PhD manuscript:
Concentration of the measure and random matrices to study data processessing algorithms

Cosme Louart

February, 28 - 2022

Part I

Concentration of the resolvent of sample covariance matrix

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, we denote $\mathrm{Sp}(\frac{1}{n}XX^T)$ the spectrum of $\frac{1}{n}XX^T$. The spectral distribution of $\frac{1}{n}XX^T$, denoted $\mu \equiv \frac{1}{p}\sum_{\lambda\in\mathrm{Sp}(\frac{1}{n}XX^T)}\delta_\lambda$, is classically studied through its Stieltjes transform expressed as:

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

The relevance of the Stieltjes transform has been extensively justified in some seminal works **??** by the Cauchy integral that provides for any analytical mapping f defined on a neighborhood of a subset $B \subset \mathrm{Sp}(\frac{1}{n}XX^T)$ the identity:

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

where $\gamma : [0,1] \to \mathbb{C} \setminus \mathrm{Sp}(\frac{1}{n}XX^T)$ is a closed path on which f is defined and whose interior I_γ satisfies $I_\gamma \cap \mathrm{Sp}(\frac{1}{n}XX^T) = B \cap \mathrm{Sp}(\frac{1}{n}XX^T)$. But we can go further and approximate linear functionals of the eigenvectors thanks to the resolvent. If we denote E_B the random eigenspace associated to the eigenvalues of $\frac{1}{n}XX^T$ belonging to B, and Π_B the orthogonal projector on E_B, then for any deterministic matrix $A \in \mathcal{M}_p$:

$$\mathrm{Tr}(\Pi_B A) = \frac{1}{2i\pi}\int_\gamma \mathrm{Tr}(AR(z))dz \qquad \text{with } R(z) \equiv \left(\frac{1}{n}XX^T - zI_p\right)^{-1}. \tag{1}$$

The matrix $R(z)$ is commonly called the resolvent of $\frac{1}{n}XX^T$. It satisfies in particular that, for all $z \in \mathbb{C}\setminus\mathrm{Sp}(\frac{1}{n}XX^T)$, $g(z) = \frac{1}{p}\mathrm{Tr}(R(z))$. It thus naturally becomes the central element of the study of the spectral distribution. One of the first tasks in random matrix theory is to devise a so called "deterministic equivalent" for $R(z)$ (**?**), that we will denote here $\tilde{R}(z)$. Specifically, we look for a deterministic matrix computable from the first statistics of our problem and close to $\mathbb{E}[R(z)]$. In particular, a main result of random matrix theory sets that this deterministic equivalent only expresses with means and covariances of the x_i's. Two questions then arise:

1. How close to $\mathbb{E}[R(z)]$ is $R(z)$?
2. What does this notion of closeness really mean ?

The first questions relate to concentrations properties on $R(z)$ that arise from concentration properties on X. The study of random matrices originally studied with i.i.d. entries (**?**, **?**), mere Gaussian hypotheses (**?**), or with weaker hypotheses concerning the first moments of the entries (supposed to be independent or at least independent up to an affine transformation). Some more recent works showed the concentration of the spectral distribution of Wishart or Wigner matrices with bounding assumption on the entries of the random matrix under study –or at least a linear transformation of it– allows to employ Talagrand results in **??** that also treat some log-concave hypotheses allowing to relax some independence assumptions as it is done in **?**. One can also find very light hypotheses on the quadratic functionals of the columns (**?**), improved in **?** or on the norms of the columns and the rows (**?**). In the present work, we do not consider the case of what is called convexly concentrated random vectors in **???** as it is done in **?** because it requires a different approach (see Section 2.2).

The Lipschitz concentration hypothesis we assume on X in a sense "propagates" to the resolvent that then satisfies, for all deterministic matrices A such that $\|A\|_F \equiv \sqrt{\mathrm{Tr}(AA^T)} \leq 1$ and for all z not too close to the spectrum of $\frac{1}{n}XX^T$:

$$\mathbb{P}\left(\left|\mathrm{Tr}\left(A\left(R(z) - \mathbb{E}[R(z)]\right)\right)\right| \geq t\right) \leq Ce^{-cnt^2} + Ce^{-cn}, \tag{2}$$

for some numerical constants C, c independent of n, p. We see from (2) the important benefit gained with our concentration hypothesis on X: it provides simple *quasi-asymptotic* results on the convergence of the resolvent with speed rates, while most of the results on random matrices are classically expressed in the limiting regime $n, p \to \infty$.

The condition $\|A\|_F \leq 1$ answers our second question: a specificity of our approach is to control the convergence of the resolvent with the Frobenius norm at a speed of order $O(1/\sqrt{n})$. The concentration inequality (2) means that all linear forms of $R(z)$, which are 1-Lipschitz[1] for the Frobenius norm, have a standard deviation of order $O(1/\sqrt{n})$; this is crucial to be able to estimate quantities expressed in (1). Generally, the Stieltjes transform $g(z) = -\frac{1}{p}\operatorname{Tr}(R(z))$ is classically the only studied linear forms of the resolvent (it is $1/\sqrt{p}$-Lipschitz so its standard deviation is of order $O(1/\sqrt{pn})$ which is a classical result although not exactly under a concentration of measure assumption) or projections on deterministic vectors $u^T R(z) u$, for which only the concentration in spectral norm with a speed of order $O(1/\sqrt{n})$ is needed.

Those remarks gain a real importance when we are able to estimate the expectation of $R(z)$ with a deterministic equivalent (that we can compute). We look for a closeness relation in Frobenius norm:

$$\left\|\mathbb{E}[R(z)] - \tilde{R}(z)\right\|_F \leq O\left(\frac{1}{\sqrt{n}}\right).$$

We may then replace in (2) the term "$\mathbb{E}[R(z)]$" by $\tilde{R}(z)$ which we are able to compute from the expectations and covariances of the columns $x_1, \ldots, x_n$.

Note that we do not assume that the columns are identically distributed: in particular, the means and covariances can be all different (although they have to satisfy some boundedness properties expressed in Assumptions 4 and 5 at the beginning of Chapter 2). This remark may be related to the studies made of matrices X with a variance profile: but this is here even more general because the laws of the columns are not solely defined from their means and covariances (although the spectral distribution of $\frac{1}{n}XX^T$ just depends on these quantities).

The extension of Marcenko Pastur result to non-identically distributed columns was well known; one can cite for instance **???** treating a very similar problem but with different assumptions of concentration (just some moments of the entries need to be bounded) the result is then given in the form of a limit (not a concentration inequality), **?** showing that no eigenvalues lie outside of the support with Gaussian hypotheses and **?** imposing some weak isotropic conditions on the different covariances. We will follow the proof scheme of **?** where the authors introduce two consecutive deterministic equivalents, the first one depending on the expectations of $\Lambda^z = z - \frac{1}{n}x_i^T Q_{-i}^z x_i$, and the second one being expressed through fixed point equation that approximate those quantities. We can indeed estimate the expectations of the quantities $\Lambda_1^z, \ldots, \Lambda_n^z$ with a diagonal matrix[2] $\tilde{\Lambda}^z = \operatorname{Diag}(\tilde{\Lambda}_i^z)_{i\in[n]} \in \mathcal{D}_n(\mathbb{C})$ obtained after successive iteration of the following equation (quite different from the one presented in **?**):

$$\forall i \in [n]: \ \tilde{\Lambda}_i^z = z - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i \tilde{Q}^{\tilde{\Lambda}^z}\right) \qquad \text{with} \quad \tilde{Q}^{\tilde{\Lambda}^z} \equiv \left(I_p - \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\tilde{\Lambda}_i^z}\right)^{-1}, \tag{3}$$

in which $\Sigma_i = \mathbb{E}[x_i x_i^T]$.

The difficulties are (i) to prove the existence and uniqueness of $\tilde{\Lambda}^z$ and (ii) to ensure some stability properties[3] on this equation eventually allowing us to assert that $\|\mathbb{E}[R(z)] - \tilde{R}(z)\|_F \leq O(1/\sqrt{n})$, where $\tilde{R}(z) \equiv \frac{1}{z}\tilde{Q}^{\tilde{\Lambda}^z}$. The existence and uniqueness of similar equations has already proven thanks to complex analysis justification (normal family theorem in **?** and Vitali's theorem in **??**), those approaches allows to extend convergence properties for some z to the whole convex half plane $\mathbb{H}$, however we are not only looking for an asymptotic result but for a quasi asymptotic result, meaning that we want precise convergence bounds for n and p big but not infinite.

The two aforementioned difficulties disappear with the introduction of a convenient semi-metric[4] d_s on which the fixed point equation satisfied by $\tilde{\Lambda}^z$ is contractive, leading (after still some work since a semi-metric is not as easy to treat as if d_s were a true metric) to existence, uniqueness and stability properties. This

[1] or λ-Lipschitz with $\lambda \leq O(1)$

[2] The interest to resort to a diagonal matrix of $\mathcal{M}_n$ rather than to a vector of $\mathbb{R}^n$ will be clearer later – mainly to employ Proposition **??** in a natural formalism.

[3] Conceptually, it means that if we have a diagonal matrix $L \in \mathcal{M}_n$ satisfying $\forall i \in [n]: \ L_i \equiv z - \frac{1}{n}\operatorname{Tr}\left(\Sigma_i \tilde{Q}^L\right)$ then L is "close" to $\tilde{\Lambda}^z$.

[4] A semi metric is defined as a metric that does not satisfy the triangular inequality.

semi-metric, quite similar to the one already introduced in **?** to study robust estimators, is defined for any $D, D' \in \mathcal{D}_n(\mathbb{H})$ as:

$$d_s(D, D') = \left\| \frac{D - D'}{\sqrt{\Im(D)\Im(D')}} \right\|.$$

This semi-metric appears as a central object in random matrix theory, one can indeed point out the fact that any Stieltjes transform is 1-Lipschitz under this semi-metric (see Proposition 1.18). It relates to Poincaré metric, the hyperbolic metric writes indeed $d_{\mathbb{H}}(z, z') = \cosh(d_s(z, z')^2 - 1)$. Apart from the book **?** that provided a groundwork for the introduction of such a metric, this approach already gained some visibility in **??** or in a context closer to ours for the study of Wishart matrices and in a squared form in **?**. We prefer an expression proportional to $\|D - D'\|$ because it is more adapted to comparison with classical distance on $\mathcal{D}_n(\mathbb{H})$, as it is done in Proposition 1.23 that subsequently provides bounds to the convergence speed. Let us outline that once the appropriate semi-metric is identified, contractivity properties are not sufficient to prove the existence and uniqueness to (3): one also needs to introduce the correct space over which the mapping is contractive ($\mathcal{D}_{I^z} \equiv \{D \in \mathcal{D}_n(\mathbb{H}), \frac{D}{z} \in \mathcal{D}_n(\mathbb{H})\}$).

Chapter 1

Stable semi metric

1.1 Definition and first properties

The stable semi-metric which we define here is a convenient object which allows us to set Banach-like fixed point theorems. It has a crucial importance to prove the existence and uniqueness of $\hat{C}$ but also to obtain some random matrix identities on $\hat{C}$, such as the estimation of its limiting spectral distribution.

Definition 1. *We call the* stable semi-metric *on* $\mathcal{D}_n^+ = \{D \in \mathcal{D}_n, \forall i \in [n],\ D_i > 0\}$ *the function:*

$$\forall \Delta, \Delta' \in \mathcal{D}_n^+ : \quad d_s(\Delta, \Delta') \equiv \left\| \frac{\Delta - \Delta'}{\sqrt{\Delta \Delta'}} \right\|. \tag{1.1}$$

In particular, this semi-metric can be defined on $\mathbb{R}^+$, identifying $\mathbb{R}^+$ with $\mathcal{D}_1^+$.

The function d_s is not a metric because it does not satisfy the triangular inequality, one can see for instance that:

$$d_s(4,1) = \frac{3}{2} > \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = d_s(4,2) + d_s(2,1) \tag{1.2}$$

More precisely, for any $x, z \in \mathbb{R}_+$ such that $x < z$, if one differentes twice the mapping $g : y \to \frac{(xy-x)^2}{xy} + \frac{(z-y)^2}{xy}$, one obtains:

$$g'(y) = \frac{1}{y} - \frac{y}{x^2} + \frac{1}{z} - \frac{z}{x^3} \qquad \text{and} \qquad g''(y) = \frac{3y}{x^3} + \frac{3z}{x^3} > 0,$$

which proves that g is strictly convex on $[x, z]$ and therefore it admits a minimum y_0 on $]x, z[$ (since $g(x) = g(z)$). In particular, one can bound:

$$d_s(x,z) > \sqrt{d_s(x,y_0)^2 + d_s(y_0,z)^2}$$

One can however sometimes palliate this weakness when needed thanks to the following inequality proved in Section 1.5.

Proposition 1.1 (Pseudo triangular inequality)**.** *Given* $x, z, y \in \mathbb{R}^+$*:*

$$|x - y| \leq |y - z| \qquad \Longrightarrow \qquad d_s(x,y) \leq d_s(x,z).$$

In addition, for any $p \in \mathbb{N}^*$ *and* $y_1, \ldots, y_{p-1} \in \mathbb{R}^+$*, we have the inequalities*[1]*:*

$$d_s(x, y_1) + \cdots + d_s(y_{p-1}, z) \geq d_s\left(x^{\frac{1}{p}}, z^{\frac{1}{p}}\right) \geq d_s(x,z)^{1/p}$$

and the left inequality turns into an equality in the case $y_i = x^{\frac{p-i}{p}} z^{\frac{i}{p}}$ *for* $i \in \{1, \ldots, p-1\}$*.*

[1] The mapping $x \mapsto x^p$ is not Lipschitz for the semi-metric d_s (unlike $x \mapsto x^{\frac{1}{p}}$).

Proof. For a given integer $p \geq 1$, let us differentiate the mapping:

$$\begin{array}{rccc} f_p : & \mathbb{R}_+^{p-1} & \longrightarrow & \mathbb{R} \\ & (y_1, \dots, y_{p-1}) & \longmapsto & \frac{y_1 - x}{\sqrt{y_1 x}} + \cdots \frac{z - y_{p-1}}{\sqrt{z y_{p-1}}} \end{array}$$

one can compute for any $y_1, \dots, y_{p-1} \in \mathbb{R}^+$ and $i \in [p-1]$:

$$\frac{\partial f_p(y_1, \dots, y_{p-1})}{\partial y_i} = \frac{1}{2} \frac{1}{\sqrt{y_i y_{i-1}}} \left(1 + \frac{y_{i-1}}{y_i}\right) - \frac{1}{2} \frac{1}{\sqrt{y_{i+1} y_i}} \left(1 + \frac{y_{i+1}}{y_i}\right) \tag{1.3}$$

(where y_0 and y_p designate respectively x and z) In particular, when $p = 1$, for any $y \geq x > 0$:

$$\frac{\partial}{\partial y} \left(\frac{y - x}{\sqrt{yx}}\right) = \frac{1}{2} \frac{1}{\sqrt{xy}} \left(1 + \frac{x}{y}\right) \geq 0$$

which proves the first result of the proposition. Now if we assume that $y \leq x \leq z$:

$$d_s(x, y) + d_s(y, z) \geq d_s(x, z),$$

and the same inequality holds if one assumes that $x \leq z \leq y$. Returning to the setting of the proposition, we can therefore place ourselves in the open space:

$$\mathcal{U}_{x,z}^p = \{(y_1, \dots, y_p) \in \mathbb{R}_+^{p-1}, x < y_1 < \cdots < y_{p-1} < z\}.$$

If one fixes $x, z \in \mathbb{R}^+$, then $f_p(x, y_1, \dots, y_{p-1}, z) = d_s(x, y_1) + \cdots + d_s(y_{p-1}, z)$ is minimum for $y_1, \dots, y_{p-1}$ satisfying:

$$\frac{1}{\sqrt{y_i y_{i-1}}} \left(1 - \frac{y_{i-1}}{y_i}\right) = \frac{1}{\sqrt{y_{i+1} y_i}} \left(1 - \frac{y_i}{y_{i+1}}\right)$$

which is equivalent to $y_i = \sqrt{y_{i-1} y_{i+1}}$. Noting $\tilde{x} = \log(x), \tilde{y}_1 = \log(y_1), \dots, \tilde{y}_n = \log(y_n), \tilde{z} = \log(z)$, we see that this identity writes $\tilde{y}_i = \frac{1}{2}(\tilde{y}_{i-1} \tilde{y}_{i+1})$, which implies $\tilde{y}_i = \tilde{x} + \frac{i}{p}(\tilde{z} - \tilde{x})$, or in other words:

$$y_i = x^{\frac{p-i}{p}} z^{\frac{i}{p}}.$$

In that case:

$$d_s(y_i, y_{i+1}) = \left| \frac{x^{\frac{p-i}{2p}} z^{\frac{i}{2p}}}{x^{\frac{p-i-1}{2p}} z^{\frac{i+1}{2p}}} - \frac{x^{\frac{p-i-1}{2p}} z^{\frac{i+1}{2p}}}{x^{\frac{p-i}{2p}} z^{\frac{i}{2p}}} \right| = \left| \frac{x^{\frac{1}{2p}}}{z^{\frac{1}{2p}}} - \frac{z^{\frac{1}{2p}}}{x^{\frac{1}{2p}}} \right| = d_s\left(x^{\frac{1}{p}}, z^{\frac{1}{p}}\right),$$

and the same holds for $d_s(x, y_1)$ and $d_s(y_{p-1}, z)$.

The last inequality is just a consequence of the concavity of $t \to t^{1/p}$:

$$d_s\left(x^{\frac{1}{p}}, z^{\frac{1}{p}}\right) = \frac{z^{\frac{1}{p}} - x^{\frac{1}{p}}}{(xz)^{\frac{1}{2p}}} = \frac{\frac{1}{p} \int_0^{z-x} (t + x)^{\frac{1-p}{p}} dt}{(xz)^{\frac{1}{2p}}} \leq \frac{\frac{1}{p} \int_0^{z-x} t^{\frac{1-p}{p}} dt}{(xz)^{\frac{1}{2p}}} = \left(\frac{z - x}{(xz)^{\frac{1}{2}}}\right)^{\frac{1}{p}} = d_s(x, z)^{\frac{1}{p}}.$$

∎

The semi-metric d_s is called stable due to its many interesting stability properties.

Property 1.2. *Given $\Delta, \Delta' \in \mathcal{D}_n^+$ and $\Lambda \in \mathcal{D}_n^+$:*

$$d_s(\Lambda\Delta, \Lambda\Delta') = d_s(\Delta, \Delta') \qquad \text{and} \qquad d_s\left(\Delta^{-1}, \Delta'^{-1}\right) = d_s(\Delta, \Delta').$$

Property 1.3. *Given four diagonal matrices $\Delta, \Delta', D, D' \in \mathcal{D}_n^+$:*

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

To prove this property one needs two elementary results.

Lemma 1.4. *Given four positive numbers $a, b, \alpha, \beta \in \mathbb{R}^+$:*

$$\sqrt{ab} + \sqrt{\alpha\beta} \leq \sqrt{(a+\alpha)(b+\beta)} \qquad \text{and} \qquad \frac{a+\alpha}{b+\beta} \leq \max\left(\frac{a}{b}, \frac{\alpha}{\beta}\right)$$

Proof. For the first result, we deduce from the inequality $4ab\alpha\beta \leq (a\alpha + b\beta)^2$:

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

For the second result, we simply bound:

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

□

Proof of Property 1.3. For any $\Delta, \Delta', D, D' \in \mathcal{D}_n^+$, there exists $i_0 \in [n]$ such that:

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

thanks to Lemma 1.4. □

1.2 Stable class

Definition 2 (Stable class)**.** *The set of* 1*-Lipschitz functions for the stable semi-metric is called the stable class. We denote it:*

$$\mathcal{S}\left(\mathcal{D}_n^+\right) \equiv \left\{f : \mathcal{D}_n^+ \to \mathcal{D}_n^+ \mid \forall \Delta, \Delta' \in \mathcal{D}_n^+,\ \Delta \neq \Delta' :\ d_s(f(\Delta), f(\Delta')) \leq d_s(\Delta, \Delta')\right\}.$$

The elements of $\mathcal{S}\left(\mathcal{D}_n^+\right)$ are called the stable mappings.

Let us then provide the properties which justify why we call $\mathcal{S}\left(\mathcal{D}_n^+\right)$ a *stable* class: this class indeed satisfies far more stability properties than the usual Lipschitz mappings (for a given norm). Those stability properties are direct consequences to Properties 1.2 and 1.3.

Property 1.5. *Given $\Lambda, \Gamma \in \mathcal{D}_n^+$ and $f, g \in \mathcal{S}(\mathcal{D}_n^+)$:*

$$(\Delta \mapsto \Lambda f(\Gamma\Delta)) \in \mathcal{S}(\mathcal{D}_n^+), \qquad \frac{1}{f} \in \mathcal{S}(\mathcal{D}_n^+), \qquad f \circ g \in \mathcal{S}(\mathcal{D}_n^+), \qquad f + g \in \mathcal{S}(\mathcal{D}_n^+).$$

1.3 The sub-monotonic class

The stable class has a very simple interpretation when $n = 1$. Given a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ we introduce two characteristic functions $f_/, f_\times : \mathbb{R}^+ \to \mathbb{R}^+$:

$$f_/\ :\ x \mapsto \frac{f(x)}{x} \qquad \text{and} \qquad f_\times\ :\ x \mapsto xf(x).$$

Property 1.6. *A function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a stable mapping if and only if $f_/$ is non-increasing and $f_\times$ is non-decreasing.*

Proof. Let us consider $x, y \in \mathbb{R}^+$, such that, say, $x \leq y$. We suppose in a first time that $f_/$ is non-increasing and that $f_\times$ is non-decreasing. We know that $\frac{f(x)}{x} \geq \frac{f(y)}{y}$, and subsequently:

$$f(y) - f(x) \leq \frac{f(y)}{y}(y-x) \qquad \text{and} \qquad f(y) - f(x) \leq \frac{f(x)}{x}(y-x) \tag{1.4}$$

The same way, since $f(x)x \leq f(y)y$ we also have the inequalities:

$$f(x) - f(y) \leq \frac{f(y)}{x}(y-x) \qquad \text{and} \qquad f(x) - f(y) \leq \frac{f(x)}{y}(y-x) \tag{1.5}$$

Now if $f(y) \geq f(x)$, we can take the root of the product of the two inequalities of (1.4) and if $f(y) \leq f(x)$, we take the root of the product of the two inequalities of (1.5), to obtain, in both cases:

$$|f(x) - f(y)| \leq \sqrt{\frac{f(y)f(x)}{xy}}\,|x-y|$$

That means that $f \in \mathcal{S}(\mathbb{R}^+)$.

Conversely, if we now suppose that $f \in \mathcal{S}(\mathbb{R}^+)$, we then use the bound:

$$|f(y) - f(x)| \leq \sqrt{\frac{f(y)f(x)}{xy}}(y-x).$$

First, if $f(x) \leq f(y)$, then $f(x)x \leq f(y)y$ and we can bound:

$$f(y) - f(x) \leq \max\left(\frac{f(x)}{x}, \frac{f(y)}{y}\right)(y-x) \leq \max\left(\left(\frac{y}{x} - 1\right)f(x), \left(1 - \frac{x}{y}\right)f(y)\right)$$

which directly implies $\frac{f(y)}{y} \leq \frac{f(x)}{x}$. Second, if $f(x) \geq f(y)$, $\frac{f(x)}{x} \geq \frac{f(y)}{y}$ and we can then bound in the same way:

$$f(x)xy - f(y)xy \leq \max\left(xf(x)(y-x), (y-x)\,yf(y)\right)$$

which implies $xf(x) \leq yf(y)$. In both cases ($f(x) \leq f(y)$ and $f(y) \leq f(x)$), and we see that $f_/(x) \geq f_/(y)$ and $f_\times(x) \leq f_\times(y)$, proving the result.

□

This property allows us to understand directly that the stability of a function is a local behavior. We then conclude staightforwardly that the supremum or the infimum of stable mappings is also stable.

Corollary 1.7. *Given a family of stable mappings $(f_\theta)_{\theta\in\Theta} \in \mathcal{S}(\mathbb{R}^+)^\Theta$, for a given set Θ, the mappings $\sup_{\theta\in\Theta} f_\theta$ and $\inf_{\theta\in\Theta} f_\theta$ are both stable.*

Given $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$, we can introduce by analogy to the case of mappings on $\mathbb{R}^+$, the mappings $f_/, f_\times : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ defined with:

$$f_/ \ : \ \Delta \mapsto \mathrm{Tr}\left(\frac{f(\Delta)}{\Delta}\right) \qquad \text{and} \qquad f_\times \ : \ \Delta \mapsto \mathrm{Tr}\left(\Delta f(\Delta)\right)$$

Inspiring from Property 1.6, one can then define:

Definition 3 (Sub-monotonuous class). *A mapping $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ is said to be sub-monotonuous if and only if $f_\times$ is non decreasing and $f_/$ is non-increasing, we note this class of mappings $\mathcal{S}_m(\mathcal{D}_n^+)$.*

Remark 1.8. *We know from Property 1.6 that $\mathcal{S}(\mathbb{R}_+) = \mathcal{S}_m(\mathbb{R}_+)$ but for $n > 1$, none of the classes $\mathcal{S}_m(\mathcal{D}_n^+)$ and $\mathcal{S}(\mathcal{D}_n^+)$ contains strictly the other one. On the first hand, introducing:*

$$\begin{aligned} f : \mathcal{D}_2^+ &\longrightarrow \quad \mathcal{D}_2^+ \\ \Delta &\longmapsto \mathrm{Diag}\left(\frac{1}{\Delta_2}, \frac{1}{\Delta_1}\right), \end{aligned}$$

we see that $f \in \mathcal{S}(\mathcal{D}_n^+)$ *but* $f \notin \mathcal{S}_m(\mathcal{D}_n^+)$ *(for* $\Delta = \mathrm{Diag}(1,2)$ *and* $\Delta' = \mathrm{Diag}(2,2)$, $\Delta \leq \Delta'$ *but* $\mathrm{Tr}(f(\Delta)\Delta) = \frac{5}{2} > 2 = \mathrm{Tr}(f(\Delta')\Delta')$*). On the other hand, the mapping:*

$$\begin{aligned} g : \mathcal{D}_2^+ &\longrightarrow \quad \mathcal{D}_2^+ \\ \Delta &\longmapsto \quad \mathrm{Diag}\left(\frac{\Delta_1\Delta_2}{1+\Delta_2}, 1\right) \end{aligned}$$

is in $\mathcal{S}_m(\mathcal{D}_n^+)$ *because:*

$$\begin{cases} \dfrac{\partial g_\cdot}{\partial \Delta_1} = \dfrac{2\Delta_1\Delta_2}{1+\Delta_2} \geq 0 \\ \dfrac{\partial g_\cdot}{\partial \Delta_2} = \dfrac{\Delta_1^2}{(1+\Delta_2)^2} + 1 \geq 0 \end{cases} \quad \textit{and} \quad \begin{cases} \dfrac{\partial g_/}{\partial \Delta_1} = 0 \leq 0 \\ \dfrac{\partial g_/}{\partial \Delta_2} = \dfrac{1}{(1+\Delta_2)^2} - \dfrac{1}{\Delta_2^2} \leq 0. \end{cases}$$

However, we can see that g *is not stable if we introduce the diagonal matrices* $\Delta = \mathrm{Diag}(1,2)$ *and* $\Delta' = \mathrm{Diag}(2,3)$ *since we have then:*

$$d_s(g(\Delta), g(\Delta')) = d_s\left(\frac{2}{3}, \frac{3}{2}\right) = \frac{5}{6} > \frac{1}{\sqrt{2}} = \max\left(\frac{|3-2|}{\sqrt{6}}, \frac{|2-1|}{\sqrt{2}}\right) = d_s(\Delta, \Delta')$$

1.4 Fixed Point theorem for stable and sub-monotonic mappings

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 $\mathcal{D}_n^+$, 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 of the semi-metric space $(\mathcal{D}_n^+, d_s)$ is left in Section 1.5 since we will not need it. Let us start with a first bound.

Lemma 1.9. *Given a mapping* $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$*, contracting for the semi metric* d_s*, any sequence of diagonal matrices* $(\Delta_n)_{n\in\mathbb{N}}$ *satisfying* $\Delta^{(p+1)} = f(\Delta^{(p)})$ *is bounded from below and above and satisfies for all* $p \in \mathbb{N}$ *and* $i \in [n]$*:*

$$\exp\left(-\frac{\lambda d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)}{2(1-\lambda)}\right)\Delta_i^{(1)} \leq \Delta_i^{(p)} \leq \exp\left(\frac{\lambda d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)}{2(1-\lambda)}\right)\Delta_i^{(1)}$$

Proof. Noting $\lambda > 0$, the Lipschitz parameter of f for the semi-metric d_s, let us first show that, for all $i \in [n]$:

$$\forall p \in \mathbb{N}, \sqrt{\frac{\Delta^{(p+1)_i}}{\Delta_i^{(p)}}} \leq 1 + \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right). \tag{1.6}$$

When $\Delta_i^{(p+1)} \leq \Delta_i^{(p)}$, it is obvious and when $\Delta^{(p+1)} \geq \Delta^{(p)}$ the contractivity of f allows us to set $d_s(\Delta^{(p+1)}, \Delta^{(p)}) \leq \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)$, which implies:

$$\sqrt{\frac{\Delta_i^{(p+1)}}{\Delta_i^{(p)}}} \leq \sqrt{\frac{\Delta_i^{(p)}}{\Delta_i^{(p+1)}}} + \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right) \leq 1 + \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right).$$

Multiplying (1.6) for all $p \in \{1, \ldots, P\}$, we obtain:

$$\begin{aligned} \sqrt{\frac{\Delta_i^{(P)}}{\Delta_i^{(1)}}} \leq 1 \leq \prod_{p=1}^{P}\left(1 + \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)\right) &= \exp\left(\sum_{p=1}^{P} \log\left(1 + \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)\right)\right) \\ &\leq \exp\left(\sum_{p=1}^{P} \lambda^p d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)\right) \leq \exp\left(\frac{\lambda d_s\left(\Delta^{(1)}, \Delta^{(0)}\right)}{1-\lambda}\right). \end{aligned}$$

With a similar approach, we can eventually show the result of the lemma. $\square$

Let us now present two fixed point results that will justify the definition of the robust scatter matrix C but also of the deterministic diagonal matrix U introduced in Section **??**.

Theorem 1.10. *Any mapping $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$, contracting for the stable semi-metric d_s, admits a unique fixed point $\Delta^* \in \mathcal{D}_n(\mathbb{R}^+ \cup \{0\})$ satisfying $\Delta^* = f(\Delta^*)$.*

Proof. We cannot repeat exactly the proof of the Banach fixed point theorem since d_s does not satisfy the triangular inequality. Noting $\lambda \in (0,1)$ the parameter such that $\forall \Delta, \Delta' \in \mathcal{D}_n^+$, $d_s(f(\Delta), f(\Delta')) \leq \lambda d_s(\Delta, \Delta')$, we show that the sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying:

$$\Delta^{(0)} = I_n \qquad \text{and} \qquad \forall k \geq 1 : \; \Delta^{(k)} = f(\Delta^{(k-1)})$$

is a Cauchy sequence in $(\overline{\mathcal{D}}_n^+, \|\cdot\|)$, where $\overline{\mathcal{D}}_n^+ \equiv \mathcal{D}_n(\mathbb{R}^+ \cup \{0\})$.

We know from Lemma 1.9 that there exists $\delta > 0$ such that $\forall p \in \mathbb{N}$, $\|\Delta^{(p)}\| \leq \delta$. One can then bound for any $p \in \mathbb{N}$:

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

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

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

That allows us to conclude that $(\Delta^{(p)})_{p\in\mathbb{N}}$ is a Cauchy sequence, and therefore that it converges to a diagonal matrix $\Delta^* \equiv \lim_{p\to\infty} \Delta^{(p)} \in \overline{\mathcal{D}}_n^+$ which is complete (closed in a complete set). But since $\Delta^{(p)}$ is bounded from below and above thanks to Lemma 1.9, we know that $\Delta^* \in \mathcal{D}_n^+$. By contractivity of f, it is clearly unique. □

It is possible to relax a bit the contracting hypotheses on f if one supposes that f is monotonic. We express rigorously this result in next theorem, but it will not be employed in our applications in Chapter **??** about the robust estimation of scatter matrix since we preferred to assume u bounded to obtain the contracting properties of the fixed point satisfied by $\hat{\Delta}$.

Theorem 1.11. *Let us consider a weakly monotonic mapping $f : \mathcal{D}_n^+ \to \mathcal{D}_n^+$ bounded from below and above. If we suppose that f is stable and verifies:*

$$\forall \Delta, \Delta' \in \mathcal{D}_n^+ : \; d_s(f(\Delta), f(\Delta')) < d_s(\Delta, \Delta') \tag{1.7}$$

then there exists a unique fixed point $D \in \mathcal{D}_n^+$ satisfying $\Delta^* = f(\Delta^*)$.

1.5 Supplementary inferences on the stable semi-metric and topological properties

Remark 1.12. *Not all the stable mappings admit a continuous continuation on $\overline{\mathcal{D}}_n^+$. To construct a counter example, for any $n \in \mathbb{N}$, let us note*

- $e_n : x \mapsto \frac{3}{2} - 2^{n-1}x$ *(it satisfies $e_n(1/2^n) = 1$ and $e_n(3/2^{n+1}) = \frac{3}{4}$),*
- $d_n : x \mapsto 2^{n-1}x$ *(it satisfies $d_n(1/2^{n-1}) = 1$ and $e_n(3/2^{n+1}) = \frac{3}{4}$),*
- $v_n : \mathbb{R}^+ \to \mathbb{R}^+$ *satisfying for all $x \in \mathbb{R}^+$, $v_n(x) = \max(e_n, d_n)$ (in particular, $v_n(2^n) = v_n(2^{n-1}) = 1$),*
- $f : \mathbb{R}^+ \to \mathbb{R}^+$ *satisfying for all $x \in \mathbb{R}^+$, $f(x) = \inf_{n\in\mathbb{N}} v_n(x)$.*

We know from Property 1.6 that for all $n \in \mathbb{N}$, d_n is stable and that e_n is stable on $[0, 3/2^{n+1}]$ (where $x \mapsto xe_n(x)$ is non decreasing) which eventually allows us to set that f is stable, thanks to Corollary 1.7.

However f does not admit continuous continuation on 0 since:

$$\lim_{n\to\infty} f\left(\frac{1}{2^n}\right) = 1 \neq \frac{3}{4} = \lim_{n\to\infty} f\left(\frac{3}{2^n}\right).$$

Lemma 1.13. *Any Cauchy sequence of $(\mathcal{D}_n^+, d_s)$ is bounded from below and above (in $\mathcal{D}_n^+$).*

Proof. Considering a Cauchy sequence of diagonal matrices $\Delta^{(k)} \in \mathcal{D}_n^+$, we know that there exists $K \in \mathbb{N}$ such that:

$$\forall p, q \geq K,\ \forall i \in \{1, \dots, n\}: \quad |\Delta_i^{(p)} - \Delta_i^{(q)}| \leq \sqrt{\Delta_i^{(p)} \Delta_i^{(q)}}.$$

For $k \in \mathbb{N}$, let us introduce the indexes $i_M^k, i_m^k \in \mathbb{N}$, satisfying:

$$\Delta_{i_M^k}^{(k)} = \max\left(\Delta_i^{(k)}, 1 \leq i \leq n\right) \qquad \text{and} \qquad \Delta_{i_M^k}^{(k)} = \min\left(\Delta_i^{(k)}, 1 \leq i \leq n\right).$$

If we suppose that there exists a subsequence $(\Delta_{i_M^k}^{(\phi(k))})_{k\geq 0}$ such that $\Delta_{i_M^k}^{(\phi(k))} \underset{k\to\infty}{\longrightarrow} \infty$, then

$$\sqrt{\Delta_{i_M^{\phi(k)}}^{(\phi(k))}} \leq \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} + \frac{\Delta_{i_M^{\phi(k)}}^{(N)}}{\sqrt{\Delta_{i_M^{\phi(k)}}^{(\phi(k))}}} \underset{k\to\infty}{\longrightarrow} \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} < \infty$$

which is absurd. Therefore $(\Delta_{i_M^k}^{(k)})_{k\geq 0}$ and thus also $(\Delta^{(k)})_{k\geq 0}$ are bounded from above. For the lower bound, we consider the same way a subsequence $(\Delta_{i_m^k}^{(\psi(k))})_{k\geq 0}$ such that $\Delta_{i_m^k}^{(\psi(k))} \underset{k\to\infty}{\longrightarrow} 0$. We have:

$$\Delta_{i_M^{\phi(k)}}^{(\phi(k))} \geq \Delta_{i_M^{\phi(k)}}^{(N)} - \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)} \Delta_{i_M^{\phi(k)}}^{(\phi(k))}} \underset{k\to\infty}{\longrightarrow} \sqrt{\Delta_{i_M^{\phi(k)}}^{(N)}} > 0$$

which is once again absurd. □

Property 1.14. *The semi-metric space $(\mathcal{D}_n^+, d_s)$ is complete.*

Proof. Given a Cauchy sequence of diagonal matrices $\Delta^{(k)} \in \mathcal{D}_n^+$, we know from the preceding lemma that there exists $\delta_M, \delta_m \in \mathbb{R}^+$ such that $\forall k \geq 0 : \delta_m I_n \leq \Delta^{(k)} \leq \delta_M I_n$. Thanks to the Cauchy hypothesis:

$$\forall \varepsilon > 0, \exists K \geq 0 \mid \forall p, q \geq K: \ \forall i \in \{1, \dots, n\}: \left|\Delta_i^{(p)} - \Delta_i^{(q)}\right| \leq \varepsilon \delta_M$$

and, as a consequence, $(\Delta^{(k)})_{k\geq 0}$ is a Cauchy sequence in the complete space $(\mathcal{D}_n^{0,+}, \|\cdot\|)$: it converges to a matrix $\Delta^{(\infty)} \in \mathcal{D}_n^{0,+}$. Moreover, $\Delta^{(\infty)} \geq \delta_k I_n$ (as any $\Delta^{(k)}$) for all $k \in \mathbb{N}$, so that $\Delta^{(\infty)} \in \mathcal{D}_n^+$ and we are left to showing that $\Delta^{(k)} \underset{k\to\infty}{\longrightarrow} \Delta^{(\infty)}$ for the semi-metric d_s. It suffices to write:

$$d_s(D^{(k)}, D^{(\infty)}) = \left\|\frac{D^{(k)} - D^{(\infty)}}{\sqrt{D^{(k)} D^{(\infty)}}}\right\| \leq \delta_m \left\|D^{(k)} - D^{(\infty)}\right\| \underset{k\to\infty}{\longrightarrow} 0.$$

□

Proof of Theorem 1.11. We first suppose that f is non-decreasing. As before, let us consider $\delta_M, \delta_m \in \mathbb{R}^+$ such that $\forall \Delta \in \mathcal{D}_n^+\ \delta_m I_n \leq f(\Delta) \leq \delta_M I_n$. The sequence $(\Delta^{(k)})_{k\geq 0}$ satisfying $\Delta^{(0)} = \Delta_m I_n$, and for all $k \geq 1$, $\Delta^{(k)} = f(\Delta^{(k-1)})$ is a non-decreasing sequence bounded superiorly with δ_M, thus it converges to $\Delta^* \in \mathcal{D}_n^+$ and $\Delta^* = f(\Delta^*)$. This fixed point is clearly unique thanks to (1.7).

Now if f is non-increasing then $\Delta \mapsto f^2(\Delta)$ is non-decreasing and bounded inferiorly and superiorly thus it admits a unique fixed point $\Delta^* \in \mathcal{D}_n^+$ satisfying $\Delta^* = f^2(\Delta^*)$. We can deduce that $f(\Delta^*) = f^2(f(\Delta^*))$ which implies by uniqueness of the fixed point that $f(\Delta^*) = \Delta^*$ and the uniqueness of such a Δ^* is again a consequence of (1.7). □

1.6 Complex stable semi-metric and some consequences in random matrix theory

Let us for any $D, D' \in \mathcal{D}_n(\mathbb{H})$ as:

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

(it lacks the triangular inequality to be a true metric). This semi-metric is introduced to set Banach-like fixed point theorems.

Definition 4. *Given $\lambda > 0$, we denote $\mathcal{C}_s^\lambda(\mathcal{D}_n(\mathbb{H}))$ (or more simply $\mathcal{C}_s^\lambda$ when there is no ambiguity), 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 1.15. *Given three parameters $\alpha, \lambda, \theta > 0$ and two mappings $f \in \mathcal{C}_s^\lambda$ and $g \in \mathcal{C}_s^\theta$,*

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

The stability towards the sum is a mere adaptation of Property 1.3

Lemma 1.16. *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 of Lemma 1.16. For any $\Delta, \Delta', D, D' \in \mathcal{D}_n(\mathbb{H})$, there exists $i_0 \in [n]$ such that:

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

thanks to Lemma 1.4. $\square$

We can now present our fixed point theorem that has been demonstrated once again in **?**:

Theorem 1.17 (**?** Theorem 3.13). *Given a subset $\mathcal{D}_f$ of $\mathcal{D}_n(\mathbb{H})$ and a mapping $f : \mathcal{D}_f \to \mathcal{D}_f$ with an imaginary part bounded from above and below (in $\mathcal{D}_f$), if it is furthermore contracting for the stable semi-metric d_s on D_f, then there exists a unique fixed point $\Delta^* \in \mathcal{D}_f$ satisfying $\Delta^* = f(\Delta^*)$.*

We will now employ the semi-metric d_s indifferently on diagonal matrices $\mathcal{D}_n(\mathbb{H})$ or vectors of $\mathbb{H}^n$ or more simply with variables of $\mathbb{H}$ as in next proposition.

Proposition 1.18. *All the Stieltjes transforms are 1-Lipschitz for the semi-metric d_s on $\mathbb{H}$.*

Proof. We consider a Stieltjes transform $g : z \to \int \frac{d\mu(t)}{t-z}$ for a given measure μ on $\mathbb{R}$. Given $z, z' \in \mathbb{H}$, we can bound thanks to Cauchy-Schwarz inequality:

$$\begin{aligned} |g(z) - g(z')| &\leq \left|\int \frac{z' - z}{(t-z)(t-z')} d\mu(t)\right| \leq \left|\frac{z'-z}{\sqrt{\Im(z)\Im(z')}}\right| \left|\int \frac{\sqrt{\Im(z)\Im(z')}}{(t-z)(t-z')} d\mu(t)\right| \\ &\leq \left|\frac{z'-z}{\sqrt{\Im(z)\Im(z')}}\right| \sqrt{\int \frac{\Im(z)}{|t-z|^2} d\mu(t) \int \frac{\Im(z')}{|t-z'|^2} d\mu(t)} \\ &= \sqrt{\Im(g(z))\Im(g(z'))} d_s(z, z') \end{aligned}$$

$\square$

We did not find any particular use of this proposition (the idea could be to solve $g(z) = z$) but the stability it introduces looks interesting.

We deduce directly from Lemma 1.16 that if $f, g : \mathcal{D}_n(\mathbb{H}) \to \mathcal{D}_n(\mathbb{H})$ are λ-Lipschitz for d_s then $f + g$ is also λ-Lipschitz. This property gives a very fast proof to show the existence and uniqueness of solutions to the equations studied in **?** (however their proof is not much longer). We start with a preliminary proposition:

Proposition 1.19. *Given a matrix $S \in \mathcal{M}_{n,p}(\mathbb{R}_+)$, $z \mapsto Sz$ goes from $\mathbb{H}^p$ to $\mathbb{H}^n$ and it is 1-Lipschitz for the semi-metric d_s.*

Proof. For any $z \in \mathbb{H}^p$, $\Im(Sz) = S\Im(z) \in \mathbb{R}_+^p$ since all the entries of S are positive. If we denote $s_1, \dots, s_n$, the columns of S, we can decompose $Sz = \sum_{i=1}^n s_i \pi_i(z)$, where $\pi_i(z) = z_i$. Each mapping $s_i\pi_i$ is 1-Lipschitz for d_s, since we have for any $z, z' \in \mathbb{H}^p$:

$$d_s(s_i\pi_i(z), s_i\pi_i(z')) = \sup_{j\in[p]} \left| \frac{[s_i]_j z_i - [s_i]_j z_i'}{\sqrt{\Im([s_i]_j z_i)\Im([s_i]_j z_i')}} \right| = \frac{z_i - z_i'}{\sqrt{\Im(z_i)\Im(z_i')}},$$

therefore, as a sum of 1-Lipschitz operators, we know that $z \to Sz$ is also 1-Lipschitz for d_s. □

Proposition 1.20. *Given any $a \in \mathbb{R}^n$ and any matrix[2] $S \in \mathcal{M}_n(\mathbb{R}_+)$, and $z \in \mathbb{H}$, the equation:*

$$-\frac{1}{m} = z\mathbb{1} + a + Sm$$

admits a unique solution $m \in \mathbb{C}_+^n$.

Proof. Let us introduce $I : x \to z\mathbb{1} + a - S\frac{\mathbb{1}}{x}$. To employ Theorem 1.17, let us first show that the imaginary part of $I(x)$ is bounded from below and above for all $x \in \mathbb{H}^n$. Given $x \in \mathbb{H}^n$ we see straightforwardly that $\Im(I(x)) \geq \Im(z)$, we can furthermore bound:

$$\Im(I(x)) \leq \Im(z)\mathbb{1} + S\frac{\Im(x)}{|x|^2} \leq \Im(z)\mathbb{1} + S\frac{\mathbb{1}}{\Im(x)} \leq \left(\Im(z)I_p + \frac{1}{\Im(z)}S\right)\mathbb{1} \equiv \kappa_I\mathbb{1}.$$

Besides, we already know from Proposition 1.15 that I is 1-Lipschitz for d_s but we need a Lipschitz parameter lower than 1. Given $x, y \in \mathbb{H}^n$ we can bound thanks to Proposition 1.15 and 1.19:

$$|I(x) - I(y)| \leq \left|S\left(\frac{1}{x} - \frac{1}{y}\right)\right| \leq \sqrt{\Im\left(S\left(\frac{1}{x}\right)\right)\Im\left(S\left(\frac{1}{y}\right)\right)}d_s(x, y)$$

that implies that the Lipschitz parameter of I is lower than:

$$\sqrt{\left(1 - \frac{\Im(z)}{\Im(I(x))}\right)\left(1 - \frac{\Im(z)}{\Im(I(y))}\right)} \leq 1 - \frac{\Im(z)}{\kappa_I} < 1$$

We conclude then with Theorem 1.17 that there exists a unique $x \in \mathbb{H}^n$ such that $x = I(x)$, from which we deduce the existence and uniqueness of $m = \frac{1}{x}$. □

1.7 Stability of the stable semi-metric towards perturbations

A classical problem is to be able to bound the difference between two solution of two contractive equations for the stable semi-metric. This is possible when one of the equation is a small perturbation of the other one. For the absolute value on $\mathbb{R}$ for instance, if one is given two $1 - \varepsilon$ Lipschitz mapping $f, g : \mathbb{R} \to \mathbb{R}$, and two fixed point $x, y \in \mathbb{R}$ satisfying:

$$x = f(x) \qquad \text{and} \qquad y = g(y),$$

[2]Note that unlike in **?**, we do not suppose that S is symmetric.

then one can bound thanks to the triangular inequality:

$$|x-y| \le |f(x)-f(y)| + |f(y)-g(y)| \le (1-\varepsilon)|x-y| + |f(y)-g(y)|, \tag{1.8}$$

which implies $|x-y| \le \frac{|f(y)-g(y)|}{\varepsilon}$. When working with the semi-metric d_s, the triangular inequality is not valid (see (1.2)), one therefore needs more elaborated inferences displayed below. Note than the contractiveness of g is actually completely useless.

Proposition 1.21. *Given a diagonal matrices $\Gamma \in \mathcal{D}_n(\mathbb{H})$, a mapping $f :\to \mathcal{D}_n(\mathbb{H})$ λ-Lipschitz*[3] *for the semi-metric d_s with $\lambda < 1$ and admitting the fixed point $\tilde\Gamma = f(\tilde\Gamma)$, we have the bound:*

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

The same result is valid replacing $\mathbb{H}$ with $\mathbb{R}_+$ and removing the symbols $\Im$.

This proposition is a consequence of the following elementary result that we set for the semi-metric defined on $\mathcal{D}_n(\mathbb{H})$ since it needs more justification than to set the result on $\mathcal{D}_n(\mathbb{R}_+)$ (which merely expresses removing the symbols $\Im$ in the inequality).

Lemma 1.22. *Given three diagonal matrices $\Gamma^1, \Gamma^2, \Gamma^3 \in \mathcal{D}_n(\mathbb{H})$:*

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

Proof. Let us simply bound:

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

we can then conclude with the bound $\sqrt{\Im(\Gamma^1)}\left(\sqrt{\Im(\Gamma^2)} + \sqrt{\Im(\Gamma^1)}\right) \ge \sqrt{\Im(\Gamma^1)\Im(\Gamma^2)}$. □

Proof of Proposition 1.23. Let us simply bound thanks to Lemma 1.22:

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

since $d_s(\Im(\Gamma), \Im(f(\Gamma))) \le o(1) \le \frac{1-\lambda}{2\lambda}$ for s sufficiently big.

□

[3]Actually, f does not need to be λ-Lipschitz on the whole set $\mathcal{D}_n(\mathbb{H})$, but we need to be able to bound:

$$d(f(\tilde\Gamma), f(\Gamma)) \le \lambda d(\tilde\Gamma, \Gamma)$$

Proposition 1.21 will be employed to set asymptotic results when $n \to \infty$ and to show continuity properties with a parameter $t \to 0$ in Chapters 2 and **??**.

Proposition 1.23. *Let us consider a family of mappings of $\mathcal{D}_{n_s}(\mathbb{H})$, $(f^s)_{m\in\mathbb{N}}$, each f^s being λ-Lipschitz for the semi-metric d_s with $\lambda < 1$ and admitting the fixed point $\tilde{\Gamma}^s = f^s(\tilde{\Gamma}^s)$ and a family of diagonal matrices Γ^s. If one assumes that*[4] $d_s(\Im(\Gamma^s), \Im(f^s(\Gamma^s))) \leq o_{s\to\infty}(1)$, *then:*

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

The same result is valid replacing $\mathbb{H}$ with $\mathbb{R}_+$ and removing the symbols $\Im$.

Proof. It suffices to bound for s sufficiently big:

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

since $d_s(\Im(\Gamma), \Im(f(\Gamma))) \leq o(1) \leq \frac{1-\lambda}{2\lambda}$ when s tends to infinity. $\square$

[4] Usually the notations $O(1)$ and $o(1)$ are used for quasi-asymptotic studies when n tends to infinity but in this proposition, the relevent parameter is s, thus $d_s(f^s(\Gamma^s), \Gamma^s) \leq o_{s\to\infty}(1)$ means that for all $K > 0$, there exists $S \in \mathbb{N}$ such that for all $s \geq S$, $d_s(f^s(\Gamma^s), \Gamma^s) \leq K$.

Chapter 2

Statistical study of the resolvent

Denoting $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ the p eigenvalues of $\frac{1}{n}XX^T$ the spectral distribution is defined followingly:

$$\mu = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i}.$$

2.1 Concentration and estimation of the resolvent

We work here under the formalisme of Levy families where the random matrix under study, $X \in \mathcal{M}_{p,n}$ is seen as a sequence depending on n and p is also a sequence depending on n but that should satisfy the following relation:

Assumption 1. $p \leq O(n)$.

We need of course a concentration hypothesis on X:

Assumption 2. $X \propto \mathcal{E}_2$.

A third natural and fundamental hypothesis is to assume that the n columns of $X = (x_1, \ldots, x_n)$ are independent. Again, we do not assume that $x_1, \ldots, x_n$ are identically distributed: we can possibly have n different distributions for the columns of X.

Assumption 3. *X has independent columns $x_1, \ldots, x_n \in \mathbb{R}^p$.*

Let us note for simplicity, for any $i \in [n]$:

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

It is easy to deduce from Assumption 2 (see Proposition **??**) that there exists a constant $K > 0$ such that for all $n \in \mathbb{N}$, $\|C_i\| \leq O(1)$. But to have the best convergence bounds, we also need to impose:[1]

Assumption 4. $\sup_{i\in[n]} \|\mu_i\| \leq O(1)$.

We conclude with a last assumption which seems important to precisely approximate the support of the spectral distribution.[2] Although we are unsure of its importance, our line of arguments could not avoid it; it is nonetheless quite a weak constraint in view of the practical use of our result.

Assumption 5. $\inf_{i\in[n]} \Sigma_i \geq O(1)I_p$.

[1] In **?**, we only supposed that $\|\mu_i\| \leq O(\sqrt{n})$, however, then we only had an approximation of the resolvent with the spectral norm, here we will provide a similar convergence result with the Frobenius norm.

[2] the values of the Stieltjes distribution $g(z)$ can be approximated for $z \in \mathbb{C}$ sufficiently far from the real axis or for $z \in \mathbb{R}$ sufficiently far from the support without this assumption.

2.1.1 Resorting to a "concentration zone" for Q^z

Before studying the matricial case, let us first place ourselves in $\mathbb{R}$. We consider $X \in \mathbb{R}$, a Gaussian random variable with zero mean and variance equal to σ^2 ($X \sim \mathcal{N}(0, \sigma^2)$). In particular, although we work with unidimensional variables, there can still be a possible dependence on n, and we can write $X \propto \mathcal{E}_2$. The random variable $Q \equiv 1/(1 - X)$ is only defined if $X \neq 1$ and its law f_Q can be computed on $\mathbb{R} \setminus \{1\}$ and satisfies:

$$f_Q(q) = \frac{e^{-(1-\frac{1}{q})^2/\sigma^2}}{\sqrt{2\pi}\sigma q^2}.$$

Thus Q is clearly not exponentially concentrated (when $q \to \infty$, $f_Q(q) \sim \frac{e^{-1/\sigma^2}}{q^2}$ therefore the expectation of Q is not even defined). However, if σ is small enough (at least $\sigma \leq o(1)$), it can be interesting to consider the event $\mathcal{A}_Q \equiv \{X \leq \frac{1}{2}\}$ satisfying $\mathbb{P}(\mathcal{A}_Q^c) \leq Ce^{-1/2\sigma^2}$. The mapping $f : z \mapsto \frac{1}{1-z}$ being 4-Lipschitz on $(-\infty, \frac{1}{2}]$, one sees that $(Q \mid \mathcal{A}_Q) \in \mathcal{E}_2$. Following this setting, in the matricial cases, we also need to place ourselves in a concentration zone $\mathcal{A}_Q$ where the fixed point Q is defined; sufficiently small to retrieve an exponential concentration with $Q \mid \mathcal{A}_Q$ but large enough to be highly probable.

The same resort to a *concentration zone* will take place for the resolvent matrix, for that purpose, we introduce in this section an event of high probability, $\mathcal{A}_Q$, under which the eigen values of $\frac{1}{zn}XX^T$ are far from 1, for all $z \in S^\varepsilon$. Let us start with a bound on $\|X\|$, Assumption 4 leads us to:

$$\|\mathbb{E}[X]\| \leq \sqrt{n} \sup_{1 \leq i \leq n} \|\mathbb{E}[x_i]\| \leq O(\sqrt{n})$$

then, we deduce from Example **??** applied to Assumption 2 that $\mathbb{E}[\|X\|] \leq \|\mathbb{E}[X]\| + O(\sqrt{n}) \leq O(\sqrt{n})$. Now, introducing a constant $\varepsilon > 0$ such that $O(1) \leq \varepsilon \leq O(1)$ and $\nu > 0$ defined with:

$$\sqrt{\nu} \equiv \mathbb{E}\left[\frac{1}{\sqrt{n}}\|X\|\right] \leq O(1),$$

we denote:

$$\mathcal{A}_\nu \equiv \left\{\frac{1}{n}\left\|XX^T\right\| \leq \nu + \varepsilon\right\}. \tag{2.1}$$

Since $\|X\|/\sqrt{n} \in \nu \pm \mathcal{E}_2(1/\sqrt{n})$, we know that there exist two constants $C, c > 0$ such that $\mathbb{P}\left(\mathcal{A}_\nu^c\right) \leq Ce^{-cn}$. The mapping $M \to \frac{1}{n}MM^T$ is $O(1/\sqrt{n})$ Lipschitz on $X(\mathcal{A}_\nu) \subset \mathcal{M}_{p,n}$ and therefore, thanks to Lemma **??** and Remark **??**:

$$\frac{1}{n}XX^T \mid \mathcal{A}_\nu \propto \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right)$$

Let us note $\sigma : \mathcal{M}_p \to \mathbb{R}^p$, the mapping that associates to any matrix the sequence of its eigen values in decreasing order. It is well known (see (**?**, Theorem 8.1.15)) that σ is 1-Lipschitz (from $(\mathcal{M}_p, \|\cdot\|_F)$ to $(\mathbb{R}^p, \|\cdot\|)$) and therefore if we denote $\lambda_1, \ldots, \lambda_p$ the eigen values of $\frac{1}{n}XX^T$ such that $\lambda_1 \geq \cdots \geq \lambda_n$, we have the concentration:

$$(\lambda_1, \ldots, \lambda_p) \mid \mathcal{A}_\nu \in (\mathbb{E}_{\mathcal{A}_\nu}[\lambda_1], \ldots, \mathbb{E}_{\mathcal{A}_\nu}[\lambda_p]) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right).$$

Given a set $T \subset \mathbb{C}$, we note for any $\varepsilon > 0$, $T^\varepsilon = \{z \in \mathbb{C}, \exists t \in T, |z - t| \leq \varepsilon\}$. Now, let us denote

$$S \equiv \{\mathbb{E}_{\mathcal{A}_\nu}[\lambda_1], \ldots, \mathbb{E}_{\mathcal{A}_\nu}[\lambda_p]\},$$

(with the expectation taken on $\mathcal{A}_\nu$). We show in the next lemma that the event

$$\mathcal{A}_Q \equiv \mathcal{A}_\nu \cap \{\forall i \in [\min(p, n)] : \lambda_i \in S^{\varepsilon/2}\}$$

has an overwhelming probability.

Lemma 2.1. *There exist $C, c > 0$ such that $\forall n \in \mathbb{N}$ $\mathbb{P}(\mathcal{A}_Q^c) \leq Ce^{-cn}$.*

Proof of Lemma 2.1. Starting from the identity $S^{\frac{\varepsilon}{2}} \cup_{i\in[p]} \left[\mathbb{E}_{\mathcal{A}_\nu}[\lambda_i] - \frac{\varepsilon}{2}, \mathbb{E}_{\mathcal{A}_\nu}[\lambda_i] + \frac{\varepsilon}{2}\right]$, we see that $S^{\frac{\varepsilon}{2}}$ is a union of, say, d intervals of $\mathbb{R}$. There exists $2d$ indexes $i_1 \leq \cdots \leq i_d$ and $j_1 \leq \cdots \leq j_d$ in $[\min(p,n)]$ such that:

$$S^{\frac{\varepsilon}{2}} = \bigcup_{k\in[d]} \left[\mathbb{E}_{\mathcal{A}_\nu}[\lambda_{i_k}] - \frac{\varepsilon}{2}, \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{j_k}] + \frac{\varepsilon}{2}\right].$$

Since $\mathbb{E}_{\mathcal{A}_\nu}[\lambda_{i_1}] \geq 0$ and $\mathbb{E}_{\mathcal{A}_\nu}[\lambda_{j_d}] \leq \sqrt{\nu} \pm O(e^{-cn}) \leq O(1)$, we can bound:

$$\varepsilon d \leq \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{j_d}] + \varepsilon \leq O(1),$$

That implies in particular that $d \leq O(1)$ because $\varepsilon \geq O(1)$. We can then bound thanks to the concentration of the $2d$ random variables $\lambda_{i_1}, \ldots, \lambda_{i_d}$ and $\lambda_{j_1}, \ldots, \lambda_{j_d}$:

$$\begin{aligned}\mathbb{P}(\mathcal{A}_Q^c) &= \mathbb{P}\left(\exists k \in [d], |\lambda_{i_k} - \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{i_k}]| > \frac{\varepsilon}{2} \text{ or } |\lambda_{j_k} - \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{j_k}]| > \frac{\varepsilon}{2}\right) \\ &\leq \sum_{k=1}^{d} \left(\mathbb{P}\left(|\lambda_{i_k} - \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{i_k}]| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|\lambda_{j_k} - \mathbb{E}_{\mathcal{A}_\nu}[\lambda_{j_k}]| > \frac{\varepsilon}{2}\right)\right) \quad \leq 2dCe^{-cn\varepsilon^2/4},\end{aligned}$$

where C and c are the two constants appearing in the concentration inequality of $(\lambda_1, \ldots, \lambda_p) = \sigma(\frac{1}{n}XX^T)$. $\square$

Now, let us remark that the spectrum of $\frac{1}{n}XX^T$ is closely related to the spectrum of $\frac{1}{n}X^TX$ via the equivalence:

$$\lambda \in \operatorname{Sp}\left(\frac{1}{n}XX^T\right) \setminus \{0\} \qquad \Longleftrightarrow \qquad \lambda \in \operatorname{Sp}\left(\frac{1}{n}X^TX\right) \setminus \{0\}.$$

As a consequence, one of the two matrices has $|n-p|$ supplementary zeros in its spectrum. Since those zeros do not carry any specific information about the distribution of $\frac{1}{n}XX^T$, we try to remove them from the study to re-establish the symmetry. The $\min(p,n)$ first entries of $\sigma(\frac{1}{n}XX^T)$ and $\sigma(\frac{1}{n}X^TX)$ are the same (and some of them can cancel), we thus naturally introduce the set:

$$S_{-0} \equiv \left\{\mathbb{E}_{\mathcal{A}_Q}[\lambda_i], i \in [\min(p,n)]\right\},$$

Be careful that if $0 \in \operatorname{Sp}\left(\frac{1}{n}XX^T\right) \cap \operatorname{Sp}\left(\frac{1}{n}X^TX\right)$ then $0 \in S_{-0}$.

To avoid the issue on zero, instead of studying, as it is usually done, the resolvent $(\frac{1}{n}XX^T - zI_p)^{-1}$, we rather look at:

$$Q^z \equiv \left(I_p - \frac{1}{zn}XX^T\right)^{-1},$$

that has the advantage of satisfying $\|Q^z\| \leq O(1)$, for all $z \in S_{-0}$, which will allow us to show the concentration of Q^z even for z close to zero when $n < p$ (and $0 \notin S_{-0}$).

One of the objective of the study was to establish converging bounds when z is at a constant distance from S_{-0}. We were not able to set it for all our results (we failed with Proposition 2.20 where we could only set the result for z at a constant distance from S) however, we hope to be able to set the general result after further research, we thus keep, when accessible, general formulations with S_{-0}.

2.1.2 Concentration of the resolvent $Q^z = (I_n - \frac{1}{nz}XX^T)^{-1}$.

Given a matrix $A \in \mathcal{M}_{p,n}(\mathbb{C})$, we denote $|A| = \sqrt{A\bar{A}^T}$ ($A\bar{A}^T$ is a nonnegative Hermitian matrix). With a simple diagonalization procedure, one can show for any Hermitian matrix A the simple characterization of the spectrum of $|A|$:

$$\operatorname{Sp}(|A|) = \{|\lambda|, \lambda \in \operatorname{Sp}(A)\}, \tag{2.2}$$

note also that for any real vector $u \in \mathbb{R}^p$:

$$|u^T A u| \leq u^T |A| u. \tag{2.3}$$

It is possible to go further than the mere bound $|Q^z| \leq O(1)$ when z is close to 0 and $p \leq n$, we are then able to show that $|Q^z| \leq O(|z|/(|z|+1))$. When $p \geq n$ no such bound is true and it is then convenient to rather look at the coresolvent $\check{Q}^z \equiv (I_p - \frac{1}{nz} X^T X)^{-1}$ that satisfies in that regime $|\check{Q}^z| \leq O(|z|/(|z|+1))$. To formalize this approach, we introduce two quantities that will appear in our convergence speeds:

$$\kappa_z \equiv \begin{cases} \dfrac{|z|}{1+|z|} & \text{if } p \leq n \\ 1 & \text{if } n \leq p \end{cases} \qquad \check{\kappa}_z \equiv \begin{cases} 1 & \text{if } p \leq n \\ \dfrac{|z|}{1+|z|} & \text{if } n \leq p \end{cases}$$

note that both of them are bounded by 1 but they can tend to zero with $|z|$ depending on the sign of $p - n$, besides:

$$\kappa_z \check{\kappa}_z = \frac{|z|}{1+|z|}. \tag{2.4}$$

Note than in our formalism, the parameter z, as most quantities of our manuscript, is varying with n (we do not assume that $O(1) \leq |z| \leq O(1)$ like ε. It is not a "constant").

Lemma 2.2. *Under $\mathcal{A}_Q$, Q^z and $\check{Q}^z$ can be defined on 0 and for any*[3] $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$:

$$O(\kappa_z) I_p \leq |Q^z| \leq O(\kappa_z) I_p \qquad \textit{and} \qquad O(\check{\kappa}_z) I_p \leq |\check{Q}^z| \leq O(\check{\kappa}_z) I_p$$

(for the classical order relation on hermitian matrices).

Proof. We can diagonalize the nonnegative symmeric matrix: $\frac{1}{n} X X^T = P D P^T$, with $D = \mathrm{Diag}(\lambda_1, \cdots, \lambda_p)$ and $P \in \mathcal{O}_p$, an orthogonal matrix. There exists $q \in [p]$ such that $\lambda_{q+1} = \cdots = \lambda_p = 0$, and for all $i \leq q$, $\lambda_i \neq 0$ (possibly $q = p$ or $q < \min(p, n)$). Then, if we denote $P_0 \in \mathcal{M}_{p,p-q}$ the matrix composed of the $p - q$ last columns of P, P_+ the matrix composed of the rest of the q columns, and $D_+ = \mathrm{Diag}(\lambda_1, \ldots, \lambda_q)$, we can decompose:

$$Q^z = P \left(\frac{z I_p}{z I_p - D} \right) P^T = P_0 P_0^T + P_+ \left(\frac{z I_q}{z I_q - D_+} \right) P_+^T.$$

Since $P_0 P_+^T = 0$ and $P_+ P_0^T = 0$, we have: $|Q^z|^2 = P_0 P_0^T + P_+ \left(\frac{|z|^2 I_q}{|z I_q - D_+|^2} \right) P_+^T$. We can bound:

$$O\left(\frac{|z|}{1+|z|} \right) \leq \frac{|z|}{|z| + \nu + \varepsilon} \leq P_+ \left(\frac{|z| I_q}{|z I_q - D_+|} \right) P_+^T \leq \frac{|z|}{d(z, S_{-0}) - \frac{\varepsilon}{2}} \leq O\left(\frac{|z|}{1+|z|} \right)$$

Therefore, in all cases ($p \leq n$ or $p \geq n$) $O(1) \leq |Q^z| \leq O(1)$. However, when $p \leq n$, we can precise those bounds:

- if $\mathbb{E}_{\mathcal{A}_Q}[\lambda_{\min(p,n)}] > \frac{\varepsilon}{2}$, then P_0 is empty, $P_+ = P$ and we see that $O\left(\frac{|z|}{1+|z|} \right) \leq |Q^z| \leq O\left(\frac{|z|}{1+|z|} \right)$.
- if $\mathbb{E}_{\mathcal{A}_Q}[\lambda_{\min(p,n)}] \leq \frac{\varepsilon}{2}$, the bound $d(z, S_{-0}) \geq \varepsilon$ implies that $|z| \geq \frac{\varepsilon}{2}$ and $O(\kappa_z) = O(\frac{|z|}{1+|z|}) \leq O(1) \leq O(\kappa_z)$, therefore:

$$O(\kappa_z) \leq O(\min(1, \kappa_z)) I_p \leq |Q^z| \leq O(1 + \kappa_z) I_p \leq O(\kappa_z).$$

The inequalities on $\check{Q}^z$ are proven the same way.

□

[3] In theory, both the parameter z and the set S^{ε}_{-0} depends on our asymptotic parameter n, so one should rigorously write $z_n \in S^{\varepsilon}_{-0}(n)$

Proposition 2.3. *Given $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$, we have the concentrations $Q^z \mid \mathcal{A}_Q \propto \mathcal{E}_2(\kappa_z/\sqrt{n})$ in $(\mathcal{M}_p, \|\cdot\|_F)$ and $\check{Q}^z \mid \mathcal{A}_Q \propto \mathcal{E}_2(\check{\kappa}_z/\sqrt{n})$ in $(\mathcal{M}_n, \|\cdot\|_F)$.*

Proof. Noting $\Phi : \mathcal{M}_{p,n} \to \mathcal{M}_p(\mathbb{C})$ and $\check{\Phi} : \mathcal{M}_{p,n} \to \mathcal{M}_n(\mathbb{C})$ defined as:

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

it is sufficient to show that Φ (resp. $\check{\Phi}$) is $O(\kappa_z/\sqrt{n})$-Lipschitz (resp. $O(\check{\kappa}_z/\sqrt{n})$-Lipschitz) on $\mathcal{M}_{n,p}^{\mathcal{A}_Q} \equiv X(\mathcal{A}_Q)$[4]. For any $M \in \mathcal{M}_{n,p}^{\mathcal{A}_Q}$ and any $H \in \mathcal{M}_{p,n}$, we can bound $\|M\| \leq (\nu + \varepsilon)\sqrt{n} \leq O(\sqrt{n})$ and:

$$\|d\Phi_{|M} \cdot H\|_F = \left\|\Phi\left(M\right)\frac{1}{nz}(MH^T + HM^T)\Phi\left(M\right)\right\|_F$$

We can now distingish two different cases:

- if $p \leq n$ then:

$$\|d\Phi_{|M} \cdot H\|_F \leq O\left(\frac{\|H\|_F|z|}{(1+|z|)^2\sqrt{n}}\right) \leq O\left(\frac{\|H\|_F|z|/\sqrt{n}}{1+|z|}\right).$$

- if $p \geq n$, we know that $\|\Phi(M)\| \leq O(\kappa_z) \leq O(1)$ and $\|\check{\Phi}(M)\| \leq O(\check{\kappa}_z) \leq O(\frac{|z|}{1+|z|})$. We employ then the classical identity:

$$\Phi(M)M = M^T\check{\Phi}(M) \tag{2.5}$$

  to be able to bound:

$$\begin{aligned}\|d\Phi_{|M} \cdot H\|_F &\leq \frac{2(\nu+\varepsilon)\|H\|_F}{|z|\sqrt{n}}\left\|\check{\Phi}(M)\right\|\left\|\Phi\left(M\right)\right\| \\ &\leq O\left(\frac{\|H\|_F}{(1+|z|)\sqrt{n}}\right) \leq O\left(\frac{\|H\|_F}{\sqrt{n}}\right)\end{aligned}$$

Thus, in all cases, under $X(\mathcal{A}_Q)$, Q^z is a $O(\kappa_z/\sqrt{n})$-Lipschitz transformation of $X \propto \mathcal{E}_2(1)$ and as such, it satisfies the concentration inequality of the proposition thanks to Lemma **??** and Remark **??**. The same holds for $\check{Q}^z$. □

Remark 2.4. *The preceding proof partly relies on the assertion $\|X\| \leq \sqrt{n}$ available under the event $\mathcal{A}_Q$. However one does not need it to be able to prove that Q is a Lipschitz transformation of X since one can bound the element $QX/\sqrt{n}$ as a whole thanks to the formulas:*

$$\frac{1}{zn}Q^zXX^T = Q^z - I_n \qquad \textit{and} \qquad \frac{1}{zn}\check{Q}^zX^TX = \check{Q}^z - I_n. \tag{2.6}$$

On the one hand:

$$\left\|\frac{1}{\sqrt{n}}Q^zX\right\| \leq \sqrt{|z|\left\|\frac{1}{zn}Q^zXX^TQ^z\right\|} \leq \sqrt{|z|\,\|(Q^z)^2 - Q^z\|} \leq O(\sqrt{|z|}(\kappa_z + \kappa_z^2)) \leq O(\sqrt{|z|}\kappa_z)$$

On the other hand, (2.5) *gives us:*

$$\left\|\frac{1}{\sqrt{n}}Q^zX\right\| = \|\frac{1}{\sqrt{n}}X\check{Q}^z\| \leq \sqrt{|z|\left\|\frac{1}{zn}Q^zX^TXQ^z\right\|} \leq O(\sqrt{|z|}\check{\kappa}_z)$$

[4] $\mathcal{M}_{n,p}^{\mathcal{A}_Q} \subset \{M \in \mathcal{M}_{n,p}, \frac{1}{n}\|MM^T\| \leq \nu + \varepsilon\}$

2.1.3 A first deterministic equivalent

One is often merely working with linear functionals of Q^z, and since Proposition 2.3 implies that $Q^z \mid \mathcal{A}_Q \in \mathbb{E}_{\mathcal{A}_Q} Q^z \pm \mathcal{E}_2$, one naturally wants to estimate the expectation $\mathbb{E}_{\mathcal{A}_Q}[Q^z]$.

In **?** is provided a deterministic equivalent $\tilde{Q}^z \in \mathcal{M}_p(\mathbb{C})$ satisfying $\|\mathbb{E}[Q^z] - \tilde{Q}^z\| \leq O(1/\sqrt{n})$ for any $z \in \mathbb{R}^-$, we are going to show below a stronger result,

- with a Frobenius norm replacing the spectral norm,
- for any complex $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$,
- for random vectors $x_1, \ldots, x_n$ having possibly different distributions (it was assumed in **?** that there was a finite number of classes)

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:

$$\tilde{Q}^{\Delta} = \left(I_p - \Sigma^{\Delta}\right)^{-1} \qquad \text{where} \qquad \Sigma^{\Delta} \equiv \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\Delta_i} = \frac{1}{n}\mathbb{E}[X\Delta^{-1}X^T]. \tag{2.7}$$

One can then express the difference with the expectation of Q^z under $\mathcal{A}_Q$, $\mathbb{E}_{\mathcal{A}_Q}[Q^z]$ followingly:

$$\begin{aligned}\mathbb{E}_{\mathcal{A}_Q}[Q^z] - \tilde{Q}^{\Delta} &= \mathbb{E}_{\mathcal{A}_Q}\left[Q^z\left(\frac{1}{zn}XX^T - \Sigma^{\Delta}\right)\tilde{Q}^{\Delta}\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[Q^z\left(\frac{x_i x_i^T}{z} - \frac{\Sigma_i}{\Delta_i}\right)\tilde{Q}^{\Delta}\right].\end{aligned}$$

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

$$Q^z = Q^z_{-i} + \frac{1}{zn}\frac{Q^z_{-i}x_i x_i^T Q^z_{-i}}{1 - \frac{1}{zn}x_i^T Q^z_{-i} x_i} \qquad \text{and} \qquad Q^z x_i = \frac{Q^z_{-i}x_i}{1 - \frac{1}{zn}x_i^T Q^z_{-i}x_i}, \tag{2.8}$$

for $Q^z_{-i} = (I_n - \frac{1}{zn}X_{-i}X_{-i}^T)^{-1}$ (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:

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \qquad \text{or} \qquad A^{-1} + B^{-1} = A^{-1}(A + B)B^{-1} \tag{2.9}$$

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

Introducing the notation:

$$\Lambda^z \equiv \mathrm{Diag}_{1\leq i\leq n}(z - \frac{1}{n}x_i^T Q^z_{-i}x_i), \qquad \text{satisfying:} \quad \forall i \in [n]: \ Q^z x_i = \frac{z}{\Lambda^z}Q^z_{-i}x_i,$$

one can express thanks to the independence between Q^z_{-i} and x_i:

$$\begin{aligned}\mathbb{E}_{\mathcal{A}_Q}[Q^z] - \tilde{Q}^{\Delta} &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[Q^z_{-i}\left(\frac{x_i x_i^T}{\Lambda^z_i} - \frac{\Sigma_i}{\Delta_i}\right)\tilde{Q}^{\Delta}\right] + \frac{1}{n}\sum_{i=1}^n \frac{1}{\Delta_i}\mathbb{E}_{\mathcal{A}_Q}\left[(Q^z_{-i} - Q^z)\Sigma_i\tilde{Q}^{\Delta}\right] \\ &= \varepsilon_1 + \delta_1 + \delta_2 + \varepsilon_2 \end{aligned} \tag{2.10}$$

$$\text{with :} \quad \begin{cases} \varepsilon_1 = \frac{1}{n}\mathbb{E}_{\mathcal{A}_Q}\left[\sum_{i=1}^n Q^z_{-i}x_i\left(\frac{\Delta_i - \Lambda^z_i}{\Lambda^z_i\Delta_i}\right)x_i^T\tilde{Q}^{\Delta}\right] = \frac{1}{n}\mathbb{E}_{\mathcal{A}_Q}\left[Q^z X\left(\frac{\Delta - \Lambda^z}{z\Delta}\right)X^T\tilde{Q}^{\Delta}\right] \\ \delta_1 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[Q^z_{-i}\left(\frac{x_i x_i^T - \mathbb{E}_{\mathcal{A}_Q}[x_i x_i^T]}{\Delta_i}\right)\tilde{Q}^{\Delta}\right] \\ \delta_2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[Q^z_{-i}\left(\frac{\mathbb{E}_{\mathcal{A}_Q}[x_i x_i^T] - \Sigma_i}{\Delta_i}\right)\tilde{Q}^{\Delta}\right] \\ \varepsilon_2 = -\frac{1}{zn^2}\sum_{i=1}^n \frac{1}{\Delta_i}\mathbb{E}_{\mathcal{A}_Q}\left[Q^z x_i x_i^T Q^z_{-i}\Sigma_i\tilde{Q}^{\Delta}\right], \end{cases}$$

where we recall that $Q^z - Q^z_{-i} = \frac{1}{nz} Q^z x_i x_i^T Q^z_{-i}$. From this decomposition, one is enticed into choosing, in a first step $\Delta \approx \mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] \in \mathcal{D}_n(\mathbb{C})$ so that ε_1 would be small. We will indeed take for Δ, the deterministic diagonal matrix:

$$\hat{\Lambda}^z \equiv \mathrm{Diag}\left(z - \frac{1}{n}\mathrm{Tr}(\Sigma_i \mathbb{E}_{\mathcal{A}_Q}[Q^z])\right)_{1 \leq i \leq n} \in \mathcal{D}_n(\mathbb{C}).$$

Lemma 2.5. *Given $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$, $(\Lambda^z \mid \mathcal{A}_Q) \propto \mathcal{E}_2(\kappa_z/\sqrt{n})$ in $(\mathcal{D}_n(\mathbb{C}), \|\cdot\|)$.*

Proof. Inspiring from the proof of Proposition 2.3, one can show easily that for all $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$, the mapping $X \to \Lambda^z = \mathrm{Diag}_{i\in[n]}(z - \frac{1}{n} x_i^T Q^z_{-i} x_i)$ is a $O(\kappa^z/\sqrt{n})$-Lipschitz transformation from $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ to $(\mathcal{D}_n, \|\cdot\|)$ under $\mathcal{A}_Q$ (since then $\|x_i\| \leq O(\sqrt{n})$). □

Putting the Schur identities (2.8), the relation (2.5) and the formula (2.11)

$$\frac{1}{zn}\check{Q}^z X X^T = \check{Q}^z - I_n, \tag{2.11}$$

together one obtains:

$$\frac{z}{\Lambda^z} = \mathrm{Diag}_{i\in[n]}\left(\frac{1}{1 - \frac{1}{zn} x_i^T Q^z_{-i} x_i}\right) = I_n + \frac{1}{zn}\mathrm{Diag}(X^T Q^z X) = \mathrm{Diag}(\check{Q}^z) \tag{2.12}$$

To be able to use Proposition **??** with $\hat{\Lambda}^z$, one further needs:

Lemma 2.6. *$\|\mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] - \hat{\Lambda}^z\| \leq O(\kappa_z/n)$.*

This lemma that seems quite simple actually requires three preliminary results whose main aim is to show that Q^z_{-i} is close to Q^z. Let us first try and bound Λ^z thanks to (2.12).

Lemma 2.7. *Given $z \in \mathbb{C} \setminus S^{\varepsilon}_{-0}$:*

$$O(|z|)\, I_n \leq O\left(\frac{|z|}{\check{\kappa}_z}\right) I_n \leq |\Lambda^z| \leq O\left(\frac{|z|}{\check{\kappa}_z}\right) I_n \leq O(1 + |z|)\, I_n$$

Proof. The inequalities provided in (2.2), (2.3) and Lemma 2.2 imply:

$$O(\check{\kappa}_z) \leq \inf_{i\in[n]} \mathrm{Sp}(\check{Q}^z) \leq |\,\mathrm{Diag}(\check{Q}^z)| = |\,\mathrm{Diag}_{i\in[n]}(e_i^T \check{Q}^z e_i)| \leq \sup_{i\in[n]} \mathrm{Sp}(\check{Q}^z) \leq O(\check{\kappa}_z)$$

where $e_1, \ldots, e_n$ are the n vectors of the canonical basis of $\mathbb{R}^n$ ($e_i \in \mathbb{R}^n$ is full of 0 except in the i^{th} entry where there is a 1). One can then deduce the result of the lemma thanks to (2.12). □

Then to be able to neglect the dependence relation between x_i and X under $\mathcal{A}_Q$ we introduce the following lemma.

Lemma 2.8 (Independence under $\mathcal{A}_Q$)**.** *Given two mappings $f : \mathbb{R}^p \to \mathcal{M}_p$ and $g : \mathcal{M}_{p,n} \to \mathcal{M}_p$ such that under $\mathcal{A}_Q$, $\|f(x_i)\|_F \leq O(\kappa_f)$ and $\|g(X_{-i})\| \leq O(\kappa_g)$, we can approximate:*

$$\left\|\mathbb{E}_{\mathcal{A}_Q}[f(x_i)g(X_{-i})] - \mathbb{E}_{\mathcal{A}_Q}[f(x_i)]\mathbb{E}_{\mathcal{A}_Q}[g(X_{-i})]\right\|_F \leq O\left(\kappa_f \kappa_g e^{-cn}\right),$$

for some constant $c \geq O(1)$.

Proof. Let us continue $f_{|x_i(\mathcal{A}_Q)}$ and $g_{|X_{-i}(\mathcal{A}_Q)}$ respectively on $\mathbb{R}^p$ and on $\mathcal{M}_{p,n}$ defining for any $x \in \mathbb{R}^p$ and $M \in \mathcal{M}_{p,n}$:

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in x_i(\mathcal{A}_Q) \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \tilde{g}(M) = \begin{cases} g(M) & \text{if } M \in X_{-i}(\mathcal{A}_Q) \\ 0 & \text{otherwise} \end{cases}$$

Let us estimate:

$$\mathbb{E}_{\mathcal{A}_Q}\left[f(x_i)g(X_{-i})\right] = \frac{\mathbb{E}\left[\mathbb{1}_{X\in X(\mathcal{A}_Q)}f(x_i)g(X_{-i})\right]}{\mathbb{P}(\mathcal{A}_Q)}$$
$$= \mathbb{E}\left[\mathbb{1}_{X\in X(\mathcal{A}_Q)}\tilde f(x_i)\tilde g(X_{-i})\right] + \frac{1-\mathbb{P}(\mathcal{A}_Q)}{\mathbb{P}(\mathcal{A}_Q)}\mathbb{E}\left[\tilde f(x_i)\tilde g(X_{-i})\mathbb{1}_{X\in X(\mathcal{A}_Q)}\right].$$

the far right hand term cancels since $\mathcal{A}_Q$ has a probability bigger than $1-Ce^{-cn}$ for some constants $C, c > 0$:

$$\frac{1-\mathbb{P}(\mathcal{A}_Q)}{\mathbb{P}(\mathcal{A}_Q)}\mathbb{E}\left[\tilde f(x_i)\tilde g(X_{-i})\mathbb{1}_{X\in X(\mathcal{A}_Q)}\right] \le \frac{Ce^{-cn}}{1-Ce^{-cn}}\mathbb{E}\left[\left|\tilde f(x_i)\tilde g(X_{-i})\right|\right] \le O\left(\kappa_f\kappa_g e^{-cn}\right).$$

For all $\omega \in \Omega$, we besides know that:

$$\mathbb{1}_{X\in X(\mathcal{A}_Q)}(w) = \mathbb{1}_{X(\omega)\in X(\mathcal{A}_Q)}(\omega) \le \mathbb{1}_{x_i\in x_i(\mathcal{A}_Q)}(\omega)\mathbb{1}_{X_{-i}\in X_{-i}(\mathcal{A}_Q)}(\omega),$$

and the inequality $\mathbb{1}_{\mathcal{A}_Q}(\omega) \neq \mathbb{1}_{x_i\in x_i(\mathcal{A}_Q)}(\omega)\mathbb{1}_{X_{-i}\in X_{-i}(\mathcal{A}_Q)}(\omega)$ only happens for $\omega \in \mathcal{A}_Q^c$. We can then bound:

$$\left|\mathbb{E}\left[\left(\mathbb{1}_{X\in\mathcal{A}_Q} - \mathbb{1}_{x_i\in x_i(\mathcal{A}_Q)}\mathbb{1}_{X_{-i}\in X_{-i}(\mathcal{A}_Q)}\right)\tilde f(x_i)\tilde g(X_{-i})\right]\right| \le \kappa_f\kappa_g\mathbb{E}\left[\mathbb{1}_{\mathcal{A}_Q}\right] \le O\left(\kappa_f\kappa_g e^{-cn}\right),$$

which allows us to set[5] thanks to the independence between X_{-i} and x_i:

$$\mathbb{E}_{\mathcal{A}_Q}\left[f(x_i)g(X_{-i})\right] = \mathbb{E}\left[\mathbb{1}_{x_i\in x_i(\mathcal{A}_Q)}\tilde f(x_i)\mathbb{1}_{X_{-i}\in X_{-i}(\mathcal{A}_Q)}\tilde g(X_{-i})\right] + O_{\|\cdot\|_F}(\kappa_f\kappa_g e^{-cn})$$
$$= \mathbb{E}\left[\tilde f(x_i)\right]\mathbb{E}\left[\tilde g(X_{-i})\right] + O_{\|\cdot\|_F}(\kappa_f\kappa_g e^{-cn})$$
$$= \mathbb{E}_{\mathcal{A}_Q}\left[f(x_i)\right]\mathbb{E}_{\mathcal{A}_Q}\left[g(X_{-i})\right] + O_{\|\cdot\|_F}(\kappa_f\kappa_g e^{-cn})$$

□

As it was done in Proposition 2.3, one can show that $u^TQ^z_{-i}x_i$ and $u^TQ^z_{-i}x_i$ are both κ_z-Lipschitz transformations of $X \propto \mathcal{E}_2$ and deduce:

Lemma 2.9. *Given a deterministic vector $u \in \mathbb{R}^p$, we have the concentration:*

$$u^TQ^z_{-i}x_i \mid \mathcal{A}_Q \in O(\kappa_z) \pm \mathcal{E}_2(\kappa_z)$$

Proof. As it was done in Proposition 2.3, one can show that $u^TQ^z_{-i}x_i$ is a κ_z-Lipschitz transformations of $X \propto \mathcal{E}_2$ and deduce the concentration. Besides, one can bound thanks to Lemma 2.8 and our assumptions:

$$\left|\mathbb{E}_{\mathcal{A}_Q}[u^TQ^z_{-i}x_i]\right| \le \left|u^T\mathbb{E}_{\mathcal{A}_Q}[Q_{-i}]m_i\right| + O\left(\frac{\kappa_z}{n}\right) \le O(\kappa_z)$$

□

Lemma 2.10. *Given $z \in \mathbb{C}\setminus S^{\varepsilon}_{-0}$, for all $i \in [n]$, $\|\mathbb{E}_{\mathcal{A}_Q}[Q^z - Q^z_{-i}]\| \le O(\frac{\kappa_z}{n})$, the same way, one has $\|\mathbb{E}_{\mathcal{A}_Q}[|Q^z|^2 - |Q^z_{-i}|^2]\| \le O(\frac{\kappa_z^2}{n})$.*

Proof. As we saw with Lemma 2.8, we can consider that under $\mathcal{A}_Q$, X_{-i} and x_i are almost independent. We now omit the exponent "z" on Q and Q_{-i} for simplicity. Given two deterministic vectors $u, v \in \mathbb{R}^p$, we can deduce from Lemma 2.9 and the estimation of the product of concentrated random variables given in Lemma **??**:

$$\left|u^T\mathbb{E}_{\mathcal{A}_Q}[Q-Q_{-i}]v\right| \le \left|\mathbb{E}_{\mathcal{A}_Q}\left[\frac{u^TQ_{-i}x_ix_i^TQ_{-i}v}{n\Lambda_i}\right]\right|$$
$$\le \frac{\check\kappa_z}{n|z|}\mathbb{E}_{\mathcal{A}_Q}\left[|u^TQ_{-i}x_i||x_i^TQ_{-i}v|\right] \le O\left(\frac{\check\kappa_z\kappa_z}{n|z|}\right) \le O\left(\frac{\kappa_z}{n}\right).$$

[5] for $x_p, y_p \in \mathcal{M}_p$ and $(a_p)_{p\in\mathbb{N}} \in \mathbb{R}_+^{\mathbb{N}}$ the notation $x_p = y_p + O_{\|\cdot\|_F}(a_p)$ signifies that $\|x_p - y_p\|_F \le O(a_p)$

thanks to the formula $\kappa_z \check{\kappa}_z = \frac{|z|}{1+|z|}$.

For the difference of the squares, let us bound with the same justifications:

$$\begin{aligned}
\left|u^T \mathbb{E}_{\mathcal{A}_Q}[|Q|^2 - |Q_{-i}|^2]v\right| &= \left|u^T \mathbb{E}_{\mathcal{A}_Q}[\bar{Q}Q - \bar{Q}_{-i}Q_{-i}]v\right| \\
&= \left|\mathbb{E}_{\mathcal{A}_Q}\left[\frac{u^T Q_{-i}x_i x_i^T \bar{Q} Q_{-i} v}{n\Lambda_i}\right]\right| + \left|\mathbb{E}_{\mathcal{A}_Q}\left[\frac{u^T \bar{Q}_{-i} Q_{-i} x_i x_i^T \bar{Q}_{-i} v}{n\Lambda_i}\right]\right| \\
&\leq O\left(\frac{\check{\kappa}_z \kappa_z^3}{|z|n}\right) \leq \ O\left(\frac{\kappa_z^2}{n}\right),
\end{aligned}$$

□

Proof of Lemma 2.6. The proof of the concentration of Λ^z was already presented before the lemma. We are just left to bound:

$$\begin{aligned}
&\left\|\hat{\Lambda}^z - \mathbb{E}_{\mathcal{A}_Q}[\Lambda^z]\right\| \\
&\quad \leq \frac{1}{n}\sup_{i\in[n]}\left(\left|\mathrm{Tr}\left(\Sigma_i \mathbb{E}_{\mathcal{A}_Q}[Q - Q_{-i}]\right)\right| + \left|\mathrm{Tr}\left((\Sigma_i - \mathbb{E}_{\mathcal{A}_Q}[x_i x_i^T])\mathbb{E}_{\mathcal{A}_Q}[Q_{-i}]\right)\right|\right. \\
&\quad \left. + \left|\mathrm{Tr}\left(\mathbb{E}_{\mathcal{A}_Q}\left[x_i x_i^T\right]\mathbb{E}_{\mathcal{A}_Q}[Q_{-i}] - \mathbb{E}_{\mathcal{A}_Q}[x_i x_i^T Q_{-i}]\right)\right|\right) \ \leq \ O\left(\frac{\kappa_z}{n}\right)
\end{aligned} \tag{2.13}$$

with the same arguments as in the proof of Lemma 2.10. □

One can directly deduce from the concentration of Λ^z and Lemma 2.7 a bound on its expectation $\hat{\Lambda}^z$.

Lemma 2.11. *Given $z \in \mathbb{C} \setminus S_{-0}^{\varepsilon}$:*

$$O(|z|)\, I_n \leq O\left(\frac{|z|}{\check{\kappa}_z}\right) I_n \leq \left|\hat{\Lambda}^z\right| \leq O\left(\frac{|z|}{\check{\kappa}_z}\right) I_n \leq O\left(1+|z|\right) I_n$$

Proof. One already know from Lemma 2.7 that $O\left(\frac{|z|}{\check{\kappa}_z}\right) \leq |\Lambda^z| \leq O\left(\frac{|z|}{\check{\kappa}_z}\right)$ then it suffices to bound thanks to (2.13): $\left\|\hat{\Lambda}^z - \mathbb{E}_{\mathcal{A}_Q}[\Lambda^z]\right\| \leq O\left(\frac{\kappa_z}{n}\right) \leq O\left(\frac{|z|}{n\check{\kappa}_z}\right)$. □

We can now prove the main result of this subsections that allows us to set that $\tilde{Q}^{\hat{\Lambda}^z}$ is a deterministic equivalent of Q^z (thanks to Lemma **??**).

Proposition 2.12. *Given $z \in \mathbb{C} \setminus S_{-0}^{\varepsilon}$:*

$$\left\|\tilde{Q}^{\hat{\Lambda}^z}\right\| \leq O(\kappa_z) \qquad \textit{and} \qquad \left\|\mathbb{E}_{\mathcal{A}_Q}[Q^z] - \tilde{Q}^{\hat{\Lambda}^z}\right\|_F \leq O\left(\frac{\kappa_z}{\sqrt{n}}\right).$$

To prove this proposition, we will bound the different elements of the decomposition (2.10). To bound ε_1, we will need the following lemma. The concentration is quit sharp since we have:

$$\min(\kappa_z, \check{\kappa}_z) = \kappa_z \check{\kappa}_z = \frac{|z|}{1+|z|}$$

Lemma 2.13. *Given $z \in \mathbb{C} \setminus S_{-0}^{\varepsilon}$, under $\mathcal{A}_Q$,*

$$Q^z X = X^T \check{Q}^z \mid \mathcal{A}_Q \propto \mathcal{E}_2\left(\frac{|z|}{1+|z|}\right)$$

and $\forall i \in [n]$, $\|\mathbb{E}_{\mathcal{A}_Q}[Q^z x_i]\| \leq O\left(\frac{|z|}{1+|z|}\right)$.

Proof. We follow the steps of the proof of Proposition 2.3. Depending on the sign of $p-n$, it is more convenient to work with the expression Q^zX (when $p \leq n$) or with $X^T\check{Q}^z$ (when $p \geq n$). We just treat here the case $p \leq n$ and therefore look at the variations of the mapping $\Psi : \mathcal{M}_{p,n} \to \mathcal{M}_{p,n}(\mathbb{C})$ defined as:

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

to show the concentration of $Q^zX = \Psi(X)$. For all $H, M \in \mathcal{M}_{n,p}^{\mathcal{A}_Q} \equiv X(\mathcal{A}_Q)$ (and with the notation $\Phi(M) = \left(I_p - \frac{MM^T}{zn}\right)^{-1}$ given in the proof of Proposition 2.3):

$$\begin{aligned}\|d\Psi_{|M} \cdot H\| &\leq \left\| \Psi(M) \frac{1}{nz}(MH^T + HM^T)\Psi(M)M \right\| + \|\Psi(M)H\| \\ &\leq O\left(\frac{|z|\|H\|}{(1+|z|)^2}\right) + O\left(\frac{|z|\|H\|}{1+|z|}\right) \leq O\left(\kappa_z\check{\kappa}_z\|H\|\right).\end{aligned}$$

Thus, under $\mathcal{A}_Q$, Ψ is $O(\kappa_z\check{\kappa}_z)$-Lipschitz (for the Frobenius norm) and therefore $Q^zX \propto \mathcal{E}_2(\kappa_z\check{\kappa}_z)$.

To control the norm of $\mathbb{E}_{\mathcal{A}_Q}[Q^zx_i]$, let us employ Schur identities (2.8) and bound for any deterministic $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:

$$\left|\mathbb{E}_{\mathcal{A}_Q}[u^TQ^zx_i]\right| = |z|\left|\mathbb{E}_{\mathcal{A}_Q}\left[\frac{u^TQ^z_{-i}x_i}{\Lambda^z_i}\right]\right| \leq |z|\mathbb{E}_{\mathcal{A}_Q}[|u^TQ^z_{-i}x_i|]\frac{\check{\kappa}_z}{|z|} \leq O(\kappa_z\check{\kappa}_z),$$

thanks to Lemmas 2.9 and 2.7. □

Proof of Proposition 2.12. Let us note for simplicity $\kappa_{\tilde{Q}} \equiv \|\tilde{Q}^{\hat{\Lambda}^z}\|$. Looking at decomposition (2.10) we can start with the bound:

$$\|\varepsilon_1\|_F = \left\|\frac{1}{zn}\mathbb{E}_{\mathcal{A}_Q}\left[Q^zX\left(\hat{\Lambda}^z - \Lambda^z\right)(\hat{\Lambda}^z)^{-1}X^T\right]\tilde{Q}^{\Lambda^z}\right\|_F \leq O\left(\frac{\kappa_{\tilde{Q}}\kappa_z^2\check{\kappa}_z^2}{|z|^2\sqrt{n}}\right) \leq O\left(\frac{\kappa_{\tilde{Q}}}{\sqrt{n}}\right)$$

obtained with the bound $\frac{1}{\Lambda^z} \leq O(\frac{\check{\kappa}_z}{|z|}) \leq O(1)$ given by Lemma 2.11 and applying Proposition **??** with the hypotheses:

- $X \mid \mathcal{A}_Q \propto \mathcal{E}_2$ and $\|\mathbb{E}_{\mathcal{A}_Q}[x_i]\| \leq O(1)$ Assumption 2,
- $Q^zX \mid \mathcal{A}_Q \propto \mathcal{E}_2(\kappa_z\check{\kappa}_z)$ and $\|\mathbb{E}_{\mathcal{A}_Q}[Q^zx_i]\| \leq O(\kappa_z\check{\kappa}_z)$ given by Lemma 2.13,
- $\Lambda^z \mid \mathcal{A}_Q \in \mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] \pm \mathcal{E}_2\left(\frac{\kappa_z}{\sqrt{n}}\right)$ in $(\mathcal{D}_n, \|\cdot\|)$ given by Lemma 2.5,
- $\|\mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] - \hat{\Lambda}^z\|_F \leq O(\kappa_z/\sqrt{n})$ thanks to Lemma 2.6

Second, for any matrix $A \in \mathcal{M}_p(\mathbb{C})$ satisfying $\|A\|_F \leq 1$, let us bound thanks to Cauchy-Schwarz inequality:

$$\begin{aligned}|\mathrm{Tr}(A\varepsilon_2)| &\leq \sqrt{\frac{1}{|z|^2n^2}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^zX|\hat{\Lambda}^z_n|^{-2}X^T\bar{Q}^z\bar{A}^T\right)\right]} \\ &\quad \cdot \sqrt{\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(\tilde{Q}^{\hat{\Lambda}^z}\Sigma_iQ^z_{-i}x_ix_i^TQ^z_{-i}\Sigma_i\bar{\tilde{Q}}^{\hat{\Lambda}^z}\right)\right]} \\ &\leq O\left(\frac{\kappa_z^4\check{\kappa}_z^4}{|z|^2}\sqrt{\frac{\|A\|_F^2}{n}\frac{\sup_{i\in[n]}\mathrm{Tr}\left(\Sigma_i^3\right)\kappa_{\tilde{Q}}^2}{n}}\right) \leq O\left(\frac{\kappa_{\tilde{Q}}}{\sqrt{n}}\right)\end{aligned}$$

thanks to the bounds provided by our assumptions, and Lemmas 2.2, 2.11 and 2.13.

Third, we bound easily with Lemma 2.8 the quantity:

$$\|\mathrm{Tr}(A\delta_1)\| \equiv \frac{1}{n}\sum_{i=1}^n \frac{1}{|\hat\Lambda_i^z|}\left\|\mathrm{Tr}\left(\tilde Q^{\hat\Lambda^z} A\mathbb{E}_{\mathcal{A}_Q}\left[Q_{-i}^z x_i x_i^T\right]\right)\right.$$
$$\left.- \mathrm{Tr}\left(\tilde Q^{\hat\Lambda^z} A\mathbb{E}_{\mathcal{A}_Q}\left[Q_{-i}^z\right]\mathbb{E}_{\mathcal{A}_Q}\left[x_i x_i^T\right]\right)\right\| \leq O\left(\frac{\kappa_{\tilde Q}}{\sqrt n}\right).$$

And we can bound $\|\mathrm{Tr}(A\delta_2)\| \leq O\left(\frac{\kappa_{\tilde Q}}{n}\right)$ since $\|\Sigma_i - \mathbb{E}_{\mathcal{A}_Q}[x_i x_i^T]\| \leq O(\frac{1}{n})$ (as explained in the proof of Lemma 2.10).

Taking the supremum on $A \in \mathcal{M}_{p,n}(\mathbb{C})$ and putting the bounds on $\|\varepsilon_1\|_F$, $\|\varepsilon_2\|_F$, $\|\delta_1\|_F$ and $\|\delta_2\|_F$ together, we obtain:

$$\left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right] - \tilde Q^{\hat\Lambda^z}\right\|_F \leq O\left(\frac{\kappa_{\tilde Q}}{\sqrt n}\right)$$

So, in particular $\kappa_{\tilde Q} \equiv \left\|\tilde Q^{\hat\Lambda^z}\right\| \leq \left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right]\right\| + O\left(\frac{\kappa_{\tilde Q}}{\sqrt n}\right)$, which implies that $\kappa_{\tilde Q} \leq O(\kappa_z)$ as $\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right]\|$ since $\frac{1}{\sqrt n} = o(1)$. We obtain then directly the second bound of the proposition. □

2.1.4 Definition of the second deterministic equivalent thanks to the semi-metric d_s

Proposition 2.12 slightly simplified the problem because while we initially had to estimate the expectation of the whole matrix Q^z, now, we just need to approach the diagonal matrix $\hat\Lambda^z \equiv \mathrm{Diag}(z - \frac{1}{n}\mathrm{Tr}\left(\Sigma_i Q_{-i}^z\right)])_{i\in[n]}$. One might be tempted to introduce from the pseudo identity (where $\tilde Q$ was defined in (2.7)):

$$\hat\Lambda_i^z = z - \frac{1}{n}\mathrm{Tr}(\Sigma_i \mathbb{E}_{\mathcal{A}_Q}[Q^z]) \approx z - \frac{1}{n}\mathrm{Tr}(\Sigma_i \tilde Q^{\hat\Lambda^z}) \tag{2.14}$$

a fixed point equation whose solution would be a natural estimate for $\hat\Lambda^z$. This equation is given in Theorem 2.18: we chose $\tilde\Lambda^z$ satisfying

$$\tilde\Lambda_i^z = z - \frac{1}{n}\mathrm{Tr}\left(\Sigma_i \tilde Q^{\tilde\Lambda^z}\right).$$

We are now going to prove that $\tilde\Lambda^z$ is well defined for any $z \in \mathbb{H}$ (where we recall that $\mathbb{H} \equiv \{z \in \mathbb{C}, \Im(z) > 0\}$) to prove Theorem 2.18. Introducing the mapping:

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

we want to show that $\mathcal{I}^z$ admits a unique fixed point. For that purpose, we are going to employ the Theorem 1.17 on the mapping $\mathcal{I}^z$ ($z \in \mathbb{H}$), that is, as we will see in Proposition 2.15 below, contractive for semi-metric d_s presented in last chapter and defined for any $D, D' \in \mathcal{D}_n(\mathbb{H})$ as:

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

We first need to restrict our study on a subset of $\mathcal{D}_n(\mathbb{H})$:

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

Lemma 2.14. *For any $z \in \mathbb{H}$, $\mathcal{I}^z(\mathcal{D}_{\mathcal{I}^z}) \subset \mathcal{D}_{\mathcal{I}^z}$.*

Proof. Considering $z \in \mathbb{H}$, and $L \in \mathcal{D}_{\mathcal{I}^z}$ and $i \in [n]$, the decomposition $\tilde{Q}^L = \left(I_p - \frac{1}{n}\sum_{i=1}^n \frac{\Re(L_j)\Sigma_i}{|L_j|^2} - \frac{i}{n}\sum_{i=1}^n \frac{\Im(L_j)\Sigma_i}{|L_j|^2}\right)^{-1}$ allows us to set thanks to the resolvent identity (2.9):

$$\begin{aligned}\Im\left(\mathcal{I}^z\left(L\right)_i\right) &= \Im(z) + \frac{1}{2in}\operatorname{Tr}\left(\Sigma_i\left(\tilde{Q}^L + \bar{\tilde{Q}}^L\right)\right)\\ &= \Im(z) + \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\sum_{i=1}^n\frac{\Im(L_j)\Sigma_i}{|L_j|^2}\bar{\tilde{Q}}^L\right) > 0\end{aligned}$$

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

$$\Im\left(\mathcal{I}^z\left(L\right)_i/z\right) = \frac{1}{n|z|^2}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\left(\Im(z) + \sum_{i=1}^n\frac{\Im(L_j/z)\Sigma_i}{|L_j/z|^2}\right)\bar{\tilde{Q}}^L\right) > 0$$

□

Let us now express the Lipschitz parameter of $\mathcal{I}^z$ for the semi metric d_s.

Proposition 2.15. *For any $z \in \mathbb{H}$, the mapping $\mathcal{I}^z$ is 1-Lipschitz for the semi-metric d_s on $\mathcal{D}_{\mathcal{I}^z}$ and satisfies for any $L, L' \in \mathcal{D}_{\mathcal{I}^z}$:*

$$d_s(\mathcal{I}^z(L), \mathcal{I}^z(L')) \leq \sqrt{(1-\phi(z,L))(1-\phi(z,L'))}\,d_s(L,L'),$$

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

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

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

$$\begin{aligned}|\mathcal{I}^z(L)_i - \mathcal{I}^z(L')_i| &= \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\left(\frac{1}{n}\sum_{j=1}^n\frac{L_j - L'_j}{L_jL'_j}\Sigma_j\right)\bar{\tilde{Q}}^{L'}\right)\\ &= \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\left(\frac{1}{n}\sum_{j=1}^n\frac{L_j - L'_j}{\sqrt{\Im(L_j)\Im(L'_j)}}\frac{\sqrt{\Im(L_j)\Im(L'_j)}}{L_jL'_j}\Sigma_j\right)\bar{\tilde{Q}}^{L'}\right)\\ &\leq d_s(L,L')\sqrt{\frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\left(\frac{1}{n}\sum_{i=1}^n\frac{\Im(L_j)\Sigma_i}{|L_j|^2}\right)\bar{\tilde{Q}}^L\right)}\\ &\qquad\cdot\sqrt{\frac{1}{n}\operatorname{Tr}\left(\Sigma_i\bar{\tilde{Q}}^{L'}\left(\frac{1}{n}\sum_{i=1}^n\frac{\Im(L'_j)\Sigma_i}{|L'_j|^2}\right)\tilde{Q}^{L'}\right)}\\ &\leq d_s(L,L')\sqrt{\left(\Im(\mathcal{I}^z(L)_i) - \Im(z)\right)\left(\Im(\mathcal{I}^z(L')_i) - \Im(z)\right)},\end{aligned}\tag{2.15}$$

thanks to Cauchy-Schwarz inequality and the identity

$$0 \leq \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^L\left(\sum_{i=1}^n\frac{\Im(L'_j)\Sigma_i}{|L'_j|^2}\right)\bar{\tilde{Q}}^z(L)\right) = \Im(\mathcal{I}^z(L)_i) - \Im(z)$$

issued from the proof of Lemma 2.14. Dividing both sides of (2.15) by $\sqrt{\Im(\mathcal{I}^z(L)_i)\Im(\mathcal{I}^z(L')_i)}$, we retrieve the wanted Lipschitz parameter.

□

The contractivity of I^z is not fully stated in the previous proposition because, the Lipschitz parameter depends on the values of L, L' and we want a bound uniform on $\mathcal{D}_{I^z}$. This will be done thanks to the two lemmas.

Lemma 2.16 (Commutation between inversion and modulus of matrices). *Given an invertible matrix $M \in \mathcal{M}_p$, $|M^{-1}| = |M|^{-1}$ and for any $K > 0$:*

$$\Im M^{-1} \geq K I_p \text{ or } \Re M^{-1} \geq K I_p \qquad \Longrightarrow \qquad |M| \leq \frac{1}{K} I_p.$$

Proof. We have the identity:

$$|M^{-1}|^2 = M^{-1}\bar{M}^{-1} = (\bar{M}M)^{-1} = (|M|^2)^{-1}.$$

then we take the square root on both sides to obtain the first identity (recall that the modulus of a matrix is a non negative hermitian matrix). Now let us assume that $\Im(M^{-1}) \geq K$, we know that:

$$\begin{aligned} |M^{-1}|^2 = {} & \Im(M^{-1})\Im(M^{-1})^T + \Re(M^{-1})\Re(M^{-1})^T \\ & - i\Re(M^{-1})\Im(M^{-1})^T + i\Im(M^{-1})\Re(M^{-1})^T, \end{aligned}$$

is a nonnegative hermitian matrix satisfying for all $x \in \mathbb{C}^p$ (the cross terms cancel):

$$\begin{aligned} \bar{x}^T |M^{-1}|^2 x &= \bar{x}^T \Im(M^{-1})\Im(M^{-1})^T x + \bar{x}^T \Re(M^{-1})\Re(M^{-1})^T x \\ &\geq \bar{x}^T \Im(M^{-1})\Im(M^{-1})^T x \geq K^2 \|x\|^2. \end{aligned}$$

Thus the lower eigen value of $|M^{-1}| = |M|^{-1}$ is bigger than K and therefore $|M| \leq \frac{1}{K} I_p$. □

Lemma 2.17. *Given $L \in \mathcal{D}_{I^z}$, we can bound:*

$$\Im(z) I_n \leq |\mathcal{I}^z(L)| \leq O\left(|z| + \frac{|z|}{\Im(z)}\right) I_n$$

and:

$$O\left(\frac{\Im(z)}{1+\Im(z)}\right) I_p \leq \left|Q^{I^z(L)}\right| \leq \frac{|z| I_p}{\Im(z)}.$$

Proof. The lower bound of $\mathcal{I}^z(L)$ is immediate (see the proof of Lemma 2.14). If $L \in \mathcal{D}_{I^z}$, then we know that $L/z \in \mathcal{D}_n(\mathbb{H})$, and therefore, noting that:

$$\Im\left((\tilde{Q}^L/z)^{-1}\right) = \Im\left(zI_p - \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{L_i/z}\right) = \Im(z) I_p + \frac{1}{n}\sum_{i=1}^n \frac{\Im(L_i/z)\Sigma_i}{|L_i/z|^2} \geq \Im(z) I_p,$$

we can deduce from Lemma 2.16 that $|\tilde{Q}^L/z| \leq \frac{1}{\Im(z)}$ and thus $\|\tilde{Q}^L\| \leq \frac{|z|}{\Im(z)}$ which gives us directly the upper bound on $\mathcal{I}^z(L)$ since $\mathrm{Tr}(\Sigma_i) \leq O\left(|z| + \frac{|z|}{\Im(z)}\right)$.

Finally, we can bound:

$$\left\| I_n - \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\mathcal{I}^z(L)_i} \right\| \leq 1 + \frac{1}{n}\sum_{i=1}^n \frac{\|\Sigma_i\|}{|\Im(\mathcal{I}^z(L)_i)|} \leq 1 + O\left(\frac{1}{\Im(z)}\right),$$

which provides the lower bound on $|\tilde{Q}^{\mathcal{I}^z(L)}|$ since $O\left(\frac{\Im(z)}{1+\Im(z)}\right) \leq \frac{1}{1+O(\frac{1}{\Im(z)})}$. □

Theorem 2.18. *Given n nonnegative symmetric matrices $\Sigma_1, \ldots, \Sigma_n \in \mathcal{M}_p$, for all $z \in \mathbb{H}$, the equation:*

$$\forall i \in [n], L_i = z - \frac{1}{n} \mathrm{Tr}\left(\Sigma_i \left(I_p - \frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{L_i}\right)^{-1}\right) \tag{2.16}$$

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

Proof. On the domain $\mathcal{I}^z(\mathcal{D}_{\mathcal{I}^z})$, the mapping $\mathcal{I}^z$ is bounded and contracting for the semi-metric d_s thanks to Proposition 2.15 and Lemma 2.17. The hypotheses of Theorem 1.17 are thus satisfied, and we know that there exists a unique diagonal matrix $\tilde{\Lambda}^z \in I^z(\mathcal{D}_{I^z})$ such that $I^z(\tilde{\Lambda}^z) = \tilde{\Lambda}^z$. There can not exist a second diagonal matrix $\Lambda' \in \mathcal{D}_n(\mathbb{H})$ such that $\Lambda' = I^z(\Lambda')$ because then Proposition 2.15 (true on the whole domain $\mathcal{D}_n(\mathbb{H})$) would imply that $d_s(\Lambda', \tilde{\Lambda}^z) < d_s(\Lambda', \tilde{\Lambda}^z)$. □

We end this section with an interesting result on $\tilde{\Lambda}^z$ that will however not have any use for our main results.

Lemma 2.19. $\sup_{i\in[n]} |\tilde{\Lambda}_i^z| \leq O(1 + |z|)$.

Proof. If we assume that $\forall i \in [n], |\tilde{\Lambda}_i^z| \geq 2\nu$, then we deduce that $\frac{1}{n}\sum_{i=1}^n \frac{1}{|\tilde{\Lambda}_i^z|} \leq \frac{1}{2\nu}$ and $|\frac{1}{n}\sum_{i=1}^n \frac{\Sigma_i}{\tilde{\Lambda}_i^z}| \leq \frac{1}{2}$, and therefore, $|\tilde{Q}^{\tilde{\Lambda}^z}| \leq 2$. As a consequence, $\forall i \in [n]$:

$$\left|\tilde{\Lambda}_i^z\right| \leq |z| + \frac{1}{n}\operatorname{Tr}\left(\Sigma_i |\tilde{Q}^{\tilde{\Lambda}^z}|\right) \leq |z| + \frac{2p\nu}{n}$$

We can conclude that:

$$\sup_{i\in[n]} |\tilde{\Lambda}_i^z| \leq \max\left(\frac{\nu}{2}, |z| + \frac{2p\nu}{n}\right) \leq O(1+|z|).$$

□

2.1.5 Convergence of $\hat{\Lambda}^z$ towards $\tilde{\Lambda}^z$

To show the convergence of $\hat{\Lambda}^z$ towards $\tilde{\Lambda}^z = I^z(\tilde{\Lambda}^z)$, we need Proposition 1.23 bounding the distance to a fixed point of a contracting mapping for the semi-metric d_s. This sets what we call the stability of the equation. First allowing us to bound $\|\hat{\Lambda}^z - \tilde{\Lambda}^z\|$, it will then be employed to show the continuity of $z \mapsto \tilde{\Lambda}^z$. In the former application, the convergence parameter is n, while in the latter application it is a parameter $t \in \mathbb{C}$ in the neighbourhood of 0.

To be employ Proposition 1.23 on the matrices $\tilde{\Gamma}^n = \tilde{\Lambda}^z$ and $\Gamma^n = \hat{\Lambda}^z$ and on the mapping $f^n = \mathcal{I}^z$, we first need to set the following proposition. Unfortunately, we need to assume here that $d(z, S^\varepsilon) \geq O(1)$ (and not $d(z, S_{-0}^\varepsilon) \geq O(1)$).

Proposition 2.20. *Given $z \in \mathbb{C} \setminus S^\varepsilon$:*

$$d_s\left(\Im(\mathcal{I}^z(\hat{\Lambda}^z)), \Im(\hat{\Lambda}^z)\right) \leq O\left(\frac{1}{n}\right)$$

Remark 2.21. *The bound $O(\frac{1}{n})$ comes from the fact that for any $z \in \mathbb{C} \setminus S^\varepsilon$, if $n < p$ then $0 \in S$ and $|z| \geq \varepsilon$ and if $n \geq p$ then $\kappa_z = \frac{|z|}{1+|z|}$. Therefore, in all cases $\frac{\kappa_z}{|z|} \leq O(1)$.*

To prove Proposition 2.20 but also to show later that the mapping $\mathcal{I}^z$ is contracting, we will need:

Lemma 2.22. *Given $z \in \mathbb{C} \setminus S_{-0}^\varepsilon$:*

$$\Im(z) \leq \inf_{i\in[n]} \Im\left(\hat{\Lambda}_i^z\right) \leq \sup_{i\in[n]} \Im\left(\hat{\Lambda}_i^z\right) \leq O\left(\Im(z)\right)$$

Proof. The lower bound is obvious since for all $i \in [n]$:

$$\Im(\hat{\Lambda}_i^z) = \Im(z) + \frac{\Im(z)}{|z|^2 n^2}\mathbb{E}_{\mathcal{A}_Q}[\operatorname{Tr}\left(Q_{-i}^z X_{-i} X_{-i}^T \bar{Q}_{-i}^z \Sigma_i\right)] \geq \Im(z).$$

The upper bound is obtained thanks to the bound, valid under $\mathcal{A}_Q$, $Q_{-i}^z X_{-i}/|z|\sqrt{n} \leq O(1)$ provided by Lemma 2.13.

□

Proof of Proposition 2.20. We can bound thanks to Lemma 2.22:

$$d_s\left(\Im(\mathcal{I}^z(\hat\Lambda^z)),\Im(\hat\Lambda^z)\right)=\sup_{1\le i\le n}\left|\frac{\frac{1}{n}\operatorname{Tr}\left(\Sigma_i\Im\left(\tilde Q^{\hat\Lambda^z}-\mathbb{E}_{\mathcal{A}_Q}[Q^z]\right)\right)}{\sqrt{\Im(\hat\Lambda^z_i)\Im(\mathcal{I}^z(\hat\Lambda^z)_i)}}\right|\le O\left(\frac{1}{\sqrt n}\left\|\frac{\Im(\tilde Q^{\hat\Lambda^z})}{\Im(z)}-\frac{\Im\left(\mathbb{E}_{\mathcal{A}_Q}[Q^z]\right)}{\Im(z)}\right\|_F\right),$$

since $\|\Sigma_i\|_F\le O(\sqrt n)$. The identity $\frac{1}{\bar z n}XX^T\bar Q^z=\bar Q^z-I_n$ allows us to write:

$$\frac{\Im(\mathbb{E}_{\mathcal{A}_Q}[Q^z])}{\Im(z)}=\frac{1}{\Im(z)}\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\left(\frac{\Im(z)}{n|z|^2}XX^T\right)\bar Q^z\right]=\frac{1}{z}\mathbb{E}_{\mathcal{A}_Q}[|Q^z|^2-Q^z](=\frac{1}{\bar z}\mathbb{E}_{\mathcal{A}_Q}[|Q^z|^2-\bar Q^z]).\tag{2.17}$$

We already know how to estimate $\mathbb{E}_{\mathcal{A}_Q}[Q^z]$ thanks to Proposition 2.12, we are thus left to estimate $\mathbb{E}_{\mathcal{A}_Q}[|Q^z|^2]$. We do not give the full justifications that are closely similar to those presented in the proof of Proposition 2.12 – mainly application of Schur identities (2.8), Proposition **??** and Lemmas 2.5, 2.10. To complete this estimation, we consider a deterministic matrix $A\in\mathcal{M}_p$, and we estimate:

$$\begin{aligned}
&\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\bar Q^z\left(Q^z-\tilde Q^{\hat\Lambda^z})\right)\right]\right)\\
&\quad=\frac{1}{n}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\bar Q^zQ^z\left(\frac{x_ix_i^T}{z}-\frac{\Sigma_i}{\hat\Lambda^z_i}\right)\tilde Q^{\hat\Lambda^z}\right]\right)\\
&\quad=\frac{1}{n}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^zQ^z_{-i}x_ix_i^T\tilde Q^{\hat\Lambda^z}}{\Lambda^z_i}\right]\right)-\frac{1}{n}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^zQ^z\Sigma_i\tilde Q^{\hat\Lambda^z}}{\hat\Lambda^z_i}\right]\right)\\
&\quad=\frac{1}{n}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^z_{-i}Q^z_{-i}x_ix_i^T\tilde Q^{\hat\Lambda^z}}{\hat\Lambda^z_i}\right]\right)-\frac{1}{n^2}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^zx_ix_i^T|Q^z_{-i}|^2x_ix_i^T\tilde Q^{\hat\Lambda^z}}{z\hat\Lambda^z_i}\right]\right)\\
&\qquad-\frac{1}{n}\sum_{i=1}^n\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^z_{-i}Q^z_{-i}\Sigma_i\tilde Q^{\hat\Lambda^z}}{\hat\Lambda^z_i}\right]\right)+O\left(\frac{\kappa_z}{\sqrt n}\right)\\
&\quad=-\mathbb{E}_{\mathcal{A}_Q}\left[\frac{1}{n}\operatorname{Tr}\left(A\bar Q^zX\frac{\Delta^z}{z\hat\Lambda^z_i}X^T\tilde Q^{\hat\Lambda^z}\right)\right]+O\left(\frac{\kappa_z}{\sqrt n}\right)
\end{aligned}$$

with the introduction of the notation:

$$\Delta^z\equiv\operatorname{Diag}_{i\in[n]}\left(\frac{1}{n}x_i^T|Q^z_{-i}|^2x_i\right).$$

The random diagonal matrix Δ^z being a $O(\kappa_z^2/\sqrt n)$ Lipschitz transformation of X for the spectral norm $\|\cdot\|$ on $\mathcal{D}_n$, we know that $\Delta^z\mid\mathcal{A}_Q\in\hat\Delta^z\pm\mathcal{E}_2(\kappa_z/\sqrt n)$, where we noted $\hat\Delta^z\equiv\mathbb{E}_{\mathcal{A}_Q}[\Delta^z]$. One can then pursue the estimation, thanks again to Proposition **??**:

$$\begin{aligned}
\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\bar Q^z\left(Q^z-\tilde Q^{\hat\Lambda^z})\right)\right]\right)&=\frac{1}{n}\sum_{i=1}^n\frac{\hat\Delta^z_i}{\hat\Lambda^z_i}\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\frac{\bar Q^z_{-i}x_ix_i^T\tilde Q^{\hat\Lambda^z}}{\bar\Lambda^z_i}\right]\right)+O\left(\frac{\kappa_z}{\sqrt n}\right)\\
&=\frac{1}{n}\sum_{i=1}^n\frac{\hat\Delta^z_i}{\left|\hat\Lambda^z_i\right|^2}\operatorname{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\bar Q^z_{-i}\Sigma_i\tilde Q^{\hat\Lambda^z}\right]\right)+O\left(\frac{\kappa_z}{\sqrt n}\right)
\end{aligned}$$

Now, taking the expectation under $\mathcal{A}_Q$ on the identity valid for any $i\in[n]$:

$$\Im(\Lambda^z_i)=\Im(z)-\frac{1}{n}\Im(x_i^TQ^z_{-i}x_i)=\Im(z)\left(1+\frac{1}{nz}\left(x_i^T|Q^z_{-i}|^2x_i-x_i^TQ^z_{-i}x_i\right)\right)=\frac{\Im(z)}{z}\left(\Delta^z_i+\Lambda^z_i\right),\tag{2.18}$$

we deduce that $\hat\Delta^z = \frac{z}{\Im(z)}\Im(\hat\Lambda^z) - \hat\Lambda^z + O(\kappa_z/n)$. Therefore, with the identity $\mathbb{E}_{\mathcal{A}_Q}[\bar Q^z \tilde Q^{\hat\Lambda^z}] = |\tilde Q^{\hat\Lambda^z}|^2 + O_{\|\cdot\|_F}(\kappa_z/\sqrt{n})$, we can estimate (see the proof of Lemma 2.14 for the identification of $\Im(\mathrm{Tr}(\tilde Q^{\hat\lambda^z})))$:

$$\begin{aligned}\mathrm{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[|Q^z|^2\right]\right) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{z}{\Im(z)}\frac{\Im(\hat\Lambda_i^z)}{\left|\hat\Lambda_i^z\right|^2} - \frac{1}{\bar{\hat\Lambda}_i^z}\right)\mathrm{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[\bar Q_{-i}^z\Sigma_i\tilde Q^{\hat\Lambda^z}\right]\right) \\ &\quad + \mathrm{Tr}\left(A|\tilde Q^{\hat\Lambda^z}|^2\right) + O\left(\frac{\kappa_z}{\sqrt{n}}\right) \\ &= \frac{z}{\Im(z)}\Im\left(\mathrm{Tr}\left(A\tilde Q^{\hat\Lambda^z}\right)\right) + \frac{1}{n}\mathrm{Tr}\left(A\tilde Q^{\hat\Lambda^z}\right) + O\left(\frac{\kappa_z}{\sqrt{n}}\right)\end{aligned}$$

and therefore, we can deduce from Proposition 2.12 and (2.17):

$$\frac{\Im\left(\mathrm{Tr}\left(AQ^z\right)\right)}{\Im(z)} = \frac{1}{z}\mathrm{Tr}\left(A\mathbb{E}_{\mathcal{A}_Q}\left[|Q^z|^2 - Q^z\right]\right) = \frac{\Im\left(\mathrm{Tr}\left(A\tilde Q^{\hat\Lambda^z}\right)\right)}{\Im(z)} + O\left(\frac{\kappa_z}{|z|\sqrt{n}}\right).$$

we can bound thanks to identity 2.17 and inequality $\frac{\kappa_z}{|z|} \leq O(1)$ given by Remark 2.21:

$$\left\|\frac{\Im(\mathbb{E}_{\mathcal{A}_Q}[Q^z]) - \Im(\tilde Q^{\hat\Lambda^z})}{\Im(z)}\right\|_F \leq O\left(\frac{\kappa_z}{|z|\sqrt{n}}\right) \leq O\left(\frac{1}{\sqrt{n}}\right),$$

and we retrieve the result of the proposition. □

We now have all the elements to show the convergence of $\hat\Lambda^z$ to $\tilde\Lambda^z$. Here again, we need to assume as in Proposition 2.20 that $z \in \mathbb{C}\setminus S^\varepsilon$.

Proposition 2.23. *For any $z \in \mathbb{H}$ such that $d(z,S) \geq \varepsilon$:*

$$\|\hat\Lambda^z - \tilde\Lambda^z\| \leq O\left(\frac{\kappa_z}{n}\right) \qquad \text{and} \qquad O\left(\frac{|z|}{\breve\kappa_z}\right) \leq |\tilde\Lambda^z| \leq O\left(\frac{|z|}{\breve\kappa_z}\right)$$

Proof. We already know from Proposition 2.20 that $d_s\left(\Im(\mathcal{I}^z(\hat\Lambda^z)), \Im(\hat\Lambda^z)\right) \leq o(1)$. Besides, the Lipschitz parameter λ of $\mathcal{I}^z$ on the set $\{\tilde\Lambda^z, \hat\Lambda^z, n \in \mathbb{N}\}$ is such that $1-\lambda \geq O(1)$. Recall indeed from Proposition 2.15 that:

$$\lambda \leq \sqrt{\left(1 - \frac{\Im(z)}{\sup_{i\in[n]}\Im(\tilde\Lambda_i^z)}\right)\left(1 - \frac{\Im(z)}{\sup_{i\in[n]}\Im(\hat\Lambda_i^z)}\right)} \leq \sqrt{1-O(1)} \leq 1 - O(1),$$

thanks to Lemma 2.22. Therefore, we can employ Proposition 1.23 to set that:

$$\left\|\frac{\hat\Lambda^z - \tilde\Lambda^z}{\sqrt{\Im(\hat\Lambda^z)\Im(\tilde\Lambda^z)}}\right\| = d_s(\hat\Lambda^z, \tilde\Lambda^z) \leq O\left(\left\|\frac{\hat\Lambda^z - \mathcal{I}^z(\hat\Lambda^z)}{\sqrt{\Im(\hat\Lambda^z)\Im(\tilde\Lambda^z)}}\right\|\right),$$

which implies thanks to Lemma 2.22 that:

$$\left\|\hat\Lambda^z - \tilde\Lambda^z\right\| \leq O\left(\sqrt{\frac{\sup_{i\in[n]}\Im(\tilde\Lambda_i^z)}{\inf_{i\in[n]}\Im(\tilde\Lambda_i^z)}}\left\|\hat\Lambda^z - \mathcal{I}^z(\hat\Lambda^z)\right\|\right).$$

We reach here the only point of the whole proof where we will employ Assumption 5. It is to set that:

$$\begin{aligned}\inf_{i\in[n]}\Im(\tilde\Lambda_i^z) &= \Im(z) + \inf_{i\in[n]}\sum_{j=1}^n \frac{\Im(\tilde\Lambda_j^z)}{n|\tilde\Lambda_j^z|^2}\mathrm{Tr}\left(\Sigma_i\tilde Q^{\tilde\Lambda^z}\Sigma_j\bar{\tilde Q}^{\tilde\Lambda^z}\right) \\ &\geq \Im(z) + \sum_{j=1}^n \frac{\Im(\tilde\Lambda_j)}{n^z|\tilde\Lambda_j|^2}{}^z\, O\left(\mathrm{Tr}\left(\tilde Q^{\tilde\Lambda^z}\Sigma_j\bar{\tilde Q}^{\tilde\Lambda^z}\right)\right) \geq O\left(\sup_{i\in[n]}\Im(\tilde\Lambda_i^z)\right)\end{aligned}$$

As a conclusion:

$$\left\|\hat{\Lambda}^z - \tilde{\Lambda}^z\right\| \leq O\left(\left\|\hat{\Lambda}^z - \mathcal{I}^z(\hat{\Lambda}^z)\right\|\right) \leq O\left(\frac{1}{\sqrt{n}}\left\|\hat{Q} - \tilde{Q}^{\hat{\Lambda}^z}\right\|\right) \leq O\left(\frac{\kappa_z}{n}\right).$$

we can further deduce that Λ^z and $\tilde{\Lambda}^z$ have the same upper and lower bound of order $O(|z|/\check{\kappa}_z)$ since $\kappa_z \leq O(|z|/\check{\kappa}_z)$. □

2.1.6 Concentration and final estimation of the resolvent

The estimation of Q^z is a simple consequence of the convergence of $\hat{\Lambda}^z$ towards $\tilde{\Lambda}^z$.

Corollary 2.24. *For any $z \in \mathbb{C} \setminus S_\varepsilon$, $\|\tilde{Q}^{\tilde{\Lambda}^z}\| \leq O(\kappa_z)$ and:*

$$\left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right] - \tilde{Q}^{\tilde{\Lambda}^z}\right\|_F \leq O\left(\frac{\kappa_z}{\sqrt{n}}\right)$$

Proof. We already know from Proposition 2.12 that $\left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right] - \tilde{Q}^{\hat{\Lambda}^z}\right\|_F \leq O(\kappa_z/\sqrt{n})$, thus we are left to bound:

$$\begin{aligned}
\left\|\tilde{Q}^{\hat{\Lambda}^z} - \tilde{Q}^{\tilde{\Lambda}^z}\right\|_F &\leq \left\|\tilde{Q}^{\hat{\Lambda}^z}\left(\frac{1}{n}\sum_{i=1}^n \frac{\hat{\Lambda}_i^z - \tilde{\Lambda}_i^z}{\hat{\Lambda}_i^z\tilde{\Lambda}_i^z}\Sigma_i\right)\tilde{Q}^{\tilde{\Lambda}^z}\right\|_F \\
&\leq \sup_{i\in[n]}\left|\frac{\hat{\Lambda}_i^z - \tilde{\Lambda}_i^z}{\hat{\Lambda}_i^z\tilde{\Lambda}_i^z}\right| \|\Sigma_i\|_F \left\|\tilde{Q}^{\hat{\Lambda}^z}\right\|\left\|\tilde{Q}^{\tilde{\Lambda}^z}\right\| \\
&\leq O\left(\frac{\kappa_z^3\check{\kappa}_z^2\sqrt{p}}{|z|^2 n}\right) \leq O\left(\frac{\kappa_z}{\sqrt{n}}\right)
\end{aligned}$$

(thanks to Lemma 2.7, 2.2, 2.19 and Proposition 2.23) We can then deduce that:

$$\left\|\tilde{Q}^{\tilde{\Lambda}^z}\right\|_F \leq \|\tilde{Q}^{\hat{\Lambda}^z} - \tilde{Q}^{\tilde{\Lambda}^z}\|_F + \|\tilde{Q}^{\hat{\Lambda}^z}\|_F \leq O(\kappa_z).$$

□

The concentration of Q^z nattturally implies the concentration of $R : z \mapsto \frac{1}{z}Q^z$ around $\tilde{R} : z \mapsto \frac{1}{z}\tilde{Q}^{\tilde{\Lambda}^z}$ that for integration purpose, we set with the semi-norm $\|\cdot\|_{F,S^\varepsilon}$, defined for every $f \in \mathcal{F}(\mathbb{C}, \mathcal{M}_p)$ as:

$$\|f\|_{F,S^\varepsilon} = \sup_{z\in\mathbb{C}\setminus S^\varepsilon}\|f(z)\|_F.$$

Theorem 2.25. *$R \mid \mathcal{A}_Q \in \tilde{R} \pm \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{M}_{p,n}^{\mathbb{C}\setminus S^\varepsilon}, \|f\|_{F,S^\varepsilon})$.*

Proof of Theorem 2.25 . We saw in the proof of Proposition 2.3 that the mapping $R \in (\mathcal{F}(\mathbb{C}\setminus S^\varepsilon, \mathcal{M}_p), \|\cdot\|_{F,S^\varepsilon})$ defined, under $\mathcal{A}_Q$, for any $z \in \mathbb{C}\setminus S^\varepsilon$ as $R(z) = -\frac{1}{z}Q^z$ has a Lipschitz parameter bounded by:

$$\sup_{z\in\mathbb{C}\setminus S^\varepsilon} O\left(\frac{\kappa_z}{|z|\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right)$$

Thanks to the bound $\frac{\kappa_z}{|z|} \leq O(1)$ justified in Remark 2.21. As a $O(1/\sqrt{n})$-Lipschitz transformation of $X \propto \mathcal{E}_2$, $(R \mid \mathcal{A}_Q) \propto \mathcal{E}_2(1/\sqrt{n})$, we can then conclude thanks to Corollary 2.24 with the bound:

$$\begin{aligned}
\left\|\mathbb{E}_{\mathcal{A}_Q}[R] - \tilde{R}\right\|_{F,S^\varepsilon} &\leq \sup_{z\in\mathbb{C}\setminus S^\varepsilon}\frac{1}{|z|}\left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right] - \tilde{Q}^{\tilde{\Lambda}^z}\right\| \\
&\leq \sup_{z\in\mathbb{C}\setminus S^\varepsilon} O\left(\frac{\kappa_z}{\sqrt{n}|z|}\right) \leq O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

□

The projections on deterministic vectors provide us with good estimates on isolated eigenvectors, but a concentration in spectral norm would have been sufficient for this kind of result. A key consequence of Theorem 2.25 lies in the accurate estimates of projections on high dimensional subspaces $F \subset \mathbb{R}^p$ it provides; this is shown[6] in Figure 2.1 that depicts some of these projections with increasing numbers of classes[7]. Given $k \in \mathbb{N}$, we consider $B \equiv \{\theta_1, \ldots, \theta_k\} \subset \mathbb{R}$, a (random) subset of k eigenvalues of $\frac{1}{n}XX^T$, E_B the eigenspace associated to those eigenvalues and Π_B and Π_F, respectively the orthogonal projection on E_B and F. If one can construct[8] a deterministic path γ such that $\{\mathbb{E}[\theta_i], i \in [k]\}^\varepsilon \subset \gamma \subset \mathbb{C} \setminus S^\varepsilon_{-0}$ then we can bound (since $\|\Pi_F\|_F = \sqrt{\dim(F)} \leq O(\sqrt{p})$):

$$\mathbb{P}\left(\left|\frac{1}{p}\operatorname{Tr}(\Pi_F \Pi_A) - \frac{1}{2ip\pi}\oint_\gamma \operatorname{Tr}(\Pi_F \tilde{R}(z))dz\right| \geq t\right) \leq Ce^{-cnpt^2} + Ce^{-cn},$$

Although it is not particularly needed for practical use, we are now going to show that the mapping $\tilde{g} : z \to \frac{1}{p}\operatorname{Tr}(\tilde{R}(z))$ is a Stieltjes transform that converges towards $\tilde{g}$ on $\mathbb{C} \setminus S^\varepsilon$.

2.1.7 Approach with the Stieltjes formalism

We present here some arguably well known results (see **?** for instance) about the Stieltjes transform of the eigen value distribution that allows to get some interesting inferences about its support. We start with an interesting identity that gives a direct link between the Stieltjes transforms g and $\tilde{g}$ with the diagonal matrix Λ^z and $\tilde{\Lambda}^z$. From the equality $Q^z - \frac{1}{nz}XX^TQ^z = I_p$, and the Schur identities (2.8) we can deduce that:

$$\begin{aligned} g(z) &= -\frac{1}{pz}\operatorname{Tr}(Q^z) = -\frac{1}{z} - \frac{1}{npz^2}\sum_{i=1}^n x_i^T Q^z x_i \\ &= -\frac{1}{z} - \frac{1}{npz}\sum_{i=1}^n \frac{x_i^T Q_{-i} x_i}{\Lambda_i^z} = \frac{1}{z}\left(\frac{n}{p} - 1\right) - \frac{1}{p}\sum_{i=1}^n \frac{1}{\Lambda_i^z} = g^{\Lambda^z}(z), \end{aligned}$$

with the notation g^{D^z}, defined for any mapping $D : \mathbb{H} \ni z \mapsto D^z \in \mathcal{D}_n(\mathbb{H})$ as[9]:

$$g^{D^z} : z \longmapsto \frac{1}{z}\left(\frac{n}{p} - 1\right) - \frac{1}{p}\sum_{i=1}^n \frac{1}{D^z}.$$

Interestingly enough, if we denote $\tilde{g} \equiv g^{\tilde{\Lambda}}$, then one can rapidly check that we have the equality $\tilde{g} = -\frac{1}{pz}\operatorname{Tr}(\tilde{Q}^{\tilde{\Lambda}^z})$. To show that $\tilde{g}$ is a Stieltjes transform, we will employ the following well known theorem that can be found for instance in **?**:

Theorem 2.26. *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 two reciprocal formulas:*

- $f(z) = \int \frac{\mu(d\lambda)}{\lambda - z}$,
- *for any continuous point[10] $a < b$: $\mu([a,b]) = \lim_{y \to 0^+} \frac{1}{\pi}\int_a^b \Im(f(x+iy))dx$.*

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 $\mathbb{C} \setminus (\mathbb{R}_+ \cup \{0\})$.

---

[6]It must be noted that the setting of Figure 2.1 does not exactly fall under our hypotheses (since here $\|\mathbb{E}[x_i]\| \geq O(\sqrt{p})$), as the amplitude of the signals must be sufficiently large for the resulting eigenvalues to isolate from the bulk of the distribution when the number of classes is high ($\sqrt{p} \approx 14$ is not so large). However, even in this extreme setting the prediction are good.

[7]The number of classes is the number of different distributions that can follow the column vectors of X

[8]The possibility to construct such a path and the condition on B for its existence, related to the notion of clusters, is a very interesting question that we do not address in this study.

[9]The Steiltjes transform of the spectral distribution of $\frac{1}{n}X^TX$ is $\check{g} = \check{g}^{\Lambda^z}$, where for all $D : \mathbb{H} \ni z \mapsto D^z \in \mathcal{D}_n(\mathbb{H})$, $\check{g}^{D^z} : z \longmapsto -\frac{1}{p}\sum_{i=1}^n \frac{1}{D^z}$

[10]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 $\mu(\{a\}) = \mu(\{b\}) = 0$

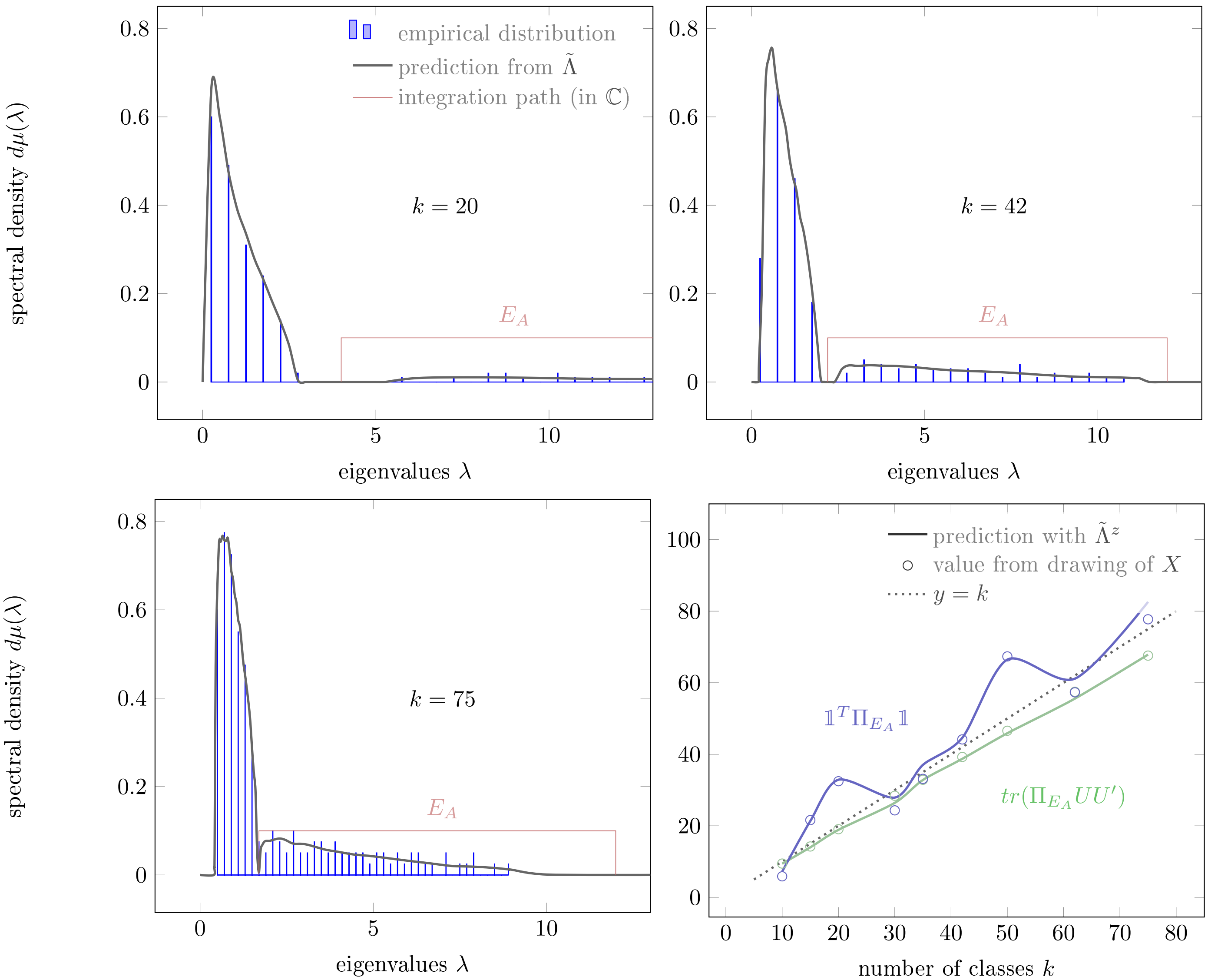

Figure 2.1: Prediction of the alignement of the signals in the data towards the eigen space of the biggest eigen values of $\frac{1}{n}XX^T$ for $p = 200$, $n = kn_k$ where $n_k = 20$ and k is the number of classes taking the values $k = 10, 15, 20, 30, 35, 42, 50, 62, 75$. The signals $u_1, \dots, u_k$ are drawn randomly and independently following a law $\mathcal{N}(0, I_p)$, and we let $U = (u_1, \dots, u_k) \in \mathcal{M}_{p,k}$. Then, for all $j \in [k]$ and $l \in [20]$, $x_{j+ln_k} \sim \mathcal{N}(u_j, I_p)$. **(top and bottom left)** representation of the spectral distribution of $\frac{1}{n}XX^T$ and its prediction with $\tilde{\Lambda}^z$.**(bottom right)** representation (with marks) of the quantities $\mathrm{Tr}(\Pi_{E_A} UU^T)$ and $\mathbb{1}^T \Pi_{E_A} \mathbb{1}$ and their prediction (smooth line) with the integration of, respectively, $\Im(\frac{1}{z\pi} \mathrm{Tr}(\tilde{Q}^{\tilde{\Lambda}^z}))$ and $\Im(\frac{1}{z\pi} \mathbb{1}^T \tilde{Q}^{\tilde{\Lambda}^z} \mathbb{1})$ on the path drawn in red on the other graphs. The line $y = k$ represents instances of $\mathrm{Tr}(\Pi_A) = k$ that can be approximated integrating $\frac{1}{z\pi} \mathrm{Tr}(\tilde{Q}^{\tilde{\Lambda}^z})$. The projection $\mathrm{Tr}(\Pi_{E_A} UU^T)$ is a little lower than k because the eigen vectors associated to the highest eigen values of $\frac{1}{n}XX^T$ are not perfectly aligned with the signal due to the randomness of X.

The first hypothesis to verify in order to employ Theorem 2.26 is the analycity of $\tilde{g}$, that originates from the analycity of $z \to \tilde{\Lambda}^z$. We can show with limiting arguments that $\tilde{\Lambda}^z$ is analytical as a limit of a sequence of analytical mappings. However, although it is slightly more laborious, we prefer here to prove the analicity from the original definition. First let us show the continuity with Proposition 1.23.

Proposition 2.27. *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 $\lim_{s\to\infty} t_s = 0$. Let us verify the assumption of Proposition 1.23 where for all $s \in \mathbb{N}$, $f^s = I^{z+t_s}$, $\tilde{\Gamma}^s = \tilde{\Lambda}^{z+t_s}$ and $\Gamma^s = \tilde{\Lambda}^z$ (it does not depend on s). We already know from Proposition 2.15 that f^s are all contracting for the stable semi-metric with a Lipschitz parameter $\lambda < 1$ that can be chosen independent from s for s big enough. Let us express for any $s \in \mathbb{N}$ and any $i \in [n]$:

$$f^s(\Gamma^s)_i - \Gamma^s_i = I^{z+t_s}(\tilde{\Lambda}^z)_i - \tilde{\Lambda}^z_i = t_s \tag{2.19}$$

Noting that for s sufficiently big, $\Im(I^{z+t_s}(\tilde{\Lambda}^z)) = \Im(t_s) + \Im(\tilde{\Lambda}^z) \geq \frac{\Im(\tilde{\Lambda}^z)}{4} \geq \frac{\Im(z)}{4}$, we see that $d_s(\Im(f^s(\Gamma^s)_i), \Im(\Gamma^s_i)) \leq \frac{4|\Im(t^s)|}{\Im(z)} \underset{s\to\infty}{\longrightarrow} 0$. Therefore, the assumptions of Proposition 1.23 are satisfied and we can conclude that there exists $K > 0$ 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:

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

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

Let us now show that $z \mapsto \tilde{\Lambda}^z$ is differentiable. Employing again the notation $f^t = \mathcal{I}^{z+t}$, we can decompose (noting for $D \in \mathcal{D}_n, R(D) \equiv (zI_p - \frac{1}{n}\sum_{i=1}^n$):

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

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

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

We have the equation:

$$a(t) = \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z+t})a(t) + t\mathbb{1}. \tag{2.20}$$

To be able to solve this equation we need:

Lemma 2.28. *Given any $z, z' \in \mathbb{H}$, $I_n - \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})$ is invertible.*

Proof. We are going to show the injectivity of $I_n - \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})$. Let us introduce a vector $x \in \mathbb{R}^n$ such that $x = \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})x$. We can bound thanks to Cauchy-Schwatz inequality, with similar calculus as in the proof of Proposition 2.15:

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

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 bound:

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

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

From the continuity of $(z,z') \mapsto \Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})$, and the limit $\lim_{x\to\infty} \|\Psi(\tilde{\Lambda}^x, \tilde{\Lambda}^x)\| = 0$ (see the proof of Proposition 2.33), we can deduce as a side result from Lemma 2.28:

Lemma 2.29. *Given any $z, z' \in \mathbb{H}$, $\|\Psi(\tilde{\Lambda}^z, \tilde{\Lambda}^{z'})\| < 1$*

The continuity of $z \mapsto \tilde{\Lambda}^z$ given by Proposition 2.27 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 equation

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

to obtain:

Proposition 2.30. *The mapping $z \mapsto \tilde{\Lambda}^z$ is analytic on $\mathbb{H}$, and satisfies:*

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

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

Proposition 2.31. *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 just check the hypotheses of Theorem 2.26. We already know that $z \mapsto -\frac{1}{\tilde{\Lambda}^z_i}$ is analytical thanks to Proposition 2.30 and the lower bound $\tilde{\Lambda}^z_i \geq \Im(z) > 0$. Besides, $\forall z \in \mathbb{H}$:

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

since $\tilde{\Lambda}^z \in \mathcal{D}_{I_z}$. Finally recalling from Lemma 2.17 that for all $y \in \mathbb{R}_+$, $\|\tilde{Q}^{\tilde{\Lambda}^{iy}}\| \leq \frac{|iy|}{\Im(iy)} = 1$, we directly see that for all $j \in [n]$:

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

we can then conclude with Theorem 2.26. □

We can then deduce easily that $\tilde{g}$ is also a Stieltjes transform with an interesting characterization of its measure with the $\tilde{\mu}_i$ (defined in Proposition 2.31).

Proposition 2.32. *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 (if $p > n$, then the measures $\tilde{\mu}_1, \dots, \tilde{\mu}_n$ contains Dirac weights on zero that cancel the term $-\frac{p-n}{p}\delta_0$).

Recall that $\tilde{\mu}$, satisfies $\tilde{g}(z) = \int_0^{+\infty} \frac{1}{\lambda - z} d\tilde{\mu}(\lambda)$, and let us denote $\tilde{S}$, its support. This formula implies that $\tilde{g}$ is analytic on $\mathbb{C} \setminus \tilde{S}$ and that for all $z \in \mathbb{C} \setminus \tilde{S}$, $\tilde{g}(\bar{z}) = \overline{\tilde{g}(z)}$. To precise the picture let us provide a result of compactness of $\tilde{S}$.

Proposition 2.33. *The measure $\tilde{\mu}$ has a compact support $\tilde{S} \subset \mathbb{R}_+$ and $\sup \tilde{S} \leq O(1)$.*

Proof. We are going to show that for x sufficiently big, $\lim_{y \to 0^+} \Im(g(x+iy)) = 0$, which will allow us to conclude thanks to the relation between $\tilde{\mu}$ and $\tilde{g}$ given in Theorem 2.26. Considering $z = x + iy \in \mathbb{H}$, for $x, y \in \mathbb{R}$ and such that

$$x \geq x_0 \equiv \max\left(\frac{8}{n}\sup_{i \in [n]} \mathrm{Tr}(\Sigma_i), 4\nu\right)$$

let us show first that $\forall i \in [n]$, $\Re(\tilde{\Lambda}^z_i) \geq \frac{x_0}{2}$. This is a consequence of the fact that $\mathcal{I}^z$ is stable on $\mathcal{A} \equiv \mathcal{D}_n(\{t \geq \frac{x}{2}\} + i\mathbb{R}^*_+) \cap \mathcal{D}_{I^z}$. Indeed, given $L \in \mathcal{A}$

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

and as we already know, since $\mathcal{D}_{I^z}$, $\Im\left((\tilde{Q}^L)^{-1}\right) \geq 0$, therefore, $\|\tilde{Q}^L\| \leq 2$. We can then bound:

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

Thus as a limit of elements of $\mathcal{A}$, $\tilde{\Lambda}^z \in \mathcal{A}$, and $\forall i \in [n]$, $\Re(\tilde{\Lambda}^z_i) \geq \frac{x}{2}$.

Besides, let us bound:

$$\begin{aligned}
\Im(\tilde{\Lambda}^z_i) &= y + \frac{1}{n}\mathrm{Tr}\left(\Sigma_i R(\tilde{\Lambda}^z)\frac{1}{n}\sum_{j=1}^{n}\frac{\Im(\tilde{\Lambda}^z_j)\Sigma_j}{|\tilde{\Lambda}^z_j|^2}\bar{R}(\tilde{\Lambda}^z)\right) \\
&\leq y + \frac{4}{n}\mathrm{Tr}\left(\Sigma_i\right)\left\|\frac{1}{n}\sum_{j=1}^{n}\Sigma_j\right\|\sup_{j \in [n]}\frac{\Im(\tilde{\Lambda}^z_j)}{\Re(\tilde{\Lambda}^z_j)^2}.
\end{aligned}$$

We can further bound $\left\|\frac{1}{n}\sum_{j=1}^{n}\Sigma_j\right\| \leq \nu$, since $\frac{1}{n}\|XX^T\| = (\|X\|/\sqrt{n})^2 \in \nu \pm \mathcal{E}_1(1/\sqrt{n})$ and therefore $\left\|\frac{1}{n}\sum_{j=1}^{n}\Sigma_j\right\| \leq \mathbb{E}[\frac{1}{n}\|XX^T\|] \leq \nu + O(1/\sqrt{n}) \leq 2\nu$. Besides $\Re(\tilde{\Lambda}^z_j)^2 \geq \frac{xx_0}{4}$, we can eventually bound $\sup_{j \in [n]}\Im(\tilde{\Lambda}^z_j) \leq y + \frac{2\nu}{x}\sup_{j \in [n]}\Im(\tilde{\Lambda}^z_j)$, which implies, for x sufficiently big:

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

We can then conclude letting y tend to 0 in the formulation $\tilde{g} = g^{\tilde{\Lambda}^z}$:

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

$\square$

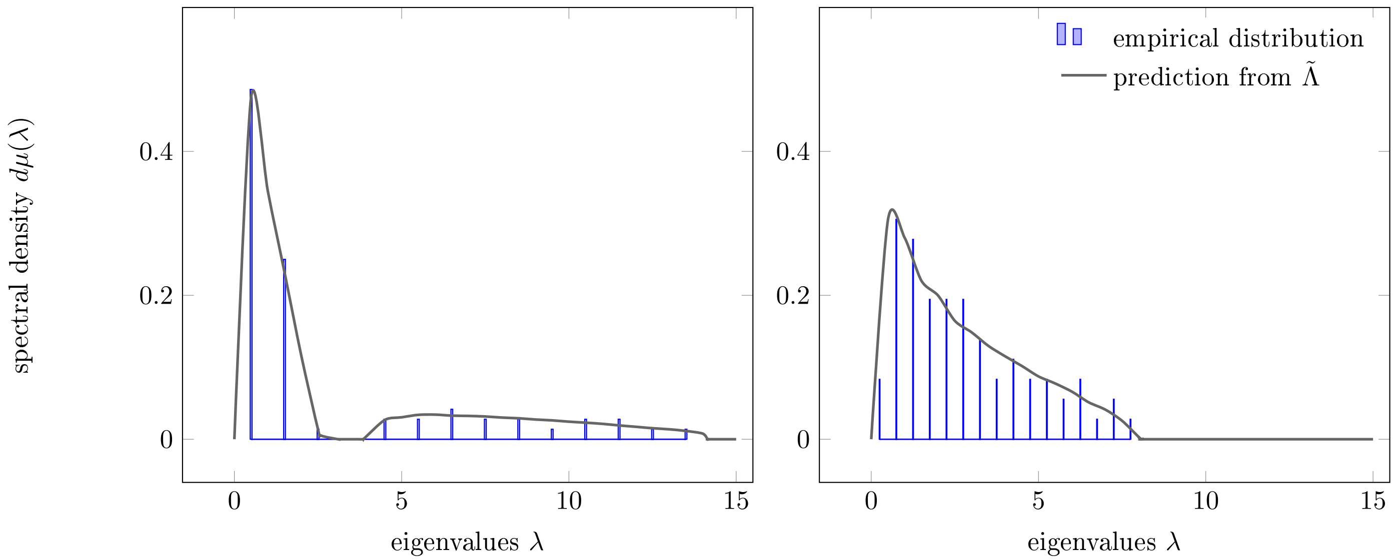

Figure 2.2: Spectral distribution of $\frac{1}{n}XX^T$ and its deterministic estimate obtained from $\tilde{\Lambda}$ for $n = 160$ and $p = 80$. Introducing P an orthogonal matrix chosen randomly and $\Sigma \in \mathcal{D}_p$ such that for $j \in \{1, \dots, 20\}$, $\Sigma_j = 1$ and for $j \in \{21, \dots, 80\}$, $\Sigma_j = 8$, we chose **(left)** $\forall i \in [n]$, $x_i \sim \mathcal{N}(0, \Sigma)$ and **(right)** $\forall i \in [n]$, $x_i \sim \mathcal{N}(0, \Sigma_i)$, where $\Sigma_1 = \Sigma$ and $\Sigma_{i+1} = P^T \Sigma_i P$ for all $i \in [n]$. The histograms would have been similar for any other concentrated vectors $x_1, \dots, x_n$ having the same covariances and comparable observation diameter (see Definition **??**)

We end this section with the proof of the convergence of the Stieltjes transform of the spectral distribution g towards $\tilde{g}$.

We introduce the semi-norm $\| \cdot \|_{S^\varepsilon}$, defined for any $f \in \mathcal{F}(\mathbb{C})$ as:

$$\|f\|_{S^\varepsilon} = \sup_{z \in \mathbb{C} \setminus S^\varepsilon} |f(z)|.$$

Theorem 2.34. *$g \mid \mathcal{A}_Q \in \tilde{g} \pm \mathcal{E}_2(1/\sqrt{pn})$ in $(\mathbb{C}^{\mathbb{C} \setminus S^\varepsilon}, \| \cdot \|_{S^\varepsilon})$.*

Proof. We know from remark 2.21 that $O(\frac{\kappa_z}{z}) \leq O(1)$ and therefore we can show as in the proof of Proposition 2.3 that the mapping g defined for any $z \in \mathbb{C} \setminus (S^\varepsilon \cup \{0\})$ with the identity $g(z) = -\frac{1}{pz} \operatorname{Tr}(Q^z)$ is, under $\mathcal{A}_Q$ a $O(1/\sqrt{pn})$-Lipschitz transformation of X, thus $(g \mid \mathcal{A}_Q) \propto \mathcal{E}_2(1/\sqrt{pn})$ in $(\mathcal{F}(\mathbb{C}), \| \cdot \|_{S^\varepsilon})$. We can then conclude thanks to the bound:

$$\begin{aligned}
\sup_{z \in \mathbb{C} \setminus S^\varepsilon} \left| \mathbb{E}_{\mathcal{A}_Q}[g(z)] - \tilde{g}(z) \right| &\leq \sup_{z \in \mathbb{C} \setminus S^\varepsilon} \frac{1}{zp} \left| \operatorname{Tr} \left(\mathbb{E}_{\mathcal{A}_Q}\left[Q^z\right] - \tilde{Q}^{\tilde{\Lambda}^z} \right) \right| \\
&\leq \sup_{z \in \mathbb{C} \setminus S^\varepsilon} O\left(\frac{\kappa_z}{|z|\sqrt{np}} \right) \leq O\left(\frac{1}{\sqrt{np}} \right).
\end{aligned}$$

□

In particular, thanks to the Cauchy identity, for any analytical mapping $f : \mathbb{C} \to \mathbb{C}$, since the integration on bounded paths of $\mathbb{C} \setminus S^\varepsilon$ is Lipschitz for the norm $\| \cdot \|_{S^\varepsilon}$, we can approximate:

$$\mathbb{P}\left(\left| \int f(\lambda) d\mu(\lambda) - \frac{1}{2i\pi} \oint_\gamma f(z) g(z) dz \right| \geq t \right) \leq Ce^{-cnpt^2} + Ce^{-cn},$$

for any closed path $S^\varepsilon \subset \gamma \subset \mathbb{C} \setminus S^\varepsilon$ with length l_γ satisfying $l_\gamma \leq O(1)$. There exists a correspondence between a distribution and its Stieltjes transform: denoting μ the spectral distribution of $\frac{1}{n}XX^T$, we have indeed for any real $a < b$:

$$\mu([a,b]) = \lim_{y \to 0} \frac{1}{\pi} \int_a^b \Im(g(x+iy))dx.$$

As seen on Figure 2.2, this measure is naturally close to $\tilde{\mu}$ defined for any real $a < b$ as $\tilde{\mu}([a,b]) = \lim_{y\to 0} \frac{1}{\pi} \int_a^b \Im(\tilde{g}(x+iy))dx$, it can indeed be shown that the Kolmogorov distance between those two measures tends to zero as stated in **?**.

2.2 Statistical study of the resolvent with convex concentration hypotheses

We want in this section to extend the result of Section 2.1 to the case of convexly concentrated matrices. Recall that convexly concentrated random vectors class is not stable through Lipschitz maps, we just set it was stable through affine transformations. As a consequence, in this setting, the concentration of the resolvent $Q^z = (I_n - X/z)^{-1}$ is no more a mere consequence of a bound on its differential on $X \in \mathcal{M}_p$. Still, as first shown by **?**, it is possible to obtain concentration properties on the sum of Lipschitz functionals of the eigen values. Here we pursue the study, looking at *linear* concentration properties of Q for which concentration inequalities are only satisfied by 1-Lipschitz *and linear* functionals f. The well known identity

$$\frac{1}{p} \sum_{\lambda \in \mathrm{Sp}(X)} f(\lambda) = -\frac{1}{2i\pi} \oint_\gamma \frac{f(z)}{z} \operatorname{Tr}(Q^z) dz, \tag{2.21}$$

is true for any analytical mapping f defined on the interior of a path $\gamma \in \mathbb{C}$ containing the spectrum of X (or any limit of such mappings), therefore, our results on the concentration of Q^z concern in particular the quantities studied in **?**.

The linear concentration of the resolvent is proven thanks to Theorem **??**, expressing it as a sum $Q^z = \frac{1}{z} \sum_{i=1}^\infty (X/z)^i$. The linear concentration of the powers of X was justified in Theorem **??**. We call this weakening of the concentration property "the degeneracy of the convex concentration through multiplication". The linear concentration of the resolvent is though sufficient for most practical applications that rely on an estimation of the Stieltjes transform $m(z) = \frac{1}{zp} \operatorname{Tr}(Q^z)$ or on projections on Q^z.

We still work under Assumptions 1-5, we just adapt Assumption 2 to the convex concentration setting and assume instead:

Assumption 2$_c$ (Convex concentration)**.** $X \propto_c \mathcal{E}_2$.

We also keep the same notations $S = \{\mathbb{E}_{\mathcal{A}_\nu}[\lambda_1], \ldots, \mathbb{E}_{\mathcal{A}_\nu}[\lambda_p]\}$ and

$$\mathcal{A}_Q \equiv \left\{ \forall i \in [p], \sigma_i \left(\frac{1}{n} X X^T \right) \in S^{\varepsilon/2} \right\},$$

for some $\varepsilon \geq O(1)$. the concentration of $\sigma(X)/\sqrt{n} \in \mathbb{E}[\sigma(X)] \pm \mathcal{E}_2(1/\sqrt{n})$, allows us to set[11] as in Lemma 2.1 that there exist two constants $C, c > 0$ such that $\mathbb{P}\left(\mathcal{A}_Q^c\right) \leq Ce^{-cn\varepsilon^2}$.

We can beside deduce from Lemma **??** that $(X \mid \mathcal{A}_Q) \propto_c \mathcal{E}_2$.

Placing ourselves under the event $\mathcal{A}_Q$, let us first show that the resolvent $Q^z \equiv (I_p - \frac{1}{nz} X X^T)^{-1}$ is concentrated if z has a big enough modulus. Be careful that the following concentration is expressed for the nuclear norm (for any deterministic matrix $A \in \mathcal{M}_p$ such that $\|A\| \leq O(1)$, $\operatorname{Tr}(AQ^z) \in \mathcal{E}_2$). The next proposition is just provided as a first direct application of Theorem **??**, a stronger result is provided in Proposition 2.36.

Proposition 2.35. *Given $z \in \mathbb{C}$ such that $|z| \geq \nu + \varepsilon$:*

$$(Q^z \mid \mathcal{A}_Q) \in \mathcal{E}_2 \qquad \textit{in} \ \ (\mathcal{M}_p, \|\cdot\|_*).$$

[11] In Lemma 2.1, the proof is conducted for Lipschitz concentration hypotheses on X. However, since only the linear concentration of $\sigma(X)$ is needed, the justification is the same in a context of convex concentration thanks to Theorem **??**.

Proof. We know from Lemma ?? that $(X \mid \mathcal{A}_Q) \propto_c \mathcal{E}_2$ and from Theorem ?? that (here $\kappa = \nu + \frac{\varepsilon}{2} \leq O(1)$, $\sigma = 1/\sqrt{n}$ and $p = O(n)$):

$$\text{Under } \mathcal{A}_Q: \qquad \left(\frac{1}{n}XX^T\right)^m \in \mathcal{E}_2\left(\left(\nu + \frac{\varepsilon}{2}\right)^m \sqrt{m}\right) \qquad \text{in } (\mathcal{M}_p, \|\cdot\|_*).$$

Let us then note that $\left(\nu + \frac{\varepsilon}{2}\right)^m \sqrt{m} = O\left(\left(\nu + \frac{3\varepsilon}{4}\right)^m\right)$ and for $z \in \mathbb{C}$ satisfying our hypotheses: $(\nu + \frac{3\varepsilon}{4})/|z| \leq 1 - \frac{\varepsilon}{4(\nu+\varepsilon)}$. We can then deduce from Corollary ?? that under $\mathcal{A}_Q$:

$$Q^z = \frac{1}{z}\left(I_p - \frac{1}{zn}XX^T\right)^{-1} = \frac{1}{z}\sum_{i=1}^{\infty}\left(\frac{1}{zn}XX^T\right)^i \in \mathcal{E}_2\left(\frac{4}{\varepsilon}(\nu+\varepsilon)\right).$$

□

Let us now try to study the concentration of Q^z when z gets close to the spectrum.

Proposition 2.36. *For all $z \in \mathbb{C} \setminus S^\varepsilon$:*

$$(Q^z \mid \mathcal{A}_Q) \in \mathcal{E}_2(\kappa_z) \qquad \textit{in} \quad (\mathcal{M}_p, \|\cdot\|_*),$$

Proof. Proposition 2.35 already set the result for $|z| \geq \nu + \varepsilon \equiv \rho$, therefore, let us now suppose that $|z| \leq \rho$.

With the notation $M \equiv \left(\Im(z)^2 + \left(\Re(z) - \frac{1}{n}XX^T\right)^2\right)^{-1} = |\frac{1}{z}Q^z|^2$, let us decompose:

$$\frac{1}{z}Q^z = \left(\Re(z) - \frac{1}{n}XX^T\right)M - \Im(z)iM. \tag{2.22}$$

We can then deduce the linear concentration of M with the same justifications as previously thanks to the Taylor decomposition:

$$M = \frac{|z|^2}{\rho^2}\sum_{m=0}^{\infty}\left(1 - \frac{\Im(z)^2}{\rho^2} - \frac{\left(\Re(z) - \frac{1}{n}XX^T\right)^2}{\rho^2}\right)^m.$$

Indeed, $\|\Re(z)I_p - \frac{1}{n}XX^T\| \leq d(\Re(z), S)$ and $d(z,S)^2 = \Im(z)^2 + d(\Re(z), S)^2 \leq \rho$ thus:

$$\left\|1 - \frac{\Im(z)^2}{\rho^2} - \frac{1}{\rho^2}\left(\Re(z)I_p - \frac{1}{n}XX^T\right)^2\right\| \leq 1 - \frac{d(z,S)^2}{\rho^2} \leq 1 - \frac{\varepsilon^2}{\rho^2} < 1.$$

We therefore deduce from (2.22) that:

$$(Q^z \mid \mathcal{A}_Q) \in \mathcal{E}_2\left(\frac{|z|}{\varepsilon^2}\left(|\Im(z)| + |\Re(z)| + \nu + \frac{\varepsilon}{2}\right)\right) = \mathcal{E}_2(\kappa_z).$$

□

One can then follow the lines of study made for the Lipschitz concentration case to show that Q^z is also concentrated around $\tilde{Q}^{\bar{\Lambda}^z}$ under convex concentration hypotheses. Although Proposition 2.36 gives us a concentration of Q^z in nuclear norm, we will estimate $\mathbb{E}[Q^z]$ with the Frobenius norm as in the Lipschitz concentration case.

Proposition 2.37. *For any $z \in \mathbb{C} \setminus S^\varepsilon$ in convex concentration setting:*

$$\left\|\mathbb{E}[Q^z] - \tilde{Q}^{\bar{\Lambda}^z}\right\|_F \leq O\left(\frac{\kappa^z}{\sqrt{n}}\right).$$

The Proof of the Theorem is done the same way as in the Lipschitz case, it just relies on the Lemmas 2.40 and 2.38 below.

Lemma 2.38. *Under $\mathcal{A}_Q$, for any $z \in \mathbb{C} \setminus S^\varepsilon$ and any $i \in [n]$:*

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

The proof is the same as the one of Lemma 2.10 and relies on the bound on Λ^z_i given by Lemma 2.7 and the concentration of $u^T Q_{-i} x_i$ provided in next lemma.

Lemma 2.39. *For any $z \in \mathbb{C} \setminus S^\varepsilon$, any $i \in [n]$ and any $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:*

$$(u^T Q^z_{-i} x_i \mid \mathcal{A}_Q) \in O(1) \pm \mathcal{E}_2.$$

Proof. We do not care about the independence issues brought by $\mathcal{A}_Q$. Let us simply bound for any $t > 0$ and under $\mathcal{A}_Q$:

$$\begin{aligned}
&\mathbb{P}\left(\left|u^T Q^z_{-i} x_i - \mathbb{E}\left[u^T Q^z_{-i} x_i\right]\right| \geq t\right) \\
&\quad \leq \mathbb{P}\left(\left|u^T Q^z_{-i}(x_i - \mu_i)\right| \geq \frac{t}{2}\right) + \mathbb{P}\left(\left|u^T\left(Q^z_{-i} - \mathbb{E}\left[Q^z_{-i}\right]\right)\mu_i\right| \geq \frac{t}{2}\right) \\
&\quad \leq \mathbb{E}\left[Ce^{-cnt^2/\|Q_{-i}\|^2}\right] + Ce^{-cnt^2} \quad \leq \ 2Ce^{-c'nt^2},
\end{aligned}$$

for some constants $C, c, c' > 0$. Besides, we can bound:

$$\left|\mathbb{E}\left[u^T Q^z_{-i} x_i\right]\right| = \left|u^T \mathbb{E}[Q^z_{-i}] m_i\right| \leq O(1),$$

thanks to Lemma 2.2.

□

The last Lemma is then important to employ similar results as Proposition **??** as in the proof of Proposition 2.12.

Lemma 2.40. *For any $z \in \mathbb{C} \setminus S^\varepsilon$:*

$$\forall i \in [n], \Lambda^z_i \mid \mathcal{A}_Q \in \tilde{\Lambda}^z_i \pm \mathcal{E}_2(\kappa_z/\sqrt{n}) + \mathcal{E}_1(\kappa_z/n).$$

Besides, for any deterministic matrix $A \in \mathcal{M}_p$:

$$(x_i^T A Q^z x_i \mid \mathcal{A}_Q) \in \mathrm{Tr}(\Sigma_i A \mathbb{E}[Q^z]) \pm \mathcal{E}_2\left(\kappa_z \|A\|_F\right) + \mathcal{E}_1\left(\kappa_z \|A\|\right).$$

Proof. We will first show the concentration of $\frac{1}{n} x_i^T A Q^z_{-i} x_i$ for any deterministic matrix $A \in \mathcal{M}_p$ that will in particular imply the concentration of Λ^z_i. Recall that under $\mathcal{A}_Q$, $\|X\| \leq O(1)$ and $\|Q^z\| \leq \kappa_z$. Given $i \in [n]$, we want to bound:

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

Now we know that, for X_{-i} fixed, we can bound thanks to Proposition **??**:

$$\begin{aligned}
\mathbb{P}\left(\left|x_i^T A Q^z_{-i} x_i - \mathrm{Tr}(\Sigma_i^T A Q^z_{-i})\right| \geq t\right) &\leq \mathbb{E}\left[Ce^{-c(t/\|Q^z_{-i}\|\|A\|_F)^2} + Ce^{-ct/\|Q^z_{-i}\|\|A\|}\right] \\
&\leq Ce^{-c'(t/\|A\|_F \kappa_z)^2} + Ce^{-c't/\|A\|\kappa_z},
\end{aligned}$$

for some constants $C, c, c' > 0$, thanks to Lemma 2.2.

Besides, we know from Proposition 2.36 and Lemma **??** that $Q^z_{-i} \in \mathbb{E}[Q^z] \pm \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{M}_p, \|\cdot\|_*)$, which allows us to bound:

$$\mathbb{P}\left(\left|\mathrm{Tr}(\Sigma_i A Q^z_{-i}) - \mathrm{Tr}(\Sigma_i A \mathbb{E}[Q^z])\right| \geq t\right) \leq Ce^{-ct^2/\|A\|_F^2}$$

for some constants $C, c > 0$, since $\|\Sigma_i\| \leq O(1)$. Putting the two concentration inequalities together, we obtain the concentration of $\frac{1}{n} x_i^T A Q^z_{-i} x_i$.

Now the identity $\frac{1}{n} x_i^T A Q^z x_i = \frac{1}{n} \frac{x_i^T A Q^z_{-i} x_i}{\Lambda^z_i}$, and the bounds $\Lambda^z_i \geq O(|z|/\breve{\kappa}_z)$ and $\frac{1}{n} x_i^T A Q^z_{-i} x_i \leq \kappa_z \rho$ allows us to deduce the concentration of $\frac{1}{n} x_i^T A Q^z x_i$ thanks to Lemma **??** setting the concentration of the product of concentrated variables. □

2.3 The deterministic equivalents of the powers of the resolvent

Our aim here is to find deterministic equivalent for all the powers $(Q^z)^m$, for $m \in \mathbb{N}$ and for $z \in \mathbb{C}$ satisfying $d(z, S^\varepsilon) \geq O(1)$ for a given constant $\varepsilon > 0$ ($\varepsilon \geq O(1)$). For simplicity, we will omit the exponent z of the resolvent. We continue to work under Assumptions 1-5 we thus won't recall them in the coming lemmas and propositions. The main goal of this section is to display the formulas given in Proposition 2.47, because they might be seen in other contexts thereby allowing to interesting analogies. The coming proofs are quite laborious and do not present all the required justification mainly to avoid complex inferences of little interest.

In the concentration inequalities, it is interesting to keep track of the dependence on m to use it later for a quasi-asymptotic formulations, we will thus assume sometimes that $m \in \mathbb{N}^{\mathbb{N}}$ is a sequence of integer (as n and p). Since it is costless to devise directly a deterministic equivalent for $QA_1QA_2, \cdots A_{m-1}Q$ where $A_1, \ldots A_{m-1}$ are all matrices of unit spectral norm we will consider this general case. Considering from now on $m-1$ matrices $A_1, \ldots, A_{m-1}$, where $\forall i \in [m]$, $\|A_i\| \leq 1$ and we note for any matrix $M \in \mathcal{M}_p$, $k \in [m]$:

$$M^{A_l^k} = \begin{cases} \text{if } k < l \ : \ I_p \\ \text{if } k = l \ : \ M \\ \text{if } k > l \ : \ MA_lMA_{l+1} \ldots A_{k-1}M \end{cases} .$$

We further note for formulation simplicity $A_m = I_p$.

Proposition 2.41. *Given a sequence of integer $m \in \mathbb{N}^{\mathbb{N}}$:*

$$Q^{A_1^m} \mid \mathcal{A}_Q \in \mathbb{E}[Q^{A^m}] \pm \mathcal{E}_2\left(\frac{m\kappa_z^m}{\sqrt{n}}\right) \qquad \text{in} \qquad (\mathcal{M}_p, \|\cdot\|_F)$$

Proof. The result is proven the same way as Proposition 2.3, employing the bound on Q^z given by Lemma 2.2. □

We now look for a computable deterministic equivalent of $Q^{A_1^m}$. Let us first provide a useful formula to start to disentangle the dependence between $Q^{A_1^m}$ and x_i; it is a consequence of the Schur formulas (2.8)

Lemma 2.42. *For any $m \in \mathbb{N}_*$:*

$$Q^{A_1^m} = Q_{-i}^{A_1^m} + \frac{1}{\Lambda_i^z n} \sum_{l=1}^m Q_{-i}^{A_1^l} x_i x_i^T Q_{-i} A_l Q^{A_{l+1}^m}$$

where $\mathcal{L}_k^m = \{(l_1, \ldots, l_k) \in \mathbb{N}^k \mid 1 \leq l_1 < \ldots < l_k \leq m\}$ and $\mathcal{L}_0^m = \{()\}$ contains only the 0-tuple (and recall that $A_m = I_p$).

Proof. The result is just a consequence of the following telescoping sum decomposition (where we set $A_0, A_m = I_p$):

$$\begin{aligned} Q^{A^m} - Q_{-i}^{A^m} &= \sum_{l=1}^m Q_{-i}^{A_1^{l-1}} A_{l-1} Q^{A_l^m} - Q_{-i}^{A_1^l} A_l Q^{A_{l+1}^m} \\ &= \sum_{l=1}^m Q_{-i}^{A_1^{l-1}} A_{l-1}(Q - Q_{-i}) A_l Q^{A_{l+1}^m} \ = \ \frac{1}{\Lambda_i^z n} \sum_{l=1}^m Q_{-i}^{A_1^l} x_i x_i^T Q_{-i} A_l Q^{A_{l+1}^m}. \end{aligned}$$

□

Lemma 2.43. *Given a deterministic vector $u \in \mathbb{R}^p$, for any $k, l \in \mathbb{N}$ such that $1 \leq k < l < m$:*

$$u^T Q_{-i}^{A_k^l} x_i \mid \mathcal{A}_Q \in O\left(\kappa_z^{l-k+1}\right) \pm \mathcal{E}_2\left((c\kappa_z)^{l-k+1}\right),$$

for some constant $c > 1$.

Corollary 2.44. *$\left\|\mathbb{E}_{\mathcal{A}_Q}[Q_{-i}^{A_1^m} - Q^{A_1^m}]\right\| = O\left(\frac{(c\kappa_z)^m}{n}\right)$ for a constant $c > 1$.*

Proof. We do not use directly the second identity of Lemma 2.42 because it would let a term m^m appear in the bound. Instead of the first result of Lemma 2.42, one could have stated the identity:

$$Q^{A_1^m} = Q_{-i}^{A_1^m} + \frac{1}{\Lambda_i^z n} \sum_{l=1}^{m} Q^{A_1^{l-1}} A_{l-1} Q_{-i} x_i x_i^T Q_{-i}^{A_l^m},$$

(recall the notation $A_0 = I_p$). Putting those two identities together one obtains:

$$\begin{aligned} Q^{A_1^m} &= Q_{-i}^{A_1^m} + \frac{1}{\Lambda_i^z n} \sum_{l=1}^{m} Q_{-i}^{A_1^l} x_i x_i^T Q_{-i}^{A_l^m} \\ &+ \frac{1}{(\Lambda_i^z n)^2} \sum_{l=1}^{m} \sum_{k=l+2}^{m} Q_{-i}^{A_1^l} x_i x_i^T Q_{-i} A_l Q^{A_{l+1}^{k-1}} A_{k-1} Q_{-i} x_i x_i^T Q_{-i}^{A_k^m} \\ &+ \frac{1}{(\Lambda_i^z n)^2} \sum_{l=1}^{m} Q_{-i}^{A_1^l} x_i x_i^T Q_{-i} A_l Q_{-i} x_i x_i^T Q_{-i}^{A_{l+1}^m} \end{aligned} \tag{2.23}$$

Thus, knowing from Lemmas 2.43 and 2.2 that given any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$, any $l \in [m]$ and any $k \in \{l+2, \dots, m\}$ we have:

- $u^T Q_{-i}^{A_1^l} x_i \mid \mathcal{A}_Q \in O(\kappa_z^l) \pm \mathcal{E}_2((c\kappa_z)^l)$
- $\frac{1}{|\Lambda_i^z n|} \left| x_i^T Q_{-i} A_l Q^{A_{l+1}^{k-1}} A_{k-1} Q_{-i} x_i \right| \leq \frac{\kappa_z^{k-l+1} \check{\kappa}_z^2}{|z|n} \leq O\left(\frac{\kappa_z^{k-l}}{n}\right)$,

one can deduce from the estimation of the product of concentrated random variables given in Lemma **??** that for all $u, v \in \mathbb{R}^p$ such that $\|u\|, \|v\| \leq 1$:

$$u^T \mathbb{E}_{\mathcal{A}_Q}\left[Q_{-i}^{A_1^m} - Q^{A_1^m}\right] v \leq O\left(\frac{(c\kappa_z)^m}{n}\right) + O\left(\frac{(c\kappa_z)^{k-l-1}}{n^2}\right) \leq O\left(\frac{(c\kappa_z)^m}{n}\right).$$

□

Before starting the estimation of $\mathbb{E}_{\mathcal{A}_Q}[Q^{A_1^m}]$, we still need a result analogous to Lemma 2.13.

Lemma 2.45. *Given $k, l \in \mathbb{N}$ such that $1 \leq k < l < m$:*

$$Q^{A_k^l} X \mid \mathcal{A}_Q \propto \mathcal{E}_2\left((c\kappa_z)^{l-k+1}\right),$$

for some constant $c > 1$.

Proof. The concentration is really proven the same way as in the proof of Proposition 2.3 (and the fact that there exists a constant $c > 0$ such that for all $i \in [m]$, $i \leq c^i$).

To bound $\|\mathbb{E}_{\mathcal{A}_Q}[Q^{A_k^l} x_i]\|$, we use the identity (2.23) and bound for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:

$$\begin{aligned} \left|\mathbb{E}_{\mathcal{A}_Q}\left[u^T Q^{A_k^l} x_i\right]\right| &= \left|\mathbb{E}_{\mathcal{A}_Q}\left[u^T Q^{A_k^l} x_i\right]\right| + \frac{1}{\hat{\Lambda}_i^z n} \sum_{j=k}^{l} \left|\mathbb{E}_{\mathcal{A}_Q}\left[u^T Q_{-i}^{A_k^j} x_i x_i^T Q_{-i}^{A_j^l} v\right]\right| \\ &+ \left|\frac{1}{(\hat{\Lambda}_i^z n)^2} \sum_{j=k}^{l} \mathbb{E}_{\mathcal{A}_Q}\left[u^T Q_{-i}^{A_k^j} x_i\right] \mathbb{E}_{\mathcal{A}_Q}\left[x_i^T Q_{-i} A_j Q_{-i} x_i\right] \mathbb{E}_{\mathcal{A}_Q}\left[x_i^T Q_{-i}^{A_{j+1}^l} v\right]\right| \\ &+ \left|\frac{1}{(\hat{\Lambda}_i^z n)^2} \sum_{j=k}^{l} \sum_{o=j+2}^{l} \mathbb{E}_{\mathcal{A}_Q}\left[u^T Q_{-i}^{A_k^j} x_i\right] \mathbb{E}_{\mathcal{A}_Q}\left[x_i^T Q_{-i} A_j Q^{A_{j+1}^{o-1}} A_{o-1} Q_{-i} x_i\right] \mathbb{E}_{\mathcal{A}_Q}\left[x_i^T Q_{-i}^{A_o^l} v\right]\right| \\ &+ O\left((c\kappa)^{l-k+1} + \frac{(l-k)(c\kappa)^{l-k+1}\check{\kappa}_z}{n|z|}\left(1 + \frac{\kappa_z \check{\kappa}_z (l-k)}{|z|n}\right)\right) \leq O\left((c\kappa_z)^{l-k+1}\right), \end{aligned}$$

for some constant $c > 0$. □

Proposition 2.46. *With the notation:*

$$T^{A_l^m} \equiv \mathrm{Diag}_{i\in[m]}\left(\frac{1}{n\Lambda_i^z} x_i^T Q_{-i}^{A_l^m} x_i\right),$$

one can estimate for any (sequence of) integer m and any set of deterministic matrices $A_1,\dots,A_{m-1}\in\mathcal{M}_p^{m-1}$:

$$\Bigg\| Q^{A_1^m} - Q^{A_1^{m-1}} A_{m-1}\tilde{Q} - \frac{1}{n}\sum_{i=1}^n \sum_{k=1}^{m-1} \frac{1}{\Lambda_i^z} \sum_{l\in L_k^{m-1}} \mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ_{-i}^{A_1^{l_1}}\Sigma_i\tilde{Q}\right)\right] \tilde{T}_i^{A_{l_1}^{l_2}}\cdots\tilde{T}_i^{A_{l_k}^m}\Bigg\|_F \le O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right),$$

where for any $k,l\in[m]$, $k<l$, $\tilde{T}^{\mathcal{A}_k^l}$ is a deterministic matrix satisfying $\|\tilde{T}_i^{A_{l_k}^m} - \tilde{T}_i^{A_{l_k}^m}\|_F \le O((c\kappa_z)^{l-k+1}/\sqrt{n})$.

Proof. Given a deterministic matrix $A\in\mathcal{M}_p$ such that $\|A\|_F\le 1$, let us try and estimate $\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^m}A_m(Q-\tilde{Q})\right)\right]$, we allow ourselves not to display all the steps of the calculus since similar inferences were already displayed in the proofs of Propositions 2.12 and 2.20. For simplicity, we note $\tilde{Q}$ instead of $\tilde{Q}^{\tilde{\Lambda}^z}$.

$$\begin{aligned}
&\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^{m-1}}A_{m-1}(Q-\tilde{Q})\right)\right]\\
&\quad= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q}\left[\frac{1}{z}\mathrm{Tr}\left(AQ^{A^m}x_ix_i^T\tilde{Q}\right)\right] - \mathbb{E}_{\mathcal{A}_Q}\left[\frac{\mathrm{Tr}\left(AQ^{A^m}\Sigma_i\tilde{Q}\right)}{\tilde{\Lambda}_i^z}\right]\\
&\quad= \frac{1}{n}\sum_{i=1}^n a_i + b_i + c_i
\end{aligned}$$

where for all $i\in[n]$:

$$\begin{aligned}
a_i &= \mathbb{E}_{\mathcal{A}_Q}\left[\frac{\tilde{\Lambda}_i^z - \Lambda_i^z}{\Lambda_i^z\tilde{\Lambda}_i^z}\,\mathrm{Tr}\left(AQ^{A^{m-1}}A_{m-1}Q_{-i}x_ix_i^T\tilde{Q}\right)\right]\\
b_i &= \frac{1}{\tilde{\Lambda}_i^z}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(A\left(Q^{A_1^{m-1}} - Q_{-i}^{A_1^{m-1}}\right)A_{m-1}Q_{-i}x_ix_i^T\tilde{Q}\right)\right]\\
c_i &= \frac{1}{\tilde{\Lambda}_i^z}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(A\left(Q_{-i}^{A^m} - Q^{A^m}\right)\Sigma_i\tilde{Q}\right)\right]
\end{aligned}$$

Let us first apply Proposition **??** with the hypotheses:

- $\tilde{Q}X \mid \mathcal{A}_Q \propto \mathcal{E}_2(\kappa_z)$ and $\|\mathbb{E}[\tilde{Q}x_i]\|\le O(1)$ thanks to Assumption 4, 2,
- $Q^{A_1^m}X \mid \mathcal{A}_Q \propto \mathcal{E}_2((c\kappa_z)^m))$ and $\|\mathbb{E}_{\mathcal{A}_Q}[Q^{A_1^m}x_i]\|\le O((c\kappa_z)^m))$ given by Lemma 2.45,
- $\Lambda^z \mid \mathcal{A}_Q \in \mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] \pm \mathcal{E}_2\left(\frac{\kappa^z}{\sqrt{n}}\right)$ in $(\mathcal{D}_n, \|\cdot\|)$ given by Lemma 2.5,
- $\|\mathbb{E}_{\mathcal{A}_Q}[\Lambda^z] - \hat{\Lambda}^z\|_F \le O(\kappa^z/\sqrt{n})$ thanks to Lemma 2.6,

to set:

$$\left|\frac{1}{n}\sum_{i=1}^n a_i\right| = \mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A^m}X\frac{\tilde{\Lambda}^z - \Lambda^z}{\tilde{\Lambda}^z}X^T\tilde{Q}\right)\right] \le O\left(\frac{\check{\kappa}_z(c\kappa_z)^{m+2}}{|z|\sqrt{n}}\right) \le \left(\frac{(c\kappa_z)^{m+1}}{\sqrt{n}}\right)$$

The bound on c_i is just a consequence of Corollary 2.44:

$$|c_i| \le \kappa_z \left\|\mathbb{E}_{\mathcal{A}_Q}\left[Q_{-i}^{A^l} - Q^{A^l}\right]\right\| = O\left(\frac{\kappa_z^m}{n}\right),$$

Because it relies on similar justifications, we will directly replace Λ_i^z with $\tilde\Lambda_i^z$ in the coming estimations.

The quantities a_i, $i \in [n]$, are not negligible as the two others, we need to evaluate it. We can again inspire from Lemma 2.42 and write:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n b_i &= \frac{1}{z\tilde\Lambda_i^z n^2}\sum_{i=1}^n\sum_{l=1}^{m-1}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^l}x_ix_i^TQ_{-i}^{A_l^m}x_ix_i^T\tilde Q\right)\right]\\ &= \frac{1}{zn^2}\sum_{l=1}^{m-1}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^l}XTX^T\tilde Q\right)\right],\end{aligned}$$

(since $\check\kappa_z\kappa_z/|z| \leq 1$). Noting that $T^{A_l^m} \propto \mathcal{E}_2\left(\frac{(c\kappa_z)^{m-l}}{\sqrt{n}}\right)$ in $(\mathcal{D}_n, \|\cdot\|)$, one can use one more time Proposition **??** to obtain for any deterministic diagonal matrix $\tilde T^{A_l^m}$ such that $\|\mathbb{E}_{\mathcal{A}_Q}[T^{A_l^m}]-\tilde T^{A_l^m}\|_F \leq O\left(\frac{(c\kappa_z)^{m-l+1}}{\sqrt{n}}\right)$:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n b_i &= \frac{1}{nz}\sum_{l=1}^{m-1}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^l}X\tilde T^{A_l^m}X^T\tilde Q\right)\right] + O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right)\\ &= \frac{1}{n}\sum_{i=1}^n\frac{1}{\tilde\Lambda_i^z}\sum_{l=1}^{m-1}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^{l-1}}A_lQ_{-i}x_i\tilde T_i^{A_l^m}x_i^T\tilde Q\right)\right] + O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right)\\ &= \frac{1}{n}\sum_{i=1}^n\frac{1}{\tilde\Lambda_i^z}\sum_{l=1}^{m-1}\mathbb{E}_{\mathcal{A}_Q}\left[x_i^T\tilde QAQ_{-i}^{A_1^l}x_i\tilde T^{A_l^m}\right]\\ &\quad + \frac{1}{nz\tilde\Lambda_i^z}\sum_{i=1}^n\sum_{l=1}^{m-1}\sum_{k=1}^{l-1}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ^{A_1^k}x_ix_iQ_{-i}^{A_k^l}x_i\tilde T_i^{A_l^m}x_i^T\tilde Q\right)\right] + O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right),\end{aligned}$$

With the same resort to Proposition **??**, on can show iteratively that:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n b_i &= \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{m-1}\frac{1}{\tilde\Lambda_i^z}\sum_{l\in L_k^{m-1}}\mathbb{E}_{\mathcal{A}_Q}\left[x_i^T\tilde QAQ_{-i}^{A_1^{l_1}}x_i\tilde T_i^{A_{l_1}^{l_2}}\cdots\tilde T_i^{A_{l_k}^m}\right] + O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right)\\ &= \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{m-1}\frac{1}{\tilde\Lambda_i^z}\sum_{l\in L_k^{m-1}}\mathbb{E}_{\mathcal{A}_Q}\left[\mathrm{Tr}\left(AQ_{-i}^{A_1^{l_1}}\Sigma_i\tilde Q\right)\right]\tilde T_i^{A_{l_1}^{l_2}}\cdots\tilde T_i^{A_{l_k}^m} + O\left(\frac{(c\kappa_z)^m}{\sqrt{n}}\right),\end{aligned}$$

for some constant $c' > 0$ (each time one uses the proposition, the bounds of the estimation are the same and we use it $\sum_{k=1}^{m-1}\binom{k}{m-1}k \leq (m-1)2^{m-1}$). □

Given two integers $k, l > 0$, $h \leq l$ and a tuple $\alpha = (\alpha_{-h}, \dots, \dots, \alpha_0, \alpha_1, \dots, \alpha_{l-1}) \in [n]^{l+h}$, let us note:

$$\Psi_\alpha^h \equiv \mathbb{E}_{\mathcal{A}_Q}\left[\frac{1}{n}\mathrm{Tr}\left(\tilde\Sigma_{\alpha_{-h}}\tilde Q\cdots\tilde\Sigma_{\alpha_{-1}}\tilde Q\tilde\Sigma_{\alpha_0}Q\cdots\tilde\Sigma_{\alpha_{l-1}}Q\right)\right],$$

where for any $i \in [n]$, we noted $\tilde\Sigma_i \equiv \Sigma_i/\tilde\Lambda_i^z$. We further introduce the tensor of shape $n \times \cdots \times n$ (l times) $\Phi_l^h \equiv \left(\Phi_\alpha^h\right)_{\alpha\in[n]^l}$. Given $h > 0$ and replacing in the last estimation A with $\frac{1}{n}\tilde\Sigma_{\alpha_{1-h}}\tilde Q\cdots\tilde\Sigma_{\alpha_{-1}}\tilde Q\tilde\Sigma_{\alpha_0}$ satisfying:

$$\left\|\frac{1}{n}\tilde\Sigma_{\alpha_{1-h}}\tilde Q\cdots\tilde\Sigma_{\alpha_{-1}}\tilde Q\tilde\Sigma_{\alpha_0}\right\|_F \leq O\left(\frac{1}{\sqrt{n}}\right),$$

we know that for any $\alpha = (\alpha_{-h}, \dots, \alpha_0, \alpha_1, \dots, \alpha_{m-1}) \in [n]^{m+h}$:

$$\Psi_\alpha^h - \Psi_\alpha^{h+1} = \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^{m-1}\sum_{l\in L_k^{m-1}}\Psi^{h+1}_{\alpha_{-h}^{l_1},i}\Psi^0_{\alpha_{l_1}^{l_2},i}\cdots\Psi^0_{\alpha_{l_k}^m,i} + O\left(\frac{(c\kappa_z)^m}{n}\right),$$

where for any $k, l \in \{h, m-1\}$, $k < l$, $\alpha_k^l = \alpha_k, \dots, \alpha_{l-1}$.

Inspiring from this equation, we are going to introduce with next proposition a set of tensors that will allow us to approximate $\mathbb{E}[Q^{A_1^m}]$.

Proposition 2.47. *There exists a unique sequence of tensors Ψ satisfying:*

$$\begin{cases} \forall m > 2, h \in \mathbb{N}, \alpha = (\alpha_{-h}, \dots, \alpha_{m-1}) \in [n]^{m+h} : \\ \qquad \tilde{\Psi}^h_\alpha - \tilde{\Psi}^{h+1}_\alpha = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{m-1} \sum_{l \in L_k^{m-1}} \tilde{\Psi}^{h+1}_{\alpha^{l_1}_{-h}, i} \tilde{\Psi}^0_{\alpha^{l_2}_{l_1}, i} \cdots \tilde{\Psi}^0_{\alpha^m_{l_k}, i}, \\ \forall l \in [n], \alpha \in [n]^l : \; \tilde{\Psi}^{l-1}_\alpha = \tilde{\Psi}^l_\alpha = \frac{1}{n} \operatorname{Tr}\left(\tilde{\Sigma}_{\alpha_1} \tilde{Q} \cdots \tilde{\Sigma}_{\alpha_l} \tilde{Q} \right). \end{cases}$$

Proof. We already know that for any $i \in [n]$:

$$\tilde{\Psi}^0_{(i)} = \tilde{\Psi}^1_{(i)} = \frac{1}{n} \operatorname{Tr}(\tilde{\Sigma}_i \tilde{Q})$$

Let us now assume that there exists $m > 1$ such that for any $k \in [m-1]$, and any $h \in \{0, \dots, k\}$, and any $\alpha \in [n]^k$, $\tilde{\Psi}^h_\alpha$ is well defined and we know how to compute it. Besides, for any $\alpha \in [m]^n$, we also know how to compute $\tilde{\Psi}^{m-1}_\alpha = \tilde{\Psi}^m_\alpha$. We then further assume that there exist $h \in [m]$ such that we know how to compute any $\tilde{\Psi}^k_\alpha$ for $k < h$. If $h \geq 1$, then we see from the iteration formula that $\tilde{\Psi}^h_\alpha$ expresses as a sum of computable elements, the only issue raises when $h = 0$, then $\tilde{\Psi}^0_\alpha$ are appearing on both side of the equality. To invoke the invertibility of $\tilde{\Psi}^2_2$, let us first consider:

$$\tilde{\Psi}^0_\alpha = \tilde{\Psi}^1_\alpha + \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^{m-1} \tilde{\Psi}^1_{\alpha^l_0, i} \tilde{\Psi}^0_{\alpha^m_l, i} + \frac{1}{n} \sum_{i=1}^n \sum_{k=2}^{m-1} \sum_{l \in L_k^{m-1}} \tilde{\Psi}^1_{\alpha^{l_1}_{-h}, i} \cdots \tilde{\Psi}^0_{\alpha^{m-h}_{l_k}, i}$$

The right hand term that raise concern is $\sum_{i=1}^n \tilde{\Psi}^1_{\alpha^1_0, i} \tilde{\Psi}^0_{\alpha^m_1, i}$. Since $\tilde{\Psi}^0_{\alpha^m_1, i} = \tilde{\Psi}^0_{i, \alpha^m_1}$ and $\tilde{\Psi}^1_2 = \tilde{\Psi}^2_2$, we can rewrite it $(\tilde{\Psi}^2_2 \tilde{\Psi}^0_m)_\alpha$ (where the right product of a matrix and a tensor is a classical matricial product on the first variable of the tensor). We can bound for any $\alpha \in [n]^2$:

$$\left| \tilde{\Psi}^2_\alpha \right| \leq \frac{\|\Sigma_{\alpha_1}\| \|\Sigma_{\alpha_2}\|}{|z|^2},$$

therefore, for $|z|$ sufficiently big, $(I_n - \tilde{\Psi}^2_2)$ is invertible and one has the identity:

$$\tilde{\Psi}^0_m = (I_n - \tilde{\Psi}^2_2)^{-1} \cdot \left(\tilde{\Psi}^1_\alpha + \frac{1}{n} \sum_{i=1}^n \sum_{l=2}^{m-1} \tilde{\Psi}^1_{\alpha^l_0, i} \tilde{\Psi}^0_{\alpha^m_l, i} \right. \tag{2.24}$$

$$\left. + \frac{1}{n} \sum_{i=1}^n \sum_{k=2}^{m-1} \sum_{l \in L_k^{m-1}} \tilde{\Psi}^1_{\alpha^{l_1}_0, i} \tilde{\Psi}^0_{\alpha^{l_2}_{l_1}, i} \cdots \tilde{\Psi}^0_{\alpha^m_{l_k}, i} \right)_{(\alpha_0, \dots, \alpha_{m-1}) \in [n]^m}. \tag{2.25}$$

Complex analysis inferences should then allow us to set the uniqueness for all the values of z.

□

Once $\tilde{\Psi}^h_m$ is defined for all $m \in \mathbb{N}$, $h \in [m]$, one can define iteratively for any sequence of deterministic matrices $A_{-h}, \dots, A_{m-1}$ followingly:

$$\begin{cases} \forall m > 2, h \in \mathbb{N}, \forall A_{-h}, \dots, A_{m-1} \in \mathcal{M}_p^{m+h} : \\ \quad \tilde{\Psi}^h_{A_{-h}, \dots, A_{m-1}} - \tilde{\Psi}^{h+1}_{A_{-h}, \dots, A_{m-1}} \\ \qquad = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{m-1} \sum_{l \in L_k^{m-1}} \tilde{\Psi}^{h+1}_{A_{-h}, \dots, A_{l_1 - 1}, \Sigma_i} \tilde{\Psi}^0_{A_{l_1}, \dots, A_{l_2 - 1}, \Sigma_i} \cdots \tilde{\Psi}^0_{A_{l_k}, \dots, A_{m-1}, \Sigma_i}, \\ \forall l \in [n], A_1, \dots, A_l \in \mathcal{M}_p^l : \; \tilde{\Psi}^{l-1}_{A_{-h}, \dots, A_{m-1}} = \tilde{\Psi}^l_{A_{-h}, \dots, A_{m-1}} = \frac{1}{n} \operatorname{Tr}\left(A_1 \tilde{Q} \cdots A_{l-1} \tilde{Q} A_l \bar{\tilde{Q}} \right). \end{cases}$$

One should need a proper proof to justify it, but we will just explain that the existence and uniqueness is proven iteratively considering sequences of matrices $(A_1, \ldots, A_i, \tilde{\Sigma}_{\alpha_{i+1}}, \ldots, \tilde{\Sigma}_{\alpha_m})$ with $(\alpha_{i+1}, \ldots, \alpha_m) \in [n]^{m-i}$ and i going from 1 to m.

We can finally properly estimate $Q^{\mathcal{A}_1^m}$. We express this result in the case m constant, but one could probably get a similar result for an asymptotic m. The proof is just a laborious use of the explicit formulation (2.24).

Proposition 2.48. *Given a constant integer $m \in \mathbb{N}$ and m matrices $A_1, \ldots, A_m$ satisfying $\|A_i\|_F \leq 1$, we can estimate:*

$$\operatorname{Tr}(A_m Q^{A_1^m}) \mid \mathcal{A}_Q \in \tilde{\Psi}^0_{A_1,\ldots,A_m} \pm \mathcal{E}_2\left(\frac{\kappa_z}{\sqrt{n}}\right).$$

Chapter 3

Concentration of the resolvent with random diagonal term

We study here the concentration of a resolvent $Q = (I_p - \frac{1}{n}XDX^T)^{-1}$ with Assumptions 1-5 for X and D (in particular D is random). Among other use, this object appears when studying robust regression **??**. In several settings, robust regression can be expressed by the following fixed point equation:

$$\beta = \frac{1}{n}\sum_{i=1}^{n} f(x_i^T\beta)x_i, \quad \beta \in \mathbb{R}^p, \tag{3.1}$$

where β is the weight vector performing the regression (to classify data, for instance). It was then shown in **?** that the estimation of the expectation and covariance of β (and therefore, of the performances of the algorithm) rely on an estimation of Q, with $D = \mathrm{Diag}(f'(x_i^T\beta))$. To obtain a sharp concentration on Q (as it is done in Theorem 3.1 below), one has to understand the dependence between Q and x_i, for all $i \in [n]$. This is performed with the notation, given for any $M = (m_1, \ldots, m_n) \in \mathcal{M}_{p,n}$ or any $\Delta = \mathrm{Diag}_{i\in[n]}(\Delta_i) \in \mathcal{D}_n$:

- $M_{-i} = (m_1, \ldots, m_{i-1}, 0, m_{i+1}, \ldots, m_n) \in \mathcal{M}_{p,n}$,
- $\Delta_{-i} = \mathrm{Diag}(\Delta_1, \ldots, \Delta_{i-1}, 0, \Delta_{i+1}, \ldots, \Delta_n) \in \mathcal{D}_n$.

Theorem 3.1. *Given a positive random diagonal matrix $D \in \mathcal{D}_n^+$ and a random matrix $X = (x_1, \ldots, x_n)$, in the regime*[1] $p \leq O(n)$ *and under the assumptions:*

- $(X, D) \propto \mathcal{E}_2$,
- *all the columns $x_1, \ldots, x_n$ are independent,*
- $O(1) \leq \sup_{i\in[n]} \|\mathbb{E}[x_i]\| \leq O(1)$,
- *for all $i \in [n]$, there exists a random positive diagonal matrix $D^{(i)} \in \mathcal{D}_n^+$, independent of x_i, such that* $\sup_{i\in[n]} \|D_{-i} - D_{-i}^{(i)}\|_F \leq O(1)$,
- *there exist*[2] *three constants $\kappa, \kappa_D, \varepsilon > 0$ ($\varepsilon \geq O(1)$ and $\kappa, \kappa_D \leq O(1)$), such that $\|X\| \leq \sqrt{n}\kappa$, $\|D^{(i)}\|, \|D\| \leq \kappa_D$ and $\kappa^2\kappa_D \leq 1 - \varepsilon$,*

[1] It is not necessary to assume that $p \leq O(n)$ but it simplifies the concentration result (if $p \gg n$, the concentration is not as good, but it can still be expressed).

[2] The assumptions $\|X\|/\sqrt{n}$ bounded and $\kappa^2\kappa_D \leq 1 - \varepsilon$ might look a bit strong (since it is not true for matrices with i.i.d. Gaussian entries), it is indeed enough to assume that $\mathbb{E}[\|X\|] \leq O(\sqrt{n})$ and introduce a parameter $z > 0$ to study the behavior of $(zI_p - \frac{1}{n}XDX^T)^{-1}$ when z is far from the spectrum of $\frac{1}{n}XDX^T$ – as it is done in Section 2.1. We however preferred here to make a relatively strong hypothesis not to have supplementary notations and proof precautions, that might have blurred the message.

the resolvent $Q \equiv (I_p - \frac{1}{n}XDX^T)^{-1}$ *follows the concentration*

$$Q \in \mathbb{E}[Q] \pm \mathcal{E}_{3/2}\left(\sqrt{\frac{\log n}{n}}\right) \qquad \text{in } (\mathcal{M}_p, \|\cdot\|_F).$$

Inspiring from the formulation of the deterministic equivalent introduced in Chapter 2, we introduce the following notation that will help us to express the deterministic equivalent of Q. Given $\delta, D \in \mathcal{M}_n$, we note:

$$\tilde{Q}^\delta(D) \equiv \left(I_p - \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\frac{D_i}{1+\delta D_i}\right]\Sigma_i\right)^{-1},$$

(where we recall that $\forall i \in [n]$, $\Sigma_i \equiv \mathbb{E}[x_i x_i^T]$).

Theorem 3.2. *For any diagonal matrix* $D' \in \mathcal{D}_n$*, the fixed point equation:*

$$\delta = \mathrm{Diag}_{i\in[n]}\left(\Sigma_i \tilde{Q}^\delta(D')\right)$$

admits a unique solution $\delta(D') \in \mathcal{D}_n$ *and, under the hypotheses of Theorem 3.1, one can estimate:*

$$\|\mathbb{E}[Q] - \tilde{Q}^{\delta(D)}(D)\|_F \leq O(\sqrt{\log n}).$$

Remark 3.3. *Note here that although there are no term* $1/\sqrt{n}$ *in the bound, this result still provides some good convergence bound for the estimation of linear forms like the Stieltjes transform or more generally any* $\frac{1}{n}\operatorname{Tr}(AQ)$ *when* $\|A\| \leq 1$ *(because then* $\frac{1}{n}\|A\|_F \leq O(1/\sqrt{n})$*).*

Remark 3.4. *Unlike in the result of Chapter 2 (for instance Theorem 2.25), there is no resort to the high probability event* $\mathcal{A}_Q$ *here. It is because, as we will see later in Lemma 3.6, in the current setting all the drawings of the resolvent* Q *are bounded thanks to the hypothesis* $\kappa^2\kappa_D \leq 1-\varepsilon$*. In the sens, the parameter* z *of Chapter 2 is here always far from the spectrum of* $\frac{1}{n}XX^T$*.*

Remark 3.5. *Let us give two examples of the matrices* $D^{(i)}$ *that one could encounter in practice:*

- *For all* $i \in [n]$, $D_i = f(x_i)$ *for* $f : \mathbb{R}^p \to \mathbb{R}$*, bounded, then,* D_i *just depends on* x_i *so one can merely take* $D^{(i)} = D_{-i}$ *for all* $i \in [n]$*.*
- *For the robust regression described by Equation 3.1, as in ?, we can assume for simplicity*[3] $\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty \leq O(1)$*. If we choose* $D = \mathrm{Diag}(f'(x_i^T\beta))$*, then it is convenient to assume* $\frac{1}{n}\|f'\|_\infty\|X\|^2 \leq 1-\varepsilon$ *(which implies in particular* $\frac{1}{n}\|X\|^2\|D\| \leq 1-\varepsilon$*) so that* β *is well defined, being solution of a contractive fixed point equation. One can further introduce* $\beta^{(i)} \in \mathbb{R}^p$*, the unique solution to*

$$\beta^{(i)} = \frac{1}{n}\sum_{\substack{1\leq j\leq n\\ j\neq i}} f(x_j^T\beta^{(i)})x_j.$$

By construction, $\beta^{(i)}$ *is independent of* x_i *and so is:*

$$D^{(i)} \equiv \mathrm{Diag}\left(f'(x_1^T\beta^{(i)}), \ldots, f'(x_{i-1}^T\beta^{(i)}), 0, f'(x_{i+1}^T\beta^{(i)}), \ldots, f'(x_n^T\beta^{(i)})\right).$$

Besides $\|D_{-i} - D_{-i}^{(i)}\|_F \leq \|f''\|_\infty\|X_{-i}^T(\beta-\beta^{(i)})\|_F$*. Now, the identities:*

$$X_{-i}^T\beta = \frac{1}{n}X_{-i}^TXf(X^T\beta) \qquad \text{and} \qquad X_{-i}^T\beta^{(i)} = \frac{1}{n}X_{-i}^TX_{-i}f(X_{-i}^T\beta^{(i)})$$

(where f *is applied entry-wise) imply:*

$$\|X_{-i}^T(\beta-\beta^{(i)})\|_F \leq \frac{1}{n}\|f'\|_\infty\|X_{-i}\|^2\|X_{-i}^T(\beta-\beta^{(i)})\|_F + \frac{1}{n}f(x_i^T\beta)X_{-i}^Tx_i.$$

We can then deduce (since $\frac{1}{n}\|f'\|_\infty\|X_{-i}\|^2 \leq 1-\varepsilon$ *by hypothesis):*

$$\|D_{-i} - D_{-i}^{(i)}\|_F \leq \|f''\|_\infty\|X_{-i}^T(\beta-\beta^{(i)})\|_F \leq \frac{\|f''\|_\infty}{n\varepsilon}f(x_i^T\beta)X_{-i}^Tx_i \leq O(1).$$

[3]The bound $\|f\|_\infty \leq O(1)$ is not necessary to set the concentration of Q, but it avoids a lot of complications.

3.1 Concentration of Q

Lemma 3.6. *Under the assumptions of Theorem 3.1,* $\|Q\| \leq \frac{1}{\varepsilon} \leq O(1)$.

Then we can show a Lipschitz concentration of Q but with looser observable diameter that the one given by Theorem 3.1 (as for XDX^T, we get better concentration speed in the linear concentration framework).

Lemma 3.7. *Under the hypotheses of Theorem 3.1:*

$$\left(Q, \frac{1}{\sqrt{n}} QX\right) \propto \mathcal{E}_2 \quad in \quad (\mathcal{M}_{p,n}, \|\cdot\|_F).$$

Proof. Let us just show the concentration of the resolvent, the tuple is treated the same way. If we note $\phi(X, D) = Q$ and we introduce $X' \in \mathcal{M}_{p,n}$ and $D' \in \mathcal{D}_n$, satisfying $\|X'\| \leq \kappa\sqrt{n}$ and $\|D'\| \leq \kappa_D$ as X, D, we can bound:

$$\begin{aligned}
&\|\phi(X, D) - \phi(X', D)\|_F \\
&\quad = \frac{1}{n} \left\|\phi(X, D)(X - X')DX^T\phi(X', D)\right\|_F + \frac{1}{n} \left\|\phi(X, D)X'D(X - X')^T\phi(X', D)\right\|_F \\
&\quad \leq \frac{2\kappa\kappa_D}{\varepsilon^2\sqrt{n}} \|X - X'\|_F \,,
\end{aligned}$$

thanks to the hypotheses and Lemma 3.6 given above. The same way, we can bound:

$$\|\phi(X, D) - \phi(X, D')\|_F \leq \frac{\kappa^2}{\varepsilon^2} \|D - D'\|_F$$

Therefore, as a $O(1)$-Lipschitz transformation of (X, D), $Q \propto \mathcal{E}_2$. □

3.2 Control on the dependency on x_i

The dependence between Q and x_i prevent us from bounding straightforwardly $\|Qx_i\|$ with Lemma 3.6 and the hypotheses on x_i. We can still disentangle this dependence thanks to the notations:

$$Q_{-i} = \left(I_p - \frac{1}{n} X_{-i}^T D X_{-i}^T\right)^{-1} \qquad \text{and} \qquad Q_{-i}^{(i)} = \left(I_p - \frac{1}{n} X_{-i}^T D^{(i)} X_{-i}^T\right)^{-1}.$$

We can indeed bound:

$$\|\mathbb{E}[Q_{-i}^{(i)} x_i]\| \leq \|\mathbb{E}[Q_{-i}^{(i)}]\mathbb{E}[x_i]\| \leq O(1), \tag{3.2}$$

and we even have interesting concentration properties that will be important later:

Lemma 3.8. *Under the assumptions of Theorem 3.1:*

$$Q_{-i}^{(i)} x_i, \ \frac{1}{\sqrt{n}} X_{-i}^T Q_{-i}^{(i)} x_i \in O(1) \pm \mathcal{E}_2.$$

Proof. Considering $u \in \mathbb{R}^p$, deterministic such that $\|u\| \leq 1$, we can bound thanks to the independence between $Q_{-i}^{(i)}$ and x_i:

$$\left|u^T Q_{-i}^{(i)} x_i - \mathbb{E}[u^T Q_{-i}^{(i)} x_i]\right| \leq \left|u^T Q_{-i}^{(i)} (x_i - \mathbb{E}[x_i])\right| + \left|u^T \left(Q_{-i}^{(i)} - \mathbb{E}[Q_{-i}^{(i)}]\right) \mathbb{E}[x_i]\right|.$$

Therefore, the concentrations $x_i \propto \mathcal{E}_2$ and $Q_{-i}^{(i)} \propto \mathcal{E}_2$ given in Lemma 3.7 imply that there exist two constants $C, c > 0$ such that $\forall t > 0$ such that if we note $\mathcal{A}_{-i}$, the sigma algebra generated by X_{-i} (it is independent with x_i):

$$\begin{aligned}
&\mathbb{P}\left(\left|u^T Q_{-i}^{(i)} x_i - \mathbb{E}[u^T Q_{-i}^{(i)} x_i]\right| \geq t\right) \\
&\quad \leq \mathbb{E}\left[\mathbb{P}\left(\left|u^T Q_{-i}^{(i)} (x_i - \mathbb{E}[x_i])\right| \geq \frac{t}{2} \mid \mathcal{A}_{-i}\right)\right] + \mathbb{P}\left(\left|u^T \left(Q_{-i}^{(i)} - \mathbb{E}\left[Q_{-i}^{(i)}\right]\right) \mathbb{E}[x_i]\right| \geq \frac{t}{2}\right) \\
&\quad \leq \mathbb{E}\left[Ce^{(t/c\|Q_{-i}^{(i)}\|)^2}\right] + Ce^{(t/c\|\mathbb{E}[x_i]\|)^2} \leq C'e^{-t^2/c'},
\end{aligned}$$

for some constants $C', c' > 0$, thanks to the bounds $\|\mathbb{E}[x_i]\| \leq O(1)$ given in the assumptions and $\|Q_{-i}^{(i)}\| \leq O(1)$ given by Lemma 3.6.

The linear concentration of $X_{-i}^T Q_{-i}^{(i)} x_i/\sqrt{n}$ is proven the same way since one can show as in Lemma 3.7 that $(X, D) \mapsto X_{-i}^T Q_{-i}^{(i)}/\sqrt{n}$ is $O(1)$-Lipschitz on $\{\|X\| \leq \kappa\sqrt{n}, \|D\| \leq \kappa_D\}$, and therefore, $X_{-i}^T Q_{-i}^{(i)}/\sqrt{n} \propto \mathcal{E}_2$. □

Let us adapt the Schur identity to the presence of the diagonal matrix D_i:

$$Q = Q_{-i} - \frac{1}{n}\frac{D_i Q_{-i} x_i x_i^T Q_{-i}}{1 + D_i \Delta_i} \qquad \text{and} \qquad Qx_i = \frac{Q_{-i} x_i}{1 + D_i \Delta_i},$$

where we noted $\Delta_i \equiv \frac{1}{n} x_i^T Q_{-i} x_i$. The link between $Q_{-i} x_i$ and $Q_{-i}^{(i)} x_i$ is made thanks to:

Lemma 3.9. *Under the hypotheses of Theorem 3.1, for all $i \in [n]$:*

$$\|Q_{-i} x_i - Q_{-i}^{(i)} x_i\| \in O(\sqrt{\log n}) \pm \mathcal{E}_2(\sqrt{\log n}).$$

Let us first prove a Lemma of independent interest:

Lemma 3.10. *Under the hypotheses of Theorem 3.1, for all $i \in [n]$*

$$\left\| \frac{1}{\sqrt{n}} X_{-i}^T Q_{-i}^{(i)} x_i \right\|_\infty \in O(\sqrt{\log n}) \pm \mathcal{E}_2(\sqrt{\log n}).$$

Proof. The control on the variation is given by Lemma 3.8 ($\|\cdot\|_\infty \leq \|\cdot\|_F$) and the bound on the expectation is a consequence of Proposition **??** and the bound:

$$\frac{1}{\sqrt{n}} \left\| \mathbb{E}\left[X_{-i}^T Q_{-i}^{(i)} x_i\right] \right\|_\infty \leq \frac{1}{\sqrt{n}} \left\| \mathbb{E}\left[X_{-i}^T Q_{-i}^{(i)}\right] \mathbb{E}\left[x_i\right] \right\| \leq O(1). \tag{3.3}$$

□

Proof of Lemma 3.9. Let us bound directly:

$$\begin{aligned}
\left\| \left(Q_{-i} - Q_{-i}^{(i)}\right) x_i \right\| &\leq \left\| \frac{1}{n} Q_{-i} X_{-i} (D_{-i}^{(i)} - D_{-i}) X_{-i}^T Q_{-i}^{(i)} x_i \right\| \\
&\leq \frac{1}{n} \|Q_{-i} X_{-i}\| \|D_{-i}^{(i)} - D_{-i}\|_F \|X_{-i}^T Q_{-i}^{(i)} x_i\|_\infty \ \leq O\left(\frac{1}{\sqrt{n}} \|X_{-i}^T Q_{-i}^{(i)} x_i\|_\infty\right).
\end{aligned}$$

We can then conclude thanks to Lemma 3.10. □

3.3 Proof of the concentration

Let us first provide a preliminary result that will allow us to set that $x_i^T QAQx_i$ behaves more or less like a $O(\sqrt{\log n})$-Lipschitz observation of (X, D, Y).

Lemma 3.11. *Under the hypotheses of Theorem 3.1, $\forall i \in [n]$, and for any deterministic matrices $U, V \in \mathcal{M}_p$ such that $\|U\|, \|V\| \leq 1$:*

$$\|VQX\|_\infty \in O\left(\sqrt{\log n}\right) \pm \mathcal{E}_2\left(\sqrt{\log n}\right).$$

Be careful that the bound would not have been so tight for $\|QXU\|_\infty$ given $U, V \in \mathcal{M}_n$. We just need a small lemma to be able to bound $1 + \Delta_i D_i$ from below, it is basically a rewriting of Lemma 2.7 bounding Λ in Chapter 2.

Lemma 3.12. *Under the hypotheses of Theorem 3.1, $\forall i \in [n]$:*

$$|\Delta_i| \leq \frac{\kappa^2}{\varepsilon} \qquad \textit{and} \qquad \varepsilon \leq 1 + D_i \Delta_i \leq 1 + \frac{\kappa \kappa_D}{\epsilon}$$

Proof. Let us simply bound for any $i \in [n]$: $\|\Delta_i\| = \frac{1}{n}|x_i^T Q x_i| \leq \frac{\kappa^2}{\varepsilon}$ thanks to Lemma 3.6. We can then directly deduce the upper bound of $|1 + D_i\Delta_i|$.

For the lower bound, let us introduce again the matrix $\check{Q} = (I_n - D^{1/2}X^TXD^{1/2})^{-1}$, we can bound as in Lemma 3.6 $\|\check{Q}\| \leq \frac{1}{\varepsilon}$, and we can show again that:

$$1 + D_i\Delta_i = 1 + \frac{D_i}{n} x_i^T Q x_i = \frac{1}{\check{Q}_i},$$

which allows us to bound $|1 + D_i\Delta_i| \geq \frac{1}{\|\check{Q}\|} \geq \varepsilon$. □

Proof. Following the same identities and arguments presented in the proof of Lemma 3.9, we can bound thanks to Lemma 3.12

$$\begin{aligned}\|VQX\|_\infty &= \sup_{i\in[n]} \left\| \frac{VQ_{-i}^{(i)}x_i + \frac{1}{n}V(Q_{-i} - Q_{-i}^{(i)})x_i}{1 + D_i\Delta_i} \right\|_\infty \\ &\leq O\left(\sup_{i\in[n]} \left(\|VQ_{-i}^{(i)}x_i\|_\infty, \frac{1}{\sqrt{n}} \|X_{-i}^T Q_{-i}^{(i)}x_i\|_\infty \right) \right).\end{aligned}$$

Introducing, as in Section ??, $(e_1, \ldots, e_p)$ and $(f_1, \ldots, f_n)$, respectively, the canonical basis of $\mathbb{R}^p$ and $\mathbb{R}^n$ we know from Lemma 3.8 that for all $k \in [p]$ and $i, j \in [n]$:

$$e_k^T VQ_{-i}^{(i)}x_i \in O(1) \pm \mathcal{E}_2 \qquad \text{and} \qquad \frac{1}{\sqrt{n}} f_j^T X_{-i}^T Q_{-i}^{(i)} x_i \in O(1) \pm \mathcal{E}_2,$$

since $|\mathbb{E}[e_k^T VQ_{-i}^{(i)}x_i]| \leq \|\mathbb{E}[Q_{-i}^{(i)}]\|\|\mathbb{E}[x_i]\| \leq O(1)$ and similarly, $|\mathbb{E}[f_j^T X_{-i}^T Q_{-i}^{(i)} x_i]/\sqrt{n}| \leq O(1)$. Following the arguments displayed in Section ??, there exist four constants K, C, c, c' (all $\leq O(1)$) such that we can bound:

$$\begin{aligned}\mathbb{P}(\|VQX\|_\infty \geq t) &\leq \mathbb{P}\left(\sup_{\substack{i,j\in[n]\\ k\in[p]}} e_k^T VQ_{-i}^{(i)}x_i + \frac{1}{\sqrt{n}} f_j^T X_{-i}^T Q_{-i}^{(i)} x_i \geq \frac{t}{K} \right) \\ &\leq \max\left(1, n^2pCe^{-t^2/c}\right) \quad \leq \max(e, C)e^{-K^2t^2/c' \log(n^2p)}.\end{aligned}$$

We can then deduce the concentration of $\|VQX\|_\infty$ since $\log(n^2p) \leq O(\log(n))$. □

Let us now prove three results of progressive difficulty. Since $Q \propto \mathcal{E}_2$ and $x_i \propto \mathcal{E}_2$, one can follow the lines of Lemma 2.40 to set:

Lemma 3.13. *Under the hypotheses of Theorem 3.1:*

$$\Delta_i \equiv \frac{1}{n} x_i^T Q_{-i} x_i \in \bar{\Delta}_i \pm \mathcal{E}_2\left(\sqrt{\frac{1}{n}}\right) + \mathcal{E}_1\left(\sqrt{\frac{1}{n}}\right),$$

where we noted $\bar{\Delta}_i \equiv \mathbb{E}[\frac{1}{n} x_i^T Q_{-i} x_i]$ *(recall that* $\forall i \in [n]: \ |\bar{\Delta}_i| \leq \frac{\kappa^2}{\varepsilon}$.

We let the reader refer to the proof of Lemma 2.40 or more simply to next result to get justifications for this lemma.

Lemma 3.14. *Under the hypotheses of Theorem 3.1, given a deterministic matrix* $A \in \mathcal{M}_{p,n}$*:*

$$x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i \in O(\sqrt{n}\|A\|_F) \pm \mathcal{E}_2(\|A\|_F) + \mathcal{E}_1(\|A\|).$$

Proof. Let us bound:

$$\begin{aligned}&\left| x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i - \mathbb{E}\left[x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i \right] \right| \\ &\quad\leq \left| x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i - \mathrm{Tr}\left(\Sigma_i Q_{-i}^{(i)} A Q_{-i}^{(i)}\right) \right| + \left| \mathrm{Tr}\left(\Sigma_i Q_{-i}^{(i)} A Q_{-i}^{(i)}\right) - \mathrm{Tr}\left(\Sigma_i \mathbb{E}\left[Q_{-i}^{(i)} A Q_{-i}^{(i)}\right]\right) \right|\end{aligned}$$

One can then deduce the result from Lemma ?? (setting that if $Z \in O(\sigma \pm \mathcal{E}_q(\sigma)$ and $0 \leq Y \leq Z$, then $Y \in O(\sigma) \pm \mathcal{E}_q(\sigma))$ applied to the concentrations:

- $x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i \mid \mathcal{A}_{-i} \in \mathrm{Tr}\left(\Sigma_i Q_{-i}^{(i)} A Q_{-i}^{(i)}\right) \pm \mathcal{E}_2(\|A\|_F/\varepsilon^2) + \mathcal{E}_1(\|A\|/\varepsilon^2)$ thanks to Hanson-Wright inequality (Proposition ??),
- $|\mathrm{Tr}\left(\Sigma_i Q_{-i}^{(i)} A Q_{-i}^{(i)}\right)|$ is a $(\frac{1}{\varepsilon}\|\Sigma_i\|\|A\|_F)$-Lipschitz transformation of $Q_{-i}^{(i)}$, and therefore, one can deduce from Lemma 3.7 that $\mathrm{Tr}\left(\Sigma_i Q_{-i}^{(i)} A Q_{-i}^{(i)}\right) \in \mathrm{Tr}\left(\Sigma_i \mathbb{E}\left[Q_{-i}^{(i)} A Q_{-i}^{(i)}\right]\right) \pm \mathcal{E}_2(\|A\|_F)$.

Besides, one can bound:

$$\left|\mathbb{E}\left[x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i\right]\right| = \left|\mathrm{Tr}\left(\Sigma_i \mathbb{E}\left[Q_{-i}^{(i)} A Q_{-i}^{(i)}\right]\right)\right| \leq O(\sqrt{p}\|A\|_F).$$

□

Lemma 3.15. *Under the hypotheses of Theorem 3.1, given a deterministic matrix $A \in \mathcal{M}_{p,n}$ such that $\|A\|_F \leq 1$:*

$$\frac{1}{n}\|X^T QAQX\|_d \in O\left(\frac{1}{\sqrt{n}}\right) \pm \mathcal{E}_1\left(\sqrt{\frac{\log n}{n}}\right).$$

This Lemma in particular gives us the concentration of any diagonal term of the random matrix $\frac{1}{n}X^T QAQX$, i.e. of any $\frac{1}{n}x_i^T QAQx_i$, $i \in [n]$.

Proof. To prove the concentration, let us introduce again the decomposition $A = U^T \Lambda V$, with $U, V \in \mathcal{O}_p$ and $\Lambda \in \mathcal{D}_p$. We are going to bound the variation of $\frac{1}{n}\|X^T QAQX\|_d$ towards the variations of $\frac{1}{\sqrt{n}}VQX \propto \mathcal{E}_2$ (see Lemma 3.7). Let us define the mapping $\phi : \mathcal{M}_{p,n}^2 \to \mathbb{R}$ satisfying for all $M, P \in \mathcal{M}_{p,n}$, $\phi(M,P) = \|M^T \Lambda P\|_d$ (with that definition, $\frac{1}{n}\|X^T QAQX\|_d = \phi(\frac{1}{\sqrt{n}}VQX, \frac{1}{\sqrt{n}}UQX))$. Given 4 variables M, P, M', P' satisfying $\|M\|, \|P\|, \|M'\|, \|P'\| \leq \frac{\kappa}{\varepsilon}$ we can bound as in the proof of Corollary ??:

$$|\phi(M,P) - \phi(M',P)| \leq \|(M-M')^T \Lambda P\|_d \leq \|M-M'\|_F \|P\|_\infty \|\Lambda\|_F \leq \|M-M'\|_F \|P\|_\infty,$$

and the same way, $|\phi(M,P) - \phi(M,P')| \leq \|P-P'\|_F \|M\|_\infty$. We further invoke Lemma 3.11 that provides the concentration:

$$\left(\frac{1}{\sqrt{n}}\|VQX\|_\infty, \frac{1}{\sqrt{n}}\|UQX\|_\infty\right) \in O\left(\sqrt{\frac{\log n}{n}}\right) \pm \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right). \tag{3.4}$$

We can then deduce from Theorem ?? the concentration $\sqrt{\frac{n}{\log n}}\phi(\frac{1}{\sqrt{n}}VQX, \frac{1}{\sqrt{n}}UQX) \propto \mathcal{E}_2 + \mathcal{E}_1 \propto \mathcal{E}_1$, from which we deduce the concentration of $\|X^T QAQX\|_d$.

For the estimation, let us first express:

$$x_i^T QAQx_i = x_i^T Q_{-i} A Q_{-i} x_i (1 + D_i \Delta_i)$$

Thanks to Lemmas 3.16, 3.14 and the assumptions on the concentration of D_i, one can bound:

$$\left\|x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i (1 + D_i \Delta_i)\right\| \leq O(\sqrt{n}).$$

To be able to replace $Q_{-i}^{(i)}$ with Q_{-i} in the previous inequality, one can bound:

$$\begin{aligned}
&\left|x_i^T QAQx_i - (1 + (1 + D_i\Delta_i))\left(x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i\right)\right| \\
&\quad \leq \left(1 + \frac{\kappa_D \kappa^2}{\varepsilon}\right)\left(\left|x_i^T\left(Q_{-i} - Q_{-i}^{(i)}\right) A \left(Q_{-i} - Q_{-i}^{(i)}\right) x_i\right| + \left|x_i^T\left(Q_{-i} - Q_{-i}^{(i)}\right) A Q_{-i}^{(i)} x_i\right|\right. \\
&\qquad \left. + \left|x_i^T Q_{-i}^{(i)} A \left(Q_{-i} - Q_{-i}^{(i)}\right) x_i\right|\right)
\end{aligned}$$

We can then invoke the concentrations:

- $\|x_i^T Q_{-i}^{(i)} A\|, \|AQ_{-i}^{(i)} x_i\| \in O(1) \pm \mathcal{E}_2$ thanks to Lemmas 3.8 and ??.
- $\|(Q_{-i} - Q_{-i}^{(i)})x_i\|, \|x_i(Q_{-i} - Q_{-i}^{(i)})\| \in O(\sqrt{\log n}) \pm \mathcal{E}_2(\sqrt{\log n})$ thanks to Lemma 3.9
- $\left|x_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i\right| \in O(\sqrt{n}) \pm \mathcal{E}_1$ thanks to Lemma 3.14.

to be able to finally bound $\mathbb{E}[\frac{1}{n}\|X^T QAQX\|_d] \leq O(\sqrt{n})$ thanks to (3.4) (with $U = V = I_p$).

□

Proof of Theorem 3.1. Let us consider $A \in \mathcal{M}_{p,n}$ such that $\|A\|_F \leq 1$ and let us note $\phi(X, D) = \mathrm{Tr}(AQ)$. We abusively work with X, D and independent copies X', D' satisfying $\|X\|, \|X'\| \leq \sqrt{n}\kappa$ and $\|D\|, \|D'\| \leq \kappa_D$ as if they were deterministic variables, and we note $Q'_X \equiv \phi(X', D)$, $Q'_D \equiv \phi(X, D')$. Let us bound the variations

$$|\phi(X, D) - \phi(X', D)| = \frac{1}{n} \left|\mathrm{Tr}\left(AQ(X - X')DXQ'_X\right)\right| \leq \frac{\kappa\kappa_D}{\varepsilon^2\sqrt{n}} \|X - X'\|_F.$$

We can also bound as in the proof of Proposition ??:

$$|\phi(X, D) - \phi(X, D')| \leq \frac{1}{n} \left\|XQ'_D AQX\right\|_d \left\|D - D'\right\|_F.$$

The concentration $1/\sqrt{n\log n} \left\|XQ'_D AQX\right\|_d \in O(1/\sqrt{\log n}) \pm \mathcal{E}_1$ is provided by Lemma 3.15 (actually Lemma 3.15 gives the concentration of $\|XQAQX\|_d$, but the proof remains the same if one replaces one of the Q with Q'_D, for a diagonal matrix D', independent with D). To simplify the use of Theorem ??, let us note $\tilde{\phi} \equiv \sqrt{\frac{n}{\log n}}\phi$. Then we have:

- $\left|\tilde{\phi}(X, D) - \tilde{\phi}(X', D)\right| \leq \Psi_1(X, X', D) \left\|X - X'\right\|_F$ with $\Psi_1(X, D, D') \propto \mathcal{E}_2 + \mathcal{E}_1$,
- $\left|\tilde{\phi}(X, D) - \tilde{\phi}(X, D')\right| \leq \Psi_2(X, D, D') \left\|D - D'\right\|_F$ with $\Psi_2(X, D, D') \propto \mathcal{E}_2 + \mathcal{E}_1$,

and we can add an other inequality with an imaginary variable to be able to apply Theorem ?? with $m = 3$, $\sigma = \mu_1 = \mu_2 = \mu_3 = 1$ which gives us $\tilde{\phi}(X, D) \propto \mathcal{E}_2 + \mathcal{E}_1 + \mathcal{E}_{3/2}$ from which we can deduce the concentration of $\mathrm{Tr}(AQ) = \sqrt{\log n/n}\tilde{\phi}(X, D)$. □

3.4 Proof of the estimation

Lemma 3.16. *Under the hypotheses of Theorem 3.1, given a deterministic matrix $A \in \mathcal{M}_{p,n}$ such that $\|A\|_F \leq 1$:*

$$x_i^T QAQx_i(1 + D_i\bar{\Delta}_i) \in O(\sqrt{n}) \pm \mathcal{E}_{2/3}\left(\sqrt{\log n}\right).$$

Proof. Let us first show the concentration of:

$$\phi(X, D) \equiv x_i^T QAQx_i(1 + D_i\Delta_i) = x_i^T QAQ_{-i}x_i = x_i^T Q_{-i}AQx_i.$$

Given $X, X' \in X(\Omega)$, we note $Q' \equiv (I_p - \frac{1}{n}X'DX'^T)^{-1}$ and $\Delta' \equiv \frac{1}{n}x_i^T Q' x_i$. One can bound:

$$\begin{aligned}
&|\phi(X, D) - \phi(X', D)| \\
&\quad \leq \left|x_i^T(Q - Q')AQ_{-i}x_i\right| + \left|x_i^T Q'AQx_iD_i(\Delta_i - \Delta'_i)\right| + \left|x_i^T Q'_i A(Q - Q')x_i\right|
\end{aligned}$$

1. One can then bound ($\left|x_i^T Q'_{-i}A(Q - Q')x_i\right|$ is treated the same way):

$$\begin{aligned}
&\left|x_i^T(Q - Q')AQ_{-i}x_i\right| \\
&\quad = \frac{1}{n}\left|x_i^T Q(X - X')DX^T Q'AQ_{-i}x_i\right| + \frac{1}{n}\left|x_i^T QX'D(X - X')^T Q'AQ_{-i}x_i\right| \\
&\quad \leq \frac{2\kappa^2\kappa_D}{\varepsilon^2}\|AQ_{-i}x_i\|\|X - X'\|,
\end{aligned}$$

2. besides:

$$\begin{aligned}|\Delta_i - \Delta_i'| &\leq \frac{1}{n^2}\left|x_i^T(Q - Q_{-i})x_i\right| + \frac{2}{n}\left|x_i^T Q'_{-i}(x_i - x_i')\right| \\ &\leq \left(\frac{2\kappa^3\kappa_D}{\varepsilon^2\sqrt{n}} + \frac{2\kappa}{\varepsilon\sqrt{n}}\right)\|X - X'\|,\end{aligned}$$

Therefore:

$$|\phi(X,D) - \phi(X',D)| \leq O\left(\frac{1}{\sqrt{n}}\left|x_i^T Q' A Q x_i\right| + \|AQ_{-i}x_i\|\right)\|X - X'\|.$$

We are then left to bound the variation towards D which is slightly more tricky. Let us consider $D' \in \mathcal{D}_n$ and note $Q'' \equiv (I_p - \frac{1}{n}XD'X^T)^{-1}$ and $\Delta' \equiv \frac{1}{n}x_i^T Q'' x_i$. This time one should decompose followingly:

$$\begin{aligned}&|\phi(X,D) - \phi(X,D')| \\ &\quad\leq \left|x_i^T(Q_i - Q_i'')AQx_i\right| + \left|x_i^T Q_i'' AQx_i D_i(\Delta_i - \Delta_i'')\right| + \left|x_i^T Q'' A(Q_i - Q_i'')x_i\right|\end{aligned}$$

1. First let us take advantage of the independence between x_i and X_{-i} to bound as in the proof of Lemma 3.9:

$$\begin{aligned}\left|x_i^T(Q_i - Q_i'')AQx_i\right| &= \frac{1}{n}\left|x_i^T Q_i X_{-i}(D - D')X_{-i}Q_i'' AQx_i\right| \\ &\leq \frac{\kappa^2}{\varepsilon^2}\left\|\frac{1}{\sqrt{n}}x_i^T Q_i X_{-i}\right\|_\infty \|AQ_{-i}x_i\|\|D - D'\|_F\end{aligned}$$

2. Second, let us bound:

$$|\Delta_i - \Delta_i''| \leq \frac{1}{n^2}\left|x_i^T Q_{-i}X_{-i}(D - D'')X_{-i}Q_{-i}x_i\right| \leq \frac{\kappa^4}{\varepsilon^2\sqrt{n}}\|D - D''\|$$

That allows us to set that:

$$|\phi(X,D) - \phi(X,D')| \leq O\left(\frac{1}{\sqrt{n}}\left|x_i^T Q' AQx_i\right| + \left\|\frac{1}{\sqrt{n}}x_i^T Q_i X_{-i}\right\|_\infty \|AQ_{-i}x_i\|\right)\|D - D'\|_F$$

We have then all the elements to apply Theorem ?? with the concentrations:

- $\frac{1}{\sqrt{\log n}}\|AQ_{-i}x_i\| \in O(1) \pm \mathcal{E}_2(1)$ thanks to Lemmas 3.8, 3.9 and the fact that $\mathbb{E}[\|AQ_{-i}^{(i)}x_i\|] \leq \sqrt{\mathbb{E}[x_i^T Q_{-i}^{(i)} AAQ_{-i}^{(i)}x_i]} \leq \frac{1}{\varepsilon}\|\mathbb{E}[x_i x_i^T]\|^{1/2}\|A\|_F \leq O(1)$.
- $\frac{1}{\sqrt{n\log n}}\left|x_i^T Q' AQx_i\right| \in O(1/\sqrt{\log n}) \pm \mathcal{E}_2$ thanks to Lemma 3.15
- $\left\|\frac{1}{\sqrt{n\log n}}x_i^T Q_i X_{-i}\right\|_\infty \in O(1) \pm \mathcal{E}_2$ thanks to Lemma 3.10.

and the parameters $\sigma = 1$ and $\mu = (1,1,1)$ to obtain:

$$x_i^T QAQx_i(1 + D_i\Delta_i) \in \mathcal{E}_2\left(\sqrt{\log n}\right) + \mathcal{E}_1\left(\sqrt{\log n}\right) + \mathcal{E}_{2/3}(\sqrt{\log n}).$$

To show the concentration of $x_i^T QAQx_i(1 + D_i\bar{\Delta}_i)$, note that:

$$\left|x_i^T QAQx_i(1 + D_i\bar{\Delta}_i) - x_i^T QAQx_i(1 + D_i\Delta_i)\right| \leq \kappa_D\left|x_i^T QAQx_i\right|\left|\Delta_i - \bar{\Delta}_i\right|,$$

Lemma ?? allows us to set the concentration of the product between $x_i^T QAQx_i \in O(\sqrt{n}) \pm \mathcal{E}_1$ and $\left|\Delta_i - \bar{\Delta}_i\right| \in 0 \pm \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1(1/n)$ satisfying $\left|\Delta_i - \bar{\Delta}_i\right| \leq \frac{\kappa^2}{\varepsilon}$:

$$\left|x_i^T QAQx_i\right|\left|\Delta_i - \bar{\Delta}_i\right| \in O(1) \pm \mathcal{E}_2(1) + \mathcal{E}_1(1/\sqrt{n})$$

Lemma ?? then allows us to conclude on the concentration of $x_i^T QAQx_i(1 + D_i\Delta_i)$. □

The existence of the deterministic parameters $\delta \in \mathcal{D}_n$ such that:

$$\delta = \frac{1}{n} \operatorname{Diag}_{i\in[n]} \operatorname{Tr}(\Sigma_i \tilde{Q}^\delta(D))$$
$$\left(\text{recall that } \tilde{Q}^\delta(D) \equiv \left(I_p - \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\frac{D_i}{1+D_i\delta_i}\right]\Sigma_i\right)^{-1}\right)$$

is a consequence of Theorem 2.18. Note that Corollary 2.24 giving a deterministic equivalent for $Q = (I_p - \frac{1}{n}XX^T)^{-1}$ imposes the columns of X to be independent but they can possibly be non identically distributed. It concerns in particular the case of matrices $(I_p - \frac{1}{n}X\tilde{D}X^T)^{-1}$ for deterministic diagonal matrices $\tilde{D}$ as stated below. It is at the basis of the estimation of $\mathbb{E}[Q] = \mathbb{E}[(I_p - \frac{1}{n}XDX^T)^{-1}]$.

Proof of Theorem 3.2. Let us introduce the resolvent $\bar{Q} \equiv (I_p - \frac{1}{n}X\bar{D}X^T)^{-1}$ where we defined

$$\bar{D} \equiv \frac{\mathbb{E}\left[\frac{D}{I_p+\bar{\Delta}D}\right]}{I_p - \bar{\Delta}\mathbb{E}\left[\frac{D}{I_p+\bar{\Delta}D}\right]}.$$

As will be understood later, this elaborated definition is taken for $\bar{D}$ to satisfy the following relation:

$$\frac{\bar{D}}{I_p + \bar{D}\bar{\Delta}} = \mathbb{E}\left[\frac{D}{I_p + \bar{\Delta}D}\right],$$

it implies in particular that $\tilde{Q}^\delta(D) = \tilde{Q}^\delta(\bar{D})$ for any $\delta \in \mathcal{D}_n$. Let us then consider a deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\|_F \leq 1$ and bound:

$$\begin{aligned}
&\left|\mathbb{E}[\operatorname{Tr}(AQ)] - \mathbb{E}[\operatorname{Tr}(A\bar{Q})]\right| \\
&= \frac{1}{n}\sum_{i=1}^n \left|\mathbb{E}\left[x_i^T QA\bar{Q}x_i\bar{\Delta}_i^{-1}\left(\bar{\Delta}_i D_i + 1 - (\bar{\Delta}_i\bar{D}_i + 1)\right)\right]\right| \\
&= \frac{1}{n}\sum_{i=1}^n \left|\mathbb{E}\left[x_i^T QA\bar{Q}x_i\bar{\Delta}_i^{-1}\left(\bar{\Delta}_i D_i + 1\right)\left(\bar{\Delta}_i\bar{D}_i + 1\right)\left(\frac{1}{\bar{\Delta}_i\bar{D}_i+1} - \frac{1}{\bar{\Delta}_i D_i + 1}\right)\right]\right| \\
&= \frac{1}{n}\sum_{i=1}^n \left|\mathbb{E}\left[x_i^T QA\bar{Q}x_i\left(\bar{\Delta}_i D_i + 1\right)\left(\bar{\Delta}_i\bar{D}_i + 1\right)\left(\frac{D_i}{\bar{\Delta}_i D_i+1} - \mathbb{E}\left[\frac{D_i}{\bar{\Delta}_i D_i + 1}\right]\right)\right]\right| \\
&= \frac{1}{n}\sum_{i=1}^n \left|\mathbb{E}\left[\left(x_i^T QA\bar{Q}x_i\left(\bar{\Delta}_i D_i + 1\right) - \mathbb{E}\left[x_i^T QA\bar{Q}x_i\left(\bar{\Delta}_i D_i + 1\right)\right]\right)\frac{D_i\left(\bar{\Delta}_i\bar{D}_i+1\right)}{\bar{\Delta}_i D_i + 1}\right]\right| \\
&= \kappa_D\left(\frac{\kappa_D\kappa^2}{\varepsilon} + 1\right)\sup_{i\in[n]}\mathbb{E}\left[\left|x_i^T QA\bar{Q}x_i\left(\bar{\Delta}_i D_i + 1\right) - \mathbb{E}\left[x_i^T QA\bar{Q}x_i\left(\bar{\Delta}_i D_i + 1\right)\right]\right|\right] \\
&\leq O\left(\sqrt{\log n}\right)
\end{aligned}$$

Thanks to Lemma 3.16. Then Corollary 2.24 allows us to state that the deterministic diagonal matrix $\delta(\bar{D}) \in \mathcal{D}_n$ solution to:

$$\delta = \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^\delta(\bar{D})\right) \qquad \left(= \frac{1}{n}\operatorname{Tr}\left(\Sigma_i\tilde{Q}^\delta(D)\right)\right),$$

satisfies the estimation:

$$\left\|\mathbb{E}[Q] - \tilde{Q}^{\delta(D)}(D)\right\|_F \leq \left\|\mathbb{E}[Q] - \mathbb{E}[\bar{Q}]\right\|_F + \left\|\mathbb{E}[\bar{Q}] - \tilde{Q}^{\delta(D)}(D)\right\|_F \leq O\left(\sqrt{\log n} + \frac{1}{\sqrt{n}}\right) \leq O\left(\sqrt{\log n}\right).$$

□

The same way that Theorem 3.1 can be linked to Proposition **??** giving the linear concentration of XDX^T, the next proposition can be linked to Proposition **??** giving the Lipschitz concentration of XDX^Tu for any deterministic $u \in \mathbb{R}^p$.

Proposition 3.17. *In the setting of Theorem 3.1, for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq O(1)$:*

$$Qu \propto \mathcal{E}_1 \left(\sqrt{\frac{\log n}{n}} \right).$$

Proof. With the same variables $X, X \in \mathcal{M}_{p,n}$, $D, D' \in \mathcal{D}_n$ and with the same notations Q, Q'_X, Q'_D as in the proof of Theorem 3.1, we bound:

$$\begin{aligned} \|Qu - Q'_X u\| &= \frac{1}{n} \left\| Q(X - X')DX^T Q'_X u \right\| + \frac{1}{n} \left\| QX'D(X - X')^T Q'_X u \right\| \\ &\leq \frac{\kappa \kappa_D}{\varepsilon^2 \sqrt{n}} \|X - X'\| , \end{aligned}$$

Second:

$$\|Qu - Q'_D u\| = \frac{1}{n} \left\| Q'_D X(D - D')X^T Qu \right\| \leq \frac{\kappa}{\varepsilon \sqrt{n}} \|D - D'\|_F \|X^T Qu\|_\infty,$$

and we know from Lemma 3.11 that $\|X^T Qu\|_\infty \in O(\sqrt{\log n}) + \mathcal{E}_2(\sqrt{\log n})$, which allows us to conclude with Theorem **??** that:

$$\sqrt{\frac{n}{\log n}} Qu \propto \mathcal{E}_2 + \mathcal{E}_1,$$

but the $\mathcal{E}_2$ decay can here be removed since the $\mathcal{E}_1$ is looser. □