Concentration of measure and generalized product of random vectors with an application to Hanson-Wright-like inequalities.

Cosme Louart[a], Romain Couillet[a]

[a] *GIPSA-lab, Université Grenoble-Alpes, France*

Abstract

Starting from concentration of measure hypotheses on m random vectors $Z_1, \dots, Z_m$, this article provides an expression of the concentration of functionals $\phi(Z_1, \dots, Z_m)$ where the variations of ϕ on each variable depend on the product of the norms (or semi-norms) of the other variables (as if ϕ were a product). We illustrate the importance of this result through various generalizations of the Hanson-Wright concentration inequality as well as through a study of the random matrix XDX^T and its resolvent $Q = (I_p - \frac{1}{n}XDX^T)^{-1}$, where X and D are random, which have fundamental interest in statistical machine learning applications.

Keywords: Concentration of Measure, Hanson Wright inequality, Random Matrix Theory.

Contents

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

Introduction

Among the various assumptions one could pose on random vectors $Z_1, \ldots, Z_m$ to study the concentration of a functional $\phi(Z_1, \ldots, Z_m)$ of limited variations (on $Z_1, \ldots, Z_m$), concentration of measure hypotheses provide flexible properties allowing one (i) to characterize a wide range of settings where, in particular, the hypothesis of independent entries is relaxed, and (ii) to retrieve rich concentration inequalities with precise convergence bounds. The historical result of concentration of measure theory was obtained on the uniform distribution on the sphere by Lévy Lévy (1951) and later formalized by Milman and Gromov Gromov and Milman (1983) who extended the approach to other families of distributions, notably involving isoperimetric inequalities and the Ricci curvature. To present the simplest picture possible, we admit for the moment that what we call “concentrated vectors” (or “Lipschitz concentrated vectors”) are transformations $X = F(Z) \in \mathbb{R}^p$ of a Gaussian vector $Z \sim \mathcal{N}(0, I_d)$ for a given 1-Lipschitz (for the Euclidean norm) mapping $F : \mathbb{R}^d \to \mathbb{R}^p$. This class of random vectors originates from a core result of concentration of measure theory (Ledoux, 2005, Corollary 2.6) which states that, for any λ-Lipschitz mapping $f : \mathbb{R}^d \to \mathbb{R}$ (where $\mathbb{R}^d$ and $\mathbb{R}$ are respectively endowed with the Euclidean norm $\| \cdot \|$ and with the absolute value $| \cdot |$),

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

where $C = 2$ and $c = \sqrt{2}$ (these constants do not depend on the dimensions d !). Note here that the speed of concentration is proportional to the Lipschitz parameter of f. This implies in particular that the standard deviation of the random variable $f(Z)$ – called a “λ-Lipschitz observation of Z” – does not depend on the dimension d (if λ stays constant when d tends to ∞). We denote this property succinctly as $Z \propto C\mathcal{E}_2(c)$ or, if we place ourselves in the quasi-asymptotic regime where the dimension d (or p) is large, we do not pay attention to the constants appearing in the exponential bound (as long as $C, c \leq_{d\to\infty} O(1)$, the result would not be much different) and we write instead $Z \propto \mathcal{E}_2$.

We can then deduce a host of concentration inequalities on any observation $g(F(Z))$ for $g : \mathbb{R}^p \to \mathbb{R}$ Lipschitz. If F is, say, σ-Lipschitz with σ possibly depending on the dimension, we have the concentration $X = F(Z) \propto \mathcal{E}_2(\sigma)$. This succinct notation ($X \propto \mathcal{E}_2(\sigma)$ is analogous to (1) with σ replacing c) only displays the central quantity describing the concentration of X, namely, the "observable diameter of X": $O(\sigma)$. Indeed, the implicit concentration inequalities constrain the standard deviations of *any* ν-Lipschitz observation of X to be of order $O(\nu\sigma)$.

The objective of this article is to go beyond the Lipschitz case and express, with the help of our shorthand notation, the concentration of products of concentrated random vectors. As an illustrative exemple, let $Y = \underbrace{X \circ \cdots \circ X}_{m \text{ times}} \in \mathbb{R}^p$, for a given product "$\circ$" satisfying:

$$\forall x_1, \dots, x_m \in \mathbb{R}^p, \quad \|x_1 \circ \cdots \circ x_m\| \leq \|x_1\| \cdots \|x_m\|. \tag{2}$$

($\circ$ could for instance be the entry-wise product and the norm would be the infinite norm). When $\mathbb{E}[X] = 0$, we will see in particular that

$$Y \propto \mathcal{E}_2(p^{\frac{m-1}{2}}\sigma^m) + \mathcal{E}_{\frac{2}{m}}(\sigma^m), \tag{3}$$

where μ satisfies $\mathbb{E}[\|X\|] \leq O(\mu)$, which means in our framework that there exist two constants $C, c > 0$ (independent of p) such that for any 1-Lipschitz mapping $f : \mathbb{R}^p \to \mathbb{R}$, $\forall t > 0$

$$\mathbb{P}\left(|f(Y) - \mathbb{E}[f(Y)]| \geq t\right) \leq C \exp\left(-\left(\frac{t}{cp^{\frac{m-1}{2}}\sigma^m}\right)^2\right) + C \exp\left(-\left(\frac{t}{c\sigma^m}\right)^{2/m}\right).$$

We see here that the term $\mathcal{E}_{\frac{2}{m}}(\sigma^m)$ in (3) controls the tail of the distribution of $f(Y)$, but its first moments are controlled by the term $\mathcal{E}_2(p^{\frac{m-1}{2}}\sigma^m)$. Specifically, its standard deviation is of order $O(p^{\frac{m-1}{2}}\sigma^m)$, the observable diameter of Y. In a sense, the result is quite intuitive looking back at the algebraic inequality (2): here $\mathbb{E}[\|X\|] \leq O(\sqrt{p})$ and the variations of Y are bounded by $\mathbb{E}[\|X\|]^{m-1}$ times the variation of X. This simple scheme generalizes to more elaborate products of random vectors $X_1, \dots, X_m$ belonging to different normed vector spaces, with Y possibly not a multilinear mapping of $(X_1, \dots, X_m)$ but still satisfying an inequality similar to (2) (with semi-norms possibly replacing some of the norms). The complete description of these possible settings is the central result of this paper: Theorem 2.

As a simple, but fundamental, application of our main result, the article then provides two Hanson-Wright inequalities expressing the concentration of

X^TAX, where $A \in \mathcal{M}_p$ and X is either a random vector of $\mathbb{R}^p$ or a random matrix of $M_{p,n}$. The common approach to this problem is to consider random vectors $X = (X_1, \dots, X_p) \in \mathbb{R}^p$ with independent subgaussian entries, say $X_i \propto \mathcal{E}_2(K)$; under this setting, from Boucheron et al. (2013) (see also Vershynin (2017)), we have the concentration:

$$X^TAX \propto \mathcal{E}_2(K^2\|A\|_F) + \mathcal{E}_1(K^2\|A\|) \tag{4}$$

(where $\|\cdot\|_F$ is the Frobenius norm). Based on concentration of measure hypotheses (allowing for dependence between the entries), good concentration inequalities were already obtained in the case $n = 1$ in Vu and Wang (2014) (with a term $\mathcal{E}_2(\sqrt{\log n}\|A\|_F)$ replacing $\mathcal{E}_2(\|A\|_F)$) and then improved in Adamczak (2015) to reach (4). We thus reprove this result in the context of Theorem 2 and even go further, looking at the concentration of X^TAY and YDX^T for $D \in \mathcal{M}_n$, diagonal and $X, Y \in \mathcal{M}_{p,n}$ satisfying the concentrations $X, Y, D \propto \mathcal{E}_2$ (as if they had i.i.d. entries following the law $\mathcal{N}(0, 1)$). However, as explained in Remark 8, unlike Vu and Wang (2014) and Adamczak (2015), we do not take convex concentration hypotheses (issued from a well known result from Talagrand) because Theorem 2 could not be proven in this setting[1].

To illustrate our central result with more general products (when $m = 3$), we study the concentration of X^TDY where $D \in \mathcal{M}_n$ is a diagonal random matrix and $X, Y \in \mathcal{M}_{p,n}$ are two random matrices, all satisfying $X, D, Y \propto \mathcal{E}_2$. With the same random objects, to go beyond the multilinear case, we look at the concentration of the resolvent $Q = (I_p - \frac{1}{n}XDX^T)^{-1}$ studied in Pajor and Pastur (2009); Guédon et al. (2014) but with a diagonal matrix D possibly depending on X, Y. This setting appears in robust regression problems El Karoui et al. (2013); Mai et al. (2019); Seddik et al. (2021). Although they might seam restrictive, with the possibly complex dependencies between the entries of x_i they allow, concentration of measure hypotheses, are very light compared to the classical Gaussian hypotheses adopted in large dimensional statistics and statistical learning Huang (2017); Deng et al. (2020). To obtain a good concentration of Q, one has to assume that the columns $x_1, \dots, x_n$ of X are all independent and that for all $i \in [n]$, there exists a diagonal random matrix $D^{(i)} \in \mathcal{M}_n$, not too far from D and independent with x_i. We show then under our realistic hypotheses that each of the linear observations $u(Q)$ lie close to $u((I_p - \frac{1}{n}X\mathbb{E}[D]Y^T)^{-1})$ that can be estimated with classical random matrix theory results.

The remainder of the article is organized as follows. After rigorously setting a probabilistic approach to the concentration of measure theory **(I)**, we introduce the class of *linearly concentrated random vectors* **(II)** and explain how their norm can be controlled in generic normed vector spaces **(III)**. We then briefly discuss the fact that the random vector $(X_1, \dots, X_m)$ (as a whole) is not always

[1]A result analog to Theorem 2 can be proven in convex concentration setting and for the entry wise product in $\mathbb{R}^p$ or the matrix product in $\mathcal{M}_{p,n}$ but this not the purpose of the present article.

concentrated if one only assumes that each of the X_i's, $i \in [m]$, is concentrated **(IV)**. This provides us with the ingredients to establish the concentration of $\phi(X_1, \ldots, X_m)$ in Theorem 2, the core result of the article, and to provide a first set of elementary consequences **(V)**. As an application of Theorem 2, we next provide a generalization of the Hanson-Wright theorem **(VI)**. Then we end the article with a study of the concentration of XDY^T **(VII)** and the resolvent $Q = (I_p - \frac{1}{n}XDX^T)^{-1}$ **(VIII)**.

1. Basics and notations of the concentration of measure framework

To discuss concentration of measure, 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, given a sequence of random vectors $(Z_p)_{p\geq\mathbb{N}}$ where each Z_p belongs to a space of dimension p (typically $\mathbb{R}^p$), we wish to obtain inequalities of the form:

$$\forall p \in \mathbb{N}, \forall t > 0 : \mathbb{P}\left(|f_p(Z_p) - a_p| \geq t\right) \leq \alpha_p(t), \tag{5}$$

where, for every $p \in \mathbb{N}$, $\alpha_p : \mathbb{R}^+ \to [0, 1]$ is called a *concentration function*: it is left-continuous, decreasing, and tends to 0 at infinity; $f_p : \mathbb{R}^p \to \mathbb{R}$ 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 (5) for all 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).

Concentrated vectors admitting an exponentially decreasing concentration function α_p are extremely flexible objects. We dedicate the next two subsections to further definitions of the fundamental notions involved under this setting. These are of central interest to the present article – this approach is primarily inspired by the Gaussian fundamental example satisfying (1).

Our main interest is in two classes of concentrated vectors, characterized by the regularity of the class of admissible sequences of functions $(f_p)_{p\in\mathbb{N}}$ satisfying (5). When (5) holds for all the 1-Lipschitz mappings f_p, Z_p is said to be *Lipschitz concentrated*; when true for all 1-Lipschitz *linear* mappings f_p, Z_p is said to be *linearly concentrated* (the convex concentration, not studied here, occurs when (5) is satisfied for all 1-Lipschitz convex mappings f_p Vu and Wang (2014)). As such, the concentration of a random vector Z_p is only defined through the concentration of what we refer to as its "observations" $f_p(Z_p)$ for all f_p in a specific class of functions.

We will work with normed (or semi-normed) vector spaces, although concentration of measure theory is classically developed in metric spaces. The presence of a norm (or a semi-norm[2]) on the vector space is particularly important when

[2]A semi-norm becomes a norm when it satisfies the implication $\|x\| = 0 \Rightarrow x = 0$.

establishing the concentration of a product of random vectors.

Definition/Proposition 1. *Given a sequence of normed vector spaces $(E_p, \|\cdot\|_p)_{p\geq 0}$, a sequence of random vectors*[3] $(Z_p)_{p\geq 0} \in \prod_{p\geq 0} E_p$*, 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 Lipschitz* q-exponentially concentrated *with* observable diameter *of order* $O(\sigma_p)$ *iff one of the following three equivalent assertions is satisfied:*[4]

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

$$\mathbb{P}\left(\left|f(Z_p) - f(Z_p')\right| \geq t\right) \leq Ce^{-(t/c\sigma_p)^q},$$

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

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

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

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

where Z_p' *is an independent copy of* Z_p *and* m_f *is a median*[5] *of* $f(Z_p)$*; the mappings* f *are* 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* $Z \propto \mathcal{E}_q(\sigma)$*).*

Remark 1 (Quasi-asymptotic regime). *Most of our results will be expressed under the quasi-asymptotic regime where p is large. Sometimes, it will be natural to index the sequences of random vectors with two (or more) indices (e.g., the numbers of rows and columns for random matrices): in these cases, the quasi-asymptotic regime is not well defined since the different indices could have different convergence speed. This issue is overcome with the extensive use of the notation* $O(\sigma_t)$*, where* $t \in \Theta$ *designates the (possibly multivariate) index. Given two sequences* $(a_t)_{t\in\Theta}, (b_t)_{t\in\Theta} \in \mathbb{R}_+^{\Theta}$*, we will denote* $a_t \leq O(b_t)$ *if there exists a constant* $C > 0$ *such that* $\forall t \in \Theta$*,* $a_t \leq Cb_t$ *and* $a_t \geq O(b_t)$ *if* $\forall t \in \Theta$*,* $a_t \geq Cb_t$*. For clarity, when dealing with a "constant"* $K > 0$*, we will often*

---

[3] A 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 $Z : \Omega \to E$, but we abusively simply denote $Z \in E$.

[4] Aside from the fact that they all give interesting interpretation of the concentration of a random vector, all three characterizations can be relevant, depending on the needs:

- the characterization with the independent copy is employed in Remark 4 and in the proof of Theorem 2;
- the characterization with the median is employed in the proof of Lemma 1;
- the characterization with the expectation, likely the most intuitive, is used to establish Proposition 11, Theorem 3 and Lemma 8.

[5] $\mathbb{P}\left(f(Z_p) \geq m_f\right) \geq \frac{1}{2}$ and $\mathbb{P}\left(f(Z_p) \leq m_f\right) \geq \frac{1}{2}$.

state that $K \leq O(1)$ and $K \geq O(1)$ (depending on the required control). For a concentrated random vector $Z_t \propto \mathcal{E}_q(\sigma_t)$, any sequence $(\nu_t)_{t\in\Theta} \in \mathbb{R}_+^\Theta$ such that $\sigma_t \leq O(\nu_t)$ is also an observable diameter of Z_t. When $\sigma_t \leq O(1)$, we simply write $Z_t \propto \mathcal{E}_q$.

The equivalence between the three definitions is proved in Ledoux (2005) (full details are given in (Louart and Couillet, 2019, Propositions 1.2, 1.18 Corollary 1.24)).

Remark 2 (Existence of the expectation). *(Ledoux, 2005, Proposition 1.7) In the last item of Definition 1, the existence of the expectation of $f_p(Z_p)$ is guaranteed if any of the other two items holds. For instance*

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

implies the bound

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

the random variable $f_p(Z_p)$ is thus integrable and admits an expectation (there always exists a median $m_{f_p} \in \mathbb{R}$).

Remark 3 (Metric versus normed spaces). *It is more natural, as done in Ledoux (2005), to introduce the notion of concentration in metric spaces, as one only needs to resort to Lipschitz mappings which merely require a metric structure on E. However, to exploit Theorem 2, we will need to control the amplitude of concentrated vectors which is easily conducted when the metric is a norm, under linear concentration assumptions.*

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 i
n addition $\|x\|' = 0 \Rightarrow x = 0$).

When a concentrated vector $Z_p \propto \mathcal{E}_q(\sigma_p)$ takes values only on some subset $A_p \equiv Z_p(\Omega) \subset E_p$ (where Ω is the universe), it might be useful to be able to establish the concentration of observations $f_p(Z_p)$ where f_p is only 1-Lipschitz on A_p (and possibly non Lipschitz on $E_p \setminus A_p$). This would be an immediate consequence of Definition 1 if one were able to extend $f_p|_{A^p}$ into a mapping $\tilde{f}_p$ Lipschitz on the whole vector space E_p; but this is rarely possible. Yet, the observation $f_p(Z_p)$ does concentrate under the hypotheses of Definition 1.

Lemma 1 (Concentration of locally Lipschitz observations). *Given a sequence of random vectors $Z_p : \Omega \to E_p$, satisfying $Z_p \propto \mathcal{E}_q(\sigma_p)$, for any sequence of mappings $f_p : E_p \to \mathbb{R}$, which are 1-Lipschitz on $Z_p(\Omega)$, we have the concentration $f_p(Z_p) \propto \mathcal{E}_q(\sigma_p)$.*

PROOF. considering a sequence of median m_{f_p} of $f_p(Z_p)$ and the (sequence of) sets $S_p = \{f_p \leq m_{f_p}\} \subset E_p$, 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 for any $z \in A$ and $t > 0$:

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

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 $\mathbb{P}(d(Z_p,S_p) \leq 0) \geq \mathbb{P}(f_p(Z_p) \leq m_{f_p}) \geq \frac{1}{2}$),

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

One could argue that, instead of Definition 1, we could have posed hypotheses on the concentration of Z_p on $Z_p(\Omega)$ only; however, we considered the present definition of concentration already quite complex as it stands. This locality aspect must be kept in mind: it will be exploited to obtain the concentration of products of random vectors.

Lemma 1 is particularly interesting when working with conditioned variables.[6]

Remark 4 (Concentration of conditioned vectors). *Given a (sequence of) random vectors $Z \propto \mathcal{E}_q(\sigma)$ and a (sequence of) events $\mathcal{A}$ such that $\mathbb{P}(\mathcal{A}) \geq O(1)$, it is straightforward to show that $(Z \mid \mathcal{A}) \propto \mathcal{E}_q(\sigma)$, since there exist two constants $C, c > 0$ such that for any $p \in \mathbb{N}$ and any 1-Lipschitz mapping $f : E_p \to \mathbb{R}$:*

$$\forall t > 0 : \mathbb{P}\left(\left|f(Z_p) - f(Z_p')\right| \geq t \mid \mathcal{A}\right) \leq \frac{1}{\mathbb{P}(A)}\mathbb{P}\left(\left|f(Z_p) - f(Z_p')\right| \geq t\right) \leq Ce^{-c(t/\sigma)^q}.$$

This being said, Lemma 1 allows us to obtain the same concentration inequality for any mapping $f : E_p \to \mathbb{R}$ 1-Lipschitz on $Z_p(\mathcal{A})$ (that will be abusively denoted $\mathcal{A}$ later on).

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

Proposition 1 (Stability through Lipschitz mappings). *In the setting of Definition 1, given a sequence $(\lambda_p)_{p \geq 0} \in \mathbb{R}_+^{\mathbb{N}}$, a supplementary sequence of*

---

[6]Letting $X : \Omega \to E$ 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 $X \mid \mathcal{A}$ designates the random vector X conditioned with $\mathcal{A}$ defined as the measurable mapping $(\mathcal{A}, \mathcal{F}_A, \mathbb{P}/\mathbb{P}(\mathcal{A})) \to (E, \|\cdot\|)$ satisfying: $\forall \omega \in \mathcal{A}$, $(X \mid \mathcal{A})(\omega) = X(\omega)$. When there is no ambiguity, we will allow ourselves to designate abusively with the same notation "$\mathcal{A}$" the actual $\mathcal{A} \subset \Omega$ and the subset $X(\mathcal{A}) \subset E$.

normed vector spaces $(E'_p, \|\cdot\|'_p)_{p\geq 0}$ and a sequence of λ_p-Lipschitz transformations $F_p : (E_p, \|\cdot\|_p) \to (E'_p, \|\cdot\|'_p)$, we have

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

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

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

Some fundamental results also give concentrations $Z \propto \mathcal{E}_1$ (when $Z \in \mathbb{R}^p$ has independent entries with density density $\frac{1}{2}e^{-|\cdot|}d\lambda_1$, (Talagrand, 1995)) or $Z \propto \mathcal{E}_q\left(p^{-\frac{1}{q}}\right)$ (when $Z \in \mathbb{R}^p$ is uniformly distributed on the unit ball of the norm $\|\cdot\|_q$, (Ledoux, 2005)). .

Remark 5 (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 an asymptotic rate difference 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 $O(1)$, that is to say $\frac{1}{\sqrt{p}}$ times less than the metric diameter (that is of order $O(\sqrt{p})$): Gaussian vectors are indeed concentrated.*

As a counter example of a non concentrated vector, one may consider the random vector $Z = [X, \ldots, X] \in \mathbb{R}^p$ 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 $\frac{1}{\sqrt{p}}(X + \cdots + X)$ (the mapping $(z_1, \ldots, z_p) \mapsto \frac{1}{\sqrt{p}}(z_1 + \cdots + z_p)$ is 1-Lipschitz).

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

Proposition 2 (Characterization with the centered moments). *(Ledoux, 2005, Proposition 1.10) A random vector $Z \in E$ is q-exponentially concentrated with an observable diameter of order σ (i.e., $Z \propto \mathcal{E}_q(\sigma)$) if and only if there exist two constants $C, c > 0$ such that for all $p \in \mathbb{N}$, any (sequence of) 1-Lipschitz functions $f : E_p \to \mathbb{R}$:*

$$\forall r > 0 : \mathbb{E}\left[\left|f(Z_p) - f(Z_p')\right|^r\right] \leq C \left(\frac{r}{q}\right)^{\frac{r}{q}} (c\sigma_p)^r, \tag{6}$$

where Z_p' is an independent copy of Z_p. Inequality (6) also holds if we replace $f(Z_p')$ with $\mathbb{E}[f(Z_p)]$ (of course the constants C and c might be slightly different).

The Lipschitz-concentrated vectors described in Definition 1 belong to the larger class of linearly concentrated random vectors that only requires the linear observations to concentrate. This "linear concentration" presents less stability properties than those described by Proposition 1 but is still a relevant notion because:

1. although it must be clear that a concentrated vector Z is generally far from its expectation (for instance Gaussian vectors lie on an ellipse), it can still be useful to have some control on $\|Z - \mathbb{E}[Z]\|$ to express the concentration of product of vectors; linear concentration is a sufficient assumption for this control,
2. there are some examples (Proposition 8 and 10) where we can only derive linear concentration inequalities from a Lipschitz concentration hypothesis. In that case, we say that the Lipschitz concentration "degenerates" into linear concentration that appears as a "residual" concentration property.

These properties of linear concentration are discussed in depth in the next section.

2. Linear concentration and control on high order statistics

Definition 2 (Linearly concentrated vectors). *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 $q > 0$, Z_p is said to be 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 $f \in E_p'$ ($\forall p \in \mathbb{N}$, $\forall x \in E$: $|f(x)| \leq \|x\|$):*

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

When the property holds, we write $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$. If it is unnecessary to mention the deterministic equivalent, we will simply write $Z \in \mathcal{E}_q(\sigma)$. If we just need to control its amplitude, we can write $Z \in O(\theta) \pm \mathcal{E}_q(\sigma)$ when $\|\tilde{Z}_p\| \leq O(\theta_p)$.

When $q = 2$, we retrieve the well known class of sub-Gaussian random vectors. We need this definition with generic q to prove Proposition 8 which involves a weaker than $\mathcal{E}_2$ tail decay.

Of course linear concentration is stable through affine transformations.

Proposition 3 (Stability through affine mappings). *Given two (sequences of) normed vector spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, a (sequence of) random vectors $Z \in E$, a (sequence of) deterministic vectors $\tilde{Z} \in E$ and a (sequence of) affine mappings $\phi : E \to F$ such that $\forall x \in E : \|\phi(x) - \phi(0)\|_F \leq \lambda \|x\|_E$:*

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

When the expectation can be defined, there exists an implication link between Lipschitz concentration (Definitions 1) and linear concentration (Definition 2).

Lemma 2. *Given a normed space $(E, \|\cdot\|)$ and a random vector $Z \in E$ admitting an expectation, we have the implication:*

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

This implication becomes an equivalence in law dimensional spaces (i.e. when the sequence index "p" is not linked to the dimension of the vector spaces E_p); then the distinction between linear concentration and Lipschitz concentration is not relevant anymore. To simplify the hypotheses, we will assume that the normed vector space does not change at all with p.

Proposition 4. *Given a normed vector space of finite dimension $(E, \|\cdot\|)$, a sequence of random vectors $(Z_p)_{p\in\mathbb{N}} \in E^{\mathbb{N}}$ and a sequence of positive values $(\sigma_p)_{p\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, we have the equivalence:*

$$Z \propto \mathcal{E}_q(\sigma) \qquad \Longleftrightarrow \qquad Z \in \mathbb{E}[Z] \pm \mathcal{E}_q(\sigma)$$

Proof. We already know from Lemma 2 that $Z \propto \mathcal{E}_q(\sigma) \Rightarrow Z \in \mathbb{E}[Z] \pm \mathcal{E}_q(\sigma)$, so let us now assume that $Z \in \mathbb{E}[Z] \pm \mathcal{E}_q(\sigma)$. Let us note d, the dimension of E and introducing $(e_1, \ldots, e_d) \in E^d$, a basis of d, we note $\|\cdot\|_{\ell^\infty}$, the norm defined for any $x = \sum_{i=1}^d x_i e_i$ as $\|x\|_{\ell_\infty} = \max_{i\in[d]} |x_i|$. There exists a constant α ($\alpha \leq O(1)$) such that for all $x \in E$, $\|x\| \leq \alpha \|x\|_{\ell^\infty}$ and therefore, one can bound for any 1-lipschitz mapping $f : E \to \mathbb{R}$ and any sequence of random vectors Z', independent with Z:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \mathbb{P}\left(\|Z - Z'\| \geq t\right) \leq \mathbb{P}\left(\alpha \sup_{i\in[d]} |Z_i - Z_i'| \geq t\right) \leq dCe^{-(t/c\sigma\alpha)^q},$$

where $C, c > 0$ are two constants. Since $dC, \alpha c \leq O(1)$, we retrieve the Lipschitz concentration of Z.

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 3. *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 the equivalence:*

$$Z \in \tilde{Z}' \pm \mathcal{E}_q(\sigma) \qquad \Longleftrightarrow \qquad \left\| \tilde{Z} - \tilde{Z}' \right\| \leq O(\sigma)$$

Definition 3 (Centered 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} centered moment of X" the symmetric r-linear form $C_r^X : (\mathbb{R}^p)^r \to \mathbb{R}$ defined for any $u_1, \ldots, u_r \in \mathbb{R}^p$ by*

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

When $r = 2$, the centered moment is the covariance matrix.

We define the operator norm of an r-linear form S of $\mathbb{R}^p$ as

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

When S is symmetric, we employ the simpler formula $\|S\| = \sup_{\|u\| \leq 1} S(u, \ldots, u)$. We then have the following characterization, similar to Proposition 2 (refer to (Louart and Couillet, 2019, Proposition 1.21, Lemma 1.21) for the technical arguments required to go from a bound on $r \in \mathbb{N}$ to a bound on $r > 0$).

Proposition 5 (Moment characterization of linear concentration). *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:*

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

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 $\|C\| \leq O(\sigma^2)$, if in addition $X \in O(\sigma) \pm \mathcal{E}_q(\sigma)$ (which means that $\|\mathbb{E}[X]\| \leq O(\sigma)$), then $\|\mathbb{E}[XX^T]\| \leq O(\sigma^2)$

With these results at hand, we are in particular in position to explain how a control on the norm can be deduced from a linear concentration hypothesis.

3. Control of the norm of linearly concentrated random vectors

Given a random vector $Z \in (E, \|\cdot\|)$, if $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$, the control of $\|Z - \tilde{Z}\|$ 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 $(e_1,\ldots,e_p)$ is the canonical basis of $\mathbb{R}^p$). We can then bound:

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

for some constants $c, C > 0$ ($C \leq O(1)$, $c \geq O(1)$).

To manage the infinity norm, the supremum is taken on a finite set $\{e_1, \ldots e_p\}$. Problems arise when this supremum must be taken on an infinite set. For instance, for the Euclidean norm, the supremum is taken over the whole unit ball $\mathcal{B}_{\mathbb{R}^p} \equiv \{u \in \mathbb{R}^p, \|u\| \leq 1\}$ since for any $x \in \mathbb{R}^p$, $\|x\| = \sup\{u^T x, \|u\| \leq 1\}$. 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.

One can then show that there exist two constants $C, c > 0$ such that:

$$\mathbb{P}(\|Z-\tilde{Z}\| \geq t) \leq \max(C,e)\exp\left(-\frac{ct^q}{p\sigma^q}\right). \tag{7}$$

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

$$\forall x \in E : \|x\| = \sup_{\substack{f\in H\\ \|f\|\leq 1}} f(x), \quad \text{with } H \subset E' \text{ and } \dim(\mathrm{Vect}(H)) < \infty, \tag{8}$$

for a given $H \subset E'$ (for E', the dual space of H) and with $\mathrm{Vect}(H)$ the subspace of E' generated by H. Such a $H \subset E'$ exists in particular when $(E, \|\cdot\|)$ is a reflexive[7] space James (1957).

[7]Introducing the mapping $J : E \to E''$ (where E'' is the bidual of E) satisfying $\forall x \in E$ and $\phi \in E'$: $J(x)(\phi) = \phi(x)$, the normed vector space E is said to be "reflexive" if J is onto.

When $(E, \|\cdot\|)$ is of infinite dimension, it is possible to establish (8) for some $H \subset E$ when E is reflexive thanks to a result from James (1957), or for some choice of semi-norms $\|\cdot\|$. Without going into details, we introduce the notion of *norm degree* which will help us adapt the concentration rate p appearing in the exponential term of concentration inequality (7) (concerning $(\mathbb{R}^p, \|\cdot\|)$) to other normed vector spaces.

Definition 4 (Norm degree). *Given a normed (or semi-normed) 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,*
- $\eta_H \equiv \dim(\operatorname{Vect} H)$ *if H is infinite.*

If there exists a subset $H \subset E'$ such that (8) is satisfied, we 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 :*

- $\eta\left(\mathbb{R}^p, \|\cdot\|_\infty\right) = \log(p)$ *($H = \{x \mapsto e_i^T x, 1 \leq i \leq p\}$),*
- $\eta\left(\mathbb{R}^p, \|\cdot\|\right) = p$ *($H = \{x \mapsto u^T x, u \in \mathcal{B}_{\mathbb{R}^p}\}$),*
- $\eta\left(\mathcal{M}_{p,n}, \|\cdot\|\right) = n + p$ *($H = \{M \mapsto u^T M v, (u, v) \in \mathcal{B}_{\mathbb{R}^p} \times \mathcal{B}_{\mathbb{R}^n}\}$),*
- $\eta\left(\mathcal{M}_{p,n}, \|\cdot\|_F\right) = np$ *($H = \{M \mapsto \operatorname{Tr}(AM), A \in \mathcal{M}_{n,p}, \|A\|_F \leq 1\}$),*
- $\eta\left(\mathcal{M}_{p,n}, \|\cdot\|_*\right) = np$ *($H = \{M \mapsto \operatorname{Tr}(AM), A \in \mathcal{M}_{n,p}, \|A\| \leq 1\}$).*[8]

Just to give some justification, if $E = \mathbb{R}^p$ or $E = \mathcal{M}_{p,n}$, the dual space E' can be identified with E through the representation with the scalar product. Given a subset $H' \subset E$ such that:

$$\forall x \in \mathbb{R}^p, \|x\|_\infty = \sup_{u \in H'} u^T x,$$

we can set that all $u \in H'$ satisfy $\|u\|_1 = \sum_{i=1}^p |u_i| \leq 1$ because if we note $u' = (sign(u_i))_{i \in [p]}$, we can bound $\|u\|_1 = u^T u' \leq \sup_{v \in H'} v^T u' \leq \|u'\|_\infty \leq 1$. Then, noting $H = \{e_1, \ldots, e_p\}$, we know that $H \subset H'$, otherwise, if, say $e_i \notin H'$, then one could bound $\|e_i\|_\infty = \sup_{u \in H'} u^T e_i < 1$ (because if $\|u\|_1 \leq 1$ and $u \neq e_i$, then $u_i < 1$). Therefore $H \subset H'$ and it consequently reaches the minimum of $\eta_{H'}$. The value of the other norm indexes is justified with the same arguments.

[8] $\|\cdot\|_*$ is the nuclear norm defined for any $M \in \mathcal{M}_{p,n}$ by $\|M\|_* = \operatorname{Tr}(\sqrt{MM^T})$; it is the dual norm of $\|\cdot\|$, which means that for any $A, B \in \mathcal{M}_{p,n}$, $\operatorname{Tr}(AB^T) \leq \|A\|\|B\|_*$. One must be careful that Proposition 6 is rarely useful to bound the nuclear norm as explained in footnote 10.

Depending on the ambient vector space, one can employ one of these examples along with the following proposition borrowed from (Louart and Couillet, 2019, Proposition 2.9,2.11, Corollary 2.13) to establish the concentration of the norm of a random vector.

Proposition 6. *Given a reflexive vector space $(E, \|\cdot\|)$ and a concentrated vector $Z \in E$ satisfying $Z \in \tilde{Z} \pm \mathcal{E}_q(\sigma)$:*

$$\|Z - \tilde{Z}\| \propto \mathcal{E}_q\left(\eta_{\|\cdot\|}^{1/q}\sigma\right) \qquad \text{and} \qquad \mathbb{E}\left[\|Z - \tilde{Z}\|\right] \leq O\left(\eta_{\|\cdot\|}^{1/q}\sigma\right).$$

Remark 6. *In Proposition 6, if $Z \propto \mathcal{E}_q(\sigma)$, the norm staisfies the same concentration as it is a Lipschitz observation, and one gets*[9]*:*

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

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

- *if $Z \propto \mathcal{E}_2$ in $(\mathbb{R}^p, \|\cdot\|)$, then $\mathbb{E}\|Z\| \leq \|\mathbb{E}[Z]\| + O(\sqrt{p})$,*
- *if $X \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|)$, then $\mathbb{E}\|X\| \leq \|\mathbb{E}[X]\| + O(\sqrt{p+n})$,*
- *if $X \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$, then $\mathbb{E}\|X\| \leq \|\mathbb{E}[X]\|_F + O(\sqrt{pn})$.*
- *if $X \propto \mathcal{E}_2(\sqrt{\min(p,n)})$ in $(\mathcal{M}_{p,n}, \|\cdot\|_*)$, then $\mathbb{E}\|X\|_* \leq \|\mathbb{E}[X]\|_* + O(\sqrt{pn}\sqrt{\min(p,n)})$.*[10]

Example 3. *Let us consider the semi norm $\|\cdot\|_d$ that will be useful later and that satisfies:*

$$\forall M \in \mathcal{M}_n : \quad \|M\|_d = \left(\sum_{i=1}^{n} M_{i,i}^2\right)^{\frac{1}{2}} = \sup_{\substack{D \in \mathcal{D}_n \\ \|D\|_F \leq 1}} \mathrm{Tr}(DM).$$

where D_n is the set of diagonal matrices of $\mathcal{M}_p$, note that for all $D \in \mathcal{D}_n$, $\|D\|_d = \|D\|_F$. We see directly that $\eta_{(\mathcal{M}_n, \|\cdot\|_d)} = \#\mathcal{D}_n = n$ and therefore for a given $X \in \mathcal{M}_n$ such that $X \propto \mathcal{E}_2$, we can bound $\mathbb{E}\|X\|_d \leq \|\mathbb{E}[X]\|_d + O(\sqrt{n})$.

Proposition 6 is not always the optimal way to bound norms. For instance, given a vector $Z \in \mathbb{R}^p$ and a deterministic matrix $A \in \mathcal{M}_p$, if $Z \propto \mathcal{E}_q$, one is tempted to bound naively thanks to Proposition 6:

---

[9] The notation $Z \in O(\theta) \pm \mathcal{E}_q(\sigma)$ was presented in Definition 2 for linearly concentrated vectors, it can be extended to concentrated random variables.

[10] One must be careful here that Theorem 1 just provides concentration in the Euclidean spaces $(\mathbb{R}^p, \|\cdot\|)$ or $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ from which one can deduce concentration in $(\mathbb{R}^p, \|\cdot\|_\infty)$ or $(\mathcal{M}_{p,n}, \|\cdot\|)$ since for all $x \in \mathbb{R}^p$, $\|x\|_\infty \leq \|x\|$ and for all $M \in \mathcal{M}_{p,n}$, $\|M\| \leq \|M\|_F$. However one cannot obtain a better bound than $\|M\|_* \leq \sqrt{\min(n,p)}\|M\|_F$: this for instance implies that a random matrix $X = (x_1, \ldots, x_n)$ with $x_1, \ldots, x_n$ i.i.d. satisfying $\forall i \in [n]$, $x_i \sim \mathcal{N}(0, I_p)$ follows the concentration $X \propto \mathcal{E}_2(\sqrt{\min(p,n)})$ in $(\mathcal{M}_{p,n}, \|\cdot\|_*)$.

- if $\|\mathbb{E}[Z]\| \leq O(p^{1/q})$, $\mathbb{E}[\|AZ\|] \leq \|A\| \mathbb{E}[\|Z\|] \leq O(\|A\| p^{\frac{1}{q}})$;
- if $\|\mathbb{E}[Z]\| \leq O(1)$, decomposing $A = P^T \Lambda Q$, where $P, Q \in \mathcal{O}_p$, $\Lambda = \mathrm{Diag}(\lambda)$, $\lambda = (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$ and setting $\check{Z} = (\check{Z}_1, \ldots, \check{Z}_p) \equiv QZ$:

  $$\mathbb{E}[\|AZ\|] = \mathbb{E}[\|\Lambda QZ\|] = \mathbb{E}\left[\sqrt{\sum_{i=1}^{p} \lambda_i^2 \check{Z}_i^2}\right] \leq \|\lambda\| \mathbb{E}\left[\|\check{Z}\|_\infty\right] \leq \|A\|_F O\left((\log p)^{\frac{1}{q}}\right).$$

  Note indeed that $\check{Z} \propto \mathcal{E}_2$ and therefore $\mathbb{E}[\|\check{Z}\|_\infty] \leq \|\mathbb{E}[\check{Z}]\|_\infty + O\left((\log p)^{\frac{1}{q}}\right) \leq \|\mathbb{E}[Z]\| + O\left((\log p)^{\frac{1}{q}}\right)$.

However, here, Proposition 6 is suboptimal: one can reach a better bound thanks to the following lemma. We give a result for random vectors and random matrices, they are actually equivalent.

Lemma 4. *Given a random vector $Z \in \mathcal{E}_q$ in $(\mathbb{R}^p, \|\cdot\|_\infty)$ such that $\|\mathbb{E}[Z]\| \leq O(1)$ and a deterministic matrix $A \in \mathcal{M}_p$:*

$$\mathbb{E}[\|AZ\|] \leq O\left(\|A\|_F\right).$$

and given a random matrix $X \in \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ such that $\|\mathbb{E}[X]\|_F \leq O(1)$ and a supplementary deterministic matrix $B \in \mathcal{M}_n$:

$$\mathbb{E}[\|AXB\|_F] \leq O\left(\|A\|_F \|B\|_F\right).$$

Proof. Denoting $\Sigma = \mathbb{E}[ZZ^T] = C_2^Z + \mathbb{E}[Z]\mathbb{E}[Z]^T$, we know from Proposition 5 that $\|\Sigma\| \leq O(1)$; we can then bound with Jensen's inequality:

$$\mathbb{E}[\|AZ\|] \leq \sqrt{\mathbb{E}[Z^T A^T A Z]} = \sqrt{\mathbb{E}[\mathrm{Tr}(\Sigma A^T A)]} \leq \sqrt{\|\Sigma\|} \|A\|_F \leq O(\|A\|_F).$$

The second result is basically the same. If we introduce $\check{X} \in \mathbb{R}^{pn}$ satisfying $\check{X}_{i(j-1)+j} = X_{i,j}$, we know that $\tilde{X} \propto \mathcal{E}_2$ like X (since $\|\tilde{X}\| = \|X\|_F$) and thanks to the previous result we can bound:

$$\mathbb{E}[\|AXB\|_F] = \mathbb{E}[\|A \otimes B\tilde{X}\|] \leq O\left(\|A \otimes B\|_F\right) = O\left(\|A\|_F \|B\|_F\right).$$

Returning to Lipschitz concentration, in order to control the concentration of the sum $X + Y$ or the product XY of two random vectors X and Y, a first step is to express the concentration of the concatenation (X, Y). This last result is easily obtained for the class of linearly concentrated random vectors but a tight concentration of the product with good observable diameter is in general not accessible. In the class of Lipschitz concentrated vectors, the concentration of (X, Y) is far more involved, and assumptions of independence here play a central role (unlike for linear concentration).

4. Concentration of vector concatenation

To understand the issue, consider the example where X and Y are concentrated but not (X, Y). Let X be uniformly distributed on the sphere $\sqrt{p}\mathbb{S}^{p-1}$ and $Y = f(X)$ where, for any $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, $f(x) = x$ if $x_1 \geq 0$ and $f(x) = -x$ otherwise. All linear observations of $X + Y$ are concentrated (it can be shown that the linear concentration is stable through summation) but this is not the case for all Lipschitz observations. The observation $\|X + Y\|$ has a one-out-of-two chance to be equal to 0 or to be equal to $2\sqrt{p}$. This means that the observable diameter is at least of order $O(\sqrt{p})$, which is the metric diameter of $X + Y$, thus contradicting Remark 5. This effect is due to the fact that f is clearly not Lipschitz, and Y in a sense "defies" X (see (Louart and Couillet, 2019, Remark 2.26, Example 2.27) for more details).

Still, there exist two simple ways to obtain the concentration of (X, Y). The first one follows 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.

Proposition 7 (Stability through independent concatenation). *(Ledoux, 2005, Proposition 1.11) Given $(E, \|\cdot\|)$ a sequence of normed vector spaces and two sequences of independent random vectors $X, Y \in E$, such 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), then:*

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

where, for all $x, y \in E^2$, $\|(x, y)\|_{\ell^\infty} = \max(\|x\|, \|y\|)$.[11]

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 $f : (E_p^2, \|\cdot\|_{\ell^\infty}) \to (\mathbb{R}, |\cdot|)$, $\forall t > 0$:

$$\mathbb{P}\left(\left|f(X_p, Y_p) - f(X_p', Y_p')\right| \geq t\right) \leq Ce^{(t/c\sigma_p)^q} + Ce^{(t/c\rho_p)^r}.$$

The sum being a 2-Lipschitz operation (for the norm $\|\cdot\|_{\ell^\infty}$), the concentration of $X + Y$ is easily handled with Proposition 1 and directly follows from the concentration of (X, Y). For products of vectors, more work is required.

5. Concentration of generalized products of random vectors

To treat the product of vectors, we provide a general result of concentration of what could be called "*multilinearly m-Lipschitz mappings*" on normed vector spaces. Instead of properly defining this class of mappings we present it

[11] One could also have considered a number of equivalent norms such as $\|(x, y)\|_{\ell^1} = \|x\| + \|y\|$ or $\|(x, y)\|_{\ell^2} = \sqrt{\|x\|^2 + \|y\|^2}$.

directly in the hypotheses of the theorem. Briefly, these mappings are multivariate functions which are Lipschitz on each variable, with a Lipschitz parameter depending on the product of the norms (or semi-norms) of the other variables and/or constants. To express the observable diameter of such an observation, one needs a supplementary notation.

Given a vector of parameters $(\nu_l)_{l\in[m]} \in \mathbb{R}_+^m$, we denote for any $k \in [m]$:

$$\nu^{(k)} \equiv \max_{1\leq l_1<\cdots<l_k\leq m} \nu_{l_1}\cdots\nu_{l_k} = \nu_{(m-k+1)}\cdots\nu_{(m)},$$

where $\{\nu_{(l)}\}_{l\in[m]} = \{\nu_l\}_{l\in[m]}$ and $\nu_{(1)} \leq \cdots \leq \nu_{(m)}$.

Theorem 2 (Concentration of generalized product). *Given a constant m ($m \leq O(1)$), let us consider:*

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

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

for some (sequence of) positive numbers $\sigma \in \mathbb{R}_+$, and for both norms[12] $\|z_1, \ldots, z_m\|_{\ell^\infty} = \sup_{i=1}^m \|z_i\|_i$ and $\|(z_1, \ldots, z_m)\|'_{\ell^\infty} = \sup_{i=1}^m \|z_i\|'_i$ defined on $E = E_1 \times \cdots \times E_m$.

- *a (sequence of) normed vector spaces $(F, \|\cdot\|)$, a (sequence of) mappings $\phi : E_1, \ldots, E_m \to F$, such that $\forall (z_1, \ldots, z_m) \in E_1 \times \cdots \times E_m$ and $z'_i \in E_i$:*

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

where $\mu_i > 0$ is a (sequence of) positive reals such that $\mu_i \geq \mathbb{E}[\|Z_i\|'_i]$. We further assume $\mu_i \geq O(\sigma)$.[13]

Then we have the concentration :[14]

$$\phi(Z) \propto \max_{l\in[m]} \mathcal{E}_{q/l}\left(\sigma^l \mu^{(m-l)}\right) \tag{9}$$

[12]One just needs to assume the concentration $\|Z_i\|'_i \in \mu_i \pm \mathcal{E}_q(\sigma)$; the global concentration of Z_i for the norm (or seminorm) $\|\cdot\|'_i$ is not required.

[13]This is a very light assumption: it is hard to find any practical example where $\mu_i \ll \sigma$.

[14]Which means that there exist two constants $C, c > 0$ such that for all indexes and for all 1-Lipschitz mapping $f : F \to \mathbb{R}$, and $\forall t > 0$, (10) is satisfied. Here since $m \leq O(1)$, taking the maximum over $l \in [m]$ is equivalent to taking the sum, up to a small change of the constants; we will thus indifferently write $\phi(Z) \propto \max_{l\in[m]} \mathcal{E}_{lq/m}\left(\left(\sigma\mu^{(l-1)}\right)^{\frac{m}{l}}\right)$ or $\phi(Z) \propto \sum_{l=1}^m \mathcal{E}_{lq/m}\left(\left(\sigma\mu^{(l-1)}\right)^{\frac{m}{l}}\right)$.

The proof of the theorem is interesting but somehow technical; we thus defer it to Appendix A.1.

It is explained in Appendix B that, in the setting of Theorem 2, the standard deviation (resp. the r^{th} centered moment with $r \leq O(1)$) of any 1-Lipschitz observation of $\phi(Z)$ is of order $O\left(\sigma\mu^{(m-1)}\right)$ (resp. $O\left((\sigma\mu^{(m-1)})^r\right)$) thus the observable diameter, in the sense given by Remark 5 is given by the first exponential decay, $\mathcal{E}_q\left(\sigma\mu^{(m-1)}\right)$, which represent the guiding term of (9).

We give later a more general setting with Offshot 1 where the variations of ϕ are not controlled with norms (or semi-norms) $(\|\cdot\|'_j)_{j\in[m]\setminus\{i\}}$ but with concentrated variables. The setting is somehow more complex, but sometimes more easy to apply; the proof is morally the same.

One can relax the hypothesis "m constant" and reach good concentration rates when m tends to infinity: this is done in Appendix C.

Remark 7 (Regime decomposition). *Let us rewrite the concentration inequality* (9) *to let appear the implicit parameter t. There exist two constants $C, c > 0$ (in particular, $C, c \leq O(1)$) such that for any 1-Lipschitz mapping $f : F \to \mathbb{R}$, for any $t > 0$:*

$$\mathbb{P}\left(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t\right) \leq C \max_{l\in[m]} \exp\left(-\left(\frac{t/(c\sigma)^l}{\mu^{(m-l)}}\right)^{\frac{q}{l}}\right), \qquad (10)$$

This expression displays m regimes of concentration, depending on $l \in [m]$: the first one, $Ce^{-c(t/\sigma\mu^{(m-1)})^q}$ ($l = 1$), controls the probability for the small values of t, and the last one, $Ce^{-ct^{q/m}/\sigma}$ ($l = m$), controls the tail. Let us define:

$$t_1 = 0; \qquad \forall i \in [m] \setminus \{1\} : t_i \equiv \mu^{(m-i)}\mu_{(i)}^i = \frac{(\mu^{(m-i+1)})^i}{(\mu^{(m-i)})^{i-1}}; \qquad t_{m+1} = \infty.$$

Recalling that $\mu_{(1)} \leq \cdots \leq \mu_{(m)}$, we see that $t_1 \leq \cdots \leq t_m$. One can then show that for any $i \in [m]$, we have the equivalence:

$$t \in [t_i, t_{i+1}] \iff \forall j \in [m] \setminus \{i\} : \exp\left(-\left(\frac{t/(c\sigma)^j}{\mu^{(m-j)}}\right)^{\frac{q}{j}}\right) \leq \exp\left(-\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right).$$

Now, if, for a given $i \in [m]$, $i \geq 2$, $\mu_{(i)} = \mu_{(i+1)}$, then $t_i = t_{i+1}$, therefore the term $C\exp\left(-(\frac{(t/(c\sigma)^i}{\mu^{(m-i)}})^{\frac{q}{i}}\right)$ can be removed from the expression of the concentration inequality since it never reaches the maximum.

In particular, when $\mu_{(1)} = \cdots = \mu_{(m)} \equiv \mu_0$,[15] *$\forall i \in [m]$, $t_i = \mu_0^m$. In this case, there are only two regimes and we can more simply write :*

$$\Phi(Z) \propto \mathcal{E}_q(\sigma\mu_0^{m-1}) + \mathcal{E}_{\frac{q}{m}}(\sigma^m).$$

[15] To be precise, it is sufficient to assume $\mu_{(2)} = \cdots = \mu_{(m)}$, since $\mu_{(1)}$ never appears in the definition of the t_i for $i \in [m]$.

Corollary 1. *In the setting of Theorem 2, when* $\forall i \in [m]$, $\|E[Z_i]\|_i \leq O(\sigma \eta_{\|\cdot\|'}^{1/q})$*, we have the simpler concentration:*

$$\phi(Z) \propto \max_{l \in [m]} \mathcal{E}_{q/l}\left(\sigma^m \eta^{(m-l)}\right),$$

where we denoted $\eta = \left(\eta_{\|\cdot\|_1'}^{1/q}, \dots, \eta_{\|\cdot\|_m'}^{1/q}\right)$.

Proof. Proposition 6 allows us to choose $\mu_i = C' \sigma \eta_{\|\cdot\|_i'}^{1/q}$ for some constant $C' > 0$; we thus retrieve the result thanks to Theorem 2.

Let us give examples of "multilineary Lipschitz mappings" that would satisfy the hypotheses of Theorem 2.

Example 4 (Entry-wise product). *Letting* $\odot$ *be the entry-wise product in* $\mathbb{R}^p$ *defined as* $[x \odot y]_i = x_i y_i$ *(it is the Hadamard product for matrices),* $\phi : (\mathbb{R}^p)^m \ni (x_1, \dots, x_m) \mapsto x_1 \odot \cdots \odot x_m \in \mathbb{R}^p$ *is multilinearly Lipschitz since we have for all* $i \in [m]$*:*

$$\|x_1 \odot \cdots \odot x_{i-1} \odot (x_i - x_i') \odot x_{i+1} \odot \cdots \odot x_m)\| \leq \left(\prod_{\substack{j=1 \\ j \neq i}}^{m} \|x_j\|_\infty\right) \|x_i - x_i'\|,$$

for all vectors $x_1, \dots, x_m, x_1', \dots, x_m' \in \mathbb{R}^p$*. As a practical case, if* $(Z_1, \dots, Z_m) \propto \mathcal{E}_2$ *and* $\forall i \in [m]$, $\|\mathbb{E}[Z_i]\|_\infty \leq O(1)$*, then Corollary 1 and Remark 7 imply:*

$$Z_1 \odot \cdots \odot Z_m \propto \mathcal{E}_2\left(\log(p)^{\frac{m-1}{2}}\right) + \mathcal{E}_{\frac{2}{m}}.$$

It is explained in Appendix B (Remark 14) that in this case, the observable diameter of $Z_1 \odot \cdots \odot Z_m$ *is provided by* $\mathcal{E}_2\left(\log(p)^{\frac{m-1}{2}}\right)$*: as such, under this very common setting, the entry-wise product has almost no impact on the rate of concentration.*

Example 5 (Matrix product). *The mapping* $\phi : (\mathcal{M}_p)^q \ni (M_1, \dots, M_q) \mapsto M_1 \cdots M_m \in \mathcal{M}_p$ *is multilinearly Lipschitz since for all* $i \in [m]$ *:*[16]

$$\|M_1 \cdots M_{i-1}(M_i - M_i')M_{i+1} \cdots M_m)\|_F \leq \left(\prod_{\substack{j=1 \\ j \neq i}}^{m} \|M_j\|\right) \|M_i - M_i'\|_F,$$

[16] One could have equivalently considered, for even $m \in \mathbb{N}$, the mapping $\phi : \mathcal{M}_{p,n}^m \to \mathcal{M}_{p,n}$ satisfying $\forall M_1, \dots, M_m \in \mathcal{M}_{p,n}$, $\phi(M_1, \dots, M_m) = M_1 M_2^T \cdots M_{m-1} M_m^T$.

for all matrices $M_1, \dots, M_m, M_1', \dots, M_m' \in \mathcal{M}_p$. Given m random matrices $X_1, \dots, X_m \in \mathcal{M}_p$ such that $(X_1, \dots, X_m) \propto \mathcal{E}_2$ and $\forall i \in [m]$, $\|\mathbb{E}[X_i]\| \leq O(\sqrt{p})$, Corollary 1 implies:

$$X_1 \cdots X_m \propto \mathcal{E}_2\left(p^{\frac{m-1}{2}}\right) + \mathcal{E}_{\frac{2}{m}}.$$

In particular, for a "data" matrix[17] *$X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$ satisfying $X \propto \mathcal{E}_2$ and $\mathbb{E}[\|X\|] \leq O(\sqrt{p+n})$, the sample covariance matrix satisfies the concentration:*

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

which provides an observable diameter of order $O(1/\sqrt{n})$ when $p \leq O(n)$.

Example 6 (Composition). *Beyond the linear case, consider the composition of functions defined on $\mathcal{L}ip(\mathbb{R})$, the set of Lipschitz and bounded functions of $\mathbb{R}$. Given $f \in \mathcal{L}ip(\mathbb{R})$, we denote:*

$$\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)| \qquad\qquad \|f\|_{\mathcal{L}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|}$$

($\|f\|_{\mathcal{L}}$ is not a norm but a semi-norm). The mapping $\phi : (\mathcal{L}ip(\mathbb{R}))^m \ni (f_1, \dots, f_m) \mapsto f_1 \circ \cdots \circ f_m \in \mathcal{L}ip(\mathbb{R})$ is multilinearly Lipschitz and $\forall f_1 \dots f_m, f_i' \in \mathcal{L}ip(\mathbb{R})$:

$$\begin{aligned}\|f_1 \circ \cdots \circ f_m - f_1 \circ \cdots \circ f_{i-1} \circ f_i' \circ f_{i+1} \circ \cdots \circ f_m\|_\infty& \\ \leq \|f_1\|_{\mathcal{L}} \cdots \|f_{i-1}\|_{\mathcal{L}} \|f_i - f_i'\|_\infty&.\end{aligned}$$

Therefore, assuming $f_1, \dots, f_m$ all $O(1)$-Lipschitz and satisfying the concentration $(f_1, \dots, f_m) \propto \mathcal{E}_q(\sigma)$, thanks to Theorem 2, the following concentration holds :

$$f_1 \circ \cdots \circ f_m \propto \mathcal{E}_q(\sigma) + \mathcal{E}_{q/m}(\sigma^m).$$

When $f_1 = \cdots = f_m \equiv f$ and f is λ-Lipschitz with $1 - \lambda \geq O(1)$, we can follow the dependence of the concentration of $f \circ \cdots \circ f$ on m thanks to Offshot 3 that induces for any sequence of integers m the concentration inequality:

$$f^m \propto \mathcal{E}_q\left(\sigma(1-\varepsilon)^m\right) + \mathcal{E}_{q/m}\left((\kappa\sigma)^m\right)$$

for some constants $\kappa, \varepsilon > 0$.

[17] That is, a matrix whose columns contain vectors of "data", as per data science terminology.

6. Generalized Hanson-Wright theorems

To give some more elaborate consequences of Theorem 2, let us first provide a matricial version of the popular Hanson-Wright concentration inequality, Hanson and Wright (1971).

Proposition 8 (Hanson-Wright). *Given two random matrices $X, Y \in \mathcal{M}_{p,n}$, assume that $(X, Y) \propto \mathcal{E}_2$ and $\|\mathbb{E}[X]\|_F, \|\mathbb{E}[Y]\|_F \leq O(1)$ (as $n, p \to \infty$). Then, for any deterministic matrix $A \in \mathcal{M}_p$, we have the linear concentration (in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$):*

$$Y^T AX \in \mathcal{E}_2\left(\|A\|_F\right) + \mathcal{E}_1(\|A\|).$$

This proposition, which provides a result in terms of linear concentration, points out an instability of the class of Lipschitz concentrated vectors which (here through products) degenerates into a mere linear concentration. This phenomenon fully justifies the introduction of the notion of linear concentration: it will occur again in Proposition 10 and Lemma 8. We present the proof directly here as it is a short and convincing application of Theorem 2.

Proof. Let us first assume that $\|A\| \leq 1$ and consider a deterministic matrix $B \in \mathcal{M}_n$ such that $\|B\|_F \leq 1$. We then introduce the semi-norm $\|\cdot\|_{A,B}$ defined on $\mathcal{M}_{p,n}$ and satisfying for all $M \in \mathcal{M}_{p,n}$, $\|M\|_{A,B} \equiv \|AMB\|_F$. Note that for any $M, P \in \mathcal{M}_{p,n}$:

$$\mathrm{Tr}(BMAP^T) \leq \begin{cases} \|M\|_{A,B}\|P\|_F \\ \|M\|_F\|P\|_{A,B}. \end{cases}$$

Thanks to Lemma 4, $\mathbb{E}[\|X\|_{A,B}], \mathbb{E}[\|Y\|_{A,B}] \leq O(\|A\|_F)$. Besides, since $(X, Y) \propto \mathcal{E}_q$ in $((\mathbb{R}^p)^2, \|\cdot\|_{\ell^\infty})$, and $\|A\|, \|B\| \leq 1$, we also know that $(X, Y) \propto \mathcal{E}_q$ in $((\mathbb{R}^p)^2, \|\cdot\|_{A,B,\ell^\infty})$ where $\|(M, P)\|_{A,B,\ell^\infty} = \max(\|M\|_{A,B}, \|P\|_{A,B})$. Therefore, the hypotheses of Theorem 2 are satisfied with a tuple $\mu = (\mu_1, \mu_2) \in \mathbb{R}_+^2$ for which $\mu_1, \mu_2 \leq O(\|A\|_F)$; we then deduce :

$$X^T AY \in \mathcal{E}_q\left(\|A\|_F\right) + \mathcal{E}_{\frac{q}{2}}.$$

If $\|A\| > 1$, one can still show that $\frac{1}{\|A\|} X^T AY \in \mathcal{E}_q\left(\|A\|_F/\|A\|\right) + \mathcal{E}_{q/2}$ and retrieve the result thanks to Proposition 1.

Remark 8. *In Vu and Wang (2014) and Adamczak (2015), the result is obtained assuming convex concentration for $X = Y \in \mathbb{R}^p$, i.e., the inequalities of Definition 1 are satisfied for all 1-Lipschitz and convex functionals. This definition is less constrained, thus the class of convexly concentrated random vector is larger*[18] *than the class of Lipschitz concentrated random vectors. A well-known*

[18] It is even strictly larger as it was shown in Talagrand (1988) that the uniform distribution on $\{0, 1\}^p$ is convexly concentrated but not Lipschitz concentrated (with interesting concentration speed).

theorem of Talagrand (1995) provides the concentration of the Lipschitz and convex observations of any random vector X built as an affine transformation of a random vector with bounded (with respect to p) and independent entries.

These looser hypotheses are not very hard to handle in this particular case of quadratic functionals since these observations exhibit convex properties. The main issue is to find a result analogous to Lemma 1 to show the convex concentration of $X^T AX$ on events $\{\|X\| \leq K\}$ for $K > 0$ (note that these events are associated to convex subsets of $\mathbb{R}^p$). These details go beyond the scope of the article: we have shown in the ongoing work Louart and Couillet (2021) that Theorem 2 can extend to entry-wise products of m convexly random vectors and to matrix products of m convexly concentrated random matrices (for the latter operations, the concentration is not as good as in the Lipschitz case).

Remark 9. *In Adamczak (2015), the concentration is even expressed on the random variable $\sup_{A\in\mathcal{A}} X^T AX$ where $\mathcal{A}$ is a bounded set of matrices (and $X \in \mathbb{R}^p$). The author indeed obtains the concentration inequality :*

$$\sup_{A\in\mathcal{A}} X^T AX \propto \mathcal{E}_2(\|X\|_{\mathcal{A}}) + \mathcal{E}_1(\sup_{A\in\mathcal{A}} \|A\|),$$

where $\|X\|_{\mathcal{A}} \equiv \mathbb{E}[\sup_{A\in\mathcal{A}} \|(A+A^T)X\|]$. This result can also be obtained – in the Lipschitz concentration case – thanks to Theorem 2 since for any $x_1, x_2, y \in \mathbb{R}^p$:

$$\left| \sup_{A\in\mathcal{A}} x_1^T Ay - \sup_{A\in\mathcal{A}} x_2^T Ay \right| \leq \|x_1 - x_2\| \sup_{A\in\mathcal{A}} \|Ay\|$$

and the same kind of inequality naturally holds for the variations over y.

Let us end this section with a useful consequence of Proposition 8.

Corollary 2. *Given a deterministic matrix $A \in \mathcal{M}_p$ satisfying $\|A\|_F \leq 1$ and two random matrices $X = (x_1, \ldots, x_n), Y = (y_1, \ldots, y_n) \in \mathcal{M}_{p,n}$ satisfying $(X, Y) \propto \mathcal{E}_2$ and $\sup_{i\in[n]} \|\mathbb{E}[x_i]\|, \|\mathbb{E}[y_i]\| \leq O(1)$ such that we have the concentration:*

$$\left\|Y^T AX\right\|_d \in \mathcal{E}_2\left(\sqrt{\log(np)}\right) + \mathcal{E}_1.$$

If[19]*, in addition, $\|A\|_* \leq O(1)$ or $\sup_{i\in[n]} \|\mathbb{E}[y_i x_i^T]\|_F \leq O(1)$*[20]*, then $\mathbb{E}\left[\left\|X^T AY\right\|_d\right] \leq O(\sqrt{n})$.*

[19] Actually, to bound the expectation we just need the concentration of each of the couples (x_i, y_i) but not of the matrix couple (X, Y). Recall (see Section 4) that the concentration of each the $x_1, \ldots, x_n$ does not imply the concentration of the whole matrix X, even if the columns are independent. to trackle this issue, some authors Pajor and Pastur (2009) require a logconcave distribution for all the columns because the product of logconcave distribution is also logconcave. However our assumptions are more general because they allow to take for x_i any $O(1)$-Lipschitz transformation of a Gaussian vector which represents a far larger class of random vectors.

[20] To be precise, one just needs $\sup_{i\in[n]} |\mathbb{E}[x_i^T Ay_i]| \leq O(1)$.

Remark 10. *Recall from Proposition 5 that if* $(x_i, y_i) \in \mathcal{E}_2$ *and* $\|\mathbb{E}[(x_i, y_i)]\|_F \leq O(1)$ *(as under the hypotheses of Corollary 2),* $\|\mathbb{E}[y_i x_i^T]\| \leq O(1)$ *and therefore,* $\|\mathbb{E}[y_i x_i^T]\|_F \leq \sqrt{p}\|\mathbb{E}[y_i x_i^T]\| \leq O(\sqrt{p})$*. In particular, when* x_i *is independent with* y_i *(and* $\|\mathbb{E}[x_i]\|, \|\mathbb{E}[y_i]\| \leq O(1)$*), we can bound* $\|\mathbb{E}[y_i x_i^T]\|_F \leq \|\mathbb{E}[y_i]\|\|\mathbb{E}[x_i]\| \leq O(1)$*.*

Proof (Proof of Corollary 2). Let us decompose $A = U^T \Lambda V$ with $\Lambda = \mathrm{Diag}(\lambda) \in \mathcal{D}_n$ and $U, V \in \mathcal{O}_p$, noting $\check{X} \equiv VX$ and $\check{Y} \equiv UY$ we have the identity:

$$\|Y^T A X\|_d \leq \sup_{\substack{D \in \mathcal{D}_n \\ \|D\|_F \leq 1}} \mathrm{Tr}(D \check{Y}^T \Lambda \check{X}) \leq \sup_{\|d\| \leq 1} d^T (\check{X} \odot \check{Y}) \lambda \leq \|\check{X} \odot \check{Y}\| \leq \|\check{X}\|_F \|\check{Y}\|_\infty,$$

and the same way, $\|Y^T A X\|_d \leq \|\check{X}\|_\infty \|\check{Y}\|_F$. Now we can bound thanks to Proposition 6:

$$\mathbb{E}[\|\check{X}\|_\infty] \leq \left\|\mathbb{E}[\check{X}]\right\|_\infty + O(\sqrt{\log(pn)}) \leq \|\mathbb{E}[X]\| + O(\sqrt{\log(pn)}) \leq O(\sqrt{\log(pn)}),$$

and the same holds for $\mathbb{E}[\|\check{Y}\|_\infty]$. Therefore, applying Theorem 2 to the mapping $\phi : (\check{X}, \check{Y}) \in \mathcal{M}_{p,n}^2 \mapsto \|\check{Y}^T \Lambda \check{X}\|_d$, we obtain the looked for concentration.

To bound the expectation, we start with the identity $\left\|X^T A Y\right\|_d = \sqrt{\sum_{i=1}^n (x_i^T A y_i)^2}$, and we note that the hypotheses of Proposition 8 are satisfied and therefore $x_i^T A y_i \in \mathcal{E}_1$. Now, if $\|A\|_* \leq 1$, we can bound:

$$\left|\mathbb{E}[x_i^T A y_i]\right| = \left\|\mathbb{E}[y_i x_i^T]\right\| \|A\|_* \leq O(1),$$

thanks to Proposition 5 ($(x_i, y_i) \propto \mathcal{E}_2$ and $\|\mathbb{E}[x_i]\mathbb{E}[y_i]^T\| \leq O(1)$). The same bound is true when $\left\|\mathbb{E}[y_i x_i^T]\right\|_F \leq O(1)$ because $\left|\mathbb{E}[x_i^T A y_i]\right| \leq \left\|\mathbb{E}[y_i x_i^T]\right\|_F \|A\|_F$. As a consequence, $x_i^T A y_i \in O(1) \pm \mathcal{E}_1$ (with the same concentration constants for all $i \in [n]$), and we can bound:

$$\mathbb{E}\left[\left\|X^T A Y\right\|_d\right] \leq \sqrt{\sum_{i=1}^n \mathbb{E}\left[(x_i^T A y_i)^2\right]} \leq O(\sqrt{n}).$$

Let us now give an example of application of Theorem 2 when $m \geq 3$.

7. Concentration of XDY^T

Considering 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 $\|\mathbb{E}[D]\|, \|\mathbb{E}[X]\|_F, \|\mathbb{E}[Y]\|_F \leq O(1)$ we wish to study the concentration of XDY^T. Theorem 2 just allows us to obtain the concentration $XDY^T \propto \mathcal{E}_2(n) + \mathcal{E}_1(\sqrt{n}) + \mathcal{E}_{2/3}$ since we cannot get a better bound than $\|XDY^T\|_F \leq \|X\|\|D\|_F\|Y\|$. However, considering some particular observations on XDY^T, it appears that the observable diameter can be smaller than n. Next Propositions reveal indeed that for any deterministic $u \in \mathbb{R}^p$ and $A \in \mathcal{M}_p$:

1. $XDY^T u$ is Lipschitz concentrated with an observable diameter of order $O(\|u\|\sqrt{(n+p)\log(n)})$
2. $\mathrm{Tr}(AXDY^T)$ is concentrated with a standard deviation of order $O(\|A\|_F\sqrt{(n+p)\log(np)})$ if $\sup_{i\in[n]}\|\mathbb{E}[x_i y_i^T]\|_F \leq O(1)$ and $O(\|A\|_*\sqrt{(n+p)\log(np)})$ otherwise.

Proposition 9. *Given three random matrices $X, Y = (y_1, \ldots, y_n) \in \mathcal{M}_{p,n}$ 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 $\|\mathbb{E}[D]\|, \sup_{i\in[n]}\|\mathbb{E}[y_i]\| \leq O(\sqrt{\log n})$, for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$:*

$$XDY^T u \propto \mathcal{E}_2\left(\sqrt{(p+n)\log n}\right) + \mathcal{E}_1\left(\sqrt{p+n}\right) + \mathcal{E}_{2/3} \quad in \quad (\mathbb{R}^p, \|\cdot\|).$$

Proof. The Lipschitz concentration of $XDY^T u$ is obtained thanks to the inequalities:

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

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 $\mathbb{R}^n$), we then have:

- $\eta_X \equiv \mathbb{E}[\|X\|] \leq \|\mathbb{E}[X]\| + O(\sqrt{p+n}) \leq O(\sqrt{p+n})$,
- $\eta_D \equiv \mathbb{E}[\|D\|] \leq \|\mathbb{E}[D]\| + O(\sqrt{\log(n)}) \leq O(\sqrt{\log(n)})$,
- $\eta_{Y^T u} \equiv \mathbb{E}[\|Y^T u\|_\infty] \leq \sup_{i\in[n]}\|\mathbb{E}[y_i^T u]\| + O(\sqrt{\log(n)}) \leq O(\sqrt{\log(n)})$.

To obtain the result, we then employ Corollary 1 for the tuple:

$$\eta = \left(O(\sqrt{n}), O(\sqrt{\log(np)}, O(\sqrt{\log(np)}\right)$$

satisfying:

- $\eta^{(3-1)} = O(\max(\sqrt{(p+n)\log(n)}, \log(n))) \leq O(\sqrt{(p+n)\log(n)})$
- $\eta^{(3-2)} = O(\max(\sqrt{p+n}, \sqrt{\log(n)}) \leq O(\sqrt{p+n})$.

To express the concentration of $\mathrm{Tr}(AXDY^T)$, it is convenient to introduce the following offshot of Theorem 2 based on somehow elaborate but actually simpler assumptions.

Offshot 1. *In the setting of Theorem 2, let us assume that there exist a mapping for all $i \in [m]$ $\Psi_i : E_{+i} \to \mathbb{R}$, where $E_{+i} = E_1 \times \cdots \times E_m \times E_i$, such that $\forall i \in [n]$, $\forall z = (z_1, \ldots, z_m) \in E = E_1 \times \cdots \times E_m$ and $z_i' \in E_i$, we have the bound::*

$$\|\phi(z_1, \ldots, z_m) - \phi(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_m)\| \leq \Psi_i(z_{+i})\left\|z_i - z_i'\right\|_i,$$

where $z_{+i} = (z_1, \dots, z_m, z_i') \in E_{+i}$ and for any independent copy $Z' = (Z_1', \dots, Z_m')$ of Z, we have the concentration[21]*:*

$$\Psi_i(Z_{+i)}) \in O(\mu^{(m-1)}) \pm \max_{l \in [m-1]} \mathcal{E}_{q/l}\left(\sigma^l \mu^{(m-l-1)}\right)$$

where $Z_{+i} = (Z_1, \dots, Z_i, Z_{i+1}', \dots, Z_m', Z_i') \in E_{+i}$, for a parameter vector $\mu \in \mathbb{R}_+^m$ such that $1 \leq \mu_{(1)}$ $(\leq \mu_i, \forall i \in [m])$. Then:

$$\phi(Z) \propto \max_{l \in [m]} \mathcal{E}_{q/l}\left(\sigma^l \mu^{(m-l)}\right).$$

The proof is quite similar to the proof of Theorem 2 and is left to Appendix A.2. In the remainder, this result is sometimes applied in cases where:

$$\Psi_l(Z) \in O(\mu^{(m-1)}) \pm \max_{l \in [m-1]} \mathcal{E}_{q/l}\left(\frac{\sigma^l \mu^{(m-l)}}{\mu_{(m)}}\right),$$

as for instance in Proposition 10 below. Although this corresponds to a stronger setting, the most important element, which provides the observable diameter of $\phi(Z)$, is the order of the expectation of $\Psi_l(Z)$ which still equals $O(\mu^{(m-1)})$.

Remark 11. *Theorem 2 could be seen as an iterative consequence to Offshot 1 as if we assume Offshot 1 and Theorem 2 up to $m = m_0 - 1$ and want to show its validity for $m = m_0$. We can show with the iteration hypothesis for $m = m_0 - 1$ that for all $i \in [m]$:*

$$\|Z_1\|_1' \cdots \|Z_{i-1}\|_{i-1}' \|Z_{i+1}\|_{i+1}' \cdots \|Z_m\|_m' \in O(\mu_{-i}^{(m-1)}) + \sup_{l \in [m-1]} \mathcal{E}_l\left(\mu_{-i}^{(m-1-i)} \sigma^i\right),$$

where $\mu_{-i} \equiv \mu_1 \cdots \mu_{i-1} \mu_{i-1} \cdots \mu_m$. Now, since for all $k \in [m-1]$, $\mu_{-i}^{(l)} \leq O(\mu^{(l)})$, we retrieve the hypotheses of Offshot 1 and we can prove Theorem 2 for $m = m_0$.

The concentration of $\mathrm{Tr}(AXDY^T)$ signifies a linear concentration of XDY^T, demonstrating as in Proposition 8 the relevance of the notation of linear concentration. Note besides that this result can be seen as a weak offshot of Hanson-Wright concentration inequality if one takes $D = \sqrt{n} E_{1,1}$, where $[E_{1,1}]_{i,j} = 0$ for all $(i,j) \neq (1,1)$ and $[E_{1,1}]_{1,1} = 1$.

Proposition 10. *Given three random matrices $X = (x_1, \dots, x_n), Y = (y_1, \dots, y_n) \in \mathcal{M}_{p,n}$ and $D \in \mathcal{D}_n$ such that $(X, Y, D) \propto \mathcal{E}_2$, $\|\mathbb{E}[D]\|_F \leq O(\sqrt{n})$, $\|\mathbb{E}[X]\|_F, \|\mathbb{E}[Y]\|_F \leq O(1)$, we have the linear concentration*[22]*:*

$$XDY^T \in \mathcal{E}_1\left(\sqrt{n}\right) + \mathcal{E}_{2/3} \qquad \text{in } (\mathcal{M}_p, \|\cdot\|).$$

[21] As explained inDefinition 2, this expression means that $\Psi_l(Z, Z^{(i)}) \in \max_{l \in [m-1]} \mathcal{E}_{q/l}\left(\sigma^l \mu^{(m-l-1)}\right)$ and $\mathbb{E}[\Psi_l(Z, Z')] \leq O(\mu^{(m-1)})$.

[22] The estimation of $\mathbb{E}[XDY^T]$ is done in Proposition 11

If, in addition, $\sup_{i\in[n]} \|\mathbb{E}[y_i x_i^T]\|_F \leq O(1)$*:*

$$XDY^T \in \mathcal{E}_1\left(\sqrt{n}\right) + \mathcal{E}_{2/3} \qquad\qquad \textit{in } (\mathcal{M}_p, \|\cdot\|_F).$$

Before proving this corollary let us give a preliminary lemma of independent interest.

Lemma 5. *Given two random matrices* $X \in \mathcal{M}_{p,n}$ *and* $D \in \mathcal{D}_n$ *such that* $(X, D) \propto \mathcal{E}_2$*,* $\|\mathbb{E}[X]\| \leq O(1)$ *and* $\|\mathbb{E}[D]\|_F \leq O(\sqrt{n})$ *and a deterministic matrix* $A \in \mathcal{M}_{p,n}$ *such that* $\|A\|_F \leq 1$*, we have the concentration:*

$$\|AXD\|_F \in O(\sqrt{n}) \pm \mathcal{E}_2(\sqrt{\log(np)}) + \mathcal{E}_1.$$

Proof. With the same decomposition $A = U^T \Lambda V$ and notation $\check{X} \equiv VX$ as in the proof of Corollary 2, we have the identity:

$$\|AXD\|_F = \|\Lambda \check{X} D\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^p \lambda_j^2 \check{X}_{i,j}^2 D_i^2} \leq \|\check{X}\|_\infty \|D\|_F,$$

and besides, $\|AXD\|_F \leq \|\check{X}\|\|D\|$, we can thus employ Theorem 2 to the mapping $\phi : (\check{X}, D) \in \mathcal{M}_{p,n} \times \mathcal{D}_n \to \|U^T \Lambda \check{X} D\|_F$, to set:

$$\|AXD\|_F \propto \mathcal{E}_2(\sqrt{\log(np)}) + \mathcal{E}_1.$$

To bound the expectation, let us note that for all $i \in [n], j \in [p]$, $\check{X}_{i,j} \in O(1) \pm \mathcal{E}_2$ and $D_i \in O(O(|\mathbb{E}[D_i]|)) \pm \mathcal{E}_2$, therefore, $\check{X}_{i,j} D_i \in O(|\mathbb{E}[D_i]|) \pm \mathcal{E}_2(|\mathbb{E}[D_i]|+1) + \mathcal{E}_1$, and we can estimate:

$$\begin{aligned}
\mathbb{E}[\check{X}_{i,j}^2 D_i^2] &= \mathbb{E}\left[\left(\check{X}_{i,j} D_i - \mathbb{E}[\check{X}_{i,j} D_i]\right)^2\right] + \mathbb{E}[\check{X}_{i,j} D_i]^2 \\
&\leq O((1 + \mathbb{E}[D_i])^2) + \mathbb{E}[\check{X}_{i,j}^2]\mathbb{E}[(D_i - \mathbb{E}[D_i])^2]^2 + \mathbb{E}[\check{X}_{i,j}^2]\mathbb{E}[D_i]^2 \\
&\leq O(1 + \mathbb{E}[D_i]^2) + O(1) + O(\mathbb{E}[D_i]^2),
\end{aligned}$$

with constants independent of i, j. Finally, we can bound:

$$\mathbb{E}[\|AXD\|_F] \leq \sqrt{\sum_{i=1}^n \sum_{j=1}^p \lambda_j^2 \mathbb{E}[\check{X}_{i,j}^2 D_i^2]} \leq O\left(\sqrt{\left(\sum_{i=1}^n 1 + \mathbb{E}[D_i]^2\right)\left(\sum_{j=1}^p \lambda_j^2\right)}\right) \leq O(\sqrt{n})$$

since $\|\mathbb{E}[D]\|_F \leq O(\sqrt{n})$ and $\|\lambda\| = \|A\|_F \leq 1$.

Proof (Proof of Proposition 10). Considering a deterministic matrix $A \in \mathcal{M}_{p,n}$, we will assume that $\|A\|_F \leq 1$ if $\sup_{i\in[n]} \|\mathbb{E}[y_i x_i^T]\|_F \leq O(1)$ (to show a concentration in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$) and that $\|A\|_* \leq 1$ otherwise (to show a concentration in $(\mathcal{M}_{p,n}, \|\cdot\|)$). In both cases, Corollary 2 and Lemma 5 allows us to set:

$$\|Y^T AX\|_d,\ \|AXD\|_F,\ \|DY^T A\|_F \in O(\sqrt{n}) \pm \mathcal{E}_2(\sqrt{n}) + \mathcal{E}_1.$$

Besides, we can bound:

$$\mathrm{Tr}(AXDY^T) \leq \begin{cases} \|AXD\|_F\|Y\|_F \\ \|DY^TA\|_F\|X\|_F \\ \|Y^TAX\|_d\|D\|_d, \end{cases}$$

which allows us to conclude thanks to Offshot 1 applied with the parameter vector $\mu = (1, 1, \sqrt{n})$:

$$\mathrm{Tr}(AXDY^T) \propto \mathcal{E}_2(\sqrt{n}) + \mathcal{E}_1(\sqrt{n}) + \mathcal{E}_{\frac{2}{3}} \propto \mathcal{E}_1(\sqrt{n}) + \mathcal{E}_{\frac{2}{3}}.$$

In the setting of Proposition 9, once one knows that XDY is concentrated it is natural to look for a simple deterministic equivalent. The next proposition help us for such a design. Note that the hypotheses are far lighter, in particular, we just need the linear concentration of D.

Proposition 11. *Given three random matrices $D \in \mathcal{D}_n$, $X = (x_1, \ldots, x_n), Y = (y_1, \ldots, y_n) \in \mathcal{M}_{p,n}$ and a deterministic matrix $\tilde{D} \in \mathcal{D}_n$, such that $D \in \tilde{D} \pm \mathcal{E}_2$ in $(\mathcal{D}_n, \|\cdot\|)$ and for all*[23] *$i \in [n]$, $(x_i, y_i) \propto \mathcal{E}_2$ and $\sup_{i\in[n]} \|\mathbb{E}[x_i]\|, \|\mathbb{E}[y_i]\| \leq O(1)$, we have the estimate :*

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

We can precise the estimation with supplementary assumptions:

- *if $\|\tilde{D} - \mathbb{E}[D]\|_F \leq O(1)$ then $\left\|\mathbb{E}[X(D - \tilde{D})Y^T]\right\|_F \leq O\left(\sqrt{n\max(p,n)}\right)$*
- *if $\sup_{i\in[n]} \|\mathbb{E}[x_iy_i^T]\|_F \leq O(1)$ then $\left\|\mathbb{E}[XDY^T] - \mathbb{E}[X\tilde{D}Y^T]\right\|_F \leq O\left(n\right)$.*

Proof. Considering a deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\|_F \leq 1$:

$$\begin{aligned} &\left|\mathbb{E}[\mathrm{Tr}(AXDY^T)] - \mathbb{E}[\mathrm{Tr}(AX\mathbb{E}[D]Y^T)]\right| \\ &\qquad \leq \sum_{i=1}^n \left|\mathbb{E}\left[D_ix_i^TAy_i - \mathbb{E}[D_i]x_i^TAy_i\right]\right| \\ &\qquad = \sum_{i=1}^n \left|\mathbb{E}\left[\left(x_i^TAy_i - \mathbb{E}[x_i^TAy_i]\right)(D_i - \mathbb{E}[D_i])\right]\right| \\ &\qquad \leq \sum_{i=1}^n \sqrt{\mathbb{E}\left[\left|x_i^TAy_i - \mathbb{E}[x_i^TAy_i]\right|^2\right]\mathbb{E}\left[\left|D_i - \mathbb{E}[D_i]\right|^2\right]} \quad \leq \quad O\left(n\right) \end{aligned}$$

[23] If we adopt the stronger assumptions $(X,Y) \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|)$ and $\|\mathbb{E}[X]\|_F, \|\mathbb{E}[X]\|_F \leq O(1)$, we can show more directly thanks to Propositions 8 and 6:

$$\begin{aligned} &\left|\mathbb{E}[\mathrm{Tr}(AXDY^T)] - \mathbb{E}[\mathrm{Tr}(AX\mathbb{E}[D]Y^T)]\right| \\ &\qquad = \left|\mathbb{E}\left[\mathrm{Tr}\left(\left(Y^TAX - \mathbb{E}[Y^TAX]\right)(D - \mathbb{E}[D])\right)\right]\right| \\ &\qquad \leq \sqrt{\mathbb{E}\left[\|Y^TAX - \mathbb{E}[Y^TAX]\|_d^2\right]\mathbb{E}\left[\|D - \mathbb{E}[D]\|_d^2\right]} \leq O(n) \end{aligned}$$

thanks to Hölder's inequality applied to the bounds given by Proposition 13 (we know that $D_i \in \mathbb{E}[D_i] \pm \mathcal{E}_2$ and from Proposition 8 that $x_i^T A y_i \in \mathbb{E}[x_i^T A y_i] \pm \mathcal{E}_2 + \mathcal{E}_1$; note that the concentration constants are the same for all $i \in [n]$).

Now, $|\mathbb{E}[x_i^T A y_i]| \leq \|A\|_F \|\mathbb{E}[y_i x_i^T]\|_F \leq O(\sqrt{p})$ thanks to Remark 10 and if $\|\mathbb{E}[D] - \tilde{D}\|_F \leq O(1)$, we can bound:

$$\left|\mathbb{E}\left[\mathrm{Tr}\left(AX\left(\mathbb{E}[D] - \tilde{D}\right)Y^T\right)\right]\right| \leq \sum_{i=1}^n \left|\mathbb{E}\left[x_i^T A y_i\right]\left(\mathbb{E}[D_i] - \tilde{D}_i\right)\right|$$
$$\leq \sup_{i \in [n]} \left|\mathbb{E}\left[x_i^T A y_i\right]\right| \sqrt{n} \left\|\mathbb{E}[D] - \tilde{D}\right\|_F \leq O\left(\sqrt{np}\right).$$

If, $\|\mathbb{E}[D] - \tilde{D}\|_F$ is possibly of order far bigger than $O(1)$, but $\sup_{i \in [n]} \|\mathbb{E}[y_i x_i^T]\|_F \leq O(1)$, then $\sup_{i \in [n]} |\mathbb{E}[x_i^T A y_i]| \leq O(1)$, and we can still bound:

$$\left\|\mathbb{E}\left[\mathrm{Tr}\left(AX\left(\mathbb{E}[D] - \tilde{D}\right)Y^T\right)\right]\right\|_F \leq n \sup_{i \in [n]} \left|\mathbb{E}\left[x_i^T A y_i\right]\right| \left\|\mathbb{E}[D] - \tilde{D}\right\| \leq O(n).$$

Let us end this article with a non multi-linear application of Theorem 2.

8. Concentration of the resolvent $(I_p - \frac{1}{n} X D Y^T)^{-1}$

We study here the concentration of a resolvent $Q = (I_p - \frac{1}{n} X D Y^T)^{-1}$ with the assumption of Proposition 10 for X, Y and D (in particular D is random). This object appears in particular when studying robust regression El Karoui et al. (2013); Mai et al. (2019). In several settings, robust regression can be expressed by the following fixed point equation:

$$\beta = \frac{1}{n} \sum_{i=1}^n f(x_i^T \beta) x_i, \quad \beta \in \mathbb{R}^p, \tag{11}$$

where β is the weight vector performing the regression (to classify data, for instance). It was then shown in Seddik et al. (2021) that the estimation of the expectation and covariance of β (and therefore, of the performances of the algorithm) rely on an an estimation of Q, with $D = \mathrm{Diag}(f'(x_i^T \beta))$. To obtain a sharp concentration on Q (as it is done in Theorem 3 below), one has to understand the dependence between Q and x_i, for all $i \in [n]$. This is performed with the notation, given for any $M = (m_1, \ldots, m_n) \in \mathcal{M}_{p,n}$ or any $\Delta = \mathrm{Diag}_{i \in [n]}(\Delta_i) \in \mathcal{D}_n$:

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

The structure of the study of the resolvent is very similar to the one conducted in Section 7 and we will try to draw the maximum of analogy between the two sections. The first theorem should for instance be compared to Proposition 11 and Proposition 10.

Theorem 3. *Given a random diagonal matrix $D \in \mathcal{D}_n$ and two random matrices $X = (x_1, \ldots, x_n), Y = (y_1, \ldots, y_n) \in \mathcal{M}_{p,n}$, in the regime*[24] *$p \leq O(n)$ and under the assumptions:*

- *$(X, D, Y) \propto \mathcal{E}_2$,*
- *all the couples (x_i, y_i) are independent,*
- *$O(1) \leq \sup_{i\in[n]} \|\mathbb{E}[x_i]\|, \sup_{i\in[n]} \|\mathbb{E}[y_i]\| \leq O(1)$,*
- *for all $i \in [n]$, there exists a random diagonal matrix $D^{(i)}$, independent of (x_i, y_i), such that $\sup_{i\in[n]} \|D_{-i} - D_{-i}^{(i)}\|_F \leq O(1)$,*
- *there exist*[25] *three constants $\kappa, \kappa_D, \varepsilon > 0$ ($\varepsilon \geq O(1)$ and $\kappa, \kappa_D \leq O(1)$), such that $\|X\|, \|Y\| \leq \sqrt{n}\kappa$, $\|D\| \leq \kappa_D$ and $\kappa^2\kappa_D \leq 1 - \varepsilon$,*

the expectation of resolvent $Q \equiv (I_p - \frac{1}{n}XDY^T)^{-1}$ can be estimated by the expectation of the random matrix $\tilde{Q} \equiv \left(I_p - \frac{1}{n}X\tilde{D}Y^T\right)^{-1}$, for a deterministic matrix $\tilde{D}$ satisfying[26] *$D \in \tilde{D} \pm \mathcal{E}_2$:*

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

Under this setting, we then have the linear concentration[27]*:*

$$Q \in \tilde{Q} \pm \mathcal{E}_1\left(\frac{\log(n)}{\sqrt{n}}\right) \qquad \textit{in } (\mathcal{M}_p, \|\cdot\|),$$

and if we further assume that $\|\mathbb{E}[y_i x_i^T]\|_F \leq O(1)$:

$$Q \in \tilde{Q} \pm \mathcal{E}_1\left(\frac{\log(n)}{\sqrt{n}}\right) \qquad \textit{in } (\mathcal{M}_p, \|\cdot\|_F).$$

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

[25] The assumptions $\|X\|/\sqrt{n}$ and $\|Y\|/\sqrt{n}$ bounded might look a bit strong (since it is not true for matrices with i.i.d. Gaussian entries) and it is indeed enough to assume that $\mathbb{E}[\|X\|], \mathbb{E}[\|Y\|] \leq \sqrt{n}\kappa - \eta$, for $\eta \geq O(1)$ small and place ourselves – as it is done in – on the event $\{\|Y\|, \|X\| \leq \sqrt{n}\kappa\}$ that has an overwhelming probability to happen since $\|X\|/\sqrt{n} \in \mathbb{E}[\|X\|]/\sqrt{n} \pm \mathcal{E}_2(1/\sqrt{n})$ and the same holds for Y. We however preferred here to make a relatively strong hypothesis not to have supplementary notations and proof precautions, that might have blurred the message.

[26] We already assumed $D \propto \mathcal{E}_2$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$, so we just add here the hypothesis $\|\mathbb{E}[D] - \tilde{D}\|_F \leq O(1)$.

[27] In Definition 2, this notation was introduced for deterministic matrix $\tilde{Q}$. When $\tilde{Q}$ is random, nothing changes, be just careful that this concentration inequality does not directly imply that $Q \in \mathbb{E}[Q] \pm \mathcal{E}_2\left(\frac{1}{n}\sqrt{(p+n)\log(pn)}\right) + \mathcal{E}_1\left(\frac{\sqrt{p+n}}{n}\right) + \mathcal{E}_{\frac{2}{3}}\left(\frac{1}{n}\right)$ (in particular, for any (sequence of) positive value $\sigma > 0$, $Q \in Q \pm \mathcal{E}_2(\sigma)$. However since $\tilde{Q} \propto \mathcal{E}_2(1/\sqrt{n})$ in $(\mathcal{M}_{p,n}, \|\cdot\|)$, we can deduce the linear concentration of Q

Remark 12. *Theorem 3 simplifies greatly the study of the resolvent* $Q = (I_p - \frac{1}{n} XDY^T)^{-1}$ *because it states that it basically behaves like the random matrix* $\tilde{Q} = (I_p - \frac{1}{n} X\tilde{D}Y^T)^{-1}$ *which was far more studied in the random matrix literature (see Silverstein and Bai (1995); Pajor and Pastur (2009), for instance). It can be shown indeed that under the hypotheses of Theorem 3:*

$$\tilde{Q} \in \check{Q}_\delta \pm \mathcal{E}_2 \qquad\qquad \text{in } (\mathcal{M}_{p,n}, \|\cdot\|_F),$$

where for all $\nu \in \mathbb{R}^n$, $\check{Q}_\nu \equiv \left(I_p - \frac{1}{n}\sum_{i=1}^n \frac{\tilde{D}_i \Sigma_i}{1 - \tilde{D}_i \nu_i}\right)^{-1}$ *(when it is defined),* $\forall i \in [n]$, $\Sigma_i \equiv \mathbb{E}[x_i y_i^T]$ *and* $\delta \in \mathbb{R}^n$ *is the unique solution to:*

$$\forall i \in [n] : \delta_i = \frac{1}{n} \operatorname{Tr}\left(\Sigma_i \check{Q}_\delta\right).$$

Remark 13. *Let us give two examples of the matrices* $D^{(i)}$ *that one could choose, depending on the cases:*

- *For all* $i \in [n]$, $D_i = f(x_i, y_i)$ *for* $f : \mathbb{R}^{2p} \to \mathbb{R}$, *bounded, then,* D_i *just depends on* (x_i, y_i) *so one can merely take* $D^{(i)} = D_{-i}$ *for all* $i \in [n]$.
- *For the robust regression described by Equation 11, as in Seddik et al. (2021), we can assume for simplicity*[28] $\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty \leq O(1)$. *If we choose* $D = \operatorname{Diag}(f'(x_i^T\beta))$, *then it is convenient to assume* $\frac{1}{n}\|f'\|_\infty \|X\|^2 \leq 1 - \varepsilon$ *(which implies in particular* $\frac{1}{n}\|X\|\|D\|\|Y\| \leq 1 - \varepsilon$*) so that* β *is well defined, being solution of a contractive fixed point equation. One can further introduce* $\beta^{(i)} \in \mathbb{R}^p$, *the unique solution to*

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

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

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

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

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

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

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

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

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

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

The proof of Theorem 3 follows the scheme of the proof of Proposition 11 which employs Corollary 2, showing the concentration of $\|Y^T AX\|_d$ and requiring in particular:

1. $\|x_i\|, \|y_i\| \leq O(1)$,
2. if $\|A\|_*$ is not of order $O(1)$ then $\|x_i y_i^T\|_F \leq O(1)$.

The preliminary lemmas to the proof of Theorem 3 are here to prove a similar result to Corollary 2, namely the concentration of $\|Y^T QAQX\|_d$, given in Lemma 12. To this end, we fist bound $\|\mathbb{E}[Qx_i]\|$ and $\|\mathbb{E}[Q^T y_i]\|$ from the bound given on $\|\mathbb{E}[x_i]\|$ and $\|\mathbb{E}[y_i]\|$ in Appendix D.2. To treat the case where $\|A\|_* \gg O(1)$, we show that, when $\|x_i y_i^T\|_F \leq O(1)$, then $\|\mathbb{E}[Qx_i y_i^T Q]\|_F \leq O(1)$ directly in Lemma 12, in Appendix D.3. We then have all the elements to prove Theorem 3 in Appendix D.4.

The same way that Theorem 3 can be linked to Proposition 10 giving the linear concentration of XDY^T, the next proposition can be linked to Proposition 9 giving the Lipschitz concentration of $XDY^T u$ for any deterministic $u \in \mathbb{R}^p$. the proof is left in Appendix D.5.

Proposition 12. *In the setting of Theorem 3, for any deterministic vector $u \in \mathbb{R}^p$ such that $\|u\| \leq O(1)$:*

$$Qu \propto \mathcal{E}_1 \left(\sqrt{\frac{\log n}{n}} \right).$$

Conclusion

With the complexity of nowadays machine learning algorithms, it becomes crucial to devise simple and efficient notations to comprehend their structural logic. For that purpose, the present work provides a systematic approach to comprehend the probabilistic issues involving concentrated vectors, as a model for real data, and their use in statistical learning methods. Indeed, on the one hand, as justified in Seddik et al. (2019), the very realistic artificial images created by generative adversarial networks are concentrated random vectors by construction: this strongly suggests that most commonly studied databases satisfy our hypotheses. On the other hand, the flexibility of the hypotheses of Theorem 2 and of Offshot 1 ensures that a wide range of real functionals involved in machine learning problems are concerned by those results.

As such, in essence, the article provides a catalogue of ready-to-use results for a probabilistic approach of machine learning. To summarize, establishing a concentration inequality on a given random quantity Y generally follows the steps:

1. Identify the random vectors $X_1, \ldots, X_m$ (independent or not) upon which Y is built, and verify that $(X_1, \ldots, X_m) \propto \mathcal{E}_2$;
2. Bound the variations of Y with a functional δ_i when X_i varies, $\forall i \in [m]$;

3. Express the concentration of δ_i, for all $i \in [m]$ and deduce the concentration of Y from Offshot 1, or from Theorem 2, depending on δ_i.

Our work strongly relates to general log-concave settings (very similar to the setting we proposed here: in Adamczak (2011) for Wigner matrices and in Pajor and Pastur (2009) for Wishart matrices) for which the asymptotic behavior of the spectral distribution of random matrices was shown only to depend on the first moments of the entries. In these probabilistic contexts, the random objects behave as if the initial data were Gaussian because the only relevant statistics of the asymptotic behavior is composed of the means and covariances of the data. The laborious Gaussian calculus (with the Stein method as in Pastur (2005), possibly combined with Poincaré inequalities as in Chatterjee (2017)) then appears as superfluous and can be replaced by concentration of measure arguments in more general settings. This being said, a result from Klartag (2007); Fleury et al. (2007) establishing a central limit theorem (CLT) for deterministic projections of concentrated random vectors allowed us in a parallel contribution to employ Gaussian inference for the estimation of quantities depending on such projections Seddik et al. (2021). Nonetheless, in this case, a small number of projections do not satisfy the CLT[29], which restricts the application of the argument.

Appendix A. Proof of the concentration of generalized products

Appendix A.1. Proof of Theorem 2

For simplicity, let us reorder the indexing of the argument of ϕ such that $(\mu_1, \ldots, \mu_m) = (\mu_{(1)}, \ldots, \mu_{(m)})$ (i.e., $\mu_1 \leq \cdots \leq \mu_m$). Let us consider $Z_1', \ldots, Z_m'$, m independent copies of $Z_1, \ldots, Z_m$ and a 1-Lipschitz (for the norm $\|\cdot\|$) mapping $f : F \to \mathbb{R}$. For simplicity, we will note $\phi(Z) \equiv \phi(Z_1, \ldots, Z_m)$ (and $\phi(Z') = \phi(Z_1', \ldots, Z_m')$). Given $t > 0$, we wish to bound:[30]

$$\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t\right). \tag{A.1}$$

The map $Z \mapsto f(\phi(Z))$ is not Lipschitz, unless Z is bounded. We thus decompose the probability argument into two events, one with bounded $\|Z\|_{\ell^\infty}'$ and $\|Z'\|_{\ell^\infty}'$ and the complementary with small probability. For all $i \in [m]$, since

[29] Given a random variable $z_1 \sim \mathrm{Unif}([0,1])$ and $p-1$ i.i.d. random variables $z_2, \ldots, z_p \sim \mathcal{N}(0,1)$, we know that $Z = (z_1, \ldots, z_p) \sim \mathcal{E}_2$ but $e_1^T Z \sim \mathrm{Unif}([0,1])$ is not Gaussian. It is stated in Klartag (2007) that for most $u \in \mathbb{S}^{p-1}$, $u^T Z$ is quasi-Gaussian (the measure of the complementary set to such u is exponentially decreasing with the maximal distance in infinity norm between the Gaussian CDF and the CDFs of $u^T Z$).

[30] We choose here to employ the characterization of the concentration with the independent copy, because, at some point of the proof, we restrict ourselves to an event $\mathcal{A}_K$ and then $\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t \mid \mathcal{A}_K\right)$ can be bounded directly from the concentration $(\phi(Z) \mid \mathcal{A}_K) \propto \mathcal{E}_q(K^{(m-1)}\sigma)$ resulting from Lemma 1 and Remark 4. To bound $\mathbb{P}\left(|f(\phi(Z)) - \mathbb{E}[f(\phi(Z))]| \geq t \mid \mathcal{A}_K\right)$, one would have needed to show first that $|\mathbb{E}[f(\phi(Z))] - \mathbb{E}[f(\phi(Z)) \mid \mathcal{A}_K]| \leq O(K^{(m-1)}\sigma)$ to then employ Lemma 3.

$\|Z\|_i' \in \mathbb{E}[\|Z\|_i'] \pm E_q(\sigma)$ and $\mathbb{E}[\|Z\|_i'] \leq \mu_i$ (and the same hold for Z_i'), we can then employ for all $t \geq \mu_i$the identity:

$$\mathbb{P}\left(\|Z\|_i' \geq 2t, \|Z'\|_i' \geq 2t\right) \leq \mathbb{P}\left(\left|\|Z\|_i' - \mathbb{E}\left[\|Z\|_i'\right]\right| \geq 2t - \mu_i\right)^2 \leq Ce^{-(t/c\sigma)^q},$$

for two constants $C, c > 0$. The bounds on each of the $\|Z_i\|_i'$ and $\|Z_i'\|_i'$, $i \in [m]$, depend on the value of t. Let us note for all $i \in [m]$:

$$t_i \equiv \mu^{(m-i)} \mu_i^i = \mu^{(m-i+1)} \mu_i^{i-1} = \frac{(\mu^{(m-i+1)})^i}{(\mu^{(m-i,)})^{i-1}} \qquad \text{and} \qquad t_{m+1} = \infty,$$

and remark that $t_1 \leq \cdots \leq t_m$. If $t \in [t_i, t_{i+1}]$, we decompose the probability (A.1) playing on the realization the event:

$$\mathcal{A}_i \equiv \left\{ \|Z_1\|_1', \|Z_1'\|_1' \leq 2\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}}; \ldots; \|Z_i\|_i', \|Z_i'\|_i' \leq 2\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}}; \right.$$
$$\left. \|Z_{i+1}\|_{i+1}', \|Z_{i+1}'\|_{i+1}' \leq 2\mu_{i+1}; \ldots; \|Z_m\|_m', \|Z_m'\|_m' \leq 2\mu_m \right\}.$$

Noting that for $j \in [m]$:

- if $j \leq i$: $\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}} \geq \left(\frac{t_i}{\mu^{(m-i)}}\right)^{\frac{1}{i}} = \mu_i \geq \mu_j$,
- if $j \geq i+1$: $\mu_j \geq \mu_{i+1} = \left(\frac{t_{i+1}}{\mu^{(m-i)}}\right)^{\frac{1}{i}} \geq \left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}}$,

we can first bound, on the first hand:

$$\begin{aligned}
&\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t, Z \in \mathcal{A}_i^c\right) \leq \mathbb{P}\left(Z \in \mathcal{A}_i^c\right) \\
&\quad \leq \sum_{j=1}^{i-1} \mathbb{P}\left(\|Z_j\| \geq 2\mu_j\right)^2 + \sum_{k=i}^{m} \mathbb{P}\left(\|Z_k\| \geq 2\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}}\right)^2 \\
&\quad \leq mC \exp\left(-\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}. \qquad \text{(A.2)}
\end{aligned}$$

On the other hand, one can show that $f \circ \phi|_{\mathcal{A}_i}$ is λ_i-Lipschitz with:

$$\lambda_i \equiv m2^{m-1}\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{i-1}{i}} \mu^{(m-i)}.$$

Therefore, following Remark 4, since $\mathbb{P}(\mathcal{A}_i) \geq O(1)$:

$$\left(f(\phi(Z)) \mid \mathcal{A}_i\right) \propto \mathcal{E}_q\left(\lambda_i\right),$$

which allows us to bound (for all $t \in [t_i, t_{i-1}]$):

$$\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t, Z \in \mathcal{A}_i\right) \leq C \exp\left(-\frac{t}{(2^m c\sigma)^i \mu^{(m-i)}}\right)^{\frac{q}{i}}. \qquad \text{(A.3)}$$

If $t \in (0, t_1] = (0, \mu^{(m)}]$, we introduce:

$$\mathcal{A}_0 \equiv \{\|Z_1\|_1', \|Z_1'\|_1' \leq 2\mu_1; \ldots ; \|Z_m\|_m', \|Z_m'\|_m' \leq 2\mu_m\},$$

one can still bound $\forall i \in [m]$, $\mu_i \geq \mu_1 = \frac{t_1}{\mu^{(m-1)}} \geq \frac{t}{\mu^{(m-1)}}$, and therefore $\mathbb{P}(\mathcal{A}_0^c) \leq mCe^{-(\frac{t}{c\sigma\mu^{(m-1)}})^q}$, besides, $\phi|_{\mathcal{A}_0}$ is $m2^{m-1}\mu^{(m-1)}$-Lipschitz, therefore we retrieve:

$$\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t, Z \in \mathcal{A}_0\right) \leq C \exp\left(-t/2^m c\sigma\mu^{(m-1)}\right)^q.$$

Combining these bounds for $t \in (0, t_1]$ with (A.2) and (A.3), we obtain the result of the theorem (since $mC, 2^m c \leq O(1)$).

Appendix A.2. Proof of Offshot 1

We assume for simplicity that for all $i \in [m]$, $\mathbb{E}[\Psi_i(Z_{+i})] \leq \mu^{(m-1)}$. With the same notation as before for $t \in [0, t_1]$:

$$\mathcal{A}_0 \equiv \left\{\forall i \in [m] : \Psi_i(Z_{+i}) \leq 2\mu^{(m-1)}\right\},$$

then, since $\mu^{(m-1)} \geq t$, there exist two constants $C, c > 0$ such that we can bound:

$$\begin{aligned}\mathbb{P}\left(A_0^c\right) &\leq m \sup_{i\in[m]} \mathbb{P}\left(|\Psi_i(Z_{+i}) - \mathbb{E}[\Psi_i(Z_{+i})]| \geq \mu^{(m-1)}\right) \\ &\leq m \sup_{\substack{i\in[m]\\ l\in[m-1]}} C \exp\left(-\left(\frac{\mu^{(m-1)}}{c\sigma^l\mu^{(m-l-1)}}\right)^{q/l}\right) \leq m \sup_{\substack{i\in[m]\\ l\in[m-1]}} Ce^{-(t/c\sigma^l\mu^{(m-l-1)})^{q/l}}.\end{aligned}$$

Besides, since ϕ is $m\mu^{(m-1)}$-Lipschitz on $\mathcal{A}_0$, we can bound:

$$\mathbb{P}\left(|\phi(Z) - \phi(Z')| \geq t \mid \mathcal{A}_0\right) \leq Ce^{-(t/mc\mu^{(m-1)}\sigma)^q}.$$

When $t \in [t_i, t_{i+1}] = [\mu^{(m-i)}\mu_{(i)}^i, \mu^{(m-i)}\mu_{(i+1)}^i]$ for $i \in [m-1]$ or $t \in [t_m, \infty) = [\mu_{(m)}^m, \infty)$ for $i = m$, we rather work with the event:

$$\mathcal{A}_i \equiv \left\{\forall i \in [m] : \Psi_i(Z_{+i}) \leq 2t\left(\frac{\mu^{(m-i)}}{t}\right)^{\frac{1}{i}}\right\}.$$

On the first hand, since $t\left(\frac{\mu^{(m-i)}}{t}\right)^{\frac{1}{i}} \geq t_i^{\frac{i-1}{i}}(\mu^{(m-i)})^{\frac{1}{i}} \geq \mu^{(m-i)}\mu_{(i)}^{i-1} \geq \mu^{(m-1)}$, we can bound:

$$\begin{aligned}\mathbb{P}\left(A_i^c\right) &\leq m \sup_{j\in[m]} \mathbb{P}\left(|\Psi_j(Z_{+j}) - \mathbb{E}[\Psi_j(Z_{+j})]| \geq 2t\left(\frac{\mu^{(m-i)}}{t}\right)^{\frac{1}{i}} - \mu^{(m-1)}\right) \\ &\leq m \sup_{l\in[m-1]} C \exp\left(-\left(\frac{t}{c\sigma\mu^{(m-l-1)}}\left(\frac{\mu^{(m-i)}}{t}\right)^{\frac{1}{i}}\right)^q\right).\end{aligned}$$

Now, given $l \in [m-1]$ (and when possibly $i = m$ and $t \in [\mu_{(m)}^m, \infty)$):

- if $m-1 \geq l \geq i$, $\frac{t}{\mu^{(m-l-1)}} \left(\frac{\mu^{(m-i)}}{t} \right)^{\frac{1}{i}} \geq \frac{t}{\mu^{(m-l-1)} \mu_{(i+1)}} \geq \frac{t}{\mu^{(m-l)}}$,
- if $l \leq i-1$, $\frac{t}{\mu^{(m-l-1)}} \left(\frac{\mu^{(m-i)}}{t} \right)^{\frac{1}{i}} \geq \left(\frac{t}{\mu^{(m-l-1)}} \right)^{\frac{i-1}{i}} \geq \frac{t}{\mu^{(m-l)}}$,

since $\frac{t}{\mu^{(m-l-1)}} \geq \frac{t}{\mu^{(m-i)}} \geq \mu_{(i)}^i \geq 1$, by hypothesis. That allows us to bound:

$$\mathbb{P}\left(A_i^c\right) \leq m \sup_{l \in [m-1]} C e^{-(t/c\sigma^l \mu^{(m-l)})^{q/l}}. \tag{A.4}$$

Besides, since ϕ is $mt \left(\frac{\mu^{(m-i)}}{t} \right)^{\frac{1}{i}}$-Lipschitz on $\mathcal{A}_i$, we can bound:

$$\mathbb{P}\left(|\phi(Z) - \phi(Z')| \mid \mathcal{A}_i\right) \leq C e^{-(t/c^i m^i \mu^{(m-i)} \sigma^i)^{\frac{q}{i}}}. \tag{A.5}$$

We then obtain our result combining (A.4) and (A.5) (recall that $c^m, m^m \leq O(1)$).

Appendix B. Moment characterization of multi-regime concentration

Proposition 2, which provides a control of the centered moments of a concentrated vector, cannot be directly applied when the concentration follows differing exponential regimes as in Theorem 2. We give here a generalization of this result.

Proposition 13 (Moment characterization of multi-regime concentration). *Given a constant integer $m \in \mathbb{N}$ and $m+1$ (sequences of) positive values*[31] *$\sigma, \mu_1 \ldots, \mu_m$, a random vector $Z \in E$ satisfies the concentration:*

$$Z \propto \max_{l \in [m]} \mathcal{E}_{q/l}(\sigma^l \mu^{(m-l)})$$

if and only if there exist two constants $C, c > 0$ such that for all 1-Lipschitz mapping $f : E \to \mathbb{R}$, and for all $r > 0$, we have the bound:[32,33]

$$\mathbb{E}\left[|f(Z) - \mathbb{E}\left[f(Z)\right]|^r\right] \leq C \max_{l \in [m]} \left(\frac{rl}{q} \right)^{\frac{rl}{q}} (c\sigma^l \mu^{(m-l)})^r. \tag{B.1}$$

[31] Instead of this very particular setting originating from Theorem 2, we may replace in the theorem the quantities $\sigma^l \mu^{(m-l)}$ by positive variables σ_l that should satisfy $\sigma_1 \geq \cdots \geq \sigma_m$ and $\forall k \in [m-1]$ and $l \in \{2, \ldots, m\}$ such that $k \leq l$:

$$\sigma_k^{k+1}/\sigma_{k+1}^k \leq \sigma_k^l/\sigma_l^k \leq \sigma_{l-1}^l/\sigma_l^{l-1}.$$

[32] Since m, l and q are constants, one could replace (B.1) with:

$$\mathbb{E}\left[|f(Z) - \mathbb{E}\left[f(Z)\right]|^r\right] \leq C \max_{l \in [m]} \left(r^{\frac{rl}{q}} (c\sigma_l)^r \right),$$

but the formulation of (B.1) is more adapted to the proof.

[33] Here, one could have replaced in the inequality $\mathbb{E}[f(Z)]$ by any median of $f(Z)$ or by $f(Z')$, for any Z' an independent copy of Z.

Proof. This proof is mainly a rewriting of (Ledoux, 2005, Proposition 1.10) with the regime decomposition employed in Appendix A.1. Let us note for simplicity, for any $l \in [m]$, $\sigma_l = \sigma^l \mu^{(m-l)}$. We start with the direct implication which is easier to prove. Assume that there exists two constants $C, c > 0$ such that for any 1-Lipschitz mapping $f : E \to \mathbb{R}$:

$$\forall t > 0, \quad \mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq C \max_{l \in [m]} e^{-(t/c\sigma_l)^{\frac{q}{l}}}.$$

With the Fubini Theorem, one can bound for all $r > 0$ Given $r > 0$:

$$\begin{aligned}
\mathbb{E}\left[|f(Z) - \mathbb{E}[f(Z)]|^r\right] &= \int_0^\infty \mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]|^r \geq t\right) dt \\
&= \int_0^\infty r t^{r-1} \mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) dt \\
&\leq \max_{l \in [m]} Cr \int_0^\infty t^{r-1} e^{-(t/c\sigma_l)^{\frac{q}{l}}} \, dt \\
&= \max_{l \in [m]} C (c\sigma_l)^r r \int_0^\infty t^{r-1} e^{-t^{\frac{q}{l}}} \, dt,
\end{aligned}$$

and, if we assume that $r \geq q$ ($\geq \frac{q}{l}$ for all $l \in [m]$) :

$$r \int_0^\infty t^{r-1} e^{-t^{\frac{q}{l}}} \, dt = \frac{lr}{q} \int_0^\infty t^{\frac{rl}{q} - 1} e^{-t} dt = \frac{lr}{q} \Gamma\left(\frac{rl}{q}\right) \leq \left(\frac{rl}{q}\right)^{\frac{rl}{q}},$$

when $r < q$, one can still bound with Jensen's inequality (since $\frac{r}{q} \leq 1$):

$$\mathbb{E}\left[|f(Z) - \mathbb{E}[f(Z)]|^r\right] \leq \mathbb{E}\left[|f(Z) - \mathbb{E}[f(Z)]|^q\right]^{\frac{r}{q}} \leq C^{\frac{r}{q}} \max_{l \in [m]} l^{\frac{rl}{q}} (c\sigma_l)^r \leq C^{\frac{r}{q}} m^m (c\sigma_1)^r.$$

Since $m^m, \max(C, 1) \leq O(1)$, we can choose cleverly our constants to set the first implication of the proposition.

Let us now assume (B.1) for all 1-Lipschitz mappings $f : E \to \mathbb{R}$. Considering such a mapping f and $t > 0$, we deduce from Markov inequality and basic integration calculus that $\forall r > 0$:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \frac{\mathbb{E}\left[|f(Z) - \mathbb{E}\left[f(Z)\right]|^r\right]}{t^r} \leq C \max_{l \in [m]} \left(\frac{rl (c\sigma_l/t)^{\frac{q}{l}}}{q}\right)^{\frac{rl}{q}}. \tag{B.2}$$

Given $k, l \in [m]$, $k \leq l$ let us note:

$$t_{k,l} = c \left(\frac{\sigma_k^l}{\sigma_l^k}\right)^{\frac{1}{l-k}} \qquad\qquad t_{0,l} = 0 \qquad\qquad t_{m,m+1} = +\infty.$$

We will note for all $l \in [m]$, $t_l \equiv t_{l-1,l}$, the parameters $t_1, \ldots, t_m$ will play the same role as in the proof of Theorem 2 in Appendix A.1. Recalling that

$\forall l \in [m]$, $\sigma_l \equiv \sigma^l \mu^{(m-l)}$, the inequality $\mu_{(1)} \leq \dots \leq \mu_{(m)}$ allows us to bound:

$$\begin{aligned} t_{k,l} &= c\left(\frac{(\mu^{(m-k)})^l}{(\mu^{(m-l)})^k}\right)^{\frac{1}{l-k}} = c\left((\mu^{(m-k)})^{l-k}(\mu_{(k+1)}\cdots\mu_{(l)})^k\right)^{\frac{1}{l-k}} \\ &\leq c\mu^{(m-k)}\mu_{(l)}^k \leq c\mu^{(m-l+1)}\mu_{(l)}^{l-1} = c\left(\frac{(\mu^{(m-l+1)})^l}{(\mu^{(m-l)})^{l-1}}\right)^{\frac{1}{l-k}} \ = \ t_{l-1,l} = t_l \end{aligned}$$

(recall that $k \leq l$). The same way, we show that $\forall h \in [m]$ such that $h \geq l$, $t_{l+1} \leq t_{l,h}$ we know in particular that $0 = t_0 \leq t_1 \leq \dots \leq t_{m+1} = \infty$. Given $l \in [m]$ and $t \in [t_l, t_{l+1}]$ if we chose $r = \frac{q}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}$, then, for all $k \in [m]$, we want to bound with a $\mathcal{E}_l$ decay the quantity:

$$c_k(t) \equiv C\left(\frac{rk}{q}(c\sigma_k/t)^{\frac{q}{k}}\right)^{\frac{rk}{q}} = C\left(\frac{k}{me}\left(\frac{c^{l-k}\sigma_k^l}{t^{l-k}\sigma_l^k}\right)^{\frac{q}{lk}}\right)^{\frac{k}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}}$$

to be able to bound the concentration inequality (B.2).

If $k = l$, we have directly $c_k(t) = c_l(t) = Ce^{-\frac{l}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}}$. If $k \leq l-1$, then $t_{k,l} \leq t_l \leq t$ which implies $1/t^{l-k} \leq 1/t_{k,l}^{l-k}$ and:

$$c_k(t) \leq C\left(\frac{k}{me}\left(\frac{c^{l-k}\sigma_k^l}{t_{k,l}^{l-k}\sigma_l^k}\right)^{\frac{q}{lk}}\right)^{\frac{k}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}} = C\left(\frac{k}{me}\right)^{\frac{k}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}} \leq Ce^{-\frac{k}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}}.$$

And the same way, when $k \geq l+1$, $t \leq t_{l+1} \leq t_{l,k}$, thus $t^{k-l} \leq t_{l,k}^{k-l}$ and we can conclude that $c_k(t) \leq Ce^{\frac{k}{me}(\frac{t}{c\sigma_l})^{\frac{q}{l}}}$. When $t \in (0, t_1]$, choosing $r = \frac{q}{me}(\frac{t}{c\sigma_1})^q$, we show the same way that $\forall k \in [m]$, $c_k(t) \leq Ce^{-\frac{k}{me}(\frac{t}{c\sigma_1})^q}$. We eventually obtain for all $t \in \cup_{0 \leq l \leq m}(t_l, t_{l+1}] \supset \mathbb{R}_*^+$:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \max_{l \in [m]} Ce^{-(\frac{t}{c'\sigma_l})^{\frac{q}{l}}},$$

with $c' = (me)^{\frac{m}{q}}c \leq O(1)$. This is the looked for concentration.

Remark 14. *Proposition 13 is generally employed to bound the first centered moments of an observation. In this case,* $\left(\frac{rl}{q}\right)^{\frac{rl}{q}} \leq O(1)$*, and when* $\sigma \leq O(\mu_{(1)})$ *(which is generally the case), there exists a constant* $C > 0$ *such that we can bound for any constant* $r > 0$ *(*$r \leq O(1)$*):*

$$\mathbb{E}\left[|f(Z) - \mathbb{E}\left[f(Z)\right]|^r\right] \leq C(\sigma\mu^{(m-1)})^r,$$

since $\mu^{(m-1)} \geq \mu^{(m-l)}$*,* $\forall l \in [m]$*. We then see that the first exponential regime* $\mathcal{E}_q(\sigma\mu^{(m-1)})$ *controls the first statistics of the observations and we then say that the observable diameter of* Z *is of order* $O(\sigma\mu^{(m-1)})$*.*

Appendix C. Concentration of high order products

We give here some offshots of Theorem 2 when m is a quasiasymptotic variable (thus a sequence of positive real values). The result of Theorem 2 stays almost unmodified if m tends to infinity, one mainly needs a supplementary hypothesis.

Offshot 2. *Under the hypotheses of Theorem 2 but without the hypothesis that m is constant, if one further assumes that $(\log(m))^{\frac{1}{q}} \leq O(\frac{\mu_{(1)}}{\sigma})$, then there exists a constant $\kappa > 0$ such that we have the concentration:*

$$\phi(Z) \propto \max_{l \in [m]} \mathcal{E}_{q/l}\left(\kappa^m \sigma^l \mu^{(m-l)}\right).$$

Proof. One must be careful that this time m can tend to infinity. Given $i \in \{0\} \cup [m]$, we will employ as in Appendix A.1 the variables $t_i = \mu^{(m-i)} \mu_{(i)}^i$ for $i \in [m]$, $t_0 = 0$ and $t_{m+1} = \infty$, but we note this time for all $i \in \{0\} \cup [m]$:

$$\mathcal{A}_i \equiv \Bigg\{ \|Z_1\|_1', \|Z_1'\|_1' \leq (\delta+1)\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}}; \ldots; \|Z_i\|_i', \|Z_i'\|_i' \leq (\delta+1)\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{1}{i}};$$
$$\|Z_{i+1}\|_{i+1}', \|Z_{i+1}'\|_{i+1}' \leq (\delta+1)\mu_{i+1}; \ldots; \|Z_m\|_m', \|Z_m'\|_m' \leq (\delta+1)\mu_m \Bigg\}$$

for some $\delta \geq 0$. As we saw in the proof of Theorem 2, for all $1 \leq k \leq i < j \leq m$:

$$\forall t \leq t_{i+1} : \left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}} \leq \frac{\mu_j}{c\sigma} \qquad \text{and} \qquad \forall t \geq t_i : \left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{1}{i}} \geq \frac{\mu_k}{c\sigma}.$$

Therefore, given $t \in [t_i, t_{i+1}]$, one obtains $\forall j \in \{i+1, \ldots, m\}$:

$$\begin{aligned}\mathbb{P}\left(\|Z_j\|_j' \geq (\delta+1)\mu_j\right) &\leq \mathbb{P}\left(\left|\|Z_j\|_j' - \mathbb{E}[\|Z_j\|_j']\right| \geq \delta\mu_j\right) \\ &\leq C\exp\left(-\delta^q\left(\frac{\mu_j}{c\sigma}\right)^q\right) \leq C\exp\left(-\delta^q\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right),\end{aligned}$$

and the same way $\forall t \in [t_i, t_{i+1}]$ and $k \in [i]$:

$$\mathbb{P}\left(\|Z_k\|_k' \geq (\delta+1)\left(\frac{t}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right) \leq C\exp\left(-\delta^q\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right).$$

By hypothesis, there exists a constant $K > 0$ such that $\forall i \in [m]$ and $t \in [t_i, t_{i+1}]$:

$$K\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}} \geq K\left(\frac{\mu_i}{c\sigma}\right)^q \geq \log(m).$$

Therefore choosing $\delta = (K+1)^{\frac{1}{q}}$:

$$\mathbb{P}(\mathcal{A}_i^c) \leq C\exp\left(\log m - (K+1)\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right) \leq C\exp\left(-\left(\frac{t/(c\sigma)^i}{\mu^{(m-i)}}\right)^{\frac{q}{i}}\right),$$

and $\forall t \in (0, t_1]$:

$$\mathbb{P}\left(\mathcal{A}_0^c\right) \leq C \exp\left(\log m - (K+1)\left(\frac{\mu_1}{c\sigma}\right)^q\right) \leq C \exp\left(\left(\frac{\mu_1}{c\sigma}\right)^q\right) \leq C \exp\left(-\left(\frac{t/c\sigma}{\mu^{(m-1)}}\right)^q\right).$$

Besides, for any $i \in \{0\} \cup [m]$ and $t \in [t_i, t_{i+1}]$, we can bound in the asymptotic m case, as in Appendix A.1, the concentration inequality:

$$\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t, Z \in \mathcal{A}_i\right) \leq C \exp\left(-\frac{t}{(m(1+\delta)^m c\sigma)^i \mu^{(m-i)}}\right)^{\frac{q}{i}}.$$

Regrouping all the concentration inequalities for the different values of t and noting $\kappa \equiv (c+1)(\delta+1)e^{1/e} \leq O(1)$, we retrieve our result (note that $e^{1/e} \geq m^{1/m}$ for all $m > 0$).

When only one concentrated random vector is involved, one can obtain a better concentration inequality characterized by only two exponential regime (as explained in Remark 7).

Offshot 3. *With the hypotheses of Theorem 2 in a setting where m is quasi-asymptotic, if $Z_1 = \cdots = Z_m$, we do not have to assume anymore that $(\log(m))^{\frac{1}{q}} \leq O(\mathbb{E}[\|Z_i\|_i']/\sigma)$ and if $\mu_1 = \cdots = \mu_m \equiv \mu_0$, then, for any constant $\varepsilon > 0$ (i.e. such that $\varepsilon \geq O(1)$) we have the concentration:*

$$\phi(Z) \propto \mathcal{E}_q\left(m\sigma((1+\varepsilon)\mu_0)^{m-1}\right) + \mathcal{E}_{q/m}\left((\kappa\sigma)^m\right),$$

for some constant[34] *$\kappa > 0$.*

Note that when $1 - \mu_0 \geq O(1)$, one can choose ε sufficiently small such that $m((1+\varepsilon)\mu_0)^{m-1} \leq O(1)$ (when $m \to \infty$), then $\phi(Z) \propto \mathcal{E}_q(\sigma) + \mathcal{E}_{q/l}((\kappa\sigma)^m)$.

Proof. Let us note $K \equiv \max((1+\varepsilon)\mu, (t/m)^{1/m})$ and $\mathcal{A}_K \equiv \{\|Z\|' \geq K\}$, then, on the one hand $K - \mu \geq \frac{\varepsilon}{1+\varepsilon}$ and:

$$\mathbb{P}(\mathcal{A}_K^c) \leq \mathbb{P}\left(|\|Z\|' - \mathbb{E}[\|Z\|']| \geq K - \mu\right) \leq Ce^{-\left(\frac{\frac{\varepsilon}{1+\varepsilon}K}{c\sigma}\right)^q} \leq Ce^{-\left(\frac{t}{m(\frac{1+\varepsilon}{\varepsilon}c\sigma)^m}\right)^{\frac{q}{m}}},$$

and on the second hand ϕ is mK^{m-1}-Lipschitz on $\mathcal{A}_K$ thus:

$$\begin{aligned}\mathbb{P}\left(|f(\phi(Z)) - f(\phi(Z'))| \geq t \mid \mathcal{A}_K\right) &\leq Ce^{-(t/mc\sigma K^{m-1})^q} \\ &\leq Ce^{-(t/mc\sigma((1+\varepsilon)\mu)^{m-1})^q} + Ce^{-(t/m(c\sigma)^m)^{q/m}}.\end{aligned}$$

Therefore, we obtain the concentration:

$$\phi(Z) \propto \mathcal{E}_q\left(m\sigma((1+\varepsilon)\mu)^{m-1}\right) + \mathcal{E}_{q/m}\left(\left(ce^{1/e}\frac{1+\varepsilon}{\varepsilon}\sigma\right)^m\right), \tag{C.1}$$

which provides us the wanted inequality since $ce^{1/e}\frac{1+\varepsilon}{\varepsilon} \leq O(1)$.

[34] The presence of this constants basically imposes that if $\sigma \geq O(1)$, then the observable diameter of the $\mathcal{E}_{q/m}((\kappa\sigma)^m)$ decay can not tend to zero. A better inequality might be obtained, if one computes precisely the concentration constants, starting from a sharp concentration inequality on Z.

Appendix D. Proofs of resolvent concentration properties

Appendix D.1. Lipschitz concentration of Q

Lemma 6. *Under the assumptions of Theorem 3, $\|Q\| \leq \frac{1}{\varepsilon} \leq O(1)$.*

Then we can show a Lipschitz concentration of Q but with looser observable diameter that the one given by Theorem 3 (as for XDY^T, we get better concentration speed in the linear concentration framework).

Lemma 7. *Under the hypotheses of Theorem 3:*

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

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

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

thanks to the hypotheses and Lemma 6 below. The same way, we can bound:

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

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

Appendix D.2. Control on $\|Qx_i\|$ and $\|Q^Ty_i\|$

The dependence between Q and (x_i, y_i) prevent us from bounding straightforwardly $\|Qx_i\|$ and $\|Q^Ty_i\|$ with Lemma 6 and the hypotheses on x_i, y_i. We can still disentangle this dependence thanks to the notations:

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

We can indeed bound:

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

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

Lemma 8. *Under the assumptions of Theorem 3:*

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

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

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

Therefore, the concentrations $x_i \propto \mathcal{E}_2$ and $Q_{-i}^{(i)} \propto \mathcal{E}_2$ given in Remark 7 imply that there exist two constants $C, c > 0$ such that $\forall t > 0$ such that if we note $\mathcal{A}_{-i}$, the sigma algebra generated by X_{-i} and Y_{-i} (it is independent with x_i):

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

for some constants $C', c' > 0$, thanks to the bounds $\|\mathbb{E}[x_i]\| \leq O(1)$ given in the assumptions and $\|Q_{-i}^{(i)}\| \leq O(1)$ given by Lemma 6.

The linear concentration of $Y_{-i}^T Q_{-i}^{(i)} x_i / \sqrt{n}$ is proven the same way since one can show as in Remark 7 that $(X, D, Y) \mapsto Y_{-i}^T Q_{-i}^{(i)} / \sqrt{n}$ is $O(1)$-Lipschitz on $\{\|X\|, \|Y\| \leq \kappa\sqrt{n}, \|D\| \leq \kappa_D\}$, and therefore, $Y_{-i}^T Q_{-i}^{(i)} / \sqrt{n} \propto \mathcal{E}_2$.

The link between Qx_i and $Q_{-i}x_i$ is made possible thanks to classical Schur identities:

$$Q = Q_{-i} - \frac{1}{n} \frac{D_i Q_{-i} x_i y_i^T Q_{-i}}{1 + \frac{1}{n} D_i x_i^T Q_{-i} y_i} \qquad \text{and} \qquad Q x_i = \frac{Q_{-i} x_i}{1 + \frac{1}{n} D_i y_i^T Q_{-i} x_i}, \tag{D.2}$$

and the link between $Q_{-i}x_i$ and $Q_{-i}^{(i)} x_i$ is made thanks to

Lemma 9. *Under the hypotheses of Theorem 3, for all $i \in [n]$:*

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

Proof. Let us bound directly:

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

We can then conclude thanks to Proposition 6 combined with the concentration provided Lemma 8 and the bound:

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

We can then bound $\|Qx_i\|$ and $\|Q^T y_i\|$ combining Lemmas 8 and 9 with (D.1).

Lemma 10. *Under the hypotheses of Theorem 3 we can bound:*

$$\|\mathbb{E}[Qx_i]\| \leq O(\sqrt{\log n}) \qquad \textit{and} \qquad \|\mathbb{E}[Qy_i]\| \leq O(\sqrt{\log n}).$$

Proof. The Schur identities (D.2) allow us to write:

$$\mathbb{E}[Qx_i] = \mathbb{E}\left[\delta_i Q_{-i} x_i\right] \qquad \text{where } \ \delta_i \equiv 1 + \frac{1}{n} D_i y_i^T Q_{-i} x_i.$$

For any deterministic $u \in \mathbb{R}^p$ such that $\|u\| \leq 1$, we can bound:

$$\begin{aligned}\left|u^T \mathbb{E}[Qx_i]\right| &\leq \left|\mathbb{E}\left[\delta_i u^T Q_{-i}^{(i)} x_i\right] + \mathbb{E}\left[\delta_i u^T \left(Q_{-i} - Q_{-i}^{(i)}\right) x_i\right]\right| \\ &\leq O\left(\mathbb{E}\left[\left|u^T Q_{-i}^{(i)} x_i\right|\right] + \mathbb{E}\left[\left\|\left(Q_{-i} - Q_{-i}^{(i)}\right) x_i\right\|\right]\right)\end{aligned}$$

(since $|\delta_i| \leq \frac{\kappa^2 \kappa_D}{\varepsilon}$). We can then conclude that $\|\mathbb{E}[Qx_i]\| \leq \sup_{\|u\| \leq 1} |u^T \mathbb{E}[Qx_i]| \leq O(\sqrt{\log n})$ thanks to Lemma 9 and the concentration $u^T Q_{-i}^{(i)} x_i \in O(1) \pm \mathcal{E}_2$. The same holds for $\|\mathbb{E}[Qy_i]\|$.

Appendix D.3. Concentration of $y_i^T QAQx_i$

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

Lemma 11. *Under the hypotheses of Theorem 3, $\forall i \in [n]$, and for any deterministic matrices $U, V \in \mathcal{M}_p$ such that $\|U\|, \|V\| \leq 1$:*

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

Be careful that the bound would not have been so tight for $\|QXU\|_\infty$ and $\|QYV\|_\infty$ given $U, V \in \mathcal{M}_n$.

Proof. Following the same identities and arguments presented in the proof of Lemma 9, we can bound (since $|\delta_i| \leq \frac{\kappa^2 \kappa_D}{\varepsilon}$)

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

Introducing, as in Section 3, $(e_1, \ldots, e_p)$ and $(f_1, \ldots, f_n)$, respectively, the canonical basis of $\mathbb{R}^p$ and $\mathbb{R}^n$ we know from Lemma 8 that for all $k \in [p]$ and $i, j \in [n]$:

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

since $|\mathbb{E}[e_k^T VQ_{-i}^{(i)} x_i]| \leq \|\mathbb{E}[Q_{-i}^{(i)}]\| \|\mathbb{E}[x_i]\| \leq O(1)$ and similarly, $|\mathbb{E}[f_j^T Y_{-i}^T Q_{-i}^{(i)} x_i]/\sqrt{n}| \leq O(1)$. Following the arguments displayed in Section 3, there exist four constants K, C, c, c' (all $\leq O(1)$) such that we can bound:

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

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

We then have this last Lemma, quite close to Corollary 2:

Lemma 12. *Under the hypotheses of Theorem 3, given a deterministic matrix $A \in \mathcal{M}_{p,n}$ such that $\|A\|_F \leq 1$:*

$$\frac{1}{n} \|Y^T QAQX\|_d \propto \mathcal{E}_1 \left(\sqrt{\frac{\log n}{n}} \right).$$

If we assume in addition that $\|A\|_* \leq 1$ or $\|\mathbb{E}[x_i y_i^T]\|_F \leq O(1)$, then $\mathbb{E}[\frac{1}{n}\|Y^T QAQX\|_d] \leq O(\log n/\sqrt{n})$.

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

Proof. To prove the concentration, let us introduce again the decomposition $A = U^T \Lambda V$, with $U, V \in \mathcal{O}_p$ and $\Lambda \in \mathcal{D}_p$. We are going to bound the variation of $\frac{1}{n}\|Y^T QAQX\|_d$ towards the variables $(\frac{1}{\sqrt{n}} VQX, \frac{1}{\sqrt{n}} UQY) \propto \mathcal{E}_2$ (see Lemma 7). Let us define the mapping $\phi : \mathcal{M}_{p,n}^2 \to \mathbb{R}$ satisfying for all $M, P \in \mathcal{M}_{p,n}$, $\phi(M, P) = \|M^T \Lambda P\|_d$ (with that definition, $\frac{1}{n}\|Y^T QAQX\|_d = \phi(\frac{1}{\sqrt{n}} VQX, \frac{1}{\sqrt{n}} UQY)$). Given 4 variables M, P, M', P' satisfying $\|M\|, \|P\|, \|M'\|, \|P'\| \leq \frac{\kappa}{\varepsilon}$ we can bound as in the proof of Corollary 2:

$$|\phi(M, P) - \phi(M', P)| \leq \|(M - M')^T \Lambda P\|_d \leq \frac{1}{\sqrt{n}\varepsilon^2} \|M - M'\|_F \|P\|_\infty,$$

and the same way, $|\phi(M, P) - \phi(M, P')| \leq \frac{1}{\sqrt{n}\varepsilon^2} \|P - P'\|_F \|M\|_\infty$. Here, we invoke Lemma 11, to employ the concentrations:

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

We can then deduce from Offshot 1 the concentration $\sqrt{\frac{n}{\log n}} \phi(\frac{1}{\sqrt{n}} VQX, \frac{1}{\sqrt{n}} UQY) \propto \mathcal{E}_2 \pm \mathcal{E}_1 \propto \mathcal{E}_1$, from which we deduce the concentration of $\|Y^T QAQX\|_d$.

To bound the expectation of $\|Y^T QAQX\|_d$, recall from the proof of Corollary 2, that we just need to show that for all $i \in [n]$:

$$y_i^T QAQx_i \in O(\log n) \pm \mathcal{E}_1(\log n). \tag{D.3}$$

Since the projection on any of the diagonal elements of a matrix is a 1-Lipschitz mapping for the semi-norm $\|\cdot\|_d$, we already know that $y_i^T QAQx_i \propto \mathcal{E}_1\left(\sqrt{\log n}\right)$, we are thus left to bound $\left|\mathbb{E}[y_i^T QAQx_i]\right|$, for all $i \in [n]$. Let us decompose with the same calculus as in the proof of Lemma 10:

$$\begin{aligned} y_i^T QAQx_i = {} & \delta_i^2 y_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i + \delta_i^2 y_i^T (Q_{-i} - Q_{-i}^{(i)}) A (Q_{-i} - Q_{-i}^{(i)}) x_i \\ & + \delta_i^2 y_i^T (Q_{-i} - Q_{-i}^{(i)}) A Q_{-i}^{(i)} x_i + \delta_i^2 y_i^T Q_{-i}^{(i)} A (Q_{-i} - Q_{-i}^{(i)}) x_i. \end{aligned}$$

We already know that (assuming only $\|A\|_F \leq 1$):

- $|\delta_i| \leq O(1)$ and $y_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i \propto \mathcal{E}_1(\sqrt{\log n})$,
- $\left|y_i^T (Q_{-i} - Q_{-i}^{(i)}) A (Q_{-i} - Q_{-i}^{(i)}) x_i\right| \leq \|(Q_{-i} - Q_{-i}^{(i)}) x_i\| \|(Q_{-i} - Q_{-i}^{(i)}) y_i\| \in O(\log n) \pm \mathcal{E}_1(\log n)$ thanks to Lemma 9 and because $\|A\| \leq \|A\|_F \leq 1$,
- we can bound: $|y_i^T (Q_{-i} - Q_{-i}^{(i)}) A Q_{-i}^{(i)} x_i| \leq \|V Q_{-i}^{(i)} x_i\|_\infty \|\Lambda\|_F \|U (Q_{-i} - Q_{-i}^{(i)}) y_i\| \in O(\log n) \pm \mathcal{E}_1(\log n)$, thanks to Lemma 9 and the concentration $\|V Q_{-i}^{(i)} x_i\|_\infty \in O(\sqrt{\log n}) + \mathcal{E}_2(\sqrt{\log n})$ given by Lemma 11.

In addition, noting $\Sigma \equiv \mathbb{E}[x_i y_i^T]$, we already know from Proposition 5 and the hypotheses on (x_i, y_i) that $\|\Sigma\| \leq 1$ and:

- if $\|A\|_* \leq 1$, $\left|\mathbb{E}[y_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i]\right| \leq \frac{\|\Sigma\|}{\varepsilon^2} \|A\|_* \leq O(1)$,
- if $\|\Sigma\|_F \leq 1$, $\left|\mathbb{E}[y_i^T Q_{-i}^{(i)} A Q_{-i}^{(i)} x_i]\right| \leq \frac{\|\Sigma\|_F}{\varepsilon^2} \|A\|_F \leq O(1)$.

Therefore, in all cases, Hölder inequalities allow us to show that $\mathbb{E}[|y_i^T QAQx_i|] \leq O(\log n)$ and (D.3) is true, we can then conclude as in the proof of Corollary 2 that $\mathbb{E}[\|Y^T QAQX\|_d] \leq O(\log n / \sqrt{n})$.

Appendix D.4. Proof of Theorem 3

Noting $\bar{Q} \equiv (I_p - \frac{1}{n} X \mathbb{E}[D] Y^T)^{-1}$, we consider a deterministic matrix $A \in \mathcal{M}_p$, such that $\|A\|_F \leq 1$ and we bound in a first step, as in the proof of Proposition 11:

$$\begin{aligned} & \left|\mathbb{E}[\mathrm{Tr}(AQ)] - \mathbb{E}[\mathrm{Tr}(A\bar{Q})]\right| \\ & \qquad \leq \frac{1}{n} \sum_{i=1}^n \left|\mathbb{E}\left[\left(y_i^T QAQxi - \mathbb{E}[y_i^T QAQx_i]\right)(D_i - \mathbb{E}[D_i])\right]\right| \leq \ O(\log n), \end{aligned}$$

thanks to Hölder's inequality applied to the concentrations $D_i \propto \mathcal{E}_2$ and $y_i^T QAQx_i \propto \mathcal{E}_1(\log n)$ thanks to Lemma 12; (with the same concentration constants for all $i \in [n]$). We can further bound:

$$\left\| \mathbb{E}[\tilde{Q}] - \mathbb{E}[\bar{Q}] \right\|_F \leq \frac{1}{n} \left\| \mathbb{E}\left[\bar{Q} X (\tilde{D} - \mathbb{E}[D]) Y^T \tilde{Q} \right] \right\|_F \leq \frac{\kappa^2}{\varepsilon^2} \left\| \tilde{D} - \mathbb{E}[D] \right\|_F \leq O(1),$$

which eventually allows us to set that $\left\| \mathbb{E}[Q] - \mathbb{E}[\tilde{Q}] \right\|_F \leq O(\log n)$.

To show the concentration of Q, we note $\phi(X, D, Y) = \mathrm{Tr}(AQ)$. We abusively work with X, D, Y and independent copies X', D', Y' satisfying $\|X\|, \|Y\|, \|X'\|, \|Y'\| \leq \sqrt{n}\kappa$ and $\|D\|, \|D'\| \leq \kappa_D$ as if they were deterministic variables, and we note $Q'_X \equiv \phi(X', D, Y)$, $Q'_D \equiv \phi(X, D', Y)$ and $Q'_Y \equiv \phi(X, D, Y')$. Let us bound the variations

$$|\phi(X, D, Y) - \phi(X', D, Y)| = \frac{1}{n} \left| \mathrm{Tr}\left(AQ(X - X') DY Q'_X \right) \right| \leq \frac{\kappa \kappa_D}{\varepsilon^2 \sqrt{n}} \|X - X'\|_F.$$

The same way, $|\phi(X, D, Y) - \phi(X, D, Y')| \leq \frac{\kappa\kappa_D}{\varepsilon^2\sqrt{n}} \|Y - Y'\|_F$, and we can also bound as in the proof of Proposition 10:

$$|\phi(X, D, Y) - \phi(X, D', Y)| \leq \frac{1}{n} \left\| Y Q'_D A Q X \right\|_d \|D - D'\|_F.$$

We can then conclude applying Offshot 1 to the variation control we provided and the concentration of $\frac{1}{n} \|YQ'_D AQX\|_d$ given by Lemma 12 (actually Lemma 12 gives the concentration of $\|YQAQX\|_d$, but the proof remains the same if one replaces one of the Q with Q'_D, for a diagonal matrix D', independent with D).

Appendix D.5. Proof of Proposition 12

With the same variables $X, Y, X', Y' \in \mathcal{M}_{p,n}$, $D, D' \in \mathcal{D}_n$ and with the same notations Q, Q'_X, Q'_Y, Q'_D as in the proof of Theorem 3, we bound:

$$\|Qu - Q'_X u\| = \frac{1}{n} \left\| Q(X - X') D Y^T Q'_X u \right\| \leq \frac{\kappa \kappa_D}{\varepsilon^2 \sqrt{n}} \|X - X'\|,$$

and the same way, $\|Qu - Q'_Y u\| \leq \frac{\kappa\kappa_D}{\varepsilon^2\sqrt{n}} \|Y - Y'\|$. Second:

$$\|Qu - Q'_D u\| = \frac{1}{n} \left\| Q'_D X (D - D') Y^T Q u \right\| \leq \frac{\kappa}{\varepsilon \sqrt{n}} \|D - D'\|_F \, \|Y^T Q u\|_\infty,$$

and we know from Lemma 11 that $\|Y^T Qu\|_\infty \in O(\sqrt{\log n}) + \mathcal{E}_2(\sqrt{\log n})$, which allows us to conclude with Offshot 1 that:

$$\sqrt{\frac{n}{\log n}} Qu \propto \mathcal{E}_2 + \mathcal{E}_1,$$

but the $\mathcal{E}_2$ decay can here be removed since the $\mathcal{E}_1$ is looser.

References

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

Adamczak, R., 2015. A note on the hanson-wright inequality for random vectors with dependencies. Electronic Communications in Probability 20, 1–13. doi:10.1214/ECP.v20-3829.

Boucheron, S., Lugosi, G., Massart, P., 2013. Concentration Inequalities: a Nonasymptotic Theory of Independence. Oxford University Press.

Chatterjee, S., 2017. Fluctuations of eigenvalues and second order poincaré inequalities. Probability Theory and Related Fields 143, 1–40.

Deng, Z., Kammoun, A., Thrampoulidis, C., 2020. A model of double descent forhigh-dimensional binary linear classification. ICASSP'2020 IEEE International Conference on Acoustics, Speech and Signal Processing .

El Karoui, N., Bean, D., Bickel, P.J., Lim, C., Yu, B., 2013. On robust regression with high-dimensional predictors. Proceedings of the National Academy of Sciences 110, 14557–14562.

Fleury, B., Guédon, O., Paouris, G., 2007. A stability result for mean width of l p -centroid bodies. Advances in Mathematics 214, 865–877.

Gromov, M., Milman, V., 1983. . a topological application of the isoperimetric inequality. American Journal of Mathematics 105, 843–854.

Guédon, O., Lytova, A., Pajor, A., Pastur, L., 2014. The central limit theorem for linear eigenvalue statistics of the sum of independent matrices of rank one. American Mathematical Society Translations: Series 2 233, 145–164.

Hanson, D.L., Wright, F.T., 1971. A bound on tail probabilities for quadratic forms in independent random variables. Annals of Mathematical Statistics 42, 1079–1083.

Huang, H., 2017. Asymptotic behavior of support vector machine for spiked population model. Journal of Machine Learning Research 18, 1–21.

James, R.C., 1957. Reflexivity and the supremum of linear functionals. Annals of Mathematics 66, 159 – 169.

Klartag, B., 2007. A central limit theorem for convex sets. Inventiones mathematicae 168, 91–131.

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., 2021. Concentration of the product of random vectors and of the resolvent in convex concentration settings. unpublished .

Lévy, P., 1951. Problemes concrets d'analyse fonctionnelle. Gauthier-Villars.

Mai, X., Liao, Z., Couillet, R., 2019. A large scale analysis of logistic regression: Asymptotic performance and new insights. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing .

Pajor, A., Pastur, L., 2009. On the limiting empirical measure of the sum of rank one matrices with log-concave distribution. Studia Mathematica, Institut Matematyczny Polska Akademia nauk 195, 11–29.

Pastur, L., 2005. A simple approach to the global regime of gaussian ensembles of random matrices. Ukrainian Mathematical Journal 57, 936–966.

Seddik, M.E.A., Louart, C., Tamaazousti, M., Couillet, R., 2021. The unexpected deterministic and universal behavior of large softmax classifiers. AISTATS .

Seddik, M.E.A., Tamaazousti, M., Couillet, R., 2019. Kernel random matrices of large concentrated data : the example of gan-generated images. ICASSP'19 .

Silverstein, J.W., Bai, Z.D., 1995. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. Journal of Multivariate Analysis 54, 175–192.

Talagrand, M., 1988. An isoperimetric theorem on the cube and the kintchine-kahane inequalities. Proceedings of the American Mathematical Society 81, 73—-205.

Talagrand, M., 1995. Concentration of measure and isoperimetric inequalities in product spaces. Publications mathématiques de l'IHÉS 104, 905–909.

Vershynin, R., 2017. High dimensional probability. Cambridge University Press.

Vu, V., Wang, K., 2014. Random weighted projections, random quadratic forms and random eigen vectors. Random Structures and Algorithms doi:10.1002/rsa.20561.