Eigen behaviour of Random matrices with Heavy tailed independent columns

Cosme Louart* Menglin Wang*

Abstract

We study the resolvent $Q^z = (zI_p - \frac{1}{n}XX^T)^{-1}$ for $z \in \mathbb{C}$, with $\Im(z) > 0$ and where $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$ is a random matrix with independent but not necessarily identically distributed columns. Following the concentration of measure framework, we assume each column vector x_i has a "concentration function" $\alpha(\cdot/\eta_p)$ for a certain $\eta_p = o(\sqrt{p})$ and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ possibly slowly decaying non-increasing mapping. In particular an hypothesis on independence between the entries of x_i is not required. Quasi-asymptotically, Q^z has a natural deterministic equivalent $\tilde{Q}^z$, which depends on the second moments of the column vectors $x_1, \dots, x_n$. We demonstrate that for a given deterministic matrix $A \in \mathcal{M}_p$, the projection $\mathrm{Tr}(AQ^z)$ is concentrated around $\mathrm{Tr}(A\tilde{Q}^z)$, with decay speed proportional to $\eta_p \|A\|$ when $\int t\alpha(t)\,dt < \infty$ and proportional to $\eta_p \|A\|_{\mathrm{HS}}/\sqrt{n}$ when $\int t^3\alpha(t)\,dt < \infty$.

Keywords: Random matrix theory; Heavy tailed concentration; Hanson-Wright inequality; Limiting spectral distribution.
MSC2020 subject classifications: 60-08, 60B20, 62J07.

Notations

Let us introduce the notations $\mathbb{R}_+ \equiv [0, \infty)$, $\mathbb{R}_+^* \equiv (0, +\infty)$ and $\mathbb{H} \equiv \{z \in \mathbb{C}, \Im(z) > 0\}$ (the complex half plane). Given $n, p \in \mathbb{N}$, $[n] \equiv \{1, \dots, n\}$, the entries of a vector $x \in \mathbb{C}^p$ are generally denoted $x_1, \dots, x_p$, the columns of a complex matrix $A \in \mathcal{M}_{p,n}$ are denoted $a_1, \dots, a_n$. Let us denote $\mathcal{M}_n$, the set of square matrices $\mathcal{M}_{n,n}$, H_n, the set of Hermitian matrices and D_n, the set of diagonal matrices. Given $M \in \mathcal{M}_{p,n}$ the transpose of M i denoted M^T and the transpose conjugate is denoted $M^* \equiv \bar{M}^T$. We introduce the natural order relation on H_n, given $A, B \in H_n$:

$$A \leq B \qquad \Longleftrightarrow \qquad \forall x \in \mathbb{C}^n : \quad x^*(B - A)x \geq 0.$$

Given $x \in \mathbb{C}^n$, $D = \mathrm{Diag}(x) \in D_n$ is the diagonal matrix having the elements $x_1, \dots, x_n$ on the diagonal then one usually denote $\forall i \in [n]$, $D_i \equiv x_i$. Given a square matrix $A \in \mathcal{M}_p(\mathbb{C})$, the spectrum of A is $\mathrm{Sp}(A)$ and we denote $|A| = \sqrt{AA^*} \in \mathcal{H}_p$.

The ℓ_2 norm on $\mathbb{C}^p$ is denoted $\|\cdot\|$ ($\|x\| \equiv \sqrt{\sum_{i=1}^p |x_i|^2}$), then the Hilbert-Schmidt norm is denoted $\|\cdot\|_{\mathrm{HS}}$ ($\forall M \in \mathcal{M}_{p,n}$: $\|M\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(MM^*)} = \sup_{\|A\|_{\mathrm{HS}} \leq 1} |\mathrm{Tr}(AM)|$),

*School of Data Science, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

the spectral norm is denoted $\|\cdot\|$ ($\|M\| = \sup_{\|x\|=1} \|Mx\|$) and the nuclear norm is denoted $\|\cdot\|_*$ ($\|M\| = \operatorname{Tr}(\sqrt{MM^*}) = \sup_{\|A\|\leq 1} |\operatorname{Tr}(AM)|$). Given two normed vector spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, and a linear mapping $u : E \to F$, the operator norm of u is denoted $\|u\| \equiv \sup_{\|x\|_E \leq 1} \|u(x)\|_F$.

Given an index set Θ and two family of parameters $(a_\theta)_{\theta\in\Theta} \in \mathbb{R}_+^\Theta$ and $(b_\theta)_{\theta\in\Theta} \in \mathbb{R}_+^\Theta$, we denote: "$a_\theta \leq O(b_\theta),\ \theta \in \Theta$" or "$a \leq O(b)$" or, abusively, "$a \leq O(b),\ \theta \in \Theta$" iif there exists a constant $C > 0$ such that $\forall \theta \in \Theta$: $a_\theta \leq Cb_\theta$ (and we note $a \geq O(b)$ iif $\exists C > 0$ such that $\forall \theta \in \Theta$, $a_\theta \geq Cb_\theta$). We further denote $a \leq o(1)$ iif for any constant $\varepsilon > 0$, there exists a finite subset $T \subset \Theta$ such that $\forall \theta \in \Theta \setminus T$: $a_\theta \leq \varepsilon$ and $a \leq o(b)$ signifies $\frac{a}{b} \leq o(1)$. If $A, B \in \prod_{\theta\in\Theta} H_{n_\theta}$ are two families of Hermitian matrices, $A \leq O(B)$ means that there exists a constant $C > 0$ such that:

$$\forall \theta \in \Theta : \qquad B_\theta - A_\theta \geq CI_{n_\theta}.$$

Given a normed vector space E, random variables $Y \in E$ are typically measurable mapping from a certain probability space Ω to E but Ω is omitted for simplicity. Without any further specification, Y' designates an independent copy of Y.

Introduction

Considering a non centered sample covariance matrix $\frac{1}{n}XX^T$, where $X = (x_1, \ldots, x_n) \in \mathcal{M}_{p,n}$ is the data matrix, The spectral distribution of $\frac{1}{n}XX^T$, denoted by μ is given by the expression $\mu \equiv \frac{1}{p}\sum_{\lambda\in\operatorname{Sp}(\frac{1}{n}XX^T)} \delta_\lambda$ and is classically studied through its Stieltjes transform, which is defined as:

$$\begin{aligned} g : \mathbb{C} \setminus \operatorname{Sp}\left(\frac{1}{n}XX^T\right) \quad &\longrightarrow \quad \mathbb{C} \\ z \quad &\longmapsto \int_{\mathbb{R}} \frac{d\mu(\lambda)}{z-\lambda}. \end{aligned}$$

Note that actually The Steiltjes transform g is linked to the so-called resolvent of $\frac{1}{n}XX^T$ $Q^z \equiv (I_p - \frac{1}{nz}XX^T)^{-1}$ through the formula $g(z) = \frac{1}{zp}\operatorname{Tr}(Q^z)$ valid for all $z \in \mathbb{C} \setminus \operatorname{Sp}\left(\frac{1}{n}XX^T\right)$. The importance of the Stieltjes transform has been thoroughly established in seminal works [MP67, Sil86], through the Cauchy integral formula. This formula allows the evaluation of integrals of analytical functions f defined on a neighborhood of a subset $B \subset \operatorname{Sp}(\frac{1}{n}XX^T)$ by:

$$\int_B f(\lambda)d\mu(\lambda) = \frac{1}{2i\pi}\oint_\gamma f(z)g(z)dz,$$

where $\gamma : [0, 1] \to \mathbb{C} \setminus \operatorname{Sp}(\frac{1}{n}XX^T)$ is a closed contour in $\mathbb{C} \setminus \operatorname{Sp}\left(\frac{1}{n}XX^T\right)$ on which f is defined. The interior I_γ satisfies $I_\gamma \cap \operatorname{Sp}(\frac{1}{n}XX^T) = B \cap \operatorname{Sp}(\frac{1}{n}XX^T)$.

Historically, much of the foundational work in sample covariance matrices focused on matrices with independent and identically distributed (i.i.d.) entries [MP67, Wac78, Yin86, SB95] or on Gaussian hypotheses allowing to access independence assumptions through linear transformation [BKV96, KA16]. The readers are referred to the monographs [BS09], [AGZ10], [PŜ11] for further details.

In practical scenarios taken from telecommunication [WCDS12] or machine learning, data matrices often exhibit strong correlation between the entries. This has prompted significant interest in extending random matrix theory to account for dependent entries like in [Ada11, DKL22] where they mainly assume concentration hypotheses on the second moments of the columns of the matrix. These results are already quite strong but mainly concentrate on the limiting behavior when we try here to provide quasi-asymtotic results on the eigenvalue distribution. This is done through a concentration of measure framework (for comprehensive presentation refer to [Led05, BLM13] or to [Tao12, Ver18] for specific applications to random matrices). To reach “heavy tailed” settings, we consider an example recently presented in [Lou24] of random vectors whose 1-Lipschitz observations have only finite moment up to a certain order.

We are able to show Marcenko-Pastur convergence under our hypotheses assuming that the second moment is bounded and we can also reach convergence of the resolvent in Hilbert-Schmidt norm under the hypothesis of bounded fourth moment. This last result seems to be very close to the local law convergence which was first set for Wigner matrices in [ESY09] and adapted to Wishart matrices in [ESYY12, KM23]. However we still do not manage to get the typical decay proportional to $\frac{1}{n\Im(z)}$ when z is getting close to the spectrum. To avoid presenting unaesthetic and unusable powers of $\Im(z)$ in our result, we thus decided to take z as a fixed complex number of either positive imaginary part or negative real part.

Contents

1 Main results

Our approach naturally decomposes in two almost independent steps:

1. A purely deterministic approach relying on complex analysis and basic topology and algebra to express the so called “deterministic equivalent” of the resolvent.
2. A probabilistic approach relying on concentration of measure assumption to set concentration of the resolvent binding around its deterministic equivalent.

1.1 Deterministic results

The ultimate goal of our paper is to express *quasi-asymptotic results*, meaning that our concentration inequality will be valid for big *but finite* p and n. However we adopt, although misleading, the denomination “limiting distribution” that can be found in random matrix litterature. For us it will simply be a deterministic approxiamte of the eigen value distribution of $\frac{1}{n}XX^T$, given $n, p \in \mathbb{N}$. It expresses thanks to the non centered population covariance matrices:

$$\forall i \in [n] : \qquad \Sigma_i \equiv \mathbb{E}[x_i x_i^T]$$

and the following uniquely defined diagonal matrix:

Theorem 1.1. *Given n nonnegative symmetric matrices $\Sigma_1, \ldots, \Sigma_n \in \mathcal{M}_p$, for all $z \in \mathbb{H}$, the equation:*

$$\forall i \in [n], L_i = z - \frac{1}{n} \operatorname{Tr} \left(\Sigma_i \left(I_p - \frac{1}{nz} \sum_{j=1}^n \frac{\Sigma_j}{L_j} \right)^{-1} \right) \tag{1.1}$$

admits a unique solution $L \in \mathcal{D}_n(\mathbb{H})$ which we denote by $\tilde{\Lambda}^z$.

Then the mapping

$$\tilde{g} : z \mapsto \frac{1}{p} \operatorname{Tr}(\tilde{Q}^{\tilde{\Lambda}^z}) = \frac{1}{z} \left(1 - \frac{n}{p} \right) + \frac{1}{p} \sum_{i=1}^n \frac{1}{\tilde{\Lambda}_i^z}$$

happens to be an analytical mapping that can be shown to be the Stieltjes transform of the limiting distribution.

Theorem 1.2. *The mapping $\tilde{g}$ is the Stieltjes transform of the measure:*

$$\tilde{\mu} = \left(\frac{n}{p} - 1 \right) \delta_0 - \frac{1}{p} \sum_{i=1}^n \tilde{\mu}_i,$$

where δ_0 is the Dirac measure on 0 and the measures $\tilde{\mu}_1, \ldots, \tilde{\mu}_n$ have as Stieltjes transform $z \mapsto -\frac{1}{\tilde{\Lambda}_i^z}$. And $\tilde{\mu}$ has a compact support $\tilde{S} \subset \mathbb{R}_+$.

Once those object are well defined, one can introduce for all $z \in \mathbb{C} \setminus \tilde{S}$ the deterministic equivalent:

$$\tilde{Q}^{\tilde{\Lambda}^z} \equiv \left(I_p - \frac{1}{nz} \sum_{j=1}^n \frac{\Sigma_j}{\tilde{\Lambda}_j^z} \right)^{-1},$$

and next part of the paper will aim at showing the convergence of the resolvent Q^z around $\tilde{Q}^{\tilde{\Lambda}^z}$.

1.2 Probabilistic results

In what follows, we study the asymptotic properties of the eigenvalue distribution when p and n tend to infinity, the random matrix X should then be seen as a family of random matrices depending on the asymptotic parameters $p, n \in \mathbb{N}$, however, for simplicity we will not index the family of matrices X and of columns $x_1, \ldots, x_n, \ldots$ with n and p.

To stay as simple as possible in this first presentation, we only present here two representative settings, one of highly concentrated random vectors and the other one allowing heavy-tailed decays. More general assumptions will be presented in Section **??**.

Setting (S). There exist some parameters $\lambda, C, K, r > 0$ such that for all $p, n \in \mathbb{N}$, there exist a set of n independent Gaussian random vectors $z^{(1)}, \ldots, z^{(n)} \sim \mathcal{N}(0, I_p)$, a set of n λ-Lipschitz mappings $G^{(i)} : \mathbb{R}^p \to \mathbb{R}^p$, $i \in [n]$ such that:

(S1) $\forall i \in [n] : \quad x_i = G^{(i)}(z^{(i)})$.

(S2) $\forall i \in [n]$, there exist $g_1, \dots, g_p : \mathbb{R} \to \mathbb{R}$, such that:

$$x_i = G^{(i)}((g_1^{(i)}(z_1^{(i)}), \dots, g_p^{(i)}(z_p^{(i)})), \quad \text{and} \quad \sup_{i\in[n], j\in[p]|v|,|w|\leq u} \frac{\|g_j^{(i)}(v) - g_j^{(i)}(w)|}{|v-w|} \leq \frac{C}{u} e^{\frac{u^2}{2r}},$$

and we assume in both case that $\|\mathbb{E}[x_i]\| \leq K$. Later, we will restrict ourselves to the cases $r > 2$ and $r > 4$.

Under this setting, it can be proven that ([**?**]) for all $f : \mathbb{R}^p \to \mathbb{R}$ and all $F : \mathcal{M}_{p,n} \to \mathbb{R}$ 1-Lipschitz (respectively for the euclidean and for the Hilbert-Schmidt norm):

$$\sup_{i\in[n]} \mathbb{P}\left(|f(x_i) - f(x_i')| \geq t\right) \leq \alpha\left(\frac{t}{\eta_p}\right) \quad \text{and} \quad \mathbb{P}\left(|F(X) - F(X')| \geq t\right) \leq \alpha\left(\frac{t}{\eta_p\eta_n}\right), \quad (1.2)$$

where the concentration function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ and the parameters η_k, $k \in \mathbb{N}$ vary depending on the exact setting:

- in setting **(S1)**: $\alpha(t) = 2e^{-(t/\lambda)^2/2}$ and $\forall k \in \mathbb{N}$ $\eta_k = 1$
- in setting **(S2)**: $\alpha(t) = 6(\frac{C\lambda}{t})^r$ and $\forall k \in \mathbb{N}$: $\eta_k = k^{1/r}$.

These concentration of measure results being given, the assumptions of next theorems will then sound more natural. First when the second or fourth moment of α is bounded one can set the concentration of the projections of the resolvent.

Theorem 1.3. *Considering a constant $\gamma > 0$, a family of random matrices $(X^{(p,n)})_{p,n\in\mathbb{N}} \in \prod_{p,n\in\mathbb{N}} \mathcal{M}_{p,n}$ satisfying Setting (S), let us assume that the concentration function $\alpha : \mathbb{R} \to \mathbb{R}$ introduced in* (1.2) *satisfies that*[1] $\int t\alpha(t)dt < \infty$ *then given $z \in \mathbb{H}$, there exist some constants $C, c > 0$ such that for all $n, p \in \mathbb{N}$ such that $p \leq \gamma n$:*

$$\mathbb{P}\left(\left|\operatorname{Tr}(A(Q^z - \tilde{Q}^{\tilde{\Lambda}^z}))\right| \geq t\right) \leq C\alpha\left(\frac{c\sqrt{n}t}{\eta_p p\|A\|}\right).$$

In the same setting, if we assume further that[2] $\int t^3\alpha(t)dt < \infty$*, then the concentration is expressed as:*

$$\mathbb{P}\left(\left|\operatorname{Tr}(A(Q^z - \tilde{Q}^{\tilde{\Lambda}^z}))\right| \geq t\right) \leq C\alpha\left(\frac{c\sqrt{n}t}{\eta_p^3\|A\|_{\mathrm{HS}}}\right).$$

As a corollary, assuming that the second moment of α is bounded, one can set the concentration of the Stieltjes transform of Q^z.

Corollary 1.4. *In the same setting as in 1.3, if we assume that α has bounded second moment $\int t\alpha(t)dt < \infty$, then the concentration of the Stieltjes transform is expressed as:*

$$\mathbb{P}\left(\left|g(z) - \frac{1}{zp}\operatorname{Tr}(\tilde{Q}^{\tilde{\Lambda}^z})\right| \geq 0\right) \leq C\alpha(c\sqrt{p}t/\eta_p).$$

After introducing some basic lemmas to control the norm and the concentration of the resolvent, the two results of Theorem 1.3 will be proven in two different subsections of Section 3 since they rely on quite different approach.

[1] That happens in Setting **(S1)** or in Setting **(S2)** when $r > 2$.
[2] That happens in Setting **(S1)** or in Setting **(S2)** when $r > 4$.

2 Deterministic study: Definition of the limiting distribution

2.1 Definition of the deterministic equivalent of the resolvent

We consider in the whole section a set of n nonnegative symmetric matrices $\Sigma_1, \ldots, \Sigma_n \in \mathcal{M}_p$ (they will be in next section the non-centered empirical covariance matrices of the columns of X). Let us then introduce the mapping Φ^z:

$$\forall L \in \mathcal{D}_n(\mathbb{H}) : \ \Phi^z(L) \equiv \mathrm{Diag}\left(z - \frac{1}{n}\mathrm{Tr}\left(\Sigma_i \tilde{Q}^L\right)\right)_{1\leq i\leq n}$$

we want to show that Φ^z admits a unique fixed point $\tilde{\Lambda}^z$. For that purpose,let us introduce a semi-metric $d_{\mathbb{H}}$ for which Φ^z happens to be contractive. Then a Banach-like theorem (Theorem B.6 given in the Appendix B) will provide the existence and uniqueness of $\tilde{\Lambda}^z$. For any $D, D' \in \mathcal{D}_n(\mathbb{H})$ let us define:

$$d_{\mathbb{H}}(D, D') = \sup_{1\leq i\leq n} \frac{|D_i - D'_i|}{\sqrt{\Im(D_i)\Im(D'_i)}}.$$

To be able to bound the variation of Φ^z for the semi-metric $d_{\mathbb{H}}$, one first needs to restrict the study on a subset:

$$\mathcal{D}_{\Phi^z} \equiv \{D \in \mathcal{D}_n(\mathbb{H}), D/z \in \mathcal{D}_n(\mathbb{H})\} \subset \mathcal{D}_n(\mathbb{H}).$$

Lemma 2.1. *For any $z \in \mathbb{H}$, $\Phi^z(\mathcal{D}_{\Phi^z}) \subset \mathcal{D}_{\Phi^z}$.*

Proof. Considering $z \in \mathbb{H}$, and $L \in \mathcal{D}_{\mathcal{I}^z}$ and $i \in [n]$, the decomposition $\tilde{Q}^L = \left(I_p - \frac{1}{n}\sum_{j=1}^n \frac{\Re(L_j)\Sigma_j}{|L_j|^2} + \frac{i}{n}\sum_{j=1}^n \frac{\Im(L_j)\Sigma_j}{|L_j|^2}\right)^{-1}$ allows us to set:

$$\begin{aligned}
\Im\left(\Phi^z(L)_i\right) &= \frac{1}{2i}(\Phi^z(L)_i - \overline{\Phi^z(L)_i}) \\
&= \Im(z) + \frac{1}{2in}\mathrm{Tr}\left(\Sigma_i\left(\overline{\tilde{Q}^L} - \tilde{Q}^L\right)\right) \\
&= \Im(z) + \frac{1}{n}\mathrm{Tr}\left(\Sigma_i \sum_{j=1}^n \bar{\tilde{Q}}^L\left(\frac{\Im(L_j)\Sigma_j}{|L_j|^2}\right)\tilde{Q}^L\right) > 0
\end{aligned}$$

(since $\bar{\tilde{Q}}^L\Sigma_i\tilde{Q}^L$ is a non negative Hermitian matrix). The same way, one can show:

$$\Im\left(\Phi^z(L)_i/z\right) = \frac{1}{2i}\mathrm{Tr}\left(\Sigma_i\left(\frac{\tilde{Q}^L}{z} - \frac{\bar{\tilde{Q}}^L}{\bar{z}}\right)\right) = \frac{1}{2i|z|^2}\mathrm{Tr}\left(\Sigma_i\bar{\tilde{Q}}^L\left(\sum_{j=1}^n \frac{\Im(L_j/z)\Sigma_j}{|L_j/z|^2}\right)\tilde{Q}^L\right) > 0$$

□

Let us now express the Lipschitz parameter of Φ^z for the semi metric $d_{\mathbb{H}}$.

Proposition 2.2. *For any $z \in \mathbb{H}$, the mapping Φ^z is 1-Lipschitz for the semi-metric $d_{\mathbb{H}}$ on $\mathcal{D}_{\Phi^z}$ and satisfies for any $L, L' \in \mathcal{D}_{\Phi^z}$:*

$$d_{\mathbb{H}}(\Phi^z(L), \Phi^z(L')) \leq \sqrt{(1-\zeta(z,L))(1-\zeta(z,L'))}\,d_{\mathbb{H}}(L, L'),$$

where for any $w \in \mathbb{H}$ and $L \in \mathcal{D}_{\Phi^w}$:

$$\zeta(w, L) = \frac{\Im(w)}{\sup_{1\leq i\leq n}\Im(\Phi^w(L))_i} \in (0, 1).$$

Proof. Let us bound for any $L, L' \in \mathcal{D}_{\Phi^z}$:

$$\begin{aligned}
|\Phi^z(L)_i - \Phi^z(L')_i| &= \frac{1}{n} \operatorname{Tr}\left(\Sigma_i \tilde{Q}^L \left(\frac{1}{n} \sum_{j=1}^n \frac{L_j - L'_j}{L_j L'_j} \Sigma_j \right) \tilde{Q}^{L'} \right) \\
&= \frac{1}{n} \operatorname{Tr}\left(\Sigma_i \tilde{Q}^L \left(\frac{1}{n} \sum_{j=1}^n \frac{L_j - L'_j}{\sqrt{\Im(L_j)\Im(L'_j)}} \frac{\sqrt{\Im(L_j)\Im(L'_j)}}{L_j L'_j} \Sigma_j \right) \tilde{Q}^{L'} \right) \\
&\leq d_{\mathbb{H}}(L, L') \sqrt{\frac{1}{n} \operatorname{Tr}\left(\Sigma_i \tilde{Q}^L \left(\frac{1}{n} \sum_{j=1}^n \frac{\Im(L_j)\Sigma_j}{|L_j|^2} \right) \bar{\tilde{Q}}^L \right)} \\
&\qquad \cdot \sqrt{\frac{1}{n} \operatorname{Tr}\left(\Sigma_i \tilde{Q}^{L'} \left(\frac{1}{n} \sum_{j=1}^n \frac{\Im(L'_j)\Sigma_j}{|L'_j|^2} \right) \bar{\tilde{Q}}^{L'} \right)} \\
&< d_{\mathbb{H}}(L, L') \sqrt{\left(\Im(\Phi^z(L)_i) - \Im(z)\right)\left(\Im(\Phi^z(L')_i) - \Im(z)\right)}, \qquad (2.1)
\end{aligned}$$

thanks to Cauchy-Schwarz inequality and the identity

$$0 \leq \frac{1}{n} \operatorname{Tr}\left(\Sigma_i \sum_{j=1}^n \bar{\tilde{Q}}^L \left(\frac{\Im(L_j)\Sigma_j}{|L_j|^2} \right) \tilde{Q}^L \right) = \Im(\Phi^z(L)_i) - \Im(z) \qquad (2.2)$$

issued from the proof of Lemma 2.1. By dividing both sides of (2.1) by $\sqrt{\Im(\Phi^z(L)_i)\Im(\Phi^z(L')_i)}$, we obtain the desired Lipschitz constant. □

Lemma 2.3. *Given* $\Delta \in \mathcal{D}_{\Phi^z}$, $\|\tilde{Q}^\Delta\| \leq O\left(\frac{|z|}{\Im(z)}\right)$.

Here, one could have merely assumed $\frac{\Delta}{z} \in \mathcal{D}_n(\mathbb{H})$ instead of $\Delta \in \mathcal{D}_{\Phi^z}$.

Proof. Let us consider the inverse matrix of $\frac{1}{z}\tilde{Q}^\Delta = (zI_p - \frac{1}{n}\sum_{i=1}^n \frac{z\Sigma_i}{\Delta_i})^{-1}$ applied to a vector $u \in \mathbb{R}^p$

$$\begin{aligned}
&\left\| \left(zI_p - \frac{1}{n} \sum_{i=1}^n \frac{\Sigma_i}{\Delta_i/z} \right) u \right\|^2 \\
&\quad = \left\| \left(\Re(z)I_p - \frac{1}{n} \sum_{i=1}^n \frac{\Re(\Delta_i/z)}{|\Delta_i/z|^2} \Sigma_i \right) u + i \left(\Im(z)I_p + \frac{1}{n} \sum_{i=1}^n \frac{\Im(\Delta_i/z)}{|\Delta_i/z|^2} \Sigma_i \right) u \right\|^2 \\
&\quad \geq \left\| \left(\Im(z)I_p + \frac{1}{n} \sum_{i=1}^n \frac{\Im(\Delta_i/z)}{|\Delta_i/z|^2} \Sigma_i \right) u \right\|^2 \ \geq\ \|\Im(z)u\|^2
\end{aligned}$$

Reversing this inequality, we deduce the result of the Lemma. □

Lemma 2.4. *Given* $L \in \mathcal{D}_{\Phi^z}$*, we can bound:*

$$\Im(z)I_n \leq |\Phi^z(L)| \leq \left(|z| + \frac{|z|p\|\Sigma\|}{n\Im(z)} \right) I_n \quad \text{and} \quad \left(\frac{\Im(z)}{\|\Sigma\| + |z|\Im(z)} \right) I_p \leq \left| \tilde{Q}^{\Phi^z(L)} \right| \leq \frac{|z|I_p}{\Im(z)}.$$

where $\|\Sigma\| \equiv \sup_{i \in [n]} \|\Sigma_i\|$.

Proof. The lower bound of $\Phi^z(L)$ is an immediate consequence of the inequality $\Im(\Phi^z(L)) \geq \Im(z)$ (see the proof of Lemma 2.1). Given $L \in \mathcal{D}_{\Phi^z}$, we know from Lemma 2.3 that

$|\tilde{Q}^L| \leq \frac{|z|}{\Im(z)}$ which directly provides the upper bound on $\Phi^z(L)$. Finally, we can bound:

$$\left\| I_p - \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\Phi^z(L)_i} \right\| \leq |z| + \frac{1}{n}\sum_{i=1}^n \frac{\|\Sigma_i\|}{|\Im(\Phi^z(L)_i)|} \leq |z| + \frac{\|\Sigma\|}{\Im(z)},$$

which gives us the lower bound on $|\tilde{Q}^{\Phi^z(L)}|$. □

Theorem 2.5. *For all $z \in \mathbb{H}$, the equation:*

$$\forall i \in [n], L_i = \Phi^z(L)_i \equiv z - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i \left(I_p - \frac{1}{n}\sum_{j=1}^n \frac{\Sigma_j}{L_j}\right)^{-1}\right) \tag{2.3}$$

admits a unique solution $L \in \mathcal{D}_n(\mathbb{H})$ which we denote by $\tilde{\Lambda}^z$, it satisfies in particular $\Im(\Lambda^z/z) > 0$.

Proof. On the domain $\mathcal{D}_{\Phi^z}$, the mapping Φ^z is bounded and contracting with respect to the semi-metric $d_{\mathbb{H}}$ thanks to Proposition 2.2 and Lemma 2.4 (indeed $\inf_{L\in\mathcal{D}_{\Phi^z}} \zeta(z, L) > 0$). Therefore the hypotheses of Theorem B.6 are satisfied, ensuring the existence of a unique diagonal matrix $\tilde{\Lambda} \in \mathcal{D}_{\Phi^z}$ such that $\Phi^z(\tilde{\Lambda}) = \tilde{\Lambda}$. Besides, Proposition 2.2 being true on the whole domain $\mathcal{D}_n(\mathbb{H})$, we are sure that there are no other fixed point in the entire set $\mathcal{D}_n(\mathbb{H})$. □

2.2 Definition of the limiting distribution

We present here some arguably well known results (see [KA16] for instance) about the Stieltjes transform of the eigenvalue distribution which provide valuable insights into its support. To show that $\tilde{g}$ is a Stieltjes transform, we will use the following theorem that can be found, for instance, in [Bol97]:

Theorem 2.6. *Given an analytic mapping $f : \mathbb{H} \to \mathbb{H}$, if $\lim_{y\to+\infty} -iyf(iy) = 1$ then f is the Stieltjes transform of a probability measure μ that satisfies the following two reciprocal formulas:*

- $f(z) = \int \frac{\mu(d\lambda)}{\lambda - z}$,
- *for any continuous point*[3] $a < b$: $\mu([a, b]) = \lim_{y\to 0^+} \frac{1}{\pi}\int_a^b \Im(f(x + iy))dx$.

If, in particular, $\forall z \in \mathbb{H}$, $zf(z) \in \mathbb{H}$, then $\mu(\mathbb{R}^-) = 0$ and f admits an analytic continuation on $\mathbb{C} \setminus (\mathbb{R}_+ \cup \{0\})$.

The first hypothesis to verify for applying Theorem 2.6 is the analyticity of $\tilde{g}$, which follows from the analyticity of the mapping $z \to \tilde{\Lambda}^z$. While it is possible to prove the analyticity of $\tilde{\Lambda}^z$ using limiting arguments, by viewing it as the limit of a sequence of analytic mappings, we opt instead to derive this property directly from its original definition, even though the approach is somewhat more laborious. We start with establishing continuity by using Proposition B.8.

Proposition 2.7. *The mapping $z \mapsto \tilde{\Lambda}^z$ is continuous on $\mathbb{H}$.*

Proof. Given $z \in \mathbb{H}$, we consider a sequence $(t_s)_{s\in\mathbb{N}} \in \{w \in \mathbb{C} \mid w + z \in \mathbb{H}\}$ such that $\lim_{s\to\infty} t_s = 0$. We now verify the assumption of Proposition B.8 where for all $s \in \mathbb{N}$, $f^s = \Phi^{z+t_s}$, $\tilde{\Gamma}^s = \tilde{\Lambda}^{z+t_s}$ and $\Gamma^s = \tilde{\Lambda}^z$ (which does not depend on s). From Proposition 2.2 we already know that each f^s is contracting for the stable semi-metric with a Lipschitz

[3] We can add the property $\forall x \in \mathbb{R}$, $\mu(\{x\}) = \lim_{y\to 0^+} y\Im(f(x + iy))$, here for μ to be continuous in a, b, we need $\mu(\{a\}) = \mu(\{b\}) = 0$

parameter $\lambda < 1$ which can be chosen independent from s for sufficiently large s. Next, we express for any $s \in \mathbb{N}$ and any $i \in [n]$:

$$f^s(\Gamma^s)_i - \Gamma_i^s = \Phi^{z+t_s}(\tilde{\Lambda}^z)_i - \tilde{\Lambda}_i^z = t_s \tag{2.4}$$

Noting that for sufficiently large s, $\Im(\Phi^{z+t_s}(\tilde{\Lambda}^z)) = \Im(t_s) + \Im(\tilde{\Lambda}^z) \geq \frac{\Im(\tilde{\Lambda}^z)}{4} \geq \frac{\Im(z)}{4}$, we observe that $d_{\mathbb{H}}(\Im(f^s(\Gamma^s)_i), \Im(\Gamma_i^s)) \leq \frac{4|\Im(t^s)|}{\Im(z)} \underset{s\to\infty}{\longrightarrow} 0$. Thus, the assumptions of Proposition B.8 are satisfied, and we can conclude that there exists a constant $K > 0$ such that for all $s \in \mathbb{N}$:

$$\left\| \frac{\tilde{\Lambda}^{z+t_s} - \tilde{\Lambda}^z}{\sqrt{\Im(\tilde{\Lambda}^{z+t_s})\Im(\tilde{\Lambda}^z)}} \right\| \leq \frac{K|t_s|}{\inf_{i\in[n]} \sqrt{\Im(\tilde{\Lambda}^{z+t_s})\Im(\tilde{\Lambda}^z)}} \leq \frac{2K|t_s|}{\Im(z)}.$$

Besides, we can also bound:

$$\sqrt{\Im(\tilde{\Lambda}^{z+t_s})} \leq \frac{2\sqrt{\Im(\tilde{\Lambda}^z)}}{\Im(z)}(\Im(\tilde{\Lambda}^z) + Kt_s) \leq O(1),$$

This directly implies that $\|\tilde{\Lambda}^{z+t_s} - \tilde{\Lambda}^z\| \leq O(t_s) \underset{s\to\infty}{\longrightarrow} 0$, and consequently, the mapping $z \mapsto \tilde{\Lambda}^z$ is continuous on $\mathbb{H}$. □

Let us now show that $z \mapsto \tilde{\Lambda}^z$ is differentiable. Using the notation $f^t = \Phi^{z+t}$, we can decompose:

$$\begin{aligned}\left(\tilde{\Lambda}^{z+t} - \tilde{\Lambda}^z\right) &= \left(f^t(\tilde{\Lambda}^{z+t}) - f^t(\tilde{\Lambda}^z) + f^t(\tilde{\Lambda}^z) - f^0(\tilde{\Lambda}^z)\right) \\ &= \mathrm{Diag}_{i\in[n]}\left(\frac{1}{n}\mathrm{Tr}\left(\Sigma_i \tilde{Q}^{\tilde{\Lambda}^{z+t}} \frac{1}{n}\sum_{j=1}^n \frac{\tilde{\Lambda}_j^{z+t} - \tilde{\Lambda}_j^z}{\tilde{\Lambda}_j^{z+t}\tilde{\Lambda}_j^z}\Sigma_j \tilde{Q}^{\tilde{\Lambda}^z}\right)\right) + tI_n\end{aligned}$$

Now, we introduce the vector $a(t) = \left(\tilde{\Lambda}_i^{z+t} - \tilde{\Lambda}_i^z\right)_{1\leq i\leq n} \in \mathbb{C}^n$, and for any $D, D' \in \mathcal{D}_n(\mathbb{H})$, the matrix:

$$\Psi(D, D') = \left(\frac{1}{n}\frac{\mathrm{Tr}\left(\Sigma_i \tilde{Q}^D \Sigma_j \tilde{Q}^{D'}\right)}{D_j D'_j}\right)_{1\leq i,j\leq n} \in \mathcal{M}_n$$

we then have the equation:

$$a(t) = \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z+t})a(t) + t\mathbb{1}. \tag{2.5}$$

To be able to solve this equation we need the following property:

Lemma 2.8. *For all $z, z' \in \mathbb{H}$, $I_n - \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})$ is invertible.*

Proof. We are going to show that $I_n - \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})$ is injective. We introduce a vector $x \in \mathbb{C}^n$ such that $x = \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})x$. We can bound using the Cauchy-Schwatz inequality, with similar calculations as those in the proof of Proposition 2.2:

$$\begin{aligned}|x_i| &= \left| \frac{1}{n}\mathrm{Tr}\left(\Sigma_i \tilde{Q}^{\tilde{\Lambda}^z} \sum_{j=1}^n \frac{x_j \Sigma_j}{\sqrt{\Im(\tilde{\Lambda}^z)\Im(\tilde{\Lambda}^{z'})}} \frac{\sqrt{\Im(\tilde{\Lambda}^z)\Im(\tilde{\Lambda}^{z'})}}{\tilde{\Lambda}_j^z \tilde{\Lambda}_j^{z'}} \tilde{Q}^{\tilde{\Lambda}^{z'}}\right)\right| \\ &\leq \sup_{j\in[n]} \left| \frac{x_j}{\sqrt{\Im(\tilde{\Lambda}^z)\Im(\tilde{\Lambda}^{z'})}} \right| \sqrt{\Im(\tilde{\Lambda}_i^z) - \Im(z)}\sqrt{\Im(\tilde{\Lambda}_i^{z'}) - \Im(z')}\end{aligned}$$

therefore, if we denote $\|x\|_{\tilde\Lambda} \equiv \sup_{i\in[n]} \left| \frac{x_j}{\sqrt{\Im(\tilde\Lambda^{z'})\Im(\tilde\Lambda^{z'})}} \right|$, we have then the following bound:

$$\|x\|_{\tilde\Lambda^{z'},\tilde\Lambda^{z'}} \leq \|x\|_{\tilde\Lambda^{z'},\tilde\Lambda^{z'}} \sqrt{(1-\zeta(z,\tilde\Lambda^{z'}))(1-\zeta(z,\tilde\Lambda^{z'}))}$$

This directly implies that $x=0$ since we know that $\zeta(z,\tilde\Lambda^{z^{(\prime)}}) = \frac{\Im(w)}{\sup_{1\leq i\leq n}\Im(\Phi^w(\tilde\Lambda^{z^{(\prime)}}))_i} \in (0,1)$. □

The continuity of $z\mapsto\tilde\Lambda^z$ given by Proposition 2.7 and the continuity of the inverse operation on matrices (around $I_n - \Psi(\tilde\Lambda^z,\tilde\Lambda^z)$ which is invertible), allows us to let t tend to zero in the following equation

$$\frac{1}{t}a(t) = (I_n - \Psi(\tilde\Lambda^z,\tilde\Lambda^{z+t}))^{-1}\mathbb{1},$$

to obtain:

Proposition 2.9. *The mapping $z\mapsto\tilde\Lambda^z$ is analytic on $\mathbb{H}$, and satisfies:*

$$\frac{\partial\tilde\Lambda^z}{\partial z} = \mathrm{Diag}\left((I_n - \Psi(\tilde\Lambda^z,\tilde\Lambda^z))^{-1}\mathbb{1}\right)$$

We can then conclude that for all $i\in[n]$, the mappings $z\to -\frac{1}{\tilde\Lambda^z_i}$ are Stieltjes transforms.

Proposition 2.10. *For all $i\in[n]$, there exists a distribution $\tilde\mu_i$ with support on $\mathbb{R}_+$ whose Stieltjes transform is $z\mapsto -\frac{1}{\tilde\Lambda^z_i}$*

Proof. We now check the hypotheses of Theorem 2.6. We already know that $z\mapsto -\frac{1}{\tilde\Lambda^z_i}$ is analytical thanks to Proposition 2.9 and the lower bound $\Im(\tilde\Lambda^z_i)\geq\Im(z)>0$. Furthermore, $\forall z\in\mathbb{H}$:

$$\Im\left(-\frac{1}{\tilde\Lambda^z_i}\right) = \frac{\Im(\tilde\Lambda^z_i)}{|\tilde\Lambda^z_i|} > 0 \qquad \text{and} \qquad \Im\left(-\frac{z}{\tilde\Lambda^z_i}\right) = \frac{\Im(\tilde\Lambda^z_i/z)}{|\tilde\Lambda^z_i/z|} > 0,$$

since $\tilde\Lambda^z\in\mathcal{D}_{\Phi_z}$. Finally recalling from Lemma 2.4 that for all $y\in\mathbb{R}_+$, $\|\tilde Q^{\tilde\Lambda^{iy}}\|\leq\frac{|iy|}{\Im(iy)}=1$, we directly see that for all $j\in[n]$:

$$\frac{\tilde\Lambda^{iy}_j}{iy} = 1 + \frac{1}{iyn}\mathrm{Tr}(\Sigma_j\tilde Q^{\tilde\Lambda^{iy}}) \underset{y\to+\infty}{\longrightarrow} 1.$$

we can thus conclude with Theorem 2.6. □

We can then easily deduce from Proposition 2.10 that $\tilde g$ is a Stieltjes transform of the measure:

$$\tilde\mu = \left(\frac{n}{p}-1\right)\delta_0 - \frac{1}{p}\sum_{i=1}^n\tilde\mu_i,$$

where δ_0 is the Dirac measure on 0, and $\tilde\mu_i$ is the measure whose Stieltjes transform is $z\mapsto -\frac{1}{\tilde\Lambda^z_i}$. To show Theorem 1.2, one is just left to show that the supports of all the measures μ_i, $i\in[n]$ (and the measure $\tilde\mu$) are bounded. Actually, they are all bounded with[4]:

$$x_\Sigma \equiv \max\left(\frac{8p}{n},4\right)\|\Sigma\|.$$

We need the following preliminary Lemma.

[4]Recall that $\|\Sigma\|\equiv\sup_{i\in[n]}\|\Sigma_i\|$.

Lemma 2.11. *For all $z \in \mathbb{H}$ such that $\Re(z) \geq x_\Sigma$:*

$$\Re(\tilde\Lambda^z) \geq \frac{\Re(z) I_n}{2} \qquad \text{and} \qquad \|\tilde Q^{\tilde\Lambda^z}\| \leq 2$$

Proof. Let us show first that given $z \in \mathbb{H}$ such that $\Re(z) \geq x_\Sigma$, Φ^z is stable on $\mathcal{A} \equiv \mathcal{D}_n(\{z \in \mathbb{H} : \Re(z) \geq \frac{x}{2}\}) \cap \mathcal{D}_{\Phi^z}$. Given $L \in \mathcal{A}$:

$$\Re\left((\tilde Q^L)^{-1}\right) = I_p - \frac{1}{n}\sum_{i=1}^n \frac{\Re(\Lambda_i)\Sigma_i}{|L_i|^2} \geq \left(1 - \frac{2\|\Sigma\|}{x_\Sigma}\right) I_p \geq \frac{I_p}{2},$$

which implies (as in the proof of Lemma 2.3) that $\|\tilde Q^L\| \leq 2$. We can then bound:

$$\begin{aligned}\Re(\Phi^z(L)_i) &= x - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i \tilde Q^L \left(1 - \frac{1}{n}\sum_{j=1}^n \frac{\Re(L_j)\Sigma_j}{|L_j|^2}\right)\bar{\tilde Q}^L\right) \\ &\geq x - \frac{4}{n}\operatorname{Tr}(\Sigma_i) \geq \frac{x}{2} + \frac{x_\Sigma}{2} - \frac{4p\|\Sigma\|}{n} \geq \frac{x}{2} \geq \frac{x}{2}.\end{aligned}$$

Now, setting $L_0 \equiv (\Re(z) + i) I_n$ and $L_k = (\Phi^z)^k(L_0)$ for all $k > 0$, we know from the proof of Theorem 2.5 that $\tilde\Lambda^z = \lim_{k\to\infty} L_k$, thus, $\mathcal{A}$ being a closed set, $\tilde\Lambda^z \in \mathcal{A}$, and in particular, $\forall i \in [n]$, $\Re(\tilde\Lambda_i^z) \geq \frac{\Re(z)}{2}$. □

Note then that we can deduce, as an interesting (but not useful) side result from Lemma 2.8:

Lemma 2.12. *Given any $z, z' \in \mathbb{H}$, $\|\Psi(\tilde\Lambda^z, \tilde\Lambda^{z'})\| < 1$*

Proof. We know from Lemma 2.11 that $\forall z, z' \in \mathbb{H}$ such that $\Re(z), \Re(z') \geq x_0$:

$$\left|\Psi(\tilde\Lambda^z, \tilde\Lambda^{z'})_{i,j}\right| \leq \frac{1}{n}\left|\frac{\operatorname{Tr}\left(\Sigma_i \tilde Q^{\tilde\Lambda}\Sigma_j \tilde Q^{\tilde\Lambda'}\right)}{\tilde\Lambda_j \tilde\Lambda_j'}\right| \leq \frac{\|\Sigma\|^2 p}{n}\frac{4}{\Im(z)},$$

which can be as small as needed if we let $\Im(z)$ tend to $+\infty$. Therefore, from the continuity of $(z, z') \mapsto \|\Psi(\tilde\Lambda^z, \tilde\Lambda^{z'})\|$, one can deduce that $\forall z, z' \in \mathbb{H}$, $\|\Psi(\tilde\Lambda^z, \tilde\Lambda^{z'})\| \leq 1$ (otherwise, there would exists $z, z' \in \mathbb{H}$ such that $\Psi(\tilde\Lambda^z, \tilde\Lambda^{z'})$ would not be invertible, which would contradict Lemma 2.8). □

Let us finally conclude.

Proof of Theorem 1.2. We are just left to prove that $\tilde\mu$ has a compact support. Given $z \in \mathbb{H}$ let us denote for simplicity $x \equiv \Re(z)$ and $y \equiv \Im(z)$ such that $z = x + iy$. Then, with the notations of Lemma 2.11, assuming that $x \geq x_\Sigma$, we know that:

$$\begin{aligned}\Im(\tilde\Lambda_i^z) &= y + \frac{1}{n}\operatorname{Tr}\left(\Sigma_i \tilde Q^{\tilde\Lambda_i^z}\frac{1}{n}\sum_{j=1}^n \frac{\Im(\tilde\Lambda_j^z)\Sigma_j}{|\tilde\Lambda_j^z|^2}\bar{\tilde Q}^{\tilde\Lambda_i^z}\right) \\ &\leq y + \frac{4p\|\Sigma\|^2}{n^2}\sup_{j\in[n]}\frac{\Im(\tilde\Lambda_j^z)}{\Re(\tilde\Lambda_j^z)^2}\end{aligned}$$

Besides $\Re(\tilde\Lambda_j^z)^2 \geq \frac{x x_\Sigma}{4}$, and denoting $\nu \equiv \frac{16p\|\Sigma\|^2}{n^2 x_\Sigma}$ which allows us to eventually bound:

$$\sup_{j\in[n]}\Im(\tilde\Lambda_j^z) \leq y + \frac{\nu}{x}\sup_{j\in[n]}\Im(\tilde\Lambda_j^z)$$

This implies, for $x \geq 2\nu$:

$$\sup_{j\in[n]} \Im(\tilde\Lambda_j^z) \leq \frac{y}{1-\frac{\nu}{x}} \leq 2y \underset{y\to 0^+}{\longrightarrow} 0.$$

returning to the Stieltjes transform, that gives us:

$$\Im(\tilde g(x+iy)) = \frac{y}{x^2+y^2}\left(\frac{n}{p}-1\right) + \frac{1}{p}\sum_{i=1}^n \frac{\Im(\tilde\Lambda_i^z)}{\Re(\tilde\Lambda_i^z)^2+\Im(\tilde\Lambda_i^z)^2} \underset{y\to 0^+}{\longrightarrow} 0$$

That allows us to conclude that $\tilde\mu$ has compact support thanks to the relation between $\tilde\mu$ and $\tilde g$ given in Theorem 2.6. □

Let us end this subsection with an interesting connection between the deterministic Stieltjes transform $\tilde g$ and the random one defined for all $z \in \mathbb{H}$ as:

$$g(z) = \frac{1}{p}\operatorname{Tr}(Q^z) = \frac{1}{z} + \frac{1}{npz}\sum_{i=1}^n x_i^T Q^z x_i \tag{2.6}$$

Remark 2.13. A classical procedure in Random matrix theory is to disentangle the dependence between Q and x_i thanks to the introduction of a resolvent deprived of the contribution of x_i, namely: $Q_{-i}^z \equiv (I_n - \frac{1}{nz}X_{-i}X_{-i}^T)^{-1}$ where $X_{-i} = (x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n) \in \mathcal{M}_{p,n}$. The Schur identity then provides

$$Q^z = Q_{-i}^z + \frac{1}{nz}Q^z x_i x_i^T Q_{-i}^z \qquad \text{and} \qquad \frac{1}{z}Q^z x_i = \frac{Q_{-i}^z x_i}{z - \frac{1}{n}x_i^T Q_{-i}^z x_i} \tag{2.7}$$

To get closer to the form of $\tilde g$, one naturally introduce the diagonal matrix $\Lambda^{\in}\mathcal{D}_n(\mathbb{H})$, defined for any $i \in [n]$ as:

$$\Lambda_i^z \equiv z - \frac{1}{n}x_i^T Q_{-i} z_i,$$

and that will play a central role in next section (similar to the role $\tilde\Lambda_i^z$ played in this section). The identity $zQ^z - \frac{1}{n}XX^TQ^z = I_p$ and (2.7) finally give us from (2.6):

$$g(z) = \frac{1}{z} + \frac{1}{npz}\sum_{i=1}^n \frac{x_i^T Q_{-i} x_i}{\Lambda_i^z} = \frac{1}{z}\left(1-\frac{n}{p}\right) + \frac{1}{p}\sum_{i=1}^n \frac{1}{\Lambda_i^z} = g^{\Lambda^z}(z),$$

Given a mapping $\Delta : \mathbb{H} \ni z \mapsto \Delta^z \mathcal{D}_n(\mathbb{H})$, if one denotes:

$$\mathfrak{g}^{\Delta} : z \longmapsto \frac{1}{z}\left(1-\frac{n}{p}\right) + \frac{1}{p}\sum_{i=1}^n \frac{1}{\Delta_i^z},$$

then, finally[5]:

$$g = \mathfrak{g}^{\Lambda} \qquad \text{and} \qquad \tilde g \equiv \mathfrak{g}^{\tilde\Lambda}.$$

[5]Similarly, the Stieltjes transform of the spectral distribution of $\frac{1}{n}X^TX$ is $\check g = \check{\mathfrak{g}}^{\Lambda^z}$, where for all $D : \mathbb{H} \ni z \mapsto D^z \in \mathcal{D}_n(\mathbb{H})$, $\check g^{D^z} : z \longmapsto \frac{1}{p}\sum_{i=1}^n \frac{1}{D^z}$. This is straightforward from the identity $\frac{1}{\lambda_i^z} \equiv \check Q_{i,i}$ given later in 3.5. we see here that the spectrum of $\frac{1}{n}X^TX$ has $|n-p|$ less or more 0 eigenvalues than $\frac{1}{n}XX^T$.

3 Probabilistic study: concentration around the deterministic equivalent

3.1 Problem settlement

Notations

We will set in this section *quasi-asymptotic* results on random matrices, meaning that we will express convergence results for inequalities or concentration inequalities when important quantities like the number of rows p and the number of columns n converge to ∞. Just the rate of convergence is relevant, therefore, in order to remove smoothly the constants from the quasi-asymptotic result, we will introduce several notations. Below, the set of indexes Θ could be thought to be $\mathbb{N} \times \mathbb{N} \times \mathbb{C}$ or even something more elaborate like $\{(p, n, z) \in \mathbb{N} \times \mathbb{N} \times \mathbb{C}, p \leq n, \Im(z) \in (0, 1]\}$.

Following a previous work done in [**?**] , we will express concentration inequalities with operators which are set valued mappings and conversely (which gives natural meaning to $a \leq \alpha(t)$ for some $a, t \in \mathbb{R}$ and $\alpha : \mathbb{R} \mapsto 2^{\mathbb{R}}$). Classical mappings from $\mathbb{R}$ to $\mathbb{R}$ are identified as singleton-valued operators. An operator $\alpha : \mathbb{R} \mapsto 2^{\mathbb{R}}$ is said to be a positive probabilistic operator or a "concentration function" and we denote $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ iif it is maximally decreasing[6] $\{1\} \subset \mathrm{Ran}(\alpha)$ and $\mathrm{Dom}(\alpha) \subset \mathbb{R}_+$. Given a parameter $r > 0$, the r-moment of a concentration function $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ is defined as[7]:

$$M_\alpha^{(k)} \equiv r \int t^{r-1} \alpha(t) dt = \int \alpha\left(t^{\frac{1}{r}}\right) dt.$$

Given two families of probability operators $\alpha, \beta \in \mathcal{M}_{\mathbb{P}_+}^{\Theta}$, we will denote $\alpha \subset \beta$ iif there exist two constants $C, c > 0$ such that:

$$\forall \theta \in \Theta, \forall t \geq 0 : \qquad \alpha_\theta(t) \leq C \beta_\theta(ct),$$

A good illustration of this order relation is given in Lemma A.1.

Let us consider a family of random variables $(X_\theta)_{\theta \in \Theta} \in \mathbb{R}^\Theta$ and a family of positive probabilistic operators $(\alpha_\theta)_\theta \in \mathcal{M}_{\mathbb{P}_+}^{\Theta}$. If there exists some constants $C, c > 0$ such that $\forall \theta \in \Theta$:

$$\forall t \geq 0 : \quad \mathbb{P}(|X_\theta - X'_\theta| \geq t) \leq C \alpha_\theta(ct),$$

where $(X'_\theta)_{\theta \in \Theta}$ is a family of independent copies of X_θ, $\theta \in \Theta$, then we denote $X \in \alpha$ or if one needs to describe more precisely the dependence on Θ:

$$X_\theta \in \alpha_\theta, \quad \theta \in \Theta.$$

When there exists a family of deterministic parameters $(\tilde{X}_\theta)_{\theta \in \Theta}$ such that $\forall \theta \in \Theta$:

$$\forall t \geq 0 : \quad \mathbb{P}(|X_\theta - \tilde{X}_\theta| \geq t) \leq C \alpha_\theta(ct),$$

for some constants $C, c > 0$, one denotes $X \in \tilde{X} \pm \alpha$ or more simply $X \in O(m) \pm \alpha$, for any $(m_\theta)_{\theta \in \Theta}$ such that $|\tilde{X}| \leq O(m)$.

[6] Following the monotone operator theory (see for instance [BC11]), given an operator $\alpha : \mathbb{R} \mapsto 2^{\mathbb{R}}$, one denotes $\mathrm{Gra}(\alpha) \equiv \{(x, y) \in \mathbb{R}^2 : y \in \alpha(x)\}$, the graph of α, $\mathrm{Dom}(\alpha) \equiv \{x \in \mathbb{R}, f(x) \neq \emptyset\}$, the domain of α and $\mathrm{Ran}(\alpha) \equiv \{y \in \mathbb{R}, \exists x \in \mathrm{Dom}(\alpha) : y \in \alpha(x)\}$ then α is maximally decreasing iif it satisfies the implication $\forall x, y \in \mathbb{R}^2$:

$$\forall (w, z) \in \mathrm{Gra}(\alpha) : (x - w)(z - y) \geq 0 \qquad \Longrightarrow \qquad (x, y) \in \mathrm{Gra}(\alpha).$$

[7] One can naturally define the integral of maximally monotone operators as the integral on continuous points.

We rely on real-valued functional to extend those notations to random vectors. Given a family of normed vector spaces $(E_\theta, \|\cdot\|)_{\theta\in\Theta}$ and a family of random vectors $(X_\theta)_\theta \in \prod_{\theta\in\Theta} E_\theta$ we will work on concentrations (for assumptions and results)[8]:

$$f(X_\theta) \in \alpha_\theta, \qquad \theta \in \Theta,\ f : E_\theta \to \mathbb{R},\ \text{1-Lipschitz.}$$

that we will rather denote for simplicity:

$$X \in \alpha.$$

This choice of notation is not ambiguous because in $\mathbb{R}$, the concentration of Lipschitz observation is equivalent to the concentration of the random variables themselves.

If, in addition, we are given a family of deterministic vectors $(\tilde{X}_\theta)_{\theta\in\Theta} \in \prod_{\theta\in\Theta} E_\theta$ such that $X \in \alpha$ and[9]:

$$u(X_\theta) \in u(\tilde{X}_\theta) \pm \alpha_\theta, \qquad \theta \in \Theta,\ u : E_\theta \to \mathbb{R},\ \text{linear},\ \|u\| \leq 1,$$

then we denote:

$$X \in \tilde{X} \pm \alpha,$$

and when there exists some family of positive parameters $(\sigma_\theta)_{\theta\in\Theta} \in \mathbb{R}_+^\Theta$ such that $\|\tilde{X}\| \leq O(\sigma)$, we denote $X \in O(\sigma) \pm \alpha$.

We denote Id, the identity operator $t \mapsto \{t\}$, and for simplicity, given a parameter $\lambda \in \mathbb{R}$ and an operator α, $\alpha(\lambda\,\mathrm{Id}) \equiv \alpha \circ (\lambda\,\mathrm{Id})$. The same way, $\sqrt{\mathrm{Id}} : t \mapsto \{\sqrt{t}\}$, it satisfies $\mathrm{Dom}(\sqrt{\mathrm{Id}}) \subset \mathbb{R}_+$. Given two operators $\alpha, \beta : \mathbb{R} \to 2^{\mathbb{R}}$, the parallel sum and product of α and β are respectively denoted $\alpha \boxplus \beta = (\alpha^{-1} + \beta^{-1})^{-1}$ and $\alpha \boxtimes \beta = (\alpha^{-1} \cdot \beta^{-1})^{-1}$ (where the inverse of an operator α is classically defined as $\alpha^{-1} : y \mapsto \{x \mid y \in \alpha(x)\}$).

Setting

By default, the sets of matrices $\mathcal{M}_{p,n}$ (in particular $\mathcal{D}_n \subset \mathcal{M}_n$), $p, n \in \mathbb{N}$ are endowed with Hilbert-Schmidt norms $\|\cdot\|_{\mathrm{HS}}$ and the sets of random vectors $\mathbb{R}^p$, $p \in \mathbb{N}$ are endowed with the ℓ_2 norm.

In what follows, we consider a constant $\gamma > 0$ and introduce:

$$\Theta_\gamma \equiv \{(n, p, z) \in \mathbb{N}^2 \times \mathbb{H}, n \geq \gamma p\}.$$

the index set that will direct our quasi-asymptotic results. With the notations we introduced, $\gamma \leq O(1)$, there for seeing p, n as families of parameters indexed with Θ_γ, one has $p \leq O(n)$.

Considering a family of random matrices $X = (X_{(n,p)})_{(n,p)\in\Theta_\gamma}$, given $i \in \mathbb{N}$, let us naturally denote $x_i \equiv (x_i^{(n,p)})_{(n,p)\in\Theta_\gamma, n\geq i}$, the family of the i^{th} column of X and recall the family of means, of centered and non-centered empirical covariance matrices for all $i \in \mathbb{N}$:

$$\Sigma_i \equiv \mathbb{E}[x_i x_i^T].$$

[8] That means that there exists some constants $C, c > 0$ such that $\forall \theta \in \Theta$ for all 1-Lipschitz mappings $f : E_\theta \to \mathbb{R}$:

$$\forall t \geq 0 : \quad \mathbb{P}(|f(X_\theta) - f(X'_\theta)| \geq t) \leq C\alpha_\theta(ct),$$

[9] that means that there exist some constants $C, c > 0$ such that $\forall \theta \in \Theta$ for all linear form $u : E_\theta \to \mathbb{R}$ such that $\|u\| \leq 1$:

$$\forall t \geq 0 : \quad \mathbb{P}(|u(X_\theta - \tilde{X}_\theta)| \geq t) \leq C\alpha_\theta(ct).$$

Assumptions

Let us assume the following properties are satisfied:

(A1) $x_i, \dots, x_n$ are independent,

(A2) there exists some concentration function $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ (independent with n, p) and a sequence $\eta \in \mathbb{R}_+^{\mathbb{N}}$ such that:

$$x_i \in \alpha \circ \left(\frac{\mathrm{Id}}{\eta} \right), i \in \mathbb{N} \qquad X \in \alpha \circ \left(\frac{\mathrm{Id}}{\sqrt{n}\eta} \right), \qquad O(1) \leq \eta \leq O(\sqrt{p}),$$

(A3) $\|\Sigma_i\| \leq O(1)$, $i \in [n]$.

Later, one will have to make one further assumption on the existence of moments of α, depending on the result (Assumption **(A4.a)** and **(A4.b)**).

3.2 First concentration results and general strategy

Concentration of the resolvent

It is somehow convenient to study simultaneously the so-called "co-resolvent" $\check{Q}$ defined as:

$$\check{Q}^z = \left(I_n - \frac{1}{zn} X^T X \right)^{-1} \in \prod_{(n,p) \in \Theta_\gamma, \Im(z) \in (0,1]} \mathcal{M}_{p,n}.$$

Lemma 2.3 already provided the bounds:

$$\|Q^z\| \leq \frac{|z|}{\Im(z)} \leq O(1) \qquad \text{and} \qquad \|\check{Q}^z\| \leq O(1) \tag{3.1}$$

(since in our regime: $O(1) \leq \Im(z) \leq |z| \leq O(1)$). Further note that the identity $Q \frac{1}{zn} X X^T = Q - I_p$ provides:

$$\left\| \frac{1}{n} Q^z X \right\| \leq \frac{|z|}{\sqrt{n}} \sqrt{\left\| \frac{1}{zn} Q^z X X^T Q^z \right\|} \leq \frac{|z|}{\sqrt{n}} \sqrt{\|(Q^z)^2 - Q^z\|} \leq O\left(\frac{1}{\sqrt{n}} \right) \tag{3.2}$$

Proposition 3.1. $Q^z, \check{Q}^z \in \alpha(\sqrt{n}\,\mathrm{Id})$.

Proof. Introducing the mappings $\mathcal{Q} : \mathcal{M}_{p,n} \to \mathcal{M}_p$ and $\check{\mathcal{Q}} : \mathcal{M}_{p,n} \to \mathcal{M}_n$ defined as:

$$\mathcal{Q}(M) = \left(I_p - \frac{MM^T}{zn} \right)^{-1} \qquad \text{and} \qquad \check{\mathcal{Q}}(M) = \left(I_n - \frac{M^T M}{zn} \right)^{-1},$$

it is sufficient to show that $\mathcal{Q}$ and $\check{\mathcal{Q}}$ are both $O(1/\sqrt{n}\Im(z)^2)$-Lipschitz (for the Hilbert-Schmidt norm). For any $M \in \mathcal{M}_{n,p}$ and any $H \in \mathcal{M}_{p,n}$, we can bound:

$$\left\| d\mathcal{Q}|_M \cdot H \right\|_{\mathrm{HS}} = \left\| \mathcal{Q}(M) \frac{1}{zn} (MH^T + HM^T) \mathcal{Q}(M) \right\|_{\mathrm{HS}} \leq O\left(\frac{\|H\|_{\mathrm{HS}}}{\sqrt{n}} \right),$$

thanks to (3.1) and (3.2). The same holds for $\check{Q}^z$. □

We also provide here the expression of the concentration of QX and $X^T \check{Q}$ that will be useful later.

Lemma 3.2. $Q^z X = X \check{Q}^z \in \alpha \left(\Im(z)^2 \,\mathrm{Id} \right)$

Proof. Let us look at the variations of the mapping $\mathcal{R} : \mathcal{M}_{p,n} \to \mathcal{M}_{p,n}(\mathbb{C})$ defined as:

$$\mathcal{R}(M) = \left(I_p - \frac{MM^T}{zn}\right)^{-1} M.$$

to show the concentration of $Q^z X = \mathcal{R}(X)$. For all $H, M \in \mathcal{M}_{p,n}$ (and with the notation $\mathcal{Q}(M) = \left(I_p - \frac{MM^T}{zn}\right)^{-1}$ given in the proof of Proposition 3.1), let us bound:

$$\left\| d\mathcal{R}|_M \cdot H \right\|_{\mathrm{HS}} \leq \left\| \mathcal{Q}(M) \frac{1}{zn}(MH^T + HM^T)\mathcal{Q}(M) M \right\|_{\mathrm{HS}} + \|\mathcal{Q}(M) H\|_{\mathrm{HS}} \leq O\left(\|H\|_{\mathrm{HS}}\right).$$

□

Strategy to approach the expectation of the resolvent

Once Proposition 3.1 is set, Lemma A.6 provides the concentration around the expectation (since α has the first moment bounded):

$$Q^z \in \mathbb{E}[Q^z] \pm \alpha(\sqrt{n}\,\mathrm{Id}) \qquad \text{(in Hilbert Schmidt norm)}.$$

We are then left to bound $\|\mathbb{E}[Q^z] - \tilde{Q}^{\tilde{\Lambda}^z}\|_*$ and $\|\mathbb{E}[Q^z] - \tilde{Q}^{\tilde{\Lambda}^z}\|_{\mathrm{HS}}$ to be able to set concentration around our deterministic equivalent thanks to Lemma A.3. It seems difficult to prove this result straightforwardly, one will rather rely on two proxies for $\tilde{\Lambda}^z$ (namely $\check{\Lambda}$ and $\hat{\Lambda}$) to progressively approach $\tilde{Q}^{\tilde{\Lambda}^z}$ from $\mathbb{E}[Q^z]$. To identify these proxies let us consider a certain deterministic diagonal matrix $\Delta \in \mathcal{D}_n(\mathbb{H})$ and express the difference:

$$\mathbb{E}[Q^z] - \tilde{Q}^{\Delta} = \mathbb{E}\left[Q^z \left(\frac{1}{nz} XX^T - \Sigma^{\Delta}\right) \tilde{Q}^{\Delta}\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Q^z \left(\frac{x_i x_i^T}{z} - \frac{\Sigma_i}{\Delta_i}\right) \tilde{Q}^{\Delta}\right].$$

To pursue the estimation of the expectation, one needs to control the dependence between Q^z and x_i. For that purpose, recall the Schur identity given in (2.7) $\frac{1}{z}Q^z x_i = \frac{Q^z_{-i}x_i}{\Lambda^z_i}$, where we also recall that $\forall i \in [n]$: $\Lambda^z_i \equiv z - \frac{1}{n}x_i^T Q^z_{-i} x_i$. It is then possible to express:

$$\begin{aligned}
\mathbb{E}[Q^z] - \tilde{Q}^{\Delta} &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Q^z_{-i}\left(\frac{x_i x_i^T}{\Lambda^z_i} - \frac{\Sigma_i}{\Delta_i}\right)\tilde{Q}^{\Delta}\right] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(Q^z_{-i} - Q^z)\frac{\Sigma_i}{\Delta_i}\tilde{Q}^{\Delta}\right] \\
&= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^{\Delta} + \delta_i^{\Delta}, \qquad (3.3)
\end{aligned}$$

where we introduced, for any $i \in \mathbb{N}$, $\Delta \in \mathcal{D}_n(\mathbb{H})$ the matrices $\varepsilon_i^{\Delta}, \delta_i^{\Delta}$ defined as:

- $\varepsilon_i^{\Delta} = \mathbb{E}\left[Q^z_{-i} x_i x_i^T \left(\frac{1}{\Lambda^z_i} - \frac{1}{\Delta_i}\right)\tilde{Q}^{\Delta}\right]$
- $\delta_i^{\Delta} = \frac{1}{\Delta}\mathbb{E}\left[(Q^z_{-i} - Q^z)\Sigma_i \tilde{Q}^{\Delta}\right]$.

It can be shown easily that the matrix δ_i^{Δ}(when $\Im(\Delta) > 0$) is of small size:

Lemma 3.3.

$$\|\delta_i^{\Delta}\| \leq O\left(\frac{1}{n\Im(z)^4\Im(\Delta)}\right)$$

This result is a consequence of Lemma 2.3 that provides the bound $\|\tilde{Q}^{\Delta}\| \leq O(1)$ and thanks to the following result:

Lemma 3.4. $\|\mathbb{E}[Q^z_{-i}] - \mathbb{E}[Q^z]\| \leq O\left(\frac{1}{n\Im(z)^3}\right)$.

This Lemma relies on the formula:

$$Q^z - Q^z_{-i} = \frac{1}{n\Lambda^z_i} Q^z_{-i} x_i x_i^T Q^z_{-i} \tag{3.4}$$

consequence to the Schur identities (2.7). Naturally one then want to look for upper bounds on $\frac{1}{\Lambda^z}$. That can be obtained through two ways. The first one relies on the identity: (3.1) to express:

$$\begin{aligned}\frac{1}{\Lambda^z} = \operatorname{Diag}_{i\in[n]}\left(\frac{1}{z - \frac{1}{n}x_i^T Q^z_{-i} x_i}\right) &= \frac{1}{z}\operatorname{Diag}_{i\in[n]}\left(1 + \frac{\frac{1}{n}x_i^T Q^z_{-i}x_i}{z - \frac{1}{n}x_i^T Q^z_{-i}x_i}\right)\\ &= \frac{1}{z}\left(I_n + \frac{1}{zn}\operatorname{Diag}(X^T Q^z X)\right)\\ &= \operatorname{Diag}\left(\frac{1}{z}I_n + \frac{1}{nz}X^T X \check{Q}^z\right) = \frac{1}{z}\operatorname{Diag}(\check{Q}^z).\end{aligned} \tag{3.5}$$

The second one relies on the inequality $|\Lambda^z_i| = \frac{|z\Lambda^z_i|}{|z|} \geq \frac{\Im(z\Lambda^z_i)}{|z|}$ and the identity:

$$\Im(\Lambda^z_i) = \Im(z) - x_i^T \Im(Q^z_{-i}) x_i = \Im(z) + x_i^T Q^z_{-i}\frac{\Im(z)}{n|z|^2} X_{-i}X_{-i}^T \bar{Q}^z_{-i} x_i \geq \Im(z). \tag{3.6}$$

Both bounds (3.5) and (3.6) can lead to:

Lemma 3.5. $\Lambda^z \in \mathcal{D}_n(\mathbb{H})$ *and* $\frac{1}{\Lambda^z} \leq \frac{1}{\Im(z)} \leq O(1)$.

We are now able to prove Lemma 3.4:

Proof of Lemma 3.4. For any deterministic vector $u, v \in \mathbb{C}^p$, employing (3.4), let us bound with Cauchy Schwarz inequality:

$$\begin{aligned}|u^*(\mathbb{E}[Q^z_{-i}] - \mathbb{E}[Q^z])v| &= \frac{1}{n}\left|\mathbb{E}\left[u^* Q^z_{-i} x_i x_i^T Q^z_{-i} v \frac{1}{\Lambda^z_i}\right]\right|\\ &= \frac{1}{n}\sqrt{\mathbb{E}\left[u^* Q^z_{-i} x_i x_i^T \bar{Q}^z_{-i} u\left|\frac{1}{\Lambda^z_i}\right|\right]\mathbb{E}\left[v^* \bar{Q}^z_{-i} x_i x_i^T Q^z_{-i} v\left|\frac{1}{\Lambda^z_i}\right|\right]}\\ &= \sqrt{\mathbb{E}\left[u^* Q^z_{-i}\Sigma_i \bar{Q}^z_{-i} u\right]\mathbb{E}\left[v^* \bar{Q}^z_{-i}\Sigma_i Q^z_{-i} v\right]}\, O\left(\frac{1}{n\Im(z)}\right) \leq O\left(\frac{1}{n\Im(z)^3}\right)\end{aligned}$$

thanks to Lemmas 3.1 and 3.5. □

Now that we know how to bound δ^Δ_i for all $i \in [n]$ and all $\Delta \in \mathcal{D}_n(\mathbb{H})$, we are left to bound $\varepsilon^\Delta_i = \mathbb{E}\left[Q^z_{-i} x_i x_i^T \left(\frac{1}{\Lambda^z_i} - \frac{1}{\Delta_i}\right)\tilde{Q}^\Delta\right]$ which seems possible if Δ either takes the values:

$$\hat{\Lambda}^z \equiv \mathbb{E}[\Lambda^z] \qquad \text{or} \qquad \check{\Lambda}^z \equiv \frac{1}{\mathbb{E}[1/\Lambda^z]}.$$

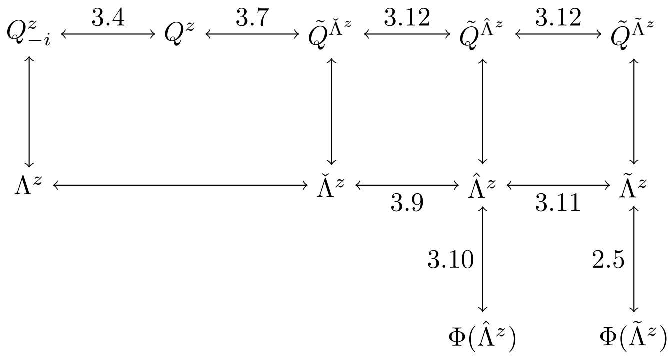
We will see that both choice are relevant, but not used the same way if one tries to bound the nuclear norm or the Hilbert Schmidt norm. The next two subsections will present those two approaches —- the first one requires the second moment of α to be bounded, the second one requires the fourth moment of α to be bounded.

3.3 Convergence in nuclear norm when α has bounded second moment

This subsection relies on the supplementary assumption to **(A1-3)**:

(A4.a) $\int_{\mathbb{R}_+} t\alpha(t)dt < \infty$

The proof process follows the diagram:



Reaching $\check\Lambda^z \equiv 1/\mathbb{E}[1/\Lambda^z]$

Lemma 3.6. *Given any sequence of deterministic matrices* $A \in \prod_{(n,p)\in\Theta_\gamma} \mathcal{M}_p$*:*

$$x_i^T Q^z_{-i} A x_i \in \alpha \circ \min\left(\frac{\mathrm{Id}}{\eta\|A\|_{\mathrm{HS}}}, \sqrt{\frac{\mathrm{Id}}{\eta^2\|A\|}}\right) \subset \alpha \circ \sqrt{\frac{\mathrm{Id}}{\eta\sqrt{p}\|A\|}}.$$

Proof. Note that $\mathbb{E}[x_i^T Q^z_{-i} A x_i] = \mathrm{Tr}(\Sigma_i \mathbb{E}[Q^z_{-i}]A)$ and let us bound:

$$\begin{aligned}&\left|x_i^T Q^z_{-i} A x_i - \mathrm{Tr}(\Sigma_i \mathbb{E}[Q^z_{-i}]A)\right|\\ &\qquad \le \left|x_i^T Q^z_{-i} A x_i - \mathrm{Tr}(\Sigma_i Q^z_{-i} A)]\right| + \left|\mathrm{Tr}(\Sigma_i Q^z_{-i} A) - \mathrm{Tr}(\Sigma_i \mathbb{E}[Q^z_{-i}]A)\right|.\end{aligned}$$

Therefore, the Hanson-Wright inequality provided in Theorem A.7 gives the existence of some constants $C, c > 0$ such that for all $\theta \in \Theta_\gamma$, for all $z \in \mathbb{H}$:

$$\begin{aligned}&\mathbb{P}\left(\left|x_i^T Q^z_{-i} A x_i - \mathbb{E}[x_i^T Q^z_{-i} A x_i]\right| \ge t\right)\\ &\qquad \le \mathbb{E}\left[\mathbb{P}\left(\left|x_i^T Q^z_{-i} A x_i - \mathrm{Tr}(\Sigma_i Q^z_{-i} A)]\right| \ge \frac{t}{2} \mid X_{-i}\right)\right]\\ &\qquad\qquad + \mathbb{P}\left(\left|\mathrm{Tr}(\Sigma_i Q^z_{-i} A) - \mathrm{Tr}(\Sigma_i \mathbb{E}[Q^z_{-i}]A)\right| \ge \frac{t}{2}\right)\\ &\qquad \le \mathbb{E}\left[\alpha \circ \min\left(\frac{ct}{\eta\|Q^z_{-i}A\|_{\mathrm{HS}}}, \sqrt{\frac{ct}{\eta^2\|Q^z_{-i}A\|}}\right)\right] + \alpha\left(\frac{c\sqrt{n}t}{\eta\sqrt{n}\sqrt{|z|}\|A\|}\right),\\ &\qquad \le \alpha \circ \min\left(\frac{\mathrm{Id}}{\eta\|A\|_{\mathrm{HS}}}, \sqrt{\frac{\mathrm{Id}}{\eta^2\|A\|}}\right) + \alpha\left(\frac{ct}{\eta\|A\|}\right)\end{aligned}$$

one can then conclude thanks to the inequality $\|A\|_{\mathrm{HS}} \le \sqrt{p}\|A\|$ and Lemma A.1. □

We now have all the preliminary results to prove:

Proposition 3.7. *When* α *has bounded second moment, one can bound:*

$$\left\|\mathbb{E}[Q^z] - \tilde Q^{\check\Lambda^z}\right\|_* \le O\left(\eta\sqrt{p}\right)$$

Proof. Let us consider a deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\| \leq 1$. With the notation introduced to set (3.3), we can decompose:

$$\operatorname{Tr}\left(A\left(\mathbb{E}\left[Q^z\right]-\tilde{Q}^{\check{\Lambda}^z}\right)\right)=\frac{1}{n}\sum_{i=1}^{n}\operatorname{Tr}(A\varepsilon_i^{\check{\Lambda}^z})+\operatorname{Tr}(A\delta_i^{\check{\Lambda}^z})$$

First the term with $\delta_i^{\check{\Lambda}^z}$ can be bounded with Lemma 3.3:

$$|\operatorname{Tr}(A\delta_i^{\check{\Lambda}^z})| \leq O\left(\frac{p\|A\|}{n}\right) \leq O(1)$$

Second, let us estimate:

$$\begin{aligned}\left|\operatorname{Tr}(A\varepsilon_i^{\check{\Lambda}^z})\right| &= \left|\mathbb{E}\left[\left(x_i^T\tilde{Q}^{\check{\Lambda}^z}AQ_{-i}^z x_i-\mathbb{E}[x_i^T\tilde{Q}^{\check{\Lambda}^z}AQ_{-i}^z x_i]\right)\frac{1}{\Lambda_i^z}\right]\right| \\ &\leq \mathbb{E}\left[\left|x_i^T\tilde{Q}^{\check{\Lambda}^z}AQ_{-i}^z x_i-\mathbb{E}[x_i^T\tilde{Q}^{\check{\Lambda}^z}AQ_{-i}^z x_i]\right|\right]O\left(1\right).\end{aligned}$$

The integration of the concentration:

$$x_i^T\tilde{Q}^{\check{\Lambda}^z}AQ_{-i}^z x_i \in \alpha\circ\left(\sqrt{\frac{\operatorname{Id}}{\eta\sqrt{p}\|A\|}}\right)$$

provided by Lemma 3.6 leads to $\left|\operatorname{Tr}(A\varepsilon_i^{\check{\Lambda}^z})\right| \leq O\left(\eta\sqrt{p}\|A\|\right)$ thanks to Lemma A.2 and the hypothesis $\eta \leq O(\sqrt{p})$. Combining the bounds on $\delta_i^{\check{\Lambda}^z}$ and the bounds on $\varepsilon_i^{\check{\Lambda}^z}$ for $i \in [n]$ (the bounding constants are the same), one finally obtains the result of the proposition. □

Reaching $\hat{\Lambda}^z \equiv \mathbb{E}[\Lambda^z]$

Let us now go from $\check{\Lambda}^z \equiv 1/\mathbb{E}[1/\Lambda^z]$ to:

$$\hat{\Lambda}^z \equiv \mathbb{E}[\Lambda^z] \in \mathcal{D}_n(\mathbb{C}),$$

Applying Lemma 3.6 in the case $A \equiv I_p$ one gets:

Lemma 3.8. $\forall i \in [n]:\ \Lambda_i^z \in \hat{\Lambda}_i^z\alpha\circ\left(\sqrt{\frac{n\operatorname{Id}}{\eta\sqrt{p}}}\right)$.

Lemma 3.9. $\left\|\frac{1}{\hat{\Lambda}^z}-\frac{1}{\check{\Lambda}^z}\right\| \leq O\left(\frac{\eta\sqrt{p}}{n}\right)$.

Proof. We know from (3.6) that $\frac{1}{\Lambda_i^z}, \frac{1}{\mathbb{E}[\Lambda_i^z]} \leq O(1)$, the concentration given by Lemma 3.8 then allows us to bound:

$$\begin{aligned}\left|\mathbb{E}\left[\frac{1}{\Lambda_i^z}\right]-\frac{1}{\mathbb{E}[\Lambda_i^z]}\right| &\leq \mathbb{E}\left[\left|\frac{1}{\Lambda_i^z}-\frac{1}{\mathbb{E}[\Lambda_i^z]}\right|\right] \leq \mathbb{E}\left[\frac{|\Lambda_i^z-\mathbb{E}[\Lambda_i^z]|}{|\Lambda_i^z\mathbb{E}[\Lambda_i^z]|}\right] \\ &\leq O\left(\mathbb{E}[|\Lambda_i^z-\mathbb{E}[\Lambda_i^z]|]\right) \leq O\left(\frac{\eta\sqrt{p}}{n}\right).\end{aligned}$$

□

Finally reaching $\tilde{\Lambda}^z$

To show the convergence of $\hat{\Lambda}^z$ towards $\tilde{\Lambda}^z$ solution to $\tilde{\Lambda}^z = \Phi(\tilde{\Lambda}^z)$, we need Proposition B.8 bounding the distance to a fixed point of a contracting mapping for the semi-metric $d_{\mathbb{H}}$.

Lemma 3.10. $\|\Phi(\hat{\Lambda}^z) - \hat{\Lambda}^z\| \leq O\left(\frac{\eta\sqrt{p}}{n}\right)$.

Proof. Let us bound:

$$|\Phi(\hat{\Lambda}^z)_i - \hat{\Lambda}^z_i| = \left|\mathbb{E}\left[\frac{1}{n}\operatorname{Tr}(\Sigma_i \tilde{Q}^{\hat{\Lambda}^z}) - \frac{1}{n} x_i^T Q^z_{-i} x_i\right]\right| = \frac{1}{n}\left|\operatorname{Tr}\left(\Sigma_i(\tilde{Q}^{\hat{\Lambda}^z} - \mathbb{E}[Q^z_{-i}])\right)\right|$$
$$\leq O\left(\frac{1}{n}\|\tilde{Q}^{\hat{\Lambda}^z} - \tilde{Q}^{\check{\Lambda}^z}\|_* + \frac{1}{n}\|\tilde{Q}^{\check{\Lambda}^z} - \mathbb{E}[Q^z]\|_* + \frac{1}{n}\|\mathbb{E}[Q^z] - \mathbb{E}[Q^z_{-i}]\|_*\right).$$

One can then conclude thanks to Lemmas 3.4, Proposition 3.7 and the bound:

$$\|\tilde{Q}^{\hat{\Lambda}^z} - \tilde{Q}^{\check{\Lambda}^z}\|_* = \sup_{\|A\|\leq 1}\left\|\operatorname{Tr}\left(A\tilde{Q}^{\check{\Lambda}^z}\frac{1}{n}\sum_{i=1}^n\left(\mathbb{E}\left[\frac{1}{\Lambda^z_i}\right] - \frac{1}{\hat{\Lambda}^z_i}\right)\tilde{Q}^{\hat{\Lambda}^z}\right)\right\| \tag{3.7}$$
$$\leq O\left(p\sup_{i\in[n]}\left|\mathbb{E}\left[\frac{1}{\Lambda^z_i}\right] - \frac{1}{\hat{\Lambda}^z_i}\right|\right) \leq O\left(\frac{p^{3/2}\eta}{n}\right),$$

thanks to Lemma 3.9. One then conclude thanks to the bounds $\frac{p^{3/2}}{n^2}, \frac{p}{n^2} \leq \frac{\sqrt{p}}{n}$. □

We now have all the elements to show the convergence of $\hat{\Lambda}^z$ to $\tilde{\Lambda}^z$.

Proposition 3.11. $\|\hat{\Lambda}^z - \tilde{\Lambda}^z\| \leq O\left(\frac{\eta}{\sqrt{n}}\right)$

Proof. We know from Lemma 2.4 that $\sup_{i\in[n]}\Im(\tilde{\Lambda}^z_i) \leq \|\Phi(\tilde{\Lambda}^z)\| \leq O(1)$, then Proposition 2.2 provides:

$$d_{\mathbb{H}}(\Phi^z(\tilde{\Lambda}^z), \Phi^z(\hat{\Lambda}^z)) \leq \sqrt{\left(1 - \frac{\Im(z)}{\sup_{i\in[n]}\Im(\tilde{\Lambda}^z_i)}\right)\left(1 - \frac{\Im(z)}{\sup_{i\in[n]}\Im(\hat{\Lambda}^z_i)}\right)} d_{\mathbb{H}}(\tilde{\Lambda}^z, \hat{\Lambda}^z)$$
$$\leq \lambda d_{\mathbb{H}}(\tilde{\Lambda}^z, \hat{\Lambda}^z),$$

with $1 - \lambda \geq O(1)$. Moreover, Lemma 3.10 and the bounds $\Im(\Phi^z(\hat{\Lambda}^z)), \Im(\hat{\Lambda}^z) \geq \Im(z)$ provide:

$$d_{\mathbb{H}}(\Phi^z(\hat{\Lambda}^z), \hat{\Lambda}^z) \leq o(1) \qquad \text{and consequently:} \qquad d_{\mathbb{R}^*_+}(\Im(\Phi^z(\hat{\Lambda}^z)), \Im(\hat{\Lambda}^z)) \leq o(1)$$

(recall that by assumption $\eta \leq o(\sqrt{p}) \leq o(\sqrt{n})$). The conditions of Proposition B.8 are satisfied and one can finally set that:

$$d_{\mathbb{H}}(\hat{\Lambda}^z, \tilde{\Lambda}^z) \leq O\left(\left\|\frac{\hat{\Lambda}^z - \Phi(\hat{\Lambda}^z)}{\sqrt{\Im(\hat{\Lambda}^z)\Im(\tilde{\Lambda}^z)}}\right\|\right) \leq O\left(\frac{\eta\sqrt{p}}{n}\right)$$

Then, recalling that $\forall i \in [n]$, $\Im(\hat{\Lambda}^z_i) \geq O(1)$ and from Lemma 2.4 that $\Im(\tilde{\Lambda}^z_i) \leq O(1)$, one can bound:

$$\forall i \in [n]: \quad \sqrt{\frac{\Im(\hat{\Lambda}^z_i)}{\Im(\tilde{\Lambda}^z_i)}} \leq \sqrt{\frac{\Im(\tilde{\Lambda}^z_i)}{\Im(\hat{\Lambda}^z_i)}} + d_{\mathbb{R}^*_+}(\Im(\hat{\Lambda}^z), \Im(\tilde{\Lambda}^z)) \leq O(1),$$

which implies that $\Im(\hat{\Lambda}_i^z) \leq O(1)$. One then has all the elements to conclude that:

$$\|\hat{\Lambda}^z - \tilde{\Lambda}^z\| \leq O\left(d_{\mathbb{H}}(\hat{\Lambda}^z, \tilde{\Lambda}^z)\right) \leq O\left(\frac{\eta\sqrt{p}}{n}\right)$$

□

Similar calculus as the one done in (3.7) to deduce a bound $\|\tilde{Q}^{\hat{\Lambda}^z} - \tilde{Q}^{\check{\Lambda}^z}\|_*$ from a bound on $\|\hat{\Lambda}^z - \check{\Lambda}^z\|$ then leads to:

Proposition 3.12 (Convergence in nuclear norm)**.** $\left\|\mathbb{E}\left[Q^z\right] - \tilde{Q}^{\tilde{\Lambda}^z}\right\|_* \leq O\left(\frac{p^{3/2}\eta}{n}\right)$.

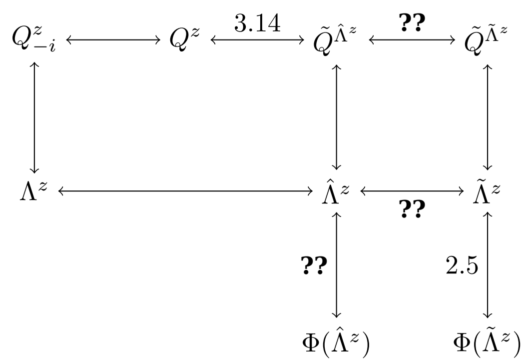
Finally Lemma A.3 allows us to combine this result with the concentration of Q^z given in Proposition 3.1 to finally obtain the first result of Theorem 1.3.

3.4 Convergence in the HS norm when α has bounded fourth moment

This subsection relies on the supplementary assumption to **(A1-3)**:

(A4.b) $\int_{\mathbb{R}_+} t^3\alpha(t)dt < \infty$

The proof process for this section is shown in the following diagram.



Recall that $\hat{\Lambda} \equiv \mathbb{E}[\Lambda]$, we will demonstrate that the distance between $\mathbb{E}[Q^z]$ and $\tilde{Q}^{\hat{\Lambda}^z}$ in Hilbert Schmidt norm is $1/n$ better than in the nuclear norm. The proof follows generally the same structure as that of Proposition 3.7. To get the best bound possible, one needs a refined version of Lemma 3.6.

Lemma 3.13. *Given any sequence of deterministic matrices $A \in \prod_{(n,p)\in\Theta_\gamma} \mathcal{M}_p$:*

$$x_i^T A Q^z x_i \in \alpha \circ \min\left(\frac{\mathrm{Id}}{\eta\|A\|_{\mathrm{HS}}}, \sqrt{\frac{\mathrm{Id}}{\eta^2\|A\|}}\right) \subset \alpha \circ \sqrt{\frac{\mathrm{Id}}{\eta^2\|A\|_{\mathrm{HS}}}}.$$

Note that outside from the heavy-tailed case, in the specific case where the concentration of X is the same as x_i one can use the concentration of $Q^z x_i = Q^z X e_i$ provided in Lemma 3.2 to employ directly Hanson-Wright theorem and set the result of Lemma 3.13.

Proof. The proof is slightly more elaborate than the proof of Lemma 3.6 because x_i is not independent with Q^z. Still, it follows the same scheme and starts with the decomposition:

$$\begin{aligned}&\left|x_i^T A Q^z x_i - \mathbb{E}[x_i^T A Q^z x_i]\right| \\ &\qquad \leq \left|x_i^T A Q^z x_i - \mathbb{E}[x_i^T A Q^z x_i | X_{-i}]\right| + \left|\mathbb{E}[x_i^T A Q^z x_i | X_{-i}] - \mathbb{E}[x_i^T A Q^z x_i]\right|.\end{aligned}$$

Observe that $x_i \in \alpha(\mathrm{Id}/\eta)$ and when X_{-i} is fixed, $x_i \mapsto Q^z x_i$ is $o(1)$-Lipschitz thanks to the bound $\|Q^z\| \leq \frac{|z|}{\Im(z)}$. One can therefore employ Hanson-Wright inequality given in theorem A.7 and obtain the existence of some constants $C, c > 0$ such that $\forall (n,p) \in \Theta_\gamma$:

$$\mathbb{P}\left(\left|x_i^T A Q^z x_i - \mathbb{E}[x_i^T A Q^z x_i | X_{-i}]\right| \geq t\right) \leq C\alpha \circ \min\left(\frac{c \,\mathrm{Id}}{\eta_p \|A\|_{\mathrm{HS}}}, \sqrt{\frac{c\,\mathrm{Id}}{\eta_p^2 \|A\|}}\right)$$

To control the second part, we introduce for any $i \in [n]$ the following mapping:

$$q_{-i} : M_{-i} \to \mathbb{E}[x_i^T A Q^z x_i | X_{-i} = M_{-i}]$$

Note that q_{-i} is defined on the set of matrices of $\mathcal{M}_{p,n}$ that have only zeros in the i^{th} column. Given such $M_{-i}, M'_{-i} \in \mathcal{M}_{p,n}$ and $x_i \in \mathbb{R}^p$, we denote below:

$$M = (m_1, \ldots, m_{i-1}, x_i, m_{i+1}, \ldots, m_n) \quad \text{and} \quad M' = (m'_1, \ldots, m'_{i-1}, x_i, m'_{i+1}, \ldots, m'_n),$$

such that now, with the notation $\mathcal{Q} : M \mapsto \left(I_p - \frac{MM^T}{zn}\right)^{-1}$ already introduced in the proof of Proposition 3.1, one has the identities $q_{-i}(M_{-i}) = \mathbb{E}[x_i^T A \mathcal{Q}(M) x_i]$ and $q_{-i}(M'_{-i}) = \mathbb{E}[x_i^T A \mathcal{Q}(M') x_i]$. Then:

$$\begin{aligned}
&\|q_{-i}(M_{-i}) - q_{-i}(M'_{-i})\|_{\mathrm{HS}} \\
&\quad = \left|\mathbb{E}[x_i^T A(\mathcal{Q}(M) - \mathcal{Q}(M'))x_i]\right| \\
&\quad = \left|\mathbb{E}\left[x_i^T A\mathcal{Q}(M)\left(\frac{1}{zn}M_{-i}M_{-i}^T - \frac{1}{zn}M'_{-i}{M'_{-i}}^T\right)\mathcal{Q}(M')x_i\right]\right| \\
&\quad \leq \left|\mathbb{E}\left[x_i^T A\mathcal{Q}(M)\left(\frac{1}{zn}M_{-i}(M_{-i}^T - {M'_{-i}}^T) + \frac{1}{zn}(M_{-i} - M'_{-i}){M'_{-i}}^T\right)\mathcal{Q}(M')x_i\right]\right|
\end{aligned}$$

By symmetry, it is sufficient to bound one of the two terms, it is done with Cauchy-Schwarz inequality:

$$\begin{aligned}
&\frac{1}{n}\left|\mathbb{E}\left[x_i^T A\mathcal{Q}(M)M_{-i}(M_{-i} - M'_{-i})^T\mathcal{Q}(M')x_i\right]\right| \\
&\quad \leq \sqrt{\mathbb{E}\left[x_i^T A\frac{\mathcal{Q}(M)M_{-i}}{n}\frac{M_{-i}^*\mathcal{Q}(M)^*}{n}A^T x_i\right]}\sqrt{\mathbb{E}\left[x_i^T\mathcal{Q}(M')^*|M_{-i} - M'_{-i}|^2\mathcal{Q}(M')x_i\right]} \\
&\quad \leq \frac{1}{\sqrt{n}}O\sqrt{\mathbb{E}[x_i^T AA^T x_i]}\sqrt{\mathbb{E}[x_i^T|M_{-i} - M'_{-i}|^2 x_i]} \quad \leq O\left(\frac{\|A\|_{\mathrm{HS}}\|\Sigma_i\|}{\sqrt{n}}\right)\|M_{-i} - M'_{-i}\|_{\mathrm{HS}}
\end{aligned}$$

Therefore, q_{-i} is $O\left(\frac{\|A\|_{\mathrm{HS}}}{\sqrt{n}}\right)$-Lipschitz and, as in the proof of Lemma 3.6, Theorem A.7 provides the existence of some constants $C, c > 0$ such that for all $\theta \in \Theta_\gamma$, for all $z \in \mathbb{H}$:

$$\begin{aligned}
&\mathbb{P}\left(\left|x_i^T A Q^z x_i - \mathbb{E}[x_i^T A Q^z x_i]\right| \geq t\right) \\
&\quad \leq \mathbb{E}\left[\mathbb{P}\left(\left|x_i^T A Q^z x_i - \mathbb{E}[x_i^T A Q^z x_i|X_{-i}]\right| \geq \frac{t}{2}\right)\right] + \mathbb{P}\left(|q_{-i}(X_{-i}) - \mathbb{E}[q_{-i}(X_{-i})]| \geq \frac{t}{2}\right) \\
&\quad \leq \mathbb{E}\left[\alpha \circ \min\left(\frac{ct}{\eta\|A\|_{\mathrm{HS}}}, \sqrt{\frac{ct}{\eta^2\|A\|}}\right)\right] + \alpha\left(\frac{c\sqrt{n}t}{\eta\sqrt{n}\|A\|_{\mathrm{HS}}}\right)
\end{aligned}$$

one can then conclude. □

The bound on the fourth moment of α given by **(A4.b)** is only useful to set next proposition:

Proposition 3.14. $\|\mathbb{E}[Q^z] - \tilde{Q}^{\hat{\Lambda}^z}\|_{\mathrm{HS}} \leq O\left(\frac{\eta^3\sqrt{p}}{n}\right)$

Proof. Let us consider a deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\|_{\mathrm{HS}} \leq 1$. With the notation introduced to set (3.3) we can decompose:

$$\mathrm{Tr}\left(A\left(\mathbb{E}\left[Q^z\right] - \tilde{Q}^{\hat{\Lambda}^z}\right)\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{Tr}(A\varepsilon_i^{\hat{\Lambda}^z}) + \mathrm{Tr}(A\delta_i^{\hat{\Lambda}^z})$$

The bound on $\delta_i^{\check{\Lambda}^z}$ is again obtained with Lemma 3.3:

$$|\,\mathrm{Tr}(A\delta_i^{\hat{\Lambda}^z})| \leq O\left(\frac{\sqrt{p}\|A\|_{\mathrm{HS}}}{n\Im(z)^5}\right)$$

Second, let us estimate:

$$\begin{aligned}
|\,\mathrm{Tr}(A\varepsilon_i^{\hat{\Lambda}^z})| &= \left|\frac{1}{\mathbb{E}[\Lambda_i^z]}\mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda^z]}AQx_i\left(\Lambda_i^z - \mathbb{E}\left[\Lambda_i^z\right]\right)\right]\right| \\
&= \left|\frac{1}{\mathbb{E}[\Lambda_i^z]}\mathbb{E}\left[\left(x_i^T\tilde{Q}^{\mathbb{E}[\Lambda^z]}AQ^zx_i - \mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda^z]}AQ^zx_i\right]\right)(\Lambda_i^z - \mathbb{E}[\Lambda_i^z])\right]\right|
\end{aligned}$$

Lemma A.5 applied to the concentration provided in Lemmas 3.8 and 3.13 allows us to set (the distribution of the composition towards the parallel product is given by Lemma A.4):

$$\begin{aligned}
\left(x_i^T\tilde{Q}^{\mathbb{E}[\Lambda^z]}AQ_{-i}^zx_i - \mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda^z]}AQ^zx_i\right]\right)(\Lambda_i^z - \mathbb{E}[\Lambda_i^z]) &\in \alpha\circ\left(\sqrt{\frac{\mathrm{Id}}{\eta^2\|A\|_{\mathrm{HS}}}}\boxtimes\sqrt{\frac{n\,\mathrm{Id}}{\eta\sqrt{p}}}\right) \\
&\in \alpha\circ\left(\left(\frac{n\,\mathrm{Id}}{\eta^3\sqrt{p}\|A\|_{\mathrm{HS}}}\right)^{\frac{1}{4}}\right)
\end{aligned}$$

A simple integration then provides $\left|\mathrm{Tr}(A\varepsilon_i^{\hat{\Lambda}^z})\right| \leq O\left(\frac{\eta^3\sqrt{p}\|A\|_{\mathrm{HS}}}{n}\right)$.

Combining the bounds on $\hat{\delta}_i$ and the bounds on ε_i for $i \in [n]$ (the bounding constants are the same), one finally obtains both result of the proposition (take the supremum on all $A \in \mathcal{M}_p$ such that such that $\|A\|_{\mathrm{HS}} \leq 1$ to get a bound on $\|Q^z - \tilde{Q}^{\hat{\Lambda}^z}\|_{\mathrm{HS}}$). □

The rest of the study is then exactly the same as in Subsection 3.3, we thus give the next results without proof.

Proposition 3.15. $\|\hat{\Lambda}^z - \tilde{\Lambda}^z\| \leq O\left(\frac{\eta p}{n^2}\right)$

Proposition 3.16 (Convergence in HS norm)**.** $\left\|\mathbb{E}\left[Q^z\right] - \tilde{Q}^{\tilde{\Lambda}^z}\right\|_{\mathrm{HS}} \leq O\left(\frac{\eta\sqrt{p}}{n}\right)$.

One can then combine this result with Proposition 3.1 and Lemma A.3 to finally set the second result of Theorem 1.3.

A Concentration functions and inequalities

Lemma A.1. *Given two family of positive parameters $\sigma, \eta \in \mathbb{R}_+^\Theta$ such that $\eta \leq O(\sigma)$ and a concentration function $\alpha \in \mathcal{M}_{\mathbb{P}_+}^\Theta$ such that $\alpha(1) \geq O(1)$, for all parameters $0 < q \leq r$:*

$$\alpha\left(\left(\frac{\mathrm{Id}}{\eta}\right)^r\right) \subset \alpha\left(\left(\frac{\mathrm{Id}}{\sigma}\right)^q\right).$$

Proof. By definition of our notations, one knows that there exists $c > 0$ such that $\forall\theta \in \Theta$, $\eta_\theta \leq c\sigma_\theta$. Then it is immediate to check that for all $\theta \in \Theta$, $\forall t \geq 0$:

$$\alpha\left(\left(\frac{\mathrm{Id}}{\eta}\right)^r\right) \subset \alpha\left((1/c)^r\left(\frac{\mathrm{Id}}{\sigma}\right)^r\right).$$

To set that $\alpha((\frac{\mathrm{Id}}{\sigma})^r) \subset \alpha((\frac{\mathrm{Id}}{\sigma})^q)$, let us consider $\theta \in \Theta$ and start with the equivalence:

$$\alpha\left(\left(\frac{t}{\sigma_\theta}\right)^r\right) \leq \alpha\left(\left(\frac{t}{\sigma_\theta}\right)^q\right) \qquad \Longleftrightarrow \qquad t \geq \sigma_\theta.$$

Besides, for all $\theta \in \Theta$ and all $t < \sigma_\theta$:

$$\min\left(1, \alpha\left(\left(\frac{t}{\sigma_\theta}\right)^r\right)\right) \leq 1 \leq \frac{1}{\alpha(1)}\alpha\left(\left(\frac{t}{\sigma_\theta}\right)^q\right).$$

Therefore, for any $\theta \in \Theta$, any $t \geq 0$:

$$\min\left(1, \alpha\left(\left(\frac{t}{\sigma_\theta}\right)^r\right)\right) \leq \frac{1}{\min(1, C)}\alpha\left(\left(\frac{t}{\sigma_\theta}\right)^q\right).$$

□

Lemma A.2. *Given a concentration function $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ (independent of θ), a family of parameters $\sigma \in \mathbb{R}_+^\Theta$ and a family of random variables $X \in \mathbb{R}^\Theta$ such that $X \in 0 \pm \alpha \circ (\mathrm{Id}/\sigma)$:*

$$\forall r > 0 : \quad \int t^{r-1}\alpha(t)dt \leq \infty \qquad \Longleftrightarrow \qquad \mathbb{E}[|X|^r] \leq O(\sigma^r).$$

Proof. There exists $C, c > 0$ such that for all $\theta \in \Theta$, $\mathbb{P}(|X_\theta| \geq t) \leq C\alpha\left(\frac{ct}{\sigma_\theta}\right)$ one can then simply bound $\forall \theta \in \Theta$:

$$\mathbb{E}[|X_\theta|^r] = \int_0^\infty \mathbb{P}(|X_\theta|^r \geq t)dt \leq \int_0^\infty C\alpha\left(\frac{ct^{1/r}}{\sigma_\theta}\right)dt \leq Cr\sigma_\theta^r \int_0^\infty t^{r-1}\alpha\left(ct\right)dt.$$

□

Lemma A.3. *Given a family of normed vector space $(E_\theta)_{\theta\in\Theta}$, a family of random vectors $X \in E^\Theta$, two families of deterministic vectors $\tilde{X}_1, \tilde{X}_2 \in E^\Theta$ a family of parameters $\sigma \in \mathbb{R}^\Theta$, and a concentration function (constant with $\theta \in \Theta$) $\alpha \in \mathcal{M}_{\mathbb{P}_+}$, we have the following implication:*

$$X \in \tilde{X}_1 \pm \alpha \circ \left(\frac{\mathrm{Id}}{\sigma}\right) \quad \text{and} \quad \|\tilde{X}_1 - \tilde{X}_2\| \leq O(\sigma) \qquad \Longrightarrow \qquad X \in \tilde{X}_2 \pm \alpha \circ \left(\frac{\mathrm{Id}}{\sigma}\right)$$

Lemma A.4. *Given three operators $f, g, h : \mathbb{R} \to 2^{\mathbb{R}}$:*

$$f \circ g \boxplus f \circ h = f \circ (g \boxplus h) \qquad \text{and} \qquad f \circ g \boxtimes f \circ h = f \circ (g \boxtimes h).$$

Lemma A.5. *Given two families of random variables $X, Y \in \mathbb{R}^\Theta$ and two family of concentration functions $\alpha, \beta \in \mathcal{M}_{\mathbb{P}_+}^\Theta$:*

$$X \in 0 \pm \alpha \quad \text{and} \quad Y \in 0 \pm \beta \qquad \Longrightarrow \qquad XY \in 0 \pm \alpha \boxtimes \beta$$

Lemma A.6. *Given a concentration functions $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ and a family of random variables $X \in \mathbb{R}^\Theta$:*

$$X \in \alpha \qquad \Longrightarrow \qquad X \in \mathbb{E}[X] \pm \alpha.$$

Theorem A.7 (Hanson-Wright Theorem)**.** *Let us consider a family of random vectors $X \in \prod_{\theta\in\Theta} \mathbb{R}^{p_\theta}$, a family of parameters $\sigma \in \mathbb{R}_+^\Theta$ and a concentration functions $\alpha \in \mathcal{M}_{\mathbb{P}_+}$. If we assume that $X \in \alpha \circ (\mathrm{Id}/\sigma)$ and $\mathbb{E}[XX^T] \leq O(\sigma^2)$, then for any sequence of deterministic matrices $A \in \prod_{\theta\in\Theta\in\mathbb{N}} \mathcal{M}_{p_\theta}$:*

$$X^T AY \in \alpha \circ \min\left(\frac{\mathrm{Id}}{\|A\|_{\mathrm{HS}}\sigma^2}, \sqrt{\frac{\mathrm{Id}}{\|A\|\sigma^2}}\right).$$

B Semi metric on $\mathcal{D}_n(\mathbb{H})$

We introduce the semi-metric $d_{\mathbb{H}}$ on $\mathcal{D}_n(\mathbb{H}) = \{D \in \mathcal{D}_n, \forall i \in [n], \Im D_i > 0\}$:

$$d_{\mathbb{H}}(\Delta, \Delta') = \sup_{1 \leq i \leq n} \frac{|\Delta - \Delta'|}{\sqrt{\Im(\Delta)\Im(\Delta')}}$$

The distance $d_{\mathbb{H}}$ is not a metric because it does not satisfy the triangular inequality, see the following counter-example:

$$d_{\mathbb{H}}(4i, i) = \frac{3}{2} > \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = d_{\mathbb{H}}(4i, 2i) + d_{\mathbb{H}}(2i, 1i)$$

Indeed, one has the counter-triangular inequality when certain conditions are met:

Lemma B.1. *Given $x, y, z \in \mathbb{R}$, $x < y < z$ implies that:*

$$d_{\mathbb{H}}^2(a + xi, a + zi) > d_{\mathbb{H}}^2(a + xi, a + yi) + d_{\mathbb{H}}^2(a + yi, a + zi)$$

Proof. Here we construct the function

$$g : y \to \frac{(y - x)^2}{xy} + \frac{(z - y)^2}{yz}$$

and we differentiate it twice to get:

$$g'(y) = \frac{y^2 - x^2}{xy^2} + \frac{y^2 - z^2}{y^2 z} = \frac{1}{x} - \frac{x}{y^2} + \frac{1}{z} - \frac{z}{y^2}$$

$$g''(y) = \frac{3y}{x^3} + \frac{3z}{x^3} > 0$$

This shows that g is strictly convex on $[x, z]$, and the statement follows from the fact that $g(x) = g(z) = d_{\mathbb{H}}^2(a + xi, a + yi)$ and that $g(y) = d_{\mathbb{H}}^2(a + xi, a + yi) + d_{\mathbb{H}}^2(a + yi, a + zi)$ □

Lemma B.2. *Given $\Delta, \Delta' \in \mathcal{D}_n(\mathbb{H})$ and $\Lambda \in \mathcal{D}_n^+$*

$$d_{\mathbb{H}}(\Lambda\Delta, \Lambda\Delta') = d_{\mathbb{H}}(\Delta, \Delta')$$

$$d_{\mathbb{H}}(-\Delta^{-1}, -\Delta'^{-1}) = d_{\mathbb{H}}(\Delta, \Delta')$$

Lemma B.3. *Given four diagonal matrices $\Delta, \Delta', D, D' \in \mathcal{D}_n(\mathbb{H})$:*

$$d_{\mathbb{H}}(\Delta + D, \Delta' + D') \leq \max(d_{\mathbb{H}}(\Delta, \Delta'), d_{\mathbb{H}}(D, D'))$$

Proof. For any $\Delta, \Delta', D, D' \in \mathcal{D}_n(\mathbb{H})$:, there exist $i_0 \in [n]$ such that:

$$\begin{aligned} d_{\mathbb{H}}(\Delta + D, \Delta' + D') &= \frac{|\lambda_{i_0} - \Lambda'_{i_0} + D_{i_0} - D'_{i_0}|}{\sqrt{\Im(\Delta_{i_0} + D_{i_0})\Im(\Delta'_{i_0} + D'_{i_0})}} \\ &\leq \frac{|\lambda_{i_0} - \Lambda'_{i_0}| + |D_{i_0} - D'_{i_0}|^2}{\sqrt{\Im(\Delta_{i_0})\Im(\Delta'_{i_0})} + \sqrt{\Im(D_{i_0})\Im(D'_{i_0})}} \\ &\leq \max\left(\frac{|\lambda_{i_0} - \Lambda'_{i_0}|}{\sqrt{\Im(\Delta_{i_0})\Im(\Delta'_{i_0})}}, \frac{|D_{i_0} - D'_{i_0}|}{\sqrt{\Im(D_{i_0})\Im(D'_{i_0})}}\right) \end{aligned}$$

□

In proving this property we have used the following elementary inequality results.

Lemma B.4. *Given four positive real numbers a, b, α, β:*

$$\sqrt{ab} + \sqrt{\alpha\beta} \leq \sqrt{(a+\alpha)(b+\beta)}$$

$$\frac{a+\alpha}{b+\beta} \leq max(\frac{a}{b}, \frac{\alpha}{\beta})$$

Proof. For the first result, we deduce from the inequality $2\sqrt{ab\alpha\beta} \leq a\beta + b\alpha$:

$$(\sqrt{ab} + \sqrt{\alpha\beta})^2 = ab + \alpha\beta + 2\sqrt{ab\alpha\beta} \leq ab + \alpha\beta + a\beta + b\alpha$$

For the second result, we simply bound:

$$\frac{a+\alpha}{b+\beta} = \frac{a}{b}\frac{b}{b+\beta} + \frac{\alpha}{\beta}\frac{\beta}{b+\beta} \leq max\left(\frac{a}{b}, \frac{\alpha}{\beta}\right).$$

□

Proposition B.5. *Given three parameters $\alpha, \lambda, \theta > 0$ and two mappings $f, g :: \mathcal{D}_n(\mathbb{H}) \to \mathcal{D}_n(\mathbb{H})$, λ-Lipschitz for the semi-metric $d_{\mathbb{H}}$, the mappings*

$$\frac{-1}{f}, \qquad \alpha f, \qquad f \circ g, \qquad \text{and} \qquad f + g$$

are also λ-Lipschitz for the semi-metric $d_{\mathbb{H}}$.

The Banach fixed point theorem states that a contracting function on a complete space admits a unique fixed point. The extension of this result to contracting mappings on $\mathcal{D}_n(\mathbb{H})$, for the semi-metric $d_{\mathbb{H}}$, is not obvious: first, because $d_{\mathbb{H}}$ does not verify the triangular inequality and second because the completeness needs to be proven. The completeness is guaranteed by a boundedness condition that we impose on the matrices. One can rely on the natural topology on $\mathcal{D}_n(\mathbb{H})$ endowed by any norm of the finite dimension vector space $\mathcal{D}_n$. Below the notion of closeness are introduced for this topology.

Theorem B.6. *Let us consider a closed*[10] *subset $\mathcal{D}_b \subset \mathcal{D}_n(\mathbb{H})$ such that there exists $\delta > 0$ satisfying: $\forall \Delta \in \mathcal{D}_b$, $\forall i \in [n] : \Im(\Delta_i) \leq \delta$ and a mapping $f : \mathcal{D}_b \to \mathcal{D}_b$. If f is contracting for the stable semi-metric $d_{\mathbb{H}}$ on $\mathcal{D}_b$, then there exists a unique fixed point $\Delta^* \in \mathcal{D}_b$ satisfying $\Delta^* = f(\Delta^*)$.*

Proof. Let us denote $\lambda \in (0,1)$ the Lipschitz constant such that $\forall \Delta, \Delta' \in \mathcal{D}_n(\mathbb{H}), d_{\mathbb{H}}(f(\Delta), f(\Delta')) \leq \lambda d_{\mathbb{H}}(\Delta, \Delta')$. Then let us show that the sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying $\Delta^{(0)} \in \mathcal{D}_b$ and:

$$\forall k \geq 1 : \qquad \Delta^{(k)} = f(\Delta^{(k-1)})$$

is a Cauchy sequence in $\mathcal{D}_b$. We can bound for any $p \in \mathbb{N}$:

$$\|\Delta^{(p+1)} - \Delta^{(p)}\| \leq \delta d_{\mathbb{H}}(\Delta^{(p+1)}, \Delta^{(p)}) \leq \lambda^p \delta d_{\mathbb{H}}(\Delta^{(1)}, \Delta^{(0)}).$$

Therefore, thanks to the triangular inequality in $(\mathcal{D}_n(\mathbb{H}), \|\cdot\|)$, for any $n \in \mathbb{N}$:

$$\begin{aligned}\|\Delta^{(p+n)} - \Delta^{(p)}\| &\leq \|\Delta^{(p+n)} - \Delta^{(p+n-1)}\| + \cdots + \|\Delta^{(p+1)} - \Delta^{(p)}\| \\ &\leq \frac{\delta d_{\mathbb{H}}(\Delta^{(1)}, \Delta^{(0)})}{1-\lambda}\lambda^p \to 0.\end{aligned}$$

As a Cauchy sequence, $(\Delta^{(p)})_{p\in\mathbb{N}}$ converges to a diagonal matrix $\Delta^* \equiv \lim_{p\to\infty} \Delta^{(p)} \in \mathcal{D}_b$ which is closed thus complete in $\mathcal{D}_n(\mathbb{H})$. By contractivity of f, Δ^* is clearly the unique fixed point of f. □

[10] say, as a subset of $(\mathcal{D}_n(\mathbb{C}), \|\cdot\|)$.

To set the two next result, one needs to introduce a semi-metric $d_{\mathbb{R}_+^*}$ analogous to $d_{\mathbb{H}}$ but defined on $\mathcal{D}_n(\mathbb{R}_+^*)$:

$$\forall \Delta, \Delta' \in \mathcal{D}_n(\mathbb{R}_+^*): \qquad d_{\mathbb{R}_+^*}(\Delta, \Delta') \equiv \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta\Delta'}} \right\|,$$

it satisfies the same stability properties as $d_{\mathbb{H}}$ depicted in Proposition B.5 and allows to set the following elementary result.

Lemma B.7. *Given three positive diagonal matrices $\Gamma^1, \Gamma^2, \Gamma^3 \in \mathcal{D}_n(\mathbb{R}_+)$:*

$$\left\| \frac{\Gamma^3}{\sqrt{\Gamma^1}} \right\| \leq \left\| \frac{\Gamma^3}{\sqrt{\Gamma^2}} (1 + d_{\mathbb{R}_+^*}(\Gamma^1, \Gamma^2)) \right\|.$$

Proof. We simply bound for any $i \in [n]$:

$$\left| \frac{\Gamma_i^3}{\sqrt{\Gamma_i^1}} \right| \leq \left| \frac{\Gamma_i^3}{\sqrt{\Gamma_i^2}} \right| + \left| \frac{\Gamma_i^3 \left(\sqrt{\Gamma_i^2} - \sqrt{\Gamma_i^1} \right)}{\sqrt{\Gamma_i^2 \Gamma_i^1}} \right| \leq \left| \frac{\Gamma_i^3}{\sqrt{\Gamma_i^2}} \right| + \left| \frac{\Gamma_i^3}{\sqrt{\Gamma_i^2}} \right| \left| \frac{\Gamma_i^2 - \Gamma_i^1}{\sqrt{\Gamma_i^1} \left(\sqrt{\Gamma_i^2} + \sqrt{\Gamma_i^1} \right)} \right|.$$

□

Next we give the result to bound the distance between a diagonal matrix and the other one which is obtained as a fixed point. The contractivity can be difficult to set on the whole set, we will thus introduce a sufficient weaker condition. Given a mapping $\Psi : \mathcal{D}(\mathbb{H}) \to \mathcal{D}(\mathbb{H})$, and a diagonal matrix $\Delta \subset \mathcal{D}(\mathbb{H})$, we say that Ψ is "λ-Lipschitz from $\tilde{\Delta}$" iif:

$$\forall \Delta' \in \mathcal{D}(\mathbb{H}) : d_{\mathbb{H}}(\Psi(\tilde{\Delta}), \Psi(\Delta')) \leq \lambda d_{\mathbb{H}}(\tilde{\Delta}, \Delta').$$

Proposition B.8. *Let us consider a family of integers $(n_\theta)_{\theta \in \Theta} \in \mathbb{N}^\Theta$, and for all $\theta \in \Theta$, a mappings $f_\theta : \mathcal{D}_{n_m}(\mathbb{H}) \to \mathcal{D}_{n_m}(\mathbb{H})$, λ-Lipschitz from a fixed point $\tilde{\Gamma}_\theta = f_\theta(\tilde{\Gamma}_\theta)$ and a family of diagonal matrices $(\Gamma_\theta)_{\theta \in \Theta)} \in \prod_{\theta \in \Theta} \in \mathcal{D}_{n_m}(\mathbb{H})^\Theta$. If one assumes that $d_{\mathbb{R}_+^*}(\Im(\Gamma), \Im(f(\Gamma))) \leq o(1)$, then*

$$d_{\mathbb{H}}(\Gamma, \tilde{\Gamma}) \leq O\left(\left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| \right),$$

Proof. Applying Lemma B.7 with $\Gamma_3 = \frac{f(\tilde{\Gamma}) - f(\Gamma)}{\sqrt{\Im(f(\tilde{\Gamma}))}}$, $\Gamma_1 = \Im(\Gamma)$ and $\Gamma_2 = \Im(f(\Gamma))$

$$\begin{aligned} d_{\mathbb{H}}(\Gamma, \tilde{\Gamma}) &\leq \left\| \frac{\tilde{\Gamma} - f(\Gamma)}{\sqrt{\Im(\Gamma)\Im(\tilde{\Gamma})}} \right\| + \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| = \left\| \frac{f(\tilde{\Gamma}) - f(\Gamma)}{\sqrt{\Im(f(\tilde{\Gamma}))}\sqrt{\Im(\Gamma)}} \right\| + \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| \\ &\leq d_{\mathbb{H}}\left(f(\tilde{\Gamma}), f(\Gamma) \right) (1 + d_{\mathbb{H}}(\Im(\Gamma), \Im(f(\Gamma))) + \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| \\ &\leq \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\Gamma)\Im(\tilde{\Gamma})}} \right\| / (1 - \lambda - \lambda d_{\mathbb{H}}(\Im(\Gamma), \Im(f(\Gamma))) \end{aligned} \tag{B.1}$$

Now, we know that there exists a finite subset $T \subset \Theta$ such that $\forall \theta \in \Theta \backslash T$: $d_{\mathbb{H}}(\Im(\Gamma_\theta), \Im(f_\theta(\Gamma_\theta))) \leq \frac{1-\lambda}{2\lambda}$ then one can conclude from (B.1) that for all $\forall \theta \in \Theta \setminus T$:

$$d_{\mathbb{H}}(\Gamma_\theta, \tilde{\Gamma}_\theta) \leq \frac{2}{1 - \lambda} \left\| \frac{f_\theta(\Gamma_\theta) - \Gamma_\theta}{\sqrt{\Im(\tilde{\Gamma_\theta})\Im(\Gamma_\theta)}} \right\|,$$

which can be generalized to any $\theta \in \Theta$ with a constant $C > 0$ replacing $\frac{2}{1-\lambda}$. □

References

[Ada11] Radoslaw Adamczak. On the marchenko-pastur and circular laws for some classes of random matrices with dependent entries. *Electronic Journal of Probability*, 16:1065–1095, 2011.

[AGZ10] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2010.

[BC11] H. Bauschke and P. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer Science+ Business Media, 2011.

[BKV96] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Processes and Related Fields*, 2(4):607–636, 1996.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[Bol97] Vladimir Bolotnikov. On a general moment problem on the half axis. *Linear algebra and its application*, 255:57–112, 1997.

[BS09] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of large dimensional Random Matrices*. Springer, 2009.

[DKL22] Alicja Dembczak-Kołodziejczyk and Anna Lytova. On the empirical spectral distribution for certain models related to sample covariance matrices with different correlations. *Random Matrices: Theory and Applications*, 11(03):2250030, 2022.

[ESY09] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for wigner random matrices. *Communications in Mathematical Physics*, 287(2):641–655, 2009.

[ESYY12] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1–46, 2012.

[HLN07] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Annals of Applied Probability*, 17(3):875–930, 2007.

[KA16] Abla Kammoun and Mohamed-Slim Alouini. No eigenvalues outside the limiting support of generally correlated gaussian matrices. *IEEE Transactions on Information Theory*, 62(7):4312–4326, 2016.

[KM23] Anastasis Kafetzopoulos and Anna Maltsev. Local marchenko–pastur law at the hard edge of the sample covariance ensemble. *Journal of Mathematical Physics*, 64(12), 2023.

[Led05] Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2005.

[Lou24] Cosme Louart. Operation with concentration inequalities and conjugate of parallel sum. *arXiv preprint arXiv:2402.08206*, 2024.

[MP67] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR-Sbornik*, 1(4):457–483, 1967.

[PŜ11] L. Pastur and M. Ŝerbina. *Eigenvalue distribution of large random matrices*. American Mathematical Society, 2011.

[SB95] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995.

[Sil86] J. W. Silverstein. Eigenvalues and eigenvectors of large dimensional sample covariance matrices. *Random Matrices and their Applications*, pages 153–159, 1986.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[Ver18] Roman Vershynin. *High-Diensional Probability*. Cambridge University Press, 2018.

[Wac78] Kenneth W. Wachter. The Strong Limits of Random Matrix Spectra for Sample Matrices of Independent Elements. *The Annals of Probability*, 6(1):1 – 18, 1978.

[WCDS12] Sebastian Wagner, Romain Couillet, Mérouane Debbah, and Dirk TM Slock. Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback. *IEEE transactions on information theory*, 58(7):4509–4537, 2012.

[Yin86] Y. Yin. Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, 20:50–68, 1986.