Eigen behaviour of Random matrices with Heavy tailed independent columns

Authors*

Abstract

Abstract

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

1 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 $n, p \in \mathbb{N}$, $[n] \equiv \{1, \ldots, n\}$, the entries of a vector $x \in \mathbb{C}^p$ are generally denoted $x_1, \ldots, x_p$, the columns of a complex matrix $A \in \mathcal{M}_{p,n}$ are denoted $a_1, \ldots, a_n$. Let us denote $\mathcal{M}_n$, the set of square matrices $\mathcal{M}_{n,n}$, S_n, the set of (possibly non-real) symmetric matrices, H_n, the set of Hermitian matrices, $\mathcal{O}_n$, the set of orthogonal matrices, $\mathcal{U}_n$, the set of unitary matrices and D_n, the set of diagonal matrices. We introduce the natural order relation on H_n, given $A, B \in H_n$:

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

Given $x \in \mathbb{C}^n$, $D = \mathrm{Diag}(x) \in D_n$ is the diagonal matrix having the elements $x_1, \ldots, x_n$ on the diagonal then one usually denote $\forall i \in [n]$, $D_i \equiv x_i$. The conjugate transpose of a matrix M is denoted $M^* = \bar{M}^T$. 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 $\|\cdot\|_{\mathrm{HS}}$ ($\forall M \in \mathcal{M}_{p,n}$: $\|M\|_{\mathrm{HS}} = \mathrm{Tr}(MM^*)$) and the spectral norm is denoted $\|\cdot\|$ ($\|M\| = \sup_{\|x\|=1} \|Mx\|$). Given two normed vector spaces $(E, \|\cdot\|)$ and $(E', \|\cdot\|')$, and a linear mapping $u : E \to F$, the operator norm of u is denoted $\|u\| \equiv \sup_{\|x\| \leq 1} \|u(x)\|'$.

We will set in this paper *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 ∞ or the imaginary part of the complex parameter $z \in \mathbb{C}$ appearing in the definition of the Stieltjes Transform tends to zero. 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 I could be thought to be $\mathbb{N} \times \mathbb{N} \times \mathbb{C}$

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

or even something more elaborate like $\{(p,n,z) \in \mathbb{N} \times \mathbb{N} \times \mathbb{C}, p \leq n, \Im(z) > 0\}$ (see Assumption **??**).

Given an index set Θ and two family of parameters $(a_\theta)_{\theta\in\Theta} \in \mathbb{R}_+$ and $(b_\theta)_{\theta\in\Theta} \in \mathbb{R}_+^\Theta$, we denote: "$a_\theta \leq O(b_\theta),\ \theta \in \Theta$" or more simply "$a \leq O(b)$" 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 $\forall \theta \in \Theta$, $a_\theta \geq Cb_\theta$). 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:

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

Following a previous work done in [**?**, **?**] we will express concentration inequalities with operators which are set valued mappings. An operator $\alpha : \mathbb{R} \mapsto 2^{\mathbb{R}}$ is said to be a positive probabilitic operator and we denote $\alpha \in \mathcal{M}_{\mathbb{P}_+}$ iif it is maximally decreasing[1] $\{1\} \subset \mathrm{Ran}(\alpha)$ and $\mathrm{Dom}(\alpha) \subset \mathbb{R}_+$. Let us consider a family of random variables $(X_\theta)_{\theta\in\Theta} \in \mathbb{R}^\Theta$ and a family of positive probabilitic 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:

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

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 $|\tilde{X}| \leq O(m)$.

We rely on real-valued functional to extend those notations to random vectors. Given a family of normed vector spaces $(E_\theta, \|\cdot\|)_{\theta\in\Theta}$, a family of random vectors $(X_\theta)_\theta \in \prod_{\theta\in\Theta} E_\theta$, the notation "$X \in \alpha$" means that there exists some constants $C, c > 0$ such that[2] $\forall \theta \in \Theta$ for all 1-Lipschitz mappings $f : E_\theta \to \mathbb{R}$:

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

where for all $\theta \in \Theta$, X'_θ is an independent copy of X_θ. 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 exists some constants $C, c > 0$ such that[3] $\forall \theta \in \Theta$ for all linear form $u : E_\theta \to \mathbb{R}$ such that $\|u\| \leq 1$:

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

We denote Id, the identity operator $t \mapsto \{t\}$, then $\sqrt{\mathrm{Id}} : t \mapsto \{\sqrt{t}\}$, it satisfies $\mathrm{Dom}(\sqrt{\mathrm{Id}}) \subset \mathbb{R}_+$.

[1] Following the monotone operator theory (see for instance [**?**]), 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 $\forall x, y \in \mathbb{R}^2$:

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

[2] With the random variable notations, that means that:

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

[3] With previous notations, that means that $X \in \alpha$ and:

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

2 Setting

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

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

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

the index set that will direct our quasi-asymptotic results.

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 introduce the family of means, of centered and non-centered empirical covariance matrices for all $i \in \mathbb{N}$:

$$\mu_i \equiv \mathbb{E}[x_i] \qquad \Sigma_i \equiv \mathbb{E}[x_i(x_i)^T]. \qquad C_i \equiv \Sigma_i - \mu_i(\mu_i)^T.$$

Considering a family of positive probability operators $\alpha \in \mathcal{M}_{\mathbb{P}_+}^{\Theta_\gamma}$, we will assume the following properties are satisfied:

- for all $(n,p) \in \Theta_\gamma$: $x_i^{(n,p)}, \ldots, x_n^{(n,p)}$ are independent,
- $X \in \alpha$,
- $\sigma_\alpha \equiv \int t\alpha(t)dt \leq \infty$ and $\alpha(\sigma\alpha) \leq O(1)$,
- $\|\mu_i\| \leq O(1)$, $i \in [n]$,
- $\Sigma_i \geq O(1)$, $i \in [n]$.

Remark 2.1. • Possible alpha...
- bound on $\|\mu\|$...
- lower bound on Σ...

3 Concentration of the resolvent.

To study the spectral distribution of $\frac{1}{n}XX^T$:

$$\nu \equiv \frac{1}{n} \sum_{\lambda \in \text{Sp}(\frac{1}{n}XX^T)} \delta_\lambda,$$

the classical approach is to look at the Stieltjes transform defined for any $z \notin \text{Sp}(\frac{1}{n}XX^T)$ as:

$$m(z) \equiv \int \frac{1}{z-\lambda} d\nu(\lambda).$$

To deduce properties on ν, it is sufficient to study $m(z)$ for $z \in \mathbb{C}$ such that $\Im(z) \in (0,1]$, we will thus restrict our study to this range to simplify the bounds in the convergence results. Introducing the family of resolvents $Q \equiv (Q_{n,p}^z)_{(n,p)\in\Theta_\gamma, \Im(z)\in(0,1]} \in \prod_{(p,n)\in\Theta_\gamma, \Im(z)\in(0,1]} \mathcal{M}_p$ defined for any $(p,n) \in \Theta_\gamma$, $\forall z \in \mathbb{H}$ as:

$$Q^z = \left(I_p - \frac{1}{n}XX^T\right)^{-1}$$

one will rely on the identity:

$$m(z) = \frac{1}{p} \text{Tr}(Q^z).$$

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

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

To set the concentration of Q and $\check{Q}$, let us first bound them, it is a trivial and classical result of Random matrix theory that we provide here without proof.

Lemma 3.1. $|Q|, |\check{Q}| \leq O\left(\frac{1}{\Im z}\right)$.

Let us note that from the identity $Q\frac{1}{n}XX^T = Q - I_p$, one can also bound:

$$\left\|\frac{1}{n}QX\right\| \leq \frac{1}{\sqrt{n}}\sqrt{\left\|\frac{1}{n}QXX^TQ\right\|} \leq \frac{1}{\sqrt{n}}\sqrt{\|Q^2-Q\|} \leq O\left(\frac{1}{\sqrt{n}\Im(z)}\right) \tag{3.1}$$

Proposition 3.2. $Q^z, \check{Q}^z \in \alpha(\frac{\cdot}{\Im(z)^2\sqrt{n}})$.

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

$$\Phi(M) = \left(zI_p - \frac{MM^T}{n}\right)^{-1} \qquad \text{and} \qquad \check{\Phi}(M) = \left(zI_n - \frac{M^TM}{n}\right)^{-1},$$

it is sufficient to show that Φ and $\check{\Phi}$ 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:

$$\left\|d\Phi|_M \cdot H\right\|_{\mathrm{HS}} = \left\|\Phi(M)\frac{1}{n}(MH^T + HM^T)\Phi(M)\right\|_{\mathrm{HS}} \leq O\left(\frac{1}{\Im(z)^2\sqrt{n}}\right)\|H\|_{\mathrm{HS}},$$

thanks to lemma 3.1 and (3.1). 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 3.3. $QX = X^T\check{Q} \in \alpha\left(\Im(z)^2\right)$

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

$$\Psi(M) = \left(zI_p - \frac{MM^T}{n}\right)^{-1} M.$$

to show the concentration of $QX = \Psi(X)$. For all $H, M \in \mathcal{M}_{n,p}$ (and with the notation $\Phi(M) = \left(zI_p - \frac{MM^T}{n}\right)^{-1}$ given in the proof of Proposition 3.2), let us bound:

$$\|d\Psi|_M \cdot H\| \leq \left\|\Psi(M)\frac{1}{n}(MH^T + HM^T)\Psi(M)M\right\| + \|\Psi(M)H\| \leq O\left(\frac{\|H\|_{\mathrm{HS}}}{\Im(z)^2}\right).$$

□

4 A first deterministic equivalent

In this subsection, we provide a first estimator of $\mathbb{E}[Q]$.

An efficient approach, developed in particular in [**?**, **?**] is to look for a deterministic equivalent of Q^z depending on a deterministic diagonal matrix $\Delta \in \mathbb{R}^n$ and having the form:

$$\tilde{Q}^\Delta = \left(zI_p - \Sigma^\Delta\right)^{-1} \qquad \text{where} \qquad \Sigma^\Delta \equiv \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\Delta_i} = \frac{1}{n}\mathbb{E}[X\Delta^{-1}X^T]. \tag{4.1}$$

One can then express the difference with the expectation $\mathbb{E}[Q^z]$ followingly:

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

To pursue the estimation of the expectation, one needs to control the dependence between Q and x_i. For that purpose, one uses classically the Schur identities:

$$Q = Q_{-i} + \frac{1}{n}\frac{Q_{-i}x_ix_i^TQ_{-i}}{1 - \frac{1}{n}x_i^TQ_{-i}x_i} \qquad \text{and} \qquad Qx_i = \frac{Q_{-i}^zx_i}{1 - \frac{1}{n}x_i^TQ_{-i}^zx_i}, \tag{4.2}$$

for $Q_{-i} = (I_n - \frac{1}{zn}X_{-i}X_{-i}^T)^{-1}$ (recall that $X_{-i} = (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \in \mathcal{M}_{p,n}$). The Schur identities can be seen as simple consequences to the so called "resolvent identity" that can be generalized to any, possibly non commuting, square matrices $A, B \in \mathcal{M}_p$ with the identity:

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \qquad \text{or} \qquad A^{-1} + B^{-1} = A^{-1}(A + B)B^{-1} \tag{4.3}$$

(it suffices to note that $A(A^{-1} + A^{-1}(B - A)B^{-1})B = I_p$).

Introducing the notation:

$$\Lambda \equiv \mathrm{Diag}_{1\leq i\leq n}\left(1 - \frac{1}{n}x_i^TQ_{-i}x_i\right) \in \prod_{(n,p)\in\Theta_\gamma, \Im(z)\in(0,1]}\mathcal{D}_n,$$

one has the identity $Qx_i = \frac{1}{\Lambda_i}Q_{-i}x_i$. It is then possible to express:

$$\mathbb{E}[Q] - \tilde{Q}^{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Q_{-i}\left(\frac{x_ix_i^T}{\Lambda_i} - \frac{\Sigma_i}{\Delta_i}\right)\tilde{Q}^{\Delta}\right] + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta_i}\mathbb{E}\left[(Q_{-i} - Q)\Sigma_i\tilde{Q}^{\Delta}\right] \tag{4.4}$$

where we recall that $Q - Q_{-i} = \frac{1}{n}Qx_ix_i^TQ_{-i}$.

From this decomposition, one is enticed into choosing, in a first step $\Delta = \mathbb{E}[\Lambda] \in \mathcal{D}_n(\mathbb{C})$ so that ε_1 would be small.

Proposition 4.1. $\|Q - \tilde{Q}^{\mathbb{E}[\Lambda]}\|_{\mathrm{HS}} \leq O\left(\frac{1}{\Im(z)^4\sqrt{n}}\right)$

The proof of this proposition relies on the next four preliminary lemmas.

Lemma 4.2. $\Lambda \geq O\left(\frac{1}{\Im(z)}\right)$.

Proof. It is a simple consequence of Lemma 3.1 and the identity:

$$\frac{1}{\Lambda^z} = \mathrm{Diag}_{i\in[n]}\left(\frac{1}{1 - \frac{1}{n}x_i^TQ_{-i}x_i}\right) = I_n + \frac{1}{n}\mathrm{Diag}(X^TQX) = \mathrm{Diag}(\check{Q}). \tag{4.5}$$

□

Lemma 4.3. $\|\mathbb{E}[Q_{-i}] - \mathbb{E}[Q]\| \leq O\left(\frac{1}{n\Im(z)^2}\right)$.

Note from **[?]** that this lemma implies in particular from Proposition 3.2 that $Q_{-i} \in \mathbb{E}[Q] \pm \alpha \circ \left(\Im(z)^2\sqrt{n}\ \mathrm{Id}\right)$.

Proof. Let us bound for any deterministic vector $u \in \mathbb{C}$:

$$\begin{aligned}|u^*(\mathbb{E}[Q_{-i}] - \mathbb{E}[Q])u| &= \frac{1}{n}\left|\mathbb{E}\left[\frac{u^*Q_{-i}x_ix_i^*Q_{-i}u}{\Lambda_i}\right]\right| \\ &\leq \mathbb{E}[u^*Q_{-i}\Sigma_iQ_{-i}u]O\left(\frac{1}{n\Im(z)}\right) \leq O\left(\frac{1}{n\Im(z)^2}\right)\end{aligned}$$

thanks to Lemmas 3.1 and 4.2 □

Lemma 4.4. *Given* $\Delta \in \mathcal{D}_n$, $\|\tilde{Q}^{\Delta}\| \leq O\left(\frac{1}{\min_{i\in[n]}(\Im(z\Delta_i))}\right)$.

Proof. TODO □

Lemma 4.5. $\Im(z\Lambda) \leq \Im(z)$.

One can deduce directly from Lemmas 4.4 and 4.5 that $\|\tilde{Q}^{\mathbb{E}[\Lambda]}\| \leq O\left(\frac{1}{\Im(z)}\right)$.

Proof. Let us compute:

$$\Im(z\Lambda) = \Im(z) - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i(Q - \bar{Q})\right) = \Im(z) - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i Q(\bar{z}I_p - zI_p)\bar{Q}\right) \geq \Im(z),$$

since $\operatorname{Tr}\left(\Sigma_i Q\bar{Q}\right) \geq 0$. □

proof of Proposition 4.1. To prove our bound let us consider $A \in \mathcal{M}_p$, and start with the first component of (4.4):

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\operatorname{Tr}\left(AQ_{-i}\left(\frac{x_i x_i^T}{\Lambda_i} - \frac{\Sigma_i}{\mathbb{E}[\Lambda_i]}\right)\tilde{Q}^{\mathbb{E}[\Lambda]}\right)\right]\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQ_{-i}x_i\left(\frac{1}{\Lambda_i} - \frac{1}{\mathbb{E}\left[\Lambda_i\right]}\right)\right]\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQx_i\left(\frac{\mathbb{E}\left[\Lambda_i\right] - \Lambda_i}{\mathbb{E}\left[\Lambda_i\right]}\right)\right]\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\mathbb{E}[\Lambda_i]}\mathbb{E}\left[\left(x_i^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQx_i - \mathbb{E}\left[x_i^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQx_i\right]\right)\right]\\
&\quad= \frac{1}{n}\mathbb{E}\left[\operatorname{Tr}\left(\frac{1}{\mathbb{E}[\Lambda]}X^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQX\right) - \mathbb{E}\left[\operatorname{Tr}\left(\frac{1}{\mathbb{E}[\Lambda]}X^T\tilde{Q}^{\mathbb{E}[\Lambda]}AQX\right)\right]\right] \leq O\left(\frac{1}{\sqrt{n}\Im(z)^4}\right),
\end{aligned}$$

thanks to the matricial, heavy-tailed form of Hanson-Wright result ($\sigma_\alpha \int t\alpha \leq \infty$ and $\alpha(\sigma_\alpha) \geq O(1)$) applied to:

- the concentration $\tilde{Q}^{\mathbb{E}[\Lambda]}X \in \alpha$ (by hypothesis on X and thanks to the bound provided by Lemmas 4.4 and 4.5),
- $QX \in \alpha(\frac{1}{\Im(z)^2})$ (see Lemma 3.3),
- $\|\frac{1}{\mathbb{E}[\Lambda]}_{HS} \leq O\left(\frac{\sqrt{n}}{\Im(z)}\right)$ (see Lemma 4.2).

The second component of (4.4) is simply bounded thanks to Lemma 4.3 that implies:

$$\mathbb{E}\left[\operatorname{Tr}\left(A(Q_{-i} - Q)\Sigma_i\tilde{Q}^{\mathbb{E}[\Lambda]}\right)\right] \leq O\left(\|\mathbb{E}[Q_{-i}] - \mathbb{E}[Q]\|\frac{\sqrt{p}}{\Im(z)}\right) \leq O\left(\frac{1}{\sqrt{n}\Im(z)^2}\right)$$

Combining those two bounds with (4.4), one obtains the result of the proposition. □

In next section, we will see that it is more convenient to work with the deterministic diagonal matrix

$$\hat{\Lambda} \equiv \operatorname{Diag}\left(1 - \frac{1}{n}\operatorname{Tr}(\Sigma_i\mathbb{E}[Q])\right)_{1\leq i\leq n} \in \mathcal{D}_n(\mathbb{C}),$$

which is close to $\mathbb{E}[\Lambda]$ thanks to Lemma 4.3.

Proposition 4.6. $\|\tilde{Q}^{\mathbb{E}[\Lambda]} - \tilde{Q}^{\hat{\Lambda}}\| \leq O\left(\frac{1}{\Im(z)^6 n}\right)$.

Proof. Let us bound for any deterministic vector $u \in \mathbb{C}^p$:

$$\left|u^*(\tilde{Q}^{\mathbb{E}[\Lambda]} - \tilde{Q}^{\hat{\Lambda}})u\right| = \frac{1}{n}\sum_{i=1}^{n}\left|u^*\tilde{Q}^{\mathbb{E}[\Lambda]}\Sigma_i\tilde{Q}^{\hat{\Lambda}}u\right|\left|\frac{\check{\Lambda}_i - \mathbb{E}[\Lambda_i]}{\check{\Lambda}_i\mathbb{E}[\Lambda_i]}\right|.$$

One can then directly conclude thanks to the bounds given in Lemmas 3.1 and 4.4 and the bound:

$$|\check{\Lambda}_i - \mathbb{E}[\Lambda_i]| = \frac{1}{n}\left|\mathrm{Tr}(\Sigma_i(Q - Q_{-i}))\right| \leq O\left(\frac{1}{\Im(z)^2 n}\right).$$

□

5 A second deterministic equivalent

A The semi-metric and Lipschitz mapping

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

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

The distance d_s is not a metric because it does not satisfy the triangular inequality, see the following counter-example:

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

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

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

$$d_s^2(a + xi, a + zi) > d_s^2(a + xi, a + yi) + d_s^2(a + yi, a + zi)$$

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 $g(x) = g(z) = d_s^2(a + xi, a + yi)$ and that $g(y) = d_s^2(a + xi, a + yi) + d_s^2(a + yi, a + zi)$ □

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

$$d_s(\Lambda\Delta, \Lambda\Delta') = d_s(\Delta, \Delta')$$

$$d_s(-\Delta^{-1}, -\Delta'^{-1}) = d_s(\Delta, \Delta')$$

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

$$d_s(\Delta + D, \Delta' + D') \leq \max(d_s(\Delta, \Delta'), d_s(D, D'))$$

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

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

□

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

Lemma A.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:

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

□

Definition A.5. *Given $\lambda > 0$, we denote $\mathcal{C}_s^\lambda(\mathcal{D}_n(\mathbb{H}))$, the class of functions $f : \mathcal{D}_n(\mathbb{H}) \to \mathcal{D}_n(\mathbb{H})$, λ-Lipschitz for the semi-metric d_s; i.e. satisfying for all $D, D' \in \mathcal{D}_n(\mathbb{H})$:*

$$d_s(f(D), f(D')) \leq \lambda d_s(D, D').$$

When $\lambda < 1$, we say that f is contracting for the semi-metric d_s.

Proposition A.6. *Given three parameters $\alpha, \lambda, \theta > 0$ and two mappings $f \in \mathcal{C}_s^\lambda$ and $g \in \mathcal{C}_s^\theta$,*

$$\frac{-1}{f} \in \mathcal{C}_s^\lambda, \quad \alpha f \in \mathcal{C}_s^\lambda, \quad f \circ g \in \mathcal{C}_s^{\lambda\theta}, \quad f + g \in \mathcal{C}_s^{\max(\lambda,\theta)}$$

B Fixed point theorem for contracting mapping

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_s, is not obvious: first, because d_s 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.

Theorem B.1. *Given a subset $\mathcal{D}_b$ of $\mathcal{D}_n(\mathbb{H})$ where each diagonal entry has an imaginary part bounded from above and below and a mapping $f : \mathcal{D}_b \to \mathcal{D}_b$, if it is furthermore contracting for the stable semi-metric d_s on $\mathcal{D}_b$, then there exists a unique fixed point $\Delta^* \in \mathcal{D}_b$ satisfying $\Delta^* = f(\Delta^*)$.*

Proof. Noting $\lambda \in (0,1)$ the Lipschitz constant such that $\forall \Delta, \Delta' \in \mathcal{D}_n(\mathbb{H}), d_s(f(\Delta), f(\Delta')) \leq \lambda d_s(\Delta, \Delta')$, we show that the sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying:

$$\Delta^{(0)} = I_n, \quad \forall k \geq 1, \Delta^{(k)} = f(\Delta^{(k-1)})$$

is a Cauchy sequence in $\bar{\mathcal{D}}_n(\mathbb{H})$, where $\bar{\mathcal{D}}_n(\mathbb{H}) \equiv \mathcal{D}_n(\mathbb{H} \bigcup \mathbb{R})$. $\forall p \in \mathbb{N}$, $\Delta^{(p)} \in \mathcal{D}_b$, i.e. there exists $\delta > 0$, such that $|\Im \Delta^{(p)}| \leq \delta$. We can then bound for any $p \in \mathbb{N}$:

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

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

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

This allows us to conclude that $(\Delta^{(p)})_{p\in\mathbb{N}}$ is a Cauchy sequence, and therefore it converges to a diagonal matrix $\Delta^* \equiv \lim_{p\to\infty} \Delta^{(p)} \in \bar{\mathcal{D}}_n(\mathbb{H})$ which is a closed thus complete set. But since $\Delta^{(p)}$ has diagonal entries which are bounded from below, we know that $\Delta^* \in \mathcal{D}_b$. By contractivity of f, it is clearly unique. □

C Stability of the stable semi-metric towards perturbations

We have first of all the following elementary inequality result.

Lemma C.1. *Given theree diagonal matrices* $\Gamma^1, \Gamma^2, \Gamma^3 \in \mathcal{D}_n(\mathbb{H})$*:*

$$\left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^1)}} \right\| \leq \left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^2)}} (1 + d_s(\Im(\Gamma^1), \Im(\Gamma^2))) \right\|$$

Proof. We simply bound:

$$\begin{aligned}\left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^1)}} \right\| &\leq \left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^2)}} \right\| + \left\| \frac{\Gamma^3 \left(\sqrt{\Im(\Gamma^2)} - \sqrt{\Im(\Gamma^1)} \right)}{\sqrt{\Im(\Gamma^2)\Im(\Gamma^1)}} \right\| \\ &\leq \left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^2)}} \right\| + \left\| \frac{\Gamma^3}{\sqrt{\Im(\Gamma^2)}} \right\| \left\| \frac{\Im(\Gamma^2) - \Im(\Gamma^1)}{\sqrt{\Im(\Gamma^1)} \left(\sqrt{\Im(\Gamma^2)} + \sqrt{\Im(\Gamma^1)} \right)} \right\|\end{aligned}$$

□

Next we give the result to bound the distance between a diagonal matrix and the other one which is obtained as a fixed point.

Proposition C.2. *Given a diagonal matrix* $\Gamma \in \mathcal{D}_n(\mathbb{H})$*, a mapping* $f : \mathcal{D}_n(\mathbb{H}) \to \mathcal{D}_n(\mathbb{H})$ λ *contractive for the semi-metric* d_s *with the Lipschitz coefficient* $\lambda < 1$ *and admitting the fixed point* $\tilde{\Gamma} = f(\tilde{\Gamma})$*, we have the bound:*

$$d_s(\Gamma, \tilde{\Gamma}) \leq \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\Gamma)\Im(\tilde{\Gamma})}} \right\| / (1 - \lambda - \lambda d_s(\Im(\Gamma)), \Im(f(\Gamma)))$$

Proof. Thanks to the above lemma, we can bound:

$$
\begin{aligned}
d_s(\Gamma, \tilde{\Gamma}) &\leq \left\| \frac{\tilde{\Gamma} - f(\Gamma)}{\sqrt{\Im(\Gamma)\Im(\tilde{\Gamma})}} \right\| + \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| \\
&\leq d_s(\tilde{\Gamma}, \Gamma)(1 + d_s(\Im(\Gamma)), \Im(f(\Gamma)) + \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\tilde{\Gamma})\Im(\Gamma)}} \right\| \\
&\leq \left\| \frac{f(\Gamma) - \Gamma}{\sqrt{\Im(\Gamma)\Im(\tilde{\Gamma})}} \right\| /(1 - \lambda - \lambda d_s(\Im(\Gamma)), \Im(f(\Gamma)))
\end{aligned}
$$

□

Proposition C.3. *Let us consider a family of mappings*$(f^m)_{m\in\mathbb{N}}$ *of* $\mathcal{D}_{n_m}(\mathbb{H})$*, each* f^m *being* λ*-Lipschitz for the semi-metric* d_s *with* $\lambda < 1$ *and admitting the fixed point* $\tilde{\Gamma}^m = f^m(\tilde{\Gamma}^m)$ *and a family of diagonal matrices* Γ^m*. If one assume that* $d_s(\Im(\Gamma^m), \Im(f^m(\Gamma^m))) \leq o_{m\to\infty}(1)$*, then*

$$
d_s(\Gamma^m, \tilde{\Gamma}^m) \leq O_{m\to\infty}\left(\left\| \frac{f^m(\Gamma^m)) - \Gamma^m}{\sqrt{\Im(\tilde{\Gamma^m})\Im(\Gamma^m)}} \right\| \right)
$$

Proof. For m sufficiently big, we have $d_s(\Im(\Gamma^m), \Im(f^m(\Gamma^m))) \leq o(1) \leq \frac{1-\lambda}{2\lambda}$, so we have:

$$
\begin{aligned}
d_s(\Gamma^m, \tilde{\Gamma}^m) &\leq \left\| \frac{f(\Gamma^m) - \Gamma^m}{\sqrt{\Im(\Gamma^m)\Im(\tilde{\Gamma}^m)}} \right\| /(1 - \lambda - \lambda d_s(\Im(\Gamma^m)), \Im(f(\Gamma^m))) \\
&\leq \left(\left\| \frac{f^m(\Gamma^m)) - \Gamma^m}{\sqrt{\Im(\tilde{\Gamma^m})\Im(\Gamma^m)}} \right\| \right)
\end{aligned}
$$

□