, one should see $f(\mathbf{Z})$ as $g(\mathbf{X}_A) = f((\mathbf{X}_A)^2)$ and thus control the Lipschitz norm of g and the convexity of g. Clearly, if f is convex increasing, g is convex so that we can then use Lemma 1.2 and thus complete the proof.

Remarks

1. The interest in Corollary 1.8 is in the case where N and M are large and N/M remains bounded and bounded away from zero.

2. For the convergence of the expected empirical measure of Wishart matrices, see [19], [21], and for rates, see [2].

3. Exponential convergence for the Wasserstein distance follows from our results by noting that the approximating set of functions in the proof of Theorem 1.3 below used convex increasing functions.

4. The results presented above extend to the case where $\mathbf{R}$ is self-adjoint non negative but not diagonal. Indeed, the key Lemma 1.2 generalizes as follows. Let Y, Y' denote $N \times M$ matrices with complex entries, with $||Y - Y'||^2 = \sum_{1 \le k \le N, 1 \le l \le M} |Y_{kl} - Y'_{kl}|^2$, writing $\mathbf{R} = U\widetilde{\mathbf{R}}U^*$ with U unitary, one has for $f \in \mathcal{F}_{\text{lip}}$ that

$$\mathrm{tr} f(Y\mathbf{R}Y^*) = \mathrm{tr} f(YU\widetilde{\mathbf{R}}U^*Y^*) = \mathrm{tr} f(U^*YU\widetilde{\mathbf{R}}U^*Y^*U)$$

Hence, with ρ denoting the spectral radius of $\mathbf{R}$, Lemma 1.2.b) and the observations above Corollary 1.8 imply

$$|\mathrm{tr} f(Y\mathbf{R}Y^*) - \mathrm{tr} f(Y'\mathbf{R}(Y')^*)| \le \frac{\sqrt{\rho}}{2} ||U^*YU - U^*Y'U|| = \frac{\sqrt{\rho}}{2} ||Y - Y'|| \, .$$

Hence, viewed as a function of the entries of Y, $f(Y\mathbf{R}Y^*)$ is Lipschitz of constant $\sqrt{\rho}/2$. A similar argument applies to the convexity considerations.

1.3 Concentration for non-commutative functionals

In this section, we present concentration inequalities for non-commutative functionals encountered in free probability theory. Let $(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^n_{A_n})$ be independent, $N \times N$ self adjoint inhomogeneous random matrices as in Section 1.1, with

$$(\mathbf{X}^p_{A_p})_{ij} = \frac{1}{\sqrt{N}} (\omega^p)_{ij} (A_p)_{ij} \, ,$$

with the A_p self adjoint and the entries $(\omega^p)_{ij} = (\omega^p)^R_{ij} + \sqrt{-1} (\omega^p)^I_{ij}$ of law P^p_{ij} such that

$$P^p_{ij}((\omega^p)_{ij} \in \cdot) = \int 1\!\!I_{u+iv \in \cdot} P^{p,R}_{ij}(du) P^{p,I}_{ij}(dv) \, .$$

We let $\mathbb{P}^{N,n} = \otimes_{1 \le i \le j \le N, 1 \le p \le n} (P^{p,R}_{ij} \otimes P^{p,I}_{ij})$, with respective expectation denoted by $\mathbb{E}^{N,n}$. In free probability theory, the law of $(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^n_{A_n})$ is determined by the family $\mathrm{tr}[Q(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^n_{A_n})]$ where Q are all possible (non-commutative) polynomials in n variables. In the particular case that $A_1 = \ldots = A_n = I$ and $P^p_{ij}(z) = 0$, $P^{p,R}_{ij}(x^2) = P^{p,I}_{ij}(x^2) = 1$, the law of $(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^n_{A_n})$ is known to converge to the law of n free semicircular variables. When P^p_{ij} are Gaussian, a central limit theorem was obtained in [8]. In order to state a large deviations upper bound (still in the Gaussian case), one cannot look at polynomials because these do not possess good enough exponential moments. Instead, in their study of large deviations upper bounds, T. Cabanal-Duvillard and one of the authors restricted themselves in [9] to the (separating) family of analytic, bounded functionals in the complex vector field generated by the family

$$\mathcal{F}_{\mathrm{NC}} = \{F(X_1, \dots, X_n) = \prod_{l=1}^{k} (z_l - \sum_{r=1}^{n} \alpha^l_r X_r)^{-1} \, , \quad z_l \in \mathbb{C} \setminus \mathbb{R}, \alpha^l_r \in \mathbb{R}, k \in \mathbb{N}\} \, .$$

However, the machinery developed in [9] is restricted to Gaussian entries. Here, with a less sharp bound, we provide the following concentration inequality.

Theorem 1.9 *Assume the laws $P_{ij}^{p,R}, P_{ij}^{p,I}$ satisfy the logarithmic Sobolev inequality with uniform constant c. Then, for any $F \in \mathcal{F}_{\mathrm{NC}}$, and any $\delta > 0$,*

$$\mathbb{P}^{N,n}(|\mathrm{tr}_N F(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^p_{A_p}) - \mathbb{E}^{N,n}\mathrm{tr}_N F(\mathbf{X}^1_{A_1}, \dots, \mathbf{X}^p_{A_p})| > \delta) \le 2e^{-\frac{\delta^2 N^2}{16ca^2||F||_{\mathcal{L}}}},$$

where, if $F(X_1, \dots, X_n) = \prod_{l=1}^k (z_l - \sum_{r=1}^n \alpha_r^l X_r)^{-1}$,

$$||F||^2_{\mathcal{L}} = \left[\sum_{p=1}^n \left(\sum_{l=1}^k |\alpha_l^p|\right)\left(\sum_{l=1}^k |\alpha_l^p| \cdot |\Im(z_l)|^{-2}\right)\right] \prod_{l=1}^k |\Im(z_l)|^{-2}.$$

2 Proofs

Let us first assume Lemma 1.2 and give the

Proof of Theorem 1.1

a) For a real valued random variable v, denote by M_v the median of v, i.e.

$$M_v = \sup\{t : P(v \le t) \le 1/2\}.$$

Following Talagrand (see [23]-[26] and theorem 6.6 in [25]), for any $F : \mathbb{R}^{N^2} \to \mathbb{R}$ Lipschitz with Lipschitz constant bounded by one, if $\mathbb{P}^N$ is supported on $[-1,1]^{N^2}$ and F is a convex function of (ω^R, ω^I), for any $\delta > 0$,

$$\mathbb{P}^N(|F - M_F| \ge \delta) \le 4e^{-\frac{\delta^2}{16}}.$$

Hence, taking $F(\omega^R, \omega^I) = \mathrm{tr}(f(\mathbf{X}_A(\omega)))$, F is convex and Lipschitz with Lipschitz norm bounded by $2a|f|_{\mathcal{L}}$ according to Lemma 1.2, and one gets by homothety that, for any $\mathbb{P}^N$ as in the statement of the theorem,

$$\mathbb{P}^N(|F - M_F| \ge \delta) \le 4e^{-\frac{\delta^2}{16|K|^2 a^2 |f|^2_{\mathcal{L}}}}. \tag{2.1}$$

In particular,

$$|\mathbb{E}^N \mathrm{tr}_N f - \frac{M_F}{N}| \le \mathbb{E}^N |\mathrm{tr}_N f - \frac{M_F}{N}| \le 4\int_0^\infty e^{-\frac{\delta^2}{16|K|^2 a^2 |f|^2_{\mathcal{L}}}} d\delta = \delta_0(N).$$

Substituting back into (2.1) yields the first part of Theorem 1.1.

b). In case where the $(P_{ij}^R, P_{ij}^I, 1 \le i \le j \le N)$ satisfy a log-Sobolev inequality with uniformly bounded constants c, by the product property of the logarithmic Sobolev inequality, for any $F : \mathbb{R}^{N^2} \to \mathbb{R}$ differentiable, we have

$$\int F^2 \log \frac{F^2}{\int F^2 d\mathbb{P}^N} d\mathbb{P}^N \le 2c \int ||\nabla F||^2 d\mathbb{P}^N,$$

where

$$||\nabla F||^2 = \sum_{1 \le i \le j \le N} (\partial_{\omega^R_{ij}} F(\omega))^2 + \sum_{1 \le i < j \le N} (\partial_{\omega^I_{ij}} F(\omega))^2.$$

Therefore, we can follow Herbst's argument (see Ledoux [18], proposition 2.3) to see that

$$\mathbb{P}^N\left(|F - \mathbb{E}^N(F)| > \delta\right) \le 2e^{-\frac{\delta^2}{2ca^2||(||\nabla F||)||^2_\infty}}.$$

Taking $F(\omega) = \mathrm{tr}(f(\mathbf{X}_A(\omega)))$ gives, with Lemma 1.2.b),

$$\mathbb{P}^N\left(|\mathrm{tr}(f(\mathbf{X}_A)) - \mathbb{E}^N(\mathrm{tr}(f(\mathbf{X}_A)))| > \delta\right) \le 2e^{-\frac{\delta^2}{8ca^2||f'||_\infty^2}}$$

for any differentiable functions f. The generalization to Lipschitz functions is obtained by approximations, the set of differentiable functions being dense in the set of Lipschitz functions for the Lipschitz norm.

∎

Very similarly, we can give the

Proof of Corollary 1.7 : Indeed, again assuming Lemma 1.2, we see that if f is convex Lipschitz, $(A_{ij})_{1\le i\le j\le N} \to \mathbb{E}^N[\mathrm{tr}_N(f(\mathbf{X}_A))]$ is convex Lipschitz with Lipschitz constant bounded by $2\sup_{x\in K}|x||f|_{\mathcal{L}}$. Hence, applying Talagrand's result (2.1) twice, we obtain

$$\mathbb{P}^N\left(|\mathrm{tr}_N(f(\mathbf{X}_A)) - \mathbb{E}^N(\mathrm{tr}_N(f(\mathbf{X}_A)))| \ge \delta + \frac{8|K|\sqrt{\pi}|f|_{\mathcal{L}}}{N}\right) \le 4e^{-\frac{1}{16|K|^2|f|_{\mathcal{L}}^2}N^2\delta^2} \tag{2.2}$$

and

$$Q^N\left(|E^N(\mathrm{tr}_N(f(\mathbf{X}_A))) - E^N\,\mathbb{E}^N(\mathrm{tr}_N(f(\mathbf{X}_A)))| \ge \delta + \frac{8\sup_{x\in K}|x|\sqrt{\pi}|f|_{\mathcal{L}}}{N}\right) \le 4e^{-\frac{1}{16\sup_{x\in K}|x|^2|f|_{\mathcal{L}}^2}N^2\delta^2} \tag{2.3}$$

Combining (2.2) and (2.3) give Corollary 1.7.

∎

Thus, let us prove the key Lemma 1.2.

Proof of Lemma 1.2.a) Note that for any $Y, X \in \mathcal{M}_{N\times N}^{sa}(\mathbb{C})$,

$$f(Y) - f(X) = \int_0^1 Df(X + \eta(Y-X))\sharp(Y-X)d\eta$$

where

$$Df(X)\sharp(Y) = \lim_{\epsilon\downarrow 0}\epsilon^{-1}(F(X+\epsilon Y) - F(X)).$$

For polynomial functions f, the non-commutative derivation D can easily be computed and one finds in particular that for any $p \in \mathbb{N}$,

$$Y^p - X^p = \int_0^1\left(\sum_{k=0}^{p-1}(X+\eta(Y-X))^k(Y-X)(X+\eta(Y-X))^{p-k-1}\right)d\eta. \tag{2.4}$$

For such a polynomial function, and taking the trace, one deduces that

$$\mathrm{tr}(X^p) - \mathrm{tr}((\frac{X+Y}{2})^p) = \int_0^1 p\mathrm{tr}\left((\frac{X+Y}{2} + \eta\frac{X-Y}{2})^{p-1}\frac{X-Y}{2}\right)d\eta\,, \tag{2.5}$$

$$\mathrm{tr}(Y^p) - \mathrm{tr}((\frac{X+Y}{2})^p) = \int_0^1 p\mathrm{tr}\left((\frac{X+Y}{2} - \eta\frac{X-Y}{2})^{p-1}\frac{Y-X}{2}\right)d\eta\,. \tag{2.6}$$

Hence, by further use of (2.4), (2.5) and (2.6) yield

$$\begin{aligned}\Delta &\equiv \mathrm{tr}(X^p)+\mathrm{tr}(Y^p)-2\mathrm{tr}((\frac{X+Y}{2})^p)\\ &= \frac{p}{2}\sum_{k=0}^{p-2}\int_0^1\int_0^1 \eta d\eta d\theta \mathrm{tr}\left((X-Y)Z_{\eta,\theta}^k(X-Y)Z_{\eta,\theta}^{p-2-k}\right) \end{aligned}\tag{2.7}$$

with

$$Z_{\eta,\theta}\equiv\frac{X+Y}{2}-\eta\frac{X-Y}{2}+\eta\theta(X-Y).$$

Next, for fixed $\eta,\theta\in[0,1]^2$, $X,Y\in\mathcal{M}_{N\times N}^{sa}(\mathbb{C})$, $Z_{\eta,\theta}\in\mathcal{M}_{N\times N}^{sa}(\mathbb{C})$, and we can find a unitary matrix $U_{\eta,\theta}$ and a diagonal matrix $D_{\eta,\theta}$ with real diagonal elements $(\lambda_{\eta,\theta}^1,..,\lambda_{\eta,\theta}^N)$ so that $Z_{\eta,\theta}=U_{\eta,\theta}D_{\eta,\theta}U_{\eta,\theta}^*$. Let $W_{\eta,\theta}=U_{\eta,\theta}^*(X-Y)U_{\eta,\theta}$. Then, the right hand side of (2.7) is given by

$$\begin{aligned}\Delta &= \frac{p}{2}\sum_{k=0}^{p-2}\int_0^1\int_0^1 \eta d\eta d\theta \mathrm{tr}\left(W_{\eta,\theta}D_{\eta,\theta}^kW_{\eta,\theta}D_{\eta,\theta}^{p-2-k}\right)\\ &= \frac{p}{2}\int_0^1\int_0^1 \eta d\eta d\theta\left(\sum_{k=0}^{p-2}\sum_{1\le i,j\le N}(\lambda_{\eta,\theta}^i)^k(\lambda_{\eta,\theta}^j)^{p-2-k}|W_{\eta,\theta}^{ij}|^2\right).\end{aligned}\tag{2.8}$$

But

$$\sum_{k=0}^{p-2}(\lambda_{\eta,\theta}^i)^k(\lambda_{\eta,\theta}^j)^{p-2-k}=\frac{(\lambda_{\eta,\theta}^i)^{p-1}-(\lambda_{\eta,\theta}^j)^{p-1}}{\lambda_{\eta,\theta}^i-\lambda_{\eta,\theta}^j}=(p-1)\int_0^1(\alpha\lambda_{\eta,\theta}^j+(1-\alpha)\lambda_{\eta,\theta}^i)^{p-2}d\alpha.$$

Hence, substituting in (2.8) gives,

$$\Delta = \frac{1}{2}\sum_{1\le i,j\le N}\int_0^1\int_0^1\int_0^1 d\alpha\eta d\eta d\theta|W_{\eta,\theta}^{ij}|^2f''(\alpha\lambda_{\eta,\theta}^j+(1-\alpha)\lambda_{\eta,\theta}^i)\tag{2.9}$$

for the polynomial function $f(x)=x^p$. Now, with (X,Y) fixed, the eigenvalues $(\lambda_{\eta,\theta}^1,..,\lambda_{\eta,\theta}^N)$ and the entries of $W_{\eta,\theta}$ are uniformly bounded. Hence, by Runge's theorem, we can deduce by approximation that (2.9) holds for any twice continuously differentiable function f. As a consequence, for any such convex function, $f''\ge 0$ and

$$\Delta=\frac{1}{2}\mathrm{tr}(f(X))+\frac{1}{2}\mathrm{tr}(f(Y))-\mathrm{tr}(f(\frac{X+Y}{2}))\ge 0.$$

The generalization to arbitrary convex functions is then obtained by approximations.

∎

Proof of Lemma 1.2.b) Let us first consider a bounded continuously differentiable function f and show that $\omega\to\mathrm{tr}(f(\mathbf{X}_A(\omega)))$ is differentiable and bound the Euclidean norm of its gradient. More precisely, we shall prove that

$$\sum_{1\le i\le j\le N}\left(\partial_{\omega_{ij}^R}\mathrm{tr}f(\mathbf{X}_A(\omega))\right)^2+\sum_{1\le i<j\le N}\left(\partial_{\omega_{ij}^I}\mathrm{tr}f(\mathbf{X}_A(\omega))\right)^2\le 4a^2|f|_{\mathcal{L}}^2\tag{2.10}$$

It is not hard to verify that

$$\partial_{\omega_{ij}^R}\mathrm{tr}f(\mathbf{X}_A)=\frac{1}{\sqrt{N}}\mathrm{tr}(f'(\mathbf{X}_A)\Delta_{ij})\in\mathbb{R}\tag{2.11}$$

with $(\Delta_{ij})_{kl} = a_{kl}$ if $(kl) = (ij)$ or (ji) and $(\Delta_{ij})_{kl} = 0$ otherwise. In fact, this is a consequence of (2.4) for polynomial functions and can then be obtained for continuously differentiable functions by approximations. In other words, (2.11) shows that

$$\partial_{\omega^R_{ij}} \mathrm{tr} f(\mathbf{X}_A) = \begin{cases} \frac{1}{\sqrt{N}}[a_{ij} f'(\mathbf{X}_A)_{ij} + a_{ji} f'(\mathbf{X}_A)_{ji}], & \text{if } i \neq j \\ \frac{a_{ii}}{\sqrt{N}} f'(\mathbf{X}_A)_{ii}, & \text{if } i = j . \end{cases}$$

Hence,

$$\sum_{1\le i\le j\le N} \left(\partial_{\omega^R_{ij}} \mathrm{tr} f(\mathbf{X}_A)\right)^2 \le \frac{2}{N} \sum_{1\le i,j\le N} |a_{ij} f'(\mathbf{X}_A)_{ij}|^2 \le \frac{2a^2}{N} \sum_{1\le i,j\le N} |f'(\mathbf{X}_A)_{ij}|^2 = 2a^2 \mathrm{tr}_N \left(f'(\mathbf{X}_A) f'(\mathbf{X}_A)^*\right). \tag{2.12}$$

But if $(\lambda_1, .., \lambda_N) \in I\!R^N$ denotes the eigenvalues of $\mathbf{X}_A$,

$$\mathrm{tr}_N \left(f'(\mathbf{X}_A) f'(\mathbf{X}_A)^*\right) = \frac{1}{N} \sum_{i=1}^N f'(\lambda_i)^2 \le ||f'||^2_\infty$$

so that with (2.12), we conclude

$$\sum_{1\le i\le j\le N} \left(\partial_{\omega^R_{ij}} \mathrm{tr} f(\mathbf{X}_A)\right)^2 \le 2a^2 ||f'||^2_\infty .$$

The same argument applies for the derivatives with respect to ω^I_{ij}, and we find

$$|\mathrm{tr}(f(\mathbf{X}_A(\omega))) - \mathrm{tr}(f(\mathbf{X}_A(\omega')))| \le 2a|f|_{\mathcal{L}} ||\omega - \omega'||$$

according to integration by parts formula and Cauchy-Schwartz 's inequality. Note that this last result for differentiable functions extends to Lipschitz functions by approximations. ∎

Proof of Theorem 1.3 a) Without loss of generality, we assume that $\min\{x : x \in \mathcal{K}\} = 0$. We will approximate a function $f \in \mathcal{F}_{\mathrm{lip},\mathcal{K}}$ by a combination of a finite number of convex functions from $\mathcal{F}_{\mathrm{lip}}$. Towards this end, fix $\Delta = \delta/4$, and define the function

$$g(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 \le x \le \Delta \\ \Delta, & x \ge \Delta . \end{cases}$$

We note that while $g(\cdot)$ is not convex, it is given as the sum of two convex functions in $\mathcal{F}_{\mathrm{lip}}$. Define recursively $f_\Delta(x) = 0$ for $x \le 0$ and

$$f_\Delta(x) = \sum_{i=0}^{\lceil x/\Delta \rceil} (2 1\!\!I_{f((i+1)\Delta) > f_\Delta(i\Delta)} - 1) g(x - i\Delta) .$$

Note that $||f - f_\Delta|| < \Delta$, and at most $2|\mathcal{K}|/\Delta$ different convex functions $\{h_k\}$ from $\mathcal{F}_{\mathrm{lip}}$ were used in this approximation, regardless of the particular function $f \in \mathcal{F}_{\mathrm{lip},\mathcal{K}}$. Thus,

$$\begin{aligned} &I\!P^N \left(\sup_{f \in \mathcal{F}_{\mathrm{lip},\mathcal{K}}} |\mathrm{tr}(f(\mathbf{X}_A)) - I\!E^N(\mathrm{tr}(f(\mathbf{X}_A)))| > \delta \right) \\ &\le \frac{2|\mathcal{K}|}{\Delta} \sup_k I\!P^N \left(|\mathrm{tr}(h_k(\mathbf{X}_A)) - I\!E^N(\mathrm{tr}(h_k(\mathbf{X}_A)))| > \frac{\Delta(\delta - 2\Delta)}{2|\mathcal{K}|} \right) \\ &\le \frac{8|\mathcal{K}|}{\Delta} \exp\left(-N^2 \frac{1}{16|K|^2 a^2} [(\frac{\Delta}{2|\mathcal{K}|}(\delta - 2\Delta) - \delta_1(N))_+]^2 \right) , \end{aligned}$$

where we have used Theorem 1.1 in the last inequality. Substituting $\Delta = \delta/4$ yields the conclusion.
b) Note that since the $P_{i,j}$'s are compactly supported in K, we have $\mathrm{tr}_N(\mathbf{X}_A^2) = N^{-2}\sum_{i,j}|A_{ij}\omega_{ij}|^2 \le Sa^2$. Thus, for any $M_1 > 0$, $\mathrm{tr}_N(1\!\!I_{|\mathbf{X}_A|>M_1}) \le Sa^2/M_1^2$. Let now

$$f^{M_1}(x) = \begin{cases} f(x), & |x| < M_1 \\ f(M_1) - \mathrm{sign}(f(M_1))(x - M_1), & M_1 < x < M_1 + |f(M_1)| \\ f(-M_1) + \mathrm{sign}(f(-M_1))(x + M_1), & -M_1 - |f(-M_1)| < x < -M_1 \\ 0, & \text{otherwise}. \end{cases}$$

Then,

$$|\mathrm{tr}_N f(\mathbf{X}_A) - \mathbb{E}^N \mathrm{tr}_N f(\mathbf{X}_A)| \le |\mathrm{tr}_N f^{M_1}(\mathbf{X}_A) - \mathbb{E}^N \mathrm{tr}_N f^{M_1}(\mathbf{X}_A)| + 4\frac{Sa^2}{M_1^2}.$$

Take now $M_1 = M/\sqrt{\delta}$. Then, using that $4Sa^2/M^2 \le 1/2$, and setting $\mathcal{K} = [-(M_1+1), (M_1+1)]$, one sees that $f^{M_1} \in \mathcal{F}_{\mathrm{lip},\mathcal{K}}$ and

$$\begin{aligned} &\mathbb{P}^N\Big(\sup_{f\in\mathcal{F}_{\mathrm{lip}}} |\mathrm{tr}_N f(\mathbf{X}_A) - \mathbb{E}^N \mathrm{tr}_N f(\mathbf{X}_A)| > \delta\Big) \\ &\le \mathbb{P}^N\Big(\sup_{g\in\mathcal{F}_{\mathrm{lip},\mathcal{K}}} |\mathrm{tr}_N g(\mathbf{X}_A) - \mathbb{E}^N \mathrm{tr}_N g(\mathbf{X}_A)| > \delta/2\Big) \\ \le\ & \frac{128(M+\sqrt{\delta})}{\delta^{3/2}} \exp\left(-N^2 \frac{1}{16|K|^2a^2}\Big[\frac{\delta^{5/2}}{128(M+\sqrt{\delta})} - \delta_1(N)\Big]^2\right), \end{aligned}$$

where part a) of the theorem was used in the last inequality. ∎

Proof of Corollary 1.4 The first part is immediate from Theorem 1.3. The second part follows from Theorem 1.1 using the same approximation by a fixed basis of Lipschitz functions as in Theorem 1.3. ∎

Proof of Theorem 1.5 This type of result is rather common in this field and comes from the great generality of concentration inequalities. In fact,

$$F(\omega) = d(\mathrm{tr}_N, \mu)(\omega) = \sup_{f\in\mathcal{F}_{\mathrm{lip}}} |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - \int f d\mu|$$

can be seen to be Lipschitz since for any $\omega, \omega' \in I\!\!R^{\frac{N(N+1)}{2}}$,

$$\begin{aligned} F(\omega) &= \sup_{f:||f||_{\mathcal{L}}\le 1} |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - \mathrm{tr}_N(f(\mathbf{X}_A(\omega'))) + \mathrm{tr}_N(f(\mathbf{X}_A(\omega'))) - \int f d\mu| \\ &\le F(\omega') + \sup_{f:||f||_{\mathcal{L}}\le 1} |\mathrm{tr}_N(f(\mathbf{X}_A(\omega))) - \mathrm{tr}_N(f(\mathbf{X}_A(\omega')))| \\ &\le F(\omega') + 2a||\omega - \omega'|| \end{aligned}$$

where we have used in the last line lemma 1.2.b). By symmetry, we deduce that F is Lipschitz. Since Herbst's argument used in the proof of Theorem 1.1 extends to Lipschitz functions by density, it immediately gives Theorem 1.5. ∎

Proof of Theorem 1.9 : As in the proof of Theorem 1.1.b), we only need to show that, if $F(X_1, \ldots, X_n) = \prod_{l=1}^k (z_l - \sum_{r=1}^n \alpha_r^l X_r)^{-1}$, $G : (I\!\!R^{N^2})^n \mapsto \mathbb{C}$ given by

$$G(\omega^1,..,\omega^n) := \mathrm{tr}_N\left(F(\mathbf{X}^1_{A_1}(\omega_1),..,\mathbf{X}^n_{A_n}(\omega_n))\right)$$

is Lipschitz and bound its Lipschitz norm. In fact, G is differentiable and for any $1 \le i \le j \le N$, $1 \le p \le n$,

$$\partial_{\omega^{p,R}_{ij}} G(\omega^1,..,\omega^n) = \sum_{l=1}^k \alpha^l_p \mathrm{tr}_N\left(\prod_{m=1}^l (z_m - \sum_{r=1}^n \alpha^m_r X_r)^{-1}\Delta_{ij}\prod_{m=l}^k (z_m - \sum_{r=1}^n \alpha^m_r X_r)^{-1}\right)$$

with $(\Delta_{ij})_{kl} = a^p_{kl}$ if $(kl) = (ij)$ or (ji) and $(\Delta_{ij})_{kl} = 0$ otherwise. Thus, as in the proof of Lemma 1.2.b), we deduce by Jensen's inequality

$$\sum_{1\le i\le j\le N}\left|\partial_{\omega^{p,R}_{ij}} G(\omega^1,..,\omega^n)\right|^2 \le 2a^2\left(\sum_{l=1}^k |\alpha^l_p|\right)\sum_{l=1}^k |\alpha^l_p| \mathrm{tr}_N\left|\prod_{m=l}^k (z_m - \sum_{r=1}^n \alpha^m_r X_r)^{-1}\prod_{m=1}^l (z_m - \sum_{r=1}^n \alpha^m_r X_r)^{-1}\right|^2$$

where for any $N\times N$ matrix M, $|M|^2 = MM^*$. Noticing that the operator norm of $(z-M)^{-1}$ is bounded by $\Im(z)^{-1}$ for any self-adjoint matrix M, we conclude

$$\sum_{1\le i\le j\le N}\left|\partial_{\omega^{p,R}_{ij}} G(\omega^1,..,\omega^n)\right|^2 \le 2a^2\left(\sum_{l=1}^k |\alpha^l_p|\right)\sum_{l=1}^k |\alpha^l_p| \prod_{m=1}^k |\Im(z_m)|^{-2}|\Im(z_l)|^{-2}$$

By similar computations for the derivations with respect to the $\omega^{p,I}_{ij}$ and summing over p, we complete the proof of Theorem 1.9. ∎

Acknowledgment We wish to thank Neil O'Connell for pointing out to one of us that Hamming type concentration inequalities, together with a perturbation argument for the eigenvalues, can be applied to Wigner matrices and yield exponential convergence speed of order N. The current paper can be seen as an attempt to understand and improve on his observation. We thank David Tse for raising the question of exponential convergence of speed N^2 for inhomogeneous Wishart matrices, and Michel Ledoux for discussions on concentration inequalities and their relation to logarithmic Sobolev inequalities. We are also very grateful to Dimitry Shlyakhtenko for pointing out to us that Wishart matrices can be seen as the square of inhomogeneous Wigner matrices, an observation which resulted in considerable shortening of the paper.

BIBLIOGRAPHY

[1] L. ARNOLD ; On the asymptotic distribution of eigenvalues of random matrices, *J. Math. Anal. Appl.* **20**, 262-268 (1967).

[2] Z. D. BAI ; Convergence rate of expected spectral distributions of large random matrices. Part I: Wigner Matrices, Part II: sample covariance matrices, *Annals of Probability* **21**, No 2, 625–672 (1993)

[3] Z.D. BAI ; Methodologies in spectral analysis of large dimensional random matrices: a review, *Statistica Sinica* **9**, No 3, 611–661 (1999).

[4] G. BENAROUS, A. GUIONNET, Large deviations for Wigner's law and Voiculescu's non-commutative entropy, *Prob. Th. Rel. Fields 108*, 517–542 (1997).

[5] F.A. BEREZIN ; Some remarks on the Wigner distribution, *Teo. Mat. Fiz* **17**, N. 3, 1163–1171 (English) (1973)

[6] S.G. BOBKOV , F.G. GÖTZE ; Exponential integrability and transportation cost related to log-Sobolev inequalities, *J. Funct. Anal.* **163** , No 1 (1999)

[7] L.V. BOGACHEV, S.A. MOLCHANOV, L.A. PASTUR ; On the density of states of random band matrices, *Math Notes* **50** , 832-842 (1991).

[8] T. CABANAL-DUVILLARD ; Fluctuations de la loi spectrale des grandes matrices aléatoires, Submitted to Ann. Inst. H. Poincaré. (1999).

[9] T. CABANAL-DUVILLARD, A. GUIONNET ; Large deviations upper bounds and non commutative entropies for some matrices ensembles, Preprint (1999).

[10] G. CASATI, V. GIRKO ; Generalized Wigner law for band random matrices, *Random Oper. Stochastic Equations* **1**, 279-286 (1993).

[11] L. GROSS ; Logarithmic Sobolev inequalities, *Amer. J. Math. 97*, 1061–1083 (1975).

[12] A. GUIONNET ; Large deviation upper bounds and central limit theorems for band matrices, Preprint (1999).

[13] A. GUIONNET, B. ZEGARLINSKI ; COURSE NOTES IHP 98.

[14] F. HIAI, D. PETZ ; Eigenvalue density of the Wishart matrix and large deviations, *Inf. Dim. Anal. Quantum Probab. Rel. Top.* **1**, 633–646 (1998).

[15] K. JOHANSSON ; On fluctuations of eigenvalues of random Hermitian matrices, *Duke J. Math.* **91**, 151–204 (1998).

[16] A. M. KHORUNZHY, B. A. KHORUZHENKO, L. A. PASTUR, M.V. SHCHERBINA ; The large n-limit in Statistical Mechanics and the Spectral Theory of Disordered Systems, *Phase Transition and critical phenomena* ,**15**, 73-239 ed. C. Domb, J.L. Lebowitz, Academic Press, London (1992)

[17] A. M. KHORUNZHY, B. A. KHORUZHENKO, L. A. PASTUR ; Asymptotic properties of large random matrices with independent entries, *J. Math. Phys.* **37**, 5033-5060 (1996).

[18] M. LEDOUX ; Concentration of measure and logarithmic Sobolev inequalities, *Séminaire de probabilité XXXIII* LNM **1709**, 80-216 (1999)

[19] L.A. PASTUR, V.A MARTCHENKO ; The distribution of eigenvalues in certain sets of random matrices, *Math. USSR-Sbornik* **1**, 457-483 (1967)

[20] D. SHLYAKHTENKO ; Random Gaussian Band matrices and Freeness with Amalgation, *Int. Math. Res. Not.***20**, 1013-1025 (1996).

[21] J. W. SILVERSTEIN, Z. D. BAI ; On the empirical distribution of eigenvalues of a class of large dimensional random matrices, *J. Mult. Anal* **54**, 175–192 (1995).

[22] Ya. SINAI, A. SOSHNIKOV ; A central limit theorem for traces of large random matrices with independent matrix elements, *Bol. Soc. Brasil. Mat. (N.S.)* **29**, 1–24 (1998).

[23] M. TALAGRAND ; An isoperimetric theorem on the cube and the Khintchine-Kahane inequalities, *Proc. Amer. Math. Soc* **104** , 905-909 (1988).

[24] M. TALAGRAND ; Concentration of measure and isoperimetric inequalities in product spaces, *Pub. Math. IHES* **81** , 73-205 (1995).

[25] M. TALAGRAND ; A new look at independence, *Annals of Probab.* **24**, 1-34 (1996).

[26] M. TALAGRAND ; New concentration inequalities in product spaces, *Invent. Math.* **86** , No 3, 505-563 (1996).

[27] K.W. WACHTER ; The strong limits of random matrix spectra for sample matrices of independent elements, *Annals of Probab.* **6**, 1-18 (1978).

[28] E. WIGNER ; On the distribution of the roots of certain symmetric matrices, *Ann. Math.* **67**, 325-327 (1958).

[29] J. WISHART ; The generalized product moment distribution in samples from a normal multivariate population, *Biometrika* **20**, 35-52 (1928).