



Concentration of solutions to random equations with concentration of measure hypotheses

Romain Couillet, Cosme Louart

► To cite this version:

Romain Couillet, Cosme Louart. Concentration of solutions to random equations with concentration of measure hypotheses. 2020. hal-02973851

HAL Id: hal-02973851

<https://hal.science/hal-02973851>

Preprint submitted on 21 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Concentration of solutions to random equations with concentration of measure hypotheses

Romain Couillet^a, Cosme Louart^{a,1}

^a*GIPSA-lab, Université Grenoble-Alpes, France*

^b*CEA LIST, Site Nano-innov, Palaiseau, France*

Abstract

We propose here to study the concentration of random objects that are implicitly formulated as fixed points to equations $Y = f(X)$ where f is a random mapping. Starting from an hypothesis taken from the concentration of the measure theory, we are able to express precisely the concentration of such solutions, under some contractivity hypothesis on f . This statement has important implication to random matrix theory, and is at the basis of the study of some optimization procedures like the logistic regression for instance. In those last cases, we give precise estimations to the first statistics of the solution Y which allows us predict the performances of the algorithm.

Keywords: random matrix theory, concentration of measure, parameter optimization, Robust estimation.

Contents

1	Basics of the Concentration of Measure Theory	8
1.1	Lipschitz concentration and fundamental examples	8
1.2	Linear concentration, notations and properties of the first statistics	12
1.3	Linear concentration through sums and integrals	15
1.4	Concentration of the norm	19
1.5	Concentration of basic operations	21
1.6	Concentration under highly probable event	26
1.7	The class of convexly concentrated vectors and its degeneracy . .	30
2	Concentration of the solutions of $Y = \phi(Y)$	38
2.1	When ϕ is affine and $\mathcal{L}(\phi)^k(\phi(0)) \in \mathcal{E}_2(\sigma(1 - \varepsilon)^k)$ for all $k \in \mathbb{N}$.	39
2.2	When $\phi^k(y)$ is linearly or Lipschitz concentrated for $k \leq \log(\eta_{(E, \ \cdot\)})$ and y deterministic and bounded.	41
2.3	When $\phi \propto \mathcal{E}_2(\sigma)$ for the infinity norm	46

Email addresses: romain.couillet@gipsa-lab.grenoble-inp.fr (Romain Couillet),
cosme.louart@gipsa-lab.grenoble-inp.fr (Cosme Louart)

2.4	When $\phi = \Psi(X)$ with Ψ deterministic and X concentrated . . .	47
3	First example : Concentration of the resolvent and device of a deterministic equivalent	48
3.1	Setting, assumptions and first properties	48
3.2	A first deterministic equivalent	50
3.3	A second deterministic equivalent	54
3.4	Existence and uniqueness of Λ^z when $z \in \mathbb{C}_-^-$	56
3.5	Analyticity of $z \mapsto \Lambda^z$	59
3.6	Generalization of the definition and the analyticity of $z \mapsto \Lambda^z$ on the whole set $\mathbb{C} \setminus [0, 1 - \varepsilon]$ and convergence of Λ^z towards Δ^z . .	64
3.7	Proof of Propositions 16 and 17	68
3.8	Resolvent of $\frac{1}{n}XDXT$	70
3.9	Concentration of the resolvent of the sample covariance matrix of convexly concentrated data	72
4	Second example : fixed point equation depending on independent data $x_1, \dots, x_n$	75
4.1	Setting and first properties	75
4.2	Concentration of $y_{-i}^{X'}(t)$	77
4.3	Integration of $\frac{\partial y_{-i}^{X'}(t)}{\partial t}$	81
4.4	Implicit relation between x_i^TY and $x_i^TY_{-i}$	85
4.5	Expression of the mean and covariance of Y	87
4.6	Computation of the estimation of the mean and covariance of Y when X is Gaussian	88
4.7	Computation of $\mu_j \equiv m_j^T m_Y$ and $\nu_j \equiv \text{Tr}(\Sigma_j^T \Sigma_Y)$, $j \in [n]$	90
4.8	Application to the logistic regression for $x_1, \dots, x_n$ Gaussian i.i.d. vectors.	92
Appendix A	Definition of the expectation	94
Appendix B	Concentration of Y solution to $Y = \phi(Y)$ when ϕ is locally Lipschitz	96

Introduction

From random matrix theory (RMT) to optimization theory (OT), the study of fixed point equation appears as a systematic approach to solve efficiently implicitly expressed issues. In RMT, the spectrum $\text{Sp}(X)$ that implicitly underlies the statistical behavior of a random matrix $X \in \mathcal{M}_p$ is understood through the study of the resolvent $Q(z) = (zI_p - X)^{-1}$, unique solution to the fixed point equation $Q(z) = \frac{1}{z}I_p + \frac{1}{z}XQ(z)$ when z is sufficiently far from $\text{Sp}(X)$. The same way, optimization problems with regularization terms like:

$$\text{Minimize} \quad f(x_1, \dots, x_n, \beta) + \lambda \|\beta\|^2, \quad \beta \in \mathbb{R}^p$$

induce naturally fixed point equations when f is differentiable and lead to the equation:

$$\beta = -\frac{2}{\lambda} \frac{\partial f(x_1, \dots, x_n, \beta)}{\partial \beta},$$

which is contracting and thus admit a unique solution when λ is sufficiently big. In both cases, one classically aims at showing that $Q(z)$ (or β) converges towards a deterministic behavior expressed through a so-called “deterministic equivalent”. The main objective of our paper is to show how concentration properties of the random objects X or $(x_1, \dots, x_n)$ can be respectively transferred to the resolvent $Q(z)$ or to the optimized parameter β and allows for a precise understanding of their statistical properties.

Following a work initiated in Louart and Couillet (2020), our probabilistic approach is inspired from the Concentration of Measure Theory (CMT) that demonstrates an interesting flexibility allowing (i) to characterize realistic setting where, in particular, the hypothesis of independent entries is relaxed (ii) to provide rich concentration inequalities with precise convergence bounds. The study of random matrices is often conducted in the literature with mere Gaussian hypotheses or with weaker hypotheses concerning the first moments of the entries that are supposed to be independent (at least independent via an affine transformation). As an example, consider a sample covariance matrices $\frac{1}{n}XX^T$, where $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$ is the data matrix. It seems natural to assume that the columns $x_1, \dots, x_n$ are independent, nonetheless, to assume that the entries of each datum x_i (for $i \in [n]$) are independent limits greatly the range of application. Thanks to concentration of the measure hypothesis, this last independence property is no longer needed. To present the simpler picture possible, we will admit in this introduction that what we call for the moment “concentrated vectors” are transformation $F(Z)$ of a Gaussian vector $Z \sim \mathcal{N}(0, I_d)$ for a given 1-Lipschitz (for the euclidean norm) mapping $F : \mathbb{R}^d \rightarrow \mathbb{R}^p$. This class of random vectors is originated from a central result of CMT (Ledoux, 2005, Corollary 2.6) that states that for any λ -Lipschitz mapping $f : \mathbb{R}^d \rightarrow \mathbb{R}$ (where $\mathbb{R}^d$ and $\mathbb{R}$ are respectively endowed with the euclidean norm $\|\cdot\|$ and with the absolute value $|\cdot|$):

$$\forall t > 0 : \mathbb{P}(|f(Z) - \mathbb{E}[f(Z)]| \geq t) \leq Ce^{-(t/\sigma\lambda)^2}, \quad (1)$$

where $C = 2$ and $\sigma = \sqrt{2}$ (they do not depend on the dimensions d !). Note that the speed of concentration is proportional to the Lipschitz parameter of f , the random variable $f(Z)$ – that is called a “ λ -Lipschitz observation of Z ” – has a standard deviation that does not depend on the dimension d (if λ is constant when d tends to ∞). We note this property succinctly $Z \propto C\mathcal{E}_2(\sigma)$, or, if we place ourselves in the quasi-asymptotic regime where the dimension d (or p) is big, we do not pay attention to the constants appearing in the exponential bound (as long as $C, \sigma \underset{d \rightarrow \infty}{\leq} O(1)$, the result would not be much different) and we write $Z \propto \mathcal{E}_2$. Then we say that the “observable diameter of Z ” is of order $O(1)$,

which means that all the standard deviations of the 1-Lipschitz observation of Z are of order $O(1)$ (they do not increase with the dimension).

We can then deduce an infinite number of concentration inequalities on any observation $g(F(Z))$ for $g : \mathbb{R}^p \rightarrow \mathbb{R}$ Lipschitz. If F is, say, λ -Lipschitz with λ possibly depending on the dimension, we have therefore the concentration $F(Z) \propto \mathcal{E}_2(\lambda)$. In particular, considering Generative adversarial neural network (GAN) whose objective is to construct from a given noise, realistic image of a given class (images of cats, images of planes, etc...), we can assert as it was explained in Seddik et al. (2020) that the artificial (but realistic !) outputs of those networks form concentrated vectors as Lipschitz transformations of Gaussian drawings.

The notation $F(Z) \propto \mathcal{E}_2(\lambda)$ is set rigorously in the first section of this paper but we choose to introduce it already here as a foretaste because it will be extensively employed all over the paper and once it is well understood, it clarifies drastically the message (and reduce tremendously the expressions). In some aspects, this notation and the systematic attempt to employ it for each new random object that we meet constitute our main contribution to this field of research. To present it briefly, we will show in this paper that if $X \propto \mathcal{E}_2$ (and other minor hypotheses), then:

$$(Q \mid \mathcal{A}_Q) = \left(I_p - \frac{1}{n} X X^T \right)^{-1} \propto \mathcal{E}_2(1/\sqrt{n}) \quad \text{and} \quad (\beta \mid \mathcal{A}_\beta) \propto \mathcal{E}_2(1/\sqrt{n}),$$

where $\mathcal{A}_Q$ and $\mathcal{A}_\beta$ are event of high probability (bigger than $1 - Ce^{-c(n+p)}$, for two constants $C, c > 0$) under which Q and β are, at the same time, well defined and concentrated. It was shown in Louart and Couillet (2020) that Lipschitz transformation, sums or product of concentrated vectors are also concentrated vectors, here, we go a step further setting that the concentration is preserved through such implicit formulations as the one defining Q and β . Once we know that they are concentrated, we also provide means to estimate their statistics.

Our paper is organized as follows. In a first part we display briefly, but rigorously important results of Louart and Couillet (2020) and new ones that provide a solid basis to use CMT tools, we particularly insist on the distinction between three classes of concentrated vectors identified (in an increasing order for the inclusion relation):

- the Lipschitz concentrated vectors $Z : f(Z)$ satisfies (1) for any λ -Lipschitz f having value in $\mathbb{R}$, we note $Z \propto \mathcal{E}_2(\sigma)$,
- the convexly concentrated vectors $Z : f(Z)$ satisfies (1) for any λ -Lipschitz and *convex* functional f , we note $Z \propto_c \mathcal{E}_2(\sigma)$
- the linearly concentrated random vectors – $f(Z)$ satisfies (1) for any *linear forms* f bounded by λ , we note $Z \in \mathcal{E}_2(\sigma)$. For such linear mappings, we can rewrite (1):

$$\forall t > 0 : \mathbb{P}(|f(Z) - f(\mathbb{E}[Z])| \geq t) \leq Ce^{-(t/\sigma\lambda)^2}, \quad (2)$$

which leads us to introducing the notation $Z \in \mathbb{E}[Z] \pm \mathcal{E}_2(\sigma)$ to express the fact that the linear observations $f(Z)$ of Z lie around $f(\mathbb{E}[Z])$ (which has no reason to be the case for Lipschitz observation when Z is Lipschitz-concentrated). A major issue in RMT is to find a computable deterministic matrix $\tilde{Z}$ close to $\mathbb{E}[Z]$ such that $Z \in \tilde{Z} \pm \mathcal{E}_2(\sigma)$. Such deterministic matrices are then called “deterministic equivalents” of Z ($\mathbb{E}[Z]$ is of course one of them).

Although it is the most stable class, the Lipschitz concentrated vectors can degenerate into linear concentration for instance when we look at the random matrix XDY^T for $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ Lipschitz concentrated. That justifies the introduction of the notion of linear concentration that is simpler to verify and still gives some capital control on the norm. The convex concentration, although it is not so easy to treat – being only stable through affine mappings – finds some interest thanks to a well known result of Talagrand (1995) that sets its validity for any random vector of $[0, 1]^p$ with independent entries, allowing this way to consider discrete distribution quite absent of the Lipschitz concentrated vectors class. The class of convexly concentrated vectors often degenerates into a mere linear concentration when one consider the product of convexly concentrated random matrices, the entry-wise product of convexly concentrated random vectors or the resolvent $(I_p - X)^{-1}$ of a convexly concentrated random matrix $X \in \mathcal{M}_p$...

In a second part we present our main theorems allowing us to set the concentration of the solution of equation of the type:

$$Y : Y = F(X, Y)$$

where X is concentrated and $y \mapsto F(X, y)$ is contracting with high probability. In a third part we give a first application of the two first theoretical part with the design of a deterministic equivalent of the resolvent $Q^z \equiv (zI_p - \frac{1}{n}XDX^T)^{-1}$. We consider three different settings where n and p are of the same order ($O(n) = O(p)$):

- if $X \propto \mathcal{E}_2$, and $D = I_n$, then $(Q^z \mid \mathcal{A}_Q) \propto \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{M}_p, \|\cdot\|_F)$,
- if $X \propto \mathcal{E}_2$, and $D \propto \mathcal{E}_2$, then $(Q^z \mid \mathcal{A}_Q) \propto \mathcal{E}_2(\sqrt{\log n})$ in $(\mathcal{M}_p, \|\cdot\|_F)$,
- if $X \propto_c \mathcal{E}_2$, and $D = I_n$, then $(Q^z \mid \mathcal{A}_Q) \in \mathcal{E}_2$ in $(\mathcal{M}_p, \|\cdot\|_*)$.

In addition, for all those settings we also provide a similar computable deterministic equivalent. Note that the different results of concentration differ by the observable diameter, by the type of concentration *and* by the choice of the norm endowing $\mathcal{M}_p$. For a given matrix $A \in \mathcal{M}_p$, we employ the notation $\|A\|_F = \sqrt{\text{Tr}(AA^T)} = \sup_{\|B\|_F \leq 1} \text{Tr}(BA)$ and

$$\|A\|_* = \text{Tr}(\sqrt{AA^T}) = \sup_{\|B\| \leq 1} \text{Tr}(BA),$$

where $\|\cdot\|$ is the spectral norm : $\|A\| = \sup_{\|u\|, \|v\| \leq 1} u^T Av$. In particular, since for any $A \in \mathcal{M}_p$ such that $\|A\| \leq 1$, $q \mapsto \text{Tr}(Aq)$ is $\sqrt{p}$ -Lipschitz for the

Frobenius norm, for any random matrix Z , $Z \propto \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{M}_p, \|\cdot\|_F)$ implies $Z \propto \mathcal{E}_2$ in $(\mathcal{M}_p, \|\cdot\|_*)$, which justifies that the last setting gives a weaker result than the first one (but it is still important !).

Finally, in a last part, we consider a precise fixed point equation:

$$Y = \frac{\lambda}{n} \sum_{i=1}^n f(x_i^T Y) x_i,$$

where the vectors $x_1, \dots, x_n \in \mathbb{R}^p$ are all independent, $f : \mathbb{R} \rightarrow \mathbb{R}$ is twice differentiable, f' and f'' are bounded and $\|f'\|_\infty$ is chosen small enough for the equation to be contractive with high probability (i.e. bigger than $1 - Ce^{-cn}$, for two constants $C, c > 0$). We show in that case how the statistical behavior of Y can be understood and give, in particular, a precise estimation of its expectation and covariance.

Main Contributions

1. We provide in Corollary 4 the linear concentration:

$$XDY^T \in \mathcal{E}_2(\sqrt{(n+p)\log(pn)}) \quad \text{in } (\mathcal{M}_p, \|\cdot\|_F),$$

when $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ all satisfy $X, Y, D \propto \mathcal{E}_2$. This result is central for the design of a computable deterministic equivalents of Q .

2. We prove the stability of the convex concentration through entry-wise product of vectors (Proposition 12) and through matrix product (Proposition 13).
3. We present in Section 2 a detailed framework to show the concentration of the solution of a random fixed point equation $Y = \phi(Y)$ when Φ is contractive with high probability with four main results :
 - Theorems 5 gives the linear concentration of Y when ϕ is affine and all the iterates $\phi^k(0)$ are concentrated
 - Theorems 6 and 7 give the same result when just a small number of iterations of $\phi^k(y)$, for y not too far from 0 are concentrated
 - Theorem 8 gives Lipschitz concentration ϕ is possibly non affine and a small number of iterations $\phi^k(y_0)$, for y_0 well chosen, are concentrated
 - Theorem 9 gives Lipschitz concentration of Y for general ϕ when ϕ is concentrated as a random mapping and for the infinity norm
4. We justify the concentration of the random matrices $X^k(I_p - X^l)^{-1}$ when one only assumes that X is a symmetric matrix, convexly concentrated (it happens for instance when $X \in \mathcal{S}_p([0, 1])$ and has independent entries on each triangle – it is a consequence of the Theorem of Talagrand, here the Theorem 3). The same result holds for $(XX^T)^k(I_p - (XX^T)^l)^{-1}$ (or $X^T(XX^T)^k(I_p - (XX^T)^l)^{-1}$ etc...) when $X \in \mathcal{M}_{p,n}$ is convexly concentrated (see Corollary 8).

5. We design a new semi-metric d_s that seems to have a structural importance in the field of RMT. It is defined for any complex diagonal matrices $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$ (where $\mathbb{C}^+ \equiv \{z \in \mathbb{C}, \Im(z) > 0\}$) as $d_s(D, D') = \|(D - D')/\sqrt{\Im(D)\Im(D')}\|$. It allows us to set properly the validity of the definition of the deterministic equivalent $\tilde{Q}$ of $Q = (I_p - \frac{1}{n}XX^T)^{-1}$ when the $x_1, \dots, x_n$ are not identically distributed (but independent !). It is invoked in particular to employ Banach-like theorems (Theorem 11) in order to show the existence and uniqueness of solution of fixed-point equations on diagonal matrices (Proposition 16 for instance).
6. We give a precise quasi-asymptotic speed bound to the convergence of $Q^z \equiv (zI_p - \frac{1}{n}XX^T)^{-1}$ towards its deterministic equivalent $\tilde{Q}^z$ when the distance between z and $\text{Sp}(\frac{1}{n}XX^T) \subset \mathbb{R}$ stays constant with the dimension (Theorem 10). More precisely, if there exists an event $\mathcal{A}_Q$, such that $\mathcal{A}_Q \subset \{\|\frac{1}{n}XX^T\| \leq 1 - \varepsilon\}$ and $\mathbb{P}(\mathcal{A}_Q^c) \leq Ce^{-cn}$ (for two constants $C, c > 0$), we show that if $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $|z| \leq O(1)$:

$$\|\mathbb{E}[Q^z, \mathcal{A}_Q] - \tilde{Q}^z\|_F \underset{\substack{p, n \rightarrow \infty \\ n/p = \gamma}}{\leq} O\left(\sqrt{\frac{\log n}{n}}\right)$$

It is a good improvement of Louart and Couillet (2020) where we could just bound the spectral norm (and not the Frobenius norm) of the difference and just for z at a distance of order $O(1)$ from the positive real line. The convergences surpasses (but with a different setting) the results of Bai and Zhou (2008).

7. We prove the concentration of a resolvent $Q^z(\Gamma) \equiv (zI_p - \frac{1}{n}X\Gamma X^T)^{-1}$ when $\Gamma \propto \mathcal{E}_2$. This resolvent has an observable diameter of order $O(\sqrt{\log n})$, which is $O(\sqrt{n})$ time bigger than the one of $(zI_p - \frac{1}{n}XX^T)^{-1}$. It is however still sufficient for our needs (in particular, the Stieltjes transform still concentrates and can be estimated, but more importantly for us some quantities of the last section precisely need this speed of concentration). We also prove that for any deterministic vector $u \in \mathbb{R}^p$ $Q^z(\Gamma)u$ has an observable diameter of order $O(\sqrt{\log n/n})$.
8. We provide a rigorous proof to the validity of the estimators of the expectation and covariance of a robust regression parameter as was presented in El Karoui et al. (2013) and Mai et al. (2019). Those estimations rely on a pseudo-identity involving the parameter vectors $Y \in \mathbb{R}^p$ under study and $Y_{-i} \in \mathbb{R}^p$ the same parameter deprived of the contribution of the datum x_i . They satisfy the equations:

$$Y = \frac{1}{n} \sum_{i=1}^n f(x_i^T Y) x_i \quad \text{and} \quad Y_{-i} = \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^n f(x_j^T Y_{-i}) x_j$$

note that, by construction, Y_{-i} is independent with x_i (since $x_1, \dots, x_n$ are all independent). The random vector $W_i \equiv Y - Y_{-i} - \frac{1}{n}f(x_i^T Y)Q_{-i}x_i$

is very close to zero and we have the estimations:

$$(W_i \mid \mathcal{A}_Y) \propto \mathcal{E}_2\left(\frac{1}{n}\right) \quad \text{and} \quad \|\mathbb{E}[W_i \mid \mathcal{A}_Y]\| \leq O\left(\frac{\sqrt{\log n}}{n}\right),$$

where $\mathcal{A}_Y$ is an event of high probability under which Y and Y_{-i} are well defined and $Q_{-i} \equiv (I_p - \frac{1}{n} \sum_{i=1}^n f'(x_i^T Y_{-i}) x_i x_i^T)^{-1}$ (see Proposition 35).

1. Basics of the Concentration of Measure Theory

We choose here to adopt the viewpoint of Levy families where the goal is to track the influence of the vector dimension over the concentration. Specifically, we are given a sequence of random vectors $(Z_p)_{p \geq \mathbb{N}}$ where each Z_p belongs to a space of dimension p (typically $\mathbb{R}^p$) and we want to obtain inequalities of the form:

$$\forall p \in \mathbb{N}, \forall t > 0 : \mathbb{P}(|f_p(Z_p) - a_p| \geq t) \leq \alpha_p(t), \quad (3)$$

where for every $p \in \mathbb{N}$, $\alpha_p : \mathbb{R}^+ \rightarrow [0, 1]$ is called a *concentration function*, which is left-continuous, decreasing and tends to 0 at infinity, $f_p : \mathbb{R}^p \rightarrow \mathbb{R}$ is a 1-Lipschitz function, and a_p is either a deterministic variable (typically $\mathbb{E}[f_p(Z_p)]$) or a random variable (for instance $f_p(Z'_p)$ with Z'_p an independent copy of Z_p). The sequences of random vectors $(Z_p)_{p \geq 0}$ satisfying inequality (3) for any sequences of 1-Lipschitz functions $(f_p)_{p \geq 0}$ are called *Levy families* or more simply *concentrated vectors* (with this denomination, we implicitly omit the dependence on p and abusively call “vectors” the *sequences* of random vectors of growing dimension). The concentrated vectors having a concentration function α_p exponentially decreasing are extremely flexible objects. We dedicate the next two subsections to further definitions of the fundamental notions involved under this setting, which are of central interest to the present article – these notions are primarily motivated by Theorem 1, to be introduced below.

We define here three classes of concentrated vectors depending on the regularity of the class of sequences of functions $(f_p)_{p \in \mathbb{N}}$ satisfying (3). When (3) holds for all the 1-Lipschitz mappings f_p , Z_p is said to be *Lipschitz concentrated*; when only valid for all 1-Lipschitz *and quasi-convex* (see Definition 7) mappings f_p , Z_p is said to be *convexly concentrated*; and when true for all 1-Lipschitz *and linear* mappings f_p , Z_p is said to be *linearly concentrated*. As such, the concentration of a random vector Z_p is only defined through the concentration of what we call its “observations” $f_p(Z_p)$ for all f_p in a specific class of functions.

1.1. Lipschitz concentration and fundamental examples

We will work with normed (or semi-normed) vector spaces, although CMT is classically developed in metric spaces. The presence of a norm (or a semi-norm) on the vector space is particularly important when we try to show the concentration of a product of random vectors (among other applications).

Definition/Proposition 1. *Given a sequence of normed vector spaces $(E_p, \|\cdot\|_p)_{p \geq 0}$, a sequence of random vectors $(Z_p)_{p \geq 0} \in \prod_{p \geq 0} E_p^1$, a sequence of positive reals $(\sigma_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say that Z_p is Lipschitz q -exponentially concentrated with an observable diameter of order $O(\sigma_p)$ iff one of the following three equivalent assertions is satisfied:*

- $\exists C, c > 0 \mid \forall p \in \mathbb{N}, \forall 1\text{-Lipschitz } f : E_p \rightarrow \mathbb{R}, \forall t > 0 :$

$$\mathbb{P}(|f(Z_p) - f(Z'_p)| \geq t) \leq C e^{-(t/c\sigma_p)^q},$$

- $\exists C, c > 0 \mid \forall p \in \mathbb{N}, \forall 1\text{-Lipschitz } f : E_p \rightarrow \mathbb{R}, \forall t > 0 :$

$$\mathbb{P}(|f(Z_p) - m_f| \geq t) \leq C e^{-(t/c\sigma_p)^q},$$

- $\exists C, c > 0 \mid \forall p \in \mathbb{N}, \forall 1\text{-Lipschitz } f : E_p \rightarrow \mathbb{R}, \forall t > 0 :$

$$\mathbb{P}(|f(Z_p) - \mathbb{E}[f(Z_p)]| \geq t) \leq C e^{-(t/c\sigma_p)^q},$$

where Z'_p is an independent copy of Z_p and m_f is a median of $f(Z_p)$ (it satisfies $\mathbb{P}(f(Z_p) \geq m_f), \mathbb{P}(f(Z_p) \leq m_f) \geq \frac{1}{2}$); the mappings f are of course 1-Lipschitz for the norm (or semi-norm) $\|\cdot\|_p$. We denote in this case $Z_p \propto \mathcal{E}_q(\sigma_p)$ (or more simply $Z \propto \mathcal{E}_q(\sigma)$). If $\sigma = O(1)$, we simply write $Z_p \propto \mathcal{E}_q$.

The equivalence between the three definition is proven in Louart and Couillet (2020), principally thanks to results issued from Ledoux (2005).

Remark 1 (From (Ledoux, 2005, Proposition 1.7)). *In the last item, the existence of the expectation of $f_p(Z_p)$ is guaranteed assuming any of the two other assertions. For instance*

$$\forall t > 0 : \mathbb{P}(|f_p(Z_p) - m_{f_p}| \geq t) \leq C e^{-(t/c\sigma_p)^q}$$

implies the bounding:

$$\mathbb{E}[|f_p(Z_p)|] \leq |m_{f_p}| + \mathbb{E}[|f_p(Z_p) - m_{f_p}|] \leq |m_{f_p}| + \frac{Cc\sigma_p}{q^{1/q}} < \infty;$$

the random variable $f_p(Z_p)$ is thus integrable and admits an expectation (there always exists a median $m_{f_p} \in \mathbb{R}$). For the existence of the expectation of a random vector, refer to Appendix A.

Remark 2. *It is more natural, as done in Ledoux (2005), to introduce the notion of concentration in metric spaces, because one only needs to resort to Lipschitz mappings which merely require a metric structure on E to be defined.*

¹A random vector Z of E is a measurable function from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the normed vector space $(E, \|\cdot\|)$ (endowed with the Borel σ -algebra); one should indeed write $Z : \Omega \rightarrow E$, but we abusively simply denote $Z \in E$.

However, our central Theorems (Theorem 6 and 9) involve the expectation of random vectors which can only be defined in vector spaces.

At one point in the course of the article, it will be useful to invoke concentration for semi-norms in place of norms. Definition 1 is still consistent for these weaker objects. Recall that a seminorm $\|\cdot\|' : E \mapsto \mathbb{R}$ is a functional satisfying:

1. $\forall x, y \in E : \|x + y\|' \leq \|x\|' + \|y\|'$
2. $\forall x \in E, \forall \alpha \in \mathbb{R} : \|\alpha x\|' = |\alpha| \|x\|'$

(it becomes a norm if in addition $\|x\|' = 0 \Rightarrow x = 0$).

Remark 3. In cases where a concentrated vector $Z_p \propto \mathcal{E}_q(\sigma_p)$ takes values only on some subsets $A \equiv Z_p(\Omega) \subset E_p$, it might be useful to be able to set concentration of observations $f_p(Z_p)$ where f_p is only 1-Lipschitz on A (and possibly non Lipschitz on $E_p \setminus A$). This would be an immediate consequence of Definition 1 if one would be able to continue $f_p|_A$ into a mapping $\tilde{f}_p$ Lipschitz on the whole vector space E_p but this is rarely possible. Yet, the observation $f_p(Z_p)$ do concentrates under hypothesis of Definition 1. Indeed, considering a median m_{f_p} of $f_p(Z_p)$ and the set $S_p = \{f_p \leq m_{f_p}\} \subset E_p$, if we note for any $z \in E_p$ and $U \subset E_p$, $U \neq \emptyset$, $d(z, U) = \inf\{\|z - y\|, y \in U\}$, then, we have the implications for any $z \in A$ and $t > 0$:

$$\begin{aligned} f_p(z) \geq m_{f_p} + t & \implies d(z, S_p) \geq t \\ f_p(z) \leq m_{f_p} - t & \implies d(z, S_p^c) \geq t, \end{aligned}$$

since f_p is 1-Lipschitz on A . Therefore since $z \mapsto d(z, S_p)$ and $z \mapsto d(z, S_p^c)$ are both 1-Lipschitz on E and both admit 0 as a median ($\mathbb{P}(d(Z_p, S_p) \geq 0) = 1 \geq \frac{1}{2}$ and $\mathbb{P}(d(Z_p, S_p) \leq 0) \geq \mathbb{P}(f_p(Z_p) \leq m_{f_p}) \geq \frac{1}{2}$). Therefore:

$$\begin{aligned} \mathbb{P}(|f_p(Z_p) - m_{f_p}| \geq t) & \leq \mathbb{P}(d(Z_p, S_p) \geq t) + \mathbb{P}(d(Z_p, S_p) \geq t) \\ & \leq 2Ce^{-(t/c\sigma_p)}. \end{aligned}$$

One can then deduce from this inequality similar bounds for $\mathbb{P}(|f_p(Z_p) - f_p(Z'_p)| \geq t)$ and $\mathbb{P}(|f_p(Z_p) - \mathbb{E}[f_p(Z_p)]| \geq t)$.

One could then argue, that we could have taken instead of Definition 1 hypotheses concerning the concentration of Z_p on $Z_p(\Omega)$ only; we thought however that the definition was already sufficiently complex and preferred to precise the notion in a side remark. This aspect should however be kept in mind since it will be exploited to set the concentration of products of random vectors (in particular).

A simple but fundamental consequence of Definition 1 is that any Lipschitz transformation of a concentrated vector is also a concentrated vector. The Lipschitz coefficient of the transformation controls the concentration.

Proposition 1. In the setting of Definition 1, given a sequence $(\lambda_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$, a supplementary sequence of normed vector spaces $(E'_p, \|\cdot\|'_p)_{p \geq 0}$ and a sequence of λ_p -Lipschitz transformations $F_p : (E_p, \|\cdot\|_p) \rightarrow (E'_p, \|\cdot\|'_p)$, we have

$$Z_p \propto \mathcal{E}_q(\sigma_p) \implies F_p(Z_p) \propto \mathcal{E}_q(\lambda_p \sigma_p).$$

There exists a wide range of concentrated random vectors that can be found for instance in (Ledoux, 2005). We recall below some of the major examples. In the following theorems, we only consider sequences of random vectors of the vector spaces $(\mathbb{R}^p, \|\cdot\|)$. We will omit the index p to simplify the readability of the results.

Theorem 1 (Fundamental examples of concentrated vectors). *The following sequences of random vectors are concentrated and satisfy $Z \propto \mathcal{E}_2$:*

- Z is uniformly distributed on the sphere $\sqrt{p}\mathbb{S}^{p-1}$.
- $Z \sim \mathcal{N}(0, I_p)$ has independent Gaussian entries.
- Z is uniform distributed on the ball $\sqrt{p}\mathcal{B} = \{x \in \mathbb{R}^p, \|x\| \leq \sqrt{p}\}$.
- Z is uniform distributed on $[0, \sqrt{p}]^p$.
- Z has the density $d\mathbb{P}_Z(z) = e^{-U(z)}d\lambda_p(z)$ where $U : \mathbb{R}^p \rightarrow \mathbb{R}$ is a positive functional with Hessian bounded from below by, say, cI_p with $c = O(1)$ and $d\lambda_p$ is the Lebesgue measure on $\mathbb{R}^p$.

Remark 4 (Concentration and observable diameter). *The notion of “observable diameter” (the diameter of the observations) introduced in Definition 1 should be compared to the diameter of the distribution or “metric diameter” which could be naturally defined as the expectation of the distance between two independent random vectors drawn from the same distribution. The “concentration” of a random vector can then be interpreted as a difference of concentration rates between the observable diameter and the metric diameter through dimensionality. For instance, Theorem 1 states that the observable diameter of a Gaussian distribution in $\mathbb{R}^p$ is of order 1, that is to say $\frac{1}{\sqrt{p}}$ times less than the metric diameter (that is of order $\sqrt{p}$): Gaussian vectors are indeed concentrated.*

As a counter example of a non concentrated vectors, one may consider the random vector $Z = [X, \dots, X] \in \mathbb{R}^p$ where $X \sim \mathcal{N}(0, 1)$. Here the metric diameter is of order $O(\sqrt{p})$, which is the same as the diameter of the observation $\frac{1}{\sqrt{p}}(X + \dots + X)$ (the mapping $(z_1, \dots, z_p) \mapsto \frac{1}{\sqrt{p}}(z_1 + \dots + z_p)$ is 1-Lipschitz).

Remark 5 ($(q \neq 2)$ -exponential concentration). *We provide here two examples of q -exponential concentration where $q \neq 2$ and the underlying metric space is not necessarily Euclidean:*

- (Talagrand, 1995) if Z is a random vectors of $\mathbb{R}^p$ with independent entries having density $\frac{1}{2}e^{-|\cdot|}d\lambda_1$, then $Z \propto \mathcal{E}_1$.
- (Ledoux, 2005) if Z is uniformly distributed on the balls

$$\mathcal{B}_{\|\cdot\|_q} = \left\{ x \in \mathbb{R}^p \mid \|x\|_q = \left(\sum_{i=1}^p x_i^q \right)^{1/q} \leq 1 \right\} \subset \mathbb{R}^p$$

then $Z \propto \mathcal{E}_q \left(p^{-\frac{1}{q}} \right)$.

A very explicit characterization of *exponential concentration* is given by a bound on the different centered moments.

Proposition 2. *A random vector $Z \in E$ is q -exponentially concentrated with observable diameter of order σ (i.e., $Z \propto \mathcal{E}_q(\sigma)$) if and only if there exists $C \geq 1$ and $c = O(\sigma)$ such that for any (sequence of) 1-Lipschitz functions $f : E \rightarrow \mathbb{R}$:*

$$\forall r \geq q : \mathbb{E} [|f(Z) - f(Z')|^r] \leq C \left(\frac{r}{q} \right)^{\frac{r}{q}} c^r, \quad (4)$$

where Z' is an independent copy of Z . Inequality (4) also holds if we replace $f(Z')$ with $\mathbb{E}[f(Z)]$ (of course the constants C and c might be slightly different).

1.2. Linear concentration, notations and properties of the first statistics

Although it must be clear that a concentrated vector Z is generally far from its expectation (keep the Gaussian case in mind), it can still be useful to have some control on $\|Z - \mathbb{E}[Z]\|$. That can be done for a larger class than the class of Lipschitz concentrated random vector: the class of *linearly concentrated random vectors*. It is a weaker notion but still relevant since, as it will be shown in various examples (Propositions ??, 12 and 13, Theorems 6, 7 and 12), it can become the residual concentration property satisfied by non Lipschitz operations on Lipschitz concentrated (or convexly concentrated – see Subsection 1.7) random vectors.

Definition 2. *Given a sequence of normed vector spaces $(E_p, \|\cdot\|_p)_{p \geq 0}$, a sequence of random vectors $(Z_p)_{p \geq 0} \in \prod_{p \geq 0} E_p$, a sequence of deterministic vectors $(\tilde{Z}_p)_{p \geq 0} \in \prod_{p \geq 0} E_p$, a sequence of positive reals $(\sigma_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say that Z_p is q -exponentially linearly concentrated around the deterministic equivalent $\tilde{Z}_p$ with an observable diameter of order $O(\sigma_p)$ iff there exist two constants $c, C > 0$ such that $\forall p \in \mathbb{N}$ and for any unit-normed linear form $f \in E'_p$ ($\forall p \in \mathbb{N}, \forall x \in E : |f(x)| \leq \|x\|$):*

$$\forall t > 0 : \mathbb{P} \left(\left| f(Z_p) - f(\tilde{Z}_p) \right| \geq t \right) \leq C e^{(t/c\sigma_p)^q}.$$

If this holds, we write $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$. When it is not necessary to mention the deterministic equivalent, one can write simply $Z \in \mathcal{E}_q(\sigma)$.

Of course linear concentration is stable through affine transformations. Given two normed vector spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, we denote $\mathcal{L}(E, F)$ the set of continuous linear mappings from E to F that we endow with the operator norm $\|\cdot\|_{\mathcal{L}(E, F)}$ defined as:

$$\forall \phi \in \mathcal{L}(E, F) : \quad \|\phi\|_{\mathcal{L}(E, F)} = \sup_{\|x\|_E \leq 1} \|\phi(x)\|_F.$$

When $E = F$ and $\|\cdot\|_E = \|\cdot\|_F = \|\cdot\|$, we note $\|\cdot\| = \|\cdot\|_{\mathcal{L}(E, E)}$ for simplicity, and it is an algebra-norm on $\mathcal{L}(E, F)$: for any $\phi, \psi \in \mathcal{L}(E, E)$: $\|\phi \circ \psi\| \leq \|\phi\| \|\psi\|$.

We equivalently denote $\mathcal{A}(E, F)$ the set of continuous affine mappings from E to F and we endow it with the norm:

$$\forall \phi \in \mathcal{A}(E, F) : \|\phi\|_{\mathcal{A}(E, F)} = \|\mathcal{L}(\phi)\|_{\mathcal{L}(E, F)} + \|\phi(0)\|_F \quad \text{where } \mathcal{L}(\phi) = \phi - \phi(0).$$

Proposition 3. *Given two normed vector spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, a random vector $Z \in E$, a deterministic vector $\tilde{Z} \in E$ and an affine mapping $\phi \in \mathcal{A}(E, F)$ such that $\|\mathcal{L}(\phi)\|_{\mathcal{L}} \leq \lambda$:*

$$Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma) \quad \implies \quad \phi(Z) \in \phi(\tilde{Z}) \pm \mathcal{E}_q(\lambda\sigma).$$

When the expectation can be defined (see Appendix A), we can deduce from Proposition 40 and Lemma 29 that there exists an implication relation between Lipschitz concentration (Definitions 1) and linear concentration (2).

Lemma 1. *Given a reflexive space $(E, \|\cdot\|)$ and a random vector $Z \in E$, we have the implication:*

$$Z \propto \mathcal{E}_q(\sigma) \quad \implies \quad Z \in \mathbb{E}[Z] \pm \mathcal{E}_q(\sigma).$$

A similar implication exists for random mappings thanks to the basic generalization explained in Remark 29. The next lemma is a formal expression of the assessment that “any deterministic vector located at a distance smaller than the observable diameter to a deterministic equivalent is also a deterministic equivalent”.

Lemma 2. *Given a random vector $Z \in E$, a deterministic vector $\tilde{Z} \in E$ such that $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$, we have then the equivalence:*

$$Z \in \tilde{Z}' \pm \mathcal{E}_q(\sigma) \quad \iff \quad \|\tilde{Z} - \tilde{Z}'\| \leq O(\sigma)$$

When we are just interested in the size of the deterministic equivalent, we employ the notation:

$$Z \in O(\theta) \pm \mathcal{E}_q(\sigma)$$

if $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$ and $\|\tilde{Z}\| \leq O(\theta)$ (for some deterministic vector $\tilde{Z}$). The previous lemma leads to the implication:

$$Z \in O(\sigma) \pm \mathcal{E}_q(\sigma) \quad \implies \quad Z \in 0 \pm \mathcal{E}_q(\sigma).$$

Given two random vectors $Z, W \in E$, we also allow ourselves to write:

$$Z \in W \pm \mathcal{E}_2(\sigma)$$

iff $Z - W \in O(\sigma) \pm \mathcal{E}_2(\sigma)$, W is then called a *random equivalent* of Z . In $\mathbb{R}^p$, two randomly equivalent vectors have similar expectations:

$$\|\mathbb{E}[Z] - \mathbb{E}[W]\| \leq \sup_{\|u\| \leq 1} u^T \mathbb{E}[u^T(Z - W)] \leq O(\sigma),$$

thanks to Proposition 2. However their covariance could be more different, to have a similarity on higher order statistic, one needs a stronger property : the strong equivalence. We say that W is a *strong random equivalent* of Z if $\|Z - W\| \in 0 \pm \mathcal{E}_2(\sigma)$, we have a similar result to Lemma 2:

Lemma 3. *Given two (sequence of) random vectors Z, W and a (sequence of) positive number $\sigma > 0$, if $Z \in \mathcal{E}_q(\sigma)$ and $\|Z - W\| \in O(\sigma) \pm \mathcal{E}_q(\sigma)$ then $Z \in W \pm \mathcal{E}_q(\sigma)$.*

Remark 6. *In $\mathbb{R}^p$ two strongly equivalent random vectors with observable diameter of order $O(1/\sqrt{p})$ have similar covariance in nuclear norm (and similar expectation). Indeed, if $\|Z - W\| \in 0 \pm \mathcal{E}_2(1/\sqrt{p})$, and $Z, W \in \mathcal{E}_2(1/\sqrt{p})$:*

$$\begin{aligned} & \|\mathbb{E}[ZZ^T] - \mathbb{E}[Z]\mathbb{E}[Z]^T - \mathbb{E}[WW^T] + \mathbb{E}[W]\mathbb{E}[W]^T\|_* \\ & \leq \mathbb{E}[\|ZZ^T - \mathbb{E}[Z]\mathbb{E}[Z]^T - WW^T + \mathbb{E}[W]\mathbb{E}[W]^T\|_*] \\ & \leq \mathbb{E}[\|Z - W\| \|Z - \mathbb{E}[Z]\| + \|Z - W\| \|W - \mathbb{E}[W]\|] \leq O(1/\sqrt{p}), \end{aligned}$$

thanks to Hölder inequality, and thanks to Proposition 8 given below that states that $\|Z - \mathbb{E}[Z]\|, \|W - \mathbb{E}[W]\| \in O(1) \pm \mathcal{E}_2(1)$.

Those new notation will be extensively employed with random variables or low dimensional random vectors for which the notion of concentration is very simple.

Remark 7. *For random variables, or low rank random vectors, the notions of Lipschitz concentration and linear concentration are equivalent. More over, if Z is a random variable satisfying $Z \in \mathcal{E}_q(\sigma)$, for any 1-Lipschitz mapping $f : \mathbb{R} \rightarrow \mathbb{R}$, we have:*

$$f(Z) \in f(\mathbb{E}[Z]) \pm \mathcal{E}_q(\sigma).$$

Indeed $f(Z) \in \mathbb{E}[f(Z)] \pm \mathcal{E}_q(\sigma)$ and:

$$|\mathbb{E}[f(Z)] - f(\mathbb{E}[Z])| \leq \mathbb{E}[|f(Z) - f(\mathbb{E}[Z])|] \leq \mathbb{E}[|Z - \mathbb{E}[Z]|] = O(\sigma),$$

thanks to Proposition 2. The same holds for a random vector $Z = (Z_1, \dots, Z_d) \in \mathbb{R}^d$ if $d \leq n$, because we can bound:

$$\mathbb{E}[|f(Z) - f(\mathbb{E}[Z])|] \leq \sqrt{\sum_{i=1}^d \mathbb{E}[(f(Z_i) - f(\mathbb{E}[Z_i]))^2]} \leq \sqrt{d}O(\sigma) = O(\sigma),$$

thanks again to Proposition 2, since for all $i \in [d]$, $Z_i \in \mathcal{E}_q(\sigma)$

We end this subsection with a precise characterization of the linearly concentrated random vectors of the (sequence of) normed vector space $\mathbb{R}^p$ thanks to a bound on the moments, as we did in Proposition 2.

Definition 3 (Moments of random vectors). Given a random vector $X \in \mathbb{R}^p$ and an integer $r \in \mathbb{N}$, we call the “ r^{th} moment of X ” the symmetric r -linear form $C_r^X : (\mathbb{R}^p)^r \rightarrow \mathbb{R}$ defined for any $u_1, \dots, u_r \in \mathbb{R}^p$ with:

$$C_r^X(u_1, \dots, u_p) = \mathbb{E} \left[\prod_{i=1}^p (u_i^T X - \mathbb{E}[u_i^T X]) \right].$$

When $r = 2$, we retrieve the covariance matrix.

Given an r -linear form S of $\mathbb{R}^p$ we note its operator norm:

$$\|S\| \equiv \sup_{\|u_1\|, \dots, \|u_r\| \leq 1} S(u_1, \dots, u_p),$$

when S is symmetric we employ the simpler formula $\|S\| = \sup_{\|u\| \leq 1} S(u, \dots, u)$. We have then the following characterization that we give without proof since it is a simple consequence of the definition of linearly concentrated random vectors in Proposition 2.

Proposition 4. Given $q > 0$, a sequence of random vectors $X_p \in \mathbb{R}^p$, and a sequence of positive numbers $\sigma_p > 0$, we have the following equivalence:

$$X \in \mathcal{E}_q(\sigma) \iff \exists C, c > 0, \forall p \in \mathbb{N}, \forall r \geq q : \|C_r^{X_p}\| \leq C \left(\frac{r}{q} \right)^{\frac{r}{q}} (c\sigma_p)^r$$

In particular, if we note $C = \mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X]^T$, the covariance of $X \in \mathcal{E}_q(\sigma)$, we see that $\|C\| \leq O(\sigma)$.

1.3. Linear concentration through sums and integrals

The second basic result on linear concentration allows us to follow the concentration rate through concatenation of possibly dependent random vectors. An interesting fact is that independence is generally not a relevant property when trying to set concentration inequalities from linear concentration hypotheses.

Proposition 5 (Louart and Couillet (2019)). Given two sequences $m \in \mathbb{N}^{\mathbb{N}}$ and $\sigma \in \mathbb{R}_+^{\mathbb{N}}$, a constant q , m sequences of normed vector spaces $(E_i, \|\cdot\|_i)_{1 \leq i \leq m}$, m sequences of deterministic vectors $\tilde{Z}_1 \in E_1, \dots, \tilde{Z}_m \in E_m$, and m sequences of random vectors $Z_1 \in E_1, \dots, Z_m \in E_m$ (possibly dependent) satisfying, for any $i \in \{1, \dots, m\}$, $Z_i \in \tilde{Z}_i \pm \mathcal{E}_q(\sigma)$, we have the concentration :

$$(Z_1, \dots, Z_m) \in (\tilde{Z}_1, \dots, \tilde{Z}_m) \pm \mathcal{E}_q(\sigma), \quad \text{in } (E, \|\cdot\|_{\ell^\infty}),$$

where we introduced on $E \equiv E_1 \times \dots \times E_m$ the norm $\|\cdot\|_{\ell^\infty}$ satisfying, for any $(z_1, \dots, z_m) \in E$: $\|(z_1, \dots, z_m)\|_{\ell^\infty} = \sup_{1 \leq i \leq m} \|z_i\|_i$.

If we want to consider the concatenation of vectors with different observable diameter, it is more convenient to look at the concentration in a space $(\prod_{i=1}^m E_i, \ell^r)$, for a given $r > 0$, where, for any $(z_1, \dots, z_m) \in \prod_{i=1}^m E_i$:

$$\|(z_1, \dots, z_m)\|_{\ell^r} = \left(\sum_{i=1}^m \|z_i\|_i^r \right)^{1/r}.$$

Corollary 1. *Given two constants $q, r > 0$, $m \in \mathbb{N}$, $\sigma_1, \dots, \sigma_m \in (\mathbb{R}_+^{\mathbb{N}})^m$, m sequences of $(E_i, \|\cdot\|_i)_{1 \leq i \leq m}$, m sequences of deterministic vectors $\tilde{Z}_1 \in E_1, \dots, \tilde{Z}_m \in E_m$, and m sequences of random vectors $Z_1 \in E_1, \dots, Z_p \in E_p$ (possibly dependent) satisfying, for any $i \in \{1, \dots, m\}$, $Z_i \in \tilde{Z}_i \pm \mathcal{E}_q(\sigma_i)$, we have the concentration :*

$$(Z_1, \dots, Z_m) \in (\tilde{Z}_1, \dots, \tilde{Z}_m) \pm \mathcal{E}_q(\|\sigma\|_r), \quad \text{in } (E, \|\cdot\|_{\ell^r}),$$

Remark 8. *When $E_1 = \dots = E_m = E$, then for any vector $a = (a_1, \dots, a_m) \in \mathbb{R}_+^m$, we knew from Corollary 1 that:*

$$\sum_{i=1}^m a_i Z_i \in \sum_{i=1}^m a_i \tilde{Z}_i \pm \mathcal{E}_2(|a|^T \sigma),$$

where $|a| = (|a_1|, \dots, |a_m|) \in \mathbb{R}_+^m$

PROOF. We already know from Proposition 5 that:

$$\left(\frac{Z_1}{\sigma_1}, \dots, \frac{Z_m}{\sigma_m} \right) \in \left(\frac{\tilde{Z}_1}{\sigma_1}, \dots, \frac{\tilde{Z}_m}{\sigma_m} \right) \pm \mathcal{E}_q, \quad \text{in } (E, \|\cdot\|_{\ell^\infty})$$

Let us then consider the linear mapping:

$$\begin{aligned} \phi : (E, \|\cdot\|_{\ell^\infty}) &\longrightarrow (E, \|\cdot\|_{\ell^r}) \\ (z_1, \dots, z_m) &\longmapsto (\sigma_1 z_1, \dots, \sigma_m z_m), \end{aligned}$$

the Lipschitz character of ϕ is clearly $\|\sigma\|_r = (\sum_{i=1}^m \sigma_i^r)^{1/r}$, and we can deduce the concentration of $Z = \phi(\sigma_1 Z_1, \dots, \sigma_m Z_m)$.

Corollary 1 is very useful to set the concentration of infinite series of concentrated random variables. This is settled thanks to an elementary result issued from Louart and Couillet (2020) that sets that the observable diameter of a limit of random vectors is equal to the limit of the observable vectors. Be careful that rigorously, there are two indexes, p coming from Definition 1 that only describes the concentration of sequences of random vectors, and n particular to this lemma that will tend to infinity.

Lemma 4. *Given a sequence of random vectors $(Z_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ and a sequence of positive reals $(\sigma_n)_{n \in \mathbb{N}} \in \mathbb{R}_+^{\mathbb{N}}$ such that:*

$$Z_n \propto \mathcal{E}_q(\sigma_n),$$

if we assume that $(Z_n)_{n \in \mathbb{N}}$ converges in law² when n tends to infinity to a random vector $(Z_\infty) \in E$ and that $\sigma_n \xrightarrow[n \rightarrow \infty]{} \sigma_\infty$ then:

$$Z_\infty \propto \mathcal{E}_q(\sigma_\infty).$$

The result also holds if we only assume linear concentration for Z_n then we obtain the linear concentration of Z_∞ .

Corollary 2. *Given two constants $q, r > 0$, $\sigma_1, \dots, \sigma_n \dots \in (\mathbb{R}_+^{\mathbb{N}})^{\mathbb{N}}$, a (sequence of) reflexive normed vector spaces $(E, \|\cdot\|)$, $\tilde{Z}_1, \dots, \tilde{Z}_n, \dots \in E^{\mathbb{N}}$ deterministic, and $Z_1, \dots, Z_n, \dots \in E^{\mathbb{N}}$ random (possibly dependent) satisfying, for any $n \in \mathbb{N}$, $Z_n \in \tilde{Z}_n \pm \mathcal{E}_q(\sigma_n)$. If we assume that $Z \equiv \sum_{n \in \mathbb{N}} Z_n$ is pointwise convergent³, that $\sum_{n \in \mathbb{N}} \tilde{Z}_n$ is well defined and that $\sum_{n \in \mathbb{N}} \sigma_i \leq \infty$, then we have the concentration :*

$$\sum_{n \in \mathbb{N}} Z_n \in \sum_{n \in \mathbb{N}} \tilde{Z}_n \pm \mathcal{E}_q \left(\sum_{n \in \mathbb{N}} \sigma_n \right), \quad \text{in } (E, \|\cdot\|),$$

PROOF. We already know from Corollary 1 that for all $N \in \mathbb{N}$:

$$\sum_{n=1}^N Z_n \in \sum_{n=1}^N \tilde{Z}_n \pm \mathcal{E}_q \left(\sum_{n \in \mathbb{N}} \sigma_n \right), \quad \text{in } (E, \|\cdot\|).$$

Thus in order to employ Lemma 4 let us note that for any bounded continuous mapping $f : E \rightarrow \mathbb{R}$, the dominated convergence theorem allows us to set that:

$$\mathbb{E} \left[f \left(\sum_{n=1}^N Z_n \right) \right] \xrightarrow[N \rightarrow \infty]{} \mathbb{E} \left[f \left(\sum_{n=1}^{\infty} Z_n \right) \right],$$

thus $(\sum_{n=1}^N Z_n)_{N \in \mathbb{N}}$ converges in law to $\sum_{n=1}^{\infty} Z_n$, which allows us to set the result of the corollary.

To study the concentration of integrated random mapping, we are going to introduce a notion of concentration under semi-norm families. The concentration under the infinity norm ($\phi \mapsto \|\phi\|_\infty \equiv \sup_{z \in E} \|\phi(z)\|$) is also relevant but this strong notion of concentration is unnecessarily restrictive.

²For any bounded sequence of continuous mapping $(f_p)_{p \geq 0} : \prod_{p \geq 0} E_p \rightarrow \mathbb{R}^{\mathbb{N}}$:

$$\sup_{p \in \mathbb{N}} |\mathbb{E}[f_p(Z_{p,n}) - \mathbb{E}[f_p(Z_{p,\infty})]]| \xrightarrow[n \rightarrow \infty]{} 0$$

³For any $w \in \Omega$, $\sum_{n \in \mathbb{N}} \|Z_n(w)\| \leq \infty$ and we define $Z(w) \equiv \sum_{n \in \mathbb{N}} Z_n(w)$

Definition 4 (Concentration with family of semi-norms). Given a sequence of normed vector spaces $(E_p)_{p \geq 0}$, a sequence of index sets Θ_p , a sequence of semi-norm families $((\|\cdot\|_t)_{t \in \Theta_p})_{p \geq 0}$, a sequence of random vectors $(Z_p)_{p \geq 0} \in \prod_{p \geq 0} E_p$, a sequence of positive reals $(\sigma_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say again that Z_p is Lipschitz q -exponentially concentrated with an observable diameter of order $O(\sigma_p)$ iff there exist two positive constants $C, c > 0$ such that $\forall p \in \mathbb{N}, \forall t \in \Theta_p$ for any $f : E_p \rightarrow \mathbb{R}$ 1-Lipschitz for the norm $\|\cdot\|_t$:

$$\mathbb{P}(|f(Z_p) - f(Z'_p)| \geq t) \leq C e^{-(t/c\sigma_p)^q} \quad (5)$$

We denote in this case $Z_p \propto \mathcal{E}_q(\sigma_p)$ in $(E_p, (\|\cdot\|_t)_{t \in \Theta_p})$.

Remark 9. Let us comment this definition.

- The important point of the notion of concentration for family of semi-norms is that the constants C and c are the same for all the semi norms.
- As in Definition 1, (5) can be replaced by concentration inequalities around a or the expectation of $f_p(Z_p)$.
- As we will see in next proposition, this definition is particularly convenient when Z_p is a random mapping of $F_p^{G_p}$ ($Z_p : G_p \rightarrow F_p$), for a sequence of sets $(G_p)_{p \geq 0}$, and a sequence of normed vector spaces $(F_p, \|\cdot\|)_{p \geq 0}$, when $\Theta_p \subset G_p$ and the semi-norms are defined with the evaluation maps defined for any $g \in G_p$ as:

$$\forall \phi \in F_p^{G_p} : \|\phi\|_g = \|\phi(g)\|. \quad (6)$$

The concentration of Z_p in $(F_p^{G_p}, (\|\cdot\|_t)_{t \in \Theta_p})$ is then equivalent to the concentration of all the $(Z_p(t))_{t \in \Theta_p}$ in $(F_p, \|\cdot\|)$ for all $t \in \Theta_p$ and for the same constants $C > 0$ and $c > 0$.

- If we introduce the random mapping $\Phi : z \mapsto XDY^T z$ where $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ are three random matrices such that $X, Y, D \propto \mathcal{E}_2$, $\|\mathbb{E}[X]\| \leq O(\sqrt{n+p})$, $\|\mathbb{E}[Y]\| \|\mathbb{E}[D]\| \leq O(1)$, then we will see in Corollary 4 below that for any deterministic vector $z \in \mathbb{R}^p$ such that $\|z\| \leq 1$, $\Phi(z)$ is concentrated and has an observable diameter of order $O(\sqrt{(n+p)\log(n)})$, but we can not obtain a smaller observable diameter than $O(n+p)$ for the random matrix XDY^T and consequently for $\|\Phi\|_\infty$. As such, the concentration of mapping Φ with the infinity norm is looser than the concentration with the set of semi-norms $(\|\cdot\|)_{\|z\| \leq 1}$. This important remark justifies fully the introduction of this new notion, although next Proposition could have been stated with hypotheses of concentration under infinity norm.

Proposition 6 (Concentration of the integration). Given a (sequence of) finite dimensional⁴ normed vector space $(F, \|\cdot\|)$, a (sequence of) integrable

⁴So that we can integrate $\Phi(\cdot)$

random mappings $\Phi : [0, 1] \rightarrow F$, if there exists a (sequence of) positive reals $\kappa > 0$ such that for any $t \in [0, 1]$, $\|\Phi(t)\| \leq \kappa$ and $\Phi \propto \mathcal{E}_q$ in $(F^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]})$ then:

$$\int_0^1 \Phi(t) dt \in \mathcal{E}_q \quad \text{in } (F, \|\cdot\|).$$

PROOF. For any $n \in \mathbb{N}$, we already know from Corollary 8 that:

$$\frac{1}{n} \sum_{i=1}^n \phi\left(\frac{k}{n}\right) \in \mathcal{E}_q.$$

Besides, for any bounded real valued continuous mapping $f : F \rightarrow \mathbb{R}$, the dominated convergence theorem (employed two times – for the integral and the expectation) ensures the convergence:

$$\mathbb{E} \left[f \left(\frac{1}{n} \sum_{i=1}^n \phi\left(\frac{k}{n}\right) \right) \right] = \mathbb{E} \left[f \left(\int_0^1 \phi_n(t) dt \right) \right] \xrightarrow{n \rightarrow \infty} \mathbb{E} \left[f \left(\int_0^1 \phi(t) dt \right) \right],$$

where $\phi_n(t) = \phi_n(k/n)$ if $k \leq \lfloor nt \rfloor \leq k+1$ ($\mathbb{E}[f(\phi_n(\cdot))]$ pointwise converges to $\mathbb{E}[f(\phi(\cdot))]$). Therefore, $\frac{1}{n} \sum_{i=1}^n \phi\left(\frac{k}{n}\right)$ converges in law to $\int_0^1 \phi(t) dt$, and we can conclude with Lemma 4.

1.4. Concentration of the norm

Given a random vector $Z \in (E, \|\cdot\|)$, if $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$, the control on the norm $\|Z - \tilde{Z}\|$ can be done easily when the norm $\|\cdot\|$ can be defined as the supremum on a set of linear forms; for instance when $(E, \|\cdot\|) = (\mathbb{R}^p, \|\cdot\|_\infty)$: $\|x\|_\infty = \sup_{1 \leq i \leq p} e_i^T x$ (where $(e_1, \dots, e_p)$ is the canonical basis of $\mathbb{R}^p$). We can then bound:

$$\begin{aligned} \mathbb{P} \left(\|Z - \tilde{Z}\|_\infty \geq t \right) &= \mathbb{P} \left(\sup_{1 \leq i \leq p} e_i^T (Z - \tilde{Z}) \geq t \right) \\ &\leq \min \left(1, p \sup_{1 \leq i \leq p} \mathbb{P} \left(e_i^T (Z - \tilde{Z}) \geq t \right) \right) \\ &\leq \min \left(1, p C e^{-(t/c)^q} \right) \leq \max(C, e) \exp \left(-\frac{t^q}{2c^q \log(p)} \right), \end{aligned}$$

for some $c = O(\sigma)$ and some constant $C > 0$. To manage the infinity norm, the supremum is taken on a finite set $\{e_1, \dots, e_p\}$.

Problems arise when considering the Euclidean norm satisfying for any $x \in \mathbb{R}^p$ the identity $\|x\| = \sup\{u^T x, \|u\| \leq 1\}$, indeed, here the supremum is taken on the whole unit ball $\mathcal{B}_{\mathbb{R}^p} \equiv \{u \in \mathbb{R}^p, \|u\| \leq 1\}$ which is an infinite set. This loss of cardinality control can be overcome if one introduces so-called ε -nets to discretize the ball with a net $\{u_i\}_{i \in I}$ (with I finite – $|I| < \infty$) in order to simultaneously

1. approach sufficiently the norm to ensure

$$\mathbb{P} \left(\|Z - \tilde{Z}\|_\infty \geq t \right) \approx \mathbb{P} \left(\sup_{i \in I} u_i^T(Z - \tilde{Z}) \geq t \right),$$

2. control the cardinality $|I|$ for the inequality

$$\mathbb{P} \left(\sup_{i \in I} u_i^T(Z - \tilde{Z}) \geq t \right) \leq |I| \mathbb{P} \left(u_i^T(Z - \tilde{Z}) \geq t \right)$$

not to be too loose (see Tao (2012) for more detail).

One can then show:

$$\mathbb{P}(\|Z - \tilde{Z}\| \geq t) \leq \max(C, e) \exp^{-(t/c)^q/p}. \quad (7)$$

The approach with ε -nets in $(\mathbb{R}^p, \|\cdot\|)$ can be generalized to any normed vector space $(E, \|\cdot\|)$ where the norm can be written as a supremum through an identity of the kind

$$\forall x \in E : \|x\| = \sup_{f \in H} f(x) \quad \text{with } H \subset E' \text{ and } \dim(\text{Vect}(H)) < \infty, \quad (8)$$

for a given $H \subset E'$ and where $\text{Vect } H$ designates the subspace of E generated by H . Such a $H \subset E'$ exists in particular when $(E, \|\cdot\|)$ is a reflexive spaces.

Proposition 7 (James (1957)). *In a reflexive space $(E, \|\cdot\|)$:*

$$\forall x \in E : \|x\| = \sup_{f \in \mathcal{B}_{E'}} f(x) \text{ where } \mathcal{B}_{E'} = \{f \in E' \mid \|f\| \leq 1\}.$$

When $(E, \|\cdot\|)$ has an infinite dimension and is not reflexive, it is sometimes possible to establish (8) for some $H \subset E$ in some cases (most of them appearing when $\|\cdot\|$ is a semi-norm. Without going deeper into details, we introduce the notion of *norm degree* that will help us adapting to other normed vector space the concentration rate p appearing in the exponential term of concentration inequality (7) (concerning $(\mathbb{R}^p, \|\cdot\|)$).

Definition 5 (Norm degree). *Given a normed (or seminormed) vector space $(E, \|\cdot\|)$, and a subset $H \subset E'$, the degree η_H of H is defined as :*

- $\eta_H \equiv \log(|H|)$ if H is finite,
- $\eta_H \equiv \dim(\text{Vect } H)$ if H is infinite.

If there exists a subset $H \subset E'$ such that (8) is satisfied, we then denote $\eta(E, \|\cdot\|)$, or more simply $\eta_{\|\cdot\|}$, the degree of $\|\cdot\|$, defined as :

$$\eta_{\|\cdot\|} = \eta(E, \|\cdot\|) \equiv \inf \left\{ \eta_H, H \subset E' \mid \forall x \in E, \|x\| = \sup_{f \in H} f(x) \right\}.$$

Example 1. We can give some examples of norm degrees :

- $\eta(\mathbb{R}^p, \|\cdot\|_\infty) = \log(p)$ ($H = \{x \mapsto e_i^T x, 1 \leq i \leq p\}$),
- $\eta(\mathbb{R}^p, \|\cdot\|) = p$ ($H = \{x \mapsto u^T x, u \in \mathcal{B}_{\mathbb{R}^p}\}$),
- $\eta(\mathcal{M}_{p,n}, \|\cdot\|) = n + p$ ($H = \{M \mapsto u^T M v, (u, v) \in \mathcal{B}_{\mathbb{R}^p} \times \mathcal{B}_{\mathbb{R}^n}\}$),
- $\eta(\mathcal{M}_{p,n}, \|\cdot\|_F) = np$ ($H = \{M \mapsto \text{Tr}(AM), A \in \mathcal{M}_{n,p}, \|A\|_F \leq 1\}$),
- $\eta(\mathcal{M}_{p,n}, \|\cdot\|_*) = np$ ($H = \{M \mapsto \text{Tr}(AM), A \in \mathcal{M}_{n,p}, \|A\| \leq 1\}$)⁵.

Depending on the vector space we are working in, we can then employ those different examples and the following proposition to set the concentration of the norm of a random vector.

Proposition 8. Given a reflexive vector space $(E, \|\cdot\|)$ and a concentrated vector $Z \in E$ satisfying $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$:

$$\|Z - \tilde{Z}\| \in O\left(\eta_{\|\cdot\|}^{1/q} \sigma\right) \pm \mathcal{E}_q\left(\eta_{\|\cdot\|}^{1/q} \sigma\right).$$

Remark 10. When $Z \propto \mathcal{E}_q(\sigma)$ (or $Z \propto_c \mathcal{E}_q(\sigma)$, as we will see in Subsection 1.7), we have of course the better concentration $\|Z - \tilde{Z}\| \propto \mathcal{E}_q(\sigma)$ but the bound $\mathbb{E}[\|Z - \tilde{Z}\|] \leq O\left(\eta_{\|\cdot\|}^{1/q} \sigma\right)$ can not be improved.

Example 2. Given two random vectors $Z \in \mathbb{R}^p$ and $M \in \mathcal{M}_{p,n}$:

- if $Z \propto \mathcal{E}_2$ in $(\mathbb{R}^p, \|\cdot\|)$: $\mathbb{E}\|Z\| \leq \|\mathbb{E}[Z]\| + O(\sqrt{p})$,
- if $M \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|)$: $\mathbb{E}\|M\| \leq \|\mathbb{E}[M]\| + O(\sqrt{p+n})$,
- if $M \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$: $\mathbb{E}\|M\| \leq \|\mathbb{E}[M]\|_F + O(\sqrt{pn})$.
- if $M \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_*)$: $\mathbb{E}\|M\|_* \leq \|\mathbb{E}[M]\|_* + O(\sqrt{pn})$.

1.5. Concentration of basic operations

Returning to Lipschitz concentration, if we want to control the concentration of the sum $X + Y$ or the product XY of two random vectors X and Y , we first need to express the concentration of the concatenation (X, Y) . This last result is very easy to obtain in the class of linearly concentrated random vector since it is a consequence of Proposition 5 (but the concentration of the product is impossible to set with good observable diameter). In the class of Lipschitz concentrated vectors, the concentration of (X, Y) is far more complicated, and independence here plays a central role (unlike for linear concentration).

To understand the issue, let us give an example where X and Y are concentrated but not (X, Y) . Consider X , uniformly distributed on the sphere $\sqrt{p}\mathbb{S}^{p-1}$

⁵ $\|\cdot\|_*$ is the nuclear norm defined for any $M \in \mathcal{M}_{p,n}$ as $\|M\|_* = \text{Tr}(\sqrt{MM^T})$ it is the dual norm of $\|\cdot\|$, which means that for any $A, B \in \mathcal{M}_{p,n}$, $\text{Tr}(AB^T) \leq \|A\| \|B\|_*$.

and $Y = f(X)$ where for any $x = (x_1, \dots, x_p) \in \mathbb{R}^p$, $f(x) = x$ if $x_1 \geq 0$ and $f(x) = -x$ otherwise. We know that all the linear observations of $X + Y$ are concentrated thanks to Proposition 5, but it is not the case for all the Lipschitz observations. Indeed, it is straight forward to see that the diameter of the observation $\|X + Y\|$ (which is a $\sqrt{2}$ -Lipschitz transformation of (X, Y)) is of order $O(\sqrt{p})$ like the metric diameter of $X + Y$ that contradicts the description of the concentration made in Remark 4. This effect is due to the fact that the mapping f is clearly not Lipschitz, and Y in a sense “defies” X (see Louart and Couillet (2019) for more details).

Still there exists two simple ways to obtain the concentration of (X, Y) , the first one being deduced from any identity $(X, Y) = \phi(Z)$ with Z concentrated and ϕ Lipschitz. It is also possible to deduce the concentration of (X, Y) from the concentration of X and Y when they are independent.

Lemma 5. *Given $(E, \|\cdot\|)$, a sequence of normed vector spaces and two sequences of independent random vectors $X, Y \in E$, if we suppose that $X \propto \mathcal{E}_q(\sigma)$ and $Y \propto \mathcal{E}_r(\rho)$ (where $q, r > 0$ are two positive constants and $\sigma, \rho \in \mathbb{R}_+^{\mathbb{N}}$ are two sequences of positive reals):*

$$(X, Y) \propto \mathcal{E}_q(\sigma) + \mathcal{E}_r(\rho) \quad \text{in } (E^2, \|\cdot\|_{\ell^\infty}),$$

where as in Proposition 5, we note for all $x, y \in E^2$, $\|(x, y)\|_{\ell^\infty} = \max(\|x\|, \|y\|)$ ⁶. Following our formalism, this means that there exist two positive constants $C, c > 0$ such that $\forall p \in \mathbb{N}$ and for any 1-Lipschitz function $f : (E_p^2, \|\cdot\|_{\ell^\infty}) \rightarrow (\mathbb{R}, |\cdot|)$, $\exists d_p = O(\rho_p), \forall t > 0 :$

$$\mathbb{P}(|f(X_p, Y_p) - f(X'_p, Y'_p)| \geq t) \leq C e^{(t/c\sigma_p)^q} + C e^{(t/cd_p)^r}.$$

The sum being a 2-Lipschitz operation (for the norm $\|\cdot\|_{\ell^\infty}$), the concentration of $X + Y$ is easy to handle and directly follows from Lemma 5. To treat the product of two vectors, we provide a general result of concentration for the m -linear application on normed vector space.

Theorem 2. *Given a (sequence of) integers m , let us consider:*

- m (sequence of) normed vector spaces $(E_1, \|\cdot\|_1), \dots, (E_m, \|\cdot\|_m)$.
- m (sequence of) norms (or semi-norms) $\|\cdot\|'_1, \dots, \|\cdot\|'_m$ respectively defined on $E_1, \dots, E_m$.
- m (sequence of) random vectors $Z_1 \in E_1, \dots, Z_m \in E_m$ satisfying

$$Z \equiv (Z_1, \dots, Z_m) \propto \mathcal{E}_q(\sigma),$$

⁶One could have also considered a big number of equivalent norms like $\|(x, y)\|_{\ell^1} = \|x\| + \|y\|$ or $\|(x, y)\|_{\ell^2} = \sqrt{\|x\|^2 + \|y\|^2}$.

for some (sequence of) positive number $\sigma \in \mathbb{R}_+$ and for both norms $\|z_1, \dots, z_m\|_{\ell^\infty} = \sup_{i=1}^m \|z_i\|_i$ and $\|z_1, \dots, z_m\|'_{\ell^\infty} = \sup_{i=1}^m \|z_i\|'_i$ defined on $E = E_1 \times \dots \times E_m$ ⁷. For all $i \in [m]$, we denote for simplicity

$$\mu_i = \mathbb{E}[\|Z_i\|'_i].$$

- a (sequence of) normed vector space $(F, \|\cdot\|)$, a (sequence of) mapping $\phi : E_1, \dots, E_m \rightarrow F$, and a constant $C > 0$ such that $\forall (z_1, \dots, z_m) \in E_1 \times \dots \times E_m$ and $z'_i \in E_i$:

$$\|\phi(z_1, \dots, z_m) - \phi(z_1, \dots, z_{i-1}, z'_i, \dots, z_m)\| \leq C \frac{\prod_{j=1}^m \max(\|z_j\|'_j, \mu_j)}{\max(\|z_i\|'_i, \mu_i)} \|z_1 - z'_i\|_i.$$

Then there exists a constant $\kappa > 0$ such that we have the concentration:

$$\phi(Z_1, \dots, Z_m) \propto \mathcal{E}_q\left(\sigma(\kappa\mu)^{(m-1)}\right) + \mathcal{E}_{\frac{q}{m}}((\kappa\sigma)^m), \quad \text{with } \mu^{(m-1)} = \sum_{i=1}^m \frac{\mu_1 \cdots \mu_m}{\mu_i} \quad (9)$$

(the combination of exponential concentration is described precisely in Lemma 5).

Remark 11. Let us rewrite the concentration inequality (9) to let appear the implicit parameter t . There exists two constants $C, c \leq O(1)$ such that for any 1-Lipschitz mapping $f : F \rightarrow \mathbb{R}$, for any $t > 0$:

$$\mathbb{P}(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t) \leq Ce^{-(t/c\sigma\mu^{(m-1)})^q} + Ce^{-(t/c\sigma^m)^{q/m}}, \quad (10)$$

The two exponential terms $Ce^{-(t/c\sigma\mu^{(m-1)})^q}$ and $Ce^{-(t/c\sigma^m)^{q/m}}$ induce two regimes of concentration each one taking an advantage on the other one depending on the values of t . Indeed:

$$t \leq t_0 \equiv (\mu^{(m-1)})^{\frac{m}{m-1}} \iff e^{-(t/c\sigma\mu^{(m-1)})^q} \geq e^{-(t/c\sigma^m)^{q/m}}.$$

As a consequence, when $t \leq t_0$, the leading term of concentration is $\mathcal{E}_q(\sigma\mu^{(m-1)})$ and the tail of the distribution is therefore controlled by $\mathcal{E}_{\frac{q}{m}}(\sigma^m)$. With a similar result to Proposition 2, we can show that the observable diameter (i.e. the order of the standard deviation of the 1-Lipschitz observations) of $\phi(Z_1, \dots, Z_m)$ is then of order $\sigma\mu^{(m-1)}$ which means that $\mathcal{E}_q(\sigma\mu^{(m-1)})$ is the central term of the concentration inequality. One can infer from (10) a weaker concentration result that might be sufficient and far easier to handle, depending on our needs. For all $f : F \rightarrow \mathbb{R}$, 1-Lipschitz and for all $t > 0$:

$$\mathbb{P}(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t) \leq Ce^{-(t/c\sigma\mu^{(m-1)})^q} + Ce^{-(\mu/c\sigma)^q}, \quad (11)$$

where we supposed for simplicity that $\mu_1 = \dots = \mu_m$ so that $\mu^{(m-1)} = \mu^{m-1}$.

⁷One just needs to assume the concentration $\|Z_i\|'_i \in \mu_i \pm \mathcal{E}_q(\sigma)$ and does not need the global concentration of Z_i for the norm (or seminorm) $\|\cdot\|'_i$.

Remark 12. To give a precise idea of the concentration rates given by Theorem 2, consider the case where $(E_1, \|\cdot\|_1) = \dots = (E_m, \|\cdot\|_m) = (E, \|\cdot\|)$ and for all $i \in [m]$, $\|\mathbb{E}[Z_i]\| = O(\sigma\eta_{\|\cdot\|}^{1/q})$. Then we know from Proposition 8 that $\mathbb{E}[\|Z_i\|] = O(\sigma\eta_{\|\cdot\|}^{1/q})$ and therefore:

$$\mu^{(m-1)} = O\left(\sigma^{m-1}\eta_{\|\cdot\|}^{\frac{m-1}{q}}\right).$$

Following Remark 11, the observable diameter of $\phi(Z)$ would then be $\sigma^m\eta_{\|\cdot\|}^{\frac{m-1}{q}}$ (for a canonical Gaussian vector of $(\mathbb{R}^p, \|\cdot\|)$, $Z \propto \mathcal{E}_2$, the observable diameter of $\phi(Z)$ would be $O(n^{\frac{m-1}{2}})$), and (11) could be rewritten:

$$\mathbb{P}(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t) \leq Ce^{-(t/c\sigma^m)^q/\eta^{m-1}} + Ce^{-c\eta}$$

Example 3. In $(\mathbb{R}^p, \|\cdot\|)$, we can look at the concentration of the product $\odot : x, y \mapsto (x_i y_i)_{1 \leq i \leq p} \in \mathbb{R}^p$ for which $\|x \odot y\| \leq \|x\| \|y\|_\infty$. For any random vectors $Z, W \propto \mathcal{E}_2$, such that $\|\mathbb{E}Z\|_\infty, \|\mathbb{E}W\|_\infty = O(\sqrt{\log p})$, we have thanks to Theorem 2:

$$Z \odot W \propto \mathcal{E}_2(\sqrt{\log p}) + \mathcal{E}_1,$$

since Proposition 8 implies that $\mathbb{E}[\|Z\|_\infty] \leq \|\mathbb{E}Z\|_\infty + O(\sqrt{\log p}) \leq O(\sqrt{\log p})$, and the same holds for $\mathbb{E}[\|W\|_\infty]$. Now, consider a random matrix $X \propto \mathcal{E}_2$ in $\mathcal{M}_{p,n}$ such that $\mathbb{E}[X] = 0$. Since $\forall A, B \in \mathcal{M}_{n,p}$, $\|AB^T\|_F \leq \|A\|_F \|B\|$, we see that the empirical covariance matrix has an observable diameter of order $O(\frac{\sqrt{n+p}}{n})$:

$$\frac{XX^T}{n} \propto \mathcal{E}_2\left(\frac{\sqrt{p+n}}{n}\right) + \mathcal{E}_1\left(\frac{1}{n+p}\right).$$

We give now some useful consequences of Theorem 2. We start with a result very similar to Hanson-Wright concentration inequality we do not give the proof here since we will provide later a more general result in a convex concentration setting in Corollary 6.

Corollary 3. Given a deterministic matrix $A \in \mathcal{M}_p$ and two random vectors $Z, W \in \mathbb{R}^p$ satisfying $Z, W \propto \mathcal{E}_2$ and such that $\|\mathbb{E}[Z]\|, \|\mathbb{E}[W]\| \leq O(\sqrt{\log(p)})$, we have the concentration:

$$Z^T A W \in \mathcal{E}_2\left(\|A\|_F \sqrt{\log p}\right) + \mathcal{E}_1(\|A\|_F) \quad \text{in } (\mathcal{M}_n, \|\cdot\|_F).$$

If we consider three random matrices $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ such that $X, Y, D \propto \mathcal{E}_2$ and $\|\mathbb{E}[D]\|, \|\mathbb{E}[X]\|, \|\mathbb{E}[Y]\| \leq O(1)$ then, Theorem 2 just allows us to set the concentration $XDY^T \propto \mathcal{E}_2(n) + \mathcal{E}_{3/2}$ since we can not get better bound than $\|XDY^T\|_F \leq \|X\| \|D\|_F \|Y\|$. We are going to see in next proposition than we can get a better observable diameter if we project the random matrix XDY^T on a deterministic vector.

Corollary 4. *Given three random matrices $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ diagonal such that $X, Y, D \propto \mathcal{E}_2$, $\|\mathbb{E}[X]\| \leq O(\sqrt{p+n})$ and $\|\mathbb{E}[D]\|, \|\mathbb{E}[Y]\| \leq O(\sqrt{\log n})$ then for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:*

$$XDY^T u \propto \mathcal{E}_2 \left(\sqrt{\log(n)(p+n)} \right) + \mathcal{E}_{2/3} \text{ in } (\mathbb{R}^p, \|\cdot\|),$$

PROOF. The Lipschitz concentration of XDY^T is proven thanks to the inequalities:

$$\|XDY^T v\| \leq \begin{cases} \|X\| \|D\| \|Y^T v\| \\ \|X\| \|D\|_F \|Y^T v\|_\infty. \end{cases}$$

We can bound thanks to the bounds already presented in Example 2 (the spectral norm $\|\cdot\|$ on $\mathcal{D}_n$ is like the infinity norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$):

- $\mu_X \equiv \mathbb{E}[\|X\|] \leq \|\mathbb{E}[X]\| + O(\sqrt{p+n}) \leq O(\sqrt{p+n})$,
- $\mu_D \equiv \mathbb{E}[\|D\|] \leq \|\mathbb{E}[D]\| + O(\sqrt{\log(n)}) \leq O(\sqrt{\log(n)})$,
- $\mu_{Y^T} \equiv \mathbb{E}[\|Y^T\|] \leq \|\mathbb{E}[Y^T]\| + O(\sqrt{\log(n)}) \leq O(\sqrt{\log(n)})$.

The result is then a consequence of Theorem 2 applied on the mapping

$$\begin{aligned} \Phi : \mathcal{M}_{p,n} \times \mathcal{D}_n \times \mathcal{M}_{p,n} &\longrightarrow \mathbb{R}^p \\ (X, D, Y) &\longmapsto XDY^T v, \end{aligned}$$

with the tuple:

$$\mu = \left(O(\sqrt{n+p}), O(\sqrt{\log n}), O(\sqrt{\log n}) \right)$$

satisfying $\mu^{(3-1)} = O(\sqrt{(n+p) \log n} + \log n) \leq O(\sqrt{(n+p) \log n})$.

The next proposition reveals some instability of the class of Lipschitz concentrated vectors and the relevance of the notion of linear concentration. Indeed, a Lipschitz concentration hypothesis does not always lead to results in terms of Lipschitz concentration; in the last example of next proposition, it only entails linear concentration.

Corollary 5. *Given three random matrices $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ and a deterministic diagonal matrix $\tilde{D} \in \mathcal{D}_n$ such that $X, Y \propto \mathcal{E}_2$ in $(\mathcal{D}_n, \|\cdot\|_F)$, $D \in \tilde{D} \pm \mathcal{E}_2$, in $(\mathcal{D}_n, \|\cdot\|_F)$ and $\|\tilde{D}\|, \|\mathbb{E}[X]\|, \|\mathbb{E}[Y]\| \leq O(1)$ we have the linear concentration:*

$$\frac{1}{n} XDY^T \in \frac{1}{n} \mathbb{E} [X \tilde{D} Y^T] \pm \mathcal{E}_1 \left(\frac{\log p}{\sqrt{n}} \right) + \mathcal{E}_{1/2} \left(\frac{1}{\sqrt{n}} \right) \mid e^{-n}$$

PROOF. Let us note $x_1, \dots, x_n$ and $y_1, \dots, y_n$, the columns of, respectively, X and Y , considering a matrix $A \in \mathcal{M}_p$ such that $\|A\|_F \leq 1$, we can then compute:

$$\begin{aligned} & \frac{1}{n} \operatorname{Tr} \left(AXDY^T - \mathbb{E}[AX\tilde{D}Y^T] \right) \\ &= \frac{1}{n} \sum_{i=1}^n D_i y_i^T A x_i - \tilde{D}_i \mathbb{E}[y_i^T A x_i] \\ &= \frac{1}{n} \sum_{i=1}^n (y_i^T A x_i - \mathbb{E}[y_i^T A x_i]) \tilde{D}_i + \mathbb{E}[y_i^T A x_i] (D_i - \tilde{D}_i) \\ & \quad + \frac{1}{n} \sum_{i=1}^n (y_i^T A x_i - \mathbb{E}[y_i^T A x_i]) (D_i - \tilde{D}_i) \end{aligned}$$

Then, recalling from Theorem 2 that $X\tilde{D}Y^T \in \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1(1/n)$ (since $\|\tilde{D}\| \leq O(1)$) and from Corollary 3 that for all $i \in [n]$ $y_i^T A x_i \in \mathbb{E}[y_i^T A x_i] \pm \mathcal{E}_2(\sqrt{\log p}) + \mathcal{E}_1$, our hypotheses then provides the concentrations:

- $\frac{1}{n} \operatorname{Tr} \left(AX\tilde{D}Y^T - \mathbb{E}[AX\tilde{D}Y^T] \right) \in 0 \pm \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_1$,
- $\frac{1}{n} \operatorname{Tr} \left(\mathbb{E}[Y^T A X] (D - \tilde{D}) \right) \in 0 \pm \mathcal{E}_2(1/\sqrt{n})$ (since $\|\frac{1}{n} \mathbb{E}[Y^T A X]\|_F \leq 1$)
- $(y_i^T A x_i - \mathbb{E}[y_i^T A x_i]) (D_i - \tilde{D}_i) \in 0 \pm \mathcal{E}_1(\log p/\sqrt{n}) + \mathcal{E}_{1/2}(1/\sqrt{n})$.⁸

We can then conclude the result thanks to Corollary 1

1.6. Concentration under highly probable event

When a random vector satisfies

$$X \propto \mathcal{E}_q(\sigma) + \mathcal{E}_{q/k}(\sigma'),$$

implicitly, $\sigma' \leq O(\sigma)$ (otherwise one would rather work with the simpler equivalent concentration inequality $X \propto \mathcal{E}_{q/k}(\sigma')$). This is indeed the case in particular for the random vectors whose concentration is expressed thanks to Theorem 2 (when the problem is well posed, we have $\mu^{(m-1)} \geq O(\sigma^m)$). Then, following Remark 11, we know that the leading term of the concentration is the term $\mathcal{E}_q(\sigma)$ that gives an observable diameter of order $O(\sigma)$. We then suggest to work with a weaker but more flexible concentration inequality, that allows to

⁸This product concentration does not follow exactly the setting of Theorem 2 but can be treated similarly thanks to the bound:

$$\begin{aligned} & \mathbb{P} \left(\left| y_i^T A x_i - \mathbb{E}[y_i^T A x_i] \right| |D_i - \mathbb{E}[D_i]| \geq t \right) \\ & \leq \mathbb{P} \left(\left| y_i^T A x_i - \mathbb{E}[y_i^T A x_i] \right| \geq \sqrt{t} \right) + \mathbb{P} \left(|D_i - \mathbb{E}[D_i]| \geq \sqrt{t} \right) \end{aligned}$$

replace the unrellevant but precise term $\mathcal{E}_{q/k}(\sigma')$ by an error term more easy to handle as it is done in inequality (11).

Those flexible concentration inequalities, rely on the introduction of conditioned random vectors. Letting $X : \Omega \rightarrow E$ be a random vector and $\mathcal{A} \subset \Omega$ be a measurable subset of the universe Ω , $\mathcal{A} \in \mathcal{F}$, when $\mathbb{P}(\mathcal{A}) > 0$, the random vector $X | \mathcal{A}$ designates the random vector X conditioned with $\mathcal{A}$ defined as the measurable mapping $(\mathcal{A}, \mathcal{F}_{\mathcal{A}}, \mathbb{P}/\mathbb{P}(\mathcal{A})) \rightarrow (E, \|\cdot\|)$ satisfying: $\forall \omega \in \mathcal{A}$, $(X | \mathcal{A})(\omega) = X(\omega)$. When there is no ambiguity, we will allow ourselves to designate abusively with the same notation “ $\mathcal{A}$ ”, a subset of Ω and the subset $X(\mathcal{A})$ of E .

Definition 6 (Concentration under highly probable event). *Given two (sequences of) positive numbers θ and σ , a (sequence of) random vector $Z \in (E, \|\cdot\|)$ and a (sequence of) event $\mathcal{A}_Z$, we say that Z is Lipschitz q -exponentially concentrated on the concentration zone $\mathcal{A}_Z$ over θ with an observable diameter of size σ , and we write*

$$Z \overset{\mathcal{A}_Z}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta}, \quad (12)$$

if and only if there exist some constants $C, c > 0$ such that:

$$(Z | \mathcal{A}_Z) \propto \mathcal{E}_q(\sigma) \quad \text{and} \quad \mathbb{P}(\mathcal{A}_Z^c) \leq Ce^{-\theta/c}.$$

Be careful that the expectation of $f(Z)$ is not always defined if Z is not exponentially concentrated, but if (12) is satisfied, we can still consider:

$$\mathbb{E}_{\mathcal{A}_Z}[f(Z)] \equiv \mathbb{E}[f(Z) | \mathcal{A}_Z] = \frac{\mathbb{E}[\mathbb{1}_{\mathcal{A}_Z} f(Z)]}{\mathbb{P}(\mathcal{A}_Z)}. \quad (13)$$

We will sometimes omit to precise the concentration zone $\mathcal{A}_Z$ when it is not necessary. The notation $Z \propto \mathcal{E}_q(\sigma) \mid e^{-\theta}$ implicitly requires the existence of a concentration zone $\mathcal{A}_Z$ satisfying the upper inequalities. We also adapt this notation to the case of linear concentration and we then denote $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma) \mid e^{-\theta}$, (for $\tilde{Z} \in E$, a deterministic equivalent of Z).

The concentration under highly probable event is of course a *weaker* notion than the notion of concentration presented in Definition 1:

Lemma 6. *Given three (sequences of) positive number $\sigma, \theta, \mu > 0$, a (sequence of) random vector $Z \in (E, \|\cdot\|)$ and two (sequences of) events $\mathcal{A}_1$ and $\mathcal{A}_2$, if there exists two constants $C, c > 0$ such that $\mathbb{P}(\mathcal{A}_2^c) \leq Ce^{-c/\mu}$, then:*

$$Z \overset{\mathcal{A}_1}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta} \quad \implies \quad Z \overset{\mathcal{A}_1 \cap \mathcal{A}_2}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\min(\theta, \mu)}.$$

In particular, $Z \propto \mathcal{E}_q(\sigma)$ can be rewritten $Z \overset{\Omega}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta}$ for any (sequence of) positive number $\theta > 0$, therefore Definition 1 concerns a subclass of the random vectors described by Definition 6.

PROOF. It is immediate, since $\mathbb{P}(\mathcal{A}) \geq K$, for any independent copy of Z, Z' and any $f : E \rightarrow \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}(|f(Z) - f(Z')| \geq t \mid \mathcal{A}) \leq \frac{1}{\mathbb{P}(\mathcal{A})} \mathbb{P}(|f(Z) - f(Z')| \geq t) \leq \frac{C}{K} e^{-(t/c\sigma)^q},$$

for some constants $C, c > 0$.

Remark 13. Since the term $e^{-\theta/c}$ does not vanish as $t \rightarrow \infty$, this system of concentration inequalities is only interesting in the regime where $t = O(\sigma/\theta^{1/q})$. Note that when (12) is verified, then for any λ -Lipschitz mapping $\phi : E \rightarrow E$, $\phi(Z) \propto \mathcal{E}_q(\lambda\sigma) \mid e^{-\theta}$ (the term $e^{-\theta}$ remains unchanged).

If we disregard the tail behavior, the expression of the concentration of a product of random vectors is obtained straightforwardly.

Proposition 9. Under the hypotheses of Theorem 2, without supposing that m is a constant but only that $\log(m) \leq o(\min_{1 \leq i \leq m} (\frac{\mu_i^q}{\sigma^q}, \theta))$, for all $i \in [m]$, if we assume (unlike in Theorem 2) that:

$$Z = (Z_1, \dots, Z_m) \propto \mathcal{E}_q(\sigma)^9$$

then for some $\mu_i \leq O(\mathbb{E}[\|Z_i\|'_i, \mathcal{A}_Z])$, then we have the concentration:

$$\phi(Z) \propto \mathcal{E}_q\left(\sigma \mu^{(m-1)}\right) \mid e^{-\min_{1 \leq i \leq m} \mu_i^q / \sigma^q}.$$

Recall from Proposition 8 that when $(E_1, \|\cdot\|'_1) = \dots = (E_m, \|\cdot\|'_m) = (E, \|\cdot\|')$ and $\|\cdot\|'$ admits a norm degree η' and if $Z \propto \mathcal{E}_q(\sigma)$, $\|\mathbb{E}[Z_i]\|' = O(\sigma \eta'^{1/q})$, then $\mu_i = O(\sigma \eta'^{1/q})$ and we can express the concentration:

$$\phi(Z) \propto \mathcal{E}_q\left((c\sigma)^m \eta'^{\frac{m-1}{q}}\right) \mid e^{-\eta'},$$

for some constant $c > 0$.

We know from Remark 11 that this proposition is a consequence of Theorem 2 when $Z = (Z_1, \dots, Z_m) \propto \mathcal{E}_q(\sigma)$, but we still give the proof to show how to handle the notion of concentration under highly probable events.

PROOF. Let us introduce the event:

$$\mathcal{B} = \{\forall i \in [m] : \|Z_i\|'_i \leq 2\mu_i\}.$$

⁹We can also take as hypothesis that $Z \overset{\mathcal{A}_Z}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta}$ for some $\theta > 0$, then the result is morally the same, one just has to replace the term $e^{-\min_{1 \leq i \leq m} \mu_i^q / \sigma^q}$ by $e^{-\min_{1 \leq i \leq m} (\frac{\mu_i^q}{\sigma^q}, \theta)}$ in the expression of the concentration of $\phi(Z)$

We know from the concentration $\|Z_i\|'_i \in \mu_i \pm \mathcal{E}_q(\sigma) \mid e^{-\theta}$ (for $i \in [m]$) that there exist two constants $C, c > 0$ such that:

$$\mathbb{P}(\mathcal{B}^c) \leq \sum_{i=1}^m \mathbb{P}(\|Z_i\|'_i \leq 2\mu_i) \leq \sum_{i=1}^m C e^{-\frac{c\mu_i^q}{\sigma^q}} \leq C e^{\log(m) - c \min_{1 \leq i \leq m} \frac{\mu_i^q}{\sigma^q}} \leq C e^{-c' \min_{1 \leq i \leq m} \frac{\mu_i^q}{\sigma^q}},$$

for some constant $c' > 0$. Then, on $\mathcal{B}$ (recall from the discussion preceding Definition 6 that we identify the event $\mathcal{B}$ with the subset $Z(\mathcal{B}) \subset E_1 \times \dots \times E_m$), $\phi : (E, \|\cdot\|_{\ell^\infty}) \rightarrow (F, \|\cdot\|)$ is $2^{m-1}\mu^{(m-1)}$ -Lipschitz, and since $\mathbb{P}((\mathcal{B})^2) \geq O(1)$, one can invoke Remark 3 to set that:

$$\mathbb{P}(|\phi|_{\mathcal{B}}(Z) - \phi|_{\mathcal{B}}(Z')| \geq t \mid (Z, Z') \in \mathcal{B}^2) \leq C e^{-(ct/\sigma(2\mu)^{(m-1)})^q},$$

for some constant $C, c > 0$, which concludes the proof.

Remark 14. We principally displayed Proposition 9 to give the reader an easy way to apprehend the expression of the concentration of a product of random vectors. However keep in mind that Theorem 2 is a far stronger result, obtained thanks to the optimization of a trade-off parameter. It is not so obvious when $m = 2$ (Theorem 2 just has the advantage to precisely describe the tail of the concentration) but when $m = 3$, it is quite clear. For instance, we can deduce from Corollary 4 (which is a direct consequence of Theorem 2) that (under the hypotheses of Corollary 4):

$$XDY^T u \propto \mathcal{E}_2(\sqrt{\log(n)(p+n)}) \mid e^{-\log(n)(p+n)} \quad (14)$$

(see Remark 11 for more details on this inference). However, with Proposition 9, one could only have obtained the concentration:

$$XDY^T u \propto \mathcal{E}_2(\sqrt{\log(n)(p+n)}) \mid e^{-\log(n)}$$

which is far less interesting because if $c > 0$ is too small, $(C e^{-c \log(n)})_{n \in \mathbb{N}}$ converges relatively slowly to 0.

Proposition 9 is though useful to control the concentration of high order products, because it does not let appear in the observable diameter, as in Theorem 2, a term κ^k that is hard to control. For instance we provide in the following an improvement of Corollary 4, allowing us to control $(\frac{1}{n}XDY^T)^k u$, as long as $\|\frac{1}{n}XDY^T\|$ is lower than one with high probability.

Proposition 10. Given three (sequence of) random matrices $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ such that $X, Y, D \propto \mathcal{E}_2$, $\|\mathbb{E}[X]\| \leq O(\sqrt{p+n})$, $\|\mathbb{E}[D]\|, \|\mathbb{E}[Y]\| \leq O(\sqrt{\log n})$ we assume that there exists a positive constant $\varepsilon > 0$ such that:

$$\mathbb{P}\left(\frac{1}{n}\|XDY^T\| \geq 1 - \varepsilon\right) \leq C e^{-cn}$$

Then for any $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$ and a (sequence of) integers $k \in \mathbb{N}$:

$$\left(\frac{1}{n+p}XDY^T\right)^k u \propto \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n}.$$

PROOF. Let us bound as in the proof of Corollary 4:

$$\begin{cases} \left\| \left(\frac{1}{n+p} X D Y^T \right)^k u \right\| \leq \left\| \left(\frac{1}{n+p} X D Y^T \right)^{k-1} \frac{X}{\sqrt{p+n}} \right\| \|D\|_F \|Xu\|_\infty \\ \left\| \left(\frac{1}{n+p} X D Y^T \right)^k u \right\| \leq \left\| \left(\frac{1}{n+p} X D Y^T \right)^{k-1} \frac{X}{\sqrt{p+n}} \right\| \|D\| \|Xu\| \end{cases}$$

If we note $\mathcal{A} \equiv \{ \frac{1}{n} \|X D Y^T\| \geq 1 - \varepsilon \}$, we know from Proposition 9 that:

$$\begin{cases} \left(\frac{1}{n+p} X D Y^T \right)^{k-1} \frac{X}{\sqrt{p+n}} \propto \mathcal{E}_2 \mid e^{-n} \\ \mathbb{E}_{\mathcal{A}} \left[\left\| \left(\frac{1}{n+p} X D Y^T \right)^{k-1} \frac{X}{\sqrt{p+n}} \right\| \right] \leq O((1 - \varepsilon)^k), \end{cases}$$

besides, $\mathbb{E}[\|D\|], \mathbb{E}[\|Xu\|_\infty] \leq O(\log(p+n))$, thus we can apply a second time Proposition 9 to obtain our result.

1.7. The class of convexly concentrated vectors and its degeneracy

Between Lipschitz concentration and linear concentration lies the convex concentration which arises from a theorem of Talagrand (Talagrand, 1995), extending the list of examples of concentrated vectors. A random vector is said to be convexly concentrated if its Lipschitz and *quasi-convex* observations are concentrated.

Definition 7. Given a normed vector space $(E, \|\cdot\|)$, an application $f : E \mapsto \mathbb{R}$ is said to be *quasi-convex* iif for any $t \in \mathbb{R}$, the set $\{f \leq t\} \equiv \{x \in E \mid f(x) \leq t\}$ is convex.

Definition 8. Given a sequence of normed vector spaces $(E_p, \|\cdot\|_p)_{p \geq 0}$, a sequence of random vectors $(Z_p)_{p \geq 0} \in \prod_{p \geq 0} E_p$, a sequence of positive reals $(\sigma_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say that $Z = (Z_p)_{p \geq 1}$ is *convexly q-exponentially concentrated* with an observable diameter of order $O(\sigma_p)$ iiff there exist two positive constants $C, c > 0$ such that $\forall p \in \mathbb{N}$ and for any 1-Lipschitz and quasi-convex function $f : E_p \rightarrow \mathbb{R}$ (for the norms $\|\cdot\|_p$), $\forall t > 0$,

$$\mathbb{P}(|f(Z_p) - \mathbb{E}[f(Z_p)]| \geq t) \leq C e^{(t/c\sigma_p)^q},$$

We write in that case $Z_p \propto_c \mathcal{E}_q(\sigma_p)$ (or more simply $Z \propto_c \mathcal{E}_q(\sigma)$).¹⁰

The relevance of this definition is given by the next theorem. It is a combinatorial result which provides concentration inequalities for “discontinuous” distributions (we can have atoms for instance: none of the previous theorems in Subsection 1.1 allowed us to handle this scenario).

¹⁰One could have replaced in the inequality $\mathbb{E}[f(Z_p)]$ by $f(Z'_p)$ (with Z'_p , an independent copy of Z_p) or by m_f (with m_f a median of $f(Z_p)$) as in Definition 1. All those three assertions are equivalent.

Theorem 3 (Talagrand (1995)). *A random vector $Z \in [0, 1]^n$ with independent entries satisfies $Z \propto_c \mathcal{E}_2$.*

The class of convexly concentrated random vectors is far less stable than the class of Lipschitz concentrated vectors: the only simple transformations that preserve concentration are the affine transformations (as for the class of linearly concentrated vectors).

Proposition 11. *Given two normed vector spaces $(E, \|\cdot\|)$ and $(F, \|\cdot\|')$, a random vector $Z \in E$ and an affine mapping $\phi \in \mathcal{A}(E, F)$ such that $\|\mathcal{L}(\phi)\| \leq \lambda$:*

$$Z \propto_c \mathcal{E}_q(\sigma) \quad \implies \quad \phi(Z) \propto_c \mathcal{E}_q(\lambda\sigma).$$

As for Lipschitz concentration in Remark 3 (and for linear concentration – but this results was not displayed as it is not used in our study), a lemma allows us to treat the concentration of observations $f(Z)$ when f is Lipschitz and quasi-convex only on a convex subset A and $\{Z \in A\}$ has positive probability non decreasing with the dimension.

Lemma 7. *Given a (sequence of) positive numbers $\sigma > 0$, a (sequence of) random vector $Z \in E$ satisfying $Z \propto \mathcal{E}_q(\sigma)$, and a (sequence of) convex subsets $A \subset E$, if there exists a constant $K > 0$ such that $\mathbb{P}(Z \in A) \geq K$ then there exist two constants $C, s > 0$ such that for any (sequence of) 1-Lipschitz and quasi-convex mappings $f : A \rightarrow \mathbb{R}$:*

$$\forall t > 0 : \mathbb{P}(|f(Z) - \mathbb{E}[f(Z)]| \geq t \mid Z \in A) \leq Ce^{-(t/c\sigma)^q}.$$

and similar concentration occur around any median of $f(Z)$ or any independent copy of Z (under $\mathcal{A}$).

PROOF. The proof is the same as the one provided in Remark 3 except that this time, one needs the additional argument that since $S = \{f \leq m_f\}$ (for m_f , a median of f) is convex, the mappings $z \mapsto d(z, S)$ and $z \mapsto -d(z, S)$ are both quasi-convex thanks to the triangular inequality.

Given two convexly concentrated random vectors $X, Y \in E$ satisfying $X, Y \propto_c \mathcal{E}_q(\sigma)$, the convex concentration of the couple $(X, Y) \propto_c \mathcal{E}_q(\sigma)$ is ensured if:

1. X and Y are independent
2. $(X, Y) = u(Z)$ with u affine and $Z \propto_c \mathcal{E}_q(\sigma)$.

We can then in particular state the concentration of $X + Y$ as it is a linear transformation of (X, Y) . For the product it is not as simple as for the Lipschitz concentration: we will consider the special cases $E = \mathbb{R}^p$, $E = \mathcal{M}_{p,n}$ to retrieve interesting properties. All the coming study is based on a preliminary elementary result that does not need any proof.

Lemma 8. *Given a convex mapping $f : \mathbb{R} \rightarrow \mathbb{R}$, and a vector $a \in \mathbb{R}_+^p$, the mapping $F : \mathbb{R}^p \ni (z_1, \dots, z_p) \mapsto \sum_{i=1}^p a_i f(z_i) \in \mathbb{R}$ is convex (so in particular quasi-convex).*

We need a supplementary result to efficiently manage the concentration rate when multiplying a large number of random vectors.

Lemma 9. *Given m non commutative variables $a_1, \dots, a_m$ of a given algebra, we have the identity:*

$$(-1)^m \sum_{\sigma \in \mathfrak{S}_m} a_{\sigma(1)} \cdots a_{\sigma(m)} = \sum_{I \subset [m]} (-1)^{|I|} \left(\sum_{i \in I} a_i \right)^m,$$

where $|I|$ is the cardinality of I .

PROOF. The idea is to inverse the identity:

$$(a_1 + \cdots + a_m)^m = \sum_{J \subset I} \sum_{\{i_1, \dots, i_m\} = J} a_{i_1} \cdots a_{i_m},$$

thanks to the Rota formula (see Rolland (2006)) that sets for any mappings f, g defined on the set subsets of $\mathbb{N}$ and having values in a commutative group (for the sum):

$$\forall I \subset \mathbb{N}, f(I) = \sum_{J \subset I} g(J) \iff \forall I \subset \mathbb{N}, g(I) = \sum_{J \subset I} \mu_{\mathcal{P}(\mathbb{N})}(J, I) f(J).$$

where $\mu_{\mathcal{P}(\mathbb{N})}(J, I) = (-1)^{|I \setminus J|}$ is an analog of the Moëbus function for the order relation induced by the inclusions in $\mathcal{P}(\mathbb{N})$. In our case, for any $J \subset [m]$, if we set:

$$f(J) = \left(\sum_{i \in J} a_i \right)^m \quad \text{and} \quad g(J) = \sum_{\{i_1, \dots, i_m\} = J} a_{i_1} \cdots a_{i_m},$$

we see that for any $I \subset [m]$, $f(I) = \sum_{J \subset I} g(J)$, therefore taking the Rota formula in the case $I = [m]$ and multiplying on both sides by $(-1)^m$, we obtain the result of the Lemma (in that case, $\mu_{\mathcal{P}(\mathbb{N})}(J, I) = (-1)^{m-|J|}$ and $\sum_{\{i_1, \dots, i_m\} = I} a_{i_1} \cdots a_{i_m} = \sum_{\sigma \in \mathfrak{S}_m} a_{\sigma(1)} \cdots a_{\sigma(m)}$).

Proposition 12. *Given a (sequence of) integer $m \in \mathbb{N}$ and a (sequence of) positive number $\sigma > 0$ such that $m \leq O(p)$, a (sequence of) m random vectors $X_1, \dots, X_m \in \mathbb{R}^p$ such that $\sup_{1 \leq i \leq m} \|\mathbb{E}[X_i]\|_\infty = O((\log p)^{1/q})$, if we suppose that*

$$X \equiv (X_1, \dots, X_m) \propto_c \mathcal{E}_q(\sigma) \text{ in } ((\mathbb{R}^p)^m, \|\cdot\|_{\ell^\infty}),$$

with, for any $z = (z_1, \dots, z_m) \in (\mathbb{R}^p)^m$, $\|z\|_{\ell^\infty} = \sup_{1 \leq i \leq m} \|z_i\|$, then there exists a constant $\kappa \leq O(1)$ such that:

$$X_1 \odot \cdots \odot X_m \in \mathcal{E}_q \left((\kappa\sigma)^m (\log(p))^{(m-1)/q} \right) + \mathcal{E}_{q/m}((\kappa\sigma)^m) \quad \text{in } (\mathbb{R}^p, \|\cdot\|).$$

To simplify the preceding inequality, we add that if there exists a (sequence of) positive numbers $\kappa > 0$ and two constants $C, c > 0$ such that, noting $\mathcal{A}_\kappa \equiv \{\forall i \in [m] : \|X_i\|_\infty \leq \kappa\}$, $\mathbb{P}(\mathcal{A}_\kappa^c) \leq Ce^{-c\rho}$, then:

$$X_1 \odot \cdots \odot X_m \in \mathcal{E}_q((2e\kappa)^{m-1}\sigma) \mid e^{-\rho} \quad \text{in } (\mathbb{R}^p, \|\cdot\|).$$

PROOF. Let us first assume that all the X_i are equal to a vector $Z \in \mathbb{R}^p$. Considering $a = (a_1, \dots, a_p) \in \mathbb{R}^p$, we want to show the concentration of $a^T Z^{\odot m} = \sum_{i=1}^p a_i z_i^m$ where $z_1, \dots, z_p$ are the entries of Z .

The mapping $p_m : x \mapsto x^m$ is not quasi-convex when m is odd, therefore, in that case we decompose it into the difference of two convex mappings $p_m(z) = p_m^+(z) - p_m^-(z)$ where:

$$p_m^+ : z \mapsto \max(z^m, 0) \quad \text{and} \quad p_m^- : z \mapsto -\min(z^m, 0), \quad (15)$$

(say that, if m is even, then we set $p_m^+ = p_m$ and $p_m^- : z \mapsto 0$). For the same reasons, we decompose $\phi_a^+ : z \mapsto a^T p_m^+(z)$ and $\phi_a^- : z \mapsto a^T p_m^-(z)$ into:

$$\phi_a^+ = \phi_{|a|}^+ - \phi_{|a|-a}^+ \quad \text{and} \quad \phi_a^- = \phi_{|a|}^- - \phi_{|a|-a}^-$$

(for $|a| = (|a_i|)_{1 \leq i \leq p}$), so that:

$$a^T Z^{\odot m} = \phi_{|a|}^+(Z) - \phi_{|a|-a}^+(Z) - \phi_{|a|}^-(Z) + \phi_{|a|-a}^-(Z)$$

becomes a combination of quasi-convex functionals of Z . We now need to measure their Lipschitz parameter. Let us bound for any $z \in \mathbb{R}^p$:

$$\left| \phi_{|a|}^+(z) \right| = \sum_{i=1}^n |a_i| |z_i|^m \leq \|a\| \|z\| \|z\|_\infty^{m-1},$$

and the same holds for $\phi_{|a|-a}^+$, $\phi_{|a|}^-$ and $\phi_{|a|-a}^-$. Therefore, we can conclude a similar result to Theorem 2 with $\mu = \mathbb{E}[\|Z\|_\infty] \leq O(\sigma(\log p)^{1/q})$. To show the second concentration result (still when $X_1 = \cdots = X_m$), we just have to note that under $\{\|Z\|_\infty \leq \kappa\}$, $\phi_{|a|}^+$, $\phi_{|a|-a}^+$, $\phi_{|a|}^-$ and $\phi_{|a|-a}^-$ are all $\|a\|\kappa^{m-1}$ -Lipschitz.

Now staying in the second setting ($\mathbb{P}(\mathcal{A}_\kappa^c) \leq Ce^{-c\rho}$), if we assume that the $X_1, \dots, X_m$ are different, we employ Lemma 9 in this commutative case to write ($|\mathfrak{S}_m| = m!$):

$$(X_1 \odot \cdots \odot X_m) = \frac{(-1)^m}{m!} \sum_{I \subset [m]} (-1)^{|I|} \left(\sum_{i \in I} X_i \right)^{\odot m}. \quad (16)$$

Therefore, the sum $(\mathbb{R}^p)^I \ni z_1, \dots, z_{|I|} \mapsto \sum_{i \in I} z_i \in \mathbb{R}^p$ being m -Lipschitz for the norm $\|\cdot\|_\infty$, we know that $\forall I \subset [m]$, $\sum_{i \in I} X_i \propto_c \mathcal{E}_q(m\sigma) \mid e^{-\rho}$, and thus $(\sum_{i \in I} X_i)^{\odot m} \in \mathcal{E}_q(m^m \kappa^{m-1} \sigma) \mid e^{-\rho}$. We can then exploit Proposition 5 to obtain

$$\left(\left(\sum_{i \in I} X_i \right)^{\odot m} \right)_{I \subset [m]} \in \mathcal{E}_q(m^m \kappa^{m-1} \sigma) \mid e^{-\rho} \quad \text{in } \left((\mathbb{R}^p)^{2^m}, \|\cdot\|_{\ell^\infty} \right),$$

where we recall that 2^m is the number of subsets of $[m]$. Thus summing the 2^m concentration inequalities, we can conclude from Equation (16) and the Stirling formula $\frac{m^m}{m!} = \frac{e^m}{\sqrt{2\pi m}} + O(1)$ that:

$$(X_1 \odot \cdots \odot X_m) \in \mathcal{E}_q((2e\kappa)^{m-1}\sigma) \mid e^{-p}.$$

The demonstration follows exactly the same steps when we do not place ourselves on $\mathcal{A}_\kappa$ (it is just a bit longer to write).

Under convex concentration hypothesis we also have a Hanson-Wright-like concentration inequality as in Corollary 3.

Corollary 6. *Given a deterministic matrix $A \in \mathcal{M}_p$ and two random vectors $Z, W \in \mathbb{R}^p$ satisfying $Z, W \propto_c \mathcal{E}_2$ and such that $\|\mathbb{E}[Z]\|, \|\mathbb{E}[W]\| \leq O(\sqrt{\log(p)})$, we have the concentration:*

$$Z^T A W \in \mathcal{E}_2\left(\|A\|_F \sqrt{\log p}\right) + \mathcal{E}_1(\|A\|_F) \quad \text{in } (\mathcal{M}_n, \|\cdot\|_F).$$

PROOF. We are going to employ the singular decomposition:

$$A = P_A^T \Lambda Q_A$$

with $\Lambda = \text{Diag}(\lambda) \in \mathcal{D}_p$, $P_A, Q_A \in \mathcal{O}_p$. Noting $\check{Z} \equiv P_A X$ and $\check{W} \equiv Q_A Y$, since $M \rightarrow P_A M$ and $M \rightarrow Q_A M$ are both 1-Lipschitz linear transformations on $\mathbb{R}^p$, we see from Proposition 11 that $\check{Z}, \check{W} \propto_c \mathcal{E}_2$. Besides we know from Example 2 that:

$$\mathbb{E}[\|\check{Z}\|_\infty] \leq \|\mathbb{E}[\check{Z}]\|_\infty + O(\sqrt{\log p}) \leq \|\mathbb{E}[Z]\| + O(\sqrt{\log p}) \leq O(\sqrt{\log p}),$$

and the same way $\mathbb{E}[\|\check{W}\|_\infty] \leq O(\sqrt{\log p})$. Now, since for all vector $z, w \in \mathbb{R}^p$, $\|z \odot w\| = \|w \odot z\| \leq \|z\| \|w\|_\infty$, Proposition 12 implies the concentration $\check{Z} \odot \check{W} \in \mathcal{E}_2(\sqrt{\log(pn)}) + \mathcal{E}_1$ which allows us to conclude that:

$$Z^T A W = (\check{Z} \odot \check{W})^T \lambda \in \mathcal{E}_2\left(\|\lambda\| \sqrt{\log p}\right) + \mathcal{E}_1(\|\lambda\|),$$

that gives us the result of the corollary since $\|\lambda\| = \|A\|_F$.

Corollary 5 that gives the linear concentration of random matrices $\frac{1}{n} X D Y$ for X, Y, D Lipschitz concentrated and D diagonal can not be set with good convergence speed when X, D, Y are simply convexly concentrated. We can still obtain a precise estimation of $\frac{1}{n} \mathbb{E}[X D Y]$ that will be utterly important for the design of a computable deterministic equivalent of the resolvent $Q = (I_n - \frac{1}{n} X X^T)^{-1}$.

Corollary 7. *Given three random matrices $X, Y \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ such that $X, Y \propto_c \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$, $D \in \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{D}_n, \|\cdot\|)^{11}$ and*

¹¹Note than in Corollary 5, to set the concentration of $\frac{1}{n} X D Y^T$, we had to assume that $D \in \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{D}_n, \|\cdot\|_F)$ (and not in $(\mathcal{D}_n, \|\cdot\|)$)

$\|\mathbb{E}[D]\|, \|\mathbb{E}[X]\|, \|\mathbb{E}[Y]\| \leq O(1)$, we have the bound:

$$\left\| \mathbb{E} \left[\frac{1}{n} X D Y^T \right] - \mathbb{E} \left[\frac{1}{n} X \mathbb{E}[D] Y^T \right] \right\|_F \leq O \left(\sqrt{\frac{\log p}{n}} \right).$$

If in addition, $\|\mathbb{E}[D] - \tilde{D}\|_F \leq O(1/\sqrt{n})$ ¹², then:

$$\left\| \mathbb{E} \left[\frac{1}{n} X D Y^T \right] - \mathbb{E} \left[\frac{1}{n} X \tilde{D} Y^T \right] \right\|_F \leq O \left(\sqrt{\frac{\log p}{n}} \right).$$

PROOF. As in the proof of Corollary 4 note $x_1, \dots, x_n$ and $y_1, \dots, y_n$, the columns of, respectively, X and Y , considering a matrix $A \in \mathcal{M}_p$ such that $\|A\|_F \leq 1$, we can bound:

$$\left| \mathbb{E} \left[\frac{1}{n} \text{Tr} (A X D Y^T - A X \mathbb{E}[D] Y^T) \right] \right| \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [|y_i^T A x_i - \mathbb{E}[y_i^T A x_i]| |D_i - \mathbb{E}[D_i]|]$$

Corollary 6 implies that for all $i \in [n]$ $y_i^T A x_i \in \mathbb{E}[y_i^T A x_i] \pm \mathcal{E}_2(\sqrt{\log p}) + \mathcal{E}_1$, and we can bound thanks to Hölder inequality:

$$\mathbb{E} [|y_i^T A x_i - \mathbb{E}[y_i^T A x_i]| |D_i - \mathbb{E}[D_i]|] \leq O \left(\sqrt{\frac{\log p}{n}} \right).$$

which allows us to set the first result of the corollary.

For the second result, let us bound:

$$\begin{aligned} & \left| \mathbb{E} \left[\frac{1}{n} \text{Tr} (A X D Y^T - A X \tilde{D} Y^T) \right] \right| \\ & \leq \left| \mathbb{E} \left[\frac{1}{n} \text{Tr} (A X D Y^T - A X \mathbb{E}[D] Y^T) \right] \right| + \left| \frac{1}{n} \text{Tr} (\mathbb{E}[Y^T A X] (\mathbb{E}[D] - \tilde{D})) \right|, \end{aligned}$$

we can then conclude since $\|\frac{1}{n} \mathbb{E}[Y^T A X]\|_F \leq O(1)$ and $\|\mathbb{E}[D] - \tilde{D}\|_F \leq O(1/\sqrt{n})$.

A similar result to Proposition 12 holds for matrix product but for that one needs first to introduce a new notion of concentration, namely the *transversal convex concentration*. Let us give some definitions.

Definition 9. Given a sequence of normed vector spaces $(E_p, \|\cdot\|_p)_{p \geq 0}$, a sequence of groups $(G_p)_{p \geq 0}$, each G_p (for $p \in \mathbb{N}$) acting on E_p , a sequence of random vectors $(Z_p)_{p \geq 0} \in \prod_{p \geq 0} E_p$, a sequence of positive reals $(\sigma_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$ and a parameter $q > 0$, we say that $Z = (Z_p)_{p \geq 0}$ is *convexly q -exponentially concentrated transversally* to the action of G with an observable diameter of order σ and we note $Z \propto_G^T \mathcal{E}_q(\sigma)$ iff there exists two constants $C, c \leq O(1)$ such that $\forall p \in \mathbb{N}$ and for any 1-Lipschitz, quasi-convex and G -invariant¹³ function $f : E_p \rightarrow \mathbb{R}$, $\forall t > 0$:

¹²This in particular the case if $D \in \tilde{D} \pm \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{D}_n, \|\cdot\|_F)$ (see Lemma 2)

¹³For any $g \in G$ and $x \in E$, $f(x) = f(g \cdot x)$

$$\mathbb{P}(|f(Z_p) - \mathbb{E}[f(Z_p)]| \geq t) \leq Ce^{(t/c\sigma_p)^q}{}^{14}.$$

Remark 15. Given a normed vector space $(E, \|\cdot\|)$, a group G acting on E and a random vector $Z \in E$, we have the implication chain:

$$Z \propto \mathcal{E}_q(\sigma) \quad \Rightarrow \quad Z \propto_c \mathcal{E}_q(\sigma) \quad \Rightarrow \quad Z \propto_G^T \mathcal{E}_q(\sigma).$$

Considering the actions:

- $\mathfrak{S}_p$ on $\mathbb{R}^p$ where for $\sigma \in \mathfrak{S}_p$ and $x \in \mathbb{R}^p$, $\sigma \cdot x = (x_{\sigma(i)})_{1 \leq i \leq p}$,
- $\mathcal{O}_{p,n} \equiv \mathcal{O}_p \times \mathcal{O}_n$ on $\mathcal{M}_{p,n}$ where for $(P, Q) \in \mathcal{O}_{p,n}$ and $M \in \mathcal{M}_{p,n}$, $(P, Q) \cdot M = PMQ$,

the convex concentration in $\mathcal{M}_{p,n}$ transversally to $\mathcal{O}_{p,n}$ can be expressed as a concentration on $\mathbb{R}^p$ transversally to $\mathfrak{S}_p$ thanks to the introduction the mapping σ providing to any matrix the ordered sequence of its singular values :

$$\begin{aligned} \sigma : \mathcal{M}_{p,n} &\rightarrow \mathbb{R}_+^d && \text{with } d = \min(p, n) \\ M &\mapsto (\sigma_1(M), \dots, \sigma_d(M)). \end{aligned}$$

(there exists $(P, Q) \in \mathcal{O}_{p,n}$ such that $M = P\Sigma(M)Q$, where $\Sigma \in \mathcal{M}_{p,n}$ has $\sigma_1(M) \leq \dots \leq \sigma_d(M)$ on the diagonal).

Theorem 4 (Louart and Couillet (2019)). Given a random matrix $Z \in \mathcal{M}_{p,n}$:

$$Z \propto_{\mathcal{O}_{p,n}}^T \mathcal{E}_q(\sigma) \quad \Longleftrightarrow \quad \sigma(Z) \propto_{\mathfrak{S}_d}^T \mathcal{E}_q(\sigma),$$

(where the concentrations inequalities are implicitly expressed for euclidean norms: $\|\cdot\|_F$ on $\mathcal{M}_{p,n}$ and $\|\cdot\|$ on $\mathbb{R}^d$).

We can now set the result of concentration of a product of matrices convexly concentrated.

Proposition 13. Given three sequences $m \in \mathbb{N}^{\mathbb{N}}$ and $\sigma, \kappa \in \mathbb{R}_+^{\mathbb{N}}$, a random matrix $X \in \mathcal{M}_p$, if we suppose that

$$X \propto_c \mathcal{E}_q \text{ in } (\mathcal{M}_p, \|\cdot\|_F),{}^{15}$$

and that there exists two constants $C, c > 0$ such that $\mathbb{P}(\|X\| \geq \kappa) \leq Ce^{-c\kappa}$, then:

$$X^m \in \mathcal{E}_q \left(\sqrt{p} (2e\kappa)^{m-1} \sigma \right) \mid e^{-p} \quad \text{in } (\mathcal{M}_p, \|\cdot\|_*),$$

where $\|\cdot\|_*$ is the nuclear norm satisfying for any $M \in \mathcal{M}_{p,n}$ $\|M\|_* = \text{Tr}(\sqrt{MM^T})$ (it is the dual norm of the spectral norm). Now with the same

¹⁴Once again, we point out that one could have replaced here $\mathbb{E}[f(Z_p)]$ by $f(Z'_p)$ of m_f .

¹⁵nothing is changed if we rather assume $X \propto_c \mathcal{E}_q \mid e^{-p}$.

hypotheses, if we consider this time that the random matrix X belongs to $\mathcal{M}_{p,n}$ and satisfies $\|\mathbb{E}[X]\| = O((p+n)^{1/q})$, we have the concentrations

$$\begin{aligned} (XX^T)^m &\in \mathcal{E}_q \left(\sqrt{p+n} (2e\kappa)^{2m-1} \sigma \right) \mid e^{-(p+n)} && \text{in } (\mathcal{M}_p, \|\cdot\|_*), \\ (XX^T)^m X &\in \mathcal{E}_q \left(\sqrt{p+n} (2e\kappa)^{2m} \sigma \right) \mid e^{-(p+n)} && \text{in } (\mathcal{M}_{p,n}, \|\cdot\|_*). \end{aligned}$$

Remark 16. We could have given a result of concentration concerning the product of different matrices $X_1, \dots, X_m$ but the expression would have been complicated since our proof relies on Lemma 9, and without a strong hypothesis of commutativity on the matrices $X_1, \dots, X_n$, one could not have gone further than a concentration on the whole term $\sum_{\sigma \in \mathfrak{S}_p} \text{Tr}(X_{\sigma(1)} \cdots X_{\sigma(m)})$. However, if m is small, say, if $m \leq O(1)$, and the matrices $X_1, \dots, X_m$ have different sizes, we can still introduce the random matrix

$$Y = \begin{pmatrix} 0 & X_{m-1} & & \\ & \ddots & \ddots & \\ & & \ddots & X_1 \\ X_m & & & 0 \end{pmatrix} \quad \text{then} \quad Y^m = \begin{pmatrix} 0 & X_1^m & & \\ & \ddots & \ddots & \\ & & \ddots & X_3^2 \\ X_2^1 & & & 0 \end{pmatrix},$$

where for $i, j \in \{2, \dots, m-1\}$, $X_i^j \equiv X_i X_{i+1} \cdots X_m X_1 \cdots X_j$ and $X_1^m \equiv X_1 \cdots X_m$, then the concentration $Y^m \in \mathcal{E}_q(\sqrt{p}\kappa^{m-1}\sigma) \mid e^{-p}$ in $(\mathcal{M}_p, \|\cdot\|_*)$ directly implies the concentration of X_1^m .

Remark 17. Although a concentration with the nuclear norm is a stronger result than the concentration with the Frobenius norm as the one we got for a product of Lipschitz concentrated matrices (see Proposition 9), note however that here, there is a supplementary constant $\sqrt{p}$ (or $\sqrt{p+n}$) that increases greatly the observable diameter of X^m when X is only convexly concentrated.

PROOF (PROOF OF PROPOSITION 13). We know from Theorem 4 that:

$$\sigma(X) \propto_{\mathfrak{S}_p}^T \mathcal{E}_q \mid e^{-p},$$

and therefore, as a $\sqrt{p}$ -Lipschitz, linear observation of $\sigma(X)^{\odot m} \in \mathcal{E}_q(\kappa^{m-1}\sigma) \mid e^{-p}$, $\text{Tr}(X^m)$ follows the concentration:

$$\text{Tr}(X^m) = \sum_{i=1}^p \sigma_i(X)^m \overset{\mathcal{A}_\kappa}{\in} \mathcal{E}_q(\sqrt{p}\kappa^{m-1}\sigma) \mid e^{-p}.$$

However, to obtain the linear concentration of X^m , what we would like to show is the concentration of any $\text{Tr}(AX^m)$, for a deterministic matrix $A \in \mathcal{M}_p$ such that $\|A\| \leq 1$. We know from Lemma 9 applied on the matrices $(A, X, \dots, X) \in (\mathcal{M}_p)^m$ that:

$$\text{Tr}(AX^m) = \frac{(-1)^m}{m!} \sum_{k=0}^{m-1} \binom{m-1}{k} (-1)^k (k^m \text{Tr}(X^m) - \text{Tr}((A + kX)^m)).$$

Therefore, summing the concentrations:

$$k^m \text{Tr}(X^m) - \text{Tr}((A + kX)^m) \in \mathcal{E}_q \left(\sqrt{p}(\kappa m)^{m-1} \sigma \right) \mid e^{-p},$$

we obtain the first result of the proposition (thanks again to the Stirling formula).

The second result is obtained the same way looking at projections of powers of the matrix

$$Y \equiv \begin{pmatrix} 0 & X^T \\ X & 0 \end{pmatrix} \propto_c \mathcal{E}_2 \quad \text{in } (\mathcal{M}_{p+n, p+n}, \|\cdot\|_F)$$

(like X). Given $m \in \mathbb{N}$, the concentration of $(XX^T)^m$ and $(XX^T)^m X$ is then a consequence of the identities:

$$Y^{2m} = \begin{pmatrix} 0 & (X^T X)^m \\ (X X^T)^m & 0 \end{pmatrix}; \quad Y^{2m+1} = \begin{pmatrix} 0 & (X^T X)^m X^T \\ (X X^T)^m X & 0 \end{pmatrix}.$$

2. Concentration of the solutions of $Y = \phi(Y)$

Given a concentrated random vector $X \in E$ and a random vector $Y = \phi(X)$ for some mapping $\phi : E \rightarrow E$, we have seen in the previous section several results to “transfer” the concentration of X into a concentration for Y . We are now interested in the more intricate concentration of the solution of the equation:

$$Y = \phi(Y)$$

where ϕ is now a random mapping satisfying some concentration properties. The next Remark helps us circumventing this issue.

Remark 18. *To introduce the reader to the issues and settings of this section, let us consider the simple case where $E = \mathbb{R}$ and $\phi : t \mapsto 1 + Xt$ for $X \in \mathbb{R}$, a Gaussian random variable with zero mean and variance equal to σ^2 ($X \sim \mathcal{N}(0, \sigma^2)$). We know that $X \in \mathcal{E}_2$. The solution Y of the equation $Y = \phi(Y)$ is only defined if $X \neq 1$, and in that case $Y = 1/(1 - X)$. The law f_Y of Y can be computed in this particular case and one obtains:*

$$f_Y(y) = \frac{e^{-(1-\frac{1}{y})^2/\sigma^2}}{(1-y)^2}.$$

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

2.1. When ϕ is affine and $\mathcal{L}(\phi)^k(\phi(0)) \in \mathcal{E}_2(\sigma(1-\varepsilon)^k)$ for all $k \in \mathbb{N}$

Theorem 5. Given a (sequence of) reflexive vector space $(E, \|\cdot\|)$, let $\phi \in \mathcal{A}(E)$ be a (sequence of) random mapping such that and there exists a constant $\varepsilon > 0$ and two (sequences of) integers $\sigma, \theta > 0$ satisfying for all (sequence of) integer k :

$$\mathcal{L}(\phi)^k(\phi(0)) \stackrel{\mathcal{A}_\phi}{\in} \mathcal{E}_q(\sigma(1-\varepsilon)^k) \mid e^{-\theta} \text{ in } (E, \|\cdot\|) \text{ and } \mathcal{A}_\phi \subset \{\|\mathcal{L}(\phi)\| \leq 1-\varepsilon\},$$

. Then under the event $\mathcal{A}_\phi$, the random equation

$$Y = \phi(Y)$$

admits a unique solution $Y = (Id_E - \mathcal{L}(\phi))^{-1}\phi(0)$ satisfying the linear concentration:

$$Y \stackrel{\mathcal{A}_\phi}{\in} \mathcal{E}_q(\sigma) \mid e^{-\theta}.$$

Although this theorem is far easier to use than Theorems 6 or 7, we did not give directly because the complex setting of Theorem 6 can be adapted more easily to study afterwards Lipschitz concentration of solutions to non affine equation.

PROOF. Under $\mathcal{A}_\phi$, Y is well defined and expresses:

$$Y = (Id_E - \mathcal{L}(\phi))^{-1}\phi(0) = \sum_{k=0}^{\infty} \mathcal{L}(\phi)^k \phi(0)$$

One can then conclude with Corollary 2 that $(Y|\mathcal{A}_\phi) \in \mathcal{E}_q(\sigma/\varepsilon) = \mathcal{E}_q(\sigma)$.

As an important illustration of this theorem, let us employ the result of the concentration of the powers of convexly concentrated random matrices (Proposition 13) to set the linear concentration of their resolvent.

Corollary 8. Let $X \in \mathcal{S}_p$ be a symmetric random matrix satisfying $X \propto_c \mathcal{E}_2$, and $h, l \leq O(1)$ be two constant integers and $\varepsilon > 0$, a constant. If

$$\|\mathbb{E}[X]\|/\sqrt{p} \leq \frac{1-2\varepsilon}{2e},$$

then with high probability¹⁶ $Q \equiv (I_p - (X/\sqrt{p})^h)^{-1}$ is well defined and

$$\frac{1}{p^{l/2}} Q X^l \in \mathcal{E}_2((1-\varepsilon)^l) \mid e^{-p} \text{ in } (\mathcal{M}_p, \|\cdot\|_*).$$

The same concentration result holds for the random matrix $(I_p - (\frac{1}{n} X X^T)^h)^{-1} (\frac{1}{n} X X^T)^l$ or even $(I_p - (\frac{1}{n} X X^T)^h)^{-1} (\frac{1}{n} X X^T)^l X / \sqrt{n}$ when $X \in \mathcal{M}_{p,n}$ and satisfies $X \propto_c \mathcal{E}_2$ and $\|\mathbb{E}[\frac{1}{n} X X^T]\| \leq \frac{1-2\varepsilon}{(2e)^2}$.

¹⁶We will use several times the expression “with high probability”, mathematically speaking it does not mean much in general, but in our case it means that the assertion is true with a probability bigger than $1 - Ce^{-\theta}$ where θ generally follows the dimension of the objects under study, here $\theta = p$

At first sight, the concentration inequality obtained on the resolvent might seem insufficient, that is untrue. The fact that the concentration is obtained in $(\mathcal{M}_p, \|\cdot\|_*)$ allows us to set that for any matrix $A \in \mathcal{M}_p$ such that $\|A\| \leq O(1)$:

$$\frac{1}{p} \text{Tr}(AQ) \in \mathcal{E}_2\left(\frac{1}{p}\right) \mid e^{-p}$$

which is a very efficient concentration inequality, at the core of random matrix theory inferences.

PROOF. We only need to verify the hypotheses of Theorem 5 in the normed vector space $(\mathcal{M}_p, \|\cdot\|_*)$ and for the mapping $\phi : q \mapsto (X/\sqrt{p})^h q + (X/\sqrt{p})^l$. We know that if we set $\mathcal{A}_X = \{\|X\| \leq \frac{1-\varepsilon}{2e} \sqrt{p}\}$, then $\mathbb{P}(\mathcal{A}_X^c) \leq C e^{-cp}$, for some constants $C, c > 0$. Besides, we know from Proposition 13 that for any $k \in \mathbb{N}$

$$\phi^k(\phi(0)) = \left(\frac{X}{\sqrt{p}}\right)^{kh+l} \overset{\mathcal{A}_X}{\in} \mathcal{E}_2((1-\varepsilon)^l((1-\varepsilon)^h)^k) \mid e^{-p} \quad \text{in } (\mathcal{M}_p, \|\cdot\|_*),$$

which allows us to conclude on the concentration of $\frac{1}{p^{l/2}} QX^l$.

Remark 19. In the setting of Theorem 5 (or Theorems 6, 7, 8 and 9), once one knows that $Y \propto \mathcal{E}_q(\sigma) \mid e^{-\eta}$ one might be tempted to estimate $\mathbb{E}_{\mathcal{A}_Y}[Y]$ with the fixed point $\tilde{Y}_1$, solution to:

$$\tilde{Y}_1 = \mathbb{E}_{\mathcal{A}_\phi}[\phi(\tilde{Y}_1)].$$

That is not so simple : most of the time $\|\mathbb{E}_{\mathcal{A}_\phi}[Y] - \tilde{Y}_1\| \sim O(1)$, and it is more clever to take as in our proof the solution to the expectation of an $O(\log n)$ iteration of the fixed point equation (the ideal estimator being $\tilde{Y}_\infty$ satisfying $\tilde{Y}_\infty = \lim_{k \rightarrow \infty} \mathbb{E}[\phi^k(\tilde{Y}_\infty)]$, in that case of course $\|\mathbb{E}_{\mathcal{A}_\phi}[Y] - \tilde{Y}_\infty\| = O(1/\eta^{1/q})$).

However, if we are given a smaller semi-norm $\|\cdot\|$ satisfying $\forall x \in E, \|x\|' \leq \|x\|$, $\tilde{Y}_1$ can be a relevant choice when $\eta' \equiv \eta_{(E, \|\cdot\|')} \ll 1/\sigma^q$ and when the Lipschitz properties of ϕ are the same for this norm. Let us thus suppose that $\mathbb{P}(\|\phi\|_{(\mathcal{L}, \|\cdot\|')} > 1 - \varepsilon) \leq C e^{-c\eta'}$, for two constants $C, c > 0$, where $\|\phi\|_{(\mathcal{L}, \|\cdot\|')}$ designates the Lipschitz parameter of ϕ for the norm $\|\cdot\|'$. We have the same concentration inequality $Y \propto \mathcal{E}_q(\sigma) \mid e^{-\eta}$ in $(E, \|\cdot\|')$ but this time if we set $\mathcal{A}' = \mathcal{A}_\phi \cap \{\|\phi\|_{(\mathcal{L}, \|\cdot\|')} \leq 1 - \varepsilon\}$ we can bound:

$$\begin{aligned} \mathbb{E}_{\mathcal{A}'}[\|Y - \tilde{Y}_1\|'] &\leq \mathbb{E}_{\mathcal{A}'}[\|\phi(Y) - \phi(\tilde{Y}_1)\|'] + \mathbb{E}_{\mathcal{A}'}[\|\phi(\tilde{Y}_1) - \mathbb{E}_{\mathcal{A}'}[\phi(\tilde{Y}_1)]\|'] \\ &\leq (1 - \varepsilon) \mathbb{E}_{\mathcal{A}'}[\|Y - \tilde{Y}_1\|'] + O(\sigma \eta'^{1/q}), \end{aligned}$$

thus we see that $\mathbb{E}_{\mathcal{A}'}[\|Y - \tilde{Y}_1\|'] = O(\sigma \eta'^{1/q})$ and thanks to Proposition 8:

$$\|\mathbb{E}_{\mathcal{A}'}[Y] - \tilde{Y}\|' \leq O\left(\mathbb{E}_{\mathcal{A}}[\|\mathbb{E}_{\mathcal{A}'}[Y] - Y\|'] + \mathbb{E}_{\mathcal{A}'}[\|Y - \tilde{Y}_1\|']\right) \leq O(\sigma \eta'^{1/q}).$$

This inequality will be in particular used when $q = 2$, $E = \mathbb{R}^p$, $\|\cdot\|$ is the euclidean norm and $\|\cdot\|'$ is the infinite norm. In that case, if $\sigma = \sqrt{\eta_{\|\cdot\|}} = \sqrt{p}$, one has indeed:

$$\left\| \mathbb{E}_{\mathcal{A}'}[Y] - \tilde{Y} \right\|_{\infty} \leq O \left(\sqrt{\frac{\log p}{p}} \right).$$

Nevertheless, this remark does not apply to the case described by Corollary 8, because then, $q = 2$, $E = \mathcal{M}_p$ the big norm is the Frobenius norm $\|\cdot\|_F$ the small norm is the spectral norm $\|\cdot\|_{\text{Sp}}$ (usually we note it $\|\cdot\|$), $\eta_{\|\cdot\|_F} \gg \sigma = \eta_{\|\cdot\|_{\text{Sp}}}$ and therefore the solution of:

$$\tilde{Q} = I_p + \frac{1}{n} \mathbb{E}_{\mathcal{A}'}[X X^T] \tilde{Q}.$$

is not a deterministic equivalent (for the spectral norm) of the expectation of $Q = (I_p - \frac{1}{n} X X^T)^{-1}$. Indeed, the estimation of $\mathbb{E}_{\mathcal{A}_Q}[Q]$ is harder to handle and will be done precisely in Section 3 with the Frobenius norm.

2.2. When $\phi^k(y)$ is linearly or Lipschitz concentrated for $k \leq \log(\eta_{(E, \|\cdot\|)})$ and y deterministic and bounded.

When we can not get a decreasing observable diameter for the iterates of ϕ , or when ϕ is not affine, one needs a different approach that allows to treat, at the same time affine and non-affine mappings ϕ and extend very simply linear concentration inferences to Lipschitz concentration inferences. We first give the result of concentration of a mapping ϕ affine since the hypotheses express more simply.

Theorem 6. *Given a (sequence of) reflexive¹⁷ vector space $(E, \|\cdot\|)$, we note η its norm degree (of course $\eta \geq O(1)$). Let $\phi \in \mathcal{A}(E)$ be a (sequence of) random mapping such that there exists two (sequences of) integers $\sigma, \theta > 0$ satisfying for all (sequence of) integer k such that $k \leq O(\log(\eta))$ and for any $y \in E$ such that $\|y\| \leq O(1)$:*

$$\phi^k(y) \stackrel{\mathcal{A}_\phi}{\in} \mathcal{E}_q(\sigma) \mid e^{-\theta} \quad \text{in } (E, \|\cdot\|) \quad (17)$$

We additionally suppose that there exists a constant $\varepsilon \geq O(1)$ such that $\mathcal{A}_\phi \subset \{\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon\}$ (recall that $\mathcal{L}(\phi) = \phi - \phi(0) \in \mathcal{L}(E)$) and that $\mathbb{E}_{\mathcal{A}_\phi}[\|\phi(0)\|] \leq O(1)$. Then under the event $\mathcal{A}_\phi$, the random equation

$$Y = \phi(Y)$$

admits a unique solution $Y = (Id_E - \mathcal{L}(\phi))^{-1} \phi(0)$ satisfying the linear concentration:

$$Y \stackrel{\mathcal{A}_\phi}{\in} \mathcal{E}_q(\sigma) \mid e^{-\theta}.$$

¹⁷We suppose E reflexive to be able to define an expectation operator on the set of random vectors of E – see Appendix A

Remark 20. One can then generally adapt the choice of $\mathcal{A}_\phi$ to the needs of this theorem. More precisely, if we note $\mathcal{A}_k$ the concentration zone of ϕ^k , one can choose $\mathcal{A}_\phi = \mathcal{A}_k \cap \{\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon\}$ if there exist two constants $C, c > 0$, such that $\mathbb{P}(\{\|\mathcal{L}(\phi)\| \geq 1 - \varepsilon\}) \leq Ce^{-c\theta}$ (that could have been an assumption of the Theorem replacing “ $\mathcal{A}_\phi \subset \{\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon\}$ ”).

Remark 21. When $\sigma = O(1/\eta^{1/q})$ and $\theta = O(\eta)$ for $\eta = \eta_{(E, \|\cdot\|)}$ (for instance if $\phi : Y \mapsto X^T Y / \sqrt{n}$ and $X \propto 0 \pm \mathcal{E}_2$) one has often (at least when ϕ is Lipschitz or convexly concentrated) $\|\mathcal{L}(\phi)\| \in \mathbb{E}[\|\mathcal{L}(\phi)\|] \pm \mathcal{E}_q(1/\eta^{1/q})$ and therefore if $\mathbb{E}[\|\mathcal{L}(\phi)\|] \leq 1 - 2\varepsilon$, then $\mathbb{P}(\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon) \leq Ce^{-\eta/c}$ for some constants $C, c > 0$.

Remark 22. Sometimes, one could consider a situation where for $l = O(\log(\eta))$, and $\|y\| = O(1)$, $\mathcal{L}(\phi)^l(y) \in \mathcal{E}_q(\sigma) \mid e^{-\theta}$ but $\phi^l(y)$ has an observable diameter of a bigger order than σ (and $\|\phi(0)\| \gg O(1)$). In that case, if, say, $\phi(0) \in \mathcal{E}_q(\kappa) \mid e^{-\theta}$ and $\mathbb{E}_{\mathcal{A}_\phi}[\|\phi(0)\|] \leq O(\kappa)$, one then has to consider the mapping $\phi_\kappa : x \mapsto \phi(\kappa x)/\kappa$ that satisfies the hypothesis of the theorem:

- for all $y \in E$, $\mathcal{L}(\phi_\kappa)^l(y) = \mathcal{L}(\phi)^l(y)$ thus they satisfy the same hypotheses
- $\phi_\kappa^l(0) = \sum_{i=0}^l \mathcal{L}(\phi)^i(\phi(0)/\kappa) \in \mathcal{E}_q(l\sigma)(\sigma) \mid e^{-\theta}$ (we implicitly assume that the bilinear form $\mathcal{L}(\phi)^i(\phi(0)/\kappa)$ taken in $\mathcal{L}(\phi)^i$ and $\phi(0)$ follows the concentration properties described by Theorem 2, or Propositions 9, 12 or 13)
- $\mathbb{E}_{\mathcal{A}_\phi}[\|\phi_\kappa(0)\|] = \mathbb{E}_{\mathcal{A}_\phi}[\|\phi(0)\|]/\kappa \leq O(1)$

Then the solution Y_κ verifying $Y_\kappa = \phi_\kappa(Y_\kappa)$ follows the concentration $Y_\kappa \propto \mathcal{E}_2(\sigma) \mid e^{-\theta}$ and therefore $Y = \kappa Y_\kappa \propto \mathcal{E}_2(\kappa\sigma) \mid e^{-\theta}$. We found cleverer to keep the simpler setting possible for the theorem, and to do this kind of small adaptation if needed.

It is possible to prove a stronger result than Theorem 6 if there exists a supplementary norm $\|\cdot\|'$ satisfying $\forall y \in E: \|y\|' \leq \|y\|$ allowing us to weaken the hypothesis.

Theorem 7. In the setting of Theorem 6, given any supplementary norm $\|\cdot\|'$ on E , if we suppose:

- (17) is true just any $y \in E$ such that $\|y\|' \leq O(1)$ (and not for any $y \in E$ such that $\|y\| \leq O(1)$),
- $\mathbb{E}_{\mathcal{A}_\phi}[\|\phi(0)\|'] \leq O(1)$ instead of $\mathbb{E}_{\mathcal{A}_\phi}[\|\phi(0)\|] \leq O(1)$,
- $\mathcal{A}_\phi \subset \{\|\mathcal{L}(\phi)\|' \leq 1 - \varepsilon\}$ (in addition to $\mathcal{A}_\phi \subset \{\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon\}$)

then we obtain the same concentration result for Y .

Let us prove this theorem that will imply directly Theorem 7 (taking $\|\cdot\|' = \|\cdot\|$).

PROOF. Under $\mathcal{A}_\phi$, the mapping $y \mapsto \phi(y)$ is contractive, that proves the existence and uniqueness of Y (E is complete since it is reflexive – see Remark 27).

To show the linear concentration of Y thanks to Remark 27, let us consider a unit normed linear form f on E (it satisfies $\forall x \in E: |f(x)| \leq \|x\|$). For a reason that will become clear later, we let $k = \lceil -\frac{\log(\eta)}{q \log(1-2\varepsilon)} \rceil$ (in particular $k \leq O(\log(\eta))$). There exist two constants $C, c \leq O(1)$ such that for any $y \in E$ and any sequence of unit normed linear function $f_p : E_p \rightarrow \mathbb{R}$ (we write it rigorously just this time to clarify one more time our notations) $\forall p \in \mathbb{N}, \forall t > 0$:

$$\mathbb{P} \left(\left| f_p(\phi_p^{k_p}(y)) - \mathbb{E}_{\mathcal{A}_{\phi_p}} [f_p(\phi_p^{k_p}(y))] \right| \geq t \mid \mathcal{A}_{\phi_p} \right) \leq C e^{-(t/c\sigma_p \|y\|)^q},$$

where the notation “ $\mathbb{E}_{\mathcal{A}_{\phi_p}}$ ” is introduced in equation (13). Note that the observable diameter (see Remark ??) depends on the norm of y . Let us now introduce the function

$$\begin{aligned} \tilde{\phi}_k : E &\longrightarrow E \\ y &\longmapsto \mathbb{E}_{\mathcal{A}_\phi} [\phi^k(y)]. \end{aligned}$$

We can then bound:

$$\begin{aligned} \|\mathcal{L}(\tilde{\phi}_k)\| &= \|\mathcal{L}(\mathbb{E}_{\mathcal{A}_\phi} [\phi^k])\| \leq \mathbb{E}_{\mathcal{A}_\phi} [\|\mathcal{L}(\phi)^k\|,] \\ &\leq \mathbb{E} [\|\mathcal{L}(\phi)\|^k, \|\mathcal{L}(\phi)\| \leq 1 - \varepsilon] \leq 1 - \varepsilon < 1. \end{aligned}$$

The mapping $y \mapsto \tilde{\phi}_k(y)$ is thus contractive: we can introduce the (uniquely defined) deterministic vector $\tilde{Y}$ satisfying:

$$\tilde{Y} = \tilde{\phi}_k(\tilde{Y}),$$

which we will show is a deterministic equivalent for Y . To this end, let us first bound $\|\tilde{Y}\|'$ to be able to control the concentration of $\phi^k(\tilde{Y})$:

$$\begin{aligned} \|\tilde{Y}\|' &= \left\| \mathbb{E}_{\mathcal{A}_\phi} \left[\mathcal{L}(\phi)^k \tilde{Y} + \sum_{i=0}^{k-1} \mathcal{L}(\phi)^i (\phi(0)) \right] \right\|' \\ &\leq \mathbb{E}_{\mathcal{A}_\phi} \left[\|\mathcal{L}(\phi)\|'^k \|\tilde{Y}\|' + \left(\sum_{i=0}^{k-1} \|\mathcal{L}(\phi)\|'^i \right) \|\phi(0)\|' \right] \\ &\leq (1 - \varepsilon) \|\tilde{Y}\|' + \frac{1}{\varepsilon} \mathbb{E}_{\mathcal{A}_\phi} [\|\phi(0)\|']. \end{aligned}$$

Therefore $\|\tilde{Y}\|' \leq \frac{1}{\varepsilon^2} \mathbb{E}_{\mathcal{A}_\phi} [\|\phi(0)\|'] \leq O(1)$ and $(\phi^k(\tilde{Y}) \mid \mathcal{A}_\phi) \propto \mathcal{E}_q(\sigma)$.

Then let us try and bound $\|Y - \tilde{Y}\|$ under $\mathcal{A}_\phi$:

$$\begin{aligned} \|Y - \tilde{Y}\| &= \|\phi^k(Y) - \tilde{\phi}_k(\tilde{Y})\| \leq \|\phi^k(Y) - \phi^k(\tilde{Y})\| + \|\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\| \\ &\leq \|\mathcal{L}(\phi)\|^k \|Y - \tilde{Y}\| + \|\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\|. \end{aligned}$$

Under $\mathcal{A}_\phi$, $\|\mathcal{L}(\phi)\|^k \leq 1 - \varepsilon$, so that:

$$\|Y - \tilde{Y}\| \leq \frac{1}{\varepsilon} \|\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\|.$$

But we know from Proposition 8 that:

$$\left(\|\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\| \mid \mathcal{A}_\phi \right) \in O(\eta^{1/q}\sigma) \pm \mathcal{E}_q(\eta^{1/q}\sigma)$$

which allows us to conclude that $\|Y - \tilde{Y}\| \in O(\eta^{1/q}\sigma) \pm \mathcal{E}_q(\eta^{1/q}\sigma) \mid e^{-\theta}$.

Returning to our initial goal (the linear concentration of Y), we now bound, for $f \in \mathcal{L}(E)$ and under $\mathcal{A}_\phi$ (we still have $\|\mathcal{L}(\phi)\| \leq 1 - \varepsilon$),

$$\begin{aligned} |f(Y) - f(\tilde{Y})| &\leq \left| f\left(\phi^k(Y) - \phi^k(\tilde{Y})\right) \right| + \left| f\left(\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\right) \right| \\ &\leq (1 - \varepsilon)^k \|Y - \tilde{Y}\| + \left| f\left(\phi^k(\tilde{Y}) - \tilde{\phi}_k(\tilde{Y})\right) \right|. \end{aligned}$$

Further, noting that, with our choice of k , $(1 - \varepsilon)^k = O(1/\eta^{1/q})$, we conclude again from the concentration of $\phi^k(\tilde{Y})$ that

$$f(Y) \in f(\tilde{Y}) \pm \mathcal{E}_q(\sigma) \mid e^{-\theta},$$

thereby giving the sought-for concentration result.

It is possible to get a similar result to Theorem 6 with non affine mappings and in the case of Lipschitz concentration. Given a normed vector space $(E, \|\cdot\|)$, we note $\mathcal{F}(E)$, the set of mappings from E to E . If f is bounded, we denote $\|f\|_\infty = \sup_{x,y \in E} \|f(x)\|$. For a Lipschitz mapping $f \in F(E)$, we introduce the seminorm $\|\cdot\|_{\mathcal{L}}$ which provides the Lipschitz parameter and will play the role of $\|\mathcal{L}(\phi)\|$ in Theorem 6:

$$\|f\|_{\mathcal{L}} = \sup_{\substack{x,y \in E \\ x \neq y}} \frac{\|f(x) - f(y)\|}{\|x - y\|}.$$

Theorem 8. *Let us consider a (sequence of) reflexive vector space $(E, \|\cdot\|)$ admitting a finite norm degree that we note η . Given $\phi \in F(E)$, a (sequence of) random mapping, we suppose that there exists a (sequence of) integer $\sigma > 0$, a constant $\varepsilon > 0$ such that there exists a (sequence of) highly probable event $\mathcal{A}_\phi$ satisfying:*

- $\mathcal{A}_\phi \subset \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$,
- for any (sequence of) integer k such that $k \leq O(\log(\eta))$, noting $y_0 \in E$, the (sequence of) fixed point to the deterministic equation $y_0 = \mathbb{E}_{\mathcal{A}_\phi}[\phi^k](y_0)$ ¹⁸:

$$\phi^k(y_0) \overset{\mathcal{A}_\phi}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta} \text{ in } (E, \|\cdot\|).$$

¹⁸the first assumption on $\|\phi\|_{\mathcal{L}}$ ensures the existence and uniqueness of y_0 .

Then, the random equation $Y = \phi(Y)$ admits under $\mathcal{A}_\phi$ a unique solution $Y \in E$ satisfying the Lipschitz concentration:

$$Y \overset{\mathcal{A}_\phi}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta}.$$

It can happen that one wants to employ Theorem 8 for functions ϕ whose Lipschitz semi-norm are not finite. Still some property of concentration on fixed point can be inferred if the Lipschitz parameter of ϕ can be controlled around y_0 satisfying $y_0 = \mathbb{E}_{\mathcal{A}_\phi}[\phi^k](y_0)$ or, more simply $y_0 = \mathbb{E}[\phi(y_0)]$ we detail this marginal setting in Appendix B.

PROOF. The existence and uniqueness of Y under $\mathcal{A}_\phi$ is justified, as in the proof of Theorem 6 by the contractiveness of ϕ . Of course, $y \mapsto \mathbb{E}_{\mathcal{A}_\phi}[\phi^k(y)]$ is also contractive, wich justifies the existence of y_0 .

If we first want to show that Y is linearly concentrated, one can follow the last steps of the proof of Theorem 6 with y_0 replacing $\tilde{Y}$ (here, there is no need to verify that $\|\tilde{Y}\|$ is bounded), and conclude that $Y \in \mathcal{E}_q(\sigma) \mid e^{-\theta}$.

To show the Lipschitz concentration of Y , let us consider a Lipschitz map $f : E \rightarrow \mathbb{R}$ and introduce the mappings:

$$\begin{aligned} U : E &\longmapsto E \times \mathbb{R} & V : E \times \mathbb{R} &\longmapsto E \\ y &\longmapsto (y, f(y)) & (y, t) &\longmapsto y. \end{aligned}$$

Note that if we endow $E \times \mathbb{R}$ with the norm $\|\cdot\|_{\ell^\infty}$ satisfying $\forall(x, t) \in E \times \mathbb{R}$, $\|(x, t)\|_{\ell^\infty} = \max(\|x\|, |t|)$ then the mappings U and V are both 1-Lipschitz and consequently for all constant $K > 0$, $A_K \subset \{\|U \circ \phi \circ V\|_{\mathcal{L}} \leq 1 - \varepsilon\}$ and we can consider $x_0 \in E \times \mathbb{R}$ solution to:

$$x_0 = \mathbb{E}[U \circ \phi^k \circ V(x_0) \mid \|U \circ \phi \circ V\|_{\mathcal{L}} \leq 1 - \varepsilon].$$

for a given (sequence of) integer $k \leq O(\log \eta)$ (note that $V \circ U = Id_E$). Then, noting $(y_0, t_0) = x_0$:

$$(U \circ \phi \circ V)^k(x_0) = (\phi^k(y_0), f(\phi^k(y_0))) \overset{A_K}{\in} \mathcal{E}_q(\sigma) \mid e^{-\eta} \text{ in } (E, \|\cdot\|)$$

by hypothesis. Therefore, one can deduce the linear concentration of the unique solution to the fixed point equation:

$$X = U \circ \phi \circ V(X) \quad X \in E \times \mathbb{R}, \quad (18)$$

that satisfies $X \in \mathcal{E}_2(\sigma) \mid e^{-\eta}$. As a consequence, noting $\pi : E \times \mathbb{R} \rightarrow \mathbb{R}$ the projection on the last variable, we know that:

$$\pi(X) \propto \mathcal{E}_q(\sigma) \mid e^{-\eta}.$$

Since $V \circ U = Id_E$, we see from the definition of Y that $U(Y)$ is also a solution to (18), therefore, by uniqueness of the solution (when it exists), $X = U(Y)$ and, in particular $f(Y) = \pi(X)$ which allows us to conclude on the concentration of Y .

2.3. When $\phi \propto \mathcal{E}_2(\sigma)$ for the infinity norm

If one can assume the stronger hypothesis that ϕ is concentrated as a random mapping, and for the infinity norm, then it is unnecessary to consider the $O(\log(\eta))$ iterate of ϕ . We can indeed apply Theorem 2 to infer from the concentration of ϕ , the concentration of its iterates. We give here a weak setting where the mapping ϕ is only bounded around a given point y_0 , for that, for any $r > 0$, we introduce the semi-norm $\|\cdot\|_{\mathcal{B}(y_0, r)}$ defined for any $f \in \mathcal{F}(E)$ and $y \in E$ as:

$$\|f\|_{\mathcal{B}(y_0, r)} = \sup_{\|y - y_0\| \leq r} \|f(y)\|$$

The following Lemma might seem a bit complicated and artificial, it is however perfectly adapted to the requirement of Theorem 8.

Lemma 10. *Given a normed vector space $(E, \|\cdot\|)$ whose norm degree is noted η , a vector $y_0 \in E$ and a (sequence of) random mapping $\phi \in \mathcal{F}^\infty(E)$, let us suppose that there exists a (sequence of) constant $\varepsilon > 0$ ($\varepsilon \geq O(1)$) such that for any constant $K > 0$, there exists a (sequence of) event $\mathcal{A}_K$ satisfying $\mathcal{A}_K \subset \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$ and:*

$$\phi \stackrel{\mathcal{A}_K}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\eta} \text{ in } (\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K\sigma\eta^{1/q})}),$$

then for any constant K' ($K' \leq O(1)$) we have:

$$\phi^m \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\eta} \text{ in } (\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})}),$$

PROOF. Let us introduce three constants $\kappa, C, c > 0$ ($\kappa, C, c \leq O(1)$) such that:

$$\forall K \geq \kappa : \mathbb{P}(\|\phi(y_0) - y_0\| \geq K\sigma\eta^{1/q}\varepsilon \mid \mathcal{A}_1) \leq Ce^{-\eta/c},$$

Now, given a constant $K' > 0$, we choose $K = \max(K'/\varepsilon, \kappa)$ and we set:

$$\mathcal{A}_{\phi^\infty} \equiv \mathcal{A}_{K\varepsilon} \cap \mathcal{A}_1 \cap \left\{ \|\phi(y_0) - y_0\| \leq K\sigma\eta^{1/q}\varepsilon \right\}.$$

Then we see that $\mathbb{P}(\mathcal{A}_{\phi^\infty}^c) \leq C'e^{-c'\eta}$, for some constants $C', c' > 0$ and we can bound under $\mathcal{A}_{\phi^\infty}$, for any $k \in \mathbb{N}$ and $y \in \mathcal{B}(y_0, K'\sigma\eta^{1/q})$:

$$\begin{aligned} \|\phi^k(y) - y_0\| &\leq \|\phi^k(y) - \phi^k(y_0)\| + \|\phi^k(y_0) - y_0\| \\ &\leq (1 - \varepsilon)^k \|y - y_0\| + \sum_{i=0}^{k-1} (1 - \varepsilon)^i \|\phi(y_0) - y_0\| \leq K\sigma\eta^{1/q} \end{aligned}$$

Thus, for any $f \in \mathcal{A}_{\phi^\infty}$ and for all $k \in \mathbb{N}$, $f^k : \mathcal{B}(y_0, K'\sigma\eta^{1/q}) \rightarrow \mathcal{B}(y_0, K\sigma\eta^{1/q})$, and we can bound for any supplementary mapping $g \in \mathcal{A}_{\phi^\infty}$:

$$\begin{aligned} \|f^m - g^m\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})} &\leq \sum_{i=1}^m (1 - \varepsilon)^{i-1} \|f(g^{m-i}(y)) - g(g^{m-i}(y))\| \\ &\leq \frac{1}{\varepsilon} \|f - g\|_{\mathcal{B}(y_0, K\varepsilon\sigma\eta^{1/q})}. \end{aligned}$$

Thus the mapping $f \mapsto f^m$ is $\frac{1}{\varepsilon}$ -Lipschitz from $(\mathcal{A}_{\phi^\infty}, \|\cdot\|_{\mathcal{B}(y_0, K\sigma\eta^{1/q})})$ to $(\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})})$, which directly implies the result of the lemma.

The next theorem is an important improvement (allowed by Lipschitz concentration) of Theorem 8 in that it only take as hypothesis the concentration of ϕ (and not of all its iterates).

Theorem 9. *Let us consider a (sequence of) reflexive vector space $(E, \|\cdot\|)$ whose norm degree is noted η , a (sequence of) random mapping $\phi : E \rightarrow E$, a given constant $\varepsilon > 0$ ($\varepsilon \geq O(1)$) such that for any constant $K > 0$ ($K \leq O(1)$), there exists a (sequence of) highly probable event $\mathcal{A}_K$ satisfying:*

- $\mathcal{A}_K \subset \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$,
- *Noting $y_0 \in E$, the (sequence of) fixed point to the deterministic equation $y_0 = \mathbb{E}_{\mathcal{A}_K}[\phi^k](y_0)$, we have the concentration:*

$$\phi(y_0) \overset{\mathcal{A}_K}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta} \quad \text{in } (\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K\sigma\eta^{1/q})})$$

then there exists an highly probable event $\mathcal{A}_Y$, such that, under $\mathcal{A}_Y$, the random equation

$$Y = \phi(Y)$$

admits a unique solution $Y \in E$ satisfying the Lipschitz concentration:

$$Y \propto \mathcal{E}_q(\sigma) \mid e^{-\eta}.$$

2.4. *When $\phi = \Psi(X)$ with Ψ deterministic and X concentrated*

As we saw with Corollary 8, a classical setting of Theorems 6-9 is the case where the randomness of ϕ depends on a random vector $X \in F$ (for a normed vector space $(F, \|\cdot\|)$) and the fixed point equation writes:

$$Y = \Psi(X)(Y)$$

for a given deterministic mapping $\Psi : F \rightarrow \mathcal{F}(E)$. The issue is then, not only to show the concentration of Y but also the concentration of (X, Y) to be able to control the operations made on X and Y .

Corollary 9. *Given two reflexive vector space $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, a random vector $X \in F$ satisfying $X \propto \mathcal{E}_q$ and a deterministic mapping $\Psi : F \rightarrow \mathcal{F}(E)$, we assume that there exist a constant $\varepsilon > 0$ and an event $\mathcal{A}_\Psi$ such that:*

- $\mathbb{E}_{\mathcal{A}_\Psi}[\Psi(X)]$ is defined (see Definition 14 in Appendix A)
- $\mathcal{A}_\Psi \subset \{\|\Psi(X)\|_{\mathcal{L}} < 1 - \varepsilon\}$, $\mathbb{P}(\mathcal{A}_\Psi^c) \leq Ce^{-c\eta}$ (where η is the norm degree of $\|\cdot\|_E$)

- Noting $y_0 \in E$, the unique solution to $y_0 = \mathbb{E}_{\mathcal{A}_\Psi}[\Psi(X)(y_0)]$, for any constant $K > 0$ ($K \leq O(1)$), the mapping $A \mapsto \Psi(A)$ is $O(\sigma)$ -Lipschitz from $(\mathcal{A}_\Psi, \|\cdot\|_F)$ to $(\mathcal{F}(E), \|\cdot\|_{B_E(y_0, K\sigma\sqrt{\eta})})$

Then with high probability, there exists a unique random vector $Y \in E$ such that $Y = \Psi(X)(Y)$ and it satisfies:

$$(\sigma X, Y) \propto \mathcal{E}_2(\sigma) \mid e^{-\eta}.$$

PROOF. We just need to verify the hypotheses of Theorem 9 with:

$$\begin{aligned} \phi : (F \times E, \|\cdot\|_{\ell^\infty}) &\longrightarrow (F \times E, \|\cdot\|_{\ell^\infty}) \\ (x, y) &\longmapsto (\sigma X, \psi(X)(y)), \end{aligned}$$

where for $(x, y) \in F \times E$, $\|(x, y)\|_{\ell^\infty} = \max(\|x\|_F, \|y\|_E)$. For any constant $K > 0$, the random mapping ϕ is clearly a $O(\sigma)$ -Lipschitz transformation of X for the norm $\|\cdot\|_{B(z_0, K\sigma\eta^{1/q})}$, therefore, it satisfies:

$$\phi \stackrel{\mathcal{A}_\Psi}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\eta} \text{ in } (\mathcal{F}(E), \|\cdot\|_{B(z_0, K\sigma\eta^{1/q})}),$$

where $z_0 = (\sigma \mathbb{E}_{\mathcal{A}_\Psi}[X], y_0) \in F \times E$. And of course:

$$\{\|\phi\|_{\mathcal{L}} \geq 1 - \varepsilon\} = \{\|\Psi(X)\|_{\mathcal{L}} \geq 1 - \varepsilon\} \supset \mathcal{A}_\Psi$$

One can therefore employ Theorem 9 to ϕ to set the existence and concentration of $(\sigma X, Y)$.

3. First example : Concentration of the resolvent and device of a deterministic equivalent

3.1. Setting, assumptions and first properties

Let us first study the concentration of a central object appearing in random matrices, the so-called “resolvent” defined as the matrix $(I_p - Z)^{-1}$ for $Z \in \mathcal{M}_p$, a random matrix. We already gave in Corollary 8 some conditions on a convexly concentrated random matrix Z for Q to be concentrated, we will however first place ourselves in a Lipschitz concentration setting. Let us look at the matrix $Q \equiv (I_p - \frac{1}{n} X X^T)^{-1}$, where $X = (X_{n,p})_{n,p \in \mathbb{N}} \in \prod_{p,n \in \mathbb{N}^2} \mathcal{M}_{p,n}$ is a (sequences of) random matrices. We suppose:

Assumption 1. $X \propto \mathcal{E}_2^{19}$.

¹⁹In the initial Definition 1, we defined the concentration of a sequence of random vectors and here, $X_{n,p}$ is indexed by two natural numbers. A slight change of Definition 1 allows us to adapt it to any set of indexes S for $(X_s)_{s \in S}$ (in particular to $S = \mathbb{N}^2$), the two constants $C, c > 0$ appearing in the concentration inequality are assumed to be valid for any $s \in S$ (i.e. for any $n, p \in \mathbb{N}^2$).

Assumption 2. X has independent columns $x_1, \dots, x_n \in \mathbb{R}^{p20}$

Assumption 3. $O(p) \leq n \leq O(p)^{21}$.

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

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

We know from Proposition 4 that $\sup_{1 \leq i \leq n} \|C_i\| \leq O(1)$, but we also need to bound:

Assumption 4. $\sup_{1 \leq i \leq n} \|\mu_i\| \leq O(1)$ and $\inf_{1 \leq i \leq n} \|\frac{1}{n} \text{Tr} \Sigma_i\| \geq O(1)^{22}$

This assumption allows us to bound $\|X\|$ thanks to Proposition 8. We have indeed:

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

thus we know from Proposition 8 that $\mathbb{E}[\|X\|] \leq \|\mathbb{E}[X]\| + O(\sqrt{n}) \leq O(\sqrt{n})$ and therefore, $\|X\| \in O(\sqrt{n}) \pm \mathcal{E}_2$. For $(I_p - \frac{1}{n} X X^T)$ to be invertible with high probability we need a last assumption:

Assumption 5. There exists $\varepsilon \geq O(1)$ such that $\|\frac{1}{n} \sum_{i=1}^n \Sigma_i\| \leq 1 - 2\varepsilon$.

In that case, we can indeed deduce from the linear concentration of $X X^T$ given in example 3 and the inferences on the norm provided by Proposition 8 that:

$$\mathbb{P}(\mathcal{A}_Q^c) \leq C e^{-cn}, \quad \text{for} \quad \mathcal{A}_Q \equiv \left\{ \frac{1}{n} \|X X^T\| \leq 1 - \varepsilon \right\}, \quad (19)$$

for two constants $C, c > 0$.

Strangely enough, a complex relaxation of our setting is necessary to design a deterministic equivalent of Q and justify its validity. We define for that reason for any $z \in \mathbb{C} \setminus [0, 1 - \varepsilon]$ and under $\mathcal{A}_Q$ the complex random matrix:

$$Q^z \equiv \left(z I_n - \frac{1}{n} X X^T \right)^{-1} \in \mathcal{M}_p(\mathbb{C}).$$

With the notation $|M|^2 = M \bar{M}^T \in \mathcal{S}_p^+(\mathbb{R})$ for any matrix $M \in \mathcal{M}_p(\mathbb{C})$, generalizing the notion of squared modulus to any complex matrix, we have the following Lemma:

²⁰note that we do not assume that the x_i are identically distributed as it is not required.

²¹Assumption 1 introduces two asymptotic directions, $n \rightarrow \infty$ and $p \rightarrow \infty$, the notation $a_{n,p} \leq O(b_{n,p})$ (resp. $a_{n,p} \geq O(b_{n,p})$) thus means that there exists a constant $C > 0$ such that $\forall n, p \in \mathbb{N}$, $a_{n,p} \leq C b_{n,p}$ (resp. $a_{n,p} \geq C b_{n,p}$). Assumption 3 therefore restricts our study to choices of p and n satisfying an equality $cp \leq n \leq Cp$, for two constants $C, c > 0$.

²²This second hypothesis on the statistics of X is not introduced to set the concentration of Q but for the design of a deterministic equivalent. If the covariance of a vector x_i is too small, one should be able to replace it by its expectation in the construction of a deterministic equivalent of Q , however, in this quasi asymptotic regime, it is not easy to identify the correct threshold, thus we prefer to place ourselves in the most common case where the energy of every data is taken into account in our estimation.

Lemma 11. *In the set of symmetric matrices:*

$$\frac{I_p}{|z|^2 + (1 - \varepsilon)^2} \leq |Q^z|^2 \leq \frac{I_p}{d(z, [0, 1 - \varepsilon])^2},$$

where for any $w \in \mathbb{C}$ and $A \subset \mathbb{C}$, $d(w, A) = \inf\{w - v, v \in A\}$.

PROOF. If we note $A = \Re(Q^{-1})$ and $B = \Im(Q^{-1})$, then $(|Q_{-i}|^2)^{-1} = A^2 + B^2$ and:

$$A^2 = \left(\Re(z)I_p - \frac{1}{n}XX^T \right)^2 \quad \text{and} \quad B^2 = \Im(z)^2 I_p.$$

One can then deduce the result of the Lemma from the inequality $0 \leq \frac{1}{n}XX^T \leq (1 - \varepsilon)I_p$ (under $\mathcal{A}_Q$) and:

$$|z|^2 + (1 - \varepsilon)^2 \geq A^2 + B^2 \geq \Im(z)^2 + d(\Re(z), [1 - \varepsilon])^2 = d(z, [0, 1 - \varepsilon])^2.$$

Proposition 14. *If $d(z, [0, 1 - \varepsilon]) \geq O(1)$, then $Q^z \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_F)$.*

PROOF. We could see it as a consequence of Theorem 9, applied to the equation

$$Q^z = \frac{I_p}{z} + \frac{1}{zn}XX^TQ^z,$$

but it is probably more simple to see Q^z as a $O(1)$ -Lipschitz transformation of $\frac{1}{nXX^T} \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ (see Example 3). Indeed if we note $\Phi : \mathcal{M}_p \rightarrow \mathcal{M}_p(\mathbb{C})$ defined as $\Phi(M) = (zI_p - M)^{-1}$, we have for $M \in (\frac{1}{n}XX^T)(\mathcal{A}_Q)^{23}$:

$$\|d\Phi|_M\| = \|\Phi(M)M\Phi(M)\| \leq \frac{1 - \varepsilon}{d(z, [0, 1 - \varepsilon])^2} \leq O(1).$$

3.2. A first deterministic equivalent

One is often merely working with linear functionals of Q^z , and since Proposition 14 implies that $Q^z \in \mathbb{E}_{\mathcal{A}_Q} Q^z \pm \mathcal{E}_2 \mid e^{-n}$, one naturally wants to estimate the expectation $\mathbb{E}_{\mathcal{A}_Q} Q$. In Louart and Couillet (2019) is provided a deterministic equivalent $\tilde{Q}^z \in \mathcal{M}_p$ satisfying $\|Q^z - \tilde{Q}^z\| \leq O(1/\sqrt{n})$ 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 $z \in \mathbb{C}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$.

²³That means in particular that $\|M\| \leq 1 - \varepsilon$.

An efficient approach, developed in particular in Silverstein (1986), 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:

$$\tilde{Q}^z(\Delta) = (zI_p - \Sigma_\Delta)^{-1} \quad \text{where} \quad \Sigma_\Delta = \frac{1}{n} \sum_{i=1}^n \Delta_i \Sigma_i.$$

One can then express the difference with the expectation of Q^z (the natural deterministic equivalent that we try to estimate) followingly:

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

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:

$$Q^z = Q_{-i}^z + \frac{1}{n} \frac{Q_{-i}^z x_i x_i^T Q_{-i}^z}{1 - \frac{1}{n} x_i^T Q_{-i}^z x_i} \quad \text{and} \quad Q^z x_i = \frac{Q_{-i}^z x_i}{1 - \frac{1}{n} x_i^T Q_{-i}^z x_i}, \quad (20)$$

for $Q_{-i}^z = (zI_n - \frac{1}{n} X_{-i} X_{-i}^T)^{-1}$ and $X_{-i} = (x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \in \mathcal{M}_{p,n}$. Then introducing the notation $D^z \equiv \text{Diag}_{1 \leq i \leq n} (\frac{1}{n} x_i^T Q_{-i}^z x_i)$ and the mapping $\chi : \mathbb{C} \setminus \{1\} \rightarrow \mathbb{C}$, defined as:

$$\chi(z) = \frac{1}{1-z},$$

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

$$\begin{aligned} \tilde{Q}^z(\chi(\Delta^z)) - \mathbb{E}_{\mathcal{A}_Q} Q &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q} \left[Q_{-i}^z \left(\frac{x_i x_i^T}{1 - D_i^z} - \frac{\Sigma_i}{1 - \Delta_i^z} \right) \tilde{Q}^z(\chi(\Delta^z)) \right] \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \frac{1}{1 - \Delta_i^z} \mathbb{E}_{\mathcal{A}_Q} \left[Q^z x_i x_i^T Q_{-i}^z \Sigma_i \tilde{Q}^z(\chi(\Delta^z)) \right] \\ &= \mathbb{E}_{\mathcal{A}_Q} [\varepsilon_1 + \varepsilon_2] \end{aligned} \quad (21)$$

with :

$$\begin{cases} \varepsilon_1 = \frac{1}{n} \mathbb{E}_{\mathcal{A}_Q} \left[Q^z X \chi(\Delta^z) (D^z - \Delta^z) X^T \tilde{Q}^z(\chi(\Delta^z)) \right] \\ \varepsilon_2 = \frac{1}{n^2} \sum_{i=1}^n \chi(\Delta_i^z) \mathbb{E}_{\mathcal{A}_Q} \left[Q^z x_i x_i^T Q_{-i}^z \Sigma_i \tilde{Q}^z(\chi(\Delta^z)) \right], \end{cases}$$

From this decomposition, one is enticed into choosing, in a first step:

$$\Delta^z \equiv \mathbb{E}_{\mathcal{A}_Q} [D^z] \in \mathcal{D}_n(\mathbb{C})$$

(so that ε_1 would be small). The concentration of D^z can be deduced from the concentration of

$$E^z \equiv \text{Diag} \left(\frac{1}{n} X Q^z X \right) \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$$

and Schur identities (20) that imply:

$$D^z = \frac{E^z}{E^z + I_n}.$$

To show the concentration of $\frac{I_n}{E^z + I_n}$, one needs to show that $E^z + I_n = \chi(D^z)$ is bounded inferiorly. For that purpose, let us introduce the matrix:

$$\check{Q}^z \equiv \left(zI_p - \frac{1}{n} X^T X \right)^{-1} \in \mathcal{M}_n, \quad (22)$$

we have the identities:

$$X^T Q^z = \check{Q}^z X \quad \text{and} \quad \frac{1}{n} \check{Q}^z X X^T = z \check{Q}^z - I_n \quad (23)$$

from which we can deduce:

$$I_n + E^z = I_n + \frac{1}{n} \text{Diag}(X^T Q^z X) = z \check{Q}^z$$

One can then bound in a similar way as in Lemma 11:

Lemma 12. $\frac{I_p}{|z|^2 + (1-\varepsilon)^2} \leq |\check{Q}^z|^2 \leq \frac{I_p}{d(z, [0, 1-\varepsilon])^2}$ and:

$$\frac{|z|}{|z| + 1 - \varepsilon} I_n \leq |I_n + E^z| \leq \left(1 + \frac{1 - \varepsilon}{d(z, [0, 1 - \varepsilon])} \right) I_n$$

Remark 23. Schur formulas 20 imply in particular that $\chi(D^z) = I_n + E$. Then, for any z such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$, one can also bound $|z| \geq O(1)$ and deduce from Lemma 12 that:

$$O(1) \leq \left(1 - \frac{1 - \varepsilon}{|z| + 1 - \varepsilon} \right) I_n \leq \chi(D^z) \leq \left(1 + \frac{1 - \varepsilon}{d(z, [0, 1 - \varepsilon])} \right) I_n \leq O(1)$$

In other words, there is no need to control the modulus of z from above to be able to control $\chi(D^z)$, just the distance to the segment $[0, 1 - \varepsilon]$ is relevant.

Inverting the inequalities around $\chi(D^z)$, taking expectation, and inverting again, one can deduce the similar bound:

$$O(1) \leq \chi(\Delta^z) \leq O(1).$$

PROOF. The inequalities concerning $\check{Q}^z$ are inferred as in the proof of Lemma 11 since under $\mathcal{A}_Q$, $\frac{1}{n} \|X^T X\| = \frac{1}{n} \|X X^T\| \leq 1 - \varepsilon$. One can then deduce that:

$$\frac{|z| I_n}{\sqrt{|z|^2 + (1 - \varepsilon)^2}} \leq |I_n + E| = |z \check{Q}^z| \leq \frac{|z| I_n}{d(z, [0, 1 - \varepsilon])}$$

On the one hand employing the inequality, valid for any $a, b > 0$, $\sqrt{a^2 + b^2} \leq \sqrt{(a + b)^2} = a + b$, one can note that $\frac{|z|}{\sqrt{|z|^2 + (1 - \varepsilon)^2}} \geq \frac{|z|}{|z| + (1 - \varepsilon)}$. On the second hand:

- if $\Re(z) \leq 0$, $|z| = d(z, [1 - \varepsilon])$
- if $\Re(z) \in [0, 1 - \varepsilon]$, $d(z, [1 - \varepsilon]) = \Im(z)$ and $|z| = \sqrt{d(z, [1 - \varepsilon])^2 + \Re(z)^2} \leq d(z, [1 - \varepsilon]) + 1 - \varepsilon$
- if $\Re(z) \geq 1 - \varepsilon$, $|z| \leq |z - 1 + \varepsilon| + 1 - \varepsilon = d(z, [1 - \varepsilon]) + 1 - \varepsilon$,

one can conclude that in all cases, $\frac{|z|}{d(z, [0, 1 - \varepsilon])} \leq 1 + \frac{1 - \varepsilon}{d(z, [0, 1 - \varepsilon])}$, which eventually provides the result of the lemma.

We then have all the element to set:

Lemma 13. *Given $z \in \mathbb{C}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$:*

$$D^z \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n} \quad \text{in } (\mathcal{D}_n(\mathbb{C}), \|\cdot\|_F).$$

In particular, $D^z \in \Delta^z \pm \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ in $(\mathcal{D}_n, \|\cdot\|_F)$ and we can prove:

Proposition 15. *Given $z \in \mathbb{C}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$:*

$$\left\| \tilde{Q}^z(\chi(\Delta^z)) \right\| \leq O(1) \quad \text{and} \quad \left\| \mathbb{E}_{\mathcal{A}_Q} Q^z - \tilde{Q}^z(\chi(\Delta^z)) \right\|_F = O\left(\sqrt{\frac{\log n}{n}}\right).$$

PROOF. Let us note for simplicity $N_{\tilde{Q}} \equiv \|\tilde{Q}^z(\chi(\Delta^z))\|$. With the notation introduced in (21), note that we have to bound $\|\mathbb{E}_{\mathcal{A}_Q} \varepsilon_1\|_F$ and $\|\mathbb{E}_{\mathcal{A}_Q} \varepsilon_2\|_F$. We can already bound thanks to Corollary 7:

$$\|\varepsilon_1\|_F = \left\| \frac{1}{n} \mathbb{E}_{\mathcal{A}_Q} \left[Q^z X \chi(\Delta^z) (D^z - \Delta^z) X^T \tilde{Q}^z(\chi(\Delta^z)) \right] \right\|_F \leq O\left(\sqrt{\frac{\log n}{c}} N_{\tilde{Q}}\right)$$

since $Q^z X \chi(\Delta^z) \propto \mathcal{E}_2 \mid e^{-n}$, $\tilde{Q}^z(\chi(\Delta^z)) X \propto \mathcal{E}_2(N_{\tilde{Q}}) \mid e^{-n}$ and $\|D^z - \Delta^z\|_F \leq O(1/\sqrt{n})$. It is then sufficient to bound for any matrix $A \in \mathcal{M}_p$ satisfying $\|A\|_F \leq 1$ the quantity $\mathbb{E}_{\mathcal{A}_Q} [\text{Tr}(A \varepsilon_2)]$. We can bound thanks to Cauchy-Shwarz inequality that:

$$\begin{aligned} |\mathbb{E}_{\mathcal{A}_Q} [\text{Tr}(A \varepsilon_2)]| &= \sqrt{\frac{1}{n^2} \mathbb{E}_{\mathcal{A}_Q} [\text{Tr}(A Q^z X \chi(\Delta^z)^2 X^T Q^z A^T)]} \\ &\quad \cdot \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{E}_{\mathcal{A}_Q} [\text{Tr}(\tilde{Q}^z(\chi(\Delta^z)) \Sigma_i Q_{-i}^z \Sigma_i Q_{-i}^z \tilde{Q}^z(\chi(\Delta^z)))]} \\ &\leq O\left(\sqrt{\frac{\text{Tr}(A^T A)}{n} \frac{\text{Tr}(\tilde{Q}^z(\chi(\Delta^z))^2)}{n}}\right) \leq O\left(\frac{N_{\tilde{Q}}}{\sqrt{n}}\right) \end{aligned}$$

thanks to the bounds provided by our assumptions, Lemma 11 and Remark 23. Putting the bounds on $\|\mathbb{E}_{\mathcal{A}_Q} \varepsilon_1\|$ and $\|\mathbb{E}_{\mathcal{A}_Q} \varepsilon_2\|$ together, we obtain:

$$\left\| \mathbb{E}_{\mathcal{A}_Q} Q^z - \tilde{Q}^z(\chi(\Delta^z)) \right\|_F \leq O\left(N_{\tilde{Q}} \sqrt{\frac{\log n}{n}}\right)$$

So, in particular:

$$N_{\tilde{Q}} \equiv \left\| \tilde{Q}^z(\chi(\Delta^z)) \right\| \leq \left\| \mathbb{E}_{\mathcal{A}_Q} Q^z \right\| + O \left(N_{\tilde{Q}} \sqrt{\frac{\log n}{n}} \right),$$

which implies that $N_{\tilde{Q}} \leq O(1)$ as $\left\| \mathbb{E}_{\mathcal{A}_Q} [Q^z] \right\|$. We obtain then directly the second bound of the Proposition.

3.3. A second deterministic equivalent

With Proposition 15, the problem becomes much simpler because, we initially had to estimate the expectation of the whole matrix Q^z and now, we just need to approach the expectation of the diagonal matrix $D^z = \frac{1}{n} x_i Q_{-i}^z x_i$. One is tempted to introduce from the pseudo identity:

$$\Delta_i^z \approx \mathbb{E}_{\mathcal{A}_Q} \left[\frac{1}{n} x_i^T Q_{-i}^z x_i \right] \approx \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}_{-i}^z(\chi(\Delta^z)) \right) \approx \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(\chi(\Delta^z)) \right),$$

(where, naturally, $\tilde{Q}_{-i}^z(\Gamma) = (zI_p - \frac{1}{n} \sum_{\substack{1 \leq j \leq n \\ j \neq i}} \Gamma_j \Sigma_j)^{-1}$) a fixed point equation whose solution $\Lambda \in \mathcal{D}_n(\mathbb{C})$ is a natural estimate for Δ .

The two next proposition are very strong results that will be demonstrated after the multiple inferences of the next fifteen pages. Before starting this long proof we provide Theorem 10 that gives us a computable deterministic equivalent for Q^z (which was the main objective of this section once we knew that Q^z was concentrated).

Proposition 16 (Definition of Λ^z). *For any $z \in \mathbb{C} \setminus [0, 1 - \varepsilon]$, applying χ entry-wise, the system of equations :*

$$L = \text{Diag}_{1 \leq i \leq n} \left(\frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(\chi(L)) \right) \right), \quad L \in \mathcal{D}_n(\mathbb{C})$$

admits a unique solution in $\mathcal{D}_n(\mathbb{C})$ that we note Λ^z

Proposition 17. *For any $z \in \mathbb{C}$ such that $|z| \leq O(1)$ and $d(z, [0, 1 - \varepsilon]) \geq O(1)$:*

$$\|\Delta^z - \Lambda^z\| \leq O \left(\frac{\sqrt{\log n}}{n} \right).$$

Remark 24. *The bound $\|\Delta_i^z - \Lambda_i^z\| \leq O \left(\frac{\sqrt{\log n}}{n} \right)$, valid for all $i \in [n]$ might look strong since Δ_i^z is of order $O(1)$. However a lower speed convergence like $O \left(\sqrt{\frac{\log n}{n}} \right)$ is trivial to obtain since we can show for instance that if we introduce $\Gamma = \text{Diag}_{1 \leq i \leq n} (\frac{1}{zn} \text{Tr}(\Sigma_i))$, we can bound for any $i \in [n]$:*

$$\begin{aligned} |\Delta_i - \Gamma_i| &= \left| \frac{1}{n} \text{Tr} \left(\sigma_i(\mathbb{E}_{\mathcal{A}_Q} [Q_{-i}^z] - \frac{1}{z} I_p) \right) \right| \\ &= \left| \mathbb{E}_{\mathcal{A}_Q} \left[\frac{1}{zn^2} \text{Tr} (\Sigma_i Q_{-i}^z X X^T) \right] \right| \leq O \left(\sqrt{\frac{\log n}{n}} \right), \end{aligned}$$

thanks to Proposition ?? applied to the matrices Σ_i , $Q_{-i}^z X$, I_n and X .

Anyway, such a speed of order $O\left(\frac{\sqrt{\log n}}{n}\right)$ is necessary to be able to show the validity of the second deterministic equivalent of Q^z .

Theorem 10 (Computable deterministic equivalent of the resolvent).
For any $z \in \mathbb{C}$ such that $|z| \leq O(1)$ and $d(z, [0, 1 - \varepsilon]) \geq O(1)$, $\|\tilde{Q}^z(\chi(\Lambda^z))\| \leq O(1)$ and:

$$Q^z \in \tilde{Q}^z(\chi(\Lambda^z)) \pm \mathcal{E}_2 \left(\sqrt{\frac{\log n}{n}} \right) \mid e^{-n} \quad \text{in} \quad (\mathcal{M}_p, \|\cdot\|_F)$$

PROOF. We already know from Propositions 14 and 15 that:

$$Q^z \in \tilde{Q}^z(\chi(\Delta^z)) \pm \mathcal{E}_2 \left(\sqrt{\frac{\log n}{n}} \right) \mid e^{-n} \quad \text{in} \quad (\mathcal{M}_p, \|\cdot\|_F),$$

and since for an $A \in \mathcal{M}_p$, $\text{Tr}(A) \leq \sqrt{p}\|A\|_F$ and Proposition 17 imply that:

$$\begin{aligned} & \left| \text{Tr} \left(A(\tilde{Q}^z(\chi(\Delta^z)) - \tilde{Q}^z(\chi(\Lambda^z))) \right) \right| \\ & \leq \left| \text{Tr} \left(A\tilde{Q}^z(\chi(\Delta^z)) \left(\frac{1}{n} \sum_{i=1}^n (\Lambda_i^z - \Delta_i^z) \Sigma_i \right) \tilde{Q}^z(\chi(\Lambda^z)) \right) \right| \\ & \leq \|\Lambda^z - \Delta^z\| \left| \frac{\|\tilde{Q}^z(\chi(\Lambda^z))\| \text{Tr}(A)}{d(z, [0, 1 - \varepsilon])} \right| \leq O \left(\|\tilde{Q}^z(\chi(\Lambda^z))\| \sqrt{\frac{\log n}{n}} \right). \end{aligned}$$

We can then deduce that:

$$\begin{aligned} \|\tilde{Q}^z(\chi(\Lambda^z))\| & \leq \|\tilde{Q}^z(\chi(\Delta^z))\| + \|\tilde{Q}^z(\chi(\Delta^z)) - \tilde{Q}^z(\chi(\Lambda^z))\|_F \\ & \leq O(1) + O \left(\|\tilde{Q}^z(\chi(\Lambda^z))\| \sqrt{\frac{\log n}{n}} \right) \end{aligned}$$

Therefore $\|\tilde{Q}^z(\chi(\Lambda^z))\| \leq O(1)$ and $\|\tilde{Q}^z(\chi(\Delta^z)) - \tilde{Q}^z(\chi(\Lambda^z))\|_F \leq O(\sqrt{\log n/n})$.

Propositions 16 and 17 are quite hard to establish, we will need for that a new notation:

$$\mathcal{I}^z(D) \equiv \text{Diag} \left(\frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(D) \right) \right)_{1 \leq i \leq n}$$

for a diagonal matrix $D \in \mathcal{D}_n(\mathbb{C})$ such that $\tilde{Q}^z(D)$ is well defined. We will then follow the strategy:

1. Show that $\Lambda^z = I^z \circ \chi(\Lambda^z)$ admits a unique solution for $z \in \mathbb{C}^-$, where:

$$\mathbb{C}^- \equiv \{z \in \mathbb{C}, \Im(z), \Re(z) < 0\},$$

2. Show that for all $i \in [n]$, $z \mapsto \Lambda_i^z$ is analytical on an open set of $\mathbb{C}_-^+$ like $z \mapsto \Delta_i^z$,
3. Justify the possibility of an extension of the regularity of Λ^z on $\mathbb{C}_-^+$ to the whole set of definition of $z \mapsto \Delta^z$ (i.e $\mathbb{C} \setminus [0, 1 - \varepsilon]$), while keeping, at the same time, a control on the bound on $\|\Lambda^z - \Delta^z\| \leq O(\sqrt{\log n}/n)$.
4. Actual proof of Propositions 16 and 17.

Each one of these steps will form our four next subsections.

3.4. Existence and uniqueness of Λ^z when $z \in \mathbb{C}_-^+$

To show the existence and uniqueness of Λ^z , let us adapt tools already introduced in Louart and Couillet (2020) that rely on the introduction of a semi-metric d_s called “the stable semi metric” and defined²⁴ for any $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$ (where $\mathbb{C}^+ \equiv \{z \in \mathbb{C}, \text{Im}(z) > 0\}$) 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 metric is introduced to set Banach-like fixed point theorems. For that purpose, we introduce:

Definition 10. The “stable class”, denoted $\mathcal{C}_s(\mathcal{D}_n(\mathbb{C}^+))$ is defined as the class of functions $f : \mathcal{D}_n(\mathbb{C}^+) \rightarrow \mathcal{D}_n(\mathbb{C}^+)$, 1-Lipschitz for the semi-metric d_s ; i.e. satisfying for all $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$:

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

Given $f \in \mathcal{C}_s(\mathcal{D}_n(\mathbb{C}^+))$, if there exists a constant $\varepsilon > 0$ such that:

$$\forall D, D' \in \mathcal{D}_n(\mathbb{C}^+), d_s(f(D), f(D')) \leq (1 - \varepsilon)d_s(D, D'),$$

we say that f is contracting.

The stable class owes its name to an important number of stability properties that we list in the next proposition:

Proposition 18. Given $f, g \in \mathcal{C}_s(\mathcal{D}_n(\mathbb{C}^+))$, $\alpha \in \mathbb{R}_*^+$ and $h : \mathcal{D}_n(\mathbb{C}^+ \cup \mathbb{R}) \rightarrow \mathcal{D}_n(\mathbb{C}^+ \cup \mathbb{R})$ 1-Lipschitz for the spectral norm:

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

are all in the stable class.

²⁴In Louart and Couillet (2020), d_s is only defined on $\mathcal{D}_n(\mathbb{R}_+)$ and the denominator appearing in the definition is then $\sqrt{D_i D'_i}$ instead of $\sqrt{\Im(D_i)\Im(D'_i)}$. The present adaptation does not change the fundamental properties of the semi-metric; the only objective with this new choice is to be able to set that $\mathcal{I}^z$ is contractive for this semi-metric.

We then deduce easily from these properties that:

Corollary 10. $\chi(\mathbb{C}^+) \subset \mathbb{C}^+$ and $\chi \in \mathcal{C}_s(\mathbb{C}^+)$.

We can now present our fixed point theorem that has been demonstrated once again in Louart and Couillet (2020):

Theorem 11. *Given a mapping $f : \mathcal{D}_n(\mathbb{C}^+) \rightarrow \mathcal{D}_n(\mathbb{C}^+)$, contracting for the stable semi-metric d_s and such that its imaginary part is bounded from above (in $\mathcal{D}_n^+$), there exists a unique fixed point $\Delta^* \in \mathcal{D}_n^+$ satisfying $\Delta^* = f(\Delta^*)$.*

The idea is then to employ this theorem to the mapping $\chi \circ \mathcal{I}_z$. Let us first find a good choice of z for $\mathcal{I}_z \circ \chi$ to be stable on a subset of $\mathcal{D}_n(\mathbb{C}^+)$. We need for that a first lemma that will, in passing, help us verifying the hypothesis $\mathcal{I}_z \circ \chi$ bounded from above in Theorem 11.

Lemma 14. *Given $z \in \mathbb{C}_-^+ \equiv -\mathbb{R}_+ + i\mathbb{R}_+$:*

$$\Im(\chi(z)) \in (0, 1) \quad \text{and} \quad \Re(\chi(z)) \in (0, 1).$$

PROOF. Given $z \in \mathbb{C}_-^+$, note that:

$$\Re(\chi(z)) = \frac{1 - \Re(z)}{|1 - z|^2} > 0 \quad \text{and} \quad \Im(\chi(z)) = \frac{\Im(z)}{|1 - z|^2} > 0$$

Now, on the one hand $1 - \Re(z) > 1$, thus $\Re(\chi(z)) = \frac{1 - \Re(z)}{(1 - \Re(z))^2 + \Im(z)^2} < \frac{1}{1 - \Re(z)} < 1$ and on the other hand, if $\Im(z) < 1$, $\Im(\chi(z)) < \frac{1}{(1 - \Re(z))^2} < 1$ and if $\Im(z) \geq 1$, $\Im(\chi(z)) < \frac{1}{\Im(z)} \leq 1$.

The idea is then to take $z \in \mathbb{C}_-^+$ for $\mathcal{I}^z \circ \chi$ to be stable on $\mathcal{D}_n(\mathbb{C}_-^+)$.

Lemma 15. *If $z \in \mathbb{C}^-$, then $\mathcal{I}_z(\mathcal{D}_n(\mathbb{C}^+)) \subset \mathcal{D}_n(\mathbb{C}^+)$ and if $z \in \mathbb{C}_-^+$:*

$$\mathcal{I}_z \circ \chi(\mathcal{D}_n(\mathbb{C}_-^+)) \subset \mathcal{D}_n(\mathbb{C}_-^+).$$

PROOF. Considering $z \in \mathbb{C}^-$, for any $D \in \mathcal{D}_n(\mathbb{C}^+)$ and $i \in [n]$:

$$\Im(\mathcal{I}^z(D))_i = \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(D) (-\Im(z)I_p + \Sigma_{\Im(D_j)}) \tilde{Q}^z(D) \right) > 0, \quad (24)$$

since for all $j \in [n]$, $\Sigma_i^{1/2} \tilde{Q}^z(D) \Sigma_j \tilde{Q}^z(D) \Sigma_i^{1/2}$ is a nonnegative symmetric matrix (recall that for all $\Gamma \in \mathcal{D}_n$, $\Sigma_\Gamma \equiv \frac{1}{n} \sum_{i=1}^n \Gamma_i \Sigma_i$).

Besides, since $\chi(\mathcal{D}_n(\mathbb{C}^+)) \subset \mathcal{D}_n(\mathbb{C}^+)$ (see Corollary 10), we already know that for any $D \in \mathcal{D}_n(\mathbb{C}_-^+)$ and $z \in \mathbb{C}_-^+$, $\Im(\mathcal{I}^z \circ \chi(D)) > 0$. In addition, the bound $\Re(\chi(D)) \in \mathcal{D}_n(\mathbb{R}^+)$ implies that:

$$\Re(\mathcal{I}^z(\chi(D)))_i = \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(\chi(D)) (\Re(z)I_p - \Sigma_{\Re(\chi(D))}) \tilde{Q}^z(\chi(D)) \right) < 0.$$

Next proposition provides us the contractive character of $\mathcal{I}^z$ (that will imply the contractive character of $\mathcal{I}^z \circ \chi$).

Proposition 19. *For any $z \in \mathbb{C}^-$, the mapping $\mathcal{I}_z$ is 1-Lipschitz for the semi-metric d_s and satisfies for any $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$:*

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

where for any $w \in \mathbb{C}^-$ and $D \in \mathcal{D}_n(\mathbb{C}^+)$:

$$\phi(w, D) = \inf_{1 \leq i \leq n} \frac{|\Im(w)|}{n} \frac{\text{Tr} \left(\Sigma_i |\tilde{Q}^w(D)|^2 \right)}{\Im(\mathcal{I}^w(D))_i} \in (0, 1).$$

(Recall that $|\tilde{Q}^z(D)|^2 = \tilde{Q}^z(D) \overline{\tilde{Q}^z(D)} \in \mathcal{S}_p^+(\mathbb{R})$).

PROOF. Considering $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$:

$$\mathcal{I}_z(D)_i - \mathcal{I}_z(D')_i \leq \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^z(D) \left(\frac{1}{n} \sum_{j=1}^n \frac{D_j - D'_j}{\sqrt{\Im(D_j) \Im(D'_j)}} \sqrt{\Im(D_j) \Im(D'_j)} \Sigma_j \right) \tilde{Q}^z(D') \right)$$

We can then bound thanks to Cauchy-Schwarz inequality:

$$\begin{aligned} d_s(\mathcal{I}_z(D), \mathcal{I}_z(D')) &\leq d_s(D, D') \sup_{1 \leq i \leq n} \frac{1}{n} \sqrt{\frac{1}{\Im(\mathcal{I}^z(D))_i} \text{Tr} \left(\Sigma_i \tilde{Q}^z(D) \Sigma_{\Im(D)} \tilde{\bar{Q}}^z(D) \right)} \\ &\quad \cdot \sup_{1 \leq i \leq n} \sqrt{\frac{1}{\Im(\mathcal{I}^z(D'))_i} \text{Tr} \left(\Sigma_i \tilde{Q}^z(D') \Sigma_{\Im(D')} \tilde{\bar{Q}}^z(D') \right)} \end{aligned}$$

We then conclude with the useful identity obtained from (24):

$$0 \leq \text{Tr} \left(\Sigma_i \tilde{Q}^z(D) \Sigma_{\Im(D)} \tilde{\bar{Q}}^z(D) \right) = \Im(\mathcal{I}^z(D))_i + \frac{\Im(z)}{n} \text{Tr}(\Sigma_i \tilde{Q}^z(D) \tilde{\bar{Q}}^z(D)).$$

To be able to employ Theorem 11 and define correctly Λ^z as the only matrix satisfying $\Lambda^z = \mathcal{I}^z(\chi(\Lambda^z))$, one thus needs to show that $\mathcal{I}^z \circ \chi$ is contractive for the semi-metric d_s ; i.e. to bound uniformly on $\mathcal{D}_n(\mathbb{C}^+)$ $\|Q_{-i}\|^2$ from below and $\Im(\mathcal{I}^z(D))_i$ from above (we already know from Corollary 10 that for any $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$, $d_s(\chi(D), \chi(D')) \leq d_s(D, D')$). Those bounds are not easy to obtain uniformly on $\mathcal{D}_n(\mathbb{C}^+)$, but if we condition z to belong to $\mathbb{C}^+$, then the bounds can be obtained uniformly on $\mathcal{D}_n(\mathbb{C}^+)$ which is a stable set of $\mathcal{I}^z \circ \chi$ as we saw in Lemma 15.

Corollary 11. *Given a complex number $z \in \mathbb{C}^+$, there exists a unique diagonal matrix $\Lambda^z \in \mathcal{D}_n(\mathbb{C}^+)$ satisfying $\Lambda^z = \mathcal{I}^z(\Lambda^z)$.*

PROOF. We know from Lemma 14 that $\chi(\mathcal{D}_n(\mathbb{C}^+)) \subset \mathcal{D}_n(\mathbb{C}_{(0,1)}^{(0,1)})$ thus we are going to bound, for any $i \in [n]$ and $D \in \mathcal{D}_n(\mathbb{C}_{(0,1)}^{(0,1)})$, $\frac{1}{n} \text{Tr}(\Sigma_i |\tilde{Q}^z(D)|^2)$ from below and $\Im(\mathcal{I}^z(D))_i$ from above. Following the approach conducted in the proof of Lemma 12, we note $A = \Re((Q_{-i}^z)^{-1})$ and $B = \Im((Q_{-i}^z)^{-1})$, then $(|Q_{-i}^z|^2)^{-1} = A^2 + B^2$, and since

- $0 < \Re(D_j) < 1$
- $0 < \Im(D_j) < 1$
- $\Re(z) < 0$
- $\Im(z) < 0$

- $\|\frac{1}{n} \sum_{i=1}^n \Sigma_i\| \leq 1 - 2\varepsilon$ (see Assumption 5),

one can bound:

$$\begin{cases} \Re(z)^2 I_p \leq A^2 = \left(\Re(z) I_p - \frac{1}{n} \sum_{j=1}^n \Re(D_j) \Sigma_j \right)^2 \leq (|\Re(z)| + (1 - \varepsilon))^2 I_p \\ \Im(z)^2 I_p \leq B^2 = \left(\Im(z) I_p - \frac{1}{n} \sum_{j=1}^n \Im(D_j) \Sigma_j \right)^2 \leq (|\Im(z)| + (1 - \varepsilon))^2 I_p \end{cases}$$

That gives us the bounds (in the set of symmetric matrices) for any $D \in \mathcal{D}_n(\mathbb{C}_{(0,1)}^{(0,1)})$:

$$\frac{I_p}{|z|^2 + 2|\Re(z)| + 2|\Im(z)| + 2} \leq |\tilde{Q}_{-i}^z(D)|^2 = (A^2 + B^2)^{-1} \leq \frac{I_p}{|z|^2}. \quad (25)$$

One can thus bound from below, uniformly on $\mathcal{D}_n(\mathbb{C}_{[0,1]}^{[0,1]})$, the functional $\phi(z, D)$ presented in Proposition 19 that gives us the contractive character of $\mathcal{I}^z \circ \chi$. In addition, we know from Lemma 14 that this mapping is bounded (with a bound depending only on z), we can thus employ Theorem 11 to set the existence and uniqueness of Λ^z .

Remark 25. *Inequality (25) is obtain in the case where $z \in \mathbb{C}^-$ and $D \in \mathcal{D}_n(\mathbb{C}_{(0,1)}^{(0,1)})$, when we relax those hypothesis to assume only $z \in \mathbb{C}^-$ and $D \in \mathcal{D}_n(\mathbb{C}^+)$, one can still bound:*

$$\frac{I_p}{|z| + \|D\|} \leq |\tilde{Q}_{-i}^z(D)| \leq \frac{I_p}{|\Im(z)|}. \quad (26)$$

The bound $\frac{I_p}{|\Im(z)|^2}$ (that tends to ∞ when $\Im(z) \rightarrow 0$) can be improved in the case $D = \chi(\Delta^z)$, we can then set thanks to Proposition 15 that, for any $z \in \mathbb{C}^-$ satisfying $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $|z| \leq O(1)$:

$$O(1) \leq |\tilde{Q}_{-i}^z(\chi(\Delta^z))| \leq O(1). \quad (27)$$

since $\|\chi(\Delta^z)\| \leq O(1)$ thanks to Lemma 12 (more precisely thanks to Remark 23).

3.5. Analyticity of $z \mapsto \Lambda^z$

We then want to show that for all $i \in [n]$ the mappings $z \mapsto \Delta_i^z$ and $z \mapsto \Lambda_i^z$ are analytic on an open set of $\mathbb{C}^-$ (where Λ^z is well defined) to be able to employ afterwards complex analysis inferences. It is straightforward to see that $z \mapsto \Delta^z$ is differentiable thanks to its explicit form. Let us recall the existence of a parameter $\varepsilon > 0$, introduced in Assumption 5 such that the event $\mathcal{A}_Q \equiv \{\|XX^T\| \leq 1 - \varepsilon\}$ described in (19) satisfies $\mathbb{P}(\mathcal{A}_Q^c) \leq Ce^{-cn}$, for some constants $C, c > 0$ and let us set:

Proposition 20. *For all $i \in [n]$, the mapping $z \mapsto \Delta_i^z$ is analytical on $\mathbb{C} \setminus [0, 1 - \varepsilon]$*

PROOF. Recall that for all $z \in \mathbb{C} \setminus [0, 1 - \varepsilon]$,

$$\Delta^z = \text{Diag}_{1 \leq i \leq n} \left(\mathbb{E}_{\mathcal{A}_Q} \left[\frac{1}{n} x_i^T \left(z I_p - \frac{1}{n} X_{-i} X_{-i}^T \right)^{-1} x_i \right] \right),$$

under $\mathcal{A}_Q$, we can differentiate:

$$\frac{\partial \Delta^z}{\partial z} = -z \text{Diag}_{1 \leq i \leq n} \left(\mathbb{E}_{\mathcal{A}_Q} \left[\frac{1}{n} x_i^T (Q_{-i}^z)^2 x_i \right] \right).$$

It is more difficult to differentiate $z \mapsto \Lambda^z$ (on $\mathbb{C}^-$, where it is defined for the moment) with its implicit formulation. Let us first show its continuity. For that purpose we introduce a result from Louart and Couillet (2020) that allows to bound variations around a fixed point of a stable mapping.

Proposition 21. *Let us consider a set of indexes Θ and a family of mappings of $\mathcal{D}_n(\mathbb{C}^+)$, $(f_t)_{t \in \Theta}$, each f_t being λ_t -Lipschitz for the semi-metric d_s and admitting the fixed point $D_t = f_t(D_t)$ and a family of diagonal matrices Γ_t ²⁵. If one assumes that:*

1. *there exist a constants $C > 0$ such that for all $t \in \Theta$, $\forall D \in \mathcal{D}_n(\mathbb{C}^+)$:*

$$\frac{\sup_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i}{\inf_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i}, \frac{\sup_{1 \leq i \leq n} \Im(D_t)_i}{\inf_{1 \leq i \leq n} \Im(D_t)_i} \leq C$$

2. *there exists a constant $\lambda' \in (0, 1)$ such that:*

$$\forall t \in \Theta: \quad \lambda_t \left(1 + \left\| \sqrt{\frac{|\Im(f_t(\Gamma_t)) - \Im(\Gamma_t)|}{\Im(\Gamma_t)}} \right\| \right) \leq \lambda' < 1,$$

then there exists a constant $K > 0$ such that for all $t \in \Theta$:

$$\|D_t - \Gamma_t\| \leq K \|f_t(\Gamma_t) - \Gamma_t\|.$$

PROOF. Let us first bound:

$$\begin{aligned} \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\| &\leq d_s(f_t(D_t), f_t(\Gamma_t)) + \left\| \frac{f_t(\Gamma_t) - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\| \\ &\leq \lambda \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(\Gamma_t)}} \right\| + \left\| \frac{f_t(\Gamma_t) - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\|. \end{aligned}$$

²⁵Proposition 23 uses a simpler setting where Γ_t is independent with t , then the existence of two constants $c, C > 0$ such that for all $t \in \Theta$, $\frac{\sup_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i}{\inf_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i} \leq C$ is obvious

(one must be careful here that the stable semi-metric does not satisfy the triangular inequality). Besides, since for any positive number $a, b > 0$, $\sqrt{a} - \sqrt{b} \leq \sqrt{|a - b|}$, one can bound:

$$\begin{aligned} \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(\Gamma_t)}} \right\| &\leq \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)}} \left(\frac{\sqrt{\Im(f_t(\Gamma_t))} - \sqrt{\Im(\Gamma_t)}}{\sqrt{\Im(\Gamma_t)}\sqrt{\Im(f_t(\Gamma_t))}} \right) \right\| + \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\| \\ &\leq \left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\| \left(1 + \left\| \sqrt{\frac{|\Im(f_t(\Gamma_t)) - \Im(\Gamma_t)|}{\Im(\Gamma_t)}} \right\| \right). \end{aligned}$$

Thus, by hypothesis, we have the inequality:

$$\left\| \frac{D_t - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\| \leq \lambda' \left\| \frac{f_t(\Gamma_t) - \Gamma_t}{\sqrt{\Im(D_t)\Im(f_t(\Gamma_t))}} \right\|.$$

(with $\lambda' < 1$). Thus, since $\frac{\sup_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i}{\inf_{1 \leq i \leq n} \Im(f_t(\Gamma_t))_i}$ and $\frac{\sup_{1 \leq i \leq n} \Im(D_t)_i}{\inf_{1 \leq i \leq n} \Im(D_t)_i}$ are bounded independently with $t \in \Theta$, we obtain the existence of a constant $K'' > 0$ such that:

$$\|D_t - \Gamma_t\| \leq K'' \|f_t(\Gamma_t) - \Gamma_t\|$$

The result of Proposition 21 can be obtained with simpler hypothesis when one supposes that Γ_t is constant and equals $\Gamma \in \mathcal{D}_n(\mathbb{C}^+)$.

Proposition 22. *Considering $\Theta \subset \mathbb{C}$, an open set containing 0 and a family of mappings of $\mathcal{D}_n(\mathbb{C}^+)$, $(f_t)_{t \in \Theta}$, each f_t being $1 - \nu$ -Lipschitz for the semi-metric d_s with $\nu > 0$ and admitting the fixed point $D_t = f_t(D_t)$, if one assumes that there exists a diagonal matrix $\Gamma \in \mathcal{D}_n(\mathbb{C}^+)$ such that there exists a for all $C > 0$ a subset Θ_C open in $\mathbb{C}$ and containing 0 such that:*

$$\forall t \in \Theta' : \|f_t(\Gamma) - \Gamma\| \leq C,$$

then there exists a constant $K > 0$ and an open set $U \subset \mathbb{C}$, containing 0 such that:

$$\forall t \in U : \|D_t - \Gamma\| \leq K \|f_t(\Gamma) - \Gamma\|$$

PROOF. We know that there exists $\Theta' \subset \Theta$ such that $0 \in \Theta'$ and for all $t \in \Theta'$:

$$\|f_t(\Gamma) - \Gamma\| \leq \left(\frac{1}{1 - \frac{\nu}{2}} - 1 \right)^2 \sqrt{\inf_{1 \leq i \leq n} \Im(\Gamma)_i}.$$

then, for any $t \in \Theta'$:

$$(1 - \nu) \left(1 + \left\| \sqrt{\frac{|\Im(f_t(\Gamma_t)) - \Im(\Gamma_t)|}{\Im(\Gamma_t)}} \right\| \right) \leq \frac{1 - \nu}{1 - \frac{\nu}{2}} \equiv \lambda' < 1$$

Then, we can deduce as in the proof of Proposition 21 that:

$$\left\| \frac{D_t - \Gamma}{\sqrt{\Im(D_t)}} \right\| \leq C\lambda' \left\| \frac{f_t(\Gamma) - \Gamma}{\sqrt{\Im(D_t)}} \right\|, \quad \text{with } C = \sqrt{\frac{\sup_{1 \leq i \leq n} \Im(f_t(\Gamma))}{\inf_{1 \leq i \leq n} \Im(f_t(\Gamma))}}$$

Now, if we introduce the indexes $i_{\inf}, i_{\sup} \in [n]$ satisfying:

$$\Im(D_t)_{i_{\inf}} = \inf_{1 \leq i \leq n} \Im(D_t)_i \text{ and } \Im(D_t)_{i_{\sup}} = \sup_{1 \leq i \leq n} \Im(D_t)_i,$$

we see that:

$$\frac{|\Im(D_t)_{i_{\inf}} - \Im(\Gamma)_{i_{\inf}}|}{\sqrt{\Im(D_t)_{i_{\inf}}}} \leq C\lambda' \frac{\|f_t(\Gamma) - \Gamma\|}{\sqrt{\Im(D_t)_{i_{\inf}}}},$$

from which we deduce that $\Im(D_t)_{i_{\inf}} \geq \Im(\Gamma)_{i_{\inf}} - C\lambda' \|f_t(\Gamma) - \Gamma\| \geq \frac{\Im(\Gamma)_{i_{\inf}}}{2}$, for any $t \in \Theta''$, where $\Theta'' \ni 0$ is a well chosen open subset of Θ' . Besides :

$$\sqrt{\Im(D_t)_{i_{\sup}}} \leq \frac{1}{\sqrt{\Im(D_t)_{i_{\inf}}}} (\Im(\Gamma)_{i_{\sup}} + C\lambda' \|f_t(\Gamma) - \Gamma\|) \leq \frac{2\Im(\Gamma)_{i_{\sup}}}{\sqrt{\Im(D_t)_{i_{\inf}}}},$$

for all $t \in U$, where $U \ni 0$ is a well chosen open subset of Θ'' . Setting $K = C\lambda' \frac{4\Im(\Gamma)_{i_{\sup}}}{\sqrt{\Im(D_t)_{i_{\inf}}}\Im(\Gamma)_{i_{\inf}}}$, we retrieve the result of the proposition.

Proposition 23. *The mapping $z \mapsto \Lambda_i^z$ is continuous on $\mathbb{C}^-$.*

PROOF. Given $z \in \mathbb{C}^-$, let us verify the assumption of Proposition 22 for $\Theta \subset \mathbb{C}$ being an open set of $\mathbb{C}$ containing 0 and such that $z + \Theta \in \mathbb{C}^- \cap \{\Im(w) \leq 2\Im(z)\}$, for all $t \in \Theta$, $f_t = \mathcal{I}^{z+t} \circ \chi$ (and $D_t = \Lambda^{z+t}$) and $\Gamma = \Lambda^z$. We already know from Proposition 19 that f_t are all contracting for the stable semi-metric with a Lipschitz parameter $\lambda < 1$ that can be chosen independent from t for Θ small enough. Let us express for any $t \in \Theta$ and any $i \in [n]$:

$$\begin{aligned} f_t(\Gamma_t) - \Gamma_t &= \frac{1}{n} \text{Tr}(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^z))) - \frac{1}{n} \text{Tr}(\Sigma_i \tilde{Q}^z(\chi(\Lambda^z))) \\ &= \frac{t}{n} \text{Tr}(\Sigma_i \tilde{Q}^z(\chi(\Lambda^z)) \tilde{Q}^{z+t}(\chi(\Lambda^z))) \end{aligned} \quad (28)$$

Thus, thanks to Lemma 12 and bounds on $\|\tilde{Q}^z(\chi(\Lambda^z))\|^2$ provided by (25), there exists a constant $C > 0$ such that for all $t \in \Theta$:

$$\|f_t(\Gamma_t) - \Gamma_t\| \leq \frac{|t|(1-\varepsilon)}{|z||z+t|} \leq C|t| \xrightarrow{t \rightarrow 0} 0.$$

Therefore, the assumptions of Proposition 22 are thus satisfied, and there exists a constant $K > 0$ such that:

$$\forall t \in \Theta : \|\Lambda^{z+t} - \Lambda^z\| \leq K|t|,$$

that directly implies that $z \mapsto \Lambda^z$ is continuous on $\mathbb{C}^-$.

Proposition 24. *The mapping $z \mapsto \Lambda_i^z$ is analytic on a set $\mathbb{C}^- \cap \{\Im(z) \leq -\nu\}$ for a given $\nu > 0$ satisfying $\nu \leq O(1)$.*

PROOF. Employing again the notation $f_t = \mathcal{I}^{z+t} \circ \chi$, we can decompose:

$$\frac{1}{t} (\Lambda^{z+t} - \Lambda^z) = \frac{1}{t} (f_t(\Lambda^{z+t}) - f_t(\Lambda^z) + f_t(\Lambda^z) - f_0(\Lambda^z))$$

we recognize the term provided in (28), we are thus left to expressing:

$$f_t(\Lambda^{z+t}) - f_t(\Lambda^z) = \text{Diag}_{1 \leq i \leq n} \left(\frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) A_i \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right) \right)$$

with the notation, $\forall i \in [n]$:

$$A_i = \frac{1}{n} \sum_{j=1}^n \frac{\Lambda_j^{z+t} - \Lambda_j^z}{(1 - \Lambda_j^{z+t})(1 - \Lambda_j^z)} \Sigma_j.$$

Thanks to (28), we can compute for all $k \in \mathbb{N}$:

$$\begin{aligned} & \frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) A_i \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right) \\ &= \frac{1}{n^2} \sum_{j=1}^n \frac{\text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) \Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right)}{(1 - \Lambda_j^{z+t})(1 - \Lambda_j^z)} (f_t(\Lambda_j^{z+t}) - f_t(\Lambda_j^z) + f_t(\Lambda_j^z) - f_0(\Lambda_j^z)) \\ &= \frac{1}{n^2} \sum_{j=1}^n \chi(\Lambda_j^{z+t}) \cdot \chi(\Lambda_j^z) \cdot \text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) \Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right) \\ & \quad \cdot \left(\frac{1}{n} \text{Tr} \left(\Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) A_j \tilde{Q}^z(\chi(\Lambda^z)) \right) + \frac{t}{n} \text{Tr} \left(\Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^z)) \tilde{Q}^z(\chi(\Lambda^z)) \right) \right) \end{aligned}$$

Now, if introduce the vectors:

$$\begin{aligned} a(t) &= \left(\frac{1}{n} \text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) A_i \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right) \right)_{1 \leq i \leq n} \in \mathbb{R}^n, \\ b(t) &= \left(\frac{1}{n} \text{Tr} \left(\Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^z)) \tilde{Q}^z(\chi(\Lambda^z)) \right) \right)_{1 \leq j \leq n} \in \mathbb{R}^n \end{aligned}$$

and the matrices:

$$\Psi(t) = \left(\frac{1}{n^2} \chi(\Lambda_j^{z+t}) \cdot \chi(\Lambda_j^z) \cdot \text{Tr} \left(\Sigma_i \tilde{Q}^{z+t}(\chi(\Lambda^{z+t})) \Sigma_j \tilde{Q}^{z+t}(\chi(\Lambda^z)) \right) \right)_{1 \leq i, j \leq n} \in \mathcal{M}_n$$

We have the equation:

$$a(t) = \Psi(t)(a(t) + tb(t))$$

To be able to invert $I_n - \Psi(t)$ one then needs to assume that the imaginary part of $z \in \mathbb{C}^-$ is sufficiently big since the inequality $\|\tilde{Q}^z(D)\|^2 \leq \frac{1}{|z|^2}$ given in the proof of Corollary 11 allows us to bound for any $i, j \in [n]$:

$$\|\Psi(t)\| \leq n \sup_{1 \leq i, j \leq n} |[\Psi(t)]_{i,j}| \leq \frac{1}{n} \frac{\text{Tr}(\Sigma_i \Sigma_j)}{|z|^2}$$

We have then the identity

$$\frac{1}{t}a(t) = (I_n - \Psi(t))^{-1}\Psi(t)b(t) = (I_n - \Psi(t))^{-1}b(t) - b(t),$$

and by continuity of $z \mapsto \Lambda^z$ (see Proposition 23), letting t tend to 0, one obtains:

$$\frac{\partial \Lambda^z}{\partial z} = \text{Diag}(a'(0) + b(0)) = \text{Diag}((I_n - \Psi(0))^{-1}b(0))$$

3.6. Generalization of the definition and the analyticity of $z \mapsto \Lambda^z$ on the whole set $\mathbb{C} \setminus [0, 1 - \varepsilon]$ and convergence of Λ^z towards Δ^z

The goal of this subsection is to end the proofs of Propositions 16 and 17. The two demonstrations are conducted at the same time in a quite elaborate way, therefore, we first present a sketch of proof and some useful lemmas before giving the rigorous proof. Globally the idea is to transfer the properties available for $z \in \mathbb{C}^-$ to the whole space $\mathbb{C} \setminus [0, 1 - \varepsilon]$ thanks to two complex analysis arguments. Given $r > 0, w \in \mathbb{C}$ we note the open disk $\mathcal{D}_r(w) = \{z \in \mathbb{C} \mid |z - w| < r\}$ and the close disk $\bar{\mathcal{D}}_r(w) = \{z \in \mathbb{C} \mid |z - w| \leq r\}$.

Proposition 25. *(Rudin, 1986, Theorem 16.2) If an analytical mapping is defined and bounded on a disk $\mathcal{D}_r(w)$ for $r > 0$ and $w \in \mathbb{C}$, then it can be continued into an analytical mapping defined on a strictly bigger disk $\mathcal{D}_\rho(w)$ with $\rho > r$.*

Proposition 26. *(Rudin, 1986, Theorem 10.18) Given two analytical mappings f, g defined on an open set $U \subset \mathbb{C}$ if there exists a subset $Z \subset U$ such that U contains a limit point of Z ²⁶ and $\forall z \in Z, f(z) = g(z)$, then $f|_U = g|_U$.*

Proposition 25 allows us to extend the domain of $z \mapsto \Lambda^z$ (which is only analytical on set $\mathbb{C}^- \cap \{\Im(z) \leq -\nu\}$ for $\nu > 0$ thanks to proposition 24) and Proposition 26 allows us to set that on this bigger domain, $\mathcal{I}^z(\Lambda^z)$ is also equal to Λ^z . The boundedness of $z \mapsto \Lambda^z$ on the disks contained in $\mathbb{C} \setminus [0, 1 - \varepsilon]$ can be tracked from the mapping $z \mapsto \Delta^z$ that we know to be analytical on $\mathbb{C} \setminus [0, 1 - \varepsilon]$ (see Proposition 20), thus bounded on the sub-disks of $\mathbb{C} \setminus [0, 1]$. Indeed it can be showed that those two mappings are close to one-another when n tends to infinity thanks to Proposition 21 employed this time with $\Theta = \mathbb{N}$

²⁶There exists $u_0 \in U$ such that any neighborhood of u_0 contains a point of Z .

we will then have the bound $\|\Delta^z - \Lambda^z\| \leq O\left(\frac{\sqrt{\log n}}{n}\right)$ thanks to Proposition 15 ($\|\frac{1}{n}\Sigma_i\|_F \leq O(1/\sqrt{n})$) that entices:

$$\|\mathcal{I}^z \circ \chi(\Delta^z) - \Delta^z\| = \sup_{1 \leq i \leq n} \left| \frac{1}{n} \text{Tr} \left(\Sigma_i \left(\tilde{Q}^z(\chi(\Delta^z)) - \Delta^z \right) \right) \right| \leq O\left(\frac{\sqrt{\log n}}{n}\right). \quad (29)$$

To be able to employ Proposition 21, we are left to:

- bound $\frac{\sup_{1 \leq i \leq n} \Im(\mathcal{I}^z \circ \chi(\Delta^z))_i}{\inf_{1 \leq i \leq n} \Im(\mathcal{I}^z \circ \chi(\Delta^z))_i}$ and $\frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i}$ from above,
 - show that $\left\| \frac{\Im(\mathcal{I}^w \circ \chi(\Delta^w) - \Delta^w)}{\Im(\Delta^w)} \right\| \rightarrow 0$ when n is sufficiently big,
 - bound from below the functionals $\phi(z, \chi(\Delta^z)) \in (0, 1)$ and $\phi(z, \chi(\Lambda^z)) \in (0, 1)$ with a $O(1)$. Recall that ϕ appears in the inequality given by Proposition 19 satisfied for any $D, D' \in \mathcal{D}_n(\mathbb{C}^+)$ and $w \in \mathbb{C}^-$:
- $$d_s(\mathcal{I}^w \circ \chi(D), \mathcal{I}^w \circ \chi(D')) \leq \sqrt{(1 - \phi(z, \chi(D)))(1 - \phi(z, \chi(D')))} d_s(D, D'). \quad (30)$$

The results concerning Λ^z are more easy to obtain than the results concerning Δ^z , we are thus going to prove them first in the three next lemmas.

Lemma 16. *Given a compact set $K \subset \mathbb{C}^-$ such that $\sup_{w \in K} |w| \leq O(1)$ and $O(1) \leq \inf_{w \in K} d(w, [0, 1 - \varepsilon])$, we can bound:*

$$O(|\Im(w)|) \leq \Im(\Delta^w) \leq O(|\Im(w)|) \quad \text{and} \quad O(|\Im(w)|) \leq \Im(\mathcal{I}^w \circ \chi(\Delta^w)) \leq O(|\Im(w)|).$$

This lemma implies clearly:

$$\frac{\sup_{1 \leq i \leq n} \Im(\mathcal{I}^z \circ \chi(\Delta^z))_i}{\inf_{1 \leq i \leq n} \Im(\mathcal{I}^z \circ \chi(\Delta^z))_i} \leq O(1) \quad \text{and} \quad \frac{\sup_{1 \leq i \leq n} \Im(\Delta^z)_i}{\inf_{1 \leq i \leq n} \Im(\Delta^z)_i} \leq O(1) \quad (31)$$

PROOF. For any $i \in [n]$, and $w \in K$, we can bound thanks to Lemma 11 and the bound $O(1) \leq \frac{1}{n} \text{Tr}(\Sigma_i) \leq O(1)$ given by Assumption 4 (this is the only time where we need to bound $\frac{1}{n} \text{Tr}(\Sigma_i)$ from below):

$$O(|\Im(w)|) \leq \Im(\Delta_i^w) = -\Im(w) \mathbb{E}_{\mathcal{A}_Q} \left[\frac{1}{n} x_i^T |Q_{-i}^w|^2 x_i \right] \leq O(|\Im(w)|)$$

Let us note that thanks to Remark 23, $O(1) \leq |\chi(\Delta^w)| \leq O(1)$ and:

$$O(|\Im(w)|) \leq \Im(\chi(\Delta^w)) = \Im(\Delta^w) |\chi(\Delta^w)|^2 \leq O(|\Im(w)|)$$

Finally, thanks to the bounds on $\|\tilde{Q}^w(\chi(\Delta^z))\|$ given in (27), we can bound:

$$O(|\Im(w)|) \leq$$

$$\begin{aligned} \Im(\mathcal{I}^w \circ \chi(\Delta^w)) &= \text{Tr} \left(\Sigma_i \tilde{Q}^w(\chi(\Delta^z)) \left(-\Im(z) I_p + \frac{1}{n} \sum_{i=1}^n \Im(\chi(\Delta^w)) \Sigma_i \right) \tilde{Q}^w(\chi(\Delta^z)) \right) \\ &\leq O(|\Im(w)|) \end{aligned}$$

Lemma 17. *Given a compact set $K \subset \mathbb{C}^-$ such that $\sup_{w \in K} |w| \leq O(1)$ and $O(1) \leq \inf_{w \in K} d(w, [0, 1 - \varepsilon])$, we can bound:*

$$\left\| \frac{\Im(\mathcal{I}^w \circ \chi(\Delta^w) - \Delta^w)}{\Im(\Delta^w)} \right\| \leq O\left(\frac{\sqrt{\log n}}{n}\right)$$

PROOF. ²⁷ Recalling from the proof of Lemma 17 that $\Im(\Delta^w) \geq |\Im(w)|$, we can bound for any $w \in K$:

$$\begin{aligned} \left\| \frac{\Im(\mathcal{I}^w \circ \chi(\Delta^w) - \Delta^w)}{\Im(\Delta^w)} \right\| &= \sup_{1 \leq i \leq n} \left| \frac{\frac{1}{n} \text{Tr} \left(\Sigma_i \Im \left(\tilde{Q}^w(\chi(\Delta^w)) - \mathbb{E}_{A_Q} [Q^w] \right) \right)}{\Im(\Delta^w)} \right| \\ &\leq O \left(\left\| \frac{\Im(\mathbb{E}_{A_Q} [Q^w])}{\Im(w)} - \frac{\Im(\tilde{Q}^w(\chi(\Delta^w)))}{\Im(w)} \right\|_F \right) \end{aligned}$$

Let us try and estimate $\frac{\Im(\mathbb{E}_{A_Q} [Q^w])}{\Im(z)} = -\mathbb{E}_{A_Q} [|Q^w|^2] = -\mathbb{E}_{A_Q} [\tilde{Q}^w Q^w]$. Given a deterministic matrix $A \in \mathcal{M}_p$, we estimate rapidly without the necessary justifications (that are closely similar to those presented in the proof of Proposition 15

²⁷For this poof, one might be inspired to employ an analytical argument stating that the derivatives of $z \mapsto \mathbb{E}_{A_Q} [Q^z]$ and $z \mapsto \tilde{Q}^z(\chi(\Delta^z))$ are close to one-another. We have indeed $|\frac{\mathbb{E}_{A_Q} [Q^z]}{\Im(z)}| = |\frac{\mathbb{E}_{A_Q} [Q^z]}{\partial w}|$ and:

$$\begin{aligned} \frac{\partial \tilde{Q}^w(\chi(\Delta^w))}{\partial w} &= -\tilde{Q}^w(\chi(\Delta^w)) \left(I_p - \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{A_Q} [x_i^T (Q_{-i}^w)^2 x_i] \chi(\Delta_i^w)^2 \Sigma_i \right) \tilde{Q}^w(\chi(\Delta^w)) \\ \frac{\Im(\tilde{Q}^w(\chi(\Delta^w)))}{\Im(z)} &= \tilde{Q}^w(\chi(\Delta^w)) \left(-I_p + \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{A_Q} [x_i^T |Q_{-i}^w|^2 x_i] |\chi(\Delta_i^w)|^2 \Sigma_i \right) \tilde{Q}^w(\chi(\Delta^w)). \end{aligned}$$

However to set that $|\frac{\tilde{Q}^w(\chi(\Delta^w))}{\Im(z)}| \approx |\frac{\tilde{Q}^w(\chi(\Delta^w))}{\partial w}|$ one would need $\Im(\tilde{Q}^w(\chi(\Delta^w)))$ and $\Re(\tilde{Q}^w(\chi(\Delta^w)))$ to be almost commuting. This is however not so easy to show, even knowing that $\Im(\mathbb{E}_{A_Q} [Q^z])$ and $\Re(\mathbb{E}_{A_Q} [Q^z])$ commute.

– mainly application of Proposition ??):

$$\begin{aligned}
& \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\bar{Q}^w \left(Q^w - \tilde{Q}^w(\chi(\Delta^w)) \right) \right] \right) \\
&= \frac{1}{n} \sum_{i=1}^n \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}^w Q_{-i}^w x_i x_i^T \tilde{Q}^w(\chi(\Delta^w))}{1 - \frac{1}{n} x_i^T Q_{-i}^w x_i} \right] \right) - \frac{1}{n} \sum_{i=1}^n \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}^w Q^w \Sigma_i \tilde{Q}^w(\chi(\Delta^w))}{1 - \Delta_i^w} \right] \right) \\
&= \frac{1}{n} \sum_{i=1}^n \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}_{-i}^w Q_{-i}^w x_i x_i^T \tilde{Q}^w(\chi(\Delta^w))}{1 - \Delta_i^w} \right] \right) - \frac{1}{n} \sum_{i=1}^n \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}^w Q_{-i}^w \Sigma_i \tilde{Q}^w(\chi(\Delta^w))}{1 - \Delta_i^w} \right] \right) \\
&\quad + \frac{1}{n^2} \sum_{i=1}^n \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}^w x_i x_i^T Q_{-i}^w x_i x_i^T \tilde{Q}^w(\chi(\Delta^w))}{1 - \Delta_i^w} \right] \right) + O \left(\sqrt{\frac{\log n}{n}} \right) \\
&= \frac{\mathbb{E}_{A_Q} \left[\frac{1}{n} \text{Tr} \left(A \bar{Q}^w X D^w X^T \tilde{Q}^w(\chi(\Delta^w)) \right) \right]}{1 - \Delta_i^w} + O \left(\sqrt{\frac{\log n}{n}} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \Delta_i^w \chi(\Delta_i^w) \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\frac{\bar{Q}_{-i}^w x_i x_i^T \tilde{Q}^w(\chi(\Delta^w))}{1 + D_i^z} \right] \right) + O \left(\sqrt{\frac{\log n}{n}} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \Delta_i^w |\chi(\Delta_i^w)|^2 \text{Tr} \left(A\mathbb{E}_{A_Q} \left[\bar{Q}_{-i}^w x_i x_i^T \tilde{Q}^w(\chi(\Delta^w)) \right] \right) + O \left(\sqrt{\frac{\log n}{n}} \right) \\
&= \text{Tr} \left(A \bar{Q}^w(\chi(\Delta^w)) \Sigma_{\Delta^w |\chi(\Delta^w)|^2} \tilde{Q}^w(\chi(\Delta^w)) \right) + O \left(\sqrt{\frac{\log n}{n}} \right),
\end{aligned}$$

where we recall the notation $\Sigma_\Gamma \equiv \frac{1}{n} \sum_{i=1}^n \Gamma \Sigma_i$ for any diagonal matrix $\Gamma \in \mathcal{D}_n$.
In conclusion:

$$\left\| \mathbb{E}_{A_Q} [|Q^w|^2] - \bar{Q}^w(\chi(\Delta^w)) (I_p + \Sigma_{\Delta^w |\chi(\Delta^w)|^2}) \tilde{Q}^w(\chi(\Delta^w)) \right\|_F \leq O \left(\sqrt{\frac{\log n}{n}} \right)$$

We recognize here the expression of $\Im(\tilde{Q}^w(\chi(\Delta^w)))/\Im(w)$ which allows us to conclude:

$$\frac{1}{\Im(z)} \left\| \Im \left(\mathbb{E}_{A_Q} [Q^w] - \tilde{Q}^w(\chi(\Delta^w)) \right) \right\|_F \leq O \left(\sqrt{\frac{\log n}{n}} \right)$$

Lemma 18. *Given a compact set $K \subset \mathbb{C}^-$ such that $\sup_{w \in K} |w| \leq O(1)$ and $\inf_{w \in K} d(w, [0, 1 - \varepsilon]) \geq O(1)$, we can bound:*

$$\phi(w, \Delta^z) \geq O(1).$$

PROOF. From the expression

$$\phi(w, \chi(\Delta^z)) = \inf_{1 \leq i \leq n} \frac{|\Im(w)|}{n} \frac{\text{Tr} \left(\Sigma_i |\tilde{Q}^w(\chi(\Delta^z))|^2 \right)}{\Im(\mathcal{I}^w(\chi(\Delta^z)))_i},$$

we deduce directly the result of the Lemma from the lower bound on $|\tilde{Q}^w(\chi(\Delta^z))|^2$ given in (27), Assumption 4 and Lemma 16

The implicit formulation of Λ^z complicates the design of a bigger bound for $\frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i}$ and a lower bound for $\phi(z, \Lambda^z)$ but one can track it from Lemmas 16 and 18. Indeed, $D \mapsto \frac{\sup_{1 \leq i \leq n} \Im(D)_i}{\inf_{1 \leq i \leq n} \Im(D)_i}$ and Φ being uniformly continuous on any compact, we have the following Lemma:

Lemma 19. *Given two compacts $K, K' \subset \mathbb{C}^+$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $\sup_{z \in K \cap K'} |z| \leq \sup_{z \in K \cap K'} |z| \leq O(1)$ and two positive scalars $C, \eta > 0$, there exists $N \in \mathbb{N}$ such that for any $n \geq N$, $\forall D, D' \in \mathcal{D}_n(K), \forall w \in K'$:*

$$\|D - D'\| \leq \frac{C\sqrt{\log n}}{n} \implies \begin{cases} \left\| \frac{\sup_{1 \leq i \leq n} \Im(D)_i}{\inf_{1 \leq i \leq n} \Im(D)_i} - \frac{\sup_{1 \leq i \leq n} \Im(D')_i}{\inf_{1 \leq i \leq n} \Im(D')_i} \right\| < \eta \\ |\phi(w, D) - \phi(w, D')| < \eta \end{cases}$$

One could object that to set the bound $\|\Lambda^z - \Delta^z\| \leq \frac{C\sqrt{\log n}}{n}$ we precisely need Proposition 21 that needs in particular the result of Lemma 19 to be valid : it looks like the snail is eating its own tail ! Actually, these properties can be invoked in a relevant iterative way described rigorously in the next proof that puts all together the preceding results to demonstrate Propositions 16 and 17.

3.7. Proof of Propositions 16 and 17

Let us consider $z \in \mathbb{C} \setminus [0, 1 - \varepsilon]$ such that $|z| \leq O(1)$, $d(z, [0, 1 - \varepsilon]) \geq O(1)$. Let us note

$$K \equiv \left\{ w \in \mathbb{C}^-, |w| \leq 2|z|, d(w, [0, 1 - \varepsilon]) \geq \frac{d(z, [0, 1 - \varepsilon])}{2} \right\}.$$

There exists an integer $k \leq O(1)$, a set of positive scalars $r_1, \dots, r_k > 0$ and a set of complex values $z_0, \dots, z_k$ such that $\mathcal{B}_{r_0}(z_0) \in \mathbb{C}^- \cap K \cap \{\Im(z) \leq -\nu\}$ (where ν , introduced in Proposition 24, is such that $z \mapsto \Lambda^z$ is analytic on $\mathbb{C}^- \cap \{\Im(z) \leq -\nu\}$), $z_k = z$ and for all $l \in [k]$:

$$\bar{\mathcal{D}}_{r_l}(z_l) \subset K \quad \text{and} \quad z_l \in \mathcal{D}_{r_{l-1}}(z_{l-1}),$$

With the purpose of employing Proposition 21, let us introduce two constants $\kappa \leq O(1)$ and $\eta \geq O(1)$ originating from (31) and Lemma 18 and satisfying:

$$\forall w \in K : \frac{\sup_{1 \leq i \leq n} \Im(\chi(\Delta^w))_i}{\inf_{1 \leq i \leq n} \Im(\chi(\Delta^w))_i} \leq \kappa - \eta \quad \text{and} \quad \phi(w, \chi(\Delta^w)) \geq 2\eta. \quad (32)$$

Let us show by iteration on $l \in [k]$ that there exists $N \in \mathbb{N}$ such that for $n \geq N$, and for all $l \in [k]$, the following set of properties that we note P_l is satisfied for any $w \in \bar{\mathcal{D}}_{r_l}(z_l)$:

- $z \mapsto \Lambda^z$ is analytic around $z = w$,
- $\Lambda^w = \mathcal{I}^w(\chi(\Delta^w))$,
- $\frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i} \leq \kappa$,

- $\phi(w, \chi(\Lambda^w)) \geq \eta$.

The case $l = 0$ allows us to introduce a convenient $N \in \mathbb{N}$ appearing in Lemma 19, it will be the same for all iterations, since the compact set K and the constant η will not change. Working in $\mathcal{D}_{r_0}(z_0)$, let us note $\rho \in (0, r_0]$ the biggest positive radius such that:

$$\forall w \in \mathcal{D}_\rho(z_0) : \quad \frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i} \leq \kappa \quad \text{and} \quad \phi(w, \chi(\Lambda^w)) \geq \eta \quad (33)$$

We can then employ Proposition 21 with $\Theta = \mathbb{N}$, and for all $n \in \Theta$, $f_n = \mathcal{I}^w \circ \chi$ (thus $D_n = \Lambda^w$) and $\Gamma_n = \Delta^w$, for a given $w \in \mathcal{D}_\rho(z_0)$. Indeed:

- $\forall w \in \mathcal{D}_\rho(z_0) \quad \frac{\sup_{1 \leq i \leq n} \Im(\Lambda^w)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^w)_i} \leq \kappa$ and $\frac{\sup_{1 \leq i \leq n} \Im(\mathcal{I}^w \circ \chi(\Delta^w))_i}{\inf_{1 \leq i \leq n} \Im(\mathcal{I}^w \circ \chi(\Delta^w))_i} \leq \kappa$,
- inequality (30) stated by Proposition 19 combined with (32) and (33) implies that:

$$\forall w \in \mathcal{D}_\rho(z_0) : \quad d_s(\chi \circ \mathcal{I}^z(\Lambda^w), \chi \circ \mathcal{I}^z(\Delta^w)) \leq (1 - \eta) d_s(D^w, \Delta^w)$$

- we know from Lemma 17 that $\left\| \frac{1}{\Im(\Delta^w)} \Im(\mathcal{I}^w \circ \chi(\Delta^w) - \Delta^w) \right\| \xrightarrow{n \rightarrow \infty} 0$, and therefore, for n sufficiently large:

$$\forall w \in \mathcal{D}_\rho(z_0) : \quad (1 - \nu) \left(1 + \left\| \sqrt{\frac{|\Im(\mathcal{I}^w \circ \chi(\Delta^w)) - \Im(\Delta^w)|}{\Im(\Delta^w)}} \right\| \right) \leq 1 - \frac{\nu}{2} < 1$$

Thus Proposition 21 combined with (29) implies that there exists a constant $C \geq 0$ depending only on κ and ν such that:

$$\forall w \in \mathcal{D}_{r_0}(z_0) : \quad \|\Lambda^w - \Delta^w\| \leq O(\|\mathcal{I}^w \circ \chi(\Delta^w) - \Delta^w\|) \leq C \frac{\sqrt{\log n}}{n} \quad (34)$$

Then, we can conclude from Lemma 19 and (32) that there exists $N \in \mathbb{N}$ such that for any $n \geq N$:

$$\forall w \in \mathcal{D}_\rho(z_0) : \quad \frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i} < \kappa \quad \text{and} \quad \phi(w, \chi(\Lambda^w)) > \eta$$

By continuity of $w \mapsto \frac{\sup_{1 \leq i \leq n} \Im(\Lambda^z)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^z)_i}$ and $w \mapsto \phi(w, \chi(\Lambda^w))$, one can then consider a bigger $\rho > 0$ satisfying (33), that means that $\rho = r_0$.

Let us then assume that P_{l-1} is satisfied for a given $l \in [k]$. We know that $z \mapsto D_z$ is analytical around z_l since $z_l \in \mathcal{D}_{r_{l-1}}(z_{l-1})$. Therefore, one can introduce two positive scalars $\rho_1, \rho_2 > 0$, the maximum radiuses such that respectively:

- $\forall w \in \mathcal{D}_{\rho_1}(z_l), \quad \frac{\sup_{1 \leq i \leq n} \Im(\Lambda^w)_i}{\inf_{1 \leq i \leq n} \Im(\Lambda^w)_i} \leq \kappa$ and $\phi(w, \chi(\Lambda^w)) \geq \eta$,
- $w \mapsto f(w, D^w)$ is analytic on $\mathcal{D}_{\rho_2}(z_l)$.

Let us suppose in a first time that $\rho_2 \leq \rho_1$. The mapping $z \mapsto I^z(\Lambda^z)$ is also analytical on $\mathcal{D}_{\rho_2}(z_l)$, but since it is equal to $z \mapsto D_z$ on an open set around z_l (thanks to P_{l-1}), the two mappings are equal on the full open disk $\mathcal{D}_{\rho_2}(z_l)$ (see Proposition 26). As before, for $w \in \mathcal{D}_{\rho_2}(z_l) \subset \mathcal{D}_{\rho_1}(z_l)$, one can apply Proposition 21 to set that:

$$\forall w \in \mathcal{D}_{\rho_2}(z_l) : \|\Lambda^w - \Delta^w\| \leq C \frac{\sqrt{\log n}}{n}$$

(with the same constant $C > 0$ as in (34)). In particular, Δ^w is bounded on $\bar{\mathcal{D}}_{\rho_2}(z_l) \subset \mathbb{C}^+$ and therefore so is Λ^w , which directly implies that it is analytical on a strictly bigger disk than $\mathcal{D}_{\rho_2}(z_l)$ (see Proposition 25). Therefore $\rho_2 > \rho_1$. For any $w \in \mathcal{D}_{\rho_1}(z_l)$, $\Lambda^w = \mathcal{I}^w \circ \chi(\Lambda^w)$, thus one can employ Proposition 21 the same way as in the case $l = 0$ to set that $\rho_1 = \rho_2 = r_l$. That concludes the iteration and gives the result of the Proposition for any $z \in \mathbb{C}^-$.

When $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C}, \Im(z) > 0\}$ and is such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $|z| \leq O(1)$, we can deduce the same results by symmetry since for all $z \in \mathbb{C}$, $\Delta^{\bar{z}} = \overline{\Delta^z}$ and $\forall D \in \mathcal{D}_n(\mathbb{C})$, such that $\mathcal{I}^z(D)$ is well defined, $\mathcal{I}^{\bar{z}}(\bar{D}) = \overline{\mathcal{I}^z(D)}$ therefore for all $z \in \mathbb{C}^+$, Λ^z is well defined and satisfies $\Lambda^z = \overline{\Lambda^{\bar{z}}}$, then one can bound:

$$\|\Lambda^z - \Delta^z\| = \|\overline{\Lambda^{\bar{z}}} - \overline{\Delta^{\bar{z}}}\| \leq C \frac{\sqrt{\log n}}{n}$$

All those result are also true for any $z \in \mathbb{R}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $|z| \leq O(1)$, because Proposition 25 implies that $z \mapsto \Lambda^z$ and $z \mapsto \Delta^z$ are both analytical on this set and we can then deduce the bound on $\|\Lambda^z - \Delta^z\|$ by continuity.

3.8. Resolvent of $\frac{1}{n}XDX^T$

Let us now present a result of concentration of a generalization of the resolvent Q , that will be useful for the study of the solutions Y to $Y = \frac{\lambda}{n} \sum_{i=1}^n f(x_i^T Y) x_i$ conducted in next section (the diagonal matrix D will then be $\text{Diag}_{1 \leq i \leq n} f'(x_i^T Y) \in \mathcal{D}_n$). Given a diagonal matrix $D \in \mathcal{D}_n$, we note:

$$Q^z(D) \equiv \left(zI_n - \frac{1}{n}XDX^T \right)^{-1}$$

Let us consider a random diagonal matrix $\Gamma \in \mathcal{D}_n$ satisfying:

Assumption 6. *There exists a deterministic diagonal matrix $\tilde{\Gamma} \in \mathcal{D}_n$ such that*

$$\Gamma \stackrel{\mathcal{A}_\Gamma}{\propto} \tilde{\Gamma} \pm \mathcal{E}_2 \mid e^{-n} \text{ in } (\mathcal{M}_p, \|\cdot\|) \quad \text{and} \quad \|\tilde{\Gamma}\| \leq O(\sqrt{\log n})$$

In this setting, Theorem 2 (see Remark 14) just allows us to set the concentration (since $\|X\Gamma X^T\|_F \leq \|X\| \|\Gamma\| \|X\|_F$):

$$\frac{1}{n}X\Gamma X^T \propto \mathcal{E}_2(\sqrt{\log n/n}) \mid e^{-n} \quad \text{in } (\mathcal{M}_{p,n}, \|\cdot\|_F), \quad (35)$$

and no better observable diameter can be obtained in $(\mathcal{M}_{p,n}, \|\cdot\|)$. This time, instead of Assumption 5, we need the stronger hypothesis:

Assumption 5 bis. *There exists a parameter $\varepsilon \geq O(1)$, an event $\mathcal{A}_{Q(\Gamma)}$ and two constants $C, c > 0$ such that:*

$$\mathbb{P}(\mathcal{A}_{Q(\Gamma)}^c) \leq Ce^{-cn} \quad \text{for } \mathcal{A}_{Q(\Gamma)} \equiv \mathcal{A}_\Gamma \cap \left\{ \frac{1}{n} \|X\Gamma X^T\| \leq 1 - \varepsilon \right\}$$

Under this assumption, we still have (as in Lemma 11):

Lemma 20. *Under $\mathcal{A}_{Q(\Gamma)}$: $\|Q^z(\Gamma)\| \leq \frac{1}{d(z, [0, 1 - \varepsilon])}$.*

Placing ourselves under Assumptions 1 - 4, 5 bis, 6, we already know from the Lipschitz²⁸ concentration of $\frac{1}{n}X\Gamma X^T$ given in (35) and the Lipschitz character of $M \mapsto (zI_p - M)^{-1}$ (for $z \in \mathbb{C}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and under $\mathcal{A}_Q$) that:

$$Q^z(\Gamma) \propto \mathcal{E}_2\left(\sqrt{\log n}\right) \mid e^{-n} \quad \text{in } (\mathcal{M}_p, \|\cdot\|_F)$$

It is not such a good concentration, but it still allows us set approximations of quantities like $\frac{1}{n} \text{Tr}(AQ^z(\Gamma))$ when $\|A\| \leq O(1)$ thanks to next proposition.

Proposition 27. *$Q^z(\Gamma) \in \frac{1}{z}I_p \pm \mathcal{E}_2\left(\sqrt{\log n}\right) \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_F)$.*

PROOF. It is just a consequence of Proposition ?? as described in Remark 24, we can bound indeed for any $A \in \mathcal{M}_p$ such that $\|A\| \leq 1$:

$$\begin{aligned} \left| \mathbb{E}_{\mathcal{A}_{Q(\Gamma)}} \left[\text{Tr} \left(A \left(Q^z(\Gamma) - \frac{1}{z}I_p \right) \right) \right] \right| &\leq \frac{1}{n} \left| \mathbb{E}_{\mathcal{A}_{Q(\Gamma)}} \left[\frac{1}{z} \text{Tr}(AQ^z(\Gamma)X\Gamma X) \right] \right| \\ &\leq O\left(\sqrt{\frac{\log n}{n}} \|A\|_F \|\Gamma\|_F \right) \leq O(\sqrt{\log n}) \end{aligned}$$

We can obtain a better observable diameter if we project $Q^z(\Gamma)$ on a deterministic vector.

Proposition 28. *For any $z \in \mathbb{C}$ such that $|z| \leq O(1)$, and $d(z, [0, 1 - \varepsilon]) \geq O(1)$, for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:*

$$Q^z(\Gamma)u \in \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right) \mid e^{-n} \quad \text{in } (\mathbb{R}^p, \|\cdot\|).$$

PROOF. It is a simple and direct application of Theorem 5 and Proposition 10 to the mapping:

$$\phi : y \mapsto u + \frac{1}{n}X\Gamma X^T y.$$

²⁸we need here a Lipschitz concentration to control the Lipschitz transformation operated by the resolvent. That is why we can not employ the second result of Corollary 4 that could gives us the linear concentration of $\frac{1}{n}X\Gamma X^T$ in $(\mathcal{M}_p, \|\cdot\|_F)$ if Γ would be concentration in $(\mathcal{D}_n, \|\cdot\|_F)$.

3.9. Concentration of the resolvent of the sample covariance matrix of convexly concentrated data

This subsection is the only one that employs results of Section 2, to show that the resolvent is concentrated (recall that in the case of a Lipschitz concentration hypothesis on X , it is more simple to show that the resolvent is a $O(1/\sqrt{n})$ -Lipschitz transformation of X). We suppose this time:

Assumption 1 bis. $X \propto_c \mathcal{E}_2$

Placing ourselves (in this subsection) under Assumptions 1 bis, 2-5 let us first show that the resolvent $Q^z \equiv (zI_p - \frac{1}{n}XX^T)^{-1}$ is concentrated if z is far enough from the spectrum.

Proposition 29. *Given $z \in \mathbb{C}$ such that $d(z, [0, 1 - \varepsilon]) \geq O(1)^{29}$:*

$$Q^z \in \mathcal{E}_2 \mid e^{-n} \quad \text{in } (\mathcal{M}_p, \|\cdot\|_*)$$

PROOF. If $|z| \geq (2e)^2$, then $Q^z = \frac{1}{z}(I_p - \frac{1}{zn}XX^T)^{-1}$, and $\frac{X}{\sqrt{z}}$ satisfies the hypothesis of Corollary 8 ($\sqrt{z}$ is a given square root of z , $\|X/\sqrt{z}\| \leq \sqrt{p}(1-\varepsilon)/2e$ with high probability). Therefore, the concentration of Q^z is already valid in that case.

Now, considering a general complex number $z \in \mathbb{C}$ such that $|z| \leq (2e)^2$ and $d(z, [0, 1 - \varepsilon]) \geq O(1)$, we introduce a parameter $K \geq (2e)^2$ and the complex number $z_K \in \mathbb{C}$, such that:

$$|z_K| = K \quad \text{and} \quad d(z_K, [0, 1 - \varepsilon]) = |z - z_K| + d(z, [0, 1 - \varepsilon]).$$

Note then that Q^z is solution to:

$$Q^z = Q^{z_K} + (z_K - z)Q^{z_K}Q^z, \quad (36)$$

As we are going to see, this is a contractive fixed point equation that leads us to employing Theorem 7 giving the concentration of solutions to fixed point equations. To verify the hypotheses, we need in particular to understand the concentration of the powers of Q^{z_K} . Considering an integer $k \in \mathbb{N}$ such that $k \leq O(\log n)$, let us express:

$$(z_K Q^{z_K})^k = \left(I_n - \frac{1}{nz_K} XX^T \right)^{-k} = I_n + \sum_{i=0}^{\infty} \frac{(k+i) \cdots k}{(i+1)!} \left(\frac{1}{nz_K} XX^T \right)^{i+1}$$

Besides, we know from Proposition 13 that

$$\left(\frac{1}{nz_K} XX^T \right)^{i+1} \in \mathcal{E}_2 \left(\left(\frac{4e^2(1-\varepsilon)}{K} \right)^{i+1} \right) \mid e^{-n}.$$

²⁹ ε was introduced in Assumption 5

Then, assuming that $K \geq 4e^2(1 - \varepsilon)$, we can bound:

$$1 + \sum_{i=0}^{\infty} \frac{(k+i) \cdots k}{(i+1)!} \left(\frac{4e^2(1-\varepsilon)}{K} \right)^{i+1} = \left(1 - \frac{4e^2(1-\varepsilon)}{K} \right)^{-k},$$

from which we deduce from Corollary 2 that:

$$(z_K Q^{z_K})^k \in \mathcal{E}_2 \left(\left(\frac{1}{K - 4e^2(1-\varepsilon)} \right)^k \right) \mid e^{-n}$$

Returning to equation (36), we chose $K = 2 + 4e^2(1 - \varepsilon)$, then

$$\left| \frac{z_K - z}{z_K} \right| \frac{1}{K - 4e^2(1-\varepsilon)} \leq \frac{d(z_K, [0, 1-\varepsilon]) - d(z, [0, 1-\varepsilon])}{2K} \leq \frac{K-1}{2K} \leq \frac{1}{2}$$

and we see that the mapping

$$\phi : q \mapsto \frac{1}{z_K} (z_K Q^{z_K}) + \frac{(z_K - z)}{z_K} (z_K Q^{z_K}) q$$

satisfies:

- for any $q \in \mathcal{M}_p$ such that $\|q\| \leq 1$,

$$\phi^k(q) = \left(\frac{z_K - z}{z_K} \right)^k (z_K Q^{z_K})^k q + \frac{1}{z_K} \sum_{l=0}^{k-1} \left(\frac{z_K - z}{z_K} \right)^l (z_K Q^{z_K})^{l+1} \in \mathcal{E}_2 \mid e^{-n},$$

in $(\mathcal{M}_p, \|\cdot\|_*)$ (thanks again to Corollary 1).

- $\|\phi(0)\| = \|I_p\| \leq O(1)$
- $\|\mathcal{L}(\phi)\|_*, \|\mathcal{L}(\phi)\| \leq |z - z_K| \|Q^{z_K}\| \leq \frac{|z_K - z|}{d(z_K, [0, 1-\varepsilon])} \leq 1 - \frac{d(z, [0, 1-\varepsilon])}{d(z_K, [0, 1-\varepsilon])}$ thanks to to Lemma 12 and the hypothesis on z_K . Besides, $\varepsilon' \equiv \frac{d(z, [0, 1-\varepsilon])}{d(z_K, [0, 1-\varepsilon])} \geq O(1)$ and:

$$\mathbb{P}(\|\mathcal{L}(\phi)\|_* \geq 1 - \varepsilon'), \mathbb{P}(\|\mathcal{L}(\phi)\| \geq 1 - \varepsilon') \leq C e^{-c p},$$

for some constant $C, c > 0$.

We can therefore employ Theorem 7 to set the result of the proposition.

Lemma 21. $D^z \in \Delta^z \pm \mathcal{E}_2 \mid e^{-n}$ in $(\mathcal{D}_n(\mathbb{C}), \|\cdot\|_F)$.

PROOF. Given $i \in [n]$, we can bound:

$$|D_i^z - \Delta_i^z| \leq \left| \frac{1}{n} x_i^T Q_{-i}^z x_i - \frac{1}{n} \text{Tr}(\Sigma_i Q_{-i}^z) \right| + \left| \frac{1}{n} \text{Tr}(\Sigma_i (Q_{-i}^z - \mathbb{E}[Q_{-i}^z])) \right|$$

and we know respectively from Corollary 6 ($\|Q_{-i}^z\| \leq O(1)$) and Proposition 29 ($\|\frac{1}{n}\Sigma_i\| \leq O(1/n)$) that:

$$\begin{aligned} \frac{1}{n}x_i^T Q_{-i}^z x_i &\in \frac{1}{n} \text{Tr}(\Sigma_i Q_{-i}^z) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}; \\ \frac{1}{n} \text{Tr}(\Sigma_i Q_{-i}^z) &\in \frac{1}{n} \text{Tr}(\Sigma_i \mathbb{E}_{\mathcal{A}_Q}[Q_{-i}^z]) \pm \mathcal{E}_2\left(\frac{1}{n}\right) \mid e^{-n}, \end{aligned}$$

thus $D_i^z \in \Delta_i^z \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$. We then deduce from Proposition 5 that:

$$D^z \in \Delta^z \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n} \quad \text{in } (\mathcal{M}_p, \|\cdot\|),$$

which implies the result of the Lemma since $\|A\|_F \leq \sqrt{p}\|A\|$ for all $A \in \mathcal{M}_p$.

Proposition 30. *There exists a constant $\kappa \leq O(1)$ such that for any $z \in \mathbb{C}$ satisfying $d(z, [0, 1 - \varepsilon]) \geq O(1)$ and $|z| \leq O(1)$:*

$$Q^z \in \tilde{Q}^z(\chi(\Lambda^z)) \pm \mathcal{E}_2(\log(n)) \mid e^{-n} \quad \text{in } (\mathcal{M}_p, \|\cdot\|_*),$$

where Λ^z is defined in Proposition 16.

PROOF. We already know from Corollary 8 that $Q^z \in \mathcal{E}_2 \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_*)$. With notations already introduced ($\Delta^z \equiv \text{Diag}_{1 \leq i \leq n}(\frac{1}{n} \text{Tr}(\Sigma_i \mathbb{E}_{\mathcal{A}_Q}[Q^z]))$), we already know from Theorem 10 that:

$$\|\tilde{Q}^z(\chi(\Delta^z)) - \tilde{Q}^z(\chi(\Lambda^z))\|_* \leq \sqrt{p}\|\tilde{Q}^z(\chi(\Delta^z)) - \tilde{Q}^z(\chi(\Lambda^z))\|_F \leq O(1)$$

(the deterministic matrices $\tilde{Q}^z(\chi(\Delta^z))$ and $\tilde{Q}^z(\chi(\Lambda^z))$ have not changed). It is therefore sufficient to show that $\mathbb{E}_{\mathcal{A}_Q}[Q^z]$ is close to $\tilde{Q}^z(\chi(\Delta^z))$. We follow the steps of the proof of Proposition 15, starting from $\|\mathbb{E}_{\mathcal{A}_Q}[Q^z] - \tilde{Q}^z(\chi(\Delta^z))\|_* \leq \sup_{\|A\| \leq 1} \text{Tr}(A(\varepsilon_1 + \varepsilon_2))$. First we bound thanks to Proposition ??³⁰, Lemma 21 and the bound $\|\tilde{Q}^z(\chi(\Delta^z))\| \leq O(1)$:

$$\begin{aligned} \text{Tr}(A\varepsilon_1) &= \frac{1}{n} \left| \mathbb{E}_{\mathcal{A}_Q} \left[\text{Tr} \left(\tilde{Q}^z(\Delta^z) A Q^z X \chi(\Delta^z) (D^z - \Delta^z) X^T \right) \right] \right| \\ &\leq O \left(\|A\|_F \sqrt{\frac{\log n}{n}} \right) \leq O(\sqrt{\log n}) \end{aligned}$$

Second, as in the proof of Proposition 15, we can bound:

$$|\mathbb{E}_{\mathcal{A}_Q}[\text{Tr}(A\varepsilon_2)]| \leq O \left(\sqrt{\frac{\text{Tr}(A^T A)}{n} \frac{\text{Tr}(\tilde{Q}^z(\chi(\Delta^z))^2)}{n}} \right) \leq O(1)$$

We can thus conclude that $Q^z \in \tilde{Q}^z(\chi(\Delta^z)) \pm \mathcal{E}_2(\log(n)) \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_*)$ from which we deduce our result.

³⁰We do not exactly employ Proposition ?? since we can just deduce from Corollary 8 that $Q^z X \in \mathcal{E}_2(\sqrt{n}) \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_*)$, and we would need an observable diameter of order $O(1)$. However, we can adapt the proof thanks to the bound $\|Q^z X\| \leq O(\|X\|)$

4. Second example : fixed point equation depending on independent data $x_1, \dots, x_n$

4.1. Setting and first properties

We give here some conditions on a random matrix $X \in \mathcal{M}_{p,n}$ and a twice-differentiable mapping $\Psi : \mathcal{M}_{p,n} \rightarrow \mathcal{F}(E)$ satisfying the hypotheses of Corollary 9 to be able to compute the expectation of $Y = \Psi(X)(Y)$. We are going to study the very common case of a matrix of data $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$, where all the columns of X are independent but not identically distributed and Ψ acting on each column x_i “independently” through the decomposition for all $A = (a_1, \dots, a_n) \in \mathcal{M}_{p,n}$, all $y \in E$:

$$\Psi(A)(y) = \frac{1}{n} \sum_{i=1}^n H_i(a_i)(y), \quad (37)$$

where $H_1, \dots, H_n : \mathbb{R}^p \rightarrow \mathcal{Lip}(E)$.

To compute the expectation of Y , one needs to disentangle the influence of each data x_i on Y . This leads us to studying the random vector Y_{-i} , defined (when it exists) as the unique solution to:

$$Y_{-i} = \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^n H_j(a_j)(Y_{-i}),$$

it is independent with x_i by construction. To link Y with Y_{-i} we creates a “bridge” defined by a parameter $t \in [0, 1]$, through a mapping $\Psi_{-i}^t : \mathcal{M}_{p,n} \rightarrow \mathcal{Lip}(E)$, defined for any $A \in \mathcal{M}_{p,n}$ and any $y \in E$ with:

$$\Psi_{-i}^t(A)(y) \equiv \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^n H_j(a_j)(y) + t H_i(a_i)(y).$$

Then, noting y_{-i}^A , the unique solution (when it exists) to $y_{-i}^A = \Psi_{-i}^t(A)(y_{-i}^A)$, we see that:

$$Y_{-i} = y_{-i}^X(0) \quad \text{and} \quad Y = y_{-i}^X(1)$$

The next Lemma introduces a mapping $y_{-i}^A : [0, 1] \rightarrow E$ that creates a “bridge” between Y_{-i} and Y (when $A = X$).

Lemma 22. *Considering an open set $U \subset \mathcal{M}_{p,n}$, if we suppose that for all $A \in U$, $\Psi(A)$ is differentiable and satisfies $\|\Psi(A)\|_{\mathcal{L}} \leq 1 - \varepsilon$, then the mapping y_{-i}^A is differentiable and we have:*

$$y_{-i}^{A'}(t) = \frac{1}{n} \left(I_E - d_2 \Psi_{-i}(A) \big|_{y_{-i}^A(t)} \right)^{-1} \cdot \left(H_i(a_i)(y_{-i}^A(t)) + t d_2 H_i(a_i) \big|_{y_{-i}^A(t)} \cdot y_{-i}^{A'}(t) \right),$$

where we $d_2 \Psi_{-i}(A)$ and $d_2 H_i(a_i)$ are respectively the differential of the mappings $\Psi_{-i}(A) : E \rightarrow E$ and $H_i(a_i) : E \rightarrow E$.

PROOF. It is a direct application of the inverse function theorem to the $\mathcal{C}^1$ bijective mapping:

$$\begin{aligned} \Theta : \mathbb{R} \times E &\longrightarrow \mathbb{R} \times E \\ (t, y) &\longmapsto \left(t, y - \frac{1}{n} \sum_{i=1}^n t H(a_i)(y) \right) \end{aligned}$$

Indeed, $\mathbb{R} \times E$ is a Banach space, $d\Theta$ is clearly bounded ($\Psi(A)$ is Lipschitz), and $\forall (t, y) \times (s, h) \in (\mathbb{R} \times E)^2$:

$$d\Theta|_{(t,y)} \cdot (s, h) = \left(s, h - sH(a_i)(y) - d_2\Psi_{-i}^t(A)|_y \cdot h \right)$$

Thus $d\Theta|_{(t,y)}$ is invertible with:

$$d\Theta|_{(t,y)}^{-1} \cdot (s', h') = \left(s', (Id_E - d_2\Psi_{-i}^t(A)|_y)^{-1} s' H(a_i)(y) \right)$$

($\|d\Psi_{-i}^t(A)|_{y_{-i}^A(t)}\| \leq 1 - \varepsilon$ thus $Id_E - d_2\Psi_{-i}^t(A)|_y$ is invertible). Therefore, θ^{-1} is also C^1 and, we can differentiate $y_{-i}^A = \Theta^{-1}(t, 0)$ to obtain the identity:

$$\begin{aligned} y_{-i}^{A'}(t) &= d_2\Psi_{-i}^0(A)|_{y_{-i}^A(t)} \cdot y_{-i}^{A'}(t) + \frac{1}{n} H(a_i)(y_{-i}^A(t)) \\ &\quad + \frac{1}{n} t d_2 H(a_i)|_{y_{-i}^A(t)} \cdot y_{-i}^{A'}(t), \end{aligned} \tag{38}$$

from which we retrieve directly the result of the Lemma.

To conduct our probabilistic study of $Y \in E$, we keep the assumptions made in Section 3 concerning the quantities n , p and the random matrix $X \in \mathcal{M}_{p,n}$. Now, to simplify the assumptions made in the last Lemma, we suppose in addition that $E = \mathbb{R}^p$ and we assume that there exists a twice differentiable mapping $f : \mathbb{R} \rightarrow \mathbb{R}$ such that all the mappings H_i are equal and satisfy:

$$\forall i \in [n], \forall a, y \in \mathbb{R}^p : H_i(a)(y) = f(a^T y) a$$

To be able to justify the existence and uniqueness of Y , one is led to adapt Assumptions 5 and 5 bis and assume that:

Assumption 5 ter. *There exists a constant $\lambda > 0$, $\forall t \in \mathbb{R} \ |f'(t)| \leq \lambda$ with $1 - \frac{\lambda}{n} \mathbb{E}[\|XX^T\|] \geq O(1)$.*

Lemma 23. *Under Assumption 1-5 ter, there exists a constant $\varepsilon > 0$ such that if we set $\mathcal{A}_Y = \{\lambda \frac{1}{n} \|XX^T\| \leq 1 - \varepsilon\}$, then under $\mathcal{A}_Y$, the random vector $Y \in \mathbb{R}^p$ solution to $Y = \Psi(X)(Y)$ is well defined and $\mathbb{P}(\mathcal{A}_Y^c) \leq C e^{-cn}$ for some constants $C, c > 0$.*

PROOF. Noting $\varepsilon = \frac{1}{2}(1 - \lambda \frac{1}{n} \mathbb{E}[\|XX^T\|])$, we know from Example 3 that $\frac{\lambda}{n}\|XX^T\|$ like $\frac{1}{n}XX^T$ satisfies the concentration $\frac{1}{n}\|XX^T\| \propto \mathcal{E}_2(1/\sqrt{n}) + \mathcal{E}_2(1/n)$ thus, there exists some constants $C', c' > 0$ such that:

$$\mathbb{P}\left(\left|\frac{\lambda}{n}\|XX^T\| - \frac{\lambda}{n}\mathbb{E}[\|XX^T\|]\right| \geq \varepsilon\right) \leq C' e^{-c'n/\varepsilon^2} + C' e^{-c'n/\varepsilon}.$$

Therefore:

$$\mathbb{P}(\mathcal{A}_X^c) \leq \mathbb{P}\left(\frac{\lambda}{n}\|XX^T\| \geq \varepsilon + \frac{\lambda}{n}\mathbb{E}[\|XX^T\|]\right) \leq C e^{-cn},$$

with $C = 2C'$ and $c' = \min(c'/\varepsilon^2, c'/\varepsilon)$.

To prove the existence and uniqueness of Y on $\mathcal{A}_X$, note that $\forall A \in \mathcal{M}_{p,n}$ and $\forall y \in \mathbb{R}^p$:

$$d_2\Psi(A)|_y = \frac{1}{n} \sum_{i=1}^n f'(a_i^T y) a_i a_i^T = \frac{1}{n} ADA,$$

where $D \equiv \text{Diag}(f'(a_i^T y))_{1 \leq i \leq n}$ satisfies $\|D\| \leq \|f'\|_\infty \leq \lambda$. Thus, since $\forall A \in \mathcal{M}_{p,n}$, $\|\Psi(A)\|_{\mathcal{L}} = \sup_{y \in \mathbb{R}^p} \|d_2\Psi(A)|_y\|$, one sees that $\Psi(X)$ is $(1 - \varepsilon)$ -Lipschitz on $\mathcal{A}_Y$ ³¹, which implies the existence and uniqueness of Y thanks to Banach fixed point theorem.

4.2. Concentration of $y_{-i}^{X'}(t)$

The implicit statement that:

“ f does not scale with n and p ”.

induces some important features of f , among which we can mention:

- $|f(0)| \leq O(1)$,
- $\forall t > 0, f(t) \leq O(t + 1)$,

(since $f'(t) \leq \lambda \leq O(1)$). With those characteristics on f , we can already bound our fixed point:

Lemma 24. *Given $i \in [n]$, under $\mathcal{A}_Y$:*

$$\sup_{t \in [0,1]} \|y_{-i}^X(t)\| \leq O(1) \quad \text{and} \quad \sup_{t \in [0,1]} \|y_{-i}^{X'}(t)\| \leq O(1)$$

PROOF. We only prove the result for $t = 1$ i.e when $y_{-i}^X(t) = Y$, the general result is deduced the same way from our hypotheses. We know that under $\mathcal{A}_Y$:

$$\|\Psi(X)(0)\| \leq \frac{1}{n} |f(0)| \|X\| \leq O(1)$$

³¹Recall from the discussion preceding Definition 6 that we employ the same notation for the event $\mathcal{A}_Y$ and the subset $X(\mathcal{A}_Y) \subset \mathcal{M}_{p,n}$, when there are no ambiguities.

Therefore, Lemma 30 implies that $\forall k > 0$:

$$\begin{aligned} \|\Psi(X)^k(0)\| &\leq \sum_{i=1}^k \|\Psi(X)^i(0) - \Psi(X)^{i-1}(0)\| \\ &\leq \|\Psi(X)(0) - 0\| \sum_{i=1}^k \|\Psi(X)\|_{\mathcal{L}}^i \leq \frac{\|\Psi(X)(0)\|}{1 - \|\Psi(X)\|_{\mathcal{L}}} \leq O(1) \end{aligned}$$

since $\|\Psi(X)\|_{\mathcal{L}} < 1$. In particular, letting k tend to ∞ , one obtains $\|Y\| = \|\lim_{k \rightarrow \infty} \Psi(X)^k(0)\| \leq O(1)$.

We know from a slight modification of Identity (38) that $y_{-i}^{X'}(t)$ is solution to the fixed point equation:

$$y_{-i}^{X'}(t) = d_2 \Psi_{-i}^t(X) \Big|_{y_{-i}^X(t)} \cdot y_{-i}^{X'}(t) + \frac{1}{n} f(x_i^T y_{-i}^X(t)) x_i,$$

where, under $\mathcal{A}_Y$ (see Assumption 5 ter):

$$\begin{aligned} \left\| d_2 \Psi_{-i}^t(X) \Big|_{y_{-i}^X(t)} \right\| &= \left\| \frac{1}{n} \sum_{\substack{1 \leq j \leq n \\ j \neq i}} f'(x_j^T y_{-i}^X(t)) x_j x_j^T + \frac{t}{n} f'(x_i^T y_{-i}^X(t)) x_i x_i^T \right\| \\ &\leq \|f'\|_{\infty} \left\| \frac{1}{n} X X^T \right\| \leq 1 - \varepsilon \end{aligned}$$

like $\|\Psi(X)\|_{\mathcal{L}}$, which implies naturally (since $|f(x_i^T y_{-i}^X(t))| \leq |f(0)| + \|f'\|_{\infty} \|x_i^T y_{-i}^X(t)\|$):

$$\|y_{-i}^{X'}(t)\| \leq \frac{1}{\varepsilon} \left\| \frac{1}{n} f(x_i^T y_{-i}^X(t)) x_i \right\| \leq O\left(\frac{1}{\sqrt{n}} (|f(0)| + \|f'\|_{\infty} \|x_i^T y_{-i}^X(t)\|)\right) \leq O(1) \quad (39)$$

We have now all the elements to prove the concentration of $y_{-i}^X(\cdot)$.

Proposition 31. $\forall i \in [n]$:

$$\left(\frac{1}{\sqrt{n}} X, y_{-i}^X(\cdot) \right) \propto \mathcal{E}_2 \left(\frac{1}{\sqrt{n}} \right) \mid e^{-n} \quad \text{in} \quad \left(\mathcal{M}_{p,n} \times (\mathbb{R}^p)^{[0,1]}, \|\cdot\|_{\infty} \right),$$

where for any $A \in \mathcal{M}_{p,n}$ and $y(\cdot) \in (\mathbb{R}^p)^{[0,1]}$, $\|(A, y)\|_{\infty} = \|A\|_F + \sup_{t \in [0,1]} \|y(t)\|$.

PROOF. Let us note $y_0 = \mathbb{E}[y_{-i}^X(\cdot)] \in (\mathbb{R}^p)^{[0,1]}$ (see Appendix A for a definition of this expectation), we know from Lemma 24 that $\|y_0\|_{\infty} \leq O(1)$. We show in a first time that for any constant $K > 0$, and for all $y \in \mathcal{B}(y_0, K)$, the mappings $\Psi(\cdot)(y) : \mathcal{M}_{p,n} \rightarrow (\mathbb{R}^p, \|\cdot\|)$ is $O(1/\sqrt{n})$ -Lipschitz on $\mathcal{A}_X$ as the

mapping $\frac{1}{n}H_i(\cdot)(y) : \mathcal{M}_{p,n} \rightarrow (\mathbb{R}^p, \|\cdot\|)$ satisfying $\forall A = (a_1, \dots, a_n) \in \mathcal{M}_{p,n}$, $H_i(A)(y) = f(a_i^T y)a_i$. Indeed, for $A, B \in \mathcal{A}_X$, $\|A\|, \|B\|, \|y\| \leq O(1)$ and:

$$\begin{aligned} \|\Psi(A)(y) - \Psi(B)(y)\| &\leq \frac{1}{n} \|(A - B)f(A^T y)\| + \frac{1}{n} \|B(f(A^T y) - f(B^T y))\| \\ &\leq O\left(\frac{1}{\sqrt{n}}\right) \|A - B\| + \frac{\lambda}{n} \|B\| \|(A - B)^T y\| \\ &\leq O\left(\frac{1}{\sqrt{n}}\right) \|A - B\|, \end{aligned}$$

where $f(A^T y) \in \mathbb{R}^p$ is the vector having $f(a_1^T y), \dots, f(a_n^T y)$ as coordinates, f being independent with n and p and $\|f\|_{\mathcal{L}} \leq \lambda$ implies that $\|f(Ay)\| \leq \lambda \|Ay\| + \|f(0)\| \leq O(\sqrt{n})$. For the same reasons, for any $a, b \in \mathbb{R}^p$:

$$\frac{1}{n} \|H_i(a)(y) - H_i(b)(y)\| \leq O\left(\frac{1}{\sqrt{n}}\right) \|a - b\|,$$

Therefore, since, $\forall t \in [0, 1]$, $\Psi_{-i}^t = \Psi - \frac{1}{n}(1 - t)H_i$:

$$\|\Psi_{-i}(A)(y) - \Psi_{-i}(B)(y)\|_{\infty} \leq O\left(\frac{1}{\sqrt{n}}\right) \|A - B\|,$$

and naturally, $(X, \Psi_{-i}(X)(y)) \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n} \times (\mathbb{R}^p)^{[0,1]}, \|\cdot\|_{\infty})$. Hypotheses of Corollary 9 are then satisfied to set the concentration of $(X, \sqrt{n}y_{-i}^X(\cdot))$.

Introducing, for any $t \in [0, 1]$, and any $i \in [n]$, the diagonal matrix $D_{-i}(t) \in \mathcal{D}_n$ defined with

$$D_{-i}(t)_i = tf'(x_i^T y_{-i}^X(t)) \quad \text{and} \quad \forall j \in [n], j \neq i : D_{-i}(t)_j = f'(x_j^T y_{-i}^X(t))$$

the random matrix $d_2 \Psi_{-i}^t(A)|_y$ writes more simply $\frac{1}{n} X D_{-i}(t) X$ and we have the identity:

$$y_{-i}^{X'}(t) = \frac{1}{n} f(x_i^T y_{-i}^X(t)) Q_{-i}(t) x_i, \quad (40)$$

where, $Q_{-i}(t) \equiv Q(D_{-i}(t)) \equiv (I_p - \frac{1}{n} X D_{-i}(t) X^T)^{-1}$ (see subsection 3.8); Be careful that $Q_{-i}(t)$ is different from $Q_{-i}(D_{-i}(t))$! However, when $t = 0$, the random matrix $Q_{-i}(0)$, that we note Q_{-i} is then independent with x_i like $D_{-i} \equiv D_{-i}(0)$, since then $[D_{-i}]_i = 0$ and, in that case, $Q_{-i} = Q_{-i}(D_{-i})$.

The concentration given by Proposition 31 combined with the bounds on $\|X\|$ and $\|y_{-i}^X(t)\|$, respectively provided by Assumption 4 and Lemma 24 allows us to deduce from Proposition 9 the concentration:

$$X^T y_{-i}^X(\cdot) \propto \mathcal{E}_2 \mid e^{-n} \quad \text{in} \quad \left((\mathbb{R}^p)^{[0,1]}, (\|\cdot\|_{t \in [0,1]}) \right), \quad (41)$$

Here, a bound on $\|f\|_\infty$ as in Assumption 5 ter would be very useful to set the concentration of $y_{-i}^{X'}(\cdot)$, but it is not necessary, thus we rather introduce the quantity:

$$M_i \equiv \mathbb{E} \left[\sup_{t \in [0,1]} |x_i y_{-i}^X(t)| \right] \quad (42)$$

that will be shown to be of order $O(1)$ below. Till now, we just know that $M_i \leq O(\sqrt{n})$, from (41) that $x_i^T y_{-i}^X(t) \in O(M_i) \pm \mathcal{E}_2 \mid e^{-n}$ and from the bound

$$|f(x_i^T y_{-i}^X(t))| \leq |f(0)| + \|f'\|_\infty |x_i^T y_{-i}^X(t)|, \quad (43)$$

that $f(x_i^T y_{-i}^X(t)) \in O(M_i) \pm \mathcal{E}_2 \mid e^{-n}$ (since $\|f(0)\| \leq O(1)$ and $\|f'\|_\infty \leq O(1)$).

To be able to set the concentration of $Q(D_{-i}(t))x_i$ (the second term appearing in (40)), we need first to set the concentration of $D_{-i}(t)$ and assume, in place of Assumption 6:

Assumption 5 ter. $\|f''\|_\infty \leq \infty$.³²

Lemma 25. $\forall i \in [n]: D_{-i}(\cdot) \propto \mathcal{E}_2 \mid e^{-n}$ in $(\mathcal{D}_n^{[0,1]}, (\|\cdot\|_{F,t})_{t \in [0,1]})$.

PROOF. Since $\|f''\|_\infty \leq O(1)$, f' is $O(1)$ -Lipschitz. The concentration of $(\frac{1}{\sqrt{n}}X, y_{-i}^X(t))$ combined by the bounds $\mathbb{E}[\|X\|] \leq O(\sqrt{n})$ and $\mathbb{E}[\|y_{-i}^X(t)\|] \leq O(1)$ allows us to conclude from Proposition 9 that $X^T y_{-i}^X(t) \propto \mathcal{E}_2 \mid e^{-n}$. Therefore, as a $O(1)$ -Lipschitz transformation of $X^T y_{-i}^X(t)$, $D_{-i}(t) \propto \mathcal{E}_2 \mid e^{-n}$ (and the concentration constants do not depend on t).

Proposition 32. $\forall i \in [n]:$

$$y_{-i}^{X'}(\cdot) \in \mathcal{E}_2 \left(\frac{1}{\sqrt{n}} \right) \mid e^{-n} \quad \text{and} \quad X^T y_{-i}^{X'}(\cdot) \in \mathcal{E}_2 \mid e^{-n}$$

respectively in $((\mathbb{R}^p)^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]})$ and in $((\mathbb{R}^n)^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]})$.

PROOF. To set the concentration of $y_{-i}^{X'}(t)$, we just have to verify hypotheses of Theorem 5 with the equation: We can deduce from (40) that for any $t \in [0, 1]$ and any $i \in [n]$:

$$\begin{cases} y_{-i}^{X'}(t) = \frac{1}{n} f(x_i^T y_{-i}^X(t)) \sum_{k=1}^{\infty} \left(\frac{1}{n} X D_{-i}(t) X^T \right)^k x_i \\ X y_{-i}^{X'}(t) = \frac{1}{n} f(x_i^T y_{-i}^X(t)) \sum_{k=1}^{\infty} X \left(\frac{1}{n} X D_{-i}(t) X^T \right)^k x_i \end{cases}$$

³²Since f does not scale with p and n that implicitly means that $\|f\|_\infty, \|f''\|_\infty \leq O(1)$.

and, for any $k \in \mathbb{N}$, we know from Proposition 9 that:

$$\begin{cases} \frac{1}{n}f(x_i^T y_{-i}^X(t)) \left(\frac{1}{n}XD_{-i}(t)X^T \right)^k x_i \propto \mathcal{E}_2 \left(\frac{(1-\varepsilon)^k}{\sqrt{n}} \right) \mid e^{-n} \\ \frac{1}{n}f(x_i^T y_{-i}^X(t))X \left(\frac{1}{n}XD_{-i}(t)X^T \right)^k x_i \propto \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n} \end{cases}$$

indeed:

- $f(x_i^T y_{-i}^X(t)) \propto O(\sqrt{n}) \pm \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n}$ thanks to (43),
- $\frac{1}{n}XD_{-i}(t)X^T \propto \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n}$ thanks to Lemma 13 and Proposition 9, besides, under $\mathcal{A}_Y$, we know from Assumption 5 ter that $\|\frac{1}{n}XD_{-i}(t)X^T\| \leq 1-\varepsilon$,
- $x_i \propto \mathcal{E}_2$ thanks to Assumption 1.

We can then conclude thanks to Corollary 2 as in Theorem 5 (and the concentration constants do not depend on t).

4.3. Integration of $\frac{\partial y_{-i}^X(t)}{\partial t}$

Now that the concentration of the objects $y_{-i}^X(t)$ and $y_{-i}^X(t)'$ are well understood, we are able to integrate the formula provided by Lemma 22 to express the link between Y and Y_{-i} . We just some preliminary results to control the matrix $Q_{-i}(\cdot)$. In a first time, let us study $Q_{-i} \equiv Q_{-i}(0) = Q_{-i}(D_{-i})$ which is independent with x_i .

Proposition 33. *Given $i \in [n]$:*

$$\frac{1}{n}Q_{-i}x_i \in \mathcal{E}_2\left(\frac{1}{n}\right) \mid e^{-n} \quad \text{and} \quad \frac{1}{n}XQ_{-i}x_i \in \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$$

PROOF. As in the proof of Proposition 32 the concentration of:

- $Q_{-i}x_i = \sum_{k=1}^{\infty} \left(\frac{1}{n}X_{-i}D_{-i}X_{-i}^T \right)^k x_i$,
- $X_{-i}^T Q_{-i}x_i = \sum_{k=1}^{\infty} X_{-i} \left(\frac{1}{n}X_{-i}D_{-i}X_{-i}^T \right)^k x_i + x_i \left(\frac{1}{n}X_{-i}D_{-i}X_{-i}^T \right)^k x_i e_i$,

are consequences of Corollary 2. But this time, the independence between $\frac{1}{n}X_{-i}D_{-i}X_{-i}^T$ and x_i allows us to have better observable diameter. Indeed, if we note $W_{-i} \equiv \frac{1}{n}X_{-i}D_{-i}X_{-i}^T$, we know from Proposition 10 that $W_{-i}^k \in O((1-\varepsilon)^k) \pm \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n}$ and bounding, for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$,

$$\begin{aligned} & |u^T W_{-i}^k x_i - \mathbb{E}_{\mathcal{A}_Y}[u^T W_{-i}^k x_i]| \\ & \leq |u^T W_{-i}^k (x_i - \mathbb{E}_{\mathcal{A}_Y}[x_i])| + |u^T (W_{-i} - \mathbb{E}_{\mathcal{A}_Y}[W_{-i}^k]) \mathbb{E}_{\mathcal{A}_Y}[x_i]|, \end{aligned}$$

we can deduce that:

$$W_{-i}^k x_i \in \mathcal{E}_2((1-\varepsilon)^k) \mid e^{-n}.$$

the same way, we prove that $X_{-i}^T W_{-i}^k x_i \in \mathcal{E}_2(\sqrt{n}(1-\varepsilon)^k) \mid e^{-n}$ from the concentration $X_{-i} W_{-i}^k \in O(\sqrt{n}(1-\varepsilon)^k) \pm \mathcal{E}_2(\sqrt{n}(1-\varepsilon)^k) \mid e^{-n}$ and $x_i^T W_{-i}^k x_i \in \mathcal{E}_2(\sqrt{n}(1-\varepsilon)^k) \mid e^{-n}$ from the bound:

$$\begin{aligned} & |x_i^T W_{-i}^k x_i - \mathbb{E}_{\mathcal{A}_Y}[x_i^T W_{-i}^k x_i]| \\ & \leq |x_i^T W_{-i}^k x_i - \text{Tr}(\mathbb{E}_{\mathcal{A}_Y}[x_i x_i^T] W_{-i}^k)| + |\text{Tr}(\mathbb{E}_{\mathcal{A}_Y}[x_i x_i^T] (W_{-i} - \mathbb{E}_{\mathcal{A}_Y}[W_{-i}^k]))|. \end{aligned}$$

Then, Corollary 2 allows us to deduce the result of the proposition.

Proposition 34. *Given $i \in [n]$, recalling the notation M_i given in (42):*

$$\|Q_{-i}(\cdot)x_i - Q_{-i}x_i\| \in O(M_i) \pm \mathcal{E}_2(M_i) \mid e^{-n} \quad \text{in } (\mathbb{R}^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]})$$

PROOF. Now, under $\mathcal{A}_Y$ and for any $t \in [0, 1]$, let us bound:

$$\begin{aligned} \|(Q_{-i}(t) - Q_{-i})x_i\| & \leq \frac{1}{n} \|Q_{-i}(t) X_{-i} (D_{-i} - D(t)) X_{-i}^T Q_{-i} x_i\| \\ & \leq O\left(\frac{1}{\sqrt{n}}\right) \|X_{-i}^T Q_{-i} x_i\|_\infty \sup_{t \in [0,1]} \|D_{-i} - D_{-i}(t)\|_F, \end{aligned}$$

since $\|X\| \leq O(\sqrt{n})$ and $\|Q_{-i}(t)\| \leq \frac{1}{\varepsilon} (D_{-i} \equiv D_{-i}(0))$. We already saw that $X_{-i}^T Q_{-i} x_i / \sqrt{n} \in \mathcal{E}_2 \mid e^{-n}$ and, since, for all $j \neq i$, $|\mathbb{E}_{\mathcal{A}_X}[x_j^T Q_{-i} x_i]| = |\mathbb{E}_{\mathcal{A}_X}[x_j^T Q_{-i}] \mathbb{E}_{\mathcal{A}_X}[x_i]| \leq O(\sqrt{n})$, we can deduce from Proposition 8 that

$$\|X_{-i}^T Q_{-i} x_i\|_\infty / \sqrt{n} \in O(1) \pm \mathcal{E}_2 \mid e^{-n}.$$

From identities $D_{-i}(t) = \text{Diag}(f'(X^T y_{-i}^X(t)))$ and:

$$X^T y_{-i}^X(t) = \frac{1}{n} X^T X_{-i} f(X^T y_{-i}^X(t)) + \frac{t}{n} X^T x_i f(x_i^T y_{-i}^X(t)),$$

we can bound, under $\mathcal{A}_Y$ (recall that $\|X\|, \|x_i\| \leq O(\sqrt{n})$ and $\|f\|_\infty \leq O(1)$):

$$\begin{aligned} \|D_{-i} - D_{-i}(t)\|_F & \leq \|f''\|_\infty \|X^T y_{-i}^X(0) - X^T y_{-i}^X(t)\| \leq \frac{\|f''\|_\infty}{\varepsilon} \frac{t}{n} \|X^T x_i f(x_i^T y_{-i}^X(t))\| \\ & \leq O\left(f(0) + \|f'\|_\infty \sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)|\right) \in O(M_i) \pm \mathcal{E}_2 \mid e^{-n} \end{aligned}$$

which ends the proof (thanks again to Proposition 9).

Before giving the link between Y and Y_{-i} , we give a preliminary lemma that will be of multiple use in the following to invert equations.

Lemma 26. *Under $\mathcal{A}_Y$, $\forall i \in [n]$: $1 - \frac{\|f'\|_\infty}{n} x_i^T Q_{-i} x_i \geq O(1)$*

PROOF. Recalling that $[D_{-i}]_i = 0$ and for all $j \in [n]$ such that $j \neq i$, $[D_{-i}]_i = f'(x_j Y_{-i})$, we can bound:

$$\begin{aligned} Q_{-i} &\equiv Q(D_{-i}) \equiv \left(I_p - \frac{1}{n} X D_{-i} X^T \right)^{-1} = \left(I_p - \frac{1}{n} X_{-i} D_{-i} X_{-i}^T \right)^{-1} \\ &\leq \left(I_p - \frac{\|f'\|_\infty}{n} X_{-i} X_{-i}^T \right)^{-1} = Q_{-i}(\|f'\|_\infty I_n) \end{aligned}$$

The Schur identity (20) then implies that:

$$\frac{\|f'\|_\infty}{n} x_i^T Q_{-i}(t) x_i \leq \frac{\|f'\|_\infty}{n} x_i^T Q_{-i}(\|f'\|_\infty I_n) x_i = \frac{\frac{\|f'\|_\infty}{n} x_i^T Q(\|f'\|_\infty I_n) x_i}{1 + \frac{\|f'\|_\infty}{n} x_i^T Q(\|f'\|_\infty I_n) x_i}$$

Therefore, $1 - \frac{\|f'\|_\infty}{n} x_i^T Q_{-i} x_i \geq O(1)$ since $\frac{\|f'\|_\infty}{n} x_i^T Q(\|f'\|_\infty I_n) x_i \leq O(1)$.

Proposition 35. $\forall i \in [n]$, $M_i \equiv \mathbb{E}_{\mathcal{A}_Y} \left[\sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)| \right] \leq O(1)$ and:

$$\left\| Y - Y_{-i} - \frac{1}{n} f(x_i^T Y) Q_{-i} x_i \right\| \in 0 \pm \mathcal{E}_2 \left(\frac{1}{n} \right) \mid e^{-n}$$

Then, in particular, the bound $M_i \leq O(1)$ implies (thanks to Proposition 31 and (43)):

$$x_i^T y_{-i}(t) \in O(1) \pm \mathcal{E}_2 \mid e^{-n} \quad \text{and} \quad f(x_i^T y_{-i}(t)) \in O(1) \pm \mathcal{E}_2 \mid e^{-n} \quad (44)$$

PROOF. Setting $\chi(t) \equiv t f(x_i^T y_{-i}^X(t)) \in \mathbb{R}$, let us integrate between 0 and t the identity $y_{-i}^{X'}(t) = \chi'(t) \frac{1}{n} Q_{-i}(t) x_i$:

$$y_{-i}^X(t) - Y_{-i} = \frac{1}{n} f(x_i^T Y) Q_{-i} x_i + \frac{1}{n} \int_0^t \chi'(u) (Q_{-i}(u) - Q_{-i}(0)) x_i du. \quad (45)$$

Now, $\chi'(u) = f(x_i^T y_{-i}^X(u)) + t f'(x_i^T y_{-i}^X(u)) x_i^T y_{-i}^{X'}(u)$, and satisfies the concentration

$$\chi'(\cdot) \in O(M_i) \pm \mathcal{E}_2(M_i) \mid e^{-n} \quad \text{in} \quad \left(\mathbb{R}^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]} \right)$$

(thanks to the bound $\|y_{-i}^{X'}(u)\| \leq O(1)$ given in Lemma 24 that implies the concentration $x_i^T y_{-i}^{X'}(u) \in O(M_i) \pm \mathcal{E}_2 \mid e^{-n}$ and thanks to the discussion around (43)). Besides, Proposition 9 combined with Proposition 34 imply that:

$$\|\chi'(\cdot) (Q_{-i}(u) - Q_{-i}(0)) x_i\| \leq |\chi'(\cdot)| \| (Q_{-i}(\cdot) - Q_{-i}(0)) x_i \| \in O(M_i^2) \pm \mathcal{E}_2(M_i^2) \mid e^{-n},$$

in $(\mathbb{R}^{[0,1]}, (\|\cdot\|_t)_{t \in [0,1]})$. We can then deduce the result from Proposition 6 that:

$$\left\| \frac{1}{n} \int_0^t \chi'(u) (Q_{-i}(u) - Q_{-i}(0)) x_i du \right\| \in O(M_i^2) \pm \mathcal{E}_2(M_i^2) \mid e^{-n}$$

and multiplying (45) on the right by x_i^T , one obtains thanks to (43):

$$\begin{aligned} \sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)| &\leq |x_i^T Y_{-i}| + \frac{|f(0)|}{n} x_i^T Q_{-i} x_i + \frac{\|f'\|_\infty}{n} x_i^T Q_{-i} x_i \sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)| \\ &\quad + \left\| \frac{1}{n} \int_0^t \chi'(u) (Q_{-i}(u) - Q_{-i}(0)) x_i du \right\|. \end{aligned}$$

We can then regroup all the terms $\sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)|$ on the left side of the inequality thanks to the bound $1 - \frac{\|f'\|_\infty}{n} x_i^T Q_{-i} x_i \geq O(1)$ provided by Lemma 26. Taking the expectation, we infer that:

$$\begin{aligned} M_i &\equiv \mathbb{E}_{\mathcal{A}_Y} \left[\sup_{t \in [0,1]} |x_i^T y_{-i}^X(t)| \right] \\ &\leq O \left(1 + \mathbb{E}_{\mathcal{A}_Y} \left[\left\| \frac{1}{n} \int_0^t \chi'(u) x_i^T (Q_{-i}(u) - Q_{-i}(0)) x_i du \right\| \right] \right) \leq O(1) + O \left(\frac{M_i^2}{\sqrt{n}} \right), \end{aligned}$$

which directly implies that $M_i = O(1)$ and the second result of the proposition (taking $t = 1$ in (45)).

With the formalism introduced in Subsection 1.2 (before Lemma 3) we say that $Y_{-i} - \frac{1}{n} f(x_i^T Y) Q_{-i} x_i$ is a *strong random equivalent* of Y , in particular, we saw in Remark 6 that it has almost the same expectation and covariance, but since we can bound:

$$\begin{aligned} \left\| \mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n} f(x_i^T Y) Q_{-i} x_i \right] \right\| &\leq \frac{1}{n} \mathbb{E}_{\mathcal{A}_Y} [|f(x_i^T Y)| \|x_i\| \|Q_{-i}\|] \leq O \left(\frac{1}{\sqrt{n}} \right) \\ \left\| \mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n^2} f(x_i^T Y)^2 Q_{-i} x_i x_i^T Q_{-i} \right] \right\|_* &\leq \frac{1}{n^2} \mathbb{E}_{\mathcal{A}_Y} [|f(x_i^T Y)|^2 \|x_i\|^2 \|Q_{-i}\|^2] \leq O \left(\frac{1}{n} \right) \\ \left\| \mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n} f(x_i^T Y)^2 Q_{-i} x_i Y_{-i}^T \right] \right\|_* &\leq O \left(\frac{1}{\sqrt{n}} \right), \end{aligned}$$

we can also estimate:

$$\begin{cases} \|\mathbb{E}_{\mathcal{A}_Y} [Y] - \mathbb{E}_{\mathcal{A}_Y} [Y_{-i}]\| \leq O \left(\frac{1}{\sqrt{n}} \right) \\ \|\mathbb{E}_{\mathcal{A}_Y} [Y Y^T] - \mathbb{E}_{\mathcal{A}_Y} [Y_{-i} Y_{-i}^T]\|_* \leq O \left(\frac{1}{\sqrt{n}} \right). \end{cases} \quad (46)$$

One can then wonder why we set a result as complex as Proposition 35 if it was simply to obtain these simple relations between the first statistics of Y and Y_{-i} . Of course, we are going to go further.

The observable diameter of order $O(1/n)$ in Proposition 35 allows us to keep good concentration bounds when Y is multiplied on the left by x_j^T , $j \in [n]$ (indeed, under $\mathcal{A}_Y$, $\|x_j\| \leq O(\sqrt{n})$). This time, the term $\frac{1}{n} f(x_i^T Y) x_j Q_{-i} x_i$ can

be of order $O(1)$ in particular when $j = i$. For all $j \in [n]$, $\frac{1}{n}f(x_i^T Y)x_j Q_{-i}x_i \in \mathcal{E}_2(1/\sqrt{n})$ (see (44) and Proposition 33) thus if $j \neq i$:

$$\mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n}f(x_i^T Y)x_j Q_{-i}x_i \right] \leq O\left(\frac{1}{\sqrt{n}}\right)$$

but when $j = i$ this quantity can be of order $O(1)$, therefore we obtain the concentrations:

$$\begin{cases} x_j^T Y \in x_j^T Y_{-i} \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n} & \text{when } j \neq i \\ x_i^T Y \in x_i^T Y_{-i} + \frac{1}{n}x_i^T Q_{-i}x_i f(x_i^T Y) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}. \end{cases} \quad (47)$$

4.4. Implicit relation between $x_i^T Y$ and $x_i^T Y_{-i}$

The random equivalent of $x_i Y$ in (47) interests us particularly because it allows us to replace in the identity

$$Y = \frac{1}{n} \sum_{i=1}^n f(x_i^T Y)x_i,$$

the quantity $f(x_i^T Y)x_i$ with a quantity $f(\zeta_i(x_i^T Y_{-i}))x_i$ (for a given mapping $\zeta_i : \mathbb{R} \rightarrow \mathbb{R}$) that is more easy to manage thanks to the independence between x_i and Y_{-i} . For all $i \in [n]$, let us introduce the deterministic and easily computable diagonal matrix

$$\Delta \equiv \text{Diag}_{1 \leq i \leq n} \left(\frac{1}{n} \text{Tr}(\Sigma_i) \right).$$

We know from Proposition 27 that $Q_{-i} \in I_p \pm \mathcal{E}_2(\sqrt{\log n}) \mid e^{-n}$, and, since $\|\frac{1}{n}\Sigma_i\|_F \leq O(1/\sqrt{n})$, Proposition 33 implies that:

$$\begin{aligned} \frac{1}{n}x_i Q_{-i}x_i &\in \frac{1}{n} \text{Tr}(\Sigma_i Q_{-i}) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n} \\ &\in \Delta_i \pm \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right) \mid e^{-n} \end{aligned}$$

It sounds then natural to introduce the equation

$$z = x_i^T Y_{-i} + \Delta_i f(z), \quad z \in \mathbb{R} \quad (48)$$

whose solution should be close to $x_i^T Y$ as stated by next proposition.

Proposition 36. *Given $i \in [n]$, the random equation:*

$$z = x_i^T Y_{-i} + \Delta_i f(z), \quad z \in \mathbb{R}, \quad (49)$$

admits a unique solution that we note $\zeta_i(x_i^T Y_{-i})$ and that satisfies:

$$x_i^T Y \in \zeta_i(x_i^T Y_{-i}) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$$

PROOF. Let us differentiate the mapping $g : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $\forall z \in \mathbb{R}, g(z) = x_i^T Y_{-i} + \Delta_i f(z)$:

$$g'(z) = \Delta_i f'(z) \leq \mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n} x_i^T Q_{-i} x_i \|f'\|_\infty \right] < 1,$$

thanks to Lemma 26. Thus, the mapping g is contraction and it admits a unique solution $\zeta(x_i^T Y_{-i})$. Now, we can bound under $\mathcal{A}_Y$:

$$\begin{aligned} |x_i^T Y - \zeta(x_i^T Y_{-i})| &\leq |x_i^T Y - x_i^T Y_{-i} - \Delta_i f(\zeta(x_i^T Y_{-i}))| \\ &\leq |x_i^T Y - x_i^T Y_{-i} - \Delta_i f(x_i^T Y)| + \Delta_i |f(x_i^T Y) - f(\zeta(x_i^T Y_{-i}))| \\ &\leq \|f'\| \Delta_i |x_i^T Y - \zeta_i(x_i^T Y_{-i})| + O\left(\frac{1}{\sqrt{n}}\right) \leq O\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

thanks to (47) and Lemma 26. We conclude then with Lemma 3.

We end this subsection with a little result that will allow us to differentiate ζ_i .

Lemma 27. *Given $i \in [n]$, the mapping ζ_i is differentiable.*

PROOF. Considering $z, t \in \mathbb{R}$, let us express:

$$\zeta_i(z+t) - \zeta_i(z) = t + \Delta_i (f(\zeta_i(z+t)) - f(\zeta_i(z)))$$

thus $|\zeta_i(z+t) - \zeta_i(z)| \leq \frac{t}{1 - \Delta_i \|f'\|_\infty}$ (note that it implies that ζ_i is continuous). Let us bound:

$$\begin{aligned} |f(\zeta_i(z+t)) - f(\zeta_i(z)) - f'(\zeta_i(z))(\zeta_i(z+t) - \zeta_i(z))| \\ \leq \|f''\|_\infty |\zeta_i(z+t) - \zeta_i(z)|^2 \leq \frac{t^2 \|f''\|_\infty}{(1 - \Delta_i \|f'\|_\infty)^2} \end{aligned}$$

Dividing the upper identity by t we can bound:

$$\left| \frac{1}{t} (\zeta_i(z+t) - \zeta_i(z)) - 1 - \frac{\Delta_i}{t} f'(\zeta_i(z))(\zeta_i(z+t) - \zeta_i(z)) \right| \leq \frac{t \|f''\|_\infty \Delta_i}{(1 - \Delta_i \|f'\|_\infty)^2}.$$

We can then let t tend to 0 to conclude that ζ_i is differentiable and we obtain the identity:

$$\zeta'_i(t) = 1 + \Delta_i f'(\zeta_i(t)) \zeta'_i(t).$$

We can then induce from the formulation of ζ_i that $\forall z \in \mathbb{R}$:

$$\zeta'(z) = \frac{1}{1 - \Delta_i f'(\zeta_i(z))} \leq O(1), \quad (50)$$

thanks to Lemma 26.

4.5. Expression of the mean and covariance of Y .

Let us introduce the random vector:

$$\check{Z} \equiv (x_i^T Y_{-i})_{1 \leq i \leq n}$$

and the mappings $\zeta : \mathbb{R}^n \ni (z_i)_{1 \leq i \leq n} \mapsto (\zeta_i(z_i))_{1 \leq i \leq n} \in \mathbb{R}^n$, $\xi = f \circ \zeta$ ($\forall i \in [n]$, $\xi_i = f \circ \zeta_i$). With those notations, next Proposition gives us first estimations of the deterministic objects:

$$m_Y \equiv \mathbb{E}_{\mathcal{A}_Y}[Y] \quad \text{and} \quad \Sigma_Y \equiv \mathbb{E}_{\mathcal{A}_Y}[Y Y^T],$$

Proposition 37. $\|Y - \frac{1}{n} X \xi(\check{Z})\| \in O\left(\sqrt{\frac{\log n}{n}}\right) + \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right) \mid e^{-n}$ and we can estimate:

$$\begin{cases} \left\| m_Y - \mathbb{E}_{\mathcal{A}_\zeta} \left[\frac{1}{n} X \xi(\check{Z}) \right] \right\| \leq O\left(\sqrt{\frac{\log n}{n}}\right) \\ \left\| \Sigma_Y - \mathbb{E}_{\mathcal{A}_\zeta} \left[\frac{1}{n^2} X \xi(\check{Z}) \xi(\check{Z})^T X^T \right] \right\|_* \leq O\left(\sqrt{\frac{\log n}{n}}\right) \end{cases}$$

PROOF. Let us bound under $\mathcal{A}_\zeta$:

$$\begin{aligned} \left\| Y - \frac{1}{n} X \xi(\check{Z}) \right\| &\leq \left\| \frac{1}{n} X f(Z) - \frac{1}{n} X f(\zeta(\check{Z})) \right\| \\ &\leq O\left(\sup_{1 \leq i \leq n} |x_i^T Y - \zeta_i(x_i^T Y_{-i})| \right) \end{aligned}$$

Besides, we know from Propositions 36 and 5 that $(|x_i^T Y - \zeta_i(x_i^T Y_{-i})|)_{1 \leq i \leq n} \in \mathcal{E}_2\left(\sqrt{\frac{1}{n}}\right) \mid e^{-n}$ in $(\mathbb{R}^n, \|\cdot\|_\infty)$, thus Proposition 8 implies that:

$$\sup_{1 \leq i \leq n} |x_i^T Y - \zeta_i(x_i^T Y_{-i})| \in O\left(\sqrt{\frac{\log n}{n}}\right) + \mathcal{E}_2\left(\sqrt{\frac{\log n}{n}}\right) \mid e^{-n},$$

from which we deduce the first result of the proposition. The estimation of the expectation and the non-centered covariance of Y then follow from Remark 6.

For any $u \in \mathbb{R}^p$ and any $A \in \mathcal{M}_p$ such that $\|u\|, \|A\| \leq 1$:

$$u^T m_Y = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{-i} \mathbb{E}_i [\xi_i(x_i^T Y_{-i}) u^T x_i \mid \mathcal{A}_Y] + O\left(\sqrt{\frac{\log n}{n}}\right) \quad (51)$$

$$\text{Tr}(\Sigma_Y A) = \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \mathbb{E}_{-i, j} \mathbb{E}_{i, j} [\xi_i(x_i^T Y_{-i}) \xi_j(x_j^T Y_{-j}) x_i^T A x_j \mid \mathcal{A}_Y] + O\left(\sqrt{\frac{\log n}{n}}\right),$$

where $\mathbb{E}_i$ integrates only on the variable x_i and:

$$\mathbb{E}_{i, j} = \mathbb{E}_j \circ \mathbb{E}_i = \mathbb{E}_i \circ \mathbb{E}_j; \quad \mathbb{E}_{-i} = \prod_{\substack{1 \leq j \leq n \\ j \neq i}} \mathbb{E}_j \quad \text{and} \quad \mathbb{E}_{-i, j} = \prod_{\substack{1 \leq k \leq n \\ k \neq i, j}} \mathbb{E}_k.$$

4.6. Computation of the estimation of the mean and covariance of Y when X is Gaussian

The estimation given by (51) becomes particularly interesting when X is Gaussian because in that case, the random variable $z_i \equiv x_i^T Y_{-i}$ is also Gaussian (when all the random vectors x_j are fixed, for $j \neq i$) and admits the statistics:

$$\mathbb{E}_i[z_i] = m_i^T Y_{-i} \quad \text{and} \quad \mathbb{E}_i[z_i^2] = Y_{-i}^T \Sigma_i Y_{-i}$$

(where we recall that $m_i \equiv \mathbb{E}[x_i]$ and $\Sigma_i \equiv \mathbb{E}[x_i x_i^T]$). We would like to obtain from (51) fixed point equation on m_Y and Σ_Y , for that purpose, we need to express $\mathbb{E}_i [\xi(z_j) m_i^T x_i \mid \mathcal{A}_Y]$ and $\mathbb{E}_{j,k} [\xi(z_j) \xi(z_k) x_j^T \Sigma_i x_k \mid \mathcal{A}_Y]$ as functions of m_Y , C_Y , m_i , Σ_i , $i \in [n]$. Note indeed that $z_j \in x_j^T Y_{-j} \pm \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ (for the same reasons that led to (47)), we can deduce that:

$$\begin{aligned} & \mathbb{E}_{j,k} [\xi_j(x_j^T Y_{-j}) \xi_k(x_k^T Y_{-k}) x_j^T A x_k \mid \mathcal{A}_Y] \\ &= \mathbb{E}_{j,k} \left[\xi_j(x_j^T Y_{-j}) x_j \mid \mathcal{A}_Y \right]^T A \mathbb{E}_{j,k} \left[\xi_k(x_k^T Y_{-k}) x_k \mid \mathcal{A}_Y \right] + O\left(\frac{1}{\sqrt{n}}\right) \\ &= \mathbb{E}_j \left[\xi_j(x_j^T Y_{-j}) x_j \mid \mathcal{A}_Y \right]^T A \mathbb{E}_k \left[\xi_k(x_k^T Y_{-k}) x_k \mid \mathcal{A}_Y \right] + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

The only quantities that we want to explicit are thus of the form $\mathbb{E}_j [\xi(z_j) u^T x_j \mid \mathcal{A}_Y]$ and $\mathbb{E}_j [\xi(z_j)^2 x_j^T A x_j \mid \mathcal{A}_Y]$. It will be done in two steps, constituted by the two next propositions:

1. “separate” with Stein-like identities, the “functional part” ($\xi(z_j)$ and $\xi(z_j)^2$) from the “vectorial part” ($u^T x_j$ and $x_j^T A x_j$) in $\mathbb{E}_j [\xi(z_j) u^T x_j \mid \mathcal{A}_Y]$ and $\mathbb{E}_j [\xi(z_j)^2 x_j^T A x_j \mid \mathcal{A}_Y]$,
2. show that for a given functional $h : \mathbb{R} \rightarrow \mathbb{R}$, the law of $h(z_i)$ just depends on:

$$\mu_i \equiv m_i^T m_Y \quad \text{and} \quad \nu_i \equiv \text{Tr}(\Sigma_Y \Sigma_i), \quad i \in [n]. \quad (52)$$

First, we give a result characteristic to Gaussian vectors commonly called “Stein’s identities”:

Proposition 38. *Given a Gaussian vector $x \sim \mathcal{N}(\mu, C)$ for $\mu \in \mathbb{R}^p$ and $C \in \mathcal{M}_p$ positive symmetric, two deterministic vectors $w, u \in \mathbb{R}^p$, and a deterministic matrix $A \in \mathcal{M}_{p,n}$, we have the identities:*

$$\begin{aligned} \mathbb{E}[f(w^T x) u^T x] &= \mathbb{E}[f(w^T x)] u^T \mu + \mathbb{E}[f'(w^T x)] u^T C w \\ \mathbb{E}[f(w^T x) x^T A x] &= \mathbb{E}[f(w^T x)] \text{Tr}(A(\mu \mu^T + C)) + \mathbb{E}[f'(w^T x)] w^T (A + A^T) \mu \\ &\quad + \mathbb{E}[f''(w^T x)] w^T C A C w \end{aligned}$$

Therefore, for $j \in [n]$, we can express for any $u \in \mathbb{R}^p$, $A \in \mathcal{M}_p$ symmetric, such that $\|u\|, \|A\| \leq O(1)$ and under $\mathcal{A}_\zeta$ (for the drawing of Y_{-j} that is not integrated here):

- $\mathbb{E}_i [\xi_j(z_i) u^T x_i \mid \mathcal{A}_Y] = \mathbb{E}_i [\xi_i(z_i) u^T x_i] + O\left(\frac{1}{\sqrt{n}}\right)$
 $= \mathbb{E}_i[\xi_i(z_i)] u^T m_i + \mathbb{E}_i[\xi'_i(z_i)] u^T C_i Y_{-i}$
- $\frac{1}{n} \mathbb{E}_j [\xi_j(z_j)^2 x_j^T A x_j \mid \mathcal{A}_Y] = \frac{1}{n} \mathbb{E}_j [\xi_j(z_j)^2 x_j^T A x_j] + O\left(\frac{1}{\sqrt{n}}\right)$
 $= \frac{1}{n} \mathbb{E}_j [\xi_j(z_j)^2] \text{Tr}(A \Sigma_j) + \frac{4}{n} \mathbb{E}_j [\xi_j(z_j) \xi'_j(z_j)] m_j^T A C_j Y_{-j}$
 $+ \frac{2}{n} \mathbb{E}_j [\xi_j(z_j) \xi''_j(z_j) + \xi'_j(z_j)^2] Y_{-j}^T C_j A C_j Y_{-j} + O\left(\frac{1}{\sqrt{n}}\right)$
 $= \frac{1}{n} \mathbb{E}_j [\xi_j(z_j)^2] \text{Tr}(A \Sigma_j) + O\left(\frac{1}{\sqrt{n}}\right)$
(since $|\frac{1}{n} m_j^T A C_j Y_{-j}| = O(1/n)$ and $\frac{1}{n} |Y_{-j}^T C_j A C_j Y_{-j}| \leq O(1/n)$)
- $\mathbb{E}_{j,k}[\xi_j(x_j^T Y_{-j}) \xi_k(x_k^T Y_{-k}) x_j^T A x_k \mid \mathcal{A}_Y]$
 $= (\mathbb{E}_j[\xi_j(z_j)] m_j^T + \mathbb{E}_j[\xi'_j(z_j)] Y_{-j}^T C_j)^T A (\mathbb{E}_k[\xi_k(z_k)] m_k + \mathbb{E}_k[\xi'_k(z_k)] C_k Y_{-k}) + O\left(\frac{1}{\sqrt{n}}\right).$

We know from Proposition 31 (that implies in particular that $Y_{-j} Y_{-k}^T \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ in $(\mathcal{M}_p, \|\cdot\|_*)$) thanks to (46) that:

- $u^T C_i Y_{-i} \in u^T C_i m_Y \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$
- $m_j^T A C_k Y_{-k} \in m_j^T A C_k m_Y \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$
- $Y_{-j}^T C_j A C_k Y_{-k} = \text{Tr}(Y_{-k} Y_{-j}^T C_j A C_k) \in \text{Tr}(\Sigma_Y C_j A C_k) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n}$

we are thus left to express the expectations of functionals of z_j as functions of m_Y and Σ_Y .

Next proposition provides a way to manage the randomness brought in the expressions of $\mathbb{E}_i[z_i] = m_i^T Y_{-i}$ and $\mathbb{E}_i[z_i^2] = Y_{-i}^T \Sigma_i Y_{-i}$ by Y_{-i} to retrieve μ_i and ν_i (defined in (52)).

Proposition 39. *Given two (sequences of) random variables $\mu \in \mathbb{R}$ and $\nu \in \mathbb{R}$, two (sequences of) deterministic variable $\tilde{\mu} \in \mathbb{R}$ and $\tilde{\nu} > 0$ such that $O(1) \leq \tilde{\mu}, \tilde{\nu} \leq O(1)$ such that:*

$$\mu \in \tilde{\mu} \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n} \quad \text{and} \quad \nu \in \tilde{\nu} \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n},$$

if we consider a differentiable mapping $f : \mathbb{R} \rightarrow \mathbb{R}$ not scaling with n and such that for any parameter $a, b \in \mathbb{R}$: $\lim_{y \rightarrow \pm\infty} |f(a + by)| |y| e^{-y^2/2} = 0$ ³³, then for

³³Since f does not scales with infinity, that implicitly induce that if $O(1) \leq a, b \leq O(1)$, then $\sup_{y \in \mathbb{R}} |f(a + by)| |y| e^{-y^2/2} \leq O(1)$.

any Gaussian random variable $z \sim \mathcal{N}(\mu, \nu)$, independent with ν and μ , we can estimate:

$$\mathbb{E}_z[f(z)] \in \mathbb{E}[f(\tilde{z})] \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) \mid e^{-n},$$

where $\mathbb{E}_z$ is the expectation taken only on the variation of z and $\tilde{z} \sim \mathcal{N}(\tilde{\mu}, \tilde{\nu})$.

PROOF. Let us introduce a Gaussian random variable $y \sim \mathcal{N}(0, 1)$, independent with μ and ν . We can express:

$$\mathbb{E}_z[f(z)] \in \mathbb{E}[f(\tilde{z})] = \mathbb{E}_y[f(\mu + \sqrt{\nu}y)] \equiv \phi(\mu, \nu)$$

Before bounding the variations of ϕ , note that there exists two constants $C, c > 0$ such that, introducing the event $\mathcal{A}_{\mu, \nu} \equiv \{\frac{\tilde{\mu}}{2} \leq \mu \leq 2\tilde{\mu} \text{ and } \frac{\tilde{\nu}}{2} \leq \nu \leq 2\tilde{\nu}\}$, we can bound $\mathbb{P}(\mathcal{A}_{\mu, \nu}^c) \leq Ce^{-cn}$. Then $(\sqrt{\nu}|\mathcal{A}_{\mu, \nu}) \propto \mathcal{E}_2(1/\sqrt{n})$ and under $\mathcal{A}_{\mu, \nu}$, the mapping $y \mapsto f(\mu + \sqrt{\nu}y)e^{-y^2/2}$ is bounded, we can thus differentiate ϕ :

$$\frac{\partial \phi}{\partial \nu} = \mathbb{E}_y[\sqrt{\nu}f'(\mu + \sqrt{\nu}y)] = \mathbb{E}_y[yf(\mu + \sqrt{\nu}y)] \leq \sup_{\substack{a \in [\tilde{\mu}/2, 2\tilde{\mu}] \\ b \in [\sqrt{\tilde{\nu}}/2, \sqrt{2\tilde{\nu}}]}} \mathbb{E}_y[yf(a + by)] \leq O(1)$$

$$\frac{\partial \phi}{\partial \mu} = \mathbb{E}_y[f'(\mu + \sqrt{\nu}y)] = \mathbb{E}_y\left[\frac{y}{\sqrt{\nu}}f(\mu + \sqrt{\nu}y)\right] \leq O(1)$$

Therefore as $O(1)$ -Lipschitz transformations of μ, ν under $\mathcal{A}_{\mu, \nu}$, we obtain the concentration $\phi(\mu, \nu) \in \phi(\tilde{\mu}, \tilde{\nu}) \pm \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ (see Remark 7), which is exactly the result of the proposition.

4.7. *Computation of $\mu_j \equiv m_j^T m_Y$ and $\nu_j \equiv \text{Tr}(\Sigma_j^T \Sigma_Y)$, $j \in [n]$*

To employ proposition 39, let us introduce the mapping $E : \mathcal{F}(\mathbb{R}) \times \mathbb{R} \times \mathbb{R}$ defined for any $f : \mathbb{R} \rightarrow \mathbb{R}$ and $\mu, \nu \in \mathbb{R}$ as:

$$E(f, \mu, \nu) = \mathbb{E}[f(z)] \quad \text{with } z \sim \mathcal{N}(\mu, \nu - \mu^2).$$

Then, for instance $\mathbb{E}_i[\xi_j(z_j)] = E(\xi_j, \mu_j, \nu_j) \leq O(1/\sqrt{n})$, and we can rewrite (51) to obtain for any $u \in \mathbb{R}^p$ such that $\|u\| \leq O(1)$:

$$u^T m_Y = \frac{1}{n} \sum_{i=1}^n E(\xi_j, \mu_j, \nu_j) u^T m_i + E(\xi'_j, \mu_j, \nu_j) u^T C_i m_Y + O\left(\sqrt{\frac{\log n}{n}}\right),$$

To obtain a quasi-fixed point formulation on $(\mu_j)_{j \in [n]}$ (and without apparition of m_Y alone), we set:

$$\tilde{C} \equiv \frac{1}{n} \sum_{i=1}^n E(\xi'_j, \mu_j, \nu_j) C_j \quad \text{and} \quad \tilde{m} \equiv \frac{1}{n} \sum_{i=1}^n E(\xi_j, \mu_j, \nu_j) m_j,$$

then m_Y satisfies the pseudo equation $u^T m_Y = u^T \tilde{m} + u^T \tilde{C} m_Y$ and we need the following result to be able to invert $(I_p - \tilde{C})$:

Lemma 28. $\left\| \frac{1}{n} \sum_{i=1}^n E(\xi'_j, \mu_j, \nu_j) \Sigma_i \right\| < 1 - \frac{\varepsilon}{2}$

PROOF. Given $j \in [n]$, let us express with a random variable $z \sim \mathcal{N}(\mu_j, \nu_j)$:

$$E(\xi'_j, \mu_j, \nu_j) = \mathbb{E}_{\mathcal{A}_\zeta} [f'(\zeta_j(z)) \zeta'_j(z)],$$

and we know from (50) that $\zeta'_j(z) = \frac{1}{1 - \Delta_j f'(\zeta_j(z))}$ and we can bound:

$$|E(\xi'_j, \mu_j, \nu_j)| = \left\| \mathbb{E}_{\mathcal{A}_\zeta} \left[\frac{f'(\zeta_j(z))}{1 - \Delta_j f'(\zeta_j(z))} \right] \right\| \leq \frac{\|f'\|_\infty}{1 - \|f'\|_\infty \Delta_j}.$$

Now, Proposition ?? implies that

$$\Delta_j \leq \mathbb{E}_{\mathcal{A}_Y} \left[\frac{1}{n} \text{Tr}(\Sigma_j Q_{-j}(\|f'\|_\infty I_n)) \right] = \Lambda^1(\|f'\|_\infty I_n) + O\left(\sqrt{\frac{\log n}{n}}\right)$$

where $\Lambda^1(\|f'\|_\infty I_n) = \frac{1}{n} \text{Tr} \left(\Sigma_j \tilde{Q}^z \left(\tilde{\|f'\|_\infty} \cdot \Lambda^z(\|f'\|_\infty I_n) \right) \right)$ and:

$$\tilde{Q}^1(\|f'\|_\infty \Lambda^1(\|f'\|_\infty I_n)) \equiv \left(I_p - \frac{1}{n} \sum_{i=1}^n \frac{\|f'\|_\infty \Sigma_i}{1 - \|f'\|_\infty \Lambda^1(\|f'\|_\infty I_n)} \right)^{-1}.$$

We know that:

$$\left\| \tilde{Q}^1(\|f'\|_\infty I_n) \right\| \in \left\| \tilde{Q}^1(\|f'\|_\infty \Lambda^1(\|f'\|_\infty I_n)) \right\| \pm \mathcal{E}_2 \left(\frac{1}{\sqrt{n}} \right) \mid e^{-n},$$

and $\left\| \tilde{Q}^1(\|f'\|_\infty I_n) \right\| \leq \frac{1}{\varepsilon}$ under $\mathcal{A}_Y$. Thus inverting $\tilde{Q}^1(\|f'\|_\infty \Lambda^1(\|f'\|_\infty I_n))$, we obtain the bound:

$$\left\| \frac{1}{n} \sum_{i=1}^n E(\xi'_j, \mu_j, \nu_j) \Sigma_i \right\| \leq 1 - \frac{1}{\left\| \tilde{Q}^1(\|f'\|_\infty \Lambda^1(\|f'\|_\infty I_n)) \right\|} \leq 1 - \varepsilon + O\left(\frac{1}{\sqrt{n}}\right)$$

Therefore, since $\Sigma_i = C_i + m_i m_i^T$, $\|C\| \leq \left\| \frac{1}{n} \sum_{i=1}^n E(\xi'_j, \mu_j, \nu_j) \Sigma_i \right\| < 1$, we can invert $(I_p - \tilde{C})$ and, replacing u by $(I_p - \tilde{C})^{-1} m_k$, we obtain the system of n quasi-equations on $(\mu_k)_{1 \leq k \leq n}$:

$$\forall k \in [n] : \mu_k = m_k^T (I_p - \tilde{C})^{-1} \tilde{m} + O\left(\sqrt{\frac{\log n}{n}}\right)$$

Let us express from (51), for any symmetric matrix $A \in \mathcal{M}_p$ such that

$\|A\| \leq O(1)$:

$$\begin{aligned}
\text{Tr}(\Sigma_Y A) &= \frac{1}{n^2} \sum_{i=1}^n E(\xi_i^2, \mu_i, \nu_i) \text{Tr}(A \Sigma_i) \\
&\quad + \frac{1}{n^2} \sum_{1 \leq j \neq k \leq n} E(\xi_k, \mu_k, \nu_k) E(\xi_j, \mu_j, \nu_j) m_j^T A m_k \\
&\quad + 2 \frac{1}{n^2} \sum_{1 \leq j \neq k \leq n} E(\xi_j, \mu_j, \nu_j) E(\xi'_k, \mu_k, \nu_k) m_j^T A C_k m_Y \\
&\quad + \frac{1}{n^2} \sum_{1 \leq j \neq k \leq n} E(\xi'_j, \mu_j, \nu_j) E(\xi'_k, \mu_k, \nu_k) \text{Tr}(C_j A C_k \Sigma_Y) + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{n} \text{Tr}(A \tilde{\Sigma}) + \tilde{m}^T A \tilde{m} + 2 \tilde{m}^T A \tilde{C} m_Y + \text{Tr}(\tilde{C} A \tilde{C} \Sigma_Y) + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{n} \text{Tr}(A \tilde{\Sigma}) + \left(\tilde{m} + \tilde{C} m_Y\right)^T A \left(\tilde{m} + \tilde{C} m_Y\right) + \text{Tr}(\tilde{C} A \tilde{C} C_Y) + O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

Therefore, since $m_Y = \tilde{m} + \tilde{C} m_Y + O(1/\sqrt{n})$ and $\Sigma_Y = C_Y - m_Y m_Y^T$, we can deduce that:

$$\text{Tr}(C_Y A) = \frac{1}{n} \text{Tr}(A \tilde{\Sigma}) + \text{Tr}(\tilde{C} A \tilde{C} C_Y) + O\left(\frac{1}{\sqrt{n}}\right)$$

with the notation $\tilde{\Sigma} \equiv \frac{1}{n} \sum_{i=1}^n E(\xi_i^2, \mu_i, \nu_i) \Sigma_i$. Let us then introduce the linear mapping $\Theta : \mathcal{M}_p \rightarrow \mathcal{M}_p$ defined with the equation:

$$\Theta(B) = B + \tilde{C} \Theta(B) \tilde{C}, \quad B \in \mathcal{M}_p \quad (53)$$

(this equation is invertible thanks to Lemma 28), we can then deduce that for any $k \in [n]$:

$$\begin{aligned}
\nu_k &\equiv \text{Tr}(\Sigma_k \Sigma_Y) - (m_k^T m_Y)^2 = \text{Tr}(\Sigma_k C_Y) + m_Y^T C_k m_Y \\
&= \frac{1}{n} \text{Tr}(\Sigma_k \Theta(\tilde{\Sigma})) + m_Y^T C_k m_Y + O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

Remark 26. When B commutes with $\tilde{C}$, then $\Theta(B) = (I_p - \tilde{C}^2)^{-1} B$. However, this formulation is not very interesting for implementation, since $\Theta(B)$ can be approximated far more rapidly with successive iterations of the fixed point equation (53).

4.8. Application to the logistic regression for $x_1, \dots, x_n$ Gaussian i.i.d. vectors.

To illustrate our theoretical results in a simple way we present the example of a supervised classification method called the “the logistic regression” already studied by Mai et al. (2019). Considering a deterministic vector $m \in \mathbb{R}^p$ and a positive symmetric matrix $C \in \mathcal{M}_p$, we suppose we are given n Gaussian random

vectors $z_1, \dots, z_n$, each one following the law $\mathcal{N}(y_i m, C)$ where $y_1, \dots, y_n \in \{-1, 1\}$ are the “labels” of z_i that determine the classes of our data. We assume for simplicity that the classes are balanced. Our classification problem aims at deducing from the *training set* $z_1, \dots, z_n$ and the labels $y_1, \dots, y_n$ a statistical characterization of our two classes that will allow us to classify a new coming data x , independent with the training set and following one or the other law. To introduce the problem, let us express the probability conditioned on a new data x , and knowing that z_i is Gaussian, that $y_i = y$, for a given $y \in \{0, 1\}$:

$$\begin{aligned} \mathbb{P}(y_i = y \mid z_i) &= \frac{\mathbb{P}(y_i = y) \mathbb{P}(z_i \mid y_i = y)}{\mathbb{P}(y_i = y) \mathbb{P}(z_i \mid y_i = y) + \mathbb{P}(y_i = -y) \mathbb{P}(z_i \mid y_i = -y)} \\ &= \frac{e^{-(z_i - ym)^T C^{-1} (z_i - ym)/2}}{e^{-(z_i - ym)^T C^{-1} (z_i - ym)/2} + e^{-(z_i + ym)^T C^{-1} (z_i + ym)/2}} \\ &= \frac{1}{1 + e^{-y z_i^T C^{-1} m}} = \sigma(y z_i^T \beta^*), \end{aligned}$$

noting $\sigma : t \mapsto 1/(1 + e^{-t})$ and $\beta^* \equiv C^{-1} m$. The goal of the logistic regression is to try and estimate β to be able to classify the data depending on the highest value of $\mathbb{P}(y_i = 1 \mid z_i)$ and $\mathbb{P}(y_i = -1 \mid z_i)$. For that purpose, we solve a regularized maximum likelihood problem:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{p} \sum_{i=1}^n \rho(\beta^T x_i) + \frac{\lambda}{2} \|\beta\|^2$$

where $\rho(t) = \log(1 + e^{-t})$, $x_i = y_i z_i$ and $\lambda > 0$ is the regularizing parameter. Differentiating this minimizing problem, we obtain:

$$\beta = \frac{1}{\lambda n} \sum_{i=1}^n f(x_i^T \beta) x_i,$$

where $f : t \mapsto \frac{1}{1 + e^t}$. If one choses λ sufficiently big, the Assumptions 1-5 ter are all satisfied, our results thus allow us to set the concentration $\beta \propto \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n}$ and estimate its first statistics. Introducing μ and ν as the unique solutions to the system of equation:

$$\begin{cases} \mu = E(\xi, \mu, \nu) m^T Q m \\ \nu = E(\xi^2, \mu, \nu) \text{Tr}(\Sigma^2) + 4E(\xi \xi', \mu, \nu) E(\xi, \mu, \nu) m^T C^2 Q m \\ \quad + 2E(\xi \xi'' + \xi'^2, \mu, \nu) m Q C^3 Q m + E(\xi, \mu, \nu)^2 m^2 Q^2 m - \mu^2, \end{cases} \quad (54)$$

with, as before:

- $\forall g : \mathbb{R} \rightarrow \mathbb{R}, E(g, \mu, \nu) = \mathbb{E}[g(z)]$ with $z \sim \mathcal{N}(\mu, \nu)$
- $\xi = f \circ \zeta$
- $\delta = \frac{1}{n} \text{Tr}(\Sigma)$

- $\forall x \in \mathbb{R}, \zeta(x) = x + \delta f(\zeta(x))$
- $Q \equiv (I_p - E(\xi', \mu, \nu)C)^{-1}$

The system (54) can be solved by successive iteration (the quantities $E(g, \mu, \nu)$ for $g = \xi, \xi^2, \xi', \xi\xi', \xi'^2, \xi''\xi$ can all be estimated precisely with random drawings of the distribution $\mathcal{N}(\mu_k, \nu_k)$ at the k^{th} step of the iteration). We can then deduce the performances of the algorithm from the statistics of Y , we depicted on Figure...

Appendix A. Definition of the expectation

Given a concentrated random vector $Z \in (E, \|\cdot\|)$, we will need at one point (Theorems 6, 12 and 9) to be able to consider its expectation. We already know from Remark 1 that, if $Z \propto \mathcal{E}_q(\sigma)$, then for any Lipschitz mapping $f : \mathbb{E} \mapsto \mathbb{R}$, the functional $f(Z)$ admits an expectation $\mathbb{E}[f(Z)] = \int_E f(z) d\mathbb{P}(Z = z)$. This definition can be first generalized when E is a *reflexive space*.

Given a normed vector space $(E, \|\cdot\|)$, we denote $(E', \|\cdot\|)$ the so-called “strong dual” of E , composed of the continuous linear forms of E for the norm $\|\cdot\|$. The norm $\|\cdot\|$ (written the same way as the norm on E for simplicity – no ambiguity being possible) is called the *strong norm* of E' and defined as follows.

Definition 11. *Given a normed vector space $(E, \|\cdot\|)$, the strong norm $\|\cdot\|$ is defined on E' as:*

$$\forall f \in E', \|f\| = \sup_{\|x\| \leq 1} |f(x)|.$$

To be able to define an expectation in E we first assume that E is *reflexive*. To this end, we need to first define a “topological bidual” of E , denoted $(E'', \|\cdot\|)$ and defined by $E'' = (E')'$ with norm the strong norm of the dual of E' .

Definition 12. *The “natural embedding” of E into E'' is defined as the mapping:*

$$\begin{aligned} J : E &\longrightarrow E'' \\ x &\longmapsto (E' \ni f \mapsto f(x)). \end{aligned}$$

It can be shown that J is always one-to-one, but not always onto; however, when J is bijective, we say that E is reflexive.

If E is reflexive, then it can be identified with E'' (this is in particular the case of any vector space of finite dimension but also of any Hilbert space). One can then define the expectation of any concentrated vector $X \propto \mathcal{E}_q(\sigma)$ as follows:

Definition 13. *Given a random vector Z of a reflexive space $(E, \|\cdot\|)$, if the mapping $E' \ni f \mapsto \mathbb{E}[f(Z)] \in \mathbb{R}$ is continuous on E' , we define the expectation of Z as the vector:*

$$\mathbb{E}[Z] = J^{-1}(f \mapsto \mathbb{E}[f(Z)]). \quad (\text{A.1})$$

Remark 27. A reflexive space is a complete space (since it is in bijection with a dual space). It satisfies in particular the Picard Theorem which states that any contractive mapping $f : E \rightarrow E$ ($\forall x, y \in E, \|f(x) - f(y)\| \leq (1 - \varepsilon)\|x - y\|$ with $\varepsilon > 0$) admits a unique fixed point $y = f(y)$. This property will be particularly interesting when considering the concentration of fixed point of concentrated functions of Reflexive spaces in Section 2.

Lemma 29. Given a reflexive space $(E, \|\cdot\|)$, a random vector $Z \in E$ and a continuous linear form $f \in E'$:

$$f(\mathbb{E}[Z]) = \mathbb{E}[f(Z)].$$

PROOF. It is just a consequence of the identity:

$$f(\mathbb{E}[Z]) = J(\mathbb{E}[Z])(f) = J(J^{-1}(f \mapsto \mathbb{E}[f(Z)]))(f) = \mathbb{E}[f(Z)].$$

Proposition 40. Given a reflexive space $(E, \|\cdot\|)$ and a random vector $Z \in E$, if $Z \propto \mathcal{E}_q(\sigma)$, then $\mathbb{E}[Z]$ can be defined with Definition 13.

PROOF. We just need to show that $f \mapsto \mathbb{E}[f(Z)]$ is continuous. There exists $K_p > 0$ such that $\mathbb{P}(\|Z_p\| \leq K_p) \geq \frac{1}{2}$, so that for any $f \in E'$, $\mathbb{P}(f(Z_p) \leq K_p\|f\|) \geq \frac{1}{2}$. Therefore, by definition, for any median m_f of $f(Z_p)$, $m_f \leq K_p\|f\|$ and employing again the inequalities given in Remark 1 we can obtain a similar bound on $|\mathbb{E}[f(Z)]|$ which allows us to state that the mapping ϕ is continuous.

It is still possible to define a notion of expectation when E is not reflexive but is a functional vector space having value on a reflexive space $(F, \|\cdot\|)$; for instance a subspace of F^G , for a given set G .

Definition 14. Given reflexive space $(F, \|\cdot\|)$, a given set G , a subspace $E \subset F^G$, and a random vector $\phi \in E$, if for any $x \in F$, the mapping $F' \ni f \mapsto \mathbb{E}[f(\phi(x))]$ is continuous, we can defined the expectation of ϕ as:

$$\mathbb{E}[\phi] : x \mapsto \mathbb{E}[\phi(x)].$$

Remark 28. When the space $E \subset F^G$ is endowed with a norm $\|\cdot\|$ such that $(E, \|\cdot\|)$ is reflexive and $\forall x \in G, E \ni \phi \mapsto \phi(x)$ is continuous, then there is no ambiguity on the definitions. Indeed, if we note $\mathbb{E}_1[\phi]$ and $\mathbb{E}_2[\phi]$, respectively the expectation of ϕ given by Definition 13 and Definition 14, we can show for any $x \in F$ and any $f \in F'$:

$$f(\mathbb{E}_1[\phi](x)) = \tilde{f}(\mathbb{E}_1[\phi]) = \mathbb{E}[\tilde{f}(\phi)] = \mathbb{E}[f(\phi(x))] = f(\mathbb{E}[\phi(x)]) = f(\mathbb{E}_2[\phi](x)),$$

where $\tilde{f} : E \ni \psi \rightarrow f(\psi(x))$ is a continuous linear form. Since this identity is true for any $f \in E'$, we know by reflexivity of E that $\forall x \in E: \mathbb{E}_1[\phi](x) = \mathbb{E}_2[\phi](x)$. This directly implies that $\mathbb{E}_1[\phi] = \mathbb{E}_2[\phi]$.

Remark 29. Given a random mapping $\phi \in E \subset F^G$ with $(F, \|\cdot\|)$ reflexive, we can then deduce from Proposition 40 that if for all $x \in G$, $\phi(x) \propto \mathcal{E}_q(\sigma)$ in $(F, \|\cdot\|)$ then we can define $\mathbb{E}[\phi]$. With a different formalism, we can endow F^G with the family of semi-norms $(\|\cdot\|_x)_{x \in G}$ defined for any $f \in F^G$ as $\|f\|_x = \|f(x)\|$; then if for all $x \in G$, $\phi \propto \mathcal{E}_q(\sigma)$ in $(F^G, \|\cdot\|_x)$, it is straightforward to set that $\mathbb{E}[\phi]$ is well defined. This is in particular the case if $E = \mathcal{B}(G, F)$ is the set of bounded mappings from G to F and $\phi \propto \mathcal{E}_q(\sigma)$ in $(F^G, \|\cdot\|_\infty)$ where for all $f \in F^G$, $\|f\|_\infty = \sup_{x \in G} \|f(x)\|$.

Appendix B. Concentration of Y solution to $Y = \phi(Y)$ when ϕ is locally Lipschitz

For that we first need a preliminary lemma that will allow us to set that a mapping contracting in a sufficiently large compact admits a fixed point. It expresses through the introduction of a new semi-norm defined for any $f \in \mathcal{F}(E)$, locally Lipschitz as:

$$\|f\|_{\mathcal{L}, \mathcal{B}(y_0, r)} = \sup_{x, z \in \mathcal{B}(y_0, r)} \frac{\|f(x) - f(z)\|}{\|x - z\|}.$$

Lemma 30. Given a mapping $\phi \in \mathcal{F}(E)$, if there exist two constants $\delta, \varepsilon > 0$ and a vector $y_0 \in E$ such that:

$$\|\phi\|_{\mathcal{L}, \mathcal{B}(y_0, \delta)} \leq 1 - \varepsilon \quad \text{and} \quad \|\phi(y_0) - y_0\| \leq \delta\varepsilon,$$

then for any $y \in \mathcal{B}(y_0, \delta\varepsilon)$ and any $k \in \mathbb{N}$:

$$\|\phi^k(y) - y_0\| \leq \delta. \tag{B.1}$$

PROOF. That can be done iteratively. For $k = 0$, it is obvious since $\|y - y_0\| \leq \varepsilon\delta \leq \delta$ ($\varepsilon < 1$). Now if we suppose that (B.1) is true for all $l < k$ and $y \in \mathcal{B}(y_0, \varepsilon\delta)$ (thus in particular for $y = y_0$), we can bound (under $\mathcal{A}_{\phi^\infty}$):

$$\begin{aligned} \|\phi^k(y) - y_0\| &\leq \|\phi^k(y) - \phi^k(y_0)\| + \|\phi^k(y_0) - y_0\| \\ &\leq (1 - \varepsilon)^k \|y - y_0\| + \sum_{i=1}^k \|\phi^i(y_0) - \phi^{i-1}(y_0)\| \\ &\leq (1 - \varepsilon)^k \varepsilon\delta + \sum_{i=1}^k (1 - \varepsilon)^{i-1} \|\phi(y_0) - y_0\| \leq \delta. \end{aligned}$$

Remark 30. Note that in the previous proof k can tend to ∞ without any change in the concentration bounds. This is due to the fact that for any $l \in [k]$ we used the large bounds:

$$(1 - \varepsilon)^l \leq 1 \quad \text{and} \quad \sum_{i=1}^l (1 - \varepsilon)^i \leq \frac{1}{\varepsilon}.$$

Theorem 12. *Let us consider a (sequence of) reflexive vector space $(E, \|\cdot\|)$ admitting a finite norm degree that we note η . Given $\phi \in \mathcal{Lip}(E)$, a (sequence of) random mapping, we suppose that there exists a (sequences of) integer $\sigma > 0$, a constant $\varepsilon > 0$ such that for any constant $K > 0$, there exists a (sequence of) highly probable event $\mathcal{A}_K$ such that:*

- *there exists a (sequence of) deterministic vector $y_0 \in E$ satisfying:*

$$y_0 = \mathbb{E}_{\mathcal{A}_K}[\phi(y_0)]$$

- *for all $y \in \mathcal{B}(y_0, K\sigma\eta^{1/q})$ and for all (sequence of) integer k such that $k \leq O(\log(\eta))$,*

$$\phi^k(y) \stackrel{\mathcal{A}_K}{\in} \mathcal{E}_q(\sigma) \mid e^{-\eta} \text{ in } (E, \|\cdot\|).$$

- $\mathcal{A}_K \subset \{\|\phi\|_{\mathcal{L}, \mathcal{B}(y_0, K\sigma\eta^{1/q})} \leq 1 - \varepsilon\},$

then with high probability (bigger than $1 - Ce^{-c\eta}$ for some constants $C, c > 0$), the random equation

$$Y = \phi(Y)$$

admits a unique solution $Y \in E$ satisfying the linear concentration:

$$Y \in \mathcal{E}_q(\sigma) \mid e^{-\eta}.$$

PROOF. We know that, $(\phi(y_0) \mid \mathcal{A}_1) \in y_0 \pm \mathcal{E}_q(\sigma)$, so in particular Proposition 8 implies that there exist three constants $C, c, K > 0$ such that:

$$\mathbb{P}\left(\|\phi(y_0) - y_0\| \geq K\sigma\eta^{1/q}\varepsilon \mid \mathcal{A}_1\right) \leq Ce^{-\eta/c},$$

Let us then note $\mathcal{A}_Y = \mathcal{A}_1 \cap \mathcal{A}_K \cap \{\|\phi(y_0) - y_0\| \geq K\sigma\eta^{1/q}\varepsilon\}$, we can bound $\mathbb{P}(\mathcal{A}_Y) \leq C'e^{-\eta/c'}$ for some constants C', c' , and we know from Lemma 30 and our hypotheses that $\forall k \in \mathbb{N}$:

$$\phi^k(y_0) \in \mathcal{B}(y_0, K\sigma\eta^{1/q}),$$

(since $y_0 \in \mathcal{B}(y_0, K\sigma\eta^{1/q}\varepsilon)$). Therefore since $\mathcal{A}_Y \subset \mathcal{A}_K \subset \{\|\phi\|_{\mathcal{L}, \mathcal{B}(y_0, K\sigma\eta^{1/q})} \leq 1 - \varepsilon\}$, the sequence $(\phi^k(y_0))_{k \in \mathbb{N}}$ is a Cauchy sequence and it converges to a random vector $Y \in E$ satisfying $Y = \phi(Y)$ (E is complete because it is reflexive).

We now want to show that Y is concentrated. Following the steps of the proof of Theorem 6, one sees that it is sufficient to show that $\|\tilde{Y} - y_0\| = O(\sigma\eta^{1/q})$ for $\tilde{Y}$ defined as the unique solution to the equation $\tilde{Y} = \mathbb{E}_{\mathcal{A}_Y}[\phi^k(\tilde{Y})]$ already

introduced in the proof of Theorem 6 and $k = \lceil -\frac{\log(\eta)}{q \log(1-2\varepsilon)} \rceil$. Let us bound:

$$\begin{aligned}
\|\tilde{Y} - y_0\| &\leq \left\| \mathbb{E}_{\mathcal{A}_Y} [\phi^k(\tilde{Y}) - \phi^k(y_0)] \right\| + \|\mathbb{E}_{\mathcal{A}_Y} [\phi^k(y_0) - y_0]\| \\
&\leq \mathbb{E}_{\mathcal{A}_Y} \left[\left\| \phi^k(\tilde{Y}) - \phi^k(y_0) \right\| \right] + \mathbb{E}_{\mathcal{A}_Y} [\|\phi^k(y_0) - y_0\|] \\
&\leq \mathbb{E}_{\mathcal{A}_Y} [\|\phi\|^k] \|\tilde{Y} - y_0\| + \mathbb{E}_{\mathcal{A}_Y} \left[\sum_{i=1}^k \|\phi\|_{\mathcal{L}}^i \|\phi(y_0) - y_0\| \right] \\
&\leq (1 - \varepsilon) \|\tilde{Y} - y_0\| + K\sigma\eta^{1/q}.
\end{aligned}$$

Thus $\|\tilde{Y} - y_0\| \leq \frac{K\sigma\eta^{1/q}}{\varepsilon}$, so that, noting $\mathcal{A}'_Y \equiv \mathcal{A}_{\frac{K\sigma\eta^{1/q}}{\varepsilon}} \cap \mathcal{A}_Y$, by hypothesis:

$$\phi^k(\tilde{Y}) \overset{\mathcal{A}'_Y}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta},$$

the rest of the proof is then exactly the same as the proof of Theorem 6.

Let us now give a more flexible result involving the norms $\|\cdot\|_{\mathcal{B}(y_0, r)}$ for $r > 0$ but also the semi-norms $\|\cdot\|_{\mathcal{L}, \mathcal{B}(y_0, r)}$. The following Lemma might seem a bit complicated and artificial, it is however perfectly adapted to the requirement of Theorem 9 generalizing Theorem 12 to the case of Lipschitz concentrated mappings ϕ locally Lipschitz.

Lemma 31. *Given a normed vector space $(E, \|\cdot\|)$ whose norm degree is noted η , a vector $y_0 \in E$ and a (sequence of) random mapping $\phi \in \mathcal{F}^\infty(E)$, let us suppose that there exists a (sequence of) constant $\varepsilon > 0$ ($\varepsilon \geq O(1)$) such that for any constant $K > 0$, there exists a (sequence of) event $\mathcal{A}_K$:*

$$\phi \overset{\mathcal{A}_K}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\eta} \quad \text{in} \quad (\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K\sigma\eta^{1/q})}),$$

and there exist two constants C, c such that

$$\mathbb{P}(\|\phi\|_{\mathcal{L}, \mathcal{B}(y_0, K\sigma\eta^{1/q})} \geq 1 - \varepsilon \mid \mathcal{A}_K) \leq Ce^{-c\eta},$$

then for any constant K' ($K' \geq O(1)$) we have:

$$\phi^m \overset{\mathcal{A}_\phi}{\propto} \mathcal{E}_q(\sigma) \mid e^{-\theta'} \quad \text{in} \quad (\mathcal{F}^\infty(E), \|\cdot\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})}),$$

PROOF. In this particular setting it is more convenient to redo the full proof of the weaker statement of Theorem 2 given by Proposition 9. The main difficulty being, that we work in $\mathcal{F}(E)$ with the semi-norms $(\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})})$ and $\|\cdot\|_{\mathcal{L}, \mathcal{B}(y_0, K)}$, (in $(\mathcal{F}^\infty, \|\cdot\|_\infty + \|\cdot\|_{\mathcal{L}})$ there would not have been any issue). As in the proof of Theorem 12, let us introduce three constants $C, c > 0$ and $K > K'/\varepsilon$ such that:

$$\mathbb{P}(\|\Phi(y_0) - y_0\| \geq K\sigma\eta^{1/q}\varepsilon \mid \mathcal{A}_1) \leq Ce^{-\eta/c},$$

Let us set:

$$\mathcal{A}_{\phi^\infty} \equiv \mathcal{A}_{K\varepsilon} \cap \mathcal{A}_1 \cap \left\{ \|\phi\|_{\mathcal{L}, \mathcal{B}(y_0, K\sigma\eta^{1/q})} \leq 1 - \varepsilon \right\},$$

We know from Lemma 30 that for all $k \in \mathbb{N}$, and for all $f \in \mathcal{A}_{\phi^\infty}$ (recall that we identify $\mathcal{A}_{\phi^\infty}$ with $\phi(\mathcal{A}_{\phi^\infty})$):

$$f^k(y) \in \mathcal{B}(y_0, K),$$

since $\|y - y_0\| \leq K'\sigma\eta^{1/q} \leq K\sigma\eta^{1/q}\varepsilon$. For any $f, g \in \mathcal{A}_{\phi^\infty}$:

$$\begin{aligned} \|f^m - g^m\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})} &\leq \sum_{i=1}^m \|f\|_{\mathcal{L}, \mathcal{B}(y_0, K\sigma\eta^{1/q})}^{i-1} \|f(g^{m-i}(y)) - g(g^{m-i}(y))\| \\ &\leq \frac{1}{\varepsilon} \|f - g\|_{\mathcal{B}(y_0, K\varepsilon\sigma\eta^{1/q})}. \end{aligned}$$

Thus the mapping $f \mapsto f^m$ is $\frac{1}{\varepsilon}$ -Lipschitz on $(\mathcal{A}_{\phi^\infty}, \|\cdot\|_{\mathcal{B}(y_0, K'\sigma\eta^{1/q})})$, which directly implies the result of the lemma.

The next theorem is an important improvement (allowed by Lipschitz concentration) of Theorem 8 in that it only take as hypothesis the concentration of ϕ (and not of all its iterates). It also has the advantage to assume only a local control on the Lipschitz character of ϕ to describe a larger range of settings.

Theorem 13. *Let us consider a (sequence of) reflexive vector space $(E, \|\cdot\|)$ we suppose that $\|\cdot\|$ norm degree is noted η (of course $\eta \geq O(1)$), a (sequence of) random mapping $\phi \in \mathcal{Lip}(E)$, a given constant $\varepsilon > 0$ ($\varepsilon \geq O(1)$) and $\mathcal{A}_\phi$, a (sequence of) highly probable events such that $\mathbb{P}(\mathcal{A}_\phi) \leq Ce^{-c\eta}$ for two constants $C, c > 0$ and $\mathcal{A}_\phi \subset \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$. Introducing y_0 , a (sequence of) deterministic vector $y_0 \in E$ satisfying:*

$$y_0 = \mathbb{E}_{\mathcal{A}_\phi}[\phi(y_0)],$$

if we suppose that for any constant $K > 0$:

$$\phi \propto \mathcal{E}_q(\sigma) \mid e^{-\eta} \quad \text{in} \quad (\mathcal{F}(E), \|\cdot\|_{\mathcal{B}(y_0, K\sigma\eta^{1/q})}),$$

then there exists an highly probable event $\mathcal{A}_Y$, such that, under $\mathcal{A}_Y$, the random equation

$$Y = \phi(Y)$$

admits a unique solution $Y \in E$ satisfying the Lipschitz concentration:

$$Y \propto \mathcal{E}_q(\sigma) \mid e^{-\eta}.$$

References

Bai, Z., Zhou, W., 2008. Large sample covariance matrices without independence structures in columns. *Statistica Sinica* 18, 425–442.

- El Karoui, N., Derech, B., Bickel, P.J., Lim, C., Yu, B., 2013. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences of the United States of America* 36, 14557 – 14562.
- James, R.C., 1957. Reflexivity and the supremum of linear functionals. *Annals of Mathematics* 66, 159 – 169.
- Ledoux, M., 2005. The concentration of measure phenomenon. 89, American Mathematical Soc.
- Louart, C., Couillet, R., 2019. Concentration of measure and large random matrices with an application to sample covariance matrices. submitted to *Random Matrices: Theory and Applications* .
- Louart, C., Couillet, R., 2020. A concentration of measure and random matrix approach to large dimensional robust statistics. submitted .
- Mai, X., Liao, Z., Couillet, R., 2019. A large scale analysis of logistic regression: Asymptotics performance and new insights. *ICASSP* .
- Rolland, R., 2006. Fonctions de möbius - formule de rota.
- Rudin, W., 1986. *Real and Complex Analysis* (3rd Edition). McGraw-Hill Book Company.
- Seddik, M.E.A., Louart, C., Tamaazousti, M., Couillet, R., 2020. Random matrix theory proves that deep learning representations of gan-data behave as gaussian mixtures. *ICML* (submitted) .
- Silverstein, J.W., 1986. Eigenvalues and eigenvectors of large dimensional sample covariance matrices. *Random Matrices and their Applications* , 153–159.
- Talagrand, M., 1995. Concentration of Measure and Isoperimetric Inequalities in product spaces. *Publications mathématiques de l’I.H.E.S.*, tome 81.
- Tao, T., 2012. *Topics in random matrix theory*. volume 132. American Mathematical Soc.