Concentration of Measure and Large Random Matrices with an application to Sample Covariance Matrices

Cosme Louart, Romain Couillet

Concentration of Measure and Large Random Matrices with an application to Sample Covariance Matrices *

Cosme Louart[‡†] **and Romain Couillet**[§†]

Louart Cosme : 45 Rue d'Ulm 75005 Paris.
e-mail: cosme.louart@ens.fr

Romain Couillet : Bureau D1178, 11 rue des mathématiques
Domaine Universitaire BP 46, 38402 Saint Martin d'Hères cedex.
e-mail: romain.couillet@gipsa-lab.grenoble-inp.fr
url: http://romaincouillet.hebfree.org

Abstract: The present work provides an original framework for random matrix analysis based on revisiting the concentration of measure theory for random vectors. By providing various notions of vector concentration (q-exponential, linear, Lipschitz, convex), a set of elementary tools is laid out that allows for the immediate extension of classical results from random matrix theory involving random concentrated vectors in place of vectors with independent entries. These findings are exemplified here in the context of sample covariance matrices but find a large range of applications in statistical learning and beyond, starting with the capacity to easily analyze the performance of artificial neural networks and random feature maps.

Keywords and phrases: Concentration of the Measure, Talagrand Theorem, Davis Theorem, Stieltjes transform, Deterministic equivalent, Spectral distribution, Central limit Theorem.

Contents

*This work is supported by the chair IDEX G-Stats DataScience at Univ. Grenoble Alpes.
†Univ. de Paris-Saclay, CentraleSupélec, L2S, 91190 Gif-sur-Yvette, France
‡Ecole Normale Supérieure, 75005 Paris, France
§Univ. Grenoble Alpes, CNRS, Grenoble Institute of engenering, GIPSA-lab, 38000 Grenoble, France

Nomenclature

iif "if and only if"

$\mathbb{R}$ Set of real numbers ; $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$; $\mathbb{R}_- = -\mathbb{R}_+$. If $x \in \mathbb{R}^p$, one notes $[x]_i$, or more simply x_i, $1 \leq i \leq p$, the i^{th} entry of the vector x

$(a, b]$ Considering an interval $I \subset \mathbb{R}$, we employ the sign "(" if left border of I is open and "[" if it is closed ; the same rule works for the right border. To give an example, given $a, b \in \mathbb{R} : (a, b] = \{x \mid a < x \leq b\}$

$\lfloor x \rfloor$ Integer part of $x \in \mathbb{R}$, $\lfloor x \rfloor \in \mathbb{N}$ and verifies $\lfloor x \rfloor \leq x \leq \lfloor x \rfloor + 1 = \lceil x \rceil$.

E Typical normed vector space over $\mathbb{R}$, endowed with the norm $\| \cdot \|$.

$\mathbb{C}$ Set of complex numbers.

E^* Dual space of E (the set of linear maps from E to $\mathbb{R}$).

$\mathcal{A}$ Typical non commutative algebra endowed with the algebra norm $\| \cdot \|$, the possibly non commutative product is written without operation character, we have $\forall x, y \in \mathcal{A}$, $\|xy\| \leq \|x\|\|y\|$.

∂A Boundary of $A \subset E$, if $\bar{A}$ is the closure of A and $\mathring{A}$, the interior of A, $\partial A = \bar{A} \setminus \mathring{A}$.

$\text{Conv}(A)$ Convex hull of $A \subset E$ (i.e., $\text{Conv}(A) = \cap\{C \subset E,\, C \text{ convex}, C \supset A\}$).

A_* $A \setminus \{0\}$

$\mathbb{1}_A$ Indicator function of $A \subset E$, $\mathbb{1}_A : E \to \{0, 1\}$ and $\mathbb{1}_A(x) = 1 \Leftrightarrow x \in A$. If A is not a set but an assertion (like $A = A(t) = (t \geq 1)$), $\mathbb{1}_A = 1$ if A is true and $\mathbb{1}_A = 0$ if A is false. If $\mathbb{1}$ presents no index, it designates a vector full of one with a convenient size for the context.

$\mathfrak{S}_p$ Set of permutations of $\{1,\dots,p\}$; $\mathfrak{S}_{p,n} = \mathfrak{S}_p \times \mathfrak{S}_n$.

$x^{\downarrow}$ Decreasing version of $x \in \mathbb{R}^p$, $\exists \sigma \in \mathfrak{S}_p$ such that $\forall i \in \{1,\dots,p\}$, $[x^{\downarrow}]_i = x_{\sigma(i)}$ and $[x^{\downarrow}]_i \leq [x^{\downarrow}]_{i-1}$ for $i \geq 2$.

$\prec$ Majorization relation, see Definition 15

$\mathcal{M}_{p,n}$ Set of real matrices of size $p \times n$. If $M \in \mathcal{M}_{p,n}$, one notes $[M]_{i,i}$ or more simply $M_{i,j}$ the entry at the line i and column j. If $n = p$, we simply note $\mathcal{M}_p = \mathcal{M}_{p,n}$.

Tr Trace operator on $\mathcal{M}_p$, $\forall M \in \mathcal{M}_p, \operatorname{Tr} M = \sum_{i=1}^p M_{i,i}$.

$.^T$ Transpose operator on $\mathcal{M}_p$, $\forall M \in \mathcal{M}_{p,n}$, $[M^T]_{i,j} = M_{j,i}$, $1 \leq i \leq p$, $1 \leq j \leq n$.

I_p Identity matrix of $\mathcal{M}_p$, ($[I_p]_i, j = 0$ if $i \neq j$ and $[I_p]_{i,i} = 1$, $1 \leq i, j \leq j$).

$\operatorname{Sp}(M)$ Spectrum of the matrix M.

Q_C Resolvent of the matrix $C \in \mathcal{M}_p$. For $z \in \mathbb{C} \setminus \operatorname{Sp}(C)$, $Q_C(z) = (C + zI_p)^{-1}$.

Diag Diagonal operator. If $M \in \mathcal{M}_{p,n}$, $\operatorname{Diag}(M) = (M_{i,i})_{1 \leq i \leq \min(p,n)}$; if $x \in \mathbb{R}^p$, $\operatorname{Diag}_{q,n}(x) \in \mathcal{M}_{p,n}$, $[\operatorname{Diag}_{p,n}(x)]_{i,i} = x_i$ if $1 \leq i \leq \min(p,q,n)$ and $[\operatorname{Diag}_{p,n}(x)]_{i,j} = 0$ if $i \neq j$, for $1 \leq i \leq p$, $1 \leq j \leq n$.

$\mathcal{O}_p$ Set of orthogonal matrices of $\mathcal{M}_p$: $P \in \mathcal{D}_{p,n} \Leftrightarrow P^{-1} = P^T$; $\mathcal{O}_{p,n} = \mathcal{O}_p \times \mathcal{O}_n$.

$\mathcal{D}_{p,n}$ Set of diagonal matrices of $\mathcal{M}_{p,n}$: $D \in \mathcal{D}_{p,n} \Leftrightarrow D = \operatorname{Diag}_{p,n}(\operatorname{Diag}(D))$; $D \in \mathcal{D}^+_{p,n} \Leftrightarrow D_{i,i} \geq 0, 1 \leq i \leq \min(p,n)$; $D \in \mathcal{D}^-_{p,n} \Leftrightarrow -D \in \mathcal{D}^+_{p,n}$. When $n = p$, we simply note $\mathcal{D}_p = \mathcal{D}_{p,n}$.

$\mathcal{S}_p$ Set of symmetric matrices of $\mathcal{M}_p$: $S \in \mathcal{S}_n \Leftrightarrow S_{i,j} = S_{i,j}, 1 \leq i \leq p$; $S \in \mathcal{S}^+_p \Leftrightarrow \forall u \in \mathbb{R}^p, u^T S u \geq 0$; $S \in \mathcal{S}^-_p \Leftrightarrow -S \in \mathcal{S}^+_p$. Given $S_1, S_2 \in \mathcal{S}_p$, we say that S_1 is greater than S_2 and we note $S_1 \geq S_2$ if $S_1 - S_2 \in \mathcal{S}^+_p$.

$S^{1/2}$ Square root of the nonnegative symmetric matrix $S \in \mathcal{S}^+_p$ (with the diagonalization $S = P^T \Lambda P$, $P \in \mathcal{O}_p$, $\Lambda \in \mathcal{D}_p$, we define $S^{1/2} = P^T \Lambda^{1/2} P$ where $[\Lambda^{1/2}]_{i,i} = \Lambda^{1/2}_{i,i}$).

$\mathcal{P}_p$ Set of permutation matrices of $\mathcal{M}_p$: $P \in \mathcal{P}_p \Leftrightarrow P \in \mathcal{O}_n$ and ($\exists \sigma \in \mathfrak{S}_p$, $P_{i,j} = 1 \Leftrightarrow \sigma(i) = j$) ; we also define $\mathcal{P}_{p,n} = \{(U,V) \in \mathcal{P}_p \times \mathcal{P}_n \mid U I_{p,n} V^T = I_{p,n}\}$ where $I_{p,n} = \operatorname{Diag}_{p,n}(\mathbb{1})$.

$\|\cdot\|_q$ ℓ_q-norm on $\mathbb{R}^p$ for two integers $p, q \in \mathbb{N}_*$; $\|x\|_q = (\sum_{i=1}^p x_i^q)^{1/q}$.

$\|\cdot\|$ Classical norm of the vector space E one is working on : if $E = \mathbb{R}^p$, the euclidean norm $\|\cdot\|_2$; if $E = M_{p,n}$, the spectral norm ($\forall M \in \mathcal{M}_{p,n}$: $\sup_{\|u\|=1} \|Mu\|$).

$\|\cdot\|_1$ on $\mathbb{R}^p$, the ℓ_1 ; on $M_{p,n}$, the nuclear norm ($\forall M \in \mathcal{M}_{p,n}, \|M\|_1 = \operatorname{Tr}((MM^T)^{1/2})$).

$\|\cdot\|_F$ Frobenius norm, $\forall M \in \mathcal{M}_{p,n} : \|M\|_F = \sqrt{\operatorname{Tr} MM^T} = \sqrt{\sum_{i=1}^p \sum_{j=1}^n M_{i,j}^2}$.

$\|\cdot\|_*$ Dual norm on E^*. If $f \in E^*$, $\|f\|_* = \sup_{\|x\| \leq 1} f(x)$.

$d_{\|\cdot\|}$ Distance associated to the norm $\|\cdot\|$. Given two vectors $x, y \in E$ and two sets $A, B \subset E$, $d_{\|\cdot\|}(x,y) = \|x - y\|$, $d_{\|\cdot\|}(x, A) = \inf\{d_{\|\cdot\|}(x,y), y \in A\}$ and $d_{\|\cdot\|}(A,B) = \inf\{d_{\|\cdot\|}(x,B), x \in A\}$.

$\mathcal{B}_t$ Closed ball of E of size $t > 0$, $\mathcal{B}_t = \{0\}_t = \{x \in E \mid \|x\| \leq 1\}$, when $t = 1$, we note $\mathcal{B} = \mathcal{B}_1$. We also use the notation $\mathcal{B}_{\|\cdot\|}(x,t) = \{x\}_t =$

$\{y \in E \parallel \|x-y\| \leq t\}$, the index $\|\cdot\|$ could be of course a distance d or simply unspecified when we implicitly consider the classical norm of E.

A_t If $A \subset E$, A_t is the closed set $\{x \in E \mid d(x,A) \leq t\} \supset A$.

S_f^t Level set of $f : E \to \mathbb{R}$. For $t \in \mathbb{R}$, $S_f^t = \{x \in E \mid f(x) \leq t\}$.

$\mathbb{S}^p$ Sphere of $\mathbb{R}^{p+1}$ ($\mathbb{S}^p = \{x \in \mathbb{R}^{p+1} \mid \|x\| \leq 1\}$).

$\eta_{(E,\|\cdot\|)}$ Norm degree, see Definition 9.

$\mathbb{P}$ We implicitly suppose all over the paper that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ is a sigma-algebra of the set Ω and $\mathbb{P}$, a probability measure defined on the elements of $\mathcal{F}$. The random vectors we consider are then $\mathbb{P}$-measurable applications defined on Ω and taking value in normed vector spaces endowed with the Borel σ-algebra. In that setting, given a random vector $X \in E$ (i.e. $X : \Omega \to E$), for any Borel set $A \subset E$, we note $\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega,\ X(w) \in A\})$ and for any measurable function $f : E \mapsto \mathbb{R}$ and $t \in \mathbb{R}$, we note $\mathbb{P}(f(X) \geq t) = \mathbb{P}(\{\omega \in \Omega,\ f(X(\omega)) \geq t\})$...

$\mathbb{E}$ The expectation operator. For any random vector $X \in E$ and any measurable function $f : E \to \mathbb{R}$, we define $\mathbb{E}[f(X)] = \int_\Omega f \circ X d\mathbb{P}$. When E has finite dimension, it is possible to define $\mathbb{E}X$ by integrating all the coordinates of X.

$\sigma(X)$ If $X \in \mathcal{M}_{p,n}$, $\sigma(X) \in \mathbb{R}_+^{\min(p,n)}$ is the vector constituted of the singular values of X in increasing order (i.e. the eigenvalues of $(XX^T)^{1/2}$). If $X \in E$ is a random vector, $\sigma(X)$ is the σ-algebra generated by X (i.e., the σ-algebra of sets of Ω containing all the sets $X^{-1}(B)$ when B is a Borel set of E).

a.s. Almost surely, an event (i.e., an element of $\mathcal{F}$) A is true almost surely iff $\mathbb{P}(A) = 1$. We will often abusively mix up random vectors and classes of almost surely equal random vectors.

i.i.d. "independent and identically distributed".

$\mathbb{E}[\cdot|\cdot]$ Conditional expectation. Given a random variable $X \in \mathbb{R}$ and $\mathcal{G}$, a sub σ-algebra of $\mathcal{F}$, $\mathbb{E}[X|\mathcal{G}]$ is the unique random variable that

$\mathbb{P}(\cdot|\cdot)$ Conditional probability. Given a Borel set $A \subset \mathbb{R}$ and a random variable $X \in \mathbb{R}$, $\mathbb{P}(A|X) = \mathbb{E}[\mathbb{1}_A|X]$.

$\in, \pm$ Concentration around a pivot or around a deterministic equivalent, see Definitions 3 and 8.

$\propto$ Lipschitz concentration, see Definitions 2 and 10.

$\propto_c$ Convex concentration, see Definition 12.

$\mathcal{R}_X$ Observable diameter, see Remark 1.2.28

$\propto_\cdot^T$ Transversal concentration, see Definition 14.

$d\mu$ If μ is a probability law defined on E, for any function $f : E \to \mathbb{R}$, such that $\mu(f) = \int f d\mu = 1$, we note $f d\mu$ the measure verifying for all Borel set $B : f d\mu(B) = \int \mathbb{1}_B f d\mu$.

λ_p Lebesgue measure on $\mathbb{R}^p$.

β_q^p Uniform measure on the ball $\mathcal{B}_{\|\cdot\|_q}$ of $\mathbb{R}^p$.

σ_p Uniform measure on $\mathbb{S}^p$.

^{p} Exponential measure. If $p = 1$ $\nu^1 = \frac{e^{|\cdot|}}{2} d\lambda_1$, for $p \geq 1$, $\nu^p = \nu^1 \otimes \dots \nu^1$

($p$times).

$\mathcal{N}(0, I_p)$ Distribution of Gaussian vectors of $\mathbb{R}^p$ with zero mean and covariance I_p.

m_μ Stieltjes transform of the probability law μ on $\mathbb{R}$. Let $D \subset \mathbb{R}$ be the maximal Borel set such that $\mu(\mathbb{R} \setminus D) = 0$, then $\forall z \in \mathbb{C} \setminus D$, the Stieltjes transform is defined with the formula $m_\mu(z) = \int_D \frac{d\mu(w)}{z-w}$.

Introduction

Sample covariance matrices are key quantities in applied statistics in that they allow for the estimation of structural information in the second order statistics of the sampled vectors, and find a wide range of applications in fields as diverse as applied statistics (e.g., financial statistics, biostatistics), signal or data processing, wireless communications, etc. Precisely, for a set of n independent random vectors $x_1, \dots x_n \in \mathbb{R}^p$ stacked in a matrix $X = [x_1, \cdots x_n] \in \mathcal{M}_{p,n}$, the sample covariance matrix $S = \frac{1}{n}\sum_{i=1}^n x_i x_i^T = \frac{1}{n} XX^T \in \mathcal{M}_p$ provides an estimator for $\Sigma = \frac{1}{n}\mathbb{E}[XX^T]$.

If the number of independently sampled vectors n is large compared to the dimension p of the vectors, then under usually mild assumptions S converges to Σ. When p and n have the same order of magnitude though, again under classical assumptions, the operator norm difference $\|S - \Sigma\|$ usually does not vanish and S is thus not a consistent estimator for Σ. In their now famous article [MP67], Marčenko and Pastur proved that, if the vectors x_i, in addition to being independent, have independent entries of zero mean and unit variance, then, as $n, p \to \infty$ with $p/n \to c \in (0, \infty)$, the normalized counting measure of the eigenvalues of S converges a.s. to a limiting distribution having a continuous density, and now referred to as the Marčenko–Pastur distribution. From this article on, many works have provided generalizations of [MP67]. This is the case for instance of [SB95], where the authors assume that the x_i can be written under the form $x_i = \Sigma^{\frac{1}{2}} z_i$ for z_i a vector with independent zero mean and unit variance entries. It is to be noted that the independence [MP67], linear dependence [SB95], or vanishing dependence [Ada11] between the entries of X is key in the approach pursued in these articles as it provides a necessary additional degree of freedom in the derivation of the proofs.

These earlier results thus found many practical applications in scientific fields involving large dimensional matrix models with mostly linearly dependent entries, most notably in applied statistics, electrical engineering and computer science. But the renewed interest for machine learning applications, spurred by the big data era, has recently brought forward the need to understand and improve algorithms and methods relying on random matrix models involving non-linear relations between their entries. In some scenarios, as with kernel matrices (that is, matrices $K \in \mathbb{R}^{n \times n}$ with entries of the type $K_{ij} = f(x_i^T x_j)$ for some non-linear function f), an asymptotic equivalence between these matrices and classical matrices with linearly dependent entries can be proved [El 10, CB16, KC17], thereby transferring the asymptotic analysis of the former to that of the latter. In

other scenarios though, such asymptotic equivalences are not available. This is in particular the case of so-called random feature maps and neural networks. In random feature maps, the vectors x_i can be expressed under the form $x_i = \sigma(Wz_i)$ for some non-linear (referred to as the activation) function $\sigma : \mathbb{R} \to \mathbb{R}$, here applied entry-wise, W a given matrix, and z_i yet another random vector. The randomness in random feature maps arises from the fact that W is usually chosen at random, often with independent and identically distributed entries (this way allowing to produce p independent non-linear "features" of the vector z_i). Prior to training, neural networks (in particular feedforward neural nets) may usually be seen as a cascade of such random feature maps. In evaluating the performance of algorithms and methods based on random feature maps and neural networks, it is often of importance to understand the statistical behavior of S.

Still in the scope of the statistical analysis of data processing algorithms, where sample covariance matrices built upon *real data vectors* x_i are considered, it is also quite restrictive, if not disputable, to assume that x_i can be written as $x_i = \mu + \Sigma^{\frac{1}{2}} w_i$ for some deterministic μ and with w_i having independent entries.

As we shall see in the course of the article, a very convenient assumption to be made on x_i in order (i) to answer the aforementioned controversial real dataset modeling, (ii) to properly model random feature map and neural network and (iii) to largely generalize the Marčenko–Pastur and related results, is to propose that x_i *satisfies a vector-concentration inequality*. In a nutshell, concentration inequalities being stable under (bounded) linear operations *and non-linear Lipschitz operations*, the framework proposed in this study allows for a natural study of models of S with independent x_i of the form $\sigma(Wz_i)$ with z_i itself a concentrated random vector (so for instance itself of the type $\sigma_2(W_2 y_i)$ with y_i concentrated, and so on).

The objective of the article is precisely to provide a consistent method for the analysis of S with x_i independent concentrated random vectors *based on elementary concentration inequality principles*, particularly suited to practical applications in large dimensional machine learning. The article notably extends the results of El-Karoui in [El 09], who first exploited concentration of measure arguments in place of independence in the classical proof approaches of [MP67, SB95]. The technical arguments and findings of [El 09] are nonetheless quite specific and treated lemma after lemma, some of the main results being valid only for a restricted class of concentrated random vectors (such as elliptically distributed random vectors). We rather aim here at a self-contained generic framework for the manipulation of a large class of random matrix models using quite generic concentration identities. For instance, in the present article, we provide a very simple expression of the concentration of quantities of the type $x^T A x$ when x is normally concentrated and A a deterministic matrix, with a very synthetic proof (see Lemma 1.1.9 and Theorem 1.2.52). The present work also follows after a previous article by the same authors [LC17], in which a par-

ticular model of concentrated random vectors $x_i = \sigma(Wz_i)$ was studied in the aim of analysing the performances of a one hidden-layered non-linear random neural network, commonly referred to as an extreme learning machine [HZS06]. Concentration was there induced by the randomness of W and the Lipschitz character of σ. Yet, not imposing that the input data z_i are themselves concentrated random vectors led to restrictions in the results of [LC17], to which the present article easily circumvents. In a subsequent companion article, applications of the present results to the asymptotic classification and regression performance of extreme learning machines will be devised.

The present article is also quite general in that it allows to handle sample covariance matrices for data x_i arising from either of the two standard classes of concentrated random vectors, that is (i) uniformly continuous (so in particular Lipschitz) transformations of standard Gaussian random vectors, and (ii) affine transformations of random vectors with independent and bounded entries; the latter setting follows from the works of Talagrand [Tal95] and requires a more subtle approach since vector concentration of this kind is only stable through Lipschitz *and convex* functionals.

The remainder of the article is structured as follows. We introduce in the first section a somewhat original (but convenient to our analysis) approach to the notion of concentration of measure, probabilistically oriented and based on a collection of efficient lemmas concerning random variables and subsequently exported to the case of random vectors and matrices (one original specificity of our approach is to study the stability of the concentration through basic operations like sums and products in algebras). The aim of this section is not only to prepare the ground for the study of the sample covariance but also to offer a generic toolbox beyond our present scope; for this reason we maintain as general hypotheses as possible in this section. In the second section we then devise so-called deterministic equivalents for the sample covariance matrix model under study and we eventually illustrate the robustness of our results to both artificial and real datasets.

Preamble

Remark 0.0.1 (Definition of S)**.** *As a first remark, note that we abusively define the sample covariance matrix S as $S = \frac{1}{n}XX^T$ rather than the conventional $\frac{1}{n}XX^T - \frac{1}{n}\mathbb{E}X\mathbb{E}X^T$. This choice is not completely marginal in that not subtracting the matrix $\mathbb{E}X\mathbb{E}X^T$ from S in general brings additional complications (because this matrix will in general have unbounded norm as $p, n \to \infty$); our concentration of measure approach however efficiently deals with this term. From a practical standpoint, the removal of $\mathbb{E}X\mathbb{E}X^T$ means that this matrix can be computed, which is in general not the case. Alternatively, centering S by the empirical average $\frac{1}{n}X1_n1_n^TX^T$ presumes that the x_i are identically distributed which, as is common in classification applications in machine learning, is also not a desirable assumption.*

In the course of the article, we will be particularly interested in the eigenvalues $l_1, \ldots, l_p$ of S. More precisely, we will consider the spectral distribution F of S defined as the following normalized counting measure of the eigenvalues of S (a random probability measure):

$$F = \frac{1}{p} \sum_{i=1}^{p} \delta_{l_i},$$

where δ_x is the Dirac measure centered at x. As is conventional in large dimensional random matrix theory, we shall retrieve information of F through an approximation of its Stieltjes transform m_F, defined as:

$$m_F(z) = \int_w \frac{1}{w - z} dF(w) = \frac{1}{p} \operatorname{Tr} \left((S - zI_p)^{-1} \right)$$

where $z \in \mathbb{C}$ belongs to the complementary of the support of F. Since the matrix S is nonnegative definite, it suffices to study $m_F(z)$ for $z \in \mathbb{R}_-$ to recover m_F by analytic extension. For convenience, rather than working on $\mathbb{R}^-$ we shall consider $m_F(-z)$ for z in $\mathbb{R}^+$.

The random matrix $Q_S(z) = (S + zI_p)^{-1}$, referred to as the resolvent of S, will thus be an object of fundamental importance in the remainder. It shall be denoted Q when non-ambiguous. The convenience of the resolvent Q as a cornerstone of random matrix theory analysis is due in part to its simple boundedness properties:

Lemma 0.0.2. *Given a matrix $R \in \mathcal{M}_{p,n}$, a nonnegative definite symmetric matrix $C \in \mathcal{M}_p$ and $z \in \mathbb{R}_+$, we have the following bounds:*

$$\|Q_C(z)\| \leq \frac{1}{z} \qquad \|Q_C(z)C\| \leq 1 \qquad \left\| Q_{\frac{1}{n}RR^T}(z)R \right\| \leq \frac{\sqrt{n}}{\sqrt{z}}.$$

Proof. The upper bound for $\|Q_C(z)\|$ follows from the smallest eigenvalue of $C + zI_p$ being larger than z. The second result is a consequence of the simple identity $Q_C(z)C + zQ_C(z) = I_p$ and the fact that $Q_C(z)$ is symmetric positive definite. Combining the first two results, we have the bound

$$\left\| Q_{\frac{1}{n}RR^T}(z) \frac{1}{n} RR^T Q_{\frac{1}{n}RR^T}(z) \right\| \leq \frac{1}{z},$$

providing the last result. □

Beyond the eigenvalues of S, our interest (also driven by numerous applications in electrical engineering and data processing) will also be on the eigenvectors of S. For $U = [u_1, \ldots, u_l]$ an eigen-basis for the eigenspace associated to the multiplicity-l eigenvalue λ of S, remark that $UU^T = \frac{1}{2\pi\imath} \oint_{\Gamma_\lambda} Q(-z)dz$ for

Γ_λ a negatively oriented complex contour surrounding λ only. As such, beyond studying the trace of Q (and thus the Stieltjes transform of S), our interest is also on characterizing Q itself.

Precisely, we shall study so-called deterministic equivalents for Q, the precise definition of which is given in Definition 8 and that can be described as deterministic matrices $\tilde{Q}$ verifying $\operatorname{Tr} A(Q-\tilde{Q}) \to 0$ when A has unit norm (depending on the tightness of the concentration around $\tilde{Q}$, the norm of A considered will either be the Frobenius norm $\|A\|_F = \sqrt{\operatorname{Tr} AA^T}$, either the nuclear norm $\|A\|_1 = \operatorname{Tr}(AA^T)^{\frac{1}{2}}$). Note that if we control $\operatorname{Tr} A(Q-\tilde{Q})$, we also control $u^T(Q-\tilde{Q})v = \operatorname{Tr} vu^T(Q-\tilde{Q})$ for two deterministic vectors $u, v \in \mathbb{R}^p$ with unit norm (in that case $\|vu^T\|_F = \|vu^T\|_1 = \|u\|\,\|v\|$). In the following lines, we provide an outline for our subsequent development. First, let us note that a naive approach would be to think that $(\Sigma + zI_p)^{-1}$ might be a deterministic equivalent for Q. This turns out to be incorrect under general assumptions. Instead, letting $\tilde{Q} = (\Sigma' + zI_p)^{-1}$ where Σ' is some deterministic matrix to determine, we may first compute the difference:

$$\tilde{Q} - \mathbb{E}Q = \mathbb{E}\left[Q\left(\frac{1}{n}XX^T - \Sigma'\right)\tilde{Q}\right] = \sum_{i=1}^n \frac{1}{n}\mathbb{E}\left[Q(x_ix_i^T - \Sigma')\tilde{Q}\right].$$

Here, to go further, we need to make explicit the dependence between x_i and the matrix Q in order to evaluate the expectation of the product Qx_i. Let us denote $X_{-i} \in \mathcal{M}_{p,n-1}$ the matrix X deprived of its i-th column, which leads us to defining the matrices $S_{-i} = \frac{1}{n}X_{-i}X_{-i}^T$ and $Q_{-i} = (S_{X_{-i}} + zI_p)^{-1}$ (which is not $Q_{S_{X_{-i}}}$ as n is not turned in $n-1$). To handle the dependence between x_i and Q, we will massively exploit in the paper the classical Schur identities:

$$Q = Q_{-i} - \frac{1}{n}\frac{Q_{-i}x_ix_i^TQ_{-i}}{1+\frac{1}{n}x_i^TQ_{-i}x_i} \qquad \text{and} \qquad Qx_i = \frac{Q_{-i}x_i}{1+\frac{1}{n}x_i^TQ_{-i}x_i}. \quad (1)$$

The second inequality allows us to disentangle the relation between Q and x_i in the product Qx_i with a similar but easier to apprehend product $Q_{-i}x_i$ and a factor $1/\left(1+\frac{1}{n}x_i^TQ_{-i}x_i\right)$ easily controllable thanks to a first call to concentration inequalities (see subsequently Property 1.2.52). This leads us to:

$$\tilde{Q} - \mathbb{E}Q = \sum_{i=1}^n \frac{1}{n}\mathbb{E}\left[Q_{-i}\left(\frac{x_ix_i^T}{1+\frac{1}{n}x_i^TQ_{-i}x_i} - \Sigma'\right)\tilde{Q}\right] - \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[Q_{-i}x_ix_i^TQ\Sigma'\tilde{Q}\right].$$

We will see that, due to the supplementary factor $1/n$, the norm of the rightmost random matrix will be negligible compared to that of the other right-hand side matrix. Thus, if one assumes, say, that the random vectors $(x_i)_{1\le i\le n}$ follow the same law (not our general assumption in the article), one would choose naturally $\Sigma' = \mathbb{E}[x_ix_i^T]/(1+\delta)$ where $\delta = \frac{1}{n}\mathbb{E}[x_i^TQ_{-i}x_i] = \frac{1}{n}\operatorname{Tr}\mathbb{E}[Q_{-i}]\mathbb{E}[x_ix_i^T]$. Then, having established that $\frac{1}{n}\mathbb{E}\operatorname{Tr} AQ$ (and thus $\frac{1}{n}\mathbb{E}\operatorname{Tr} AQ_{-i}$) is close to $\frac{1}{n}\operatorname{Tr} A\tilde{Q}$, in particular here for $A = \mathbb{E}[x_ix_i^T]$, one may establish an implicit equation for δ not involving expectations over Q (or Q_{-i}). The remaining issue, at the core

of our present analysis, is now to find a convenient setting in the modeling of the x_i for which such a choice of Σ' would guarantee that $\frac{1}{p}\operatorname{Tr}(\mathbb{E}Q - \tilde{Q})$ indeed vanishes. That will be related to the concentration of "chaos-like" quantities (see [Ver17, Section 6.1]) such as $x_i^T Q_{-i} x_i$ but also of quantities such as $\operatorname{Tr} AQ$ or $u^T Q v$ as we shall see.

In the first part of the article, we will show that a comfortable approach is to structure our results as an outgrowth of the concentration of measure phenomenon. To give a brief insight into our approach, let us give the original theorem of the theory that we owe to Paul Pierre Levy in the beginning of the twentieth century and concerns the concentration of the uniform distribution σ_p on the sphere $\mathbb{S}^p$. To demystify the result we mention that it is closely linked to the isoperimetrical inequality.

Theorem 0.0.3 (Normal concentration of σ_p, [Led01, Theorem 2.3])**.** *Given a degree $p \in \mathbb{N}$ and a random vector $Z \sim \sigma_p$, for any 1-Lipschitz function $f : \mathbb{R}^{p+1} \to \mathbb{R}$, we have the inequality :*

$$\mathbb{P}\left(|f(Z) - m_f| \geq t\right) \leq 2e^{-(p-1)t^2/2} \tag{2}$$

where m_f is a median of $f(Z)$ verifying by definition $\mathbb{P}(f(Z) \geq m_f), \mathbb{P}(f(Z) \leq m_f) \geq \frac{1}{2}$.

There exist plenty of other distributions that verify this inequality like the uniform distribution on the ball, on $[0,1]^n$, or the Gaussian distribution $\mathcal{N}(0, I_p)$ that are presented in [Led01] ; more generally, for Riemannian manifolds, the concentration can be interpreted as a positive lower bound on the Ricci curvature (see Gromov appendix in [MS86] or [Led01, Theorem 2.4]). Our study will not evoke the design of such distributions and the validity of their concentration since this work has already been treated in the past; we will rather directly assume a similar concentration inequality to (2) for the data matrix X and then, with the tools developed in Section 1, we will infer the concentration of the Stieltjes transform $m_F(z)$ for any $z > 0$ and devise a good estimator of it. To prepare the applications to come, the dimension p must be thought to be quasi-asymptotic. In that sense the tightening of the concentration when the metric diameter of the distribution stays constant is a remarkable specificity of the concentration phenomenon ; this is furthermore a necessary condition, required for our study in Section 2. The concentration inequality verified on the sphere will structure our paper in a way that we will try to express our result with that form as often as possible to pursue Levy's idea; we will thus choose the short notation $f(Z) \in m_f \pm 2e^{-(p-1)\cdot^2/2}$ to express it in a simple way.

1. The Concentration of Measure Framework

As Milman has advocated from the beginning of the seventies ([MS86]), the concentration of a random object appears to be an essential feature leading immediately to a lot of implications and controls on the object. We do not present here the historical introduction of concentration of the measure. Rather than

the geometric, distribution-oriented approach, we directly adopt a probabilistic point of view on the notion. That does not mean that the presentation of the theory will be incomplete. On the contrary, we display here, almost always with their proofs, all the important theorems and propositions necessary to apprehend the theory and set rigorously the study of Section 2 about the sample covariance. The usual approach to the concentration of measure is to start with geometric inequalities ruling high dimensional Banach spaces and then track from these powerful results some probabilistic properties on the real functionals, the "observable world". We propose here a reversed approach where we start from the probabilistic results on $\mathbb{R}$ which offer us some interesting reasoning schemes all the same. Then, once the reader convinced by the direct computation improvements offered by the theory, we perform the fundamental step consisting in considering high dimensional concentration properties. Many of the results are derived from the complete presentation of the theory made by Ledoux in [Led01].

1.1. Concentration of a random variable

1.1.1. Definition and first Properties

Definition 1 (Concentration function)**.** *Any non-increasing and left continuous function $\alpha : \mathbb{R}_+ \to [0, 1]$, is called a concentration function.*

Given a random variable Z and a concentration function α, we choose first to express the α-*concentration* of Z through the introduction of an independent copy Z'.

Definition 2 (Concentration of a random variable)**.** *The random variable Z is said to be* α-concentrated*, and we write $Z \propto \alpha$, iff for any independent copy Z' :*

$$\forall t > 0 \ : \quad \mathbb{P}\left(|Z - Z'| \geq t\right) \leq \alpha(t). \tag{3}$$

In Definition 1, the different properties required for the concentration function ($\alpha \in [0, 1]$, non-increasing, left-continuous) are important features of any function $t \mapsto \mathbb{P}\left(|X| \geq t\right)$ when X is a random variable. In particular, if we often deal with concentration functions α that take values outside of $(-\infty, 1]$, it is implicitly understood that we consider in the calculus $t \mapsto \min\{\alpha(t), 1\}$ instead of α.

In Definition 2, we see that α basically limits the variations of the random variable Z. The faster α decreases, the lower the variations of Z. Instinctively one might hope that this limitation of the variation would be equivalent to a concentration around some central quantity that will be called a *pivot* of the concentration in the following sense :

Definition 3 (Concentration around a pivot)**.** *Given $a \in \mathbb{R}$, the random variable Z is said to be* α-concentrated around the pivot a*, and we write $Z \in a \pm \alpha$,*

iff :

$$\forall t > 0 \ : \ \mathbb{P}\left(|Z - a| \geq t\right) \leq \alpha(t).$$

We call the parameter a a *pivot* because the concentration around a provides similar concentration around other values close to a. Given $\theta > 0$, let us denote τ_θ the operator defined on the set of concentration functions that verifies for any concentration function α :

$$\tau_\theta \cdot \alpha(t) = \begin{cases} 1 \text{ if } t \leq \theta \\ \alpha(t - \theta) \text{ if } t > \theta. \end{cases}$$

Then we have the simple Lemma :

Lemma 1.1.1. *Given a random variable $Z \in \mathbb{R}$, a concentration function α, two real numbers $a, b \in \mathbb{R}$ and $\theta > 0$, we have the implication :*

$$\begin{cases} Z \in a \pm \alpha \\ \|a - b\| \leq \theta \end{cases} \qquad \Longrightarrow \qquad Z \in b \pm \tau_\theta \cdot \alpha.$$

At first sight, there exists no pivot $a \in \mathbb{R}$ such that Definition 2 and Definition 3 are equivalent. We can however find an interesting relation considering the case $a = m_Z$ where m_Z is a median of Z, verifying by definition $\mathbb{P}(Z \geq m_Z) \geq 1/2$ and $\mathbb{P}(Z \leq m_Z) \geq 1/2$.

Proposition 1.1.2 (Concentration around the median, from [Led01, Corollary 1.5])**.** *Given a random variable Z, a median m_Z of Z and a concentration function α, we have the implications :*

$$Z \propto \alpha \qquad \Longrightarrow \qquad Z \in m_Z \pm 2\alpha \qquad \Longrightarrow \qquad Z \propto 4\alpha(\,\cdot\,/2)$$

where $\alpha(\,\cdot\,/2)$ is defined as being the function $t \mapsto \alpha(t/2)$.

We see here that if Z is α-concentrated, the tail of Z, i.e., the behavior of Z far from the median, is closely linked to the decreasing speed of α.

Proof. We need to consider the fact that $\mathbb{P}\left(Z = m_Z\right) = \epsilon$ may be non zero. Therefore there exist ϵ_1, ϵ_2 such that $\epsilon = \epsilon_1 + \epsilon_2$ and :

$$\mathbb{P}\left(Z < m_Z\right) = \frac{1}{2} - \epsilon_1 \qquad\qquad \mathbb{P}\left(Z > m_Z\right) = \frac{1}{2} - \epsilon_2.$$

Let us take $t > 0$. The first result follows from the inequalities :

$$\begin{aligned}\mathbb{P}\left(|Z - Z'| \geq t\right) &= \mathbb{P}\left(|Z - Z'| \geq t, Z' < m_Z\right) + \mathbb{P}\left(|Z - Z'| \geq t, Z' > m_Z\right) \\ &\quad + \mathbb{P}\left(|Z - m_Z| \geq t, Z' = m_Z\right) \\ &\geq (1/2 - \epsilon_1)\,\mathbb{P}\left(Z \geq t + Z' \mid Z' < m_Z\right) \\ &\quad + (1/2 - \epsilon_2)\,\mathbb{P}\left(Z' \geq t + Z \mid Z' > m_Z\right) \\ &\quad + \epsilon\,\mathbb{P}\left(|Z - m_Z| \geq t\right) \\ &\geq (1/2 - \epsilon)\left(\mathbb{P}\left(Z \geq t + m_Z\right) + \mathbb{P}\left(m_Z \geq t + Z\right)\right) \\ &\quad + \epsilon\,\mathbb{P}\left(|Z - m_Z| \geq t\right) \\ &= \frac{1}{2}\,\mathbb{P}\left(|Z - m_Z| \geq t\right).\end{aligned}$$

The second result is true for any real $a \in \mathbb{R}$ replacing m_Z :

$$\begin{aligned}\mathbb{P}\left(|Z - Z'| \geq t\right) &\leq \mathbb{P}\left(|Z - m_Z + m_Z - Z'| \geq t\right) \\ &\leq 2\mathbb{P}\left(|Z - m_Z| \geq t/2\right).\end{aligned}$$

□

The reason why Definition 2 was presented firstly and thus given the importance of the naturally underlying definition, even though Definition 3 might seem more intuitive, lies in its immediate compatibility to the composition with any Lipschitz function and more generally with any uniformly continuous function. The uniform continuity of a function is sized by its *modulus of continuity*. Although we presently work on $\mathbb{R}$, we give a general definition of the continuity modulus between two normed vector spaces because it will be useful in the next sections.

Definition 4 (Uniform continuity)**.** *Any non decreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ continuous and null in 0 is called a modulus of continuity. Given two normed vector space $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$, a function $f : E \to F$ is said to be continuous under the modulus of continuity ω if :*

$$\forall x, y \in E \quad : \quad \|f(x) - f(y)\|_F \leq \omega(\|x - y\|_E).$$

When $\forall t > 0$, $\omega(t) = \lambda t^\nu$ for $\lambda > 0$ and $\nu \in (0, 1]$, we say that f is (λ, ν)-Hölder continuous, ν is called the Hölder exponent *and λ is called the* Lipschitz coefficient *; indeed, if $\nu = 1$, f is said to be λ-*Lipschitz*. Of course a function admits a modulus of continuity iff it is uniformly continuous.*

Lemma 1.1.3. *Let us consider a random variable Z, a concentration function α, and a function $f : \mathbb{R} \to \mathbb{R}$ continuous under a modulus of continuity ω. We allow ourselves to write ω^{-1} the pseudo inverse of ω defined (for ω bijective or not) as : $\omega^{-1}(w) = \inf\{t, \omega(t) \geq w\}$. We have the implication :*

$$Z \propto \alpha \implies f(Z) \propto \alpha\left(\omega^{-1}\left(\cdot\right)\right).$$

Proof. It is straightforward to write :

$$\begin{aligned}\mathbb{P}\left(|f(Z)-f(Z')|\geq t\right) &\leq \mathbb{P}\left(\omega(|Z-Z'|)\geq t\right)\\ &\leq \mathbb{P}\left(|Z-Z'|\geq \omega^{-1}(t)\right)\leq \alpha\left(\omega^{-1}\left(t\right)\right)\end{aligned}$$

since for any $w,t>0$, $\omega(t)\geq w \implies t\geq \omega^{-1}(w)$ by definition of the pseudo inverse. □

Remark 1.1.4. *Definition 3 is not so compatible with the ω-continuous transformations. Indeed, the simple developments that one finds in the proof above cannot be performed when strictly assuming that $Z\in m_Z\pm\alpha$. In this case, one would rather combine Lemma 1.1.3 with Proposition 1.1.2 to find :*

$$Z\in m_Z\pm\alpha \implies f(Z)\in m_{f_Z}\pm 2\alpha\left(\frac{1}{2}\omega^{-1}\left(\cdot/2\right)\right).$$

The stability with respect to ω-continuous functions reflects the fact that a ω-continuous function contains the spreading of the distribution up to the modulus of continuity.

One property that we could expect from α-concentration is a stability towards the sum. Up to a multiplying factor of 2, this is the case :

Lemma 1.1.5. *Given two random variable Z_1 and Z_2 and two concentration functions α,β, we have the implication :*

$$Z_1\propto\alpha \;\; \textit{and} \;\; Z_2\propto\beta \qquad\implies\qquad Z_1+Z_2\propto\alpha(\cdot/2)+\beta(\cdot/2).$$

Proof. Recall that Z_1' and Z_2' are two independent copies respectively of Z_1 and Z_2. There is no reasons for Z_1' to be independent of Z_2 (resp., Z_2' of Z_1). The idea is to decompose the threshold $t>0$ in $\frac{t}{2}+\frac{t}{2}$:

$$\begin{aligned}&\mathbb{P}\left(|Z_1+Z_2-Z_1'-Z_2'|\geq t\right)\\ &\qquad\leq \mathbb{P}\left(|Z_1-Z_1'|\geq\frac{t}{2}\right)+\mathbb{P}\left(|Z_2-Z_2'|\geq\frac{t}{2}\right) \;\leq\; \alpha\left(\frac{t}{2}\right)+\beta\left(\frac{t}{2}\right).\end{aligned}$$

□

Remark 1.1.6. *This time the idea of the demonstration works the same with the setting of Definition 3, and we have :*

$$Z_1\in a\pm\alpha \;\; \textit{and} \;\; Z_2\in b\pm\beta \quad\implies\quad Z_1+Z_2\in a+b\pm\alpha(\cdot/2)+\beta(\cdot/2).$$

In a first approach, α-concentration performs badly with the product, as it requires a bound on both random variables involved which greatly reduces the number of possible applications.

Lemma 1.1.7. *Given two bounded random variable Z_1 and Z_2 such that $|Z_1|\leq K_1$ and $|Z_2|\leq K_2$ and two concentration functions α,β :*

$$Z_1\propto\alpha \;\; \textit{and} \;\; Z_2\propto\beta \qquad\implies\qquad Z_1Z_2\propto\alpha\left(\frac{\cdot}{2K_2}\right)+\beta\left(\frac{\cdot}{2K_1}\right).$$

Proof. Given $t > 0$:

$$\begin{aligned}\mathbb{P}\left(|Z_1Z_2 - Z_1'Z_2'| \geq t\right) &\leq \mathbb{P}\left(|Z_1(Z_2 - Z_2')| \geq \frac{t}{2}\right) + \mathbb{P}\left(|(Z_1 - Z_1')Z_2'| \geq \frac{t}{2}\right) \\ &\leq \mathbb{P}\left(|Z_2 - Z_2'| \geq \frac{t}{2K_1}\right) + \mathbb{P}\left(|Z_1 - Z_1'| \geq \frac{t}{2K_2}\right).\end{aligned}$$

□

Actually, the setting of Definition 3 is more convenient here than the setting of Definition 2 because it allows us to only require one random variable to be bounded :

Lemma 1.1.8. *Given two random variables Z_1 and Z_2 such that $|Z_1| \leq K_1$, two pivot $a, b \in \mathbb{R}$ and two concentration functions α, β, if $b \neq 0$ one has the implication :*

$$Z_1 \in a \pm \alpha \quad \textit{and} \quad Z_2 \in b \pm \beta \quad \Longrightarrow \quad Z_1Z_2 \in ab \pm \alpha\left(\frac{\cdot}{2\,|b|}\right) + \beta\left(\frac{\cdot}{2K_1}\right),$$

if $b = 0$, then the concentration of Z_1 and Z_2 implies that $Z_1Z_2 \in 0 \pm \beta(\frac{\cdot}{K_1})$.

We may even go further and dispense with the bounding hypothesis to get in this case a slightly more complicated concentration form :

Proposition 1.1.9. *Given two random variable Z_1 and Z_2, two pivot $a, b \in \mathbb{R}$ and two concentration functions α, β such that $Z_1 \in a \pm \alpha$ and $Z_2 \in b \pm \beta$, if $a \neq 0$ and $b \neq 0$, then Z_1Z_2 is concentrated around ab with :*

$$Z_1Z_2 \in ab \pm \alpha\left(\sqrt{\frac{\cdot}{3}}\right) + \alpha\left(\frac{\cdot}{3\,|b|}\right) + \beta\left(\sqrt{\frac{\cdot}{3}}\right) + \beta\left(\frac{\cdot}{3\,|a|}\right).$$

If $a = 0$ and $b \neq 0$, we get $Z_1Z_2 \in 0 \pm \alpha\left(\sqrt{\frac{\cdot}{2}}\right) + \alpha\left(\frac{\cdot}{2|b|}\right) + \beta\left(\sqrt{\frac{\cdot}{2}}\right)$ and if $a = b = 0$, $Z_1Z_2 \in 0 \pm \alpha\left(\sqrt{\cdot}\right) + \beta\left(\sqrt{\cdot}\right)$.

Proof. We just prove the result for $a, b \neq 0$ since the other cases are simpler. The idea is to use the algebraic identity : $xy - ab = (x-a)(y-b) + a(y-b) + b(x-a)$ and the implication $xy \geq t \Rightarrow x \geq \sqrt{t}$ or $y \geq \sqrt{t}$ (for $x, y, t \geq 0$):

$$\begin{aligned}\mathbb{P}\left(|Z_1Z_2 - ab| \geq t\right) \leq\ & \mathbb{P}\left(|Z_1 - a| \geq \sqrt{\frac{t}{3}}\right) + \mathbb{P}\left(|Z_2 - b| \geq \sqrt{\frac{t}{3}}\right) \\ & + \mathbb{P}\left(|Z_1 - a|\,|b| \geq \frac{t}{3}\right) + \mathbb{P}\left(|Z_2 - b|\,|a| \geq \frac{t}{3}\right).\end{aligned}$$

□

In the case of the square of a random variable or even an integer power of any size, the concentration given by Proposition 1.1.9 can simplify.

Proposition 1.1.10. *Given $m \in \mathbb{N}$ and a random variable $Z \in a \pm \alpha$ with $a \in \mathbb{R}$ and α, one has the concentration :*

$$Z^m \in a^m \pm \alpha \left(\frac{\cdot}{2^m |a|^{m-1}} \right) + \alpha \left(\left(\frac{\cdot}{2} \right)^{\frac{1}{m}} \right)$$

Proof. Let us employ the algebraic identity :

$$Z^m = (Z - a + a)^m = \sum_{i=0}^{m} \binom{m}{i} a^{m-i} (Z-a)^i = a^m + a^m \sum_{i=1}^{m} \binom{m}{i} \left(\frac{Z-a}{a} \right)^i.$$

If $\left|\frac{Z-a}{a}\right| \leq 1$, for any $i \in \{1, \dots m\}$, $\left|\frac{Z-a}{a}\right|^i \leq \left|\frac{Z-a}{a}\right|$ and conversely, if $\left|\frac{Z-a}{a}\right| \geq 1$, then $\left|\frac{Z-a}{a}\right|^i \leq \left|\frac{Z-a}{a}\right|^m$. This entails :

$$|Z^m - a^m| \leq (2|a|)^m \left(\left| \frac{Z-a}{a} \right| + \left| \frac{Z-a}{a} \right|^m \right)$$

and therefore :

$$\mathbb{P}\left(|Z^m - a^m| \geq t\right) \leq \mathbb{P}\left(|Z-a| \geq \frac{t}{2^m |a|^{m-1}} \right) + \mathbb{P}\left(|Z-a| \geq \left(\frac{t}{2} \right)^{\frac{1}{m}} \right).$$

□

In the same vein, we can give the concentration of the product of m random variables, we will though obtain better concentration constants in the case of q-exponential concentration that will be studied in the next subsection. For that reason we left the formulation and the proof of this result in Appendix A. In the same appendix, one can find the expression of the concentration of Z^r when $r \in \mathbb{R}_*^+$ (and Z is concentrated).

We conclude this section by setting the continuity of the concentration property. We adopt the classical formalism for the convergence of a random variable (or vector).

Definition 5. *We say that a sequence of random variables (or random vectors) Z_n converges in law (or "in distribution" or "weakly") to Z if for any real valued continuous function f with compact support :*

$$\lim_{n \to \infty} \mathbb{E}[f(Z_n)] = \mathbb{E}[f(Z)].$$

Although we rather find in the literature the definition mentioning *continuous and bounded* functions, this equivalent definition relying on the class of continuous functions with compact support is more adapted to our needs (see Proposition 1.2.59). We start with the preliminary lemma :

Lemma 1.1.11 ([Ouv09], Proposition 14.17)**.** *Let us consider a sequence of random variable Z_n, $n \in \mathbb{N}$, and a random variable Z with cumulative distribution functions respectively noted F_{Z_n}, $n \in \mathbb{N}$, and F_Z. The sequence $(Z_n)_{n \geq 0}$ converges in law to Z iff for any $t \in \mathbb{R}$ such that F_Z is continuous on t, $(F_{Z_n}(t))_{n \geq 0}$ converges to $F_Z(t)$.*

Proposition 1.1.12. *Consider a sequence of random variables $(Z_n)_{n\geq 0}$ that converges in law to a random variable Z, a sequence of pivot $(a_n)_{n\geq 0}$ converging to a pivot $a \in \mathbb{R}$ and a sequence of concentration functions $(\alpha_n)_{n\geq 0}$ that pointwise converges to a continuous concentration function α. If we suppose that, for any $n \in \mathbb{N}$, $Z_n \in a_n \pm \alpha_n$ then $Z \in a \pm \alpha$.*

Proof. For any $n \in \mathbb{N}$, let us note $Y_n = |Z_n - a_n|$ and $Y = |Z - a|$. We wish to show first that $(Y_n)_{n\geq 0}$ converges in law to Y. Let us consider for that purpose a continuous function $f : \mathbb{R} \to \mathbb{R}$ with compact support S and $\varepsilon > 0$. We know from the Heine-Cantor theorem that f is uniformly continuous, therefore there exists $\eta > 0$ such that :

$$|x - y| \leq \eta \quad \Longrightarrow |f(x) - f(y)| \leq \frac{\varepsilon}{2}.$$

Moreover, since $\lim_{n\to\infty} a_n = a$, there exists $n_0 > 0$ such that for any $n \geq n_0$, $|a_n - a| \leq \eta$. Eventually, if we introduce the function $g : x \to |x - a|$, we know that $f \circ g$ has a compact support $S' = a + S \cup a - S$, and thus there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$:

$$|\mathbb{E}[f(|Z_n - a|)] - \mathbb{E}[f(|Z - a|)]| = |\mathbb{E}[f \circ g(Z_n)] - \mathbb{E}[f \circ g(Z)]| \leq \frac{\varepsilon}{2}.$$

Therefore if we consider $n \geq \max(n_0, n_1)$:

$$\begin{aligned} |\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y)]| \quad &\leq |\mathbb{E}[f(|Z_n - a_n|) - f(|Z_n - a|)]| \\ &\quad + |\mathbb{E}[f(|Z_n - a|) - f(|Z - a|)]| \quad \leq \ \varepsilon. \end{aligned}$$

Then we know from Lemma 1.1.11 that for any t such that the cumulative distribution function $F_Y = \mathbb{P}(|Z - a| \leq \cdot)$ is continuous around t :

$$\mathbb{P}(|Z - a| \geq t) = \lim_{n\to\infty} \mathbb{P}(|Z_n - a_n| \geq t) \leq \lim_{n\to\infty} \alpha_n(t) = \alpha(t).$$

Since α is continuous and $t \to \mathbb{P}(|Z - a| \geq t)$ is decreasing, we recover the preceding inequality for any $t > 0$, and $Z \in a \pm \alpha$. □

In the setting of Definition 2, we can show the continuity of the concentration the same way introducing this time the random variables $Y_n = |Z_n - Z'_n|$ (Z'_n being a sequence of independent copies of Z_n). We present the next proposition without proof.

Proposition 1.1.13. *In the setting of Proposition 1.1.12, if for any $n \in \mathbb{N}$ $Z_n \propto \alpha_n$, then $Z \propto \alpha$.*

1.1.2. Exponential concentration

In [Led01], Ledoux defines a random variable as *normally concentrated* when the concentration function is of the form $Z \propto Ce^{-(\,\cdot\,/\sigma)^2}$ for two given constants $C \geq 1$ and $\sigma > 0$. The form of the concentration function has no real importance on

small dimensions since the result ensues from a mere bound on the Gaussian Q-function; it is however surprising that this form naturally appears for the uniform distribution on the sphere (see Theorem 0.0.3) and others in high dimensions.

In order to present a general picture that will be helpful later (when dealing with products of random variables), let us include the general case of *q-exponential concentrations* that we present with the formalism of Definition 3.

Definition 6. *Given $q > 0$, a random variable Z is said to be* q-exponentially concentrated *with* head parameter $C \geq 1$ *and* tail parameter $\sigma > 0$ *iff there exists a pivot $a \in \mathbb{R}$ such that $Z \in a \pm Ce^{-(\cdot/\sigma)^q}$.*

Example 1.1.14. *A random variable Z following a Gaussian distribution with zero mean (i.e. zero median) and unit variance is 2-exponentially concentrated with a tail parameter equal to $\sqrt{2}$: $Z \in 0 \pm 2e^{-(\cdot)^2/2}$.*

In practice, the random variables will depend on a random vector whose dimension, say p, tends to infinity, the tail parameter is then a function of p that represents the *asymptotic* speed of concentration since it has the same order as the standard deviation of Z. If the q-exponential concentration functions are employed reasonably, the asymptotic information can be transmitted from the head parameter to the tail parameter so that the head parameter would stay mainly uninformative and close to 1. For that purpose, the following lemma gives us an easy way to bound a given concentration to a close one with a head parameter equal to e.

Lemma 1.1.15. *Given $x, q > 0$ and $C \geq e$, we have the inequality :*

$$\min(1, Ce^{-x}) \leq ee^{-x/2\log(C)}$$

Proof. If $x \leq 2\log(C)$ the inequality is clear, and if $x \geq 2\log(C) \geq 2$, we deduce the result of the lemma from the equivalence

$$\log(C) - x \leq -\frac{x}{2\log(C)} \qquad \Longleftrightarrow \qquad x \geq \frac{2\log(C)^2}{2\log(C) - 1},$$

since $\frac{2\log(C)^2}{2\log(C)-1} \leq \frac{2\log(C)^2}{\log(C)} \leq x$. □

To place ourselves under the hypotheses of Lemma 1.1.15 and as it appears rather convenient in several propositions below we will suppose from now on that the tail parameter C is greater than e.

Exponential concentrations offer simple expressions of the concentration through shifting the pivot thanks to the following lemma.

Lemma 1.1.16. *Given the parameters $C \geq e$ and $q, \sigma, \theta > 0$:*

$$\forall t > 0 \ : \quad \tau_\theta \cdot Ce^{-(\cdot/\sigma)^q} \leq \max\left(e^{(\theta/\sigma)^q}, C\right) e^{-(\cdot/2\sigma)^q}$$

Proof. The increasing behavior of $t \mapsto t^q$ ensures that $(t-\theta)^q \geq (t/2)^q$ when $t \geq 2\theta$, therefore :

$$\forall t \geq 2\theta \ : \ \mathbb{P}\left(|Z - b| \geq t\right) \leq Ce^{-(t/2\sigma)^q}$$

The result is also true when $0 \leq t \leq 2\theta$ since :

$$\mathbb{P}\left(|Z-b| \geq t\right) \leq 1 \leq e^{(\theta/\sigma)^q} e^{-(2\theta/2\sigma)^q}.$$

□

This lemma, combined with lemma 1.1.1, clarifies the notion of *tail parameters*. The next corollary shows that it can be seen as the diameter of a "black hole" centered around any pivot of the concentration, in a sense that each value inside this "black hole" can be considered as a satisfactory pivot for the concentration ; this will be called later, in the case of random vectors, the *observable diameter* of the distribution, following Gromov terminology [Gro79].

Corollary 1.1.17. *Given $C \geq e$ and three positive parameters $\sigma, \lambda, q > 0$, two real a and b such that $|a-b| \leq \lambda\sigma$ and a random variable Z, one has the implication :*

$$Z \in a \pm Ce^{-(\cdot/\sigma)^q} \qquad \Longrightarrow \qquad Z \in b \pm C' \exp\left(-\left(\frac{\cdot}{2\sigma}\right)^q\right)$$

where $C' = \max(C, e^{\lambda^q})$.

Note that the interesting aspect of the result is the independence of the head parameter $C' = \max(C, e^{\lambda^q})$ to the tail parameter σ. Moreover the tail parameter stays unmodified when $q < 1$.

We now have all the elements to show that, due to the high concentration of exponentially concentrated random vectors, every median plays a pivotal role among the different constants that can localize the concentration.

Proposition 1.1.18 ([Led01, Proposition 1.8])**.** *Given a random variable Z, and a median m_Z of Z, if we suppose that $Z \in a \pm Ce^{-(\cdot/\sigma)^q}$ for a pivot $a \in \mathbb{R}$, then :*

$$Z \in m_Z \pm 2C \exp\left(-\left(\frac{\cdot}{2\sigma}\right)^q\right).$$

Proof. For some $\varepsilon > 0$, we choose $t_0 > \sigma\left(\log(2C) + \varepsilon\right)^{1/q}$. We know that $\mathbb{P}\left(|Z-a| \geq t_0\right) < \frac{1}{2}$ and consequently $|a - m_Z| \leq t_0$. Indeed if we suppose that $m_Z \geq a + t_0$, then :

$$1/2 \leq \mathbb{P}\left(Z \geq m_Z\right) \leq \mathbb{P}\left(Z - a \geq t_0\right) \leq \mathbb{P}\left(|Z-a| \geq t_0\right),$$

and we get the same absurd result if we suppose that $a \geq m_Z + t_0$. We can thus conclude thanks to Corollary 1.1.17 (with $C' = \max(C, \exp(\frac{t_0^q}{\sigma^q})) = 2Ce^{\varepsilon}$), letting ε tend to zero.

□

The tails of q-exponentially concentrated random variables can be controlled rather easily and roughly thanks to the next lemma that is based on the same simple mathematical inequalities that lead to Corollary 1.1.17.

Lemma 1.1.19. *Given a random variable Z, two parameters $C \geq e$ and $\sigma > 0$ an exponent $q > 0$ and a pivot $a \in \mathbb{R}$, if $Z \in a \pm Ce^{-(\cdot/\sigma)^q}$ then $\forall t \geq 2|a|$:*

$$\mathbb{P}(|Z| \geq t) \leq Ce^{-(t/2\sigma)^q}.$$

Very interestingly, exponential concentration is of great computation convenience to manage Hölder's inequality. For instance a general issue is to bound :

$$\mathbb{E}\left[(Z_1 - a_1)^{r_1} \cdots (Z_m - a_m)^{r_m}\right].$$

For any $\theta_1, \cdots, \theta_m \in (0,1)$ such that $\theta_1 + \cdots + \theta_m = 1$, Hölder's inequality gives us directly :

$$\mathbb{E}\left[(Z_1 - a_1)^{r_1} \cdots (Z_m - a_m)^{r_m}\right] \leq \prod_{i=1}^{m} \left(\mathbb{E}\left|Z_i - a_i\right|^{\frac{r_i}{\theta_i}}\right)^{\theta_i}. \tag{4}$$

As we will see in the next proposition, the quantities $\mathbb{E}|Z_i - a_i|^r$ can be bounded easily when $Z_i = a_i \pm Ce^{-(\cdot/\sigma)^{q_i}}$, and we will even show in the next proposition that the bounds on $\mathbb{E}|Z_i - a_i|^r$ for $r > 0$ can become a characterization (it is actually a *pseudo-characterization* since there is no equivalence) of q-exponential concentrations.

Proposition 1.1.20 (Moment characterization of concentration, [Led01, Proposition 1.10])**.** *Given a random variable Z, a pivot $a \in \mathbb{R}$, two exponents $r, q > 0$, and two parameters $C \geq e$ and $\sigma > 0$, we have the implications :*

$$Z \propto Ce^{-(\cdot/\sigma)^q} \;\Rightarrow\; \forall r \geq q : \mathbb{E}\left[\left|Z - Z'\right|^r\right] \leq C\Gamma\left(\frac{r}{q} + 1\right)\sigma^r \;\Rightarrow\; Z \propto Ce^{-\frac{(\cdot/\sigma)^q}{e}}$$

and

$$Z \in a \pm Ce^{-(\cdot/\sigma)^q} \;\Rightarrow\; \forall r \geq q : \mathbb{E}\left[|Z - a|^r\right] \leq C\Gamma\left(\frac{r}{q} + 1\right)\sigma^r \;\Rightarrow\; Z \in a \pm Ce^{-\frac{(\cdot/\sigma)^q}{e}},$$

where $\Gamma : r \mapsto \int_0^\infty t^{r-1}e^{-t}dt$, (if $n \in \mathbb{N}$, $\Gamma(n+1) = n!$).

In both results, the first implication consists in bounding an expectation with a probability; it will involve the Fubini relation, giving for any positive random variable Z :

$$\mathbb{E}Z = \int_Z \left(\int_0^\infty \mathbb{1}_{t \leq Z} dt\right) dZ = \int_0^\infty \mathbb{P}(Z \geq t)\, dt,$$

where $\mathbb{1}_{t \leq Z}$ is equal to 1 if $t \leq Z$ and to 0 otherwise.

The second implication consists in bounding a probability with an expectation; it is a consequence of Markov's inequality, for any non decreasing function $f : \mathbb{R} \to \mathbb{R}$:

$$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}f(Z)}{f(t)}.$$

These two key indications given we can start the proof.

Proof. We just prove the first implication, as it will be clear that both can be proved the same way. Let us first suppose that $r \geq q$. Knowing that $Z \propto e^{-(ct)^q}$, we consider Z', an independent copy of Z, and we can bound :

$$\begin{aligned}\mathbb{E}\left[|Z-Z'|^r\right] &= \int_0^\infty \mathbb{P}\left(|Z-Z'|^r \geq t\right) dt = \int_0^\infty rt^{r-1}\mathbb{P}\left(|Z-Z'| \geq t\right) dt \\ &\leq Cr\int_0^\infty t^{r-1}e^{-(t/\sigma)^q}dt = C\sigma^r r\int_0^\infty t^{r-1}e^{-t^q}dt\end{aligned}$$

and, since $r \geq q$:

$$r\int_0^\infty t^{r-1}e^{-t^q}dt = \frac{r}{q}\int_0^\infty t^{\frac{r}{q}-1}e^{-t}dt = \Gamma\left(\frac{r}{q}+1\right)$$

Now, assuming the second term of the implication chain, we know from Markov's inequality that $\forall r \geq q$:

$$\mathbb{P}\left(|Z-Z'| \geq t\right) \leq \frac{\mathbb{E}\left[|Z-Z'|^r\right]}{t^r} \leq C\Gamma\left(\frac{r}{q}+1\right)\left(\frac{\sigma}{t}\right)^r \leq C\left(\frac{r}{q(t/\sigma)^q}\right)^{r/q}.$$

If $t \geq e^{\frac{1}{q}}\sigma$, we can then set $r = \frac{qt^q}{e\sigma^q} \geq q$, and we get :

$$\mathbb{P}\left(|Z-Z'| \geq t\right) \leq Ce^{-(t/\sigma)^q/e}.$$

If $t \leq e^{\frac{1}{q}}\sigma$, we still know that $\mathbb{P}\left(|Z-Z'| \geq t\right) \leq 1$, and we conclude that :

$$\forall t > 0 \quad : \quad \mathbb{P}\left(|Z-Z'| \geq t\right) \leq \max(C,e)e^{-(t/\sigma)^q/e}.$$

□

Remark 1.1.21. *In the last proposition, we did not provide a bound on* $\mathbb{E}\left[|Z-Z'|^r\right]$ *and* $\mathbb{E}\left[|Z-a|^r\right]$ *for* $0 \leq r < q$ *since it is irrelevant to the characterization of* q*-exponential concentration. We may nonetheless easily bound these quantities with* $C\sigma^r$*. Getting inspiration from our previous derivations, we can indeed obtain when* $1 \leq r \leq q$ *:*

$$\begin{aligned}\mathbb{E}\left[|Z-Z'|^r\right] &\leq C\sigma^r\int_0^\infty rt^{r-1}e^{-t^q}dt \leq C\sigma^r\int_0^\infty t^r qt^{q-1}e^{-t^q}dt \\ &\leq C\sigma^r\int_0^1 qt^{q-1}e^{-t^q}dt + C\sigma^r\int_1^\infty t^q qt^{q-1}e^{-t^q}dt \\ &= C\sigma^r\int_0^\infty qt^{q-1}e^{-t^q}dt \quad = \quad C\sigma^r\end{aligned}$$

and when $r \leq 1 \leq q$*, we conclude with Jensen's inequality :*

$$\mathbb{E}\left[|Z-Z'|^r\right] \leq \mathbb{E}\left[|Z-Z'|\right]^r \leq C^r\sigma^r \leq C\sigma^r.$$

The following Lemma gives an alternative result to the aforementioned sufficiency of bounds on $\mathbb{E}\left[|Z-a|^r\right]$ (or on $\mathbb{E}\left[|Z-Z'|^r\right]$) for $r \in \mathbb{N}$.

Lemma 1.1.22. *Given a random variable $Z \in \mathbb{R}_+$ and three parameters $C > e$, $q, \sigma > 0$ one has the implication :*

$$\forall m \in \mathbb{N},\ \mathbb{E}[Z^m] \leq C \left(\frac{m}{q}\right)^{\frac{m}{q}} \sigma^m \implies \forall r \geq 0,\ \mathbb{E}[Z^r] \leq C e^{\frac{1}{e}} \left(\frac{r}{q}\right)^{\frac{r}{q}} \left(\frac{\sigma}{\bar{q}^{\frac{1}{q}}}\right)^r$$

where $\bar{q} = \min(q, 1)$.

Proof. When $r \leq 1$, we already know thanks to Jensen's inequality, by concavity of $t \mapsto t^r$, that :

$$\mathbb{E}[Z^r] \leq (\mathbb{E}[Z])^r \leq C \left(\frac{1}{q}\right)^{\frac{r}{q}} \sigma^r \leq C e^{\frac{1}{e}} \left(\frac{r}{q}\right)^{\frac{r}{q}} \left(\frac{\sigma}{q^{\frac{1}{q}}}\right)^r,$$

since $\forall t > 0$, $t^t \geq \frac{1}{e^{\frac{1}{e}}}$.

When $r \geq 1$, one can invoke the well known result concerning ℓ^r norms, where in our case, $\|Z\|_{\ell^r} = \mathbb{E}\left[|Z|^r\right]^{1/r}$. Let us consider the general case where we are given $p_1, p_2 > 0$ such that $p_1 \leq r \leq p_2$ and we consider $\theta \in (0, 1)$ satisfying $1/r = \theta/p_1 + (1-\theta)/p_2$. We then have the inequality :

$$\|Z\|_{\ell^r} \leq \|Z\|_{\ell^{p_1}}^{\theta} \|Z\|_{\ell^{p_2}}^{1-\theta}.$$

This implies :

$$\mathbb{E}\left[Z^r\right]^{\frac{1}{r}} \leq \left(C^{\frac{1}{p_1}} \sigma \left(\frac{p_1}{q}\right)^{\frac{1}{q}}\right)^{\theta} \left(C^{\frac{1}{p_2}} \sigma \left(\frac{p_2}{q}\right)^{\frac{1}{q}}\right)^{1-\theta} \leq \frac{C^{\frac{1}{r}} \sigma}{q^{\frac{1}{q}}} p_1^{\frac{\theta}{q}} p_2^{\frac{1-\theta}{q}}.$$

We would like to bound $p_1^{\frac{\theta}{q}} p_2^{\frac{1-\theta}{q}}$ with $r^{\frac{1}{q}}$. Unfortunately, for $\theta \in]0, 1[$, $p_1^{\theta} p_2^{1-\theta} > 1/(\theta/p_1 + (1-\theta)/p_2) = r$ (this is due to the inequality of arithmetic and geometric means, itself a consequence of the concavity of the log function). However, for the particular setting under study :

$$\left(\frac{\theta}{p_1} + \frac{1-\theta}{p_2}\right) p_1^{\theta} p_2^{1-\theta} \leq \left(\frac{1}{p_2} + \frac{(p_2 - p_1)\theta}{p_1 p_2}\right) p_2 \ \leq\ 1 + \frac{p_2 - p_1}{p_1}.$$

As a consequence, taking $p_1 = \lfloor r \rfloor$ and $p_2 = \lceil r \rceil$, one obtains :

$$\mathbb{E}\left[|Z|^r\right] \leq C \sigma^r \left(\frac{2r}{q}\right)^{\frac{r}{q}}.$$

□

Remark 1.1.23. *In Proposition 1.1.20, we saw that if $Z \in a \pm C e^{-(\cdot/\sigma)^q}$ then $|\mathbb{E}Z - a| \leq \mathbb{E}|Z - a| \leq C \left(\frac{1}{q}\right)^{1/q} \sigma < \infty$. Therefore any q-exponentially concentrated random variable admits a finite expectation.*

Now that we know it exists, we are going to show that the expectation $\mathbb{E}Z$ plays the same pivotal role as any median.

Corollary 1.1.24 ([Led01, Proposition 1.9])**.** *With the notations of the previous proposition, one has :*

$$Z \in a \pm Ce^{-(\,\cdot\,/\sigma)^q} \qquad \Longrightarrow \qquad Z \in \mathbb{E}Z \pm e^{\frac{Cq}{q}} e^{-(\,\cdot\,/2\sigma)^q}.$$

Proof. We suppose that $Z \in Ce^{-(\,\cdot\,/\sigma)^q}$ for a pivot $a \in \mathbb{R}$. Proposition 1.1.20 applied in the case $r = 1$ gives us $|a - \mathbb{E}Z| \leq C(\frac{1}{q})^{\frac{1}{q}}\,\sigma$. One can then invoke Corollary 1.1.17, to get the concentration :

$$Z \in \mathbb{E}Z \pm C'e^{-(\,\cdot\,/2\sigma)^q},$$

with $C' = \max(C, e^{\frac{Cq}{q}})$. It is then interesting to note that the function $q \mapsto \frac{Cq}{q}$ has a minimum in $\frac{1}{\log(C)}$ where it takes the value $e\log C$. Then we see that

$$C = e^{\log C} \leq e^{e\log C} \leq e^{\frac{Cq}{q}},$$

and we can simplify the head parameter to obtain the result of the corollary. □

If $Z \in a \pm Ce^{-(\,\cdot\,/\sigma)^p}$, we can then employ $\mathbb{E}Z$ as a pivot of Z to get a bound on the centered moments :

Corollary 1.1.25. *With the notations of Corollary 1.1.24, if we suppose that $Z \in a \pm Ce^{-(\,\cdot\,/\sigma)^q}$, then we have :*

$$\mathbb{E}\,|Z - \mathbb{E}Z|^r \leq e^{\frac{Cq}{q}}\,(2\sigma)^r \left(\frac{r}{q}\right)^{\frac{r}{q}}.$$

Provided two identically distributed random variables $Z_1, Z_2 \in \mathbb{R}$, even if Z_1 and Z_2 are not independent, one still intuitively expects that $Z_1 + Z_2$ and Z_1Z_2 vary at most like $2Z_1$ and Z_1^2, respectively. This simple intuition can be easily proved valid in the case of q-exponential concentrations thanks to the characterization given by Proposition 1.1.20. But before getting into this aspect, let us introduce the notion of *controlling random variables* that allows us to handle the case of non identically distributed Z_1 and Z_2.

Definition 7. *Given two random variables $Z \in \mathbb{R}$ and $Y \in \mathbb{R}_+$, one says that Y controls Z iif $\forall t \geq 0$:*

$$\mathbb{P}\left(|Z| \geq t\right) \leq \mathbb{P}\left(Y \geq t\right).$$

Given a set of random variables $Z_1, \ldots, Z_p \in \mathbb{R}_+$, any random variable Y admitting the cumulative distribution function :

$$F_Y(t) = \mathbb{P}\left(Y \leq t\right) = 1 - \sup_{1 \leq i \leq p} \mathbb{P}\left(|Z_i| > t\right)$$

clearly controls $Z_1, \dots, Z_p$. If we consider a random variable Z depending on $Z_1, \dots, Z_p$ and want to express the concentration of Z thanks to Proposition 1.1.20, we are led to bound the quantities $\mathbb{E}[|Z - Z'|^r]$ or $\mathbb{E}[|Z - a|^r]$. The next lemma gives the central idea allowing to control those quantities with expectations taken on Y only, when $r \in \mathbb{N}$ and $Z = P(Z_1, \dots, Z_p)$ is a polynomial functional with positive coefficients. It simply relies on Hölder's inequality and the fact that $\mathbb{E}[Z_i^m] \leq \mathbb{E}[Y^m]$ for all $m \in \mathbb{N}$, $1 \leq i \leq p$.

Lemma 1.1.26. *Given two integers $d, p \in \mathbb{N}$, a polynomial*

$$P(X_1, \dots, X_p) = \sum_{b_1 + \cdots + b_p \leq d} c_b X_1^{b_1} \cdots X_p^{b_p}$$

with positive coefficients $c_b = c_{b_1,\dots,b_p} \geq 0$, $b_1 + \cdots + b_p \leq d$ and p positive random variables $Z_1, \dots, Z_p \in \mathbb{R}_+$ (possibly dependent), for any nonnegative random variable $Y \geq 0$ controlling $Z_1, \dots, Z_p$, we have the inequality :

$$\mathbb{E}\left[P(Z_1, \dots, Z_p)\right] \leq \mathbb{E}\left[P(Y, \dots, Y)\right].$$

Proof. It is a simple application of Hölder's inequality :

$$\begin{aligned}
\mathbb{E}\left[P(Z_1, \dots, Z_p)\right] &= \sum_{b_1 + \cdots + b_p \leq d} c_b \mathbb{E}\left[Z_1^{b_1} \dots Z_p^{b_p}\right] \\
&\leq \sum_{b_1 + \cdots + b_p \leq d} c_b \mathbb{E}\left[Z_1^{|b|}\right]^{\frac{b_1}{|b|}} \dots \mathbb{E}\left[Z_p^{|b|}\right]^{\frac{b_p}{|a|}} \\
&= \sum_{b_1 + \cdots + b_p \leq d} c_b \mathbb{E}\left[Y^{|b|}\right] \;=\; \mathbb{E}\left[P(Y, \dots, Y)\right]
\end{aligned}$$

where $|b| = \sum_{i=1}^p b_i \leq p$. □

Lemma 1.1.22 combined with Lemma 1.1.26 provides us with the concentration of the sum of p random variables with a better tail parameter than the one that would be obtained if we would employ Proposition 1.1.5 p times.

Proposition 1.1.27. *Given the parameters $C > e$, $q, \sigma > 0$, $p \in \mathbb{N}$, p random variables $Z_1, \dots, Z_p \in \mathbb{R}$ satisfying, for any $i \in \{1, \dots, m\}$, $Z_i \in a_i \pm Ce^{-(\cdot/\sigma_i)^q}$ where $a_i \in \mathbb{R}$ and $\sigma_i > 0$, we have the concentration :*

$$Z_1 + \cdots + Z_p \in a_1 + \cdots + a_p \pm e^{1/e} C \exp\left(-\frac{q}{2e}\left(\frac{\cdot}{\sigma^p}\right)^q\right)$$

where $\sigma = \sum_{i=1}^p \sigma_i$

We will not prove here this proposition since it is a particular case of Proposition 1.2.6 concerning the linear concentration of random vectors. In the same setting, it is also possible to control the concentration of the product $Z_1 \cdots Z_p$.

Proposition 1.1.28. *Let us consider three parameters $C > e$, $q, \sigma > 0$, an integer $p \in \mathbb{N}$, and p random variables $Z_1, \dots, Z_p \in \mathbb{R}$ satisfying, for any $i \in \{1, \dots, m\}$, $Z_i \in a \pm Ce^{-(\cdot/\sigma)^q}$. Then, if $a \geq \sigma$:*

$$\forall m \in \mathbb{N} \ : \ \mathbb{E}\left[|Z_1 \cdots Z_p - a^p|^m\right] \leq C(2\sigma a^{p-1})^m \left(\frac{m}{q}\right)^{\frac{m}{q}} + (2\sigma)^{pm} \left(\frac{pm}{q}\right)^{\frac{pm}{q}}$$

while, if $a \leq \sigma$:

$$\forall m \in \mathbb{N} \ : \ \mathbb{E}\left[|Z_1 \cdots Z_p - a^p|^m\right] \leq 3C(2\sigma)^{pm} \left(\frac{pm}{q}\right)^{\frac{pm}{q}}.$$

Remark 1.1.29. *We did not express as in Proposition 1.1.27 the concentration of $Z_1 \cdots Z_p$ around a^p since the bounds on $\mathbb{E}\left[|Z_1 \cdots Z_p - a^p|^m\right]$ are too complex to allow for simple optimizations of the Chernoff inequalities as in the proof of Proposition 1.1.20. Still, for $a \leq \sigma$, we can infer from Lemma 1.1.22 (and Proposition 1.1.20) :*

$$Z_1 \cdots Z_p \in a^p \pm 3e^{1/e} C \exp\left(-\frac{q}{2e}\left(\frac{\cdot}{2^p \sigma^p}\right)^{\frac{q}{p}}\right).$$

Remark 1.1.30. *We will prove below a stronger result than Proposition 1.1.28 when in presence of multiple tail parameters and pivots. In the setting of Proposition 1.1.27, we will show that for any $\lambda > 0$:*

$$\mathbb{E}\left[|Z_1 \cdots Z_p - a_1 \cdots a_p|^m\right] \leq \left(\prod_{i=1}^{p} \left(|a_i| + \frac{\sigma_i}{\lambda}\right)\right)^m \left(\lambda^m \left(\frac{m}{q}\right)^{\frac{m}{q}} + \lambda^{mp} \left(\frac{mp}{q}\right)^{\frac{mp}{q}}\right). \tag{5}$$

Then when the pivots $a_1, \dots, a_p$ are equal to a and the tail parameters $\sigma_1, \dots, \sigma_p$ are equal to σ, one can retrieve the result of Proposition 1.1.28 choosing $\lambda = 1$ when $a \leq \sigma$ and $\lambda = \frac{\sigma}{a}$ when $a \geq \sigma$.

Proof. Placing ourselves in the general setting of Proposition 1.1.27, we consider a parameter $\lambda > 0$ and aim at proving (5). To employ Lemma 1.1.26, let us introduce the polynomial

$$\begin{aligned} P(X_1, \dots, X_p) &= (\sigma_1 X_1 + |a_1|) \cdots (\sigma_p X_p + |a_p|) - |a_1 \cdots a_p| \\ &= |a_1 \cdots a_p| \sum_{l=1}^{p} \sum_{\substack{I \subset \{1,\dots,p\} \\ \#I = l}} \prod_{i \in I} \frac{\sigma_i X_i}{|a_i|} \end{aligned}$$

that has positive coefficients and verifies :

$$|Z_1 \cdots Z_p - a_1 \cdots a_p| \leq P\left(\frac{|Z_1 - a_1|}{\sigma_1}, \dots, \frac{|Z_p - a_p|}{\sigma_p}\right).$$

Given a random vector $Y \in \mathbb{R}_+$ with the cumulative distribution function $1 - Ce^{-\cdot^q}$, one knows from Lemma 1.1.26 that for any $m \in \mathbb{N}$:

$$\begin{aligned}\mathbb{E}\left[|Z_1 \cdots Z_p - a_1 \cdots a_p|^m\right] &\leq \mathbb{E}\left[P\left(Y, \ldots, Y\right)^m\right] \\ &\leq |a_1 \cdots a_p|^m \mathbb{E}\left[\left(\sum_{l=1}^{p} \sum_{\substack{I \subset \{1,\ldots,p\} \\ \#I = l}} \prod_{i \in I} \frac{\sigma_i}{|a_i|} Y^l\right)^m\right]\end{aligned}$$

since P^m has positive coefficients (since P does). Let us introduce for all $l \in \{1, \ldots, p\}$ the coefficient c_l that depends on λ and verifies :

$$c_l = \sum_{\substack{I \subset \{1,\ldots,p\} \\ \#I = l}} \prod_{i \in I} \frac{\sigma_i}{\lambda |a_i|}.$$

Decomposing the computation of the expectation between the expectation on the set of drawings where $\lambda Y \leq 1$ and $\lambda Y \geq 1$, one obtains :

$$\mathbb{E}\left[|Z_1 \cdots Z_p - a_1 \cdots a_p|^m\right] \leq \left(|a_1 \cdots a_p| \sum_{l=1}^{p} c_l\right)^m \left(\mathbb{E}\left[(\lambda Y)^m\right] + \mathbb{E}\left[(\lambda Y)^{pm}\right]\right).$$

We then easily retrieve (5) thanks to Proposition 1.1.20 since :

$$|a_1 \cdots a_p| \sum_{l=1}^{p} c_l = P\left(\frac{1}{\lambda}, \ldots, \frac{1}{\lambda}\right) = \left(\frac{\sigma_1}{\lambda} + |a_1|\right) \cdots \left(\frac{\sigma_p}{\lambda} + |a_p|\right) - |a_1 \cdots a_p|.$$

□

The previous development of q-exponential concentration primarily aims at providing a versatile and convenient "toolbox" (note that most introduced inequalities could have been enhanced, however to the expense of clarity) for the subsequent treatment of large dimensional random vectors, rather than random variables. This analysis will be performed through resorting to concentrated *functionals* of the random vectors, i.e., real images of through a mapping with different levels of regularity (linear, Lipschitz, convex..). For large dimensional vectors, one is mostly interested in the *order* of the concentration, thus the various constants appearing in most of the previous propositions and lemmas do not have any particular interest; of major interest instead is the independence of the concentration with respect to the random vector dimension, as observed for instance in Theorem 0.0.3 and that we will extend to other type of random vectors in what follows.

1.2. Concentration of a random vector of a normed vector space

In Section 2, the random vectors under study are either in $E = \mathbb{R}^p$ endowed with the Euclidean norm $\|z\| = \sqrt{\sum_{i=1}^p z_i^2}$ or the ℓ_1-norm $\|z\|_1 = \sum_{i=1}^p |z_i|$, or in $E = \mathcal{M}_{pn}$, $p, n \in \mathbb{N}$, endowed with two possible norms :

- the spectral norm defined as $\|M\| = \sup_{\|z\|\leq 1} \|Mz\|$,
- the Frobenius norm $\|M\|_F = \sqrt{\sum_{\substack{1\leq k\leq p\\ 1\leq i\leq n}} M_{k,i}^2}$

(where $M \in \mathcal{M}_{p,n}$). Note that the Frobenius norm can be seen as a Euclidean norm on $\mathbb{R}^{pn}$ with the bijection that concatenates the column of a matrix.

To generalize the notion of concentration to the case of a random vector of a normed vector space $(E, \|\cdot\|)$, one might be tempted to follow the idea of Definition 3 and say that a vector $Z \in E$ is α-concentrated if one has for instance $\mathbb{P}\left(\left\|Z - \tilde{Z}\right\| \geq t\right) \leq \alpha(t)$ for a deterministic vector $\tilde{Z}$, well chosen. This would describe a notion of a concentration *around a vector*.

However, this basic notion would not be compatible with the fundamental example of the uniform distribution on the sphere of radius $\sqrt{p}$ presented in Theorem 0.0.3 or the Gaussian vectors of identity population covariance matrices. When the dimension grows, those random vectors concentrate around a growing sphere which is the exact opposite behavior of being concentrated around a point. Yet, they present strong dimension free concentration properties that we will progressively identify through the presentation of three fundamental notions :

1. The *linear concentration* which is the concentration of $u(Z - \tilde{Z})$ for some deterministic vector $\tilde{Z} \in E$ (the so-called *deterministic equivalent*) and for any bounded linear form $u : E \mapsto \mathbb{R}$. For instance, we know from Theorem 0.0.3 that any random vector Z uniformly distributed on the sphere admits $\tilde{Z} = \mathbb{E}Z = 0$ as a deterministic equivalent. This means that most drawings of Z are close to the equator when the dimension grows.
2. The *Lipschitz concentration* which is the concentration of $f(Z) - f(Z')$ for any i.i.d. copy Z' of Z and any Lipschitz map $f : E \mapsto \mathbb{R}$.
3. The *convex concentration* which is the concentration of $f(Z) - f(Z')$ for any Lipschitz *and weakly convex* map $f : E \mapsto \mathbb{R}$. This notion is of course weaker than the Lipschitz concentration, its presentation here is only justified by the fundamental Theorem 1.2.47 owed to Talagrand that provides concentration properties on random vectors with independent and bounded entries. It is less "stable" than the Lipschitz concentration, meaning there exist very few transformations that preserve convex concentration. As a consequence one usually naturally returns to linear concentration after some refinement. For instance, supposing that X is convexly concentrated, only a linear concentration can be obtained for the resolvent $Q = (XX^T/n + zI_p)^{-1}$.

Although a seemingly basic notion, linear concentration still has quite interesting features that justify an independent treatment at the beginning of this section.

1.2.1. Linear Concentration

Considering the linear functionals of random vectors allows us, in particular, to introduce the notion of *deterministic equivalents*, which play the role of the

pivots we presented in the concentration of random variables.

Definition 8. *Given a random vector $Z \in E$, a deterministic vector $\tilde{Z} \in E$ and a concentration function α, we say that Z is* linearly α-concentrated *around the* deterministic equivalent $\tilde{Z}$ *if for any bounded linear form $u : E \to \mathbb{R}$ with a unit operator norm (i.e., $\forall z \in E,\ |u(z)| \leq \|z\|$) :*

$$u(Z) \in u(\tilde{Z}) \pm \alpha.$$

We note in that case : $Z \in \tilde{Z} \pm \alpha$ in $(E, \|\cdot\|)$.

Remark 1.2.1. *Definition 8 is clearly compatible with Definition 3 for the case where $E = \mathbb{R}$. Indeed, in $\mathbb{R}$ the linear forms are the scalar functions and for any coefficient $\lambda \in \mathbb{R}_*$ and any random variable $Z \in \mathbb{R}$ we have the equivalence :*

$$Z \in a \pm \alpha \qquad \Longleftrightarrow \qquad \lambda Z \in \lambda a \pm \alpha(\,\cdot\,/\lambda).$$

Remark 1.2.2. *The advantage of linear functionals is that they preserve the expectation which gives us the simplest deterministic equivalent. For instance if Z is linearly q-exponentially concentrated, then we know from Corollary 1.1.24 that $\mathbb{E}Z$ is a deterministic equivalent for Z since for any continuous linear form u, $\mathbb{E}[u(Z)] = u(\mathbb{E}[Z])$. For instance, thanks to Theorem 0.0.3, one knows that for any $p \in \mathbb{N}_*$, if $Z \sim \sigma_p$ then $Z \in 0 \pm 2e^{-\,\cdot\,^2/2}$ since $\mathbb{E}[Z] = 0$.*

Since the Frobenius norm is larger than the spectral norm, for any random matrix $M \in \mathcal{M}_{p,n}$, we have the implication :

$$M \in \tilde{M} \pm \alpha \ \text{ in } (\mathcal{M}_{p,n}, \|\cdot\|_F) \qquad \Longrightarrow \qquad M \in \tilde{M} \pm \alpha \ \text{ in } (\mathcal{M}_{p,n}, \|\cdot\|)$$

for some deterministic matrix $\tilde{M}$ and some concentration function α. When the choice of the norm is not ambiguous (when $E = \mathbb{R}^p$ in particular), we will allow ourselves not to specify the norm.

Remark 1.2.3. *The same way that linear forms in $\mathbb{R}^p$ are fully described with the scalar product, in $\mathcal{M}_{p,n}$ the linear forms are fully defined by the functions $f_A : M \mapsto \operatorname{Tr} AM$ for $A \in \mathcal{M}_{p,n}$ where f_A is said to have a unit norm if $\|f_A\|_* = 1$, i.e.,*

- *in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$: $\|A\|_F = \sqrt{\operatorname{Tr}(AA^T)} = 1$,*
- *in $(\mathcal{M}_{p,n}, \|\cdot\|)$: $\|A\|_1 = \operatorname{Tr}(AA^T)^{\frac{1}{2}} = 1$.*

Linear concentration is provided by classical concentration inequalities like Bernstein's or Hoeffding's inequalities.

Example 1.2.4. *[Bernstein's inequality, [Ver17, Theorem 2.8.2]] Given p independent random variables $(Z_i)_{1 \leq i \leq p} \in \mathbb{R}^p$, and three parameters $C \geq e$ and $c, q > 0$, we have the implication :*

$$\forall i \in \{1, \ldots, p\}, Z_i \in \mathbb{E}[Z_i] \pm Ce^{-\,\cdot\,/\sigma} \quad \Rightarrow \quad Z \in \mathbb{E}[Z] \pm 2e^{-c(\,\cdot\,/\sigma)^2} + 2Ce^{-c\,\cdot/\sigma},$$

where c is a numerical constant depending only on C.

Example 1.2.5. *[Hoeffding's inequality, [Ver17, Theorem 2.2.6]] Given p independent random variables $Z_1, \dots, Z_p \in [0,1]$, the random vector $Z = (Z_1, \dots, Z_p) \in \mathbb{R}^p$ is linearly concentrated and verifies $Z \in \mathbb{E}[Z] \pm 2e^{-2\cdot^2}$ in $(\mathbb{R}^p, \|\cdot\|)$.*

We will see in Subsection 1.2.3 about convex concentration a generalization of Hoeffding's theorem with the theorem of Talagrand.

Even if one does not suppose the independence between the random variables $Z_1, \dots, Z_p$, one can still get a concentration result on the random vector $Z = (Z_1, \dots, Z_p)$ in the case of q-exponential concentrations. The basic idea is that if $Z_1, \dots, Z_p$ are identically distributed then Z is at least as concentrated as the random vector $(Z_1, \dots, Z_1)$.

Proposition 1.2.6. *Given the parameters $C > e$, $q, \sigma > 0$, $p \in \mathbb{N}$, p vector spaces $(E_i, N_i)_{1 \leq i \leq p}$, p deterministic vectors $\tilde{Z}_1 \in E_1, \dots, \tilde{Z}_p \in E_p$ and p random vectors $Z_1 \in E_1, \dots, Z_p \in E_p$ (possibly dependent) satisfying, for any $i \in \{1, \dots, m\}$, $Z_i \in \tilde{Z}_i \pm Ce^{-(\cdot/\sigma)^q}$, we have the concentration :*

$$(Z_1, \dots, Z_p) \in (\tilde{Z}_1, \dots, \tilde{Z}_p) \pm e^{\frac{1}{e}} C \exp\left(-\frac{\bar{q}}{2e}\left(\frac{\cdot}{\sigma}\right)^q\right), \quad \text{in } (E, \|\cdot\|_\infty),$$

where we introduced on $E \equiv E_1 \times \cdots \times E_p$ the norm $\|\cdot\|_\infty$ verifying for any for $(z_1, \dots, z_p) \in E$:

$$\|(z_1, \dots, z_p)\|_\infty = \sup_{1 \leq i \leq p} N_i(z_i),$$

and we recall that $\bar{q} = \min(q, 1)$.

Proof. Let us consider a linear function $u : E \to \mathbb{R}$, such that

$$\|u\|_\infty \equiv \sup_{\|z\|_\infty \leq 1} |u(z)| \leq 1.$$

Given $i \in \{1, \dots, p\}$, let us note $u_i : E_i \to \mathbb{R}$ the function defined as $u_i(z) = u((0, \dots, 0, z, 0, \dots, 0))$ (where z is in the i^{th} entry). For any $z \in E$, one can write :

$$u(z) = \sum_{i=1}^{p} N_i(u_i) u_i'(z),$$

where $N_i(u_i) = \sup_{N_i(z) \leq 1} u_i(z)$ and $u_i' = u_i / N_i(u_i)$ ($N_i(u') = 1$). With the convenient notation $n_i \equiv N_i(u_i)$, we have the inequality :

$$\sum_{i=1}^{p} n_i = \sum_{i=1}^{p} n_i \sup_{N_i(z_i) \leq 1} u_i'(z_i) = \sup_{\|z\|_\infty \leq 1} u(z) \leq 1.$$

With this bound at hand, we plan to employ the characterization with the centered moments. Considering $m \in \mathbb{N}_*$, we introduce the polynomial $P(X_1, \dots, X_p) = (n_1 X_1 + \cdots + n_p X_p)^m$: it has positive coefficients which allows us to employ

Lemma 1.1.26 in order to bound :

$$\begin{aligned}\mathbb{E}\left[\left|u(Z)-u(\tilde{Z})\right|^m\right] &\leq \mathbb{E}\left[\left(\sum_{i=0}^{p} n_i \left|u_i'\left(Z_i\right)-u_i'\left(\tilde{Z}_i\right)\right|\right)^m\right] \\ &= \mathbb{E}\left[P\left(\left|u_1'\left(Z_1\right)-u_1'\left(\tilde{Z}_1\right)\right|,\ldots,\left|u_p'\left(Z_p\right)-u_p'\left(\tilde{Z}_p\right)\right|\right)\right] \\ &\leq \mathbb{E}\left[P\left(Y,\ldots,Y\right)\right] = \mathbb{E}\left[\left(\sum_{i=1}^{p} n_i Y\right)^m\right] \\ &\leq \mathbb{E}\left[Y^m\right] \ \leq\ C\left(\frac{m}{q}\right)^{\frac{m}{q}}\sigma^m,\end{aligned}$$

where $Y \geq 0$ is a nonnegative random variable having a cumulative distribution function equal to $1-\min(1, Ce^{-(\cdot/\sigma)^q})$ (such a variable controls $Z_1,\ldots,Z_p$). We can then conclude thanks to Lemma 1.1.22 (and Proposition 1.1.20). □

Given p random variables verifying $Z_i \in a_i \pm Ce^{-(\cdot/\sigma_i)^q}$, $a_i \in \mathbb{R}$ and $\sigma_i > 0$, we can apply Proposition 1.2.6 to the functional $u : z \mapsto \sum_{i=1}^p \sigma_i z_i$ taken on the p random variables $Z_i' = \frac{Z_i}{\sigma_i} \in \frac{a_i}{\sigma_i} \pm Ce^{-\cdot^q}$:

$$\sum_{i=1}^{p} Z_i = u(Z_1',\ldots,Z_p') \in \sum_{i=1}^{p} u_i a_i \pm e^{\frac{1}{e}} C \exp\left(-\frac{\bar{q}}{2e}\left(\frac{\cdot}{\sigma_1+\cdots+\sigma_p}\right)^q\right),$$

since $\|u\|_\infty = \sum_{i=1}^p \sigma_i$. One can recognize here the concentration given by Proposition 1.1.27. The following corollary generalizes this result.

Corollary 1.2.7. *Given $p \in \mathbb{N}$, two parameters $q > 0$ and $C \geq e$, and p random vectors $Z_1,\ldots Z_p \in E$, respectively concentrated around the deterministic equivalents $\tilde{Z}_1,\ldots,\tilde{Z}_p \in \mathbb{R}^p$ such that $\forall i \in \{1,\ldots,p\}$, $Z_i \in \tilde{Z}_i \pm Ce^{-(\cdot/\sigma_i)^q}$ for some parameters $\sigma_1,\ldots,\sigma_p > 0$:*

$$Z_1+\cdots+Z_p \in \tilde{Z}_1+\cdots+\tilde{Z}_p \pm e^{\frac{1}{e}} C \exp\left(-\frac{\bar{q}}{2e}\left(\frac{\cdot}{\sigma}\right)^q\right),$$

where $\sigma = \sigma_1+\cdots+\sigma_p$.

Dealing with q-exponential concentrations, we can get an analogous result to Corollary 1.1.17 that allowed us to interchange the pivot a with any other real in a ball around a with a diameter of the same order as the tail parameter. The tail parameter will be called in the case of a random vector an *observable diameter*. Given a random vector $Z \in E$, we will say that Z is *q-exponentially concentrated* around $\tilde{Z}$ with a *head parameter* C and an *observable diameter* σ if $Z \in \tilde{Z} \pm Ce^{-(\cdot/\sigma)^q}$. The *observable diameter* is the diameter of the "observations" of the distribution that could be seen as linear projections (for other types of concentrations to be introduced subsequently, they will alternatively be related to 1-Lipschitz or quasi-convex maps) on $\mathbb{R}$. Intuitively, it is the diameter of the observation in the "real world" ; refer to [Gro99] for a both more precise and more general definition.

Lemma 1.2.8. *Let us consider a random vector $Z \in E$, two deterministic vectors $\tilde{Z}, \tilde{Z}' \in E$ and three parameters $C \geq e$, $\lambda, \sigma > 0$. If $\|\tilde{Z} - \tilde{Z}'\| \leq \lambda\sigma$ then we have the implication :*

$$Z \in \tilde{Z} \pm Ce^{-(\,\cdot\,/\sigma)^q} \qquad \Longrightarrow \qquad Z \in \tilde{Z}' \pm \max(C, e^{\lambda^q})e^{-(\,\cdot\,/\sigma)^q}.$$

One might expect for concentrated random vectors a similar proposition to Proposition 1.1.9 giving the concentration of a product of random variables. In the case of vectorial objects, we are looking for the concentration of a scalar product of two random vectors $X, Y \in E$ or, closer to our present interest as announced in the preamble, for the concentration of $u^T Q x$ where u is deterministic, Q is a concentrated random matrix and x a concentrated random vector. If one looks closely at the proof of Proposition 1.1.9, it becomes obvious that although some steps can be fully adapted, the method gets stuck when trying to bound :

$$\mathbb{P}\left(|u\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right)| \geq t\right).$$

It is tempting here to invoke the Cauchy-Schwartz inequality :

$$u\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) \leq \|u\| \, \|X - \mathbb{E}X\| \, \|Y - \mathbb{E}Y\|,$$

where $\|u\|$ is the operator norm of u. However, as we explained it with the example of spherical and Gaussian vectors in the introduction of this subsection, unlike $u(X - \mathbb{E}X)$ and $u(Y - \mathbb{E}Y)$ the quantities $\|X - \mathbb{E}X\|$ and $\|Y - \mathbb{E}Y\|$ are far from being concentrated for concentrated random vectors X and Y of practical use. We will see in Proposition 1.2.10 that it is still possible to express the concentration of the norms and obtain consequently loose bounds for the concentration of $(X - \mathbb{E}X)(Y - \mathbb{E}Y)$ as presented in Examples 1.2.16 and 1.2.17.

Before that, to present a setting where the concentration is satisfactory, we suppose in addition to the concentration that the two vectors are independent and one is bounded.

Proposition 1.2.9. *Let us consider two normed vector spaces $(E_1, \|\cdot\|_1)$ and $(E_2, \|\cdot\|_2)$, two* independent *random vectors $Z_1 \in E_1$ and $Z_2 \in E_2$ and a bilinear form $f : E_1 \times E_2 \to \mathbb{R}$ such that for any $(z_1, z_2) \in E_1 \times E_2$:*

$$|f(z_1, z_2)| \leq \|z_1\|_1 \, \|z_2\|_2 .$$

If there exist two concentration functions $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$ and two deterministic vectors $(\tilde{Z}_1, \tilde{Z}_2) \in E_1 \times E_2$ such that :

$$Z_1 \in \tilde{Z}_1 \pm \alpha \text{ in } (E_1, \|\cdot\|_1) \qquad \text{and} \qquad Z_2 \in \tilde{Z}_2 \pm \beta \text{ in } (E_2, \|\cdot\|_2),$$

and if $\|Z_2\|_2$ is bounded by a real $K_2 > 0$, then :

$$f(Z_1, Z_2) \in f\left(\tilde{Z}_1, \tilde{Z}_2\right) \pm \alpha\left(\frac{\cdot}{2K_2}\right) + \beta\left(\frac{\cdot}{2\|\tilde{Z}_1\|_1}\right).$$

Proof. Given $t > 0$, we just compute :

$$\begin{aligned}\mathbb{P}\left(\left|f(Z_1,Z_2)-f(\tilde{Z}_1,\tilde{Z}_2)\right|\geq t\right)&\\ \leq \mathbb{E}\left[\mathbb{P}\left(\left|f(Z_1-\tilde{Z}_1,Z_2)\right|\geq \frac{t}{2}\mid Z_2\right)\right]+\mathbb{P}\left(\left|f(\tilde{Z}_1,Z_2-\tilde{Z}_2)\right|\geq\frac{t}{2}\right)&\\ \leq \alpha\left(\frac{t}{2K_2}\right)+\beta\left(\frac{t}{2\tilde{Z}_1\|_1}\right)&\end{aligned}$$

since $x \mapsto f(x, Z_2)$ and $y \mapsto f(\tilde{Z}_1, y)$ are both linear and respectively K_2-Lipschitz and $\|\tilde{Z}_1\|$-Lipschitz. □

If a random vector Z is linearly concentrated around a deterministic equivalent $\tilde{Z}$, it is possible to control the norm $\|Z-\tilde{Z}\|$ if the norm $\|\cdot\|$ can be defined as the supremum on a set of linear forms. For instance, in $\mathbb{R}^p$ endowed with the sup norm $\|\cdot\|_\infty$ ($\|x\|_\infty = \sup\{|x_i|, 1 \leq i \leq p\}$), if $Z \in \tilde{Z} \pm \alpha$:

$$\mathbb{P}\left(\|Z-\tilde{Z}\|_\infty \geq t\right) = \mathbb{P}\left(\sup_{1\leq i\leq p} e_i^T(Z-\tilde{Z})\geq t\right) \tag{6}$$

$$\leq p \sup_{1\leq i\leq p}\mathbb{P}\left(e_i^T(Z-\tilde{Z})\geq t\right)\leq p\alpha(t), \tag{7}$$

where $(e_1, \ldots, e_p)$ is the canonical basis of $\mathbb{R}^p$ (e_i is a vector full of 0 with 1 on the i^{th} coordinate). To manage the infinity norm, the supremum is taken on a finite set $\{e_1, \ldots e_p\}$; things are more complex if we look at the Euclidean norm because this time one comes to use the identity $\|x\| = \sup\{u^T x, \|u\| \leq 1\}$ where the supremum is taken on the unit ball. To tackle this loss of cardinality, it is convenient here to introduce the so-called ϵ-nets to discretize the ball in order to approach sufficiently the norm and at the same time find a good bound for the probability (see [Tao11]). We leave the proof in the appendix.

Proposition 1.2.10. *Let us consider a normed vector space $(E, \|\cdot\|)$ of finite dimension such that there exists a subspace H of the dual space $(E^*, \|\cdot\|_*)$, and a ball $\mathcal{B}_H = \{f \in H, \|f\|_* \leq 1\} \subset H$ verifying for any $z \in E$:*

$$\|z\| = \sup_{f\in\mathcal{B}_H} f(z). \tag{8}$$

Given a random vector $Z \in E$, a deterministic equivalent $\tilde{Z}$ and a concentration function α, if we suppose that $Z \in \tilde{Z} \pm \alpha$ then we have the concentration :

$$\left\|Z-\tilde{Z}\right\| \in 0 \pm 8^{\dim(H)}\alpha\left(\frac{\cdot}{2}\right) \tag{9}$$

where $\dim(H)$ is the dimension of H.

Remark 1.2.11. *One always has $\|z\| = \sup_{\|f\|_*\leq 1} f(z)$, so one might be tempted to systematically consider $H = E^*$. For instance in $(\mathbb{R}^p, \|\cdot\|)$, H is taken to be equal to $\mathbb{R}^p$ and $\dim(H) = p$. The same way, in the Euclidean space*

$(\mathcal{M}_{p,n}, \|\cdot\|_F)$, *we take* $H = \mathcal{M}_{p,n}$ *(and* $\dim(H) = pn$*). However, we can reduce greatly the dimension* $\dim(H)$ *in the case of* $(\mathcal{M}_{p,n}, \|\cdot\|)$ *since for any* $M \in \mathcal{M}_{p,n}$ *:*

$$\|M\| = \sup_{\|u\|,\|v\|\leq 1} \left|u^T M v\right|.$$

Thus it is wise to consider $H = \{f_{vu^T}, u, v \in \mathbb{R}^p \times \mathbb{R}^n\}$ *(see Remark 1.2.3 for a definition of* f_A *; here,* $\|f_{vu^T}\|_* = \left\|vu^T\right\|_1 = \|u\|\,\|v\|$*) and the dimension is then only* $\dim(H) = p + n$*.*

Taking into account the two results (6) and (9), we are led to introduce an indicator characteristic to the norm that gives the speed of the concentration.

Definition 9 (Norm degree)**.** *Given a normed vector space* $(E, \|\cdot\|)$*, and a subset* $H \subset E^*$*, let us define the degree* η_H *of* H *as :*

- $\eta_H = \log(\#H)$ *if* H *is finite*
- $\eta_H = \dim(\operatorname{Vect} H)$ *if* H *is infinite*

where $\#H$ *is the number of elements in* H *and* $\operatorname{Vect} H$ *is the subspace of* E *generated by* E*. We note then* $\eta(E, \|\cdot\|)$ *or more simply* $\eta_{\|\cdot\|}$ *the degree of* $\|\cdot\|$ *that is defined as :*

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

In the setting of the last proposition :

$$\left\|Z - \tilde{Z}\right\| \in 0 \pm e^{c\eta(E,\|\cdot\|)}\alpha\left(\frac{\cdot}{2}\right), \tag{10}$$

where c is a numerical constant. In the case of the q-exponential concentration, it is possible to rearrange the concentration of $\|Z - \tilde{Z}\|$ to obtain a head parameter of order 1.

Proposition 1.2.12. *Given a random vector* $Z \in E$*, a deterministic vector* $\tilde{Z} \in \mathbb{R}^p$ *and three parameters* $C \geq e$*,* $c, \sigma > 0$*, we have the implication :*

$$Z \in \tilde{Z} \pm Ce^{-(\cdot/\sigma)^q} \qquad \Longrightarrow \qquad \|Z - \tilde{Z}\| \in 0 \pm Ce^{-(\cdot/2\sigma)^q/2c\eta_{\|\cdot\|}},$$

where c *is the same numerical constant as in* (10)*.*

Reciprocally :

$$\|Z - \tilde{Z}\| \in 0 \pm Ce^{-(\cdot/\sigma)^q} \qquad \Longrightarrow \qquad Z \in \tilde{Z} \pm Ce^{-(\cdot/\sigma)^q}.$$

The second result of the proposition is trivial (and quite useless) and the first result is just a simple consequence of (10) combined with Lemma 1.1.15.

Example 1.2.13. *We can give some examples of norm degrees :*

- $\eta\left(\mathbb{R}^p, \|\cdot\|_\infty\right) = \log(p)$

- $\eta\left(\mathbb{R}^p, \|\cdot\|_r\right) = p$ *for* $r \geq 1$
- $\eta\left(\mathcal{M}_{p,n}, \|\cdot\|\right) = n + p$
- $\eta\left(\mathcal{M}_{p,n}, \|\cdot\|_F\right) = np$.

In the particular case of q-exponential concentrations, the norm degree allows us to bound the expectation of the norm of $Z - \tilde{Z}$ thanks to Proposition 1.2.12.

Corollary 1.2.14. *Given a random vector* $Z \in E$*, if we suppose that* $Z \in \tilde{Z} \pm Ce^{-(\cdot/\sigma)^q}$ *and* $q \geq 1$*, we can bound :*

$$\mathbb{E}\left\|Z - \tilde{Z}\right\| \leq C'\sigma\eta_{\|\cdot\|}^{1/q}$$

where C' *is a constant depending on* C*.*

Example 1.2.15. *Let* $Z \in \mathbb{R}^p$ *and* $M \in \mathcal{M}_{p,n}$ *be two random vectors. Then,*

- *if* $Z \in \tilde{Z} \pm 2e^{-t^2/2}$ *in* $(\mathbb{R}^p, \|\cdot\|)$ *:* $\mathbb{E}\|Z\| \leq \|\tilde{Z}\| + C\sqrt{p}$
- *if* $M \in \tilde{M} \pm 2e^{-t^2/2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|)$ *:* $\mathbb{E}\|M\| \leq \|\tilde{M}\| + C\sqrt{p+n}$,
- *if* $M \in \tilde{M} \pm 2e^{-t^2/2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ *:* $\mathbb{E}\|M\| \leq \|\tilde{M}\| + C\sqrt{pn}$.

Now that we can control the quantities $\|Z - \tilde{Z}\|$ when $\tilde{Z}$ is a deterministic equivalent of Z, it is tempting to extend Lemma 1.1.8 to the concentration of the product of any linearly concentrated random vectors. The examples presented below are just here to give an idea of what could be obtained, they are not relevant in practice since the bounds are too loose. Unlike Lipschitz concentration as it will be presented in next subsection, linear concentration is not suited to study the concentration of the product of random vectors.

Example 1.2.16. *Let us note* $\odot$ *the product in* $\mathbb{R}^p$ *verifying for any* $x = (x_1, \ldots, x_p)$ *and* $y = (y_1, \ldots, y_p)$*,* $x \odot y = (x_1y_1, \ldots, x_py_p)$*. It gives an algebra structure to* $\mathbb{R}^p$ *where* $\|\cdot\|_\infty$ *and* $\|\cdot\|_2$ *are both algebra norms (*$\|x \odot y\|_2 \leq \|x \odot y\|_1 = \sum_{i=1}^p |x_iy_i| \leq \|x\|_2\|y\|_2$ *thanks to the Cauchy Schwarz inequality). Therefore we have for any vector* $Z \in \tilde{Z} \pm 2e^{-\cdot^2/2}$ *in* $(\mathbb{R}^p, \|\cdot\|_2)$ *:*

- $\frac{Z^{\odot 2}}{p} = \frac{Z \odot Z}{p} \in \frac{\tilde{Z}^{\odot 2}}{p} \pm Ce^{-c\cdot} + Ce^{-c(\sqrt{p}\cdot/\|\tilde{Z}\|)^2}$ *in* $(\mathbb{R}^p, \|\cdot\|_2)$
- $\frac{Z^{\odot 2}}{\log p} \in \frac{\tilde{Z}^{\odot 2}}{\log p} \pm Ce^{-c\cdot} + Ce^{-c(\sqrt{\log p}\cdot/\|\tilde{Z}\|_\infty)^2}$ *in* $(\mathbb{R}^p, \|\cdot\|_\infty)$,

where $C \geq e$ *and* $c > 0$ *are two numerical constants.*

Example 1.2.17 (Concentration of the sample covariance)**.** *Given a matrix* $X \in \mathcal{M}_{p,n}$*, and three parameters* $C, q \geq 1$ *and* $c > 0$*, if we suppose that* $X \in \tilde{X} \pm 2e^{-\cdot^2/2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_F)$*, then :*

- $\frac{XX^T}{n^2} \propto Ce^{-c\cdot/\gamma} + Ce^{-c(n\cdot/\|\tilde{X}\|_F)^2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_F)$
- $\frac{XX^T}{n} \propto Ce^{-c\cdot/\bar{\gamma}} + Ce^{-c(n\cdot/\|\tilde{X}\|)^2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|)$
- $\frac{XX^T}{\log n} \propto C\exp\left(-\frac{c\cdot}{1+\frac{\log p}{\log n}}\right) + Ce^{-c(\sqrt{\log n}\cdot/\|\tilde{X}\|_\infty)^2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_\infty)$,

where $\gamma = \frac{p}{n}$*,* $\bar{\gamma} = \gamma + 1 \geq \max(\gamma, 1)$ *and* $C \geq e$*,* $c > 0$ *are two numerical constants.*

As rich it could be the notion of linear concentration is insufficient when dealing with the resolvent of random matrices, starting with the resolvent of the sample covariance $Q = (XX^T/n + I_p)^{-1}$. It is possible to infer the concentration of the sample covariance in $(\mathcal{M}_p, \|\cdot\|)$ from the concentration of X as we saw in Example 1.2.40, we can even track the concentration of the resolvent since the inverse of a matrix of $\mathcal{M}_p$ can be written as a polynomial of degree p, but the observable diameter will then be of diverging order. To solve these issues, one needs a stronger notion of concentration to be introduced next : the *Lipschitz* concentration.

1.2.2. Lipschitz Concentration

Theorem 0.0.3 given in the preamble provides us with the concentration of the Lipschitz and even uniformly continuous functionals of random vectors uniformly distributed on the sphere, that is to say, cases that go far beyond the linear case and that let us hope for interesting inference on the resolvent of random matrices

As firstly evoked in Lemma 1.1.3, the concentration of a random vector Z can be expressed through the concentration of any random variable $f(Z) \in \mathbb{R}$ when f is Lipschitz. As before, this approach of concentration has the asset of bringing back the concentration on any normed vector space to a mere concentration on $\mathbb{R}$, that we deeply studied at the beginning of the section. The following definition allows us to generalize the notion of concentration to any metric space as presented in [Led01].

Definition 10 (Lipschitz Concentration of a random vector)**.** *Given a concentration function α, a random vector Z is said to be Lipschitz α-concentrated iff one of the following three assertions is verified for any* 1*-Lipschitz function $f : E \to \mathbb{R}$:*

- $f(Z) \propto \alpha$, *and we will note in that case* $Z \propto \alpha$
- $f(Z) \in m_f \pm \alpha$, *and we will note in that case* $Z \overset{m}{\propto} \alpha$
- $f(Z) \in \mathbb{E}[f(Z)] \pm \alpha$, *and we will note in that case* $Z \overset{\mathbb{E}}{\propto} \alpha$,

where m_f is a median of $f(X)$.

The Lipschitz concentration is the strongest notion of concentration we will present and it is the one that received most of the interest from the scientific community; therefore, we allow ourselves to omit the term "Lipschitz" when mentioning this kind of concentration. In our paper, such Lipschitz concentration of random vectors can only be obtained through Theorem 0.0.3 or Theorem 1.2.20 below, and they are both set on the normed vector space $(\mathbb{R}^p, \|\cdot\|)$ (or the analogous one $(\mathcal{M}_{p,n}, \|\cdot\|_F)$. We will thus allow ourselves to omit the precision about the normed vector space on which is made the concentration when we are on $(\mathbb{R}^p, \|\cdot\|)$, on $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ or when we are on the formal normed vector space $(E, \|\cdot\|)$.

Remark 1.2.18. *We know from Lemma 1.1.3 that Definition 10 is compatible with Definition 2 when $E = \mathbb{R}$, so there are no conflicts between the different uses of the notation $\propto$ for random vectors and random variables.*

Remark 1.2.19. *Given a random vector $Z \in E$ and a concentration function α, we know thanks to Lemma 1.1.2 :*

$$Z \propto \alpha \implies \qquad Z \overset{m}{\propto} 2\alpha \qquad \implies \qquad Z \propto 4\alpha(\cdot/2).$$

Also, by definition of the linear concentration :

$$Z \overset{\mathbb{E}}{\propto} \alpha \qquad \implies \qquad Z \in \mathbb{E}Z \pm \alpha,$$

and we can link this concentration to the other two thanks to Proposition 1.1.18 and Corollary 1.1.24 in the case of a q-exponential concentration :

$$Z \overset{m}{\propto} Ce^{(\cdot/\sigma)^q} \quad \implies \quad Z \overset{\mathbb{E}}{\propto} e^{C^q/q} e^{(\cdot/2\sigma)^q} \quad \implies \quad Z \overset{m}{\propto} 2e^{C^q/q} e^{(\cdot/4\sigma)^q},$$

for any $q, \sigma > 0$ and $C \geq e$.

We thought useful to present the theorem of concentration of Gaussian vectors in our new formalism, to add this setting to the historical example of the concentration on the sphere.

Theorem 1.2.20. *A canonical Gaussian vector Z is normally concentrated independently of its dimension. For any $p \in \mathbb{N}$:*

$$Z \sim \mathcal{N}(0, I_p) \qquad \implies \qquad Z \overset{m}{\propto} 2e^{-(\,\cdot\,/2)^2} \quad \textit{and} \quad Z \overset{\mathbb{E}}{\propto} 2e^{-(\,\cdot\,/2)^2}$$

where $\mathcal{N}(0, I_p)$ is the distribution of the canonical Gaussian vectors of dimension *p that have independent zero mean and unit variance Gaussian entries.*

A structural proof with a geometrical approach from the Poincaré lemma tracks the concentration of Gaussian vectors from the concentration of the uniform distribution on the sphere. A more functional approach based on the log-Sobolev inequalities can be found in [Led01]. An alternative proof is found in [Tao11], originally proposed by Maurey and Pisier, which does not provide the optimal constants but is more efficient and simple.

Theorem 1.2.20 can be generalized to any random vector $X \in \mathbb{R}^p$ with density $d\mathbb{P}_X(x) = e^{-U(x)} d\lambda_p(x)$ where $U : \mathbb{R}^p \to \mathbb{R}$ is a positive functional with the hessian bounded inferiorly by, say cI_p, $c > 0$. In that case, $X \propto 2e^{-c\cdot^2/2}$, (see [Led01, Theorem 2.7])

Let us add for the general picture a result from Talagrand [Tal94, Theorem 2.4] (or [Led01, Proposition 4.18]) concerning the concentration of the exponential distribution, that we shall denote ν^p, which is the distribution of random vectors of $\mathbb{R}^p$ with independent entries having density $\frac{1}{2}e^{-|\cdot|} d\lambda_1$.

Theorem 1.2.21. *[Led01, Proposition 4.18] There exist two numerical constants $C \geq 1$ and $c > 0$, such that for any $p \in \mathbb{N}$:*

$$Z \sim \nu^p \qquad \implies \qquad Z \propto Ce^{-c\,\cdot}$$

As an example of q-exponential concentrations when $q \in [1,2]$, one may consider vectors uniformly distributed on the balls of $\mathbb{R}^p$, that is $\mathcal{B}_{\|\cdot\|_q} = \{x \in \mathbb{R}^p \mid \|x\|_q = (\sum x_i^q)^{1/q} \leq 1\}$; let us note β_q^p this distribution.

Theorem 1.2.22. *[Led01, Proposition 4.21] Given $q \in [1,2]$, there exist two numerical constants $C \geq 1$ and $c > 0$, such that for any $p \in \mathbb{N}$:*

$$Z \sim \beta_q^p \qquad \Longrightarrow \qquad Z \propto Ce^{-c\,.\,q}$$

Remark 1.2.23. *The independence of the concentration of a random vector to its dimension can be interpreted as a conservation of its* observable diameter *through dimensionality when the* actual diameter *increases. This second diameter, that can be referred to as the* metric diameter*, can be naturally defined as the expectation of the distance between two independent random vectors drawn from the same distribution. Theorem 1.2.20 states that the observable diameter of a Gaussian distribution in $\mathbb{R}^p$ is of order 1, that is to say $\frac{1}{\sqrt{p}}$ times less than the diameter (that is of order $\sqrt{p}$). The same result holds for the uniform distribution on the sphere of $\mathbb{R}^p$ and for any distribution that would be called for that reason* concentrated*.*

Definition 10 only presents the concentration of Lipschitz functionals of Z (if f is λ-Lipschitz then f/λ is 1-Lipschitz and the product with a constant is easy to manage, we find $f(Z) \propto \alpha(\cdot\,/\lambda)$), but it is possible to show the concentration of any uniformly continuous functional of Z :

Proposition 1.2.24 (Concentration of the uniformly continuous transformations)**.** *Given two normed vector spaces E and G, a random vector $Z \in E$, a concentration function α, a modulus of continuity ω, a function $\phi : E \to G$, ω-continuous, we have the implication :*

$$Z \overset{m}{\propto} \alpha \qquad \Longrightarrow \qquad \phi(Z) \overset{m}{\propto} \alpha(\omega^{-1}(\,\cdot\,)).$$

Proof. Let us introduce a 1-Lipschitz function $g : G \to \mathbb{R}$, we note $f = g \circ \phi$. We introduce m_f, a median of $f(Z)$ and the sets $A_- = \{z, f(z) \leq m_f\}$ and $A_+ = \{z, f(z) \geq m_f\}$ (they verify by definition $\mathbb{P}(Z \in A_+), \mathbb{P}(Z \in A_-) \geq \frac{1}{2}$). The image through f of the boundary $\partial A_+ = \partial A_- = A_+ \cap A_-$ is equal to $\{m_f\}$. Since the boundary is closed, for any $z \in A_+$, there exists a sequence $z_n \in \partial A_-$ such that $\|z - z_n\| \underset{n\to\infty}{\longrightarrow} d(z, \partial A_-) = d(z, A_-)$, then since f is uniformly continuous like ϕ, we can bound :

$$\begin{aligned} |f(z) - m_f| &= \lim_{n\to\infty} |f(z) - f(z_n)| \leq \lim_{n\to\infty} \omega(\|z - z_n\|) \\ &\leq \omega(d(z, A_-)) = \omega(|d(z, A_-) - d(z, A_+)|), \end{aligned}$$

and the same inequality is also verified for any $z \in A_-$. This entails :

$$\mathbb{P}\left(|f(Z) - m_f| \geq t\right) \leq \mathbb{P}\left(|d(Z, A_-) - d(Z, A_+)| \geq \omega^{-1}\left(t\right)\right),$$

and we can then conclude since $\tilde{f} : z \mapsto d(z, A_-) - d(z, A_+)$ is 1-Lipschitz and $\tilde{f}(Z)$ admits 0 as a median. □

Remark 1.2.25. *Theorems 0.0.3 and 1.2.20 combined with Proposition 1.2.24 give us immediately a q-exponential concentration of all the random vectors $F(Z)$ where $F : \mathbb{R}^p \to \mathbb{R}^d$ is uniformly continuous and $Z \sim \sigma_p$ or $Z \sim \mathcal{N}(0, I_p)$. This describes a wide range of random vectors that remarkably do not need to present independent entries.*

As for Corollary 1.2.7, in order to show the concentration of the sum $X + Y$ of two random vectors $X, Y \in E$, we first need to prove the concentration of the concatenation $(X, Y) \in E^2$. In Proposition 1.2.6, we saw how to show this result when one supposes that X and Y have q-exponential concentration. Interestingly enough, an assumption of independence was not necessary (we saw that if X and Y are identically distributed then (X, Y) is at least concentrated as (X, X) up to a small change of the tail parameter). In the case of the Lipschitz concentration, the independence between X and Y plays an important role but still not essential as we will see in Theorem 1.2.35. We first present a way to infer the concentration of (X, Y) when X and Y are independent and we don't make further assumptions on their concentration (the particular case of the q-exponential concentration is studied in Appendix B.4).

To get interesting inferences from the concentration of the measure theory, one has to base its initial inequalities on a theorem of the kind of Theorem 0.0.3, 1.2.20 or 1.2.21 providing distributions with an observable diameter far smaller than the metric parameter (see Remark 1.2.23 for precisions). The following proposition is just a tool that can be some help when one wants to study a *limited* (with regard to the dimension) concatenation of concentrated vectors. In what follows E and F are two normed vector spaces respectively equipped with the norms $\|\cdot\|_E$ and $\|\cdot\|_F$. We again note $\|\cdot\|_{\ell_1}$ the norm of $E \times F$ defined as $\|(x, y)\|_{\ell_1} = \|x\|_E + \|y\|_F$.

Proposition 1.2.26. *Given two independent random vectors $X \in E$ and $Y \in F$, if we suppose that X and Y are concentrated then (X, Y) is also concentrated. Given two concentration functions $\alpha, \beta : \mathbb{R}_+ \mapsto \mathbb{R}_+$, and any $\lambda \in (0, 1)$:*

$$\begin{cases} X \propto \alpha \\ Y \propto \beta \end{cases} \qquad \Longrightarrow \qquad (X, Y) \propto \alpha\left(\lambda\,\cdot\right) + \beta\left((1-\lambda)\,\cdot\right),$$

If we suppose that α is invertible and piecewise differentiable, we also have the implication :

$$\begin{cases} X \propto \alpha \\ Y \propto \beta \end{cases} \qquad \Longrightarrow \qquad (X, Y) \propto \alpha + \beta - \alpha' * \beta,$$

where $*$ is the convolution operator ($f * g(t) = \int_{\mathbb{R}} f(u)g(t-u)du$). Since α and β are only defined on $\mathbb{R}_+$, we implicitly compute the convolution with a null continuation of α and β on $\mathbb{R}_-$. The implications are also true if we work with concentrations around the medians ($X \overset{m}{\propto} \alpha$) or around the means ($X \overset{\mathbb{E}}{\propto} \alpha$).

In the second result, be careful that, since the concentration functions are decreasing, α' is a negative function.

Proof. For the proof of the first implication, refer to [Led01, Proposition 1.11]. Let us consider a 1-Lipschitz function $f : E \times F \to \mathbb{R}$. We will work with the concentration around the means since it is easier and we will note for simplicity $\mathbb{E}f = \mathbb{E}[f(X,Y)]$ and $\mathbb{E}[f|Y] = \mathbb{E}[f(X,Y) \mid Y]$; we can bound :

$$\begin{aligned}
\mathbb{P}(|f(X,Y) - \mathbb{E}f| \geq t) &\leq \mathbb{E}\left[\mathbb{P}\left(|f(X,Y) - \mathbb{E}[f|Y]| \geq t - |\mathbb{E}f - \mathbb{E}[f|Y]| \mid Y\right)\right] \\
&\leq \mathbb{E}\left[\alpha\left(t - |\mathbb{E}f - \mathbb{E}[f|Y]|\right)\right] + \mathbb{P}\left(|\mathbb{E}f - \mathbb{E}[f|Y]| \geq t\right) \\
&\leq \int_0^1 \mathbb{P}\left(\alpha\left(t - |\mathbb{E}f - \mathbb{E}[f|Y]|\right) \geq u\right) du + \beta(t) \\
&\leq \int_0^1 \mathbb{P}\left(t - |\mathbb{E}f - \mathbb{E}[f|Y]| \leq \alpha^{-1}(u)\right) du + \beta(t) \\
&\leq \int_{\alpha(t)}^1 \beta\left(t - \alpha^{-1}(u)\right) du + \alpha(t) + \beta(t) \\
&= \int_t^0 \alpha'(u)\beta(t-u)\, du + \alpha(t) + \beta(t).
\end{aligned}$$

□

In the first result of Proposition 1.2.26, λ is often chosen to be equal to $\frac{1}{2}$ as in Lemma 1.1.5 for instance. However, we allow ourselves to choose small values of λ when we want to maintain almost untouched a concentration as will be done in the following useful example that allows to transmit the absolute continuity of the Gaussian distribution to any random vector, while, at the same time, preserving the concentration (see Theorem 1.2.35 for an effective use).

Example 1.2.27. *Given an integer $p > 0$, a concentration function α, an α-concentrated random vector $X \in \mathbb{R}^p$, and an independent random vector $U \in \mathbb{R}^p$ verifying $\frac{1}{\sqrt{p-2}}U \sim \sigma_{p-1}$ or $U \sim \mathcal{N}(0, I_p)$, we have the concentration :*

$$X + \frac{1}{n}U \propto \alpha\left(\left(1 + \frac{1}{\sqrt{n}}\right)\cdot\right) + 2e^{-n\cdot^2/2}.$$

Remark 1.2.28. *The two results of Proposition 1.2.26 can be compared if we consider a generalization of the notion of observable diameters, as defined for the exponential concentration before Lemma 1.2.8, and that is slightly different from the observable diameter introduced by Gromov in [Gro99, Chapter 3.1/2]. Still, it is of the same order when the dimension is large. Given a concentration function α, we note $\mathcal{R}_\alpha = \int_0^\infty \alpha$, and for any random vector X the observable diameter $\mathcal{R}_X$ is defined as $\mathcal{R}_X = \inf\{\mathcal{R}_\alpha \mid X \propto \alpha\}$. Our definition comes from the fact that if, say, $X \propto \alpha$, then for any 1-Lipschitz function f and any independent copy X' :*

$$\mathbb{E}[|f(X) - f(X')|] = \int_0^\infty \mathbb{P}(|f(X) - f(X')| \geq t)\, dt \leq \int_0^\infty \alpha = \mathcal{R}_\alpha.$$

With the first result given by Proposition 1.2.26, we find :

$$\mathcal{R}_{(X,Y)} \leq \frac{\mathcal{R}_X}{1-\lambda} + \frac{\mathcal{R}_Y}{\lambda},$$

and with the second result :

$$\mathcal{R}_{(X,Y)} \leq \mathcal{R}_X + 2\mathcal{R}_Y,$$

since for any differentiable concentration functions α, β :

$$\int_0^\infty \alpha' * \beta = \int_{-\infty}^{+\infty} \alpha' * \beta = \int_{-\infty}^{+\infty} \alpha' \int_{-\infty}^{+\infty} \beta = -\alpha(0)\mathcal{R}_\beta.$$

The second inequality is clearly better than the first one, however we are far from the stability of the observable diameter that we find in Theorems 0.0.3 and 1.2.20.

One can assert easily that the concentration could be generalized to non-independent random vectors since we know for instance that $(X, f(X))$ is concentrated as a 2-Lipschitz transformation of X when $f : E \to E$ is 1-Lipschitz. However, there exists a lot of examples where (X, Y) is far from being concentrated despite X and Y being concentrated.

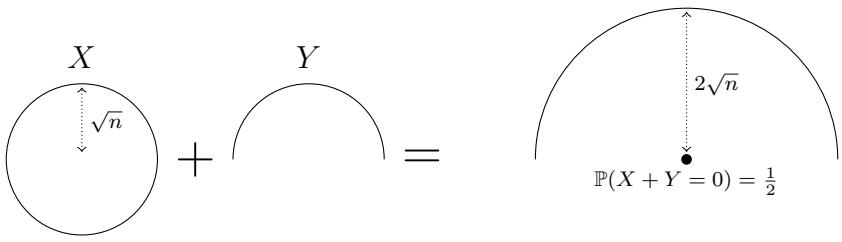

FIG 1. *The sum of two concentrated random vectors can be non concentrated*

Example 1.2.29. *Given $p \geq 0$, let us consider a random vector $Y \sim \sigma_p$. We define the random vector X as being equal to Y on $\sqrt{p}\,\mathbb{S}^p_+ = \sqrt{p}\,\mathbb{S}^p \cap (\mathbb{R}_+ \times \mathbb{R}^p)$ and equal to $-Y$ on $\sqrt{p}\,\mathbb{S}^p_- = \sqrt{p}\,\mathbb{S}^p \cap (\mathbb{R}_- \times \mathbb{R}^p)$. Note that $X \sim \sigma_p \sim 2\mathbb{1}_{\sqrt{p}\mathbb{S}^p_+}\sigma_p$ and Y are both $2e^{-\cdot^2/2}$-concentrated (see Theorem 0.0.3) and therefore $\mathcal{R}_X \leq R_Y \leq \mathcal{R}_Y \leq \int_0^\infty 2e^{-t^2/2}dt = \sqrt{2\pi}$. As we see on the schematic Figure 1, the distribution of $X+Y$ is completely different from the distribution of X and Y since it is discontinuous. If one looks at the variations of the random variable $\|X+Y\|$ that is a 1-Lipschitz functional of (X, Y), one notes that 0 is a median of $\|X+Y\|$ and we have :*

$$\mathbb{P}\left(\|X+Y\| \geq t\right) = \begin{cases} 1 \text{ if } t = 0 \\ \frac{1}{2} \text{ if } t \in \left(0, \sqrt{2p}\right] \\ 0 \text{ if } t > \sqrt{2p}. \end{cases}$$

Therefore, with the notation of the observable diameter we introduced in Remark 1.2.28, we see that :

$$\mathcal{R}_{(X,Y)} \geq \mathcal{R}_{\|(X+Y\|} = \int_0^\infty \mathbb{P}\left(\|X+Y\| \geq t\right) dt = \sqrt{\frac{p}{2}} > 4\sqrt{2\pi} \geq 2(\mathcal{R}_X + \mathcal{R}_Y)$$

For p sufficiently large, it contradicts the inequalities given in Remark 1.2.28 and we see in particular that the random vector (X,Y) has an observable diameter of the same order as the metric diameter which is of order $\sqrt{p}$. The vector (X,Y) is not concentrated in the sense given by Remark 1.2.23.

In this example, X was chosen in a way that its relationship with Y changes dramatically through the plane $\{0\}\times\mathbb{R}^p$: there is what we shall call an *unlimited defiance* of X towards Y. To be able to infer concentration properties on (X,Y) from the concentration of X and Y one needs to *limit* this defiance *under* a modulus of continuity in a way defined below.

Definition 11. *[Defiance between random vectors] Given two random variables $X \in E$ and $Y \in F$, and a continuity modulus $\omega : \mathbb{R}_+ \to \mathbb{R}_+$, one says that X defies Y under w iff for any 1-Lipschitz function $f : E \to \mathbb{R}$, any $t \in \mathbb{R}$ and any $y_1, y_2 \in F$:*

$$\mathbb{P}\left(f(X) \geq t \mid Y = y_1\right) \leq \mathbb{P}\left(f(X) \geq t - \omega(\|y_1 - y_2\|) \mid Y = y_2\right)$$

If there exist two parameters $\lambda > 0$ and $\nu \in (0,1]$ such that $\forall t > 0$, $\omega(t) = \lambda t^\nu$, one says that X (λ,ν)-Hölder defies Y and if $\nu = 1$, one says that X λ-Lipschitz defies Y.

We give some basic examples of the notion of defiance in the next lemma.

Lemma 1.2.30. *Let us consider four normed vector spaces E, F, G, H, three random vectors $X \in E$, $Y \in F$, $Z \in G$, two continuity modulus $\omega, \varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$; a ω-continuous transformation $\phi : F \to E$, a transformation $\psi : E \times G \to H$ that we suppose to be ω-continuous on the first variable (i.e. $\forall z \in G : x \mapsto \psi(x,z)$ is ω-continuous) :*

- *if Z and Y are independent then Z defies Y under 0*
- *$\phi(Y)$ defies Y under ω*
- *if X defies Y under ε then $\psi(X,Z)$ defies Y under $\omega \circ \varepsilon$.*

Example 1.2.31. *For $\omega \neq 0$, the defiance under ω (and in particular the 1-Lipschitz defiance) is a non symmetric relation. For instance if we consider a random variable $X \in [-1,1]$, say, uniformly distributed, we know that $Y = \max(X,0)$ 1-defies X but it is not possible to find a modulus of continuity ω such that X would defy Y under ω. Indeed, if we suppose that such a modulus of continuity exists, we would have for any $t > 0$:*

$$1 \leq \mathbb{P}\left(X \geq t \mid Y = t\right) \leq \mathbb{P}\left(X \geq t - \omega(t) \mid Y = 0\right),$$

and since $\mathbb{P}(X \geq t - \omega(t) \mid Y = 0)$ *is null if* $\omega(t) \leq t$ *and equal to* $\omega(t) - t$ *otherwise, it is necessary that :*

$$\forall t > 0 \ : \ \omega(t) \geq t + 1,$$

therefore, ω *is clearly not continuous at* 0*.*

However, the defiance under 0 is a symmetric relation equivalent to the relation of independence.

Proposition 1.2.32. *Given two random vectors* X *and* Y *:*

X *defies* Y *under* $0 \quad \Leftrightarrow \quad Y$ *defies* X *under* $0 \quad \Leftrightarrow \quad X$ *and* Y *are independent.*

In Definition 11, it is possible to integrate the inequality on y_2 to get a necessary condition for defiance that might help the understanding of the notion.

Lemma 1.2.33. *Given two random vectors* $X \in E$ *and* $Y \in F$*, and a modulus of continuity* ω*, if we suppose that* X *defies* Y *under* ω*, then for any* $y \in F$ *and* $\epsilon > 0$ *:*

$$\mathbb{P}(f(X) \geq t \mid Y = y) \leq \mathbb{P}(f(X) \geq t + \omega(s) \mid Y \in \mathcal{B}_s)$$

The last point of Lemma 1.2.30 gave us an easy way to build couples (X, Y) of defiant random vectors from two independent random vectors (X, Z). However, recalling that our goal is to show the concentration of (X, Y), the example $X = \phi(Y, Z)$ with ϕ ω-continuous is easy to treat since $(\phi(Y, Z), Y)$ is a 2ω-continuous transformation of the vector (Y, Z) that is concentrated when Y and Z are concentrated thanks to Proposition 1.2.26. To let the reader imagine other kinds of defiant vectors, we present an example of simple uniform distribution in $\mathbb{R}^2$ that could possibly be expressed as the ω-continuous transformation $\phi(U, V)$ of two independent random variables U, V but whose defiance can be shown straightforwardly, without Lemma 1.2.30.

Example 1.2.34. *Let us consider the random vector* $(X, Y) \in \mathbb{R}^2$ *following the uniform distribution on the set* $\{(x, y) \in \mathbb{R}_+^2 \mid x^2 + y^2 \leq 1\}$*. We are going to show that* X $(\sqrt{2}, \frac{1}{2})$*-Hölder defies* Y *and for the same reason, since* X *and* Y *play symmetric roles,* Y $(\sqrt{2}, \frac{1}{2})$*-Hölder defies* X*. Let us consider* $y_1, y_2 > 0$*, a 1-Lipschitz function* $f : \mathbb{R} \to \mathbb{R}$ *and a threshold* $t > 0$*; we want to bound* $\mathbb{P}(f(X) \geq t \mid Y = y_2)$ *with* $\mathbb{P}\left(f(X) \geq t - \sqrt{2}\,|y_1 - y_2|^{\frac{1}{2}} \mid Y = y_1\right)$ *and we note for simplicity* $w = \sqrt{2}\,|y_1 - y_2|^{\frac{1}{2}}$ *and for any* $y \in [0, 1]$*,* $r_y = \sqrt{1 - y^2}$*.*

If $y_1 \geq y_2$*, it is easy to see that :*

$$\mathbb{P}(f(X) \geq t \mid Y = y_2) \leq \mathbb{P}(f(X) \geq t \mid Y = y_1) \leq \mathbb{P}(f(X) \geq t - w \mid Y = y_1)$$

If $y_1 \leq y_2$*, if* $\mathbb{P}(f(X) \geq t - w \mid Y = y_1)$ *is equal to* 1*, there is nothing to show, we can thus suppose that it is strictly lower than* 1*. For any* $x \in [0, r_{y_1}]$ *verifying* $f(x) < t - w$*,* $\forall u \in [-w, w]$ *:*

$$f(x + u) \leq f(x + u) - f(x) + f(x) < u + t - w < t,$$

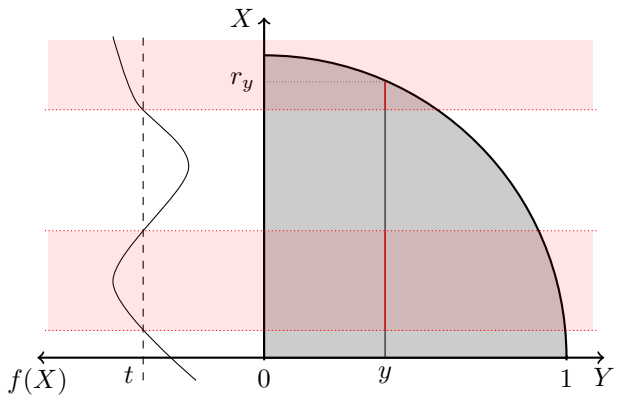

FIG 2. *An example of* $(\sqrt{2}, \frac{1}{2})$*-Hölder defiance. In shaded grey on the right, the uniform distribution on the disk quarter ; on the left of the figure, the graph of a given* 1*-Lipschitz function* f*. The red vertical lines on the right of the figure represent the set* $\{f(X) \geq t, Y = y\}$*. Keep in mind that due to the Lipschitz character of* f*, when* t *is shifted by* $w > 0$*, the red horizontal dotted boundaries are shifted at least by* w*. For any* $y_1, y_2, t \in \mathbb{R}$*,* $\mathbb{P}(f(X) \geq t \mid Y = y_1) \leq \mathbb{P}(f(X) \geq t - \sqrt{2}|y_1 - y_2|^{1/2} \mid Y = y_2)$*, the inequality is the tightest when* $1 > y_2 \geq y_1$ *and when* y_1 *is close to* 1*.*

since f *is* 1*-Lipschitz. Therefore* $\{f < t - w\}_w \subset \{f < t\}$ *(where* $A_w = \{y \mid \exists x \in A, \|x - y\| \leq w\}$ *is the closed neighborhood of order* w *of* A*) and we have the inequality :*

$$\int_0^{r_{y_1}} \mathbb{1}_{f(x) \geq t} dx + w \leq \int_0^{r_{y_1}} \mathbb{1}_{f(x) \geq t - w} dx < 1.$$

With our hypothesis, $r_{y_2} \leq r_{y_1}$*, we can then bound :*

$$\begin{aligned}
&\mathbb{P}\left(f(X) \geq t - w \mid Y = y_1\right) \\
&\quad = \frac{1}{r_{y_1}} \int_0^{r_{y_1}} \mathbb{1}_{f(x) \geq t - w} dx \geq \frac{1}{r_{y_1}} \int_0^{r_{y_2}} \mathbb{1}_{f(x) \geq t} dx + \frac{w}{r_{y_1}} \\
&\quad \geq \mathbb{P}\left(f(X) \geq t \mid Y = y_2\right) - \frac{r_{y_1} - r_{y_2}}{r_{y_1}} \frac{\int_0^{r_{y_2}} \mathbb{1}_{f(x) \geq t} dx}{r_{y_2}} + \frac{w}{r_{y_1}} \\
&\quad \geq \mathbb{P}\left(f(X) \geq t \mid Y = y_2\right) + \frac{w - (r_{y_1} - r_{y_2})}{r_{y_1}}.
\end{aligned}$$

Eventually, with the basic inequalities :

$$r_{y_1} - r_{y_2} = \sqrt{1 - y_1^2} - \sqrt{1 - y_2^2} \leq \sqrt{y_1^2 - y_2^2} \leq \sqrt{2(y_1 - y_2)} = w,$$

we can conclude that X $(\sqrt{2}, \frac{1}{2})$*-Hölder defies* Y*.*

Now that we gave some insights into the notion of defiance, we can enunciate our theorem.

Theorem 1.2.35. *Given two integers $p, q \in \mathbb{N}$ and two random vectors $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^q$, if we suppose that they are both concentrated, and that X defies Y under a modulus of continuity ω, then the concatenation (X, Y) is also concentrated in $(\mathbb{R}^{p+q}, \|\cdot\|_{\ell_1})$, where $\|(x, y)\|_{\ell_1} = \|x\|_2 + \|y\|_2$. More precisely, given two concentration functions α, β, we suppose that $X \propto \alpha$ and $Y \propto \beta$ and we note $t_0 > 0$ verifying $\alpha(t_0), \beta(t_0) < 1 - \frac{1}{\sqrt{2}}$. Then, with the shifting operator τ_{t_0} introduced in the context of Lemma 1.1.1, we have the implication :*

$$\begin{cases} X \propto \alpha \\ Y \propto \beta \end{cases} \qquad \Longrightarrow \qquad (X, Y) \propto \tau_{2t_0} \cdot 2\left(\alpha \circ (1+\omega)^{-1}(\cdot/2) + \beta(\cdot/2)\right),$$

where $(1+\omega)^{-1}$ is the pseudo inverse of $1 + w$.

Since for any $(x, y) \in \mathbb{R}^p \times \mathbb{R}^q$, $\|(x, y)\|_{\ell_1} \leq \sqrt{2}\,\|(x, y)\|_2$, we can adapt the result of Theorem 1.2.35 to a mere concentration in $\mathbb{R}^{p+q}$ endowed with the Euclidean norm just with a slight modification of the concentration function.

The translation coefficient t_0 appearing in the result arises from the expression of the concentrations around quantiles on which the proof of Theorem 1.2.35 relies. Given a random variable Z, and a parameter $\theta > 0$, we use the notation q_θ^Z to designate a θ-quantile of Z that verifies by definition $\mathbb{P}\left(Z \geq q_\theta^Z\right) \geq 1 - \theta$ and $\mathbb{P}\left(Z \leq q_\theta^Z\right) \geq \theta$. Of course there can be several quantiles; we then suppose that the assertions that contains the notation q_θ^Z are true for all the θ-quantiles of Z. Assuming that a random variable Z is concentrated, it is possible to express the behavior around a quantile.

Lemma 1.2.36 ([Led01, Lemma 1.1])**.** *Given a random variable Z, a median m_Z of Z, a concentration function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ and a threshold $\theta \in (0, 1)$, we have the implication :*

$$Z \in m_Z \pm \alpha \qquad \Longrightarrow \qquad Z \in q_\theta^Z \pm \tau_{t_\theta} \cdot \alpha,$$

where t_θ verifies $\alpha(t_\theta) < \min(\theta, 1-\theta)$ and τ_{t_θ} is the translation operator defined before Lemma 1.1.1.

Proof. By definition of the quantile q_θ^Z, we know that $|m_Z - q_\theta^Z| \leq t_\theta$, since the concentration of Z implies when $t = t_\theta$:

$$\begin{cases} \mathbb{P}\left(Z \leq m_Z - t_\theta\right) \leq \alpha(t_\theta) < \theta \\ \mathbb{P}\left(Z \geq m_Z + t_\theta\right) \leq \alpha(t_\theta) < 1 - \theta \end{cases} \qquad \Longrightarrow \qquad \begin{cases} m_Z - t_\theta \leq q_\theta^Z \\ m_Z + t_\theta \geq q_\theta^Z. \end{cases}$$

We can then conclude thanks to Lemma 1.1.1. □

Note that if we only suppose $\alpha(t_\theta) \leq \theta$, then we cannot infer any behavior around a quantile because there can be plenty of them and the inequality $\mathbb{P}\left(Z \leq m_Z - t_\theta\right) \leq \theta$ does not entail anything remarkable.

Proof of Theorem 1.2.35. In the proof we need the uniqueness of the quantiles and of the median; we will thus first suppose that the Lebesgue measure $\lambda_{\mathbb{R}^{p+q}}$ is

absolutely continuous with respect to $\mu_{(X,Y)}$, the distribution of (X,Y), which is classically written as $\lambda_{\mathbb{R}^{p+q}} \ll \mu_{(X,Y)}$ and means that for any Borel set $A \subset \mathbb{R}^{p+q}$:

$$\mathbb{P}\left((X,Y) \in A\right) = 0 \qquad \Longrightarrow \qquad \lambda_{\mathbb{R}^{p+q}}(A) = 0.$$

That allows us to say that any 1-Lipschitz function $f : E \times F \to \mathbb{R}$ and any $a, b \in \mathbb{R}$ such that $a < b$:

$$\begin{cases} \mathbb{P}\left(f(X,Y) \leq a\right) > 0 \\ \mathbb{P}\left(f(X,Y) \geq b\right) > 0 \end{cases} \qquad \Longrightarrow \qquad \mathbb{P}\left(a \leq f(X,Y) \leq b\right) > 0.$$

Indeed, if we suppose that the left assertions are true, then taking $(x_a, y_a) \in \{f \leq a\}$ and $(x_b, y_b) \in \{f \geq b\}$, the intermediate value theorem states that there exists a point $(x_0, y_0) \in [(x_a, y_a), (x_b, y_b)]$ belonging to the open set $\{a < f < b\}$. The set $\{a \leq f \leq b\}$ has then a positive Lebesgue measure as it contains a non empty open subset and $\mathbb{P}\left(a \leq f(X,Y) \leq b\right) > 0$ by hypothesis made on (X,Y). We then see that if $\theta \in (0,1)$, the θ quantile $q_\theta^{f(X,Y)}$ is unique, in particular we note m_f the median of $f(X,Y)$.

Let us introduce a parameter $\theta_t \in (0,1)$ that we will precise later and shall be considered as increasing with t. We start with the decomposition :

$$\mathbb{P}\left(f(X,Y) \geq m_f + t\right) = \int \mathbb{P}\left(f(x,Y) \geq m_f + t \mid X = x\right) d\mathbb{P}(X = x). \tag{11}$$

It is then tempting to employ Markov-like inequality distinguishing the drawings x of X where $\mathbb{P}\left(f(x,Y) \geq m_f + t \mid X = x\right) > 1 - \theta_t$ from the others. It is possible to simplify the expression of the two cases noting that for any $x \in \mathbb{R}^p$:

$$\mathbb{P}\left(f(x,Y) \geq m_f + t \mid X = x\right) > 1 - \theta_t \qquad \Longrightarrow \qquad q_{\theta_t}^x > m_f + t,$$

where $q_{\theta_t}^x = q_{\theta_t}^{f(x,Y)\ |X=x}$ is the θ_t-quantile of the random variable $f(x,Y) \mid X = x$. We get the inequality :

$$\begin{aligned} \mathbb{P}\left(f(X,Y) \geq m_f + t\right) &< \mathbb{P}\left(q_{\theta_t}^X > m_f + t\right) + \left(1 - \mathbb{P}\left(q_{\theta_t}^X > m_f + t\right)\right)(1 - \theta_t) \\ &= \mathbb{P}\left(q_{\theta_t}^X > m_f + t\right)\theta_t + (1 - \theta_t). \end{aligned}$$

Now, to find a lower bound of $\mathbb{P}\left(q_{\theta_t}^X > m_f + t\right)$, let us first consider the case $t = 0$ and for that purpose we employ again the decomposition (11) :

$$\frac{1}{2} = \mathbb{P}\left(f(X,Y) > m_f\right) > \mathbb{P}\left(q_{\theta_0}^X > m_f\right)(1 - \theta_0).$$

That leads to the inequality :

$$\mathbb{P}\left(q_{\theta_0}^X > m_f\right) < \frac{1}{2(1 - \theta_0)}.$$

As suggested in the hypothesis of the proposition, it appears natural to introduce $\theta_0 = 1 - 1/\sqrt{2}$ because in that case $1 - \theta_0 = 1/2(1 - \theta_0)$ and by definition of the quantile, we directly see that $m_f > q_{\theta_0}^{q_{\theta_0}^X}$. We could have chosen for θ_0 any value in $(\frac{1}{2}, 1)$, but the present choice symmetries and somehow optimizes the concentration – if $\alpha = \beta$. Here be careful that due to our choice of notation, $q_{\theta_0}^X$ is a random variable depending on X and $q_{\theta_0}^{q_{\theta_0}^X}$ is a constant.

When $t > 0$, the inequality $m_f > q_{\theta_0}^{q_{\theta_0}^X}$ directly implies :

$$\mathbb{P}\left(q_{\theta_t}^X > m_f + t\right) \leq \mathbb{P}\left(q_{\theta_t}^X > q_{\theta_0}^{q_{\theta_0}^X} + t\right).$$

It is interesting to bound the random variable $q_{\theta_t}^X$ with $q_{\theta_0}^X$ to let appear a concentration inequality involving a random variable $q_{\theta_0}^X$ with its quantile $q_{\theta_0}^{q_{\theta_0}^X}$ as in Lemma 1.2.36. For that purpose, considering the parameter t_0 introduced in the proposition, we set $\theta_{t+2t_0} = 1 - \beta(\frac{t}{2})$. With that choice of θ_t, thanks to Lemma 1.2.36, we know that for any $t > 0$ and $x \in \mathbb{R}$:

$$\mathbb{P}\left(f(x, Y) \geq q_{\theta_0}^x + t_0 + t\right) \leq \beta(t) = 1 - \theta_{2(t+t_0)},$$

thus $\forall t > 0,\ q_{\theta_0}^x + t + t_0 \geq q_{\theta_{2(t+t_0)}}^x$. That entails :

$$\mathbb{P}\left(q_{\theta_{t+2t_0}}^X > m_f + t + 2t_0\right) \leq \mathbb{P}\left(q_{\theta_0}^X > q_{\theta_0}^{q_{\theta_0}^X} + \frac{t}{2} + t_0\right) \leq \alpha \circ (1+\omega)^{-1}\left(\frac{t}{2}\right).$$

The last inequality was obtained with Lemma 1.2.36. Indeed, the function $x \mapsto q_{\theta_0}^x$ is $(1+\omega)$-continuous, given $x_1, x_2 \in E$:

$$\begin{aligned}
\theta_0 &\leq \mathbb{P}\left(f(x_1, Y) \leq q_{\theta_0}^{x_1} \mid X = x_1\right) \\
&\leq \mathbb{P}\left(f(x_2, Y) \leq f(x_1, Y) - f(x_2, Y) + q_{\theta_0}^{x_1} \mid X = x_1\right) \\
&\leq \mathbb{P}\left(f(x_2, Y) \leq \|x_2 - x_1\|_E + q_{\theta_0}^{x_1} \mid X = x_1\right) \\
&\leq \mathbb{P}\left(f(x_2, Y) \leq (1+\omega)\left(\|x_2 - x_1\|_E\right) + q_{\theta_0}^{x_1} \mid X = x_2\right),
\end{aligned}$$

thus $(1+\omega)\left(\|x_2 - x_1\|_E\right) + q_{\theta_0}^{x_1} \geq q_{\theta_0}^{x_2}$ and we get the same way the inequality $q_{\theta_0}^{x_1} - q_{\theta_0}^{x_2} \leq (1+\omega)\left(\|x_2 - x_1\|_E\right)$.

Combining our results all together, we can bound :

$$\mathbb{P}\left(f(X, Y) \geq m_f + t + 2t_0\right) \leq \left(1 - \beta\left(\frac{t}{2}\right)\right)\ \alpha \circ (1+\omega)^{-1}\left(\frac{t}{2}\right) + \beta\left(\frac{t}{2}\right),$$

and we get the same bound for the probability $\mathbb{P}\left(f(X, Y) \leq m_f - t - 2t_0\right)$ that eventually gives us the result of the proposition in the case where the distribution of (X, Y) controls the Lebesgue measure.

In the general case we can approximate (X, Y) with a sequence of random vectors (X_n, Y_n) verifying $\lambda_{\mathbb{R}^{p+q}} \ll \mu_{(X_n, Y_n)}\ \forall n \in \mathbb{N}$. For that purpose, let

us introduce U, V two independent Gaussian random vectors verifying $U \sim \mathcal{N}(0, I_p))$ and $V \sim \mathcal{N}(0, I_q)$ and respectively independent to X and Y. We know from Example 1.2.27 that

$$\begin{cases} X_n = X + \frac{1}{n}U \overset{m}{\propto} \alpha_n \quad \text{where} \quad \alpha_n(t) = \alpha\left(\left(1 + \frac{1}{\sqrt{n}}\right)t\right) + 2e^{-nt^2/2}, \\ Y_n = Y + \frac{1}{n}V \overset{m}{\propto} \beta_n \quad \text{where} \quad \beta_n(t) = \beta\left(\left(1 + \frac{1}{\sqrt{n}}\right)t\right) + 2e^{-nt^2/2}, \end{cases}$$

and we know that $\lambda_{\mathbb{R}^{p+q}}$ is absolutely continuous under $\mu_{(X_n,Y_n)}$. Let us consider a 1-Lipschitz function $g : \mathbb{R}^{p+q} \to \mathbb{R}$. We note m_n the median of $g(X_n, Y_n)$, $m_+ = \sup\{m, \mathbb{P}(g(X,Y) \geq m) \geq \frac{1}{2}\}$ and $m_- = \inf\{m, \mathbb{P}(g(X,Y) \leq m) \geq \frac{1}{2}\}$. For any $\varepsilon > 0$, there exists $\delta > 0$, such that we can bound :

$$\begin{aligned} \frac{1}{2} + \delta &\leq \mathbb{P}(g(X,Y) \leq m_+ + \varepsilon) \leq \mathbb{P}\left(g(X_n, Y_n) - \frac{1}{n}\|(U,V)\| \geq m_+ + \varepsilon\right) \\ &\leq \mathbb{P}(g(X_n, Y_n) \geq m_+ + 2\varepsilon) + \mathbb{P}\left(\frac{1}{n}\|(U,V)\| \geq \varepsilon\right). \end{aligned}$$

and if n is large enough, we know from Proposition 1.2.10 (and Proposition 1.2.26 : (U, V) is concentrated since U and V are independent and concentrated) that $\mathbb{P}\left(\frac{1}{n}\|(U,V)\| \geq \varepsilon\right)$ tends to 0 when n tends to infinity. Therefore we obtain for n large enough the inequality $\frac{1}{2} \leq \mathbb{P}(g(X_n, Y_n) \geq m_-)$ that directly entails $m_+ + 2\varepsilon \leq m_n$. With similar considerations, we show that the sequence $(m_n)_{n\in\mathbb{N}}$ belongs to $[m_- - 2\varepsilon, m_+ + 2\varepsilon]$ for all n large; then, replacing ε by the elements of a sequence ε_n tending to zero, we can extract a subsequence of $(m_n)_{n\in\mathbb{N}}$ that we will still abusively note $(m_n)_{n\in\mathbb{N}}$ and that converges to a real value $m_g \in [m_-, m_+]$. The limit m_g we obtain is then clearly a median of $g(X,Y)$. The hypothesis of Proposition 1.1.12 :

- m_n tends to m_g
- $g(X_n, Y_n)$ tends in law to $g(X,Y)$
- $\tau_{2t_0} \cdot 2\left(\alpha_n \circ (1+\omega)^{-1}(\cdot/2) + \beta_n(\cdot/2)\right)$ point-wise converges to $\gamma = \tau_{2t_0} \cdot 2\left(\alpha \circ (1+\omega)^{-1}(\cdot/2) + \beta(\cdot/2)\right)$

are verified, we can thus conclude that $(X,Y) \overset{m}{\propto} \gamma$.

□

In the case of exponential concentration, we can employ Lemma 1.1.16 to get the simple corollary :

Corollary 1.2.37. *Given two random vectors $X, Y \in E$ and three parameters $C \geq e$, $\sigma, q > 0$, we have the implication :*

$$X, Y \overset{m}{\propto} Ce^{-(\cdot/\sigma)^q} \qquad \Longrightarrow \qquad (X,Y) \overset{m}{\propto} 4Ce^{-(\cdot/4\sigma)^q}.$$

We can then express the concentration of a sum of two exponentially concentrated random vectors :

Corollary 1.2.38. *In the setting of Corollary 1.2.37 :*

$$X, Y \overset{m}{\propto} Ce^{-(\cdot/\sigma)^q} \qquad \Longrightarrow \qquad X + Y \overset{m}{\propto} 4Ce^{-(\cdot/4\sigma)^q}.$$

If we now consider that X and Y are two vectors of an algebra $\mathcal{A}$ endowed with an algebra norm $\|\cdot\|$ (verifying $\|XY\| \leq \|X\| \|Y\|$), we can express the concentration of the product XY thanks to Lemma 1.1.8. We present the result in the case of exponential concentration for which the expression is simple, but the proof could be adapted to any other type of concentration function.

Proposition 1.2.39 (Concentration of the vector product)**.** *Given two independent random vectors $X, Y \in \mathcal{A}$, and three parameters, $q \geq 1$, $C \geq e$ and $\sigma > 0$, if $X, Y \overset{m}{\propto} Ce^{-(\cdot/\sigma)^q}$, then noting $\tilde{X}, \tilde{Y}$, some deterministic equivalents of respectively X and Y, we have :*

$$(X - \tilde{X})(Y - \tilde{Y}) \propto C_1 e^{-(\cdot/c_1\sigma^2\eta_{\|\cdot\|}^{1/q})^q} + C_2 e^{-(\cdot/c_2\sigma^2)^{q/2}},$$

where $C_1, c_1, C_2, c_2 \geq 1$ are four numerical constants depending only on C and $\eta_{\|\cdot\|}$ is the degree of the norm $\|\cdot\|$, presented in Definition 9.

Of course if one (or both) of the random vectors X and Y has a bounded norm, then the concentration is tighter.

Proof. Let us suppose for simplicity that $\tilde{X} = \tilde{Y} = 0$. Noting $m_{\|X\|}$ and $m_{\|Y\|}$, the median of respectively $\|X\|$ and $\|Y\|$, we know from Corollary 1.2.14 that

$$m_{\|X\|}, m_{\|Y\|} \leq C'\sigma\eta_{\|\cdot\|}^{1/q},$$

for some numerical constant $C' \geq e$ proportional to C. We note $\eta = C'\sigma\eta_{\|\cdot\|}^{1/q} > 0$. Considering a 1-Lipschitz function $f : \mathcal{A} \to \mathbb{R}$, we are going to show that $f(XY)$ is concentrated as the product of two concentrated random variables $Z_1 = \|X\| + \|Y\| + \eta$ and $Z_2 = \frac{f(XY)}{\|X\|+\|Y\|+\eta}$. First, we know from Proposition 1.1.5 that $Z_1 \propto \eta + m_{\|X\|} + m_{\|Y\|} \pm 2Ce^{-(\cdot/2\sigma)^q}$. Second, Proposition B.4.1 gives us the concentration $(X, Y) \overset{m}{\propto} 4Ce^{-(\cdot/4\sigma)^q}$, and we want thus to show that the map $g : (x, y) \mapsto \frac{f(x,y)}{\|x\|+\|y\|+\eta}$ is Lipschitz. Given $(h, k) \in \mathcal{A}^2$ such that $\|h\|, \|k\| \leq \frac{\eta}{2}$, let us bound :

$$\begin{aligned} |g(x + h, y + k) - g(x, y)| &\leq \left| \frac{\|x\| - \|x + h\| + \|y\| - \|y + k\|}{(\|x + h\| + \|y + k\| + \eta)(\|x\| + \|y\| + \eta)} \right| |f(xy)| \\ &\quad + \frac{|f((x + h)(y + k)) - f(xy)|}{\|x\| + \|y\| + \eta} \\ &\leq (\|h\| + \|k\|) \frac{|f(xy)|}{\|x\| \, \|y\|} + 2(\|h\| + \|k\|). \end{aligned}$$

Let us suppose without loss of generality that $f(0) = 0$; the Lipschitz character of f gives us the inequality :

$$f(xy) \leq \|xy\| \leq \|x\| \, \|y\| .$$

Therefore g is 2-Lipschitz and we know from Proposition 1.1.9 that there exist two medians m_{Z_1} and m_{Z_2} of respectively Z_1 and Z_2 such that :

$$f(XY) = Z_1 Z_2 \in m_{Z_1} m_{Z_2} \pm Ce^{-(\cdot/4\sigma|m_{Z_2}|)^q} + 4Ce^{-(\cdot/16\sigma m_{Z_1})^q} + Ce^{-(\cdot/2\sigma^2)^{q/2}}.$$

That allows us to conclude since $m_{Z_1} \leq 2\eta$ and besides, $\frac{f(xy)}{\|x\|+\|y\|+\eta} \leq \frac{\max(\|x\|+\|y\|)}{2}$, and thus $m_{Z_2} \leq \eta$. □

Example 1.2.40. *We give as in Examples 1.2.16 and 1.2.17 some results for the product $\odot$ in $\mathbb{R}^p$ and the matricial product in $\mathcal{M}_{pn}$. Therefore we consider a random vector $Z \in \mathbb{R}^p$ and a random matrix $X \in \mathcal{M}_{pn}$ verifying $Z, X \propto 2e^{-(\cdot/2)^2}$ for the Euclidean norm (the Frobenius norm in $\mathcal{M}_{pn}$) and such that $\mathbb{E}[Z] = 0$ and $\mathbb{E}[X] = 0$:*

- $\frac{Z^{\odot 2}}{\sqrt{p}} = \frac{Z \odot Z}{\sqrt{p}} \propto Ce^{-c\cdot^2} + Ce^{-cp^{1/4}\cdot}$ *in* $(\mathbb{R}^p, \|\cdot\|_2)$
- $\frac{Z^{\odot 2}}{\sqrt{\log p}} \propto Ce^{-c\cdot^2} + Ce^{-c(\log p)^{1/4}\cdot}$ *in* $(\mathbb{R}^p, \|\cdot\|_\infty)$
- $\frac{XX^T}{n} \propto Ce^{-c(\cdot/\sqrt{\gamma})^2} + Ce^{-cn\cdot/\sigma^2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_F)$
- $\frac{XX^T}{\sqrt{n}} \propto Ce^{-c(\cdot/\sqrt{\bar{\gamma}})^2} + Ce^{-c\sqrt{n}\cdot/\sigma^2}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|)$
- $\frac{XX^T}{\sqrt{\log n}} \propto C\exp\left(-\frac{c\cdot^2}{1+\frac{\log p}{\log n}}\right) + Ce^{-c\sqrt{\log n}\cdot}$ *in* $(\mathcal{M}_{p,n}, \|\cdot\|_\infty)$,

where $\gamma = \frac{p}{n}$, $\bar{\gamma} = \gamma + 1 \geq \max(\gamma, 1)$ and $C \geq 1$ and $c > 0$ are two numerical constants. One can note in these examples that the Lipschitz concentration preserves the concentration rates better than the linear concentration through the product of random vectors.

Let us pursue our study of the concentration in algebras with the concentration of the power of a random vector as in Proposition B.3.1.

Proposition 1.2.41. *Given an integer $m \geq 2$, a random vector $Z \in \mathcal{A}$ and two parameters $C \geq e$ and $\sigma > 0$, if $Z \propto Ce^{-(\cdot/\sigma)^q}$, there exist two numerical constants $C' \geq e$ and $c > 0$ depending on C, q and m such that :*

$$Z^m \propto C'\exp\left(-c\left(\frac{\cdot}{\sigma\mathbb{E}[\|Z\|]^{m-1}}\right)^q\right) + C'e^{-c(\cdot/\sigma^m)^{\frac{q}{m}}} + C'e^{-c(\cdot/\sigma)^q}.$$

Remark 1.2.42. *If one compares Proposition 1.2.41 with Proposition B.3.1 that gives the linear concentration of Z^m when Z is linearly concentrated, one sees that the Lipschitz concentration is more respectful to the power of random vectors since the observable diameter of Z^m is $\eta_{\|\cdot\|}^{1/q}$ smaller (for any $m \geq 2$) in the case of Lipschitz concentration. To be more precise, with the notations of the propositions and in the case where $\|\mathbb{E}Z\| \sim \sigma\eta_{\|\cdot\|}^{1/q}$, the leading tail parameter in the combination of exponential concentrations is $\sigma^m\eta_{\|\cdot\|}^{m/q}$ for linear concentration and $\sigma^m\eta_{\|\cdot\|}^{(m-1)/q}$ for Lipschitz concentrations.*

Proof. Given a 1-Lipschitz function $f : \mathcal{A} \to \mathcal{A}$, we decompose as in Proposition 1.2.39 the functional $f(Z^m)$ into the product of two random variables,

$Z_1 = (\eta + \|Z\|)^m$ where $\eta = \mathbb{E}[\|Z\|]$ plays the same role as before and $Z_2 = f(Z^m)/Z_1 = g(Z)$. We know from Proposition 1.1.10 that :

$$Z_1 \in (2\mathbb{E}[\|Z\|])^m \pm C \exp\left(-\left(\frac{\cdot}{2^{2m-1}\sigma\mathbb{E}[\|Z\|]^{m-1}}\right)^q\right) + Ce^{-(\cdot/2\sigma^m)^{q/m}},$$

We are just left to show that g is Lipschitz. Let us consider a parameter $\varepsilon > 0$ that will tend to zero and two vectors $x, h \in \mathcal{A}_*$ such that $\|h\| \leq ((1+\varepsilon)^{1/m} - 1)\|x\|)$, we know that $\forall l \leq m$, $\|x+h\|^l \leq (1+\epsilon)\|x\|^l$ and we have :

$$\begin{aligned}
&|g(x+h) - g(x)| \\
&\quad \leq \left|\frac{\|x+h\|^m - \|x\|^m}{(\|x+h\|+\eta)^m(\|x\|+\eta)^m}\right| |f((x+h)^m)| + \frac{|f(x^m) - f((x+h)^m)|}{(\|x\|+\eta)^m} \\
&\quad \leq \frac{\|h\|\left(\|x\|^m + \|x\|^{m-2}\|x+h\| + \ldots + \|x+h\|^m\right)}{(\|x\|+\eta)^m} \frac{|f((x+h)^m)|}{(\|x+h\|+\eta)^m} \\
&\qquad + \left|\frac{\|x+h\|^m - \|x\|^m}{(\|x\|+\eta)^m}\right| \;\leq\; (1+\varepsilon)\frac{2m}{\eta}\|h\|.
\end{aligned}$$

As before we suppose, without loss of generality, that $f(0) = 0$ and therefore that $\forall x \in \mathcal{A}$, $f(x) \leq \|x\|$. Of course, the same inequality holds if $x = 0$ and $h \in \mathcal{A}$ is chosen arbitrarily. Therefore, letting ε tend to 0, one sees that g is $\frac{2m}{\eta}$-Lipschitz and $Z_2 = g(Z) \in \mathbb{E}[Z_2] \pm Ce^{-(\eta\cdot/2m\sigma)^{1/q}}$. Besides, knowing that $Z_2 \leq 1$, we can deduce the concentration of $f(Z^m) = Z_1 Z_2$ thanks to Lemma 1.1.8 :

$$f(Z^m) \in \mathbb{E}Z_1\mathbb{E}Z_2 \pm 2C\exp\left(-\left(\frac{\cdot/2^{2m}}{\sigma\mathbb{E}[\|Z\|]^{m-1}}\right)^q\right) + Ce^{-(\cdot/4\sigma^m)^{\frac{q}{m}}} + Ce^{-(\cdot/8m\sigma)^q}.$$

□

In the preamble, we presented the resolvent as the convenient object that one wants to study to get some insight into the spectrum of the sample covariance of a concentrated random vector. Given a matrix $X \in \mathcal{M}_{p,n}$ and a positive real number $z > 0$, recall that the resolvent Q_S of the sample covariance matrix $S = \frac{1}{n}XX^T$ is defined as :

$$Q_S(z) = \left(XX^T/n + zI_p\right)^{-1}.$$

We will simply note it Q to lighten the notations. Following Example 1.2.40, it is possible to show that the resolvent is concentrated with the same observable diameter as S (and X). However, it is possible to get a better concentration rate if we take advantage of the fact that Q is a $\frac{2}{\sqrt{z^3 n}}$-Lipschitz transformation of X; we gain this way a factor of order $\sqrt{n}$ in the concentration that will be vital in Section 2. This suggests why the resolvent is so efficient for the study of the spectral distribution of S.

Proposition 1.2.43. *Given $z > 0$, a random matrix $X \in \mathcal{M}_{p,n}$ and a concentration function α, we have the implication :*

$$X \propto \alpha \qquad \Longrightarrow \qquad Q = \left(\frac{1}{n}XX^T + zI_p\right)^{-1} \propto \alpha\left(\frac{\sqrt{nz^3}}{2}\cdot\right).$$

In particular :

$$X \overset{\mathbb{E}}{\propto} \alpha \qquad \Longrightarrow \qquad Q \in \mathbb{E}[Q] \pm \alpha\left(\frac{\sqrt{nz^3}}{2}\cdot\right).$$

We need a preliminary lemma before giving the proof to control the Frobenius norm of a product :

Lemma 1.2.44. *Given $A \in \mathcal{M}_{pn}$ and $B \in \mathcal{M}_{np}$, one has the bound :*

$$\|AB\|_F \leq \|A\| \|B\|_F \qquad \text{and} \qquad \|AB\|_F \leq \|A\|_F \|B\|.$$

One must be careful that in most cases $\|AB\|_F \neq \|BA\|_F$, which is why we need to display both inequalities. Recall in passing that the Cauchy-Schwarz inequality gives us directly $\|AB\|_F \leq \|A\|_F \|B\|_F$.

Proof. Lemma 1.2.44 is just a consequence of the computations :

$$\|AB\|_F^2 \leq \sum_{j=1}^{p} \|AB_{\cdot,j}\|_2^2 \leq \|A\|^2 \sum_{j=1}^{p} \|B_{\cdot,j}\|_2^2 \leq \|A\|^2 \|B\|_F^2.$$

The role of A and B could be inverted in the calculus without any problem. □

Proof of Proposition 1.2.43. The function $\phi : \mathcal{M}_{p,n} \to \mathcal{M}_p$ defined as $\phi(R) = \left(RR^T + zI_p\right)^{-1}$ is $2/z^{3/2}$-Lipschitz. Indeed, given $R, H \in \mathcal{M}_{p,n}^2$:

$$\begin{aligned}\phi(R+H) - \phi(R) &= \left((R+H)(R+H)^T + zI_p\right)^{-1} - \left(RR^T + zI_p\right)^{-1} \\ &= \phi(R+H)\left((R+H)H + HR^T\right)\phi(R).\end{aligned}$$

Thus $\|\phi(R+H) - \phi(R)\| \leq 2\|H\|_F / z^{3/2}$ because the Frobenius norm controls the spectral one and because of the basic result $\|\phi(R)R\| \leq 1/\sqrt{z}$ and $\|\phi(R)\| \leq \frac{1}{z}$ enunciated in Lemma 0.0.2 in the preamble. Now, since $Q(z) = \phi(X/\sqrt{n})$, we recover directly the result of the proposition thanks to the hypothesis on X. The second implication is just a consequence of Remark 1.2.19. □

If we try to characterize geometrically, and roughly speaking, the range of random vectors concerned by Theorem 0.0.3, we would describe the set of respectful modifications of the sphere where bounded dilatations or the removals of some parts are tolerated but any cut is forbidden (we have about the same statement considering Theorem 1.2.20). This represents already a good range of distributions but one might be interested in representing discrete or at least "discontinuous" distributions.

1.2.3. Convex Concentration

With a combinatorial approach, Talagrand showed in the nineties that it is possible to find a weaker notion of concentration to apprehend the concentration of partly discrete distributions. In these cases, to be concentrated the "observation" not only needs to be Lipschitz but also to be quasiconvex, in the sense of the following definition.

Definition 12 (Quasiconvexity)**.** *A function $f : E \to \mathbb{R}$ is said to be quasiconvex if for any real $t \in \mathbb{R}$, the set $\{z \in \mathbb{E} \ : \ f(z) \leq t\} = \{f \leq t\}$ is convex.*

Remark 1.2.45. *Quasiconvexity concerns of course convex functions, but also any monotonous function supported on $\mathbb{R}$. More generally, given a convex function f and a non decreasing function g, the composition $g \circ f$ is quasiconvex.*

The class of quasiconvex functions is rather interesting in the sense that it is wider than the class of merely convex functions but still verifies the property of the uniqueness of the minimum.

Definition 13 (Convex concentration)**.** *Given a random vector $Z \in (E, \|\cdot\|)$ and a concentration function α, we say that Z is convexly α-concentrated if one of the three assertions is verified for any 1-Lipschitz and quasiconvex function $f : E \to \mathbb{R}$:*

- *$f(Z) \propto \alpha$, and we will note in that case $Z \propto_c \alpha$*
- *$f(Z) \in m_f \pm \alpha$, and we will note in that case $Z \overset{m}{\propto}_c \alpha$*
- *$f(Z) \in \mathbb{E}[f(Z)] \pm \alpha$, and we will note in that case $Z \overset{\mathbb{E}}{\propto}_c \alpha$,*

where m_f is a median of $f(X)$.

Remark 1.2.46. *It is clear that the concentration of Definition 10 implies the convex concentration of Definition 13. Those two notions are equivalent when $E = \mathbb{R}$ since they are then both equivalent to Definition 2.*

Once again, when non ambiguous, we will omit the precision "in $(E, \|\cdot\|)$". We clearly have the implication :

$$Z \propto \alpha \implies Z \propto_c \alpha.$$

In the case of a q-exponential concentration, we have the implication chain :

$$Z \overset{\mathbb{E}}{\propto} Ce^{-(\,\cdot\,/\sigma)^q} \implies Z \overset{\mathbb{E}}{\propto}_c Ce^{-(\,\cdot\,/\sigma)^q} \implies Z \in \mathbb{E}Z \pm e^{-(\,\cdot\,/\sigma)^q}.$$

The fundamental example that alone justifies the interest in convex concentration is owed to Talagrand and provides to our study a supplementary setting to the "smooth" scenarios given by Theorems 0.0.3 and 1.2.20.

Theorem 1.2.47 (Convex concentration of the product of bounded distributions, [Tal95, Theorem 4.1.1])**.** *Given a random vector $Z \in [0,1]^m$, $m \in \mathbb{N}$, with independent entries :*

$$Z \overset{m}{\propto_c} 4e^{-\cdot^2/4}.$$

Considering the example of the preamble, we are tempted to look for a similar theorem where the sets $[0,1]$ are compact sets of $\mathbb{R}^p$ bounded by K and m is taken to be equal to n. It is not so interesting however, the issue being that the factor K that would appear in the tail parameter would jeopardize most of the applications since it should be of order 1 while $\|Z_i\|$ is often of order $\sqrt{p}$ when $Z_i \in \mathbb{R}^p$. This can however find some use when considering sparse random vectors.

The interesting setting is the case where $E = \mathbb{R}$, and $m = pn$. Then we do not exactly consider the concentration on $\mathbb{R}^{np}$ but the concentration on $\mathcal{M}_{p,n}$ (endowed with the Frobenius norm). Theorem 1.2.47 gives us in that case a convenient tool to build convexly q-exponentially concentrated random matrices. In that case, the bound $|Z_i| \leq K$ on the entries of a random vector Z is no more an unreachable hypothesis for applications; however we will need in that case an independence between the entries.

Theorems 0.0.3 and 1.2.20 allow us to track easily concentration properties from a vector $Z \in \mathbb{R}^p$ verifying $Z \sim \sigma_{p-1}$ or $Z \sim \mathcal{N}(0, I_p)$ to any Lipschitz transformation and even to any uniformly continuous transformation $f(Z) \in \mathbb{R}^q$. Theorem 1.2.47 is not so easy to generalize because the convexity (or the quasiconvexity) of a function is only defined for real-valued functions; indeed, most of the transformations between two vector spaces ruin the subtle structure of convexity. We can still slightly relax the hypothesis of independence in the theorem of Talagrand thanks to affine transformations :

Lemma 1.2.48. *Given two vector spaces E, F and a quasiconvex (resp., convex) function $f : E \to \mathbb{R}$, for any affine function $g : F \to E$, the composition $f \circ g$ is also quasiconvex (resp., convex).*

Remark 1.2.49. *Given two convexly concentrated random vectors $X, Y \in E$, it is not possible to adapt the proof of Proposition 1.2.35 to the convex case to show the concentration of (X, Y). For convex concentration, the independence between X and Y appears as a crucial element to deduce from a concentration on X and Y a concentration on (X, Y). We advice again the reader to take a look at Proposition B.4.1 in Appendix B.4 to see how to express the convex concentration of $(Z_1, \ldots, Z_p)$ when $Z_1, \ldots, Z_p$ are all q-exponentially convexly concentrated. This result is the same for Lipschitz and convex concentration.*

We can give a supplementary useful proposition that will allow us to keep the properties of convex concentration when thresholding a random vector. But before let us give an immediate preliminary lemma. We present it without proof since it is a direct consequence of Lemma 1.1.19.

Lemma 1.2.50. *Given a random vector Z, an exponent $q > 0$ and two positive constants $C \geq e$ and $\sigma > 0$, if we note $m_{\|Z\|}$ a median of $\|Z\|$, we have the implication :*

$$Z \overset{m}{\propto_c} Ce^{-(\,\cdot\,/\sigma)^q} \quad \Longrightarrow \quad \forall t \geq 2m_{\|Z\|} \;\; \mathbb{P}\left(\|Z\| \geq t\right) \leq Ce^{-(t/2\sigma)^q}.$$

The lemma displays a slight modification of the concentration constants, with here the behavior of the tail only beyond $2\mathbb{E}\,\|Z\|$. The next proposition allows us to say that the concentration of a q-exponentially convexly concentrated random vector occurs under a threshold of order $2m_{\|Z\|}$.

Proposition 1.2.51. *Given a random vector $Z \in E$ a constant $K \geq 2m_{\|Z\|}$, we introduce the vector $\bar{Z} = \min(1, \frac{K}{\|Z\|})Z$. If there exist an exponent $q > 0$ and two parameters $C \geq e$ and $\sigma > 0$ such that $Z \overset{m}{\propto_c} Ce^{-(\,\cdot\,/\sigma)^q}$ then*

$$\bar{Z} \propto_c 4Ce^{-(\,\cdot\,/4\sigma)^q}.$$

If $Z \propto Ce^{-(\,\cdot\,/\sigma)^q}$, it is possible to see $\bar{Z}$ as a 1-Lipschitz transformation of Z that would be naturally concentrated. However the hypothesis of quasiconvexity of the functionals $f(\bar{Z})$ required by the convex concentration cannot be extrapolated so simply.

Proof. Let us consider a function $f : E \to \mathbb{R}$ quasiconvex and 1-Lipschitz. We know from Proposition 1.1.2 that $Z \propto_c 2Ce^{-(\,\cdot\,/2\sigma)^q}$, thus introducing Z', an independent copy of Z, we can bound :

$$\begin{aligned}
\mathbb{P}\left(\left|f(\bar{Z}) - f(\bar{Z}')\right| \geq t\right) &\leq \mathbb{P}\left(\left|f(\bar{Z}) - f(\bar{Z}')\right| \geq t, \|Z\| \leq K \;\text{ and }\; \|Z'\| \leq K\right) \\
&\quad + \mathbb{P}\left(\left|f(\bar{Z}) - f(\bar{Z}')\right| \geq t, \|Z\| > K \;\text{ or }\; \|Z'\| > K\right) \\
&\leq \mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) + \mathbb{P}\left(\|Z\| > K \;\text{ or }\; \|Z'\| > K\right) \\
&\leq 2Ce^{-(t/2\sigma)^q} + 2Ce^{-(K/2\sigma)^q}.
\end{aligned}$$

The last inequality results from the hypothesis on Z and f and Lemma 1.2.50. By construction $\left\|\bar{Z}\right\|, \left\|\bar{Z}'\right\| \leq K$, thus $\left\|\bar{Z} - \bar{Z}'\right\| \leq 2K$ and f being 1-Lipschitz :

$$\text{if } t > 2K, \quad \mathbb{P}\left(\left|f(\bar{Z}) - f(\bar{Z}')\right| \geq t\right) \leq \mathbb{P}\left(\left\|\bar{Z} - \bar{Z}'\right\| \geq 2K\right) = 0.$$

Now, for any $t \leq 2K$:

$$\exp\left(-\left(\frac{K}{2\sigma}\right)^q\right) \leq \exp\left(-\left(\frac{t}{4\sigma}\right)^q\right),$$

therefore, if we rejoin the different regimes, we obtain :

$$\forall t > 0 \;\; : \quad \mathbb{P}\left(\left|f(\bar{Z}) - f(\bar{Z}')\right| \geq t\right) \leq 4C\exp\left(-\left(\frac{t}{4\sigma}\right)^q\right),$$

and we can show exactly in the same manner that :

$$\forall t > 0 \;\; : \quad \mathbb{P}\left(\left|f(\bar{Z}) - \mathbb{E}\left[f(\bar{Z})\right]\right| \geq t\right) \leq 4C\exp\left(-\left(\frac{t}{4\sigma}\right)^q\right).$$

□

Contrarily to the linear and Lipschitz concentrations, convex concentration does not seem to be preserved through the product. To derive any results on the product from a convex concentration, we will thus develop the habit of returning on the linear concentration configuration which is weaker but in a sense "stable" on algebra, like the Lipschitz concentration.

If we fail to maintain the concentration of the vectorial product, as implicitly evoked Lemma 1.2.50, we still have the concentration of the norm as it is a Lipschitz and convex map. That makes a fundamental difference with the linear concentration which entices an important result on quadratic forms quite similar to the Hanson-Wright inequality as found in [Ver17, Theorem 6.2.1] for instance. It is also a good improvement to [El 09, Lemma 8] :

Theorem 1.2.52. *Let us consider two integers $q, m \in \mathbb{N}$ and a random vector $Z \in \mathbb{R}^p$, two positive constants $C \geq e$ and $\sigma > 0$ and a matrix $A \in \mathcal{M}_p$. If we suppose that $Z \overset{\mathbb{E}}{\propto}_c Ce^{-(\cdot/\sigma)^q}$ then we have :*

$$Z^T AZ \in \mathrm{Tr}(A\mathbb{E}[ZZ^T]) \pm 2Ce^{-(\cdot/4\sigma\|A\|\mathbb{E}\|Z\|)^q} + 2Ce^{-(\cdot/2\|A\|\sigma^2)^{\frac{q}{2}}},$$

It is possible to replace the mean with the median if needed.

Proof. Let us first consider the case where A is symmetric nonnegative definite; in this case, $Z^T AZ = \|A^{\frac{1}{2}}Z\|^2$. Theorem 1.2.52 is a particular case of Proposition 1.1.10 that gives the concentration of the r-power of a random variable $\|u(Z)\|$ when u is quasiconvex and 1-Lipschitz and $r \geq 1$:

$$\|u(Z)\|^r \in \mathbb{E}[\|u(Z)\|]^r \pm Ce^{-\left(\cdot/2^r\sigma\|u\|\mathbb{E}[\|u(Z)\|]^{r-1}\right)^q} + Ce^{-\left(\frac{\cdot}{2\|u\|^r\sigma^r}\right)^{\frac{q}{r}}}. \quad (12)$$

The function $z \mapsto \|A^{\frac{1}{2}}z\|$ is $\|A\|^{\frac{1}{2}}$-Lipschitz and convex. Therefore :

$$\|A^{\frac{1}{2}}Z\| \in \mathbb{E}[\|A^{1/2}Z\|] \pm Ce^{-(\cdot/\sigma\|A\|^{\frac{1}{2}}\mathbb{E}\|A^{1/2}Z\|)^q}.$$

Since $\|A^{\frac{1}{2}}Z\| \leq \|A^{\frac{1}{2}}\|\|Z\|$, $\mathbb{E}[\|A^{1/2}Z\|] \leq \|A^{\frac{1}{2}}\|\mathbb{E}[\|Z\|]$ and we can then conclude thanks to (12).

Now if we consider a general matrix $A \in \mathcal{M}_p$, let us decompose $A = A_+ - A_- + A_0$ where A_+ is nonnegative symmetric, A_- is non positive symmetric and A_0 is antisymmetric. We have clearly $Z^T AZ = Z^T A_+ Z - Z^T A_- Z$ and we can conclude thanks to Lemma 1.1.5. □

Remark 1.2.53. *The original Hanson-Wright inequality does not take as hypothesis the concentration (or the convex concentration) of the whole vector $Z = (z_1, \ldots, z_p)$ but just the concentration of each one of its coordinates z_i. However, it assumes that the different coordinates of Z are independent which is quite a strong hypothesis and also that their means are equal to zero. The concentration result obtained under these hypotheses is not exactly the same, and relies on a quantity K that could be seen as the maximum tail parameter of the $\{z_i\}_{1\leq i\leq p}$ ($K = \sigma$ in our case). The Hanson-Wright concentration can*

indeed be written :

$$\mathbb{P}\left(\left|Z^TAZ - \mathbb{E}Z^TAZ\right| \geq t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4\left\|A\right\|_F^2}, \frac{t}{K^2\left\|A\right\|}\right)\right),$$

where $K = \max\{\|z_i\|_{\psi_2}\}_{1\leq i\leq p}$ *is the maximum of the Orlicz norms defined as :*

$$\|z\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\psi_2(|z|/t) \leq 1\} \quad \textit{with} \quad \psi_2(x) = e^{x^2} - 1.$$

The tail parameters of the 1*-exponential component is the same as the one of Theorem 1.2.52. The tail parameter of the* 2*-exponential component is proportional to* $K\left\|A\right\|_F$ *while in our result, it is proportional to* $\|A\|\,\mathbb{E}[\|Z\|] \sim \sqrt{p}\,\|A\|$*. Therefore, considering that in most cases, i.e., when* A *has a high rank and eigenvalues mainly of the same order,* $\left\|A\right\|_F \sim \sqrt{p}\,\|A\|$*, we see that the result of Theorem 1.2.52 is quite similar to the Hanson-Wright inequality if we do not take into account the hypotheses which are quite different (on the one hand they are stronger because they only require the whole vector* Z *to be concentrated, on the other hand they are weaker since they do not exploit the independence between the entries).*

Although Definition 10 is perfectly adapted to the study of the resolvent Q $(= Q_S)$, the problem is far less immediate in the setting of Definition 13, i.e., when $X \propto_c \alpha$. Indeed in the setting of convex concentration, there does not exist any analogue to Proposition 1.2.43 since it does not seem clear whether a functional of Q, $f(Q)$, with f 1-Lipschitz and quasiconvex can be written $g(X/\sqrt{n})$ with g verifying the same properties. In this case, the function ϕ cannot transfer the quasiconvexity and the study must then be conducted downstream directly with the random variables such as $\operatorname{Tr} Q$.

It is interesting to note that $\operatorname{Tr} Q = \operatorname{Tr}(XX^T/n + zI_p)^{-1}$ stays unmodified if we multiply X on the left or on the right by any orthogonal matrix. More formally, let us introduce the group $\mathcal{O}_{p,n} = \mathcal{O}_p \times \mathcal{O}_n$ where $\mathcal{O}_m$, $m \in \mathbb{N}$, is the orthogonal group of matrices of $\mathcal{M}_m$; it acts on $\mathcal{M}_{p,n}$ following the formula :

$$\text{for} \;\; (U,V) \in \mathcal{O}_{p,n},\; M \in \mathcal{M}_{p,n} \;\; : \;\; (U,V)\cdot M = UMV^T.$$

The function $f : R \in \mathcal{M}_{p,n} \mapsto \operatorname{Tr}\phi(R)$ is $\mathcal{O}_{p,n}$-invariant, in the sense that $\forall (U,V) \in \mathcal{O}_{p,n}$, $\forall R \in \mathcal{M}_{p,n}$, $f((U,V)\cdot R) = f(R)$. A result originally owed to Chandler Davis in [Dav57], and that can be found in a more general setting in [GH05], gives us the hint that such an invariance can help us showing that f is quasiconvex.

To present this theorem, we note $\mathcal{D}_{p,n}^+$ the set of nonnegative diagonal matrices of $\mathcal{M}_{p,n}$:

$$\mathcal{D}_{p,n}^+ = \left\{(M_{i,j})_{\substack{1\leq i\leq p\\ 1\leq j\leq n}} \in \mathcal{M}_{p,n} \mid i \neq j \Leftrightarrow M_{i,j} = 0 \text{ and } \forall i \in \{1,\ldots,d\} : M_{i,i} \geq 0\right\},$$

where $d = \min(p,n)$.

Theorem 1.2.54. *If a $\mathcal{O}_{p,n}$-invariant function $f : \mathcal{M}_{p,n} \to \mathbb{R}$ is quasiconvex on $\mathcal{D}^+_{p,n}$, then it is quasiconvex on the whole set $\mathcal{M}_{p,n}$.*

The original Davis' theorem is only concerned with symmetric matrices and we could not find any proof of this theorem for the case of rectangle matrices although it is not so different (it was also aiming at proving the *convexity* of f and not its *quasiconvexity* – but it is actually simpler to treat). We provide in Appendix C a rigorous proof of Theorem 1.2.54 with the help of some results borrowed from [Bha97].

Let us define the subgroup of row permutations $\mathcal{P}_p = \{U \in \mathcal{O}_p \mid U_{i,j} \in \{0,1\}, 1 \leq i,j \leq p\}$ and the subgroup of full permutations :

$$\mathcal{P}_{p,n} = \{(U,V) \in \mathcal{P}_p \times \mathcal{P}_n \ | UI_{p,n}V^T = I_{p,n}\} \subset \mathcal{O}_{p,n},$$

where $I_{p,n} \in \mathcal{D}_{p,n}$ is a matrix full of ones on the diagonal. Given a matrix $A \in \mathcal{M}_{p,n}$, we note $\mathrm{Diag}(A) = (A_{1,1}, \ldots, A_{q,q})$, the vector composed of its diagonal terms. It is tempting to identify the set $\mathcal{D}^+_{p,n}$ with $\mathbb{R}^d_+$, and the actions of $\mathcal{P}_{p,n}$ on a diagonal matrix A to the actions of the group of permutation $\mathfrak{S}_d$ on the vector $\mathrm{Diag}(A)$, where we define the action of $\mathfrak{S}_d$ on $\mathbb{R}^d$ as :

$$\forall \tau \in \mathfrak{S}_d, \forall x \in \mathbb{R}^d \quad : \quad \tau \cdot x = \big(x_{\tau(1)}, \cdots x_{\tau(d)}\big).$$

With these considerations in mind, we see that a direct interesting consequence of Theorem 1.2.54 is that there exists a link between the convex concentration of a matrix X and the convex concentration of the vector of its singular values. Recall that the sequence of singular values verifies $\sigma_1(A) \geq \cdots \sigma_d(A) \geq 0$ and for $1 \leq i \leq d$, the i^{th} singular value of M can be defined as :

$$\sigma_i(A) = \max_{\substack{F \subset \mathbb{R}^n \\ \dim F \geq i}} \min_{\substack{x \in F \\ \|x\|=1}} \|Mx\| \quad = \quad \min_{\substack{F \subset \mathbb{R}^n \\ \dim F \geq n-i+1}} \max_{\substack{x \in F \\ \|x\|=1}} \|Mx\| \tag{13}$$

where the subsets F of $\mathbb{R}^d$ on which is computed the optimization are subspaces of $\mathbb{R}^n$. We introduce the convenient function σ mapping a matrix to the ordered sequence of its singular values :

$$\begin{array}{rccc} \sigma \ : & \mathcal{M}_{p,n} & \to & \mathbb{R}^d_+ \\ & M & \mapsto & (\sigma_1(M), \ldots, \sigma_d(M)). \end{array}$$

To formalize this transfer of concentration between a matrix X and $\sigma(X)$ let us introduce a new notion of concentration in a vector space E : the *convex concentration transversally to the action of a group G* acting on E.

Definition 14. *Given a normal vector space E, a group G acting on E, a concentration function α and a random vector $Z \in E$, we say that Z is convexly α-concentrated transversally to the action of G and we note $Z \propto^T_G \alpha$ iff for any 1-Lipschitz, quasiconvex and G-*invariant *function $f : E \to \mathbb{R}$, $f(Z)$ is α concentrated.*

Remark 1.2.55. *In the setting of Definition 14, we have the induction chain :*

$$Z \propto \alpha \implies Z \propto_c \alpha \implies Z \propto^T_G \alpha.$$

Definition 14 is perfectly adapted to the next theorem whose proof can also be found in Appendix C (it is a similar result to in [Led01, Corollary 8.21] that concerns the eigenvalues of a random symmetric matrix).

Theorem 1.2.56. *Given a normal vector space E, a concentration function α and a random matrix $X \in \mathcal{M}_{p,n}$, we have the equivalence :*

$$X \propto^T_{\mathcal{O}_{p,n}} \alpha \iff \sigma(X) \propto^T_{\mathfrak{S}_d} \alpha.$$

Theorem 1.2.56 can be rather powerful to set the concentration of symmetric functionals of random singular values. For our present interest, we will prove the concentration of the Stieltjes transform $m_F = \frac{1}{p}\operatorname{Tr} Q$ of the covariance matrix $S = XX^T/n$ when $X \in \mathcal{M}_{p,n}$ is convexly concentrated (recall that F stands for the normalized counting measure of the eigenvalues of S).

Proposition 1.2.57. *Given a random matrix $X \in \mathcal{M}_{p,n}$, a concentration function α and $z > 0$:*

$$X \propto_c \alpha \implies \operatorname{Tr} Q \propto 2\alpha\left(\frac{\sqrt{nz^3}}{8d}\cdot\right) \quad \textit{with} \quad d = \min(p,n).$$

Let us introduce a simple preliminary lemma without proof :

Lemma 1.2.58. *Given an integer $d \in \mathbb{N}$, if a function $f : \mathbb{R} \to \mathbb{R}$ is convex, then the function $F : \mathbb{R}^d \to \mathbb{R}$ defined as $F(x_1, \dots, x_d) = \sum_{i=1}^d f(x_i)$ is also convex.*

Proof of Proposition 1.2.57. Assuming $X \propto_c \alpha$, we know from Remark 1.2.55 and Theorem 1.2.56 that $\sigma(X) \propto^T_{\mathfrak{S}_d} \alpha$, where $d = \min(p,n)$. Let us introduce the function :

$$f : s \longmapsto \frac{1}{s^2 + z}.$$

This function verifies $\sigma_i(Q) = f(\sigma_i(X)/\sqrt{n})$ and $\operatorname{Tr} Q = \sum_{i=1}^d f(\sigma_i(X)/\sqrt{n})$. It is also $\frac{2}{z^{3/2}}$-Lipschitz, for $s > 0$:

$$|f'(s)| = \frac{2s}{(s^2+z)^2}, \text{ thus } \quad f'(s) \leq \frac{2}{z^{3/2}}.$$

Therefore, the random variable $\operatorname{Tr} Q$ is a Lipschitz and $\mathfrak{S}_d$-invariant transformation of $\sigma(X)$. However, it is not quasiconvex. The result of convexity can be obtained decomposing $f = g - h$ with g and h both convex, then we will be able to conclude thanks to Lemma 1.1.5 giving the concentration of a sum of random variables. Let us set $h(s) = (\frac{s}{z} - \frac{1}{\sqrt{z}})^2$ if $s \in [0, \sqrt{z}]$ and $h(s) = 0$ if $s \geq \sqrt{z}$, and $g = f + h$. We have :

$$\text{if } s \in [0,\sqrt{z}] : \quad g''(s) = \frac{6s^2 - 2z}{(s^2+z)^3} + \frac{2}{z^2} \geq 0 \quad \text{and} \quad h''(s) = \frac{2}{z^2} \geq 0$$

$$\text{if } s \geq \sqrt{z} : \quad g''(s) = \frac{6s^2 - 2z}{(s^2+z)^3} \geq 0 \quad \text{and} \quad h''(s) = 0.$$

Besides, h is $\frac{2}{z^{3/2}}$-Lipschitz, therefore g is $\frac{4}{z^{3/2}}$-Lipschitz. We next introduce as in Lemma 1.2.58 the functions $G, H : \mathbb{R}^d \to \mathbb{R}$ defined as :

$$G(s_1, \cdots s_q) = \sum_{i=1}^{d} g(s_i) \qquad\qquad H(s_1, \cdots s_q) = \sum_{i=1}^{d} h(s_i).$$

The functions G and H are both $\mathfrak{S}_d$-invariant and convex from Lemma 1.2.58. Besides, G is $\frac{4d}{z^{3/2}}$-Lipschitz and H is $\frac{2d}{z^{3/2}}$-Lipschitz. Therefore, we know from Lemma 1.1.5 that

$$\operatorname{Tr} Q = G(\sigma(X)/\sqrt{n}) - H(\sigma(X)/\sqrt{n}) \propto 2\alpha\left(\frac{z^{\frac{3}{2}}\sqrt{n}\,\cdot}{8d}\right).$$

□

In the case of a convex concentration of X, an analogous to Proposition 1.2.43 setting the convex concentration of Q does not seem obvious, even with the help of Theorem 1.2.54.

It actually seems impossible to construct a quasiconvex *and Lipschitz* shift of $\operatorname{Tr} A\phi$ supported on the whole vector space $\mathcal{M}_{p,n}$ as we did in Proposition 1.2.57. On a bounded subset of $\mathcal{M}_{p,n}$, it is always possible to shift a Lipschitz function with a convex one so that the sum verifies quasiconvex and Lipschitz properties. If we place ourselves in a convex q-exponential concentration setting, Proposition 1.2.51 helps us treat first the concentration of $\operatorname{Tr} AQ$ for a bounded version of a random vector X and then we can take advantage of the contracting behavior of the resolvent (see Lemma 0.0.2) to generalize our first result to any unbounded concentrated vector X.

Proposition 1.2.59. *Given a random matrix $X \in \mathcal{M}_{p,n}$, $z > 0$, an exponent $q > 0$ and two parameters $C \geq e$ and $\sigma > 0$:*

$$X \propto_c \mathbb{E} \pm Ce^{-(\,\cdot\,/\sigma)^q} \quad \Longrightarrow \quad Q \in \mathbb{E}Q \pm 2Ce^{-\left(\frac{\sqrt{z^3 n}\,\cdot}{4\sigma}\right)^q} \text{ in } (\mathcal{M}_{p,n}, \|\cdot\|_F).$$

Before proving the result we formulate a preliminary lemma :

Lemma 1.2.60. *Given a symmetric nonnegative definite matrix $A \in \mathcal{M}_p$ and a matrix $B \in \mathcal{M}_p$, one has the bound :*

$$|\operatorname{Tr} AB| \leq \|B\| \operatorname{Tr} A.$$

Proof. There exits a diagonal matrix $\Lambda = \operatorname{Diag}(\lambda_i)_{1\leq i\leq p}$ and an orthogonal matrix $U \in \mathcal{O}_p$ such that $U^T AU = \Lambda$. If we note u_i the i^{th} column of U, we can write $A = \sum_{i=1}^{p} \lambda_i u_i u_i^T$, and for any $i \in \{1, \dots p\}$, $\|u_i\| = 1$, so that :

$$\operatorname{Tr} AB = \sum_{i=1}^{p} \lambda_i \operatorname{Tr}(Bu_i u_i^T) = \sum_{i=1}^{p} \lambda_i u_i^T Bu_i \leq \sum_{i=1}^{p} \lambda_i \|B\| \leq \|B\| \operatorname{Tr} A.$$

We show the other bound the same way ($\|u\| \leq 1 \Rightarrow u^T Bu \geq -\|B\|\|u\|$). □

Proof of Proposition 1.2.59. With the function ϕ introduced in Proposition 1.2.43 and given $A \in \mathcal{M}_p$ verifying $\|A\|_F \leq 1$, let us note :

$$f \ : \ R \ \longmapsto \ \operatorname{Tr} A\phi(R).$$

Given $R, H \in \mathcal{M}_{p,n}$, let us differentiate :

$$\nabla f|_R = -\phi(R)A\phi(R)R - R^T\phi(R)A\phi(R)$$
$$\nabla^2 f|_R(H,H) = 2\operatorname{Tr}\left(A\phi(R)L\phi(R)L\phi(R)\right) - 2\operatorname{Tr}(A\phi(R)HH^T\phi(R)),$$

with the notation $L = RH^T + HR^T$. Let us suppose first that A is nonnegative symmetric. In that case, we know from Lemma 1.2.60 ($\|A\| \leq \|A\|_F \leq 1$) that :

$$\operatorname{Tr}(A\phi(R)HH^T\phi(R)) \leq \frac{2}{z^2}\operatorname{Tr} HH^T,$$

and we recognize here the Hessian of the function $g : R \mapsto \frac{1}{z^2}\operatorname{Tr} RR^T$ taken in (H,H). If we note $h = f + g$ we know that

$$\nabla^2 h|_R(H,H) \geq 2\operatorname{Tr}\left(A\phi(R)L\phi(R)L\phi(R)\right) \geq 0.$$

Besides, on the set $\{R \in \mathcal{M}_{p,n}, \|R\| \leq \sqrt{z}\}$, the function g is $\frac{2}{z^{3/2}}$-Lipschitz and convex. If we suppose first that $\|X\| \leq \sqrt{n}$, we know from Lemma 1.1.5 that the sum $f(X/\sqrt{n}) = (h-g)(X/\sqrt{n})$ is concentrated :

$$\operatorname{Tr} AQ \in \operatorname{Tr} A\mathbb{E}[Q] \pm 2Ce^{-\frac{z^{3/2}\sqrt{n}}{4\sigma}\cdot}. \tag{14}$$

Now, if we suppose that there exists a constant $K \geq 1$ such that $\|X\| \leq z^{\frac{3}{2}}\sqrt{Kn}$, then we have thanks to (14)) and Lemma 1.1.3 :

$$\operatorname{Tr} AQ = \frac{1}{K}\operatorname{Tr} A\left(\frac{XX^T}{Kn} + \frac{z}{K}I_p\right)^{-1}.$$

And since $X/\sqrt{Kn} \overset{\mathbb{E}}{\propto} \alpha(\sqrt{Kn}\,\cdot)$ the concentration (14)) and Lemma 1.1.3 entail :

$$\operatorname{Tr} AQ \in \operatorname{Tr} A\mathbb{E}[Q] \pm 2Ce^{-\frac{\sqrt{nz^3}}{4\sigma}\cdot}.$$

In the general case, we consider the sequences of random matrices $X^{(m)}$ whose entries $X^{(m)}_{i,j}$, $1 \leq i \leq p$, $1 \leq j \leq n$, are defined as :

$$X^{(m)}_{i,j} = \min\left(1, \frac{\sqrt{nz^3}m}{\|X\|_F}\right)X_{i,j}.$$

By construction, $\left\|X^{(m)}\right\|_F \leq \sqrt{nz^3}m$. Besides we know from Remark 1.1.23 that $\mathbb{E}\|X\|_F < \infty$ and for m sufficiently large such that $2\mathbb{E}\|X\|_F \leq \sqrt{nz^3}m$, we know from Proposition 1.2.51 that $X^{(m)} \overset{\mathbb{E}}{\propto}_c \alpha$. Thus, as above :

$$\operatorname{Tr} AQ^{(m)} \in \operatorname{Tr} A\mathbb{E}[Q^{(m)}] \pm 2Ce^{-\frac{\sqrt{nz^3}}{4\sigma}\cdot}, \quad \text{with :} \quad Q^{(m)} = \phi\left(\frac{X^{(m)}}{\sqrt{n}}\right).$$

If we let m tend to ∞, $X^{(m)}$ tends in law to X and thus $\operatorname{Tr} AQ^{(m)}$ tends also in law to $\operatorname{Tr} AQ$ and we recover the result of the proposition thanks to Proposition 1.1.12 and Corollary 1.1.24. □

We set in Subsections 1.2 the comfortable environment where our subsequent results will easily unfold. We notably introduced most of the expressions of concentrated random variables needed in the subsequent section. We are thus now in position to provide results on the spectral distribution of sample covariance matrices of concentrated random vectors.

2. Spectral distribution of the sample covariance

2.1. Setup and notations

We consider here a general case of the example presented in the preamble where the n independent random vectors $x_1, \dots, x_n$ are distributed in k classes represented by k distributions μ_l, $1 \leq l \leq k$, supposedly different from one another.

All results presented below (with the exception of Propositions 2.2.8, Theorem 2.2.9 and Corollary 2.2.10) are valid for any choice of p and n, but they will of course gain more value when p and n are sufficiently large for the convergence to arise. This is the reason why we will call this study *quasi-asymptotic*. What we call a *constant* is supposed to be independent of the two quasi asymptotic quantities p and n and a *bounded* quantity is simply a quantity lower than a constant.

We place ourselves under an hypothesis of convex q-exponential concentration :

Assumption 1 (Concentration of X)**.** *There exist three constants $C \geq e$, $c > 0$ and $q > 0$ such that for any $m \in \mathbb{N}$, any $l \in \{1, \dots, k\}$, and any family of independent vectors $y_1, \dots, y_m$, each one following the law μ_l, we have the concentration :*

$$(y_1, \dots, y_m) \propto_c Ce^{-\cdot^q/c}$$

where the concentration occurs in the normed vector space $\mathcal{M}_{p,m}$ endowed with the Frobenius norm.

The parameters of the concentration C and c cannot be preserved throughout the different concentrations of the quantities that will be mentioned in this paper. However to lighten the expression of the result we will abusively keep the notations C and c to designate slight modifications of the original C and c by numerical constants.

Remark 2.1.1. *Assumption 1 is in particular verified for the distributions concerned by Theorems 0.0.3, 1.2.20 and 1.2.47. For $l \in \{1, \dots, k\}$, the law μ_l can respect one of the two settings :*

- *Setting* 1 *: μ_l is a pushforward of the canonical normal distribution on $\mathbb{R}^d$, or of the sphere $\mathbb{S}^{d+1}$ through the mapping of a uniformly continuous function (see Proposition 1.2.24).*

- *Setting 2 : μ_l is an affine pushforward of a product of distributions on $\mathbb{R}$ with support belonging to $[-1,1]$ (see Lemma 1.2.48).*

Note that μ_l can also be the sum of two independent distributions, each one coming from a different setting. In the literature, it is possible to find plenty of other cases of q-exponential concentration with $q \neq 2$ (and still with Euclidean norm!); their presentation goes far beyond the objectives of our paper and we invite the reader to refer to [Led01] for more information. We will see at the end of this paper that our results are still valid on different types of practical data. Thus we come to think that convex q-exponential concentration is a general hypothesis that can be adopted in a large range of applications where the data have a satisfactory entropy compared to the diameter of their distribution (the link with entropy is explored once again in [Led01]).

The behavior of the sample covariance matrix $S = \frac{1}{n}\sum_{i=1}^n x_i x_i^T = XX^T/n$ clearly depends on the frequency each distribution μ_l is drawn to form the different columns of the matrix X. To take into account this important feature, for $l \in \{1,\dots,k\}$, we introduce the set $I_l \subset \{1,\dots,k\}$ to index the random vectors x_i following a distribution μ_l ; the cardinality of the set I_l is denoted n_l. We further denote $X_{I_l} \in \mathcal{M}_{p,n_l}$ the matrix composed of the columns of X indexed by I_l. Assumption 1 gives us directly the concentrations $X_{I_l} \propto_c Ce^{-\cdot^q/c}$, that directly entail the concentration of the whole matrix X.

Assumption 2. *The number of classes k is bounded.*

With this last assumption, and Proposition B.4.1, we can state that the observable diameter of the random matrix $X = [x_1,\dots,x_n]$ is bounded.

Proposition 2.1.2. *There exist two constants $C \geq e$ and $c > 0$ such that :*

$$X \propto_c Ce^{-\cdot^q/c}.$$

With this proposition, we have access to the results of the propositions and remarks of Subsection 1.2 concerning the matrix X but also those concerning the random vectors x_i and the resolvent Q. To begin with, we can rewrite the first result of Proposition 1.2.59 to get the concentration of the Stieltjes transform of F, the spectral distribution of S :

$$m_F(z) = \frac{1}{p}\operatorname{Tr} Q(-z) \qquad \text{where} \qquad Q = Q(z) = \left(XX^T/n + zI_p\right)^{-1}.$$

We saw in Proposition 1.2.59 that $Q \in \mathbb{E}Q \pm Ce^{-(\sqrt{zn}\,\cdot)^q/c}$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$, for some constants $C \geq e$, $c > 0$. Since the linear form $M \in (\mathcal{M}_{p,n}, \|\cdot\|_F) \mapsto \frac{1}{p}\operatorname{Tr} M$ has an operator norm equal to $\frac{1}{p}\|I_p\|_F = 1/\sqrt{p}$, the linear concentration of Q implies directly the concentration of the Stieltjes transform with a tail parameter of order $\sqrt{z^3np}$.

Proposition 2.1.3. *There exist two numerical constants $C \geq e$ and $c > 0$ such that for every $z > 0$, $m_F(-z) \in \frac{1}{p}\operatorname{Tr}\mathbb{E}Q \pm Ce^{-(\sqrt{z^3pn}\,\cdot)^q/c}$.*

Now that we know that the Stieltjes transform is concentrated in $\mathbb{R}$, we expect to find a deterministic quantity localizing this concentration. The deterministic equivalent we presented in the preamble can inspire us for a choice of a deterministic equivalent in this more general setting. We need for that some notations.

Given a class l, $1 \leq l \leq k$, we introduce a generic random vector equal to one of the x_i following the law μ_l that will be noted y_l. All the y_l are supposed to be independent and we write by $\bar{y_l}$ the mean of the distribution μ_l. When computing an expectation involving the matrix X, this notation allows us to group the different terms depending only on their class $l \in \{1, \dots, k\}$ and not on their index in the matrix $i \in \{1, \dots, n\}$. We note C_{-y_l} or even abusively C_{-l} when $l \in \{1, \dots, k\}$ the sample covariance S deprived of the contribution of one vector of the l^{th} class (n_l is supposed in that case to be greater than 1). That leads to the notations $Q_{-l} = Q_{-y_l} = Q_{C_{-y_l}}$.

Finally, given $l \in \{1, \dots, k\}$, we note Σ_l the population covariance matrix of μ_l and :

$$\Sigma = \sum_{l=1}^{k} \frac{n_l}{n} \Sigma_l.$$

To complete the lacuna of the concentration notion that does not give any information on the order of a random vector X, as concentrated it could be, one needs to give restrictions to the size of the quantities $\|\bar{y_l}\|$, $1 \leq l \leq k$. The practical assumption that we need for our result is :

Assumption 3. $\forall l \in \{1, \dots, k\}$: $\frac{1}{\sqrt{p}} \mathbb{E} \|y_l\|$ *is bounded.*

However, thanks to Corollary 1.2.14, it is possible to make a simpler hypothesis concerning directly the mean $\bar{y_l}$ if we suppose that $q \geq 2$ (then $p^{1/q} \leq \sqrt{p}$).

Assumption 3 bis. $q \geq 2$ *and* $\forall l \in \{1, \dots, k\}$: $\frac{1}{\sqrt{p}} \|\bar{y_l}\|$ *is bounded.*

Assumption 3 (or 3 bis) allows us to control $\operatorname{Tr} \Sigma_l$.

Proposition 2.1.4. *For any* $l \in \{1, \dots, k\}$, $\frac{1}{p} \operatorname{Tr} \Sigma_l$ *is bounded.*

Proof. Since $\|y_l\| \in Ce^{-\cdot^q/c}$, we know from Proposition 1.1.28 that :

$$\operatorname{Tr} \Sigma_l = \mathbb{E}\left[\|y_l\|^2\right] = \left(\mathbb{E}\|y_l\|\right)^2 + \mathbb{E}\left(\|y_l\| - \mathbb{E}\|y_l\|\right)^2 \leq Cp,$$

for some constant $C > 0$. □

2.2. Estimation of the spectral distribution of the sample covariance

With the general setting presented above and assuming Assumptions 1, 2 and 3 (or 3 bis) we will show that the deterministic matrix :

$$\tilde{Q}_\delta = Q_{\Sigma_\delta} = (\Sigma_\delta + zI_p)^{-1} \quad \text{with} \quad \Sigma_\delta = \sum_{l=1}^{k} \frac{n_l}{n} \frac{\Sigma_l}{1+\delta_l} \quad \text{and} \quad z > 0$$

is a deterministic equivalent for the resolvent $Q(z)$ if $\delta = (\delta_1, \ldots, \delta_k) \in \mathbb{R}^k$ is chosen correctly. One must be careful that the notation Σ_x does not have the same meaning whether x is an integer of the set $\{1, \ldots, k\}$ or a vector of $\mathbb{R}_+^k$.

Following the calculus of the preamble in this more general case, one gets :

$$\tilde{Q}_\delta - \mathbb{E}Q = \sum_{l=1}^{k} \frac{n_l}{n} \mathbb{E}\left[\Delta_l + \epsilon_l\right]$$

$$\text{with :} \quad \begin{cases} \Delta_l = \left(\dfrac{1}{1 + y_l^T Q_{-l} y_l / n} - \dfrac{1}{1+\delta_l} \right) Q_{-y_l} y_l y_l^T \tilde{Q}_\delta \\ \epsilon_l = -\dfrac{1}{n} \dfrac{\mathbb{E}\left[Q_{-y_l} y_l y_l^T Q \Sigma_l \tilde{Q}_\delta \right]}{1+\delta_l} \end{cases}$$

where we took advantage of the independence between Q_{-l} and y_l to say that $\mathbb{E}[Q_{-y_l}\Sigma_l \tilde{Q}_\delta/(1+\delta_l)] = \mathbb{E}[Q_{-y_l} y_l y_l^T \tilde{Q}_\delta/(1+\delta_l)]$. The two matrices Δ_l and ϵ_l will be used several times in our proof arguments; we thus invite the reader to remember their definition. The form of Δ_l entices us to set $\delta = (\frac{1}{n}\operatorname{Tr}(\Sigma_l \mathbb{E}Q_{-y_l}))_{1\le l\le k}$ to show that $\tilde{Q}_\delta$ is a deterministic equivalent for Q. We will show afterwards that the same result holds for $\tilde{Q}_{\delta'}$ with $\delta' \in \mathbb{R}^+$ chosen as a solution of the system (see Proposition 2.2.7 for the validity of this definition) :

$$\delta_l' = \frac{1}{n} \operatorname{Tr}\left(\Sigma_l \left(\sum_{h=1}^{k} \frac{n_h}{n} \frac{\Sigma_h}{1+\delta_h'} + z I_p \right)^{-1} \right) \quad 1 \le l \le k.$$

The first choice δ can seem unsatisfactory because it relies on the computation of $\mathbb{E}Q_{-i}$ which is uneasy to treat. On the contrary, the second option δ' is much more interesting as it can be approximated by iteration of the fixed point equation as we will see in Proposition 2.2.7. In particular, this second choice reveals that the deterministic equivalent can be chosen in a way that it only depends on z and on the covariances and the means of the laws μ_l, $1 \le l \le k$, as will be fully explained in Remark 2.2.11.

2.2.1. Design of a first deterministic equivalent

Let us first show the concentration of the random variable $y_l^T Q_{-y_l} y_l/n$ around its mean δ_l. To simplify the concentration bounds, we introduce the real $1 > z_0 > 0$, and from now on, z is supposed to be greater than z_0.

Proposition 2.2.1. *Given $z > z_0$, there exists two numerical constants $C, c > 0$ such that :*

$$y_l^T Q_{-y_l} y_l/n \in \delta_l \pm C e^{-(z_0 n \cdot)^{\frac{q}{2}}/c} + C e^{-(\sqrt{z_0^3 n}\cdot/\bar{\gamma})^q/c}$$

where $\gamma = \frac{p}{n}$, and $\bar{\gamma} = \gamma + 1 \ge \max(\gamma, 1)$.

Proof. This proposition looks like Theorem 1.2.52 and Proposition 1.2.59 applied with independent random objects, respectively y_l and Q_{-y_l} that we know to be concentrated thanks to the initial concentration on X given by Proposition 2.1.2. The independence allows us to consider Q_{-y_l} as deterministic when we bound a probability involving y_l and conversely.

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}y_l^T Q_{-y_l} y_l - \frac{1}{n}\operatorname{Tr}(\Sigma_l \mathbb{E}Q_{-y_l})\right| \geq t\right) \\
&\quad \leq \mathbb{E}\left[\mathbb{P}\left(\left|y_l^T Q_{-y_l} y_l - \operatorname{Tr}\Sigma_l Q_{-y_l}\right| \geq \frac{nt}{2} \mid X_{-y_l}\right)\right] \\
&\qquad + \mathbb{P}\left(\left|\frac{1}{n}\operatorname{Tr}\Sigma_l(Q_{-y_l} - \mathbb{E}Q_{-y_l})\right| \geq \frac{t}{2}\right) \\
&\quad \leq Ce^{-(znt)^{\frac{q}{2}}/c} + Ce^{-(zn^{\frac{3}{2}}t/p)^q/c} + Ce^{-((zn)^{\frac{3}{2}}t/p)^q/c},
\end{aligned}$$

for some $C, c > 0$. We employed Theorem 1.2.52 together with Lemma 0.0.2 to control the variation of $y_l^T Q_{-y_l} y_l \mid X_{-y_l}$; and to control the variation of $\frac{1}{n}\operatorname{Tr}\Sigma_l Q_{-y_l}$, we employed Proposition 1.2.59 thanks to the fact that $\|\Sigma_l\|_F \leq \operatorname{Tr}\Sigma_l = \mathbb{E}\|y_l\|^2 \leq p$.

□

We are interested in bounding the first centered moments of $\frac{1}{n}y_l^T Q_{-y_l} y_l$, thus the exponent m is considered to be a constant of the problem and that simplifies the expression of the bound. Assumption 3 allows us to simplify the bound :

Proposition 2.2.2. *Given $l \in \{1, \ldots, k\}$ and $r > 0$, there exists a constant $C \geq e$ such that :*

$$\mathbb{E}\left[\left|\frac{y_l^T Q_{-y_l} y_l}{n} - \delta_l\right|^r\right] \leq C\left(\frac{\bar{\gamma}^2}{z_0^3 n}\right)^{r/2}.$$

Proof. It is a simple consequence of Proposition 1.1.28 applied to the concentration of $\frac{1}{n}y_l^T Q_{-y_l} y_l$ given by Proposition 2.2.1 :

$$\mathbb{E}\left[\left|\frac{y_l^T Q_{-y_l} y_l}{n} - \frac{1}{n}\operatorname{Tr}(\Sigma_l Q_{-y_l})\right|^r\right] \leq \left(\frac{C\bar{\gamma}^2}{z_0^3 n}\right)^{r/2} + \frac{C}{(z_0 n)^r},$$

and we recover the bound of the proposition thanks to :

$$\left(\frac{C\bar{\gamma}^2}{z_0^3 n}\right)^{r/2} \geq \frac{C}{(z_0 n)^r} \qquad \text{(recall that } z_0 \leq 1\text{)}.$$

□

To show the concentration of the Stieltjes transform around $\frac{1}{p}\operatorname{Tr}\tilde{Q}_\delta$ (we already know from Proposition 2.1.3 that it concentrates around $\frac{1}{p}\operatorname{Tr}\mathbb{E}Q$) but also to bound $\|\delta - \delta'\|$, it is important to control $\|\mathbb{E}Q - \tilde{Q}_\delta\|$. This quantity being defined as the maximum of $u^T(\mathbb{E}Q - \tilde{Q}_\delta)u$ for u on the unit sphere, we will see in Proposition 2.2.5 that we naturally need to consider the concentration of objects like $u^T\tilde{Q}_\delta y_l$ and $u^T Q_{-y_l} y_l$.

Proposition 2.2.3. *Given $z \geq z_0$, there exist two numerical constants $C \geq e$ and $c > 0$ such that for any $l \in \{1, \ldots, k\}$ and for any $u \in \mathbb{R}^p$ of unit norm, we have the concentrations :*

$$u^T \tilde{Q}_\delta y_l \propto C e^{-(z_0 \cdot)^q/c} \qquad \text{and} \qquad u^T Q_{-y_l} y_l \propto C e^{-(\frac{z_0^3 \cdot}{\bar{\gamma}})^q},$$

where as above, $\gamma = \frac{p}{n}$ and $\bar{\gamma} = 1 + \gamma$. Moreover, there exists a numerical constant C such that for any $r \geq 1$:

$$\mathbb{E}\left[\left|u^T \tilde{Q}_\delta y_l\right|^r\right] \leq C\left(\frac{n\bar{\gamma}}{n_l z_0^2}\right)^{\frac{r}{2}} \quad \text{and} \quad \mathbb{E}\left[\left|u^T Q_{-y_l} y_l\right|^r\right] \leq C\left(\frac{n\bar{\gamma}}{n_l z_0^3}\right)^{\frac{r}{2}}.$$

The relation between y_l and Q_{-y_l} is characterized by the multiple appearance of vectors x_i following the law μ_l in the matrix X_{-y_l}. Similarly, $\tilde{Q}_\delta$ lets appear in its definition the matrix Σ_l characteristic of the distribution of y_l. These relations allow us to bound the moments $\mathbb{E}[|u^T \tilde{Q}_\delta y_l|^r]$ and $\mathbb{E}[|u^T Q_{-y_l} y_l|^r]$, and explains the appearance of the coefficient $\frac{n_l}{n}$. Without these structural relations we could consider instead of y_l any random vector y with an expected norm of order $\sqrt{p}$. In that case $\mathbb{E}u^T Q y$ and $\mathbb{E}u^T Q_{-y_l} y$ would be rather of order $\sqrt{p}$ like the norm of y. The next preliminary lemma explains how this control can be realized when we deal with the deterministic equivalent $\tilde{Q}_\delta$.

Lemma 2.2.4. *Given $l \in \{1, \ldots, k\}$, $\left\|\tilde{Q}_\delta^{\frac{1}{2}} \Sigma_l \tilde{Q}_\delta^{\frac{1}{2}}\right\| \leq \frac{n}{n_l} \frac{\bar{\gamma}}{z_0}$.*

Proof. The result is similar to the result of Lemma 0.0.2. Recall the notation $\Sigma_\delta = \sum_{h=1}^k \frac{n_h}{n} \frac{\Sigma_h}{1+\delta_h}$ and the definition $\tilde{Q}_\delta = (\Sigma_\delta + zI_p)^{-1}$. With the order relation on the set of symmetric matrices, we deduce from the inequality $\frac{n_h}{n} \frac{\Sigma_l}{1+\delta_l} \leq \Sigma_\delta + zI_p$ that $\tilde{Q}_\delta^{\frac{1}{2}} \frac{n_l}{n} \frac{\Sigma_l}{1+\delta_l} \tilde{Q}_\delta^{\frac{1}{2}} \leq I_p$ (see [Bha97, Lemma V.1.5]). We know from Lemma 1.2.60 and Assumption 3 that :

$$|\delta_l| = \frac{1}{n} \left|\operatorname{Tr} \mathbb{E} Q_{-y_l} \Sigma_l\right| \leq \frac{1}{n} \mathbb{E} \left\|Q_{-y_l}\right\| \operatorname{Tr} \Sigma_l \leq \frac{\gamma}{z_0}. \tag{15}$$

We can then conclude thanks to the bound $(1+\delta_l) \leq \frac{\bar{\gamma}}{z_0}$ (recall that $z_0 \leq 1$). □

Proof of Proposition 2.2.3. Let us introduce the bilinear form :

$$\begin{array}{rcl} f : \mathcal{M}_p \times \mathbb{R}^p & \longrightarrow & \mathbb{R} \\ (M, y) & \longmapsto & u^T M y. \end{array}$$

We know that $f(\tilde{Q}_\delta, y_l) = u^T \tilde{Q} y_l$ is a $\frac{1}{\sqrt{z_0^3 n}}$-Lipschitz linear form on y_l, it thus concentrated as y_l. Since $Q_{-y_l} \in \mathbb{E}Q_{-y_l} + e^{-(\sqrt{z_0 n} \cdot)^q/c}$ and $y_l \in C e^{-\cdot^q/c}$ are independent, and $\|Q_{-y_l}\| \leq 1/z_0$, we know from Proposition 1.2.9 that :

$$\begin{aligned} u^T Q_{-y_l} y_l &\in u^T \mathbb{E}\left[Q_{-y_l} y_l\right] \pm C e^{-(z_0^3 \cdot)^q/c} + C e^{-(\frac{z_0^3 \cdot}{\bar{\gamma}})^q/c} \\ &\subset u^T \mathbb{E}\left[Q_{-y_l}\right] \bar{y}_l \pm C e^{-(\frac{z_0^3 \cdot}{\bar{\gamma}})^q/c} \end{aligned}$$

for some constant parameters $C \geq e$ and $c > 0$ since $\mathbb{E}\,\|y_l\| \leq \sqrt{p}$ (the inclusion between combination of q-exponential concentration functions is defined in Subsection 1.1.2).

Now that we set the concentrations, we know from Corollary 1.1.25 that for any $r > 0$ there exists a constant C such that :

$$\mathbb{E}\left[\left|u^T\tilde{Q}(y_l - \bar{y}_l)\right|^r\right] \leq \frac{C}{z_0^r}$$
$$\mathbb{E}\left[\left|u^T\left(Q_{-y_l}y_l - \mathbb{E}Q_{-y_l}\bar{y}_l\right)\right|^r\right] \leq \frac{C\bar{\gamma}^{r/2}}{z_0^{3r/2}}.$$

Therefore, we know from the convexity of $t \mapsto t^r$ (because $r \geq 1$) that :

$$\mathbb{E}\left[\left|u^T\tilde{Q}y_l\right|^r\right] \leq 2^{r-1}\left(\frac{C}{z_0^r} + \left|u^T\tilde{Q}\bar{y}_l\right|^r\right)$$
$$\left|u^TQ_{-y_l}\bar{y}_l\right|^r \leq 2^{r-1}\left(\frac{C\bar{\gamma}^{r/2}}{z_0^{2r}} + \left|u^T\mathbb{E}Q_{-y_l}\bar{y}_l\right|^r\right).$$

We thus conclude that all the centered moments are bounded as soon as we find a real $s \geq 1$ such that $\mathbb{E}[|u^T\tilde{Q}y_l|^s]$ is bounded since we know from Jensen's inequality that $|u^T\tilde{Q}\bar{y}_l|^r \leq \mathbb{E}[|u^T\tilde{Q}y_l|^s]^{\frac{r}{s}}$. This is the case of $\mathbb{E}u^T\tilde{Q}y_ly_l^T\tilde{Q}u = u^T\tilde{Q}\Sigma_l\tilde{Q}u \leq \frac{n\bar{\gamma}}{n_lz_0^2}$ (see Lemma 2.2.4). The case of $u^TQ_{-y_l}y_l$ is harder to control and we need to employ the Schur formula (1) :

$$\begin{aligned}\mathbb{E}\left[\left|u^TQ_{-y_l}y_l\right|\right] &= \mathbb{E}\left[\left|u^TQy_l\right|(1 + y_lQ_{-y_l}y_l/n)\right]\\ &\leq \sqrt{\mathbb{E}\left[u^TQy_ly_l^TQu\right]\mathbb{E}\left[(1 + y_lQ_{-y_l}y_l/n)^2\right]}\\ &\leq \frac{\bar{\gamma}}{z_0}\frac{n}{n_l}\mathbb{E}\left[\frac{1}{n}u^TQX_{I_l}X_{I_l}^TQu\right] \;\leq\; \frac{n}{n_l}\frac{\bar{\gamma}}{z_0^2},\end{aligned}$$

where we recall that $X_{I_l} \in \mathcal{M}_{p,n_l}$ is the matrix composed of the columns of X indexed by I_l (of cardinality n_l). The inequalities are consequences of Proposition 2.2.1 and Lemma 0.0.2. □

Proposition 2.2.5. *There exist a constant $C > 0$ such that :*

$$\left\|\mathbb{E}Q - \tilde{Q}_\delta\right\| \leq \frac{C\,\bar{\gamma}^2}{z_0^4\sqrt{n}}.$$

Proof. As already mentioned, it is sufficient to bound for any vector $u \in \mathbb{R}^p$ of unit norm the quantities $|u^T(\mathbb{E}Q - \tilde{Q}_\delta)u|$ since $\mathbb{E}Q$ and $\tilde{Q}_\delta$ are both symmetric matrices. Recall from the heuristic approach displayed at the beginning of this subsection the set matrices Δ_l and ϵ_l verifying :

$$u^T(\mathbb{E}Q - \tilde{Q}_\delta)u = \sum_{l=1}^k \frac{n_l}{n}\left(\left|u^T\Delta_lu\right| + \left|u^T\epsilon_lu\right|\right).$$

Thanks to Hölder's inequality and Properties 2.2.2, 2.2.5, we can bound :

$$\begin{aligned}\left|u^T\Delta_l u\right| &= \left|\mathbb{E}\left[u^T Q_{-y_l} y_l^T y_l \tilde{Q}_\delta u \frac{\delta_l - y_l^T Q_{-y_l} y_l/n}{(1+\delta_l)\left(1+y_l^T Q_{-y_l} y_l/n\right)}\right]\right| \\ &= \mathbb{E}\left[\left|u^T Q_{-y_l} y_l^T\right| \left|y_l \tilde{Q}_\delta u\right| \left|\delta_l - y_l^T Q_{-y_l} y_l/n\right|\right] \;\leq\; \frac{n_l}{n}\frac{C\bar{\gamma}^2}{z_0^4\sqrt{n}}.\end{aligned}$$

For the same reasons :

$$\left|u^T \epsilon_l u\right| \leq \frac{1}{n}\mathbb{E}\left[\left|u^T Q_{-y_l} y_l^T\right| \left|y_l Q_{-y_l}\Sigma_l \tilde{Q}_\delta u\right|\right] \;\leq\; \frac{n_l}{n}\frac{C\bar{\gamma}}{nz_0^3}\ .$$

Therefore if we sum on $l \in \{1,\ldots,k\}$, we see that there exists a constant $C \geq e$ such that :

$$\forall u \in \mathbb{R}^p \quad : \quad u^T\left(\mathbb{E}Q - \tilde{Q}_\delta\right)u \leq \frac{Ck\bar{\gamma}^2}{z_0^4\sqrt{n}}\ ,$$

and that gives us directly the bound on $\|\mathbb{E}Q - \tilde{Q}_\delta\|$ thanks to Assumption 2 (k is bounded). □

This last proposition together with Proposition 1.2.59 and Lemma 1.2.8 implies that $\tilde{Q}_\delta$ is a deterministic equivalent for Q.

Proposition 2.2.6. *$Q \in \tilde{Q}_\delta \pm Ce^{-(z_0^4\sqrt{n}\,\cdot/\bar{\gamma}^2)^q/c}$ in $(\mathcal{M}_{p,n}, \|\cdot\|)$.*

Proof. We know from Proposition 1.2.59 that $Q \in \mathbb{E}Q \pm Ce^{-(\sqrt{zn}\,\cdot)^q/c}$ in $(\mathcal{M}_{p,n}, \|\cdot\|_F)$ so in particular :

$$Q \in \mathbb{E}Q \pm Ce^{-(z_0^4\sqrt{n}\,\cdot/\bar{\gamma}^2)^q/c} \ \text{ in } (\mathcal{M}_{p,n}, \|\cdot\|).$$

But since $\|\mathbb{E}Q - \tilde{Q}_\delta\| \leq \frac{C\bar{\gamma}^2}{z_0^4\sqrt{n}}$, Lemma 1.2.8 entails the result of the proposition. □

2.2.2. A second deterministic equivalent

Let us introduce a substitute for δ that we note δ' and that will only depend on the means and covariances of the laws μ_l and on the cardinality coefficients $\frac{n_l}{n}$ of the different classes.

Proposition 2.2.7 (Definition of δ')**.** *The system of equations :*

$$\forall l \in \{1,\ldots,k\} \ : \qquad \delta_l' = \frac{1}{n}\operatorname{Tr}\left(\Sigma_l\left(\sum_{h=1}^k \frac{n_h}{n}\frac{\Sigma_h}{1+\delta_h'} + zI_p\right)^{-1}\right)$$

admits a unique solution in $\mathbb{R}_+^k$ that we note δ'.

Proof. The scheme of the proof follows the formalism of standard interference functions as presented in [Yat95]. Following the ideas of Yates, we introduce the function :

$$\begin{array}{rccc} I & : & \mathbb{R}_+^k & \longrightarrow & \mathbb{R}_+^k \\ & & (\delta_l')_{1\leq l\leq k} & \longmapsto & \dfrac{1}{n}\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta'}\right), \end{array}$$

where $\tilde{Q}_x = (\Sigma_x + zI_p)^{-1}$, and $\Sigma_x = \sum_{l=1}^k \frac{n_l}{n}\frac{\Sigma_l}{1+x_l}$ for $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$. Given $x, y \in \mathbb{R}^k$, we note $x \leq y$ iff $\forall l \in \{1, \ldots, k\}$, $x_l \leq y_l$. The function I is increasing in the sense that :

$$x \leq y \quad \Longrightarrow \quad I(x) \leq I(y).$$

This is simply due to the fact that for any $l \in \{1, \ldots, k\}$, Σ_l is symmetric nonnegative definite and $x \mapsto \tilde{Q}_x$ is increasing (with the classical order relation defined on the set of symmetric matrices see [Bha97, Section V] for more details). Besides, we know from Lemmas 0.0.2 and 1.2.60 that for any $x \in \mathbb{R}^k$, $I(x) \leq \frac{\operatorname{Tr}\Sigma_l}{nz}$. Therefore the vector

$$x_0 = \left(\frac{\operatorname{Tr}\Sigma_1}{nz}, \cdots, \frac{\operatorname{Tr}\Sigma_k}{nz}\right)$$

verifies $I(x_0) \leq x_0$. Then the monotonicity of I implies that the sequence $(I^n(x_0))_{n\geq 0}$ is decreasing and since I takes its values in $\mathbb{R}_+$, it converges to a fixed point $\delta' \geq 0$.

Let us now show the uniqueness of this fixed point. Let us suppose that there exists $\nu \neq \delta'$ such that $I(\nu) = \nu$. Given $l \in \{1, \cdots, k\}$, we have the identity :

$$\nu_l - \delta_l' = \frac{1}{n}\operatorname{Tr}\left(\sum_{h=1}^k \frac{n_h}{n}\frac{\gamma_h - \delta_h'}{(1+\delta_h')(1+\gamma_h)}\Sigma_l \tilde{Q}_{\delta'}\Sigma_h \tilde{Q}_\nu\right),$$

and if we introduce the vector $\epsilon = (\epsilon_l)_{1\leq l\leq k}$ defined as $\epsilon_l = \frac{|\nu_l - \delta_l'|}{\sqrt{(1+\nu_l)(1+\delta_l')}}$:

$$\begin{aligned} |\varepsilon_l| &= \frac{1}{n}\operatorname{Tr}\left(\sum_{h=1}^k \frac{n_h}{n}\frac{\Sigma_l^{\frac{1}{2}}\tilde{Q}_{\delta'}\Sigma_h^{\frac{1}{2}}}{\sqrt{(1+\delta_h')(1+\delta_l')}}\frac{\Sigma_h^{\frac{1}{2}}\tilde{Q}_\nu\Sigma_l^{\frac{1}{2}}}{\sqrt{(1+\gamma_h)(1+\nu_l)}}|\varepsilon_h|\right) \\ &\leq \sqrt{\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l\tilde{Q}_{\delta'}\Sigma_{\delta'}\tilde{Q}_{\delta'}\right)}{1+\delta_l'}}\quad\sqrt{\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l\tilde{Q}_\nu\Sigma_\nu\tilde{Q}_\nu\right)}{1+\nu_l}}\quad \|\varepsilon\|_\infty \end{aligned} \tag{16}$$

where for any $x \in \mathbb{R}_+^k$, we recall that $\Sigma_x = \sum_{h=1}^k \frac{n_h}{n}\frac{\Sigma_h}{1+x_h}$. Recall that by definition of the resolvent $\tilde{Q}_x = Q_{\Sigma_x}$, we have the identity

$$\tilde{Q}_x\Sigma_x + z\tilde{Q}_x = I_p.$$

This allows us to write for any $x \in \mathbb{R}_+$ such that $I(x) = (\frac{1}{n}\operatorname{Tr}(\Sigma_l \tilde{Q}_x))_{1\leq l\leq k} = x$:

$$\frac{1}{n}\operatorname{Tr}\left(\Sigma_l \tilde{Q}_x \Sigma_x \tilde{Q}_x\right) = x_l - \frac{z}{n}\operatorname{Tr}\left(\Sigma_l \tilde{Q}_x^2\right) \quad < \quad 1 + x_l.$$

Therefore, we know from the inequality (16), true for every $l \in \{1, \ldots, k\}$, that $\|\varepsilon\|_\infty = 0$, in other words $\delta' = \nu$, the fixed point is unique. □

Proposition 2.2.8. *If $\bar{\gamma}$ and $\frac{1}{z_0}$ are bounded and n is large enough, there exists a constant $C > 0$ such that :*

$$\|\delta - \delta'\|_\infty \leq \frac{C}{\sqrt{n}} \qquad \textit{and} \qquad \left\|\tilde{Q}_\delta - \tilde{Q}_{\delta'}\right\| \leq \frac{C}{\sqrt{n}}.$$

Proof. Let us employ as in the proof of Proposition 2.2.7 a vector $\varepsilon = \delta - \delta'$. Given $l \in \{1, \ldots, k\}$, we compute :

$$\begin{aligned}
|\varepsilon_l| &= \frac{1}{n}\left|\operatorname{Tr}\Sigma_l(\mathbb{E}Q - \tilde{Q}_{\delta'})\right| \\
&\leq \gamma\left\|\mathbb{E}Q - \tilde{Q}_\delta\right\| + \frac{1}{n}\left|\operatorname{Tr}\Sigma_l(\tilde{Q}_\delta - \tilde{Q}_{\delta'})\right| \\
&\leq \frac{C\,\bar{\gamma}^3}{z_0^4\sqrt{n}} + \sqrt{\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta'}\Sigma_{\delta'}\tilde{Q}_{\delta'}\right)}{1+\delta_l'}}\ \sqrt{\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta}\Sigma_{\delta}\tilde{Q}_{\delta}\right)}{1+\delta_l}}\ \|\varepsilon\|_\infty,
\end{aligned}$$

where we employed for the first inequality the result of Lemma 1.2.60 and in the second inequality the intermediate result (16)) of the proof of Proposition 2.2.7. We already know that :

$$\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta'}\Sigma_{\delta'}\tilde{Q}_{\delta'}\right)}{1+\delta_l'} < 1 \quad \text{and} \quad \frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta}\Sigma_{\delta}\tilde{Q}_{\delta}\right)}{1+\delta_l} \leq \frac{\delta_l}{1+\delta_l} + \frac{C\,\bar{\gamma}^3}{z_0^4\sqrt{n}},$$

and we know from (15) that $\frac{\delta_l}{1+\delta_l} \leq 1 - \frac{z_0}{\bar{\gamma}}$. Therefore since $\frac{z_0}{\bar{\gamma}}$ is bounded from below and $\frac{C\bar{\gamma}^3}{z_0^4}$ is bounded from above, if n is large enough :

$$\frac{1}{n}\frac{\operatorname{Tr}\left(\Sigma_l \tilde{Q}_{\delta}\Sigma_{\delta}\tilde{Q}_{\delta}\right)}{1+\delta_l} < 1 \qquad \text{and thus} \qquad \|\varepsilon\|_\infty \leq \frac{C\,\bar{\gamma}^3}{z_0^4\sqrt{n}}.$$

Now, let us bound the spectral norm of the difference $\tilde{Q}_\delta - \tilde{Q}_{\delta'}$:

$$\left\|\tilde{Q}_\delta - \tilde{Q}_{\delta'}\right\| \leq \sum_{l=1}^k \frac{n_l}{n}\frac{|\delta_l - \delta_l'|}{(1+\delta_l)(1+\delta_l')}\left\|\tilde{Q}_{\delta'}\Sigma_l\tilde{Q}_\delta\right\| \leq \frac{\|\delta - \delta'\|_\infty}{z_0} \quad \leq \quad \frac{C\,\bar{\gamma}^3}{z_0^5\sqrt{n}},$$

for some $C \geq e$. □

We now have all the elements to set the linear concentration of Q around the second deterministic equivalent $\tilde{Q}_{\delta'}$ with the same arguments as those given to justify Proposition 2.2.6.

Theorem 2.2.9. *If $\bar{\gamma}$ and $\frac{1}{z_0}$ are bounded and n is large enough, there exist two numerical constants $C \geq e$ and $c > 0$ such that :*

$$Q \in \tilde{Q}_{\delta'} \pm Ce^{-(\sqrt{n}\,\cdot)^q/c} \quad in \ (\mathcal{M}_{p,n}, \|\cdot\|)\,.$$

As for Proposition 2.1.3, we can directly deduce that $\frac{1}{p}\operatorname{Tr}\tilde{Q}_\delta$ is a pivot of the Stieltjes transform. This time the linear form $M \mapsto \frac{1}{p}\operatorname{Tr} M$ is seen as a 1-Lipschitz transformation from $(\mathcal{M}_{p,n}, \|\cdot\|)$ to $\mathbb{R}$ (see Lemma 1.2.60).

Corollary 2.2.10 (Estimation of the Stieltjes transform)**.** *In the setting of Theorem 2.2.9, for $z > 0$:*

$$m_F(-z) \in \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z) \pm Ce^{-(\sqrt{n}\,\cdot)^q/c}.$$

Remark 2.2.11 (Central limit theorem for covariance matrices)**.** *Let us define :*

$$X^{\mathcal{N}} = (\Sigma_{k(1)}x_1^{\mathcal{N}} + \bar{y}_{k(1)}, \cdots, \Sigma_{k(n)}x_n^{\mathcal{N}} + \bar{y}_{k(n)})$$

where $x_1^{\mathcal{N}}, \cdots x_n^{\mathcal{N}}$ are independent Gaussian vectors with zero mean and unit variance entries and for a given $i \in \{1, \ldots, n\}$, $k(i) \in \{1, \ldots, k\}$ designates the class of x_i.

From the definition of δ' and $\tilde{Q}_{\delta'}$ we can remark that the deterministic equivalent $\tilde{Q}_{\delta'}$ is the same for a resolvent Q constructed with the sample covariance of X or of the matrix $X^{\mathcal{N}}$. This implies that the asymptotic spectral distribution of the sample covariance matrix of X strictly depends on the means and the covariances of the laws μ_l, $1 \leq l \leq k$, but not at all on the intrinsic distribution of those laws. In that sense, the Gaussian case describes all the possible asymptotic spectral distributions of sample covariances of any concentrated data respecting our two assumptions.

In a future paper this remark will be the key idea to show a theorem of central limit for some quantities depending on the resolvent (with applications to random neural networks, such as extreme learning machines [HZS06]).

2.3. Illustration of the results

In this section, we are interested in the asymptotic spectral distribution of S, in the sense that if we index with n all the quantities characteristic to the matrix X (that we write for instance X_n), we wish here to express the limit of the spectral distributions F_n when the number of data n tends to ∞. The size p of the data can then be seen as a function p_n of n that cannot grow too fast to maintain the validity and the strength of our results. Corollary 2.2.10 requires in particular $\gamma = \frac{p_n}{n}$ to be bounded, and this will be from now on a supplementary assumption. Regarding the convergence, it appears convenient to work with the notion of the *weak convergence* of distributions that was essentially presented in Definition 5 through the convergence in law of random variables. We shall

say indeed that a sequence of measures $(H_n)_{n\in\mathbb{N}}$ tends weakly to a measure H if for any function $f : \mathbb{R} \mapsto \mathbb{R}$ with compact support :

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(t)dH_n(t) = \int_{\mathbb{R}} f(t)dH(t),$$

and we write $H_n \xrightarrow{w} H$.

Now, it is important to keep in mind that the distributions F_n are random objects and, therefore, we will be looking for a weak convergence *almost surely*. Let us note $\mathcal{X}$ the sequence of random matrices $(X_n)_{n\in\mathbb{N}}$ that we suppose to be independent. If we note $\sigma(\mathcal{X}) = \sigma(X_1) \times \cdots \sigma(X_n) \times \cdots$, the sigma-algebra generated by $\mathcal{X}$, we look for an event (i.e., an element of $\sigma(\mathcal{X})$) of probability 1 on which happens the weak convergence. As we could expect from the different results of the last sections, the convergence of the spectral distribution is a consequence of the convergence of the *Stieltjes transforms*. This is a classical results of the field of random matrices, we provide in Appendix D a proof inspired from the paper [CDS11].

Theorem 2.3.1. *With the above notations and hypotheses, if we suppose that there exists a complex function $m : \mathbb{C} \mapsto \mathbb{C}$ such that :*

$$\forall z > 0 \quad : \quad \frac{1}{p_n} \operatorname{Tr} \tilde{Q}_n(z) \longrightarrow m(z),$$

then there exists a measure H in $\mathbb{R}$ such that :

$$\forall z \in D,\ m(z) = \int_{\mathbb{R}} \frac{dH(t)}{t-z} \qquad \text{and a.s. :} \qquad F_n \xrightarrow{w} H.$$

(Recall that $\tilde{Q}_n$ is the matrix $\tilde{Q}_{\delta'}$ indexed with the number of data n)

If we go back to the original task we presented in the preamble, we need to evaluate the limit distribution H appearing in the result of Theorem 2.3.1. We can be helped by a basic proposition of the theory of large random matrices :

Proposition 2.3.2. *Given a probability distribution H in $\mathbb{R}$, we note $m_H : z \mapsto \int_{\mathbb{R}} \frac{dH(t)}{t-z}$, the Stieltjes transform of H. Given $t, s \in \mathbb{R}$, we have the identities :*

- $H(\{t\}) = \lim_{y\to 0^+} y\mathrm{Im}[m_H(x+iy)]$
- $H([t,s]) = \frac{1}{\pi} \lim_{y\to 0^+} \int_t^s \mathrm{Im}[m_H(x+iy)]dx.$

In Figure 3, we compare the spectral distribution of different sample covariance matrices with their asymptotic profile as described by Theorem 2.3.1. Inspired by Proposition 2.3.2, we basically computed $\frac{1}{p} \operatorname{Tr} \tilde{Q}_{\delta'}(x+iy)$ for small values of y to obtain the density of F. We see that the prediction is valid for different values of $\gamma = \frac{p}{n}$ and for data constructed as a mixture between Gaussian and Bernoulli random vectors.

The range of validity of our results seems to be rather wide since one can see on Figure 4 that our predictions are rather good for "raw" data as the one of the MNIST data base (left graph) or even more refined data as some feature vectors

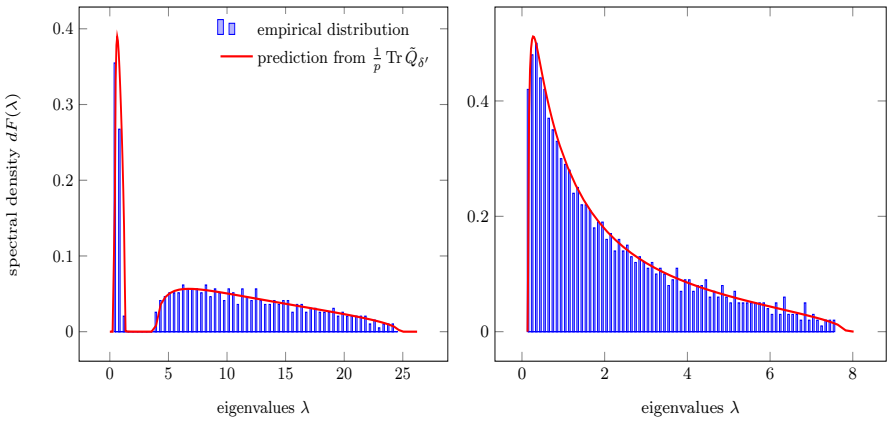

FIG 3. *Spectral distribution of the empirical covariance matrix of a sum of Gaussian and Bernoulli data. On the left,* $n = 2000$, $p = 500$ *and* Σ *has two eigenvalues :* 1 *and* 12 *with the repartition* $(\frac{1}{4}, \frac{3}{4})$. *On the right,* $p = 2000$, $n = 1000$ *and* Σ *is a symmetric Toeplitz matrix whose first line is* $(0.4, 0.4^2 \dots 0.4^p)$ *; since* $p > n$, *we removed the* $p - n$ *zero eigenvalues of* S *to simplify the graphs.*

of traffic sign images (right graph). Be careful that we drew the distribution of a push-forward of the spectral distribution dilated around zero, otherwise the distribution would be huddled against 0 and at the same time sparsely spread up to high values. The refined feature dataset satisfies our prediction better that the MNIST dataset. This statement allows us to think that in the concentration of the measure framework, the treatment of the signal though the diverse neurons tends to "normalize" the data.

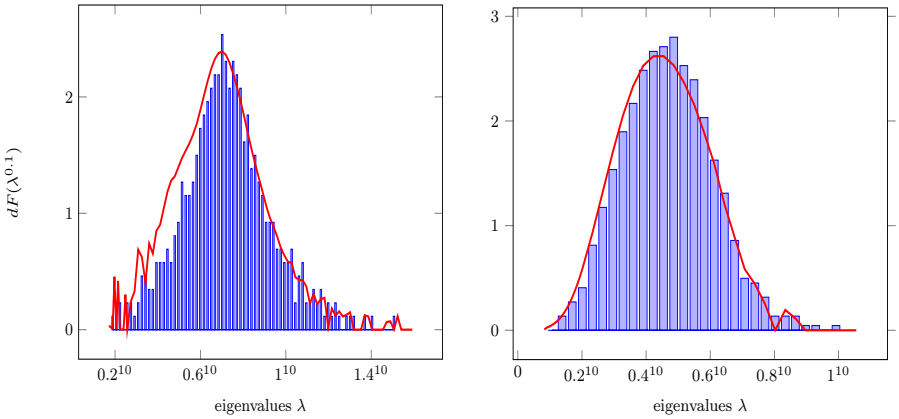

FIG 4. *Distribution of the push forward of the spectral distribution of the covariance matrix of real datasets through the map* $\lambda \mapsto \lambda^{0.1}$ *; the population covariance and the mean are empirically computed from a larger and independent dataset. On the left, images the digit* 0 *of the MNIST data set ([LCB98]);* $p = n = 784$*. On the right, "HOG" features (Histograms of Oriented Gradients) of a class of traffic sign images presented in the paper [SSSI11] ;* $p = n = 700$ *(just the first* p *features are considered)*

Appendices

Appendix A: Concentration of the product and powers of random variables

Proposition A.0.1. *Let us consider a pivot* $a \in \mathbb{R}$*, a concentration function* α *and* p *random variables* $Z_1, \dots, Z_p$*, such that for any* $i \in \{1, \dots, p\}$*,* $Z_i \in a \pm \alpha$*. Then the product* $Z_1 \cdots Z_p$ *is concentrated in the sense of Definition 3, and we have :*

$$Z_1 \cdots Z_p \in a^p \pm p\alpha\left(\frac{\cdot}{2^p\,|a|^{p-1}}\right) + p\alpha\left(\left(\frac{\cdot}{2}\right)^{\frac{1}{p}}\right)$$

Note that in the case of q-exponential concentration, if, say, $\alpha = Ce^{-(\cdot/\sigma)^q}$, $C \geq e$, $\sigma, q > 0$, we can employ Proposition 1.1.20 and obtain $\forall r \geq q$ the bound :

$$\mathbb{E}\left[|Z_1 \cdots Z_p - a^p|^r\right] \leq Cp(2|a|^{p-1}\sigma)^r \left(\frac{r}{q}\right)^{\frac{r}{q}} + Cp(2\sigma^p)^r \left(\frac{rp}{q}\right)^{\frac{rp}{q}}$$

which is looser than the one obtained in Proposition 1.1.28 (when $r \in \mathbb{N}$).

Proof. Let us bound :

$$Z_1 \cdots Z_p - a^p = a^p \sum_{l=1}^{p} \sum_{\substack{I \subset \{1,\dots,p\} \\ \#I = l}} \prod_{i \in I} \left(\frac{Z_i - a}{a}\right) \leq a^p \sum_{l=1}^{p} \binom{p}{l} \left|\frac{Z_s - a}{a}\right|^l,$$

where Z_s is the random variable verifying $\left|\frac{Z_s-a}{a}\right| = \sup_{1\leq i\leq p}\left|\frac{Z_i-a}{a}\right|$. As in the proof of Proposition 1.1.10 distinguishing the cases $\left|\frac{Z_s-a}{a}\right| \geq 1$ and $\left|\frac{Z_s-a}{a}\right| \leq 1$, we obtain the bound :

$$|Z_1 \cdots Z_p - a^p| \leq (2|a|)^p \left(\left|\frac{Z_s-a}{a}\right| + \left|\frac{Z_s-a}{a}\right|^p\right),$$

and we are left to express the concentration around 0 of the random variable $|Z_s - a|$:

$$\mathbb{P}(|Z_s-a| \geq t) \leq \sum_{i=1}^{p} \mathbb{P}(|Z_i-a| \geq t) \leq p\alpha(t)$$

and we recover the result. □

One can be interested in the concentration of Z^r when r is not an integer. If we write $r = \frac{r}{\lceil r\rceil}\lceil r\rceil$ where $\lceil r\rceil$ is the integer part of r plus 1, we see that thanks to Proposition 1.1.10 our problem can be reduced to the concentration of Z^r when $0 < r \leq 1$. That can be managed easily thanks to the following lemma :

Lemma A.0.2. *Given $a, b \in \mathbb{R}$ and $r \in [0,1]$, $|a|^r - |b|^r \leq |a-b|^r$.*

Proof. When $a, b > 0$, it is just a consequence of the triangular inequality verified by the $\ell^{\frac{1}{r}}$-norm on $\mathbb{R}$. If a or b are negative, it is just a consequence again of the triangular inequality of the absolute value : $||a| - |b|| \leq |a-b|$. □

Interestingly, when $r \leq 1$, the concentration of Z^r is easier to show in the formalism of Definition 2. Indeed we can employ Lemma 1.1.3 since we know from Lemma A.0.2 that $t \to t^r$ is $(1,r)$-Hölder continuous.

Proposition A.0.3. *Given an exponent $r \in (0,1]$ and a concentration function α, if a non-negative random variable $Z \geq 0$ is α-concentrated in the sense of Definition 2 then $Z^r \propto \alpha(\cdot^{\frac{1}{r}})$.*

Remark A.0.4. *Thanks to Proposition 1.1.2, we can adapt the result of Proposition A.0.3 to the formalism of Definition 3. This gives us when $r \in (0,1]$, with the upper notations :*

$$Z \in m_Z \pm \alpha \quad \Longrightarrow \quad Z^r \in m_Z^r \pm 4\alpha\left(\frac{\cdot^{\frac{1}{r}}}{2}\right),$$

where m_Z is a median of Z (note that implicitly mentioned in the upper implication, m_Z^r is a median of Z^r since the function $t \mapsto t^r$ is monotonous).

One can then look at the concentration of any power Z^r where $r \in \mathbb{R}_+^*$:

$$Z \in m_Z \pm \alpha \quad \Rightarrow \quad Z^r \in m_Z^r \pm 4\alpha\left(\frac{1}{2^{\lceil r\rceil+1}}\left(\frac{\cdot}{|m_Z|^{\lceil r\rceil-1}}\right)^{\frac{r}{\lceil r\rceil}}\right) + 4\alpha\left(\frac{1}{2}\left(\frac{\cdot}{2}\right)^{\frac{1}{r}}\right),$$

where $\lceil r\rceil$ designates the integer part of r plus 1.

Appendix B: Results of linear concentration

B.1. Proof of Proposition 1.2.10

Given $\varepsilon \in (0,1)$, a set $A \subset \mathcal{B}_H$ is said to be an ε-net of $\mathcal{B}_H$ if $x, y \in A \Rightarrow \|x-y\| \geq \varepsilon$. We consider here $N_{1/2}$, a maximal $\frac{1}{2}$-net of $\mathcal{B}_H$ with respect to inclusion. We know that the balls of radius $\frac{1}{4}$ centered on the points of $N_{1/2}$ are all disjoint by hypothesis, and their volume is equal to $\mathcal{V}_{\mathcal{B}_H}/4^p$ (where $\mathcal{V}_{\mathcal{B}_H}$ is the volume of $\mathcal{B}_H$). Since they all belong to the ball of radius 2 and centered at the origin, we know that their number cannot exceed $8^{\dim(H)}$.

Besides, given a drawing of Z, there exists $f_0 \in \mathcal{B}_H$ such that $Z - \tilde{Z} = f_0(Z - \tilde{Z})$ (since $\mathcal{B}_H$ is compact). Then there exists $f \in \mathcal{B}_H$ such that $\|f - f_0\|_*$ is bounded by $\frac{1}{2}$ (otherwise f_0 could be added to $N_{1/2}$). Furthermore :

$$\begin{aligned}\left\|Z-\tilde{Z}\right\| - f\left(Z-\tilde{Z}\right) &\leq \left|f\left(Z-\tilde{Z}\right) - f_0\left(Z-\tilde{Z}\right)\right| \\ &\leq \|f-f_0\|_* \left\|Z-\tilde{Z}\right\| \;\leq\; \frac{1}{2}\left\|Z-\tilde{Z}\right\|.\end{aligned}$$

Therefore :

$$\left\|Z-\tilde{Z}\right\| \leq 2\sup\left\{f \in N_{1/2} \;:\; u^T\left(Z-\tilde{Z}\right)u\right\}$$

and this inequality being true for any drawing of Z, we have then by hypothesis :

$$\forall t>0 \;:\; \mathbb{P}\left(\left\|Z-\tilde{Z}\right\| \geq t\right) \leq \sum_{N_{1/2}} \alpha\left(\frac{t}{2}\right) \leq 8^{\dim(H)}\alpha\left(\frac{t}{2}\right).$$

B.2. Linear concentration of the product

We place ourselves in an algebra $\mathcal{A}$ endowed with an algebra norm $\|\cdot\|$ (verifying $\|xy\| \leq \|x\|\,\|y\|$). To simplify the result we place ourselves in the exponential concentration setting, the reader is required to adapt the proof for generalization if needed.

Proposition B.2.1. *Given two random vectors $X, Y \in \mathcal{A}$ and three parameters $C \geq e$ and $\sigma, q > 0$, if X and Y follow the same concentration $X \in \tilde{X} \pm Ce^{-(\cdot/\sigma)^q}$ and $Y \in \tilde{Y} \pm Ce^{-(\cdot/\sigma)^q}$, then XY is also concentrated :*

$$XY \in \tilde{X}\tilde{Y} \pm C\exp\left(-\left(\frac{c\cdot}{\sigma^2\eta_{\|\cdot\|}^{\frac{2}{q}}}\right)^{\frac{q}{2}}\right) + C\exp\left(-\left(\frac{c\cdot}{\sigma(\|\tilde{X}\|+\|\tilde{Y}\|)}\right)^q\right),$$

where c is a numerical constant independent of C and σ.

Proof. As for Lemma 1.1.8, we employ the identity :

$$XY - \tilde{X}\tilde{Y} = (X-\tilde{X})(Y-\tilde{Y}) + \tilde{Y}(X-\tilde{X}) + \tilde{Y}(Y-\tilde{Y})$$

For any linear function u with an operator norm bounded by 1, we have :

$$u(XY - \tilde{X}\tilde{Y}) \leq \|X - \tilde{X}\|\|Y - \tilde{Y}\| + u_{\tilde{Y}}(X - \tilde{X}) + u_{\tilde{Y}}(Y - \tilde{Y})$$

where for any $z \in \mathcal{A}$ we defined $u_z : x \mapsto u(zx)$. To conclude, we just have to note that u_z is a linear function the operator norm of which is bounded by $\|z\|$ (since $\|u(zx)\| \leq \|u\| \, \|z\| \, \|x\|$). □

B.3. Linear concentration of the power

Let us adapt Proposition 1.1.10 to get the linear concentration of a power of a random vector, we leave to the reader the expression of the concentration of the product of m linearly concentrated random vectors.

Proposition B.3.1. *Given $m \in \mathbb{N}_*$, a random vector $Z \in \mathcal{A}$, a deterministic vector $\tilde{Z} \in \mathcal{A}_*$ and three parameters $C \geq e$, $\sigma, q > 0$, if we suppose that $Z \in \tilde{Z} \pm Ce^{-(\cdot/\sigma)^q}$, then :*

$$Z^m \in \tilde{Z}^m \pm C \exp\left(-\left(\frac{c\,\cdot}{2^m \sigma \eta_{\|\cdot\|}^{\frac{1}{q}} \|\tilde{Z}\|^{m-1}}\right)^q\right) + \exp\left(-c\left(\frac{\cdot}{2\sigma^m \eta_{\|\cdot\|}^{\frac{m}{q}}}\right)^{\frac{q}{m}}\right),$$

where c is a numerical constant depending only on q and m.

Proof. As in the proof of Proposition 1.1.10, we bound :

$$\begin{aligned}
\left\|Z^m - \tilde{Z}^m\right\| &= \left\|\sum_{k=1}^{m} \sum_{i_1+\ldots+i_k \leq m-k} (Z - \tilde{Z})^{i_1} \tilde{Z} \ldots (Z - \tilde{Z})^{i_k} \tilde{Z} (Z - \tilde{Z})^{m-i_k-\ldots-i_1}\right\| \\
&\leq \|\tilde{Z}\|^m \sum_{k=1}^{m} \binom{m}{k} \left(\frac{\|Z - \tilde{Z}\|}{\|\tilde{Z}\|}\right)^k \\
&\leq (2\|\tilde{Z}\|)^m \left(\frac{\|Z - \tilde{Z}\|}{\|\tilde{Z}\|} + \frac{\|Z - \tilde{Z}\|^m}{\|\tilde{Z}\|^m}\right).
\end{aligned}$$

Since we know from Proposition 1.2.12 that $\|Z - \tilde{Z}\| \in 0 \pm Ce^{-(c\,\cdot/\sigma)^q/\eta_{\|\cdot\|}}$ for some numerical constant $c > 0$, we get the concentration :

$$\left\|Z^m - \tilde{Z}^m\right\| \in 0 \pm C \exp\left(-\left(\frac{c/2^m\,\cdot}{\sigma \eta_{\|\cdot\|}^{\frac{1}{q}} \|\tilde{Z}\|^{m-1}}\right)^q\right) + C \exp\left(-\left(\frac{c^m\,\cdot}{2\sigma^m \eta_{\|\cdot\|}^{\frac{m}{q}}}\right)^{\frac{q}{m}}\right)$$

and we get the same concentration for Z thanks to the second result of Proposition 1.2.12. □

B.4. Concatenation of convexly concentrated random vectors

Through the characterization with the centered moments given by Proposition 1.1.20, the q-exponential concentration allows to explore the concentration of a random vector constructed as a concatenation of p independent random vectors $(Z_1, \dots, Z_p)$. This approach is indifferently adapted to the Lipschitz or convex concentration. We use the index (c) under the sign $\propto$ to specify that the proposition is valid in both settings (for Lipschitz *or* convexly concentrated random vectors).

Proposition B.4.1. *Given p normed vector spaces $(E_1, N_1), \dots, (E_p, N_p)$, consider p independent random vectors $(Z_1, \dots, Z_p) \in E = E_1 \times \cdots \times E_p$ verifying for any $i \in \{1, \dots, p\}$ that $Z_i \propto_{(c)} Ce^{-(\cdot/\sigma)^q}$, for two given parameters $C \geq e$, $\sigma > 0$. The space E can then be seen as a normed vector space endowed with the norm $\|\cdot\|_{\ell_1}$ defined as :*

$$\forall z = (z_1, \dots, z_p) \in E_1 \times \cdots \times E_p \quad : \quad \|z\|_{\ell_1} = N_1(z_1) + \ldots + N_p(z_p).$$

Then the concatenation $Z = (Z_1, \dots, Z_p)$ is q-exponentially concentrated in $(E, \|\cdot\|_{\ell_1})$ with an observable diameter lower than $p\sigma e^{\frac{1}{q}}$:

$$Z \propto_{(c)} Ce^{-(\cdot/p\sigma)^q/e}.$$

Proof. Let us consider a function $f : E \to \mathbb{R}$, 1-Lipschitz (resp. 1-Lipschitz and quasiconvex), and Z' an independent copy of Z. We plan to employ the characterization with the centered moments given by Proposition 1.1.20. Given $i \in \{0, \dots, p\}$, we note $Z^{(i)} = (Z_1, \dots Z_i, Z'_{i+1}, \dots, Z'_p)$ (with this notation : $Z^{(0)} = Z$ and $Z^{(p)} = Z'$). For any $r \geq \max(q, 1)$, let us exploit the convexity of $t \mapsto t^r$ to bound :

$$\mathbb{E}\left[|f(Z) - f(Z')|^r\right] \leq p^{r-1} \sum_{i=1}^{p} \mathbb{E}\left[\left|f\left(Z^{(i-1)}\right) - f\left(Z^{(i)}\right)\right|^r\right].$$

Therefore, since for any $(z_1, \dots, z_p) \in \mathbb{R}^p$ and for any $i \in \{1, \dots, p\}$, $z \mapsto f(z_1, \dots, z_{i-1}, z, z_{i+1}, \dots z_p)$ is Lipschitz (resp. Lipschitz and quasiconvex), we can employ Proposition 1.1.20 to bound :

$$\mathbb{E}\left[|f(Z) - f(Z')|^r\right] \leq Cp^r \left(\frac{r}{q}\right)^{\frac{r}{q}} \sigma^r.$$

If $q \leq r \leq 1$, the concavity of $t \mapsto t^r$ allows us to write thanks to Jensen's inequality :

$$\mathbb{E}\left[|f(Z) - f(Z')|^r\right] \leq \left(\mathbb{E}\left[|f(Z) - f(Z')|\right]\right)^r \leq C^r p^r \left(\frac{1}{q}\right)^{\frac{r}{q}} \sigma^r \leq \; p^r \left(\frac{r}{q}\right)^{\frac{r}{q}} \sigma^r$$

since $r \leq 1 \leq C^q$. The last implication of Proposition 1.1.20 then gives us the desired result: $f(Z) \propto Ce^{-(\cdot/p\sigma)^q/e}$. □

Appendix C: Davis theorem for rectangle matrices

C.1. Proof of Theorem 1.2.54

Let us first present basic notions to set the theorem. Given a vectorial space E and a group G acting on E, for any subset $A \subset E$, we note $G \cdot A = \{g \cdot a, g \in G, a \in A\}$. We say that a set T is *transversal* if $G \cdot T = E$ and we say that a function f is G-invariant if $\forall x \in E, \forall g \in G, f(g \cdot x) = f(x)$. In the same vein, we say that a set $A \subset E$ is *G-invariant* if $G \cdot A = A$ and that it is *G-invariant in T* if $A \subset T$ and $G \cdot A \cap T = A$. Given $U \subset T$, we note $L_T^G(U)$ the smallest convex subset of T containing U and G-invariant in T. We give here an adaptation of one of the result of Grabovsky and Hijab to the case of quasiconvex functions.

Theorem C.1.1 (cf. [GH05], Theorem 4)**.** *Let us consider a vector space E, a group G acting on E, and a* convex *and* transversal *subset $T \subset E$. We suppose that for any $U \subset T$, the set $G \cdot L_T^G(U)$ is convex ; we call this property the* convexity conservation of G from T. *Then any G-invariant function f is quasiconvex iff its restriction to T is quasiconvex.*

Intuitively, this theorem states that the quasiconvexity (or the mere convexity as in [GH05]) of any function G-invariant is "transversal" to the action of G when G preserves the convexity from T. A curious reader might be interested in simplifying the convexity conservation property as we presented it from a transversal subset $T \subset E$ to a mere conservation of the convexity of any convex subset $U \subset E$ (i.e., for any convex set $U \subset E$, $f(U)$ is convex). This would be indeed an hypothesis more than sufficient for the result of the theorem. However, in practice, and in particular for the applications we want to consider, it cannot be verified.

Proof. Let us note note $f_{|T}$ the restriction of f on T. Given any $t \in \mathbb{R}$, we know that the set $\{f_{|T} \leq t\} = \{x \in T, f(x) \leq t\}$ is convex. Then the set $\{f \leq t\} = G \cdot \{f_{|T} \leq t\} = G \cdot L_T^G(\{f_{|T} \leq t\})$ is also convex thanks to the convexity conservation of G from T. □

To simplify the application of Theorem C.1.1, Grabowsky and Hijab provide us with a useful property :

Proposition C.1.2 (Convexity conservation, [GH05], Theorem 3)**.** *With the notations of Theorem C.1.1, if for any $x, y \in T$, the set $G \cdot L_T^G(\{x, y\})$ is convex, then G conserves the convexity from T.*

Proof. Let us consider $U \subset T$, and two points $x, y \in G \cdot L_T^G(U)$. There exists $x^*, y^* \in T$ such that $x, y \in G \cdot \{x^*, y^*\}$. We know by hypothesis that the set $G \cdot L_T^G(\{x^*, y^*\})$ is convex, and moreover, it contains x and y. Therefore, any element of the segment $[x, y]$ is also in $L_T^G(\{x^*, y^*\}) \subset L_T^G(U)$. □

In the case of a matrix concentration, Theorem C.1.1 can be applied to the transversal set of nonnegative diagonal matrices $\mathcal{D}_{p,n}^+$ for the action of the group $\mathcal{O}_{p,n}$. The transversal character of $\mathcal{D}_{p,n}^+$ is a consequence of the singular value

decomposition. Indeed, for any matrix $M \in \mathcal{M}_{p,n}$, there exists $(U, V) \in \mathcal{O}_{p,n}$ such that

$$M = U\Sigma V^T \quad \text{with} \quad \Sigma = \mathrm{Diag}_{p,n}(\sigma_i(M))_{1\leq i\leq d},$$

where $d = \min(p, n)$, $\sigma_i(M)$ is the i^{th} singular value of M and the notation $\mathrm{Diag}_{p,n}(a_i)$ represents an element of $\mathcal{D}^+_{p,n}$ having the values a_i on the diagonal.

To prove Theorem 1.2.54, let us characterize the sets $L(X) = L^{\mathcal{O}_{n,p}}_{\mathcal{D}^+_{p,n}}(\{X\})$ when $X \in \mathcal{D}^+_{p,n}$. We know that $\mathcal{D}^+_{p,n}$ is invariant under the action of the subgroup of permutations $\mathcal{P}_{p,n}$. Given a subset U of a vector space, we note $\mathrm{Conv}(U)$ the convex hull of U.

Proposition C.1.3. *Given $X \in \mathcal{D}_{p,n}$, $L(X) = \mathrm{Conv}(\mathcal{P}_{p,n} \cdot \{X\})$.*

Proof. We know from the uniqueness of the singular value decomposition that for any $U \subset \mathcal{D}^+_{p,n}$, $(\mathcal{O}_{p,n} \cdot U) \cap \mathcal{D}^+_{p,n} = \mathcal{P}_{p,n} \cdot U$. Consequently, since the convexity is stable under the action of $\mathcal{P}_{p,n}$ (they are linear transformations), $L(X) = \mathrm{Conv}(\mathcal{P}_{p,n} \cdot \{X\})$. □

Here the tools of *majorization* as presented for instance in [Bha97] are perfectly adapted to the description of $\mathrm{Conv}(\mathfrak{S}_{p,n} \cdot \{x\})$ that we identify with $\mathrm{Conv}(\mathcal{P}_{p,n} \cdot \{X\}) = L(X)$. Given a vector $x = (x_1, \dots x_d) \in \mathbb{R}^d$, let us note $x^\downarrow = (x^\downarrow_1, \cdots x^\downarrow_d)$, a decreasing ordered version of x ($x^\downarrow_1 \geq \cdots x^\downarrow_d$ and $\exists \sigma \in \mathfrak{S}_d \mid x^\downarrow = \sigma \cdot x$).

Definition 15. *Given two vectors $x, y \in \mathbb{R}^d$, $d \in \mathbb{N}$, we say that y is majorized by x and we note $y \prec x$ iff :*

$$\forall k \in \{1, \dots d\} \ : \ \sum_{i=1}^{k} y^\downarrow_i \leq \sum_{i=1}^{k} x^\downarrow_i \qquad \text{and} \qquad \sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i.$$

Majorization offers a complete characterization of $\mathrm{Conv}(\mathfrak{S}_d \cdot \{x\})$, $x \in \mathbb{R}^d$:

Theorem C.1.4 ([Bha97], Theorem II.1.10)**.** *Given a vectors $x \in \mathbb{R}^d$:*

$$\left\{y \in \mathbb{R}^d, y \prec x\right\} = \mathrm{Conv}\left(\mathfrak{S}_d \cdot \{x\}\right).$$

Majorization appears to be the pertect tool to control the singular decomposition of a sum of matrices as we will see in Theorem C.1.6. This is the core argument to justify the convexity consevation of $\mathcal{O}_{p,n}$ from $\mathcal{D}^+_{p,n}$ that we need to prove Theorem 1.2.54. Let us first give an intermediate result that we originally owe to Schur and whose proof can be found in [MOA11] or [Bha97].

Proposition C.1.5 (Schur's Theorem, B.1. in [MOA11])**.** *Given a symmetric matrix $S \in \mathcal{M}_p$, we have the majorization* $\mathrm{Diag}(S) \prec \sigma(S)$.

This proposition entails a kind of triangular inequality for the set of singular values.

Theorem C.1.6 ([Bha97], Exercise II.1.15)**.** *Given two matrices $A, B \in \mathcal{M}_{p,n}$, $\sigma(A + B) \prec \sigma(A) + \sigma(B)$.*

Of course, it is important that the vectors $\sigma(A)$ and $\sigma(B)$ are both ordered when we sum them.

Proof. Given a symmetric matrix $S \in \mathcal{M}_q$, there exists $(U_S, V_S) \in \mathcal{O}_{p,n}$ such that $U_A A V_A^T = \mathrm{Diag}_{p,n}\, \sigma(S)$, thus :

$$\sum^k \mathrm{Diag}(U_S S V_S^T) = \sum^k \sigma(S)$$

where given $x \in \mathbb{R}^d$ and $k \in \{1, \dots d\}$, $\sum^k x = \sum_{i=1}^k x_i^{\downarrow}$. Besides, since $\sigma(USV^T) = \sigma(S)$, we know from Proposition C.1.5 that :

$$\sum^k \sigma(S) = \sup_{(U,V)\in\mathcal{O}_{p,n}} \sum^k \mathrm{Diag}(USV^T). \tag{17}$$

Now, if we suppose that we are given a general matrix $A\mathcal{M}_{p,n}$ and $(U_A, V_A) \in \mathcal{O}_{p,n}$ such that $U_A A V_A^T = \mathrm{Diag}_{p,n}\, \sigma(A)$. If we introduce the matrices :

$$\tilde{A} = \begin{pmatrix} (0) & A \\ A^T & (0) \end{pmatrix} \in \mathcal{M}_{p+n}, \qquad \text{and} \qquad P = \begin{pmatrix} U & (0) \\ (0) & V \end{pmatrix} \in \mathcal{M}_{p+n}$$

we have the identity :

$$P\tilde{A}P^T = \begin{pmatrix} (0) & D \\ D & (0) \end{pmatrix} \in \mathcal{M}_{p+n}, \qquad \text{with} \quad D = \mathrm{Diag}_{p,n}(\sigma(A)).$$

Depending on the relation between p and n, we introduce the invertible matrices :

$$\text{if } d = n : Q = \begin{pmatrix} I_d & (0) & I_d \\ (0) & I_d & (0) \\ -I_d & (0) & I_d \end{pmatrix} \quad \text{and if } d = p : Q = \begin{pmatrix} I_d & I_d & (0) \\ -I_d & I_d & (0) \\ (0) & (0) & I_d \end{pmatrix},$$

then if $d = n$, $QP\tilde{A}(PQ)^T = \mathrm{Diag}(\sigma(A), 0 \cdots 0, -\sigma(A))$ and if $d = p$, $QP\tilde{A}(PQ)^T = \mathrm{Diag}(\sigma(A), -\sigma(A), 0 \cdots 0)$. Thus in both cases, we obtain a diagonalisation of $\tilde{A}$ that allows us to generalize the identity (17)) for any matrix $A \in \mathcal{M}_{p,n}$ and with $0 \leq k \leq d$. The supremum of a sum being lower than the sum of a supremum, for any pair of matrices $A, B \in \mathcal{M}_{p,n}$:

$$\sigma(A + B) \prec \sigma(A) + \sigma(B).$$

□

Now that the picture is clearer, we can prove Theorem 1.2.54 :

Proof of Theorem 1.2.54. To employ Theorem C.1.1, let us show the convexity conservation property of $\mathcal{O}_{p,n}$ from $\mathcal{D}^+_{p,n}$. Inspired by Proposition C.1.2, we consider two non-negative diagonal matrices $X, Y \in \mathcal{D}^+_{p,n}$, and we note $L(X,Y) = L^{\mathcal{D}^+_{p,n}}_{\mathcal{O}_{p,n}}(\{X, Y\})$. We want to show that $\mathcal{O}_{p,n} \cdot L(X,Y)$ is convex.

Noting $x = \operatorname{Diag} X$ and $y = \operatorname{Diag} Y$, let us first show that $L(X,Y) = K(x,y)$, with :

$$K(x,y) = \{Z \in \mathcal{D}^+_{p,n}, \operatorname{Diag} Z \prec \lambda x + (1-\lambda)y,\ 1 \leq \lambda \leq 1\}.$$

We know from Theorem C.1.4 that for any $U \in \mathcal{D}^+_{p,n}$, $L(U) = \{\operatorname{Diag} Z \prec \operatorname{Diag} U\}$ and therefore, since for any $t \in [0,1]$, $tX + (1-t)Y \in L(X,Y)$, we obtain the first inclusion $K(x,y) \subset L(X,Y)$.

To prove the converse inclusion, let us show that $K(x,y)$ is convex (we already know that $\mathcal{O}_{p,n} \cdot K(x,y) \cap \mathcal{D}^+_{p,n} = \mathcal{P}_{p,n} \cdot K(x,y) = K(x,y)$ by definition of the relation $\prec$).

We consider $A, B \in K(x,y)$, $t \in [0,1]$ and we set $C = tA+(1-t)B$. Therefore, We know that there exists $t_z, t_w \in [0,1]$ such that $\operatorname{Diag} A \prec \lambda x + (1-\lambda)y$ and $\operatorname{Diag} B \prec \mu x + (1-\mu)y$, therefore :

$$\operatorname{Diag} C \prec (t\lambda + (1-t)\mu)\ x \quad + \quad (t(1-\lambda) + (1-t)(1-\mu))\ y \in [x,y].$$

In conclusion, since $X, Y \in K(x,y)$ and $K(x,y)$ is $\mathcal{P}_{p,n}$-invariant and convex we recover the second inclusion $K(x,y) \subset L(X,Y)$.

Thus we are left to show that $\mathcal{O}_{p,n} \cdot K(x,y)$ is convex. We consider this time $A, B \in \mathcal{O}_{p,n} \cdot K(x,y)$, $t \in [0,1]$, and we introduce $C = tA + (1-t)B$. We know from Theorem C.1.6 that :

$$\sigma(C) \prec t\sigma(A) + (1-t)\sigma(B),$$

and as we saw before, that implies that $\operatorname{Diag}\sigma(C) \in K(x,y)$. We can then conclude with the relations

$$C \in \mathcal{O}_{p,n} \cdot \{\operatorname{Diag}\sigma(C)\} \subset \mathcal{O}_{p,n} \cdot K(x,y) = \mathcal{O}_{p,n} \cdot L(X,Y).$$

We can apply Proposition C.1.2 to get the hypothesis of Theorem C.1.1 that entails Theorem 1.2.54 in our setting. □

C.2. Proof of Theorem 1.2.56

Let us first show the Lipschitz character of σ, it is a well known result that can be found for instance in [GL96] :

Lemma C.2.1 (Theorem 8.1.15 in [GL96])**.** *The function σ is 1-Lipschitz.*

Proof. Given $M, H \in \mathcal{M}_{p,n}$, and $i \in \{1, \dots d\}$ (where as before, $d = \min(p,n)$, we know from formula (13)) that :

$$\begin{aligned}
\lambda_i(A+H) &= \min_{\substack{F \subset \mathbb{R}^n \\ \dim F \geq n-i+1}} \max_{\substack{x \in F \\ \|x\|=1}} \|(A+H)x\| \\
&\leq \min_{\substack{F \subset \mathbb{R}^n \\ \dim F \geq n-i+1}} \left(\max_{\substack{x \in F \\ \|x\|=1}} \|Ax\| + \max_{\substack{x \in F \\ \|x\|=1}} \|Hx\| \right) \\
&\leq \min_{\substack{F \subset \mathbb{R}^n \\ \dim F \geq n-i+1}} \max_{\substack{x \in F \\ \|x\|=1}} \|Ax\| + \max_{\substack{x \in \mathbb{R}^n \\ \|x\|=1}} \|Hx\| \quad \leq \quad \lambda_i(A) + \lambda_1(H),
\end{aligned}$$

and the same way, we can show that $\lambda_i(A+H) \geq \lambda_i(A) - \lambda_n(H)$. Therefore, we get :

$$|\lambda_i(A+H) - \lambda_i(A)| \leq \max(\lambda_1(H), \lambda_n(H)) \leq \|H\|.$$

□

Proof of Theorem 1.2.56. It is a simple corollary of Theorem 1.2.54. If $X \propto^T_{\mathcal{O}_{p,n}} \alpha$, and given a 1-Lipschitz, convex and $\mathfrak{S}_n$-invariant function $f : \mathbb{R}^d \to \mathbb{R}$, one can introduce the function $\tilde{F}$ defined as :

$$\begin{array}{rccc} \tilde{F} : \mathcal{M}_{p,n} & \longrightarrow & \mathbb{R} \\ M & \longmapsto & f(\sigma(M)). \end{array}$$

The function F is 1-Lipschitz thanks to Lemma C.2.1 and $\mathcal{O}_{p,n}$-invariant because of the uniqueness of the singular decomposition. Besides we can identify the set $\mathcal{D}^+_{p,n}$ with $\mathbb{R}^d$ and introduce a function $\tilde{f} : \mathcal{D}^+_{p,n} \to \mathbb{R}$ verifying $\tilde{f}(\mathrm{diag}_{p,n}(x)) = f(x)$ for $x \in \mathbb{R}^d$. In that case, since f is $\mathfrak{S}_d$-invariant and convex, $\tilde{f}$ is also convex and since $\tilde{f} = F|_{\mathcal{D}^+_{p,n}}$, Theorem 1.2.54 allows us to set that F is also convex. Therefore, the random variable $f(\sigma(X)) = F(X)$ is α-concentrated by hypothesis on X.

Reciprocally, let us suppose that we are given a random matrix $X \in \mathcal{M}_{p,n}$ such that $\sigma(X) \propto^T_{\mathfrak{S}_d} \alpha$, and let us consider a 1-Lipschitz, convex and $\mathcal{O}_{p,n}$-invariant function $F : \mathcal{M}_{p,n} \to \mathbb{R}$. The restriction $F|_{\mathcal{D}^+_{p,n}}$ is also 1 Lipschitz, convex and $\mathcal{P}_{p,n}$-invariant. Thus, with the same identification as before between $\mathcal{D}^+_{p,n}$ and $\mathbb{R}^d$, we can assert by hypothesis that $F(X) = \tilde{F}|_{\mathcal{D}^+_{p,n}}(\sigma(X))$ is α-concentrated (we defined $\tilde{F}|_{\mathcal{D}^+_{p,n}}(x) = F|_{\mathcal{D}^+_{p,n}}(\mathrm{diag}_{p,n}(x))$, it is a $\mathfrak{S}_d$-invariant function). □

Appendix D: Proof of Theorem 2.3.1 : convergence of the spectral distribution of the sample covariance

Corollary 2.2.10 states that for any $z > z_0$:

$$\mathbb{P}\left(\left|m_{F_n}(-z) - \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z)\right| \geq t\right) \leq Ce^{-(\sqrt{n}\,\cdot)^q/c}.$$

Then if we introduce the events of $\sigma(\mathcal{X})$:

$$A_n = \left\{\left|m_{F_n}(-z) - \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z)\right| \geq \frac{1}{n^{1/4}}\right\},$$

we know that

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} Ce^{-n^{q/4}/c} < \infty.$$

Thus the Borel-Cantelli lemma ensures that

$$\mathbb{P}\left(\cap_{n\in\mathbb{N}} \cup_{k\geq n} A_n\right) = 0.$$

Therefore, a.s., for any $\varepsilon > 0$, there exists an integer $n_0 \geq 1$ such that for any $n \geq n_0$:

$$\left| m_{F_n}(-z) - \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z) \right| \leq \varepsilon.$$

In other words, a.s., the Stieltjes transform of F_n taken in z converges to $\frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z)$, and the next theorem justifies how this convergence can be exported to the sequence of spectral distributions F_n.

Theorem D.0.1 ([BCL09], Theorem B.9)**.** *Let H_n be a sequence of probability distribution on $\mathbb{R}$ and $m : \mathbb{C} \mapsto \mathbb{C}$, a complex function. If there exists a set $D \subset \mathbb{R}$ with an accumulation point such that :*

$$\forall z \in D \quad : \quad m_{H_n}(z) \longrightarrow m(z),$$

then there exists a measure H in $\mathbb{R}$ such that :

$$\forall z \in D,\ m(z) = \int_{\mathbb{R}} \frac{dH(t)}{t-z} \qquad \text{and} \qquad H_n \xrightarrow{w} H.$$

For any $k \in \mathbb{N}^*$, let us note $z_k = z_0 + \frac{1}{k}$, and A_k the event of probability one "$m_{F_n}(z_k) - \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}(z_k) \to 0$". We know that the event $A = \cap_{k\geq 1} A_k$ has probability one as a countable intersection of events of probability one. Therefore, if we consider the set $D = \{z_0 + \frac{1}{k}, k \in \mathbb{N}\}$, we know that a.s., for all $z \in D$, $m_{F_n}(z_k) - \frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'} \to 0$. To be able to exploit Theorem D.0.1, and prove Theorem 2.3.1 one ultimately needs to show that $\frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}$ is a Stieltjes transform. This is given by the next proposition.

Proposition D.0.2. *$\frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta}$ and $\frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}$ are two Stieltjes transforms.*

The proof exploits an argument already exposed in [CDS11].

Proof. Given $\theta \in \mathbb{R}^k$, we introduce for any $m \in \mathbb{N}$, $\tilde{Q}_\theta^{(m)} = \left(\Sigma_\theta^{(m)} + zI_p\right)^{-1}$, where we set the block matrix :

$$\Sigma_\theta^{(m)} = \begin{pmatrix} \Sigma_\theta & & (0) \\ & \ddots & \\ (0) & & \Sigma_\theta \end{pmatrix} \in \mathcal{M}_{pm}.$$

We clearly have $\operatorname{Tr}\tilde{Q}_\theta = \frac{1}{m}\operatorname{Tr}\tilde{Q}_\theta^{(m)}$, the idea of the proof is to show that if $\theta = \delta$ or $theta = \delta'$, $\frac{1}{mp}\operatorname{Tr}\tilde{Q}_\theta^{(m)}$ converges to a Stieltjes transform and so does $\frac{1}{p}\operatorname{Tr}\tilde{Q}_\theta$.

Let us expand our notations to our random objects, we note $Q^{(m)} = Q_{S^{(m)}}(z)$, where :

$$S^{(m)} = \frac{X_{(m)}X_{(m)}^T}{n} \quad \text{and} \quad X_{(m)} = \begin{pmatrix} X^{(1)} & & (0) \\ & \ddots & \\ (0) & & X^{(m)} \end{pmatrix} \in \mathcal{M}_{pm,nm},$$

where $X^{(1)}, \dots, X^{(m)}$ are m independent copies of X. We know that $X_{(m)}$ and all its intrinsic parameters verify the different hypothesis of Proposition 2.2.6 and we have :

$$Q^{(m)} \in \tilde{Q}_\delta^{(m)} \pm Ce^{-(z_0^4\sqrt{nm}\cdot/\bar{\gamma}^2)^q/c} \quad \text{in } (\mathcal{M}_{pm,nm}, \|\cdot\|),$$

note that we implicitly used the identity

$$\delta_l^{(m)} = \frac{1}{mn}\operatorname{Tr}\left(\Sigma_l^{(m)}\mathbb{E}Q_{-l}^{(m)}\right) = \frac{1}{n}\operatorname{Tr}\left(\Sigma_l\mathbb{E}Q_{-l}\right) = \delta_l. \tag{18}$$

As in Corollary 2.2.10, it can be shown that :

$$m_{F^{(m)}}(-z) \in \frac{1}{p}\operatorname{Tr}\tilde{Q}_\delta \pm Ce^{-(z_0^4\sqrt{nm}\cdot/\bar{\gamma}^2)^q/c},$$

where $F^{(m)}$ is the spectral distribution of $S^{(m)}$. Therefore, if we let m tend to infinity, for reasons that we exposed above this Proposition, we know thanks to the first result of Theorem D.0.1, that $\frac{1}{p}\operatorname{Tr}\tilde{Q}_\delta$ is a Stieltjes transform as a limit of Stieltjes transforms on a set with an accumulative point (e.g. $[z_0, \infty)$). To treat the case of $\frac{1}{p}\operatorname{Tr}\tilde{Q}_{\delta'}$, one needs a similar result to (18) – it is immediate considering the definition of δ'.

□

References

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

[BCL09] Z. D. Bai, Y. Chen, and Y. C. Liang. *Random matrix theory and its applications*, volume 18. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, 2009.

[Bha97] Rajendra Bhathia. *Matrix Analysis*. Springer, Graduate texts in mathematics, 1997.

[CB16] R. Couillet and F. Benaych-Georges. Kernel spectral clustering of large dimensional data. *Electronic Journal of Statistics*, 10(1):1393–1454, 2016.

[CDS11] R. Couillet, M. Debbah, and J. W. Silverstein. A deterministic equivalent for the analysis of correlated MIMO multiple access channels. 57(6):3493–3514, June 2011.

[Dav57] Chandler Davis. All convex invariant functions of hermitian matrices. *Archiv der Mathematik 8*, pages 276–278, 1957.

[El 09] N. El Karoui. Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, 19(6):2362–2405, 2009.

[El 10] N. El Karoui. The spectrum of kernel random matrices. *The Annals of Statistics*, 38(1):1–50, 2010.

[GH05] Yury Grabovsky and Omar Hijab. A generalization of the chandler davis convexity theorem. *Advances in Applied Mathematics 34*, pages 192–212, 2005.

[GL96] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins University press, 1996.

[Gro79] Mikhail Gromov. Paul lévy's isoperimetric inequality. *Preprint IHES*, 1979.

[Gro99] Mikhail Gromov. Metric structures for riemannian and non-riemannian spaces. In *Progress in Math. 152*. Birkhäuser, Boston, 1999.

[HZS06] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

[KC17] Abla Kammoun and Romain Couillet. Subspace kernel clustering of large dimensional data. *(submitted to) Journal of Machine Learning Research*, 2017.

[LC17] Cosme Louart and Romain Couillet. Harnessing neural networks: a random matrix approach. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17)*, New Orleans, USA, 2017.

[LCB98] Y. LeCun, C. Cortes, and C. Burges. The MNIST database of handwritten digits, 1998.

[Led01] Michel Ledoux. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, Number 89, 2001.

[MOA11] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer series in statistics, 2011.

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

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer-Verlag Berlin Heidelberg, 1986.

[Ouv09] Jean-Yves Ouvrard. *Probabilités, tome II*. Cassini, 2009.

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

[SSSI11] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *IEEE International Joint Conference on*

Neural Networks, pages 1453–1460, 2011.

[Tal94] Michel Talagrand. The supremum of some canonical processes. *The Johns Hopkins University Press*, pages Vol. 116, No. 2, pp. 283–325, 1994.

[Tal95] Michel Talagrand. *Concentration of Measure and Isoperimetric Inequalities in product spaces.* Publications mathématiques de l'I.H.E.S., tome 81, 1995.

[Tao11] Terence Tao. Topics in random matrix theory. Department of Mathematics, UCLA, Los Angeles, 2011.

[Ver17] R. Vershynin. *High dimensional probability.* 2017.

[Yat95] R. D. Yates. A framework for uplink power control in cellular radio systems. 13(7):1341–1347, 1995.