Concentration inequalities in Banach spaces

Matin Yadollahi, Ida Wolff
Advisor: Max Fathi

June 12, 2025

Contents

Abstract

In this report we study concentration inequalities for distributions with values in Banach spaces. We will study two different cases: the Gaussian distribution on finite-dimensional Banach spaces, and the uniform distribution on an hypercube. In the first case, we will use the invariance of the Gaussian distribution under rotation in order to prove a Sobolev-like inequality, from which our concentration inequality will follow.

However, we will not be able to extend this methods to the case of the hypercube, since we will no longer have the rotational invariance, a key ingredient in the first proof. Showing a concentration inequality on the hypercube will therefore require the use of a different method, mainly based on the idea of disturbing the symmetry of the distribution.

1 Introduction

Let E be a Banach space. E has *Radermacher type* $p \in [1,2]$ if there is a constant c such that:

$$\forall n \forall x_1, \cdots, x_n : \mathbb{E}(||\sum_{j=1}^{n} \epsilon_j x_j||^p) \leq c \sum_{j=1}^{n} ||x_j||^p$$

Enflo introduced a more general notion. We say that E is of *Enflo type* p if for every function $f : \{-1,1\}^n \to E$ we have

$$\mathbb{E}(||\frac{f(\epsilon) - f(-\epsilon)}{2}||^p) \leq c \sum_{j=1}^{n} \mathbb{E}||D_j f(\epsilon)||^p$$

where D_j is the discrete partial differentiation. These properties are of interest because they stay invariant under isometries, so they give a tool to prove that two spaces aren't isomorph. Clearly Radermacher type is the special case of Enflo type with a linear function.

However the main problem is to show that these two are equivalent. In order to prove the implication it's enough to show the following inequality:

$$\mathbb{E}(||f(\epsilon) - \mathbb{E}(f(\epsilon))||^p) \leq c\mathbb{E}||\sum_{j=1}^{n} \delta_j D_j f(\epsilon)||^p$$

If this inequality is true then assuming Radermacher type and applying the definition to the RHS we get the definition of Enflo. We will prove the gaussian case which was treated by Pisier. It's not possible to fully extend the proof to the hypercube as the constant will depend on the dimension. However in the last part we will prove a similar inequality for the general case.

2 Concentration inequalities for random Gaussian vectors

In this section, we will focus on the Gaussian case. We will first recall the definition of a Gaussian random vector, before showing a concentration inequality for such vectors. Although the results we will show are true for general Banach spaces, we will only study the finite-dimension case. Nevertheless, we will make sure to formulate the main definitions and theorems in a way that does not rely on the finite dimension of our Banach space, and can therefore be easily generalized.

In this section, let E and F be two Banach spaces of finite dimension n.

2.1 Gaussian vectors: definition and useful properties

When extending the notion of a Gaussian variable to vectors in $\mathbb{R}^n$, it is natural to ask for each coordinate of the vector to follow a normal distribution. However, this condition alone is insufficient to ensure, for example, that an affine transformation of a Gaussian vector remains a Gaussian vector. We shall thus take the following definition

Definition 2.1 (Gaussian random vector in $\mathbb{R}^n$). Let X be a random vector with values in $\mathbb{R}^n$. X is a random Gaussian vector if and only if, for every vector t in $\mathbb{R}^n$, the scalar product $\langle t \, , \, X \rangle$ follows a Gaussian distribution.

In order to extend this definition to Banach spaces, we shall note that, through the canonical isomorphism between $\mathbb{R}^n$ and $\mathbb{R}^{n*}$, the above condition is equivalent to the fact that for any linear form ξ in $\mathbb{R}^{n*}$, $\xi(X)$ is a Gaussian variable. This definition is more general and can thus be extended to vectors in Banach spaces as follows:

Definition 2.2 (Gaussian random vector in a Banach space). Let X be a random vector with values in a Banach space B. X is a random Gaussian vector if and only if, for any linear form ξ in B^*, $\xi(X)$ follows a Gaussian distribution.

To a random vector X we associate the covariance matrix $Cov(X)$, which is a generalization of the variance of a real-valued random variable, defined as follows:

Definition 2.3 (covariance matrix). Let X be a random vector in $\mathbb{R}^n$. Then $Cov(X)$ is the symmetric matrix in $M_n(\mathbb{R})$ with the following coefficients

$$Cov(X)_{i,j} = Cov(X_i X_j).$$

Remark 2.4. Note that, if we assume X to be a column vector, its covariance matrix is $Cov(X) = \mathbb{E}(XX^T) - E(X)\mathbb{E}(X)^T$.

Definition 1.2 gives a simple way to characterize Gaussian random vectors, however it lacks an explicit description of how a general Gaussian vector "looks like", and therefore is not very well suited for computation. We shall thus give an alternative description of Gaussian random vectors in the special case of finite dimensional Banach spaces, assimilated to $\mathbb{R}^n$, which will be useful later on.

Lemma 2.5 (alternative characterization of Gaussian random vectors in $\mathbb{R}^n$). *Let X be a random vector in $\mathbb{R}^n$. The following properties are equivalent:*

1. *X is a Gaussian random vector: for every vector t in $\mathbb{R}^n$, the scalar product $\langle t\ ,\ X\rangle$ follows a Gaussian distribution,*
2. *the characteristic function of X is of the form*
$$\phi_X(t) = exp(i\langle t\ ,\ m\rangle - \frac{1}{2}\langle Kt\ ,\ t\rangle)\ ,\ \textit{with } m \in \mathbb{R}^n \textit{ and } K \in M_n(\mathbb{R}),$$
3. *X is equal in distribution to $AZ + m$ where $A \in M_n(\mathbb{R})$, $m \in \mathbb{R}^n$, and Z is a Gaussian random vector with mean-value zero and $Cov(Z) = I_n$, in other terms $Z = (z_1, ..., z_n)$ where the z_i are independent standard Gaussian random variables.*

Proof.

- 1. $\implies$ 2. : Let $t \in \mathbb{R}^n$. We have
$$\phi_X(t) = \phi_{\langle t\ ,\ X\rangle}(1).$$
We suppose 1. Then $\langle t\ ,\ X\rangle$ follows a Gaussian distribution. We now compute its mean value and variance:
$$\mathbb{E}\langle t\ ,\ X\rangle = \langle t\ ,\ \mathbb{E}X\rangle = \langle t\ ,\ m\rangle$$
with $m = \mathbb{E}X$. To compute the variance we can center X, that is we rename $X - m$ to X, and we recall that in that case the covariance matrix of X which we shall call $Cov(X) = K$ is
$$K = \mathbb{E}(XX^T).$$
In consequence
$$Var(\langle t\ ,\ X\rangle) = \mathbb{E}(\langle t\ ,\ X\rangle^2) = \mathbb{E}(t^T XX^T t) = t^T\mathbb{E}(XX^T)t = t^T Kt = \langle t\ ,\ Kt\rangle = \langle Kt\ ,\ t\rangle$$
since the covariance matrix K is symmetric. From the formula for the characteristic function of a real-valued Gaussian variable, we deduce that
$$\phi_X(t) = exp(i\langle t\ ,\ m\rangle - \frac{1}{2}\langle Kt\ ,\ t\rangle)$$
which gives 2.
- 2. $\implies$ 1. : Let $\lambda \in \mathbb{R}^n$.
$$\phi_{\langle t\ ,\ X\rangle}(\lambda) = \phi_X(\lambda t) = exp(i\lambda\langle t\ ,\ m\rangle - \lambda^2\frac{1}{2}\langle Kt\ ,\ t\rangle)$$
We recognize the characteristic function of a Gaussian distribution, hence 1.
- 3. $\implies$ 1. : We suppose that $X = AZ + m$ in distribution. Then for every $t \in \mathbb{R}^n$, $\langle t\ ,\ X\rangle = \langle t\ ,\ AZ + m\rangle = \langle A^T t\ ,\ Z\rangle + \langle t\ ,\ m\rangle$. Moreover $\langle A^T t\ ,\ Z\rangle$ follows a Gaussian distribution as a linear combination of independent Gaussian variables, hence 1.

- 1. $\implies$ 3. : Let $A \in M_n(\mathbb{R})$ such that $K = AA^T$, where $K = Cov(X)$. To see that such a matrix exists we can for example use the fact that K is symmetric and diagonalize it in an orthonormal basis:

  $$K = ODO^T.$$

  with D a diagonal matrix and $O \in \mathbb{O}_n(\mathbb{R})$. Moreover, for all $y \in \mathbb{R}^n$

  $$y^T K y = \mathbb{E}(y^T X X^T y) = \mathbb{E}(X^T y)^2 \geq 0$$

  so K is a positive definite matrix, and therefor D can be written $D = \sqrt{D}^2$. We can then take $A = O\sqrt{D}O^T$. We define $m = \mathbb{E}X$ as before. We now compute the characteristic function

  $$\begin{aligned}
  \phi_{AZ+m}(t) &= e^{i\langle t,m\rangle}\mathbb{E}(e^{i\langle t,AZ\rangle}) \\
  &= e^{i\langle t,m\rangle}\mathbb{E}(e^{i\langle A^T t,Z\rangle}) \\
  &= e^{i\langle t,m\rangle}\phi_Z(A^T t) \\
  &= e^{i\langle t,m\rangle}e^{-\frac{1}{2}\langle A^T t,A^T t\rangle} \\
  &= e^{i\langle t,m\rangle}e^{-\frac{1}{2}\langle AA^T t,t\rangle} \\
  &= e^{i\langle t,m\rangle-\frac{1}{2}\langle Kt,t\rangle} = \phi_X(t)
  \end{aligned}$$

  via 2. Thus X and AX + m have the same characteristic function so they are equal in distribution, hence 1.

□

The definition of the covariance matrix clearly uses the finite-dimension aspect of our Banach space which we identify with $\mathbb{R}^n$, however the "weak moment" can be defined more generally for any random vector with values in a Banach space B:

Definition 2.6 (weak moment). Let X be a random vector with values in a Banach space B with mean value 0. Its weak moment is the real number

$$\sigma(X) = sup\{(\mathbb{E}(\xi(X)^2))^{1/2} | \xi \in B^*, ||\xi|| \leq 1\}.$$

Remark 2.7. Note that when $B = \mathbb{R}$ this definition coincides with the definition of the standard deviation for real valued vectors. More generally, for X a Gaussian random vector in $\mathbb{R}^n$, $\sigma(X)$ corresponds to the largest eigenvalue of $Cov(X)$.

Proof. For $\xi \in (\mathbb{R}^n)^*$ we will identify the linear form with the vector it is represented by, that is we write $\xi(x) = \langle \xi \ , \ x\rangle$ for $x \in \mathbb{R}^n$. Then

$$\mathbb{E}(\xi(X)^2) = \mathbb{E}(\xi^T X \xi^T X) = \mathbb{E}(\xi^T X(\xi^T X)^T) = \xi^T \mathbb{E}(XX^T)\xi.$$

But for X with expected value 0 $\mathbb{E}(XX^T)$ is exactly the covariance matrix $Cov(X) = K$, which gives us

$$\sigma(X) = sup\{\xi^T K \xi \ , ||\xi|| \leq 1\}$$

which corresponds exactly to the largest eigenvalue of the covariance matrix K. □

Lemma 2.8 (Laplace's transform of the canonical Gaussian distribution on $\mathbb{R}^n$)**.** *Let X be a standard Gaussian vector on $\mathbb{R}^n$. The Laplace's transform of X is given by*

$$\mathbb{E}(e^{\langle \lambda \ , \ X\rangle}) = e^{\frac{\|\lambda\|^2}{2}}$$

for λ in $\mathbb{R}^n$.

Proof. Noticing that $\langle \lambda \, , \, x\rangle - \frac{1}{2}\langle x \, , \, x\rangle = -\frac{1}{2}\langle x-\lambda \, , \, x-\lambda\rangle + \frac{1}{2}\|\lambda\|^2$, we can compute

$$\begin{aligned}\mathbb{E}(e^{\langle \lambda \, , \, X\rangle}) &= \int e^{\langle \lambda \, , \, x\rangle - \frac{1}{2}\langle x \, , \, x\rangle} \frac{dx_1 ... dx_n}{\sqrt{2\pi}^n} \\ &= e^{-\frac{\|\lambda\|^2}{2}} \int e^{\langle x-\lambda \, , \, x-\lambda\rangle} \frac{dx_1 ... dx_n}{\sqrt{2\pi}^n} \\ &= e^{-\frac{\|\lambda\|^2}{2}}\end{aligned}$$

since $\int e^{\langle x-\lambda \, , \, x-\lambda\rangle} \frac{dx_1 ... dx_n}{\sqrt{2\pi}^n} = \int e^{\langle x \, , \, x\rangle} \frac{dx_1 ... dx_n}{\sqrt{2\pi}^n} = 1.$ □

2.2 Concentration of Gaussian vectors

The aim of this subsection is to prove the following concentration inequality for Gaussian random vectors:

Theorem 2.9. *Let X be a Gaussian random vector with values in the Banach space E and expected value 0. Then for all $t > 0$*

$$\mathbb{P}(|\, \|X\| - \mathbb{E}\,\|X\| \ \ | > t\} \leq 2\exp(-2\pi^{-2}t^2\sigma(X)^{-2}).$$

This theorem shows that Gaussian vectors tend to concentrate around their mean value, with the probability decreasing exponentially as the distance from the mean value increases. Note as well that the rate of convergence is governed by the weak moment $\sigma(X)$ of the Gaussian vector, that is the largest eigenvalue of its covariance matrix. This is an interesting fact, given that, contrarily to the one-dimensional case, the weak moment is not enough to fully characterize the distribution of a Gaussian vector in higher dimension. Moreover, the weak moment of a Gaussian random vector X tends to be much smaller than its "strong moment" $\sqrt{\mathbb{E}(\|X\|^2)}$, as we can see in the following example:

Example 2.10. Let $E = (\mathbb{R}^n, \|\ .\|_2)$ and $Z = (z_1, ..., z_n)$ a standard Gaussian variable on $\mathbb{R}^n$. The square of the weak moment of Z is

$$\sigma(Z)^2 = sup\{\mathbb{E}(\xi(Z))^2 \ |\xi \in E^*, \ \|\xi\| \leq 1\}.$$

Here $E^* = E = (\mathbb{R}^n, \|\ .\|_2)$, and for all $\xi \in \mathbb{R}^n$, $\langle \xi \ , \ Z\rangle$ follows a Gaussian distribution with mean value 0 and variance $\|\xi\|^2$, therefore $\mathbb{E}(\xi(Z)^2) = \mathbb{E}(|\langle \xi \ , \ Z\rangle|^2) = \|\xi\|^2$, which gives us $\sigma(Z) = 1$.

On the other hand the square of the strong moment of Z is given by

$$\mathbb{E}(\|Z\|^2) = \mathbb{E}(\Sigma_{i=1}^n |z_i|^2) = \Sigma_{i=1}^n \mathbb{E}(|z_i|^2) = n$$

We therefore have $\sqrt{\mathbb{E}(\|X\|^2)} = \sqrt{n}\sigma(Z)$.

This example shows that, calculated with the Euclidean norm, the "strong moment" of a standard Gaussian vector on $\mathbb{R}^n$ grows like the square root of the dimension, whereas the weak moment stays constant to one. In particular, we can see that the concentration inequality from Theorem 2.9 would be much weaker if it had relied on the "strong moment" rather than on the weak moment of the random vector.

The proof of Theorem 2.9 relies on the following statement:

Theorem 2.11. *Let X and Y be independent Gaussian random vectors with values in the Banach space E and expected value 0, and let f be a locally Lipschitz function $f : E \to F$. Then, for any measurable convex function $\Phi : E \to F$ we have*

$$\mathbb{E}\Phi(f(X) - \mathbb{E}f(X)) \leq \mathbb{E}\Phi(\frac{\pi}{2}Df(X)(Y)).$$

Proof. Let f be a locally Lipschitz function $f : E \to F$ and $\Phi : E \to F$ a measurable convex function.

Let $X(\theta) = Xsin(\theta) + Ycos(\theta)$ and $X'(\theta) = Xcos(\theta) - Ysin(\theta)$. Then we have $X = X(\frac{\pi}{2})$ and $Y = X(0)$. Since X and Y are independent Gaussian vectors $X(\theta)$ and $X'(\theta)$ are also Gaussian vectors. Furthermore, the couple $(X(\theta), X'(\theta))$ has the same distribution as (X, Y). To show this statement, we will use the property 3. from Lemma 2.5 in order to write

$$X = AZ_1 \text{ and } Y = AZ_2$$

with Z_1 and Z_2 independent standard Gaussian vectors on $\mathbb{R}^n$. We can then write

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

and therefore

$$\begin{pmatrix} X(\theta) \\ X'(\theta) \end{pmatrix} = \begin{pmatrix} I_n sin\theta & I_n cos\theta \\ I_n cos\theta & -I_n sin\theta \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

We check that the matrices $\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$ and $\begin{pmatrix} I_n sin\theta & I_n cos\theta \\ I_n cos\theta & -I_n sin\theta \end{pmatrix}$ commute, for example by writing

$$\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} = A \otimes I_n \text{ and } \begin{pmatrix} I_n sin\theta & I_n cos\theta \\ I_n cos\theta & -I_n sin\theta \end{pmatrix} = I_n \otimes \begin{pmatrix} sin\theta & cos\theta \\ cos\theta & sin\theta \end{pmatrix}.$$

We can therefore write

$$\begin{pmatrix} X(\theta) \\ X'(\theta) \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} I_n sin\theta & I_n cos\theta \\ I_n cos\theta & -I_n sin\theta \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

From the invariance by rotation of the standard Gaussian distribution on $\mathbb{R}^n$, $\begin{pmatrix} I_n sin\theta & I_n cos\theta \\ I_n cos\theta & -I_n sin\theta \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ and $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ follow the same distribution, which shows that $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X(\theta) \\ X'(\theta) \end{pmatrix}$ in distribution.

We now use the fundamental theorem of calculus in order to write

$$\begin{aligned} f(X) - f(Y) &= f(X(\pi/2)) - f(X(0)) \\ &= \int_0^{\pi/2} \frac{d}{d\theta} f(X(\theta)) d\theta \\ &= \int_0^{\pi/2} Df(X(\theta))(X'(\theta)) d\theta \\ &= \int_0^{\pi/2} \frac{\pi}{2} Df(X(\theta))(X'(\theta)) \frac{d\theta}{\pi/2} \end{aligned}$$

Since Φ is convex, we can use Jensen's inequality.

$$\Phi(f(X) - f(Y)) \leq \int_0^{\pi/2} \Phi(\frac{\pi}{2} Df(X(\theta))(X'(\theta))) \frac{d\theta}{\pi/2}$$

and thus by linearity of the expected value

$$\mathbb{E}(\Phi(f(X) - f(Y))) \leq \int_0^{\pi/2} \mathbb{E}(\Phi(\frac{\pi}{2} Df(X(\theta))(X'(\theta)))) \frac{d\theta}{\pi/2}$$

Moreover, since $(X(\theta), X'(\theta))$ is equal in distribution to (X, Y),

$$\int_0^{\pi/2} \mathbb{E}(\Phi(\frac{\pi}{2}Df(X(\theta))(X'(\theta))))\frac{d\theta}{\pi/2} = \int_0^{\pi/2} \mathbb{E}(\Phi(\frac{\pi}{2}Df(X)(Y)))\frac{d\theta}{\pi/2} = \mathbb{E}(\Phi(\frac{\pi}{2}Df(X)(Y)))$$

and so get

$$\mathbb{E}(\Phi(f(X) - f(Y))) \leq \mathbb{E}(\Phi(\frac{\pi}{2}Df(X)(Y))).$$

To conclude, we need to show that

$$\mathbb{E}(\Phi(f(X) - \mathbb{E}f(X))) \leq \mathbb{E}(\Phi(f(X) - f(Y))).$$

This follows from the convexity of Φ, and from the fact that X and Y are independent which enables us to take the expected value first on Y and then on X, using Fubini's theorem. First fixing X and taking the expected value on Y we get, from Jensen's inequality

$$\Phi(\mathbb{E}(f(X) - f(Y))) = \Phi(f(X) - \mathbb{E}f(Y)) \leq \mathbb{E}(\Phi(f(X) - f(Y))).$$

Since X and Y follow the same distribution, we have $\mathbb{E}X = \mathbb{E}Y$, an so

$$\Phi(f(X) - \mathbb{E}f(X)) \leq \mathbb{E}(\Phi(f(X) - f(Y))).$$

Now we take the expected value over X, which gives us

$$\mathbb{E}(\Phi(f(X) - \mathbb{E}f(X))) \leq \mathbb{E}(\Phi(f(X) - f(Y)))$$

and concludes the proof. □

The following corollary will be useful to prove Theorem 1.2:

Corollary 2.12. *Suppose $E = \mathbb{R}^n$, X a standard Gaussian random vector on $\mathbb{R}^n$ and let $f : E \to F$ be a Lipschitz function. We then have*

$$\forall t > 0, \mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq 2\exp(-2\pi^{-2}t^2/\left\|f\right\|_{lip}^2).$$

Proof. Let Y be an independent copy of X. We fix $t > 0$, and we define the following family of convex functions:

$$\Phi_\lambda : x \to \exp \lambda x \quad \lambda > 0$$

We have

$$\mathbb{P}(f(X) - \mathbb{E}f(X) > t) = \mathbb{P}(\Phi_\lambda(f(X) - \mathbb{E}f(X)) > e^{\lambda t}) \leq \frac{\mathbb{E}\Phi_\lambda(f(X) - \mathbb{E}f(X))}{e^{\lambda t}}$$

by Chebyshev's inequality. Applying 2.11 with the convex function Φ_λ gives

$$\mathbb{E}\Phi_\lambda(f(X) - \mathbb{E}f(X)) \leq \mathbb{E}\Phi_\lambda(\frac{\pi}{2}Df(X)(Y)) = \mathbb{E}(e^{\lambda\frac{\pi}{2}\langle\nabla f(X)\, ,\, Y\rangle})$$

Since the random vectors X and Y are independent, by Fubini's theorem we can compute the expected value first on Y and then on X. For X fixed in $\mathbb{R}^n$ this expected value is just the Laplace's transform of a standard Gaussian random vector, which we computed in 2.8, evaluated in $\frac{\pi}{2}\lambda\nabla f(X)$. We therefore get

$$\mathbb{E}(e^{\lambda\frac{\pi}{2}\langle\nabla f(X)\, ,\, Y\rangle}) = \mathbb{E}(e^{\frac{1}{2}\left\|\lambda\frac{\pi}{2}\nabla f(X)\right\|^2})$$

Now we majorize $\|\nabla f(X)\|$ by $\|f\|_{lip}$, which gives

$$\mathbb{E}\Phi_\lambda(f(X) - \mathbb{E}f(X)) \leq e^{\frac{1}{2}(\frac{\lambda\pi}{2})^2\|f\|_{lip}^2}$$

and we finally get

$$\mathbb{P}(f(X) - \mathbb{E}f(X) > t) \leq e^{-\lambda t + \frac{1}{2}(\frac{\lambda \pi}{2})^2 \|f\|_{lip}^2}$$

for all $\lambda > 0$. Choosing $\lambda = \frac{4t}{\pi^2 \|f\|_{lip}^2}$ we get

$$\mathbb{P}(f(X) - \mathbb{E}f(X) > t) \leq \exp(-2\pi^{-2} t^2 / \|f\|_{lip}^2).$$

We conclude by applying this result to f and to $-f$, remarking that

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq \mathbb{P}(f(X) - \mathbb{E}f(X) > t) + \mathbb{P}(-f(X) - \mathbb{E}(-f(X)) > t).$$

□

We can now prove Theorem 2.9.

Proof. Let X be a Gaussian random vector with expected value 0. We will use Corollary 2.12, and to do so we need to write $\|X\|$ in the form $f(Z)$ with f a Lipschitz function and Z a standard Gaussian vector on $\mathbb{R}^n$. We will thus use the characterization 3. from Lemma 2.5 to write $X = AZ + m$ in distribution, with $A \in M_n(\mathbb{R})$ and $m \in \mathbb{R}^n$. Note that if X has values in a Banach space E, A is in fact an operator from $\mathbb{R}^n$ to $\mathbb{E}$, and m is a vector from E.

We can now define

$$f : z \in \mathbb{R}^n \to \|Az\|_E \in \mathbb{R},$$

and we have $\|X\|_E = f(Z)$ in distribution. In order to apply Corollary 2.12 we now need to compute $\|f\|_{lip}$.

Let $z_1, z_2 \in \mathbb{R}^n$, $z_1 \neq z_2$. Using the triangular inequality

$$\frac{|f(z_1) - f(z_2)|}{\|z_1 - z_2\|} \leq \frac{\|A(z_1 - z_2)\|_E}{\|z_1 - z_2\|} \leq \|A\|_{op}$$

by definition of the operator norm. Thus we have $\|f\|_{lip} \leq \|A\|_{op}$, and setting z_2 to zero shows that this is an equality, so $\|f\|_{lip} = \|A\|_{op}$. We will now show that $\|A\|_{op} = \sigma(X)$ which will conclude the proof.

We need to show that $\sigma(X) = \|A\|_{op}$, that is $\sigma(X)^2 = sup\{\mathbb{E}(\xi(X)^2) | \xi \in B^*, ||\xi|| \leq 1\} = \|A\|_{op}^2 = sup\{\|Ax\|_E^2| \ \|x\| \leq 1\}$. Remarking that $sup\{\|Ax\|_E| \ \|x\| \leq 1\} = sup\{\xi(Ax)| \ \|x\| \leq 1, \ \|\xi\| \leq 1\}$, it is sufficient to show $\mathbb{E}(\xi(X)^2) = sup\{\xi(Ax)^2 \ \|x\| \leq 1\}$ for all $\xi \in E^*$.

Let $\xi \in \mathbb{E}^*$. We have $\mathbb{E}(\xi(X)^2) = \mathbb{E}(\xi(AZ)^2) = \mathbb{E}(\tilde{\xi}(Z)^2)$ where we define $\tilde{\xi}$ by

$$\tilde{\xi} : z \in \mathbb{R}^n \to \xi(Az) \in \mathbb{R}.$$

Being a linear form on $\mathbb{R}^n$, $\tilde{\xi}$ can be represented by a vector $u \in \mathbb{R}^n$:

$$\tilde{\xi} : z \in \mathbb{R}^n \to \langle u \ , \ z \rangle \in \mathbb{R}.$$

Then we have

$$\mathbb{E}(\xi(X)^2) = \mathbb{E}(\langle u \ , \ Z \rangle^2) = \|u\|^2 \mathbb{E}(\langle \frac{u}{\|u\|} \ , \ Z \rangle^2).$$

By definition of standard Gaussian vectors on $\mathbb{R}^n$, and because $\frac{u}{\|u\|}$ is unitary, $\langle \frac{u}{\|u\|} \ , \ Z \rangle$ follows the standard Gaussian distribution on $\mathbb{R}$, and therefore $\mathbb{E}(\langle \frac{u}{\|u\|} \ , \ Z \rangle^2 = 1$, which finally gives us

$$\mathbb{E}(\xi(X)^2) = \|u\|^2.$$

But this is exactly $sup\{\xi(Ax)^2 \ \|x\| \leq 1\}$. Indeed, we can rewrite

$$sup\{\xi(Ax)^2, \ \|x\| \leq 1\} = sup\{\tilde{\xi}(x)^2, \ \|x\| \leq 1\} = sup\{(\langle u \ , \ x \rangle)^2 \ , \ \ \|x\| \leq 1\} = \|u\|^2$$

.

In consequence $\|f\|_{lip} = \|A\|_{op} = \sigma(X)$ and Theorem 2.9 follows from Corollary 2.12.

□

3 Concentration inequality on the hypercube

In this part we shall prove a theorem similar to 2.11 but with allowing the function to take any value on the hypercube of dimension n. As we are working in a discrete space, we need to define the discrete differential.

Definition 3.1 (discrete differential). Let $f : \{-1,1\}^n \to E$ be a function and $\epsilon = (\epsilon_1, \cdots, \epsilon_n) \in \{-1,1\}^n$. We define the discrete partial derivatives as:

$$D_j f(\epsilon) = \frac{f(\epsilon_1, \cdots, \epsilon_j, \cdots, \epsilon_n) - f(\epsilon_1, \cdots, -\epsilon_j, \cdots, \epsilon_n)}{2}.$$

We also define discrete Laplacian

$$\Delta f := -\sum_{j=1}^{n} D_j f$$

And the heat semigroup on the cube by

$$P_t f := (e^{t\Delta})(f) = \sum_{i=0}^{\infty} \frac{(t\Delta)^i}{i!} f.$$

Throughout the proof we suppose that ϵ is a random vector on the cube which is uniformly distributed while $\zeta(t)$ is another random vector independent of ϵ with following distribution:

Each coordinate $\zeta_i(t)$ is i.i.d. with:

$$\mathbb{P}(\zeta_i(t) = 1) = \frac{1 + e^{-t}}{2}.$$

We also standardize ζ by defining $\delta(t)$ as:

$$\delta_i(t) = \frac{\zeta_i(t) - \mathbb{E}(\zeta_i(t))}{\sqrt{\mathrm{Var}\zeta_i(t)}} = \frac{\zeta_i(t) - e^{-t}}{\sqrt{1 - e^{-2t}}}.$$

Theorem 3.2. *For any linear space E, $f : \{-1,1\}^n \to E$, and convexe function $\Phi : E \to \mathbb{R}$ we have:*

$$\mathbb{E}[\Phi(f(\epsilon) - \mathbb{E}(f(\epsilon))] \leq \int \mathbb{E}[\Phi(\frac{\pi}{2} \sum_{j=1}^{n} \delta_j(t) D_j f(\epsilon))]\mu(dt)$$

where μ is th density $\mu(dt) = \frac{2}{\pi} \frac{1}{\sqrt{e^{2t}-1}}$.

Proof. The proof relies on the following lemma which relates $f(x)$ to $\mathbb{E}f(x\zeta)$ using Kolmogorov differential equation on distributions. Before we start the proof, we will give some explanation about this equation.

3.1 Kolmogorov equation

The original equation takes place in a diffusion process. Assume a stochastic differential eqation:

$$dX_t = \sigma dW_t + b(X_t)dt$$

with W_t a brownian motion. Then the probability density p of X evolves according to:

$$\frac{d}{dt} p = -\nabla(b(x)p) + \sigma^2 \Delta p.$$

When there is no drift ($b = 0$) then the unique solution is given by

$$p = e^{\sigma^2 t \Delta} f$$

where f is the initial distribution.

Lemma 3.3. *We have*

$$P_t f(x) = \mathbb{E}[f(x\zeta(t))] \quad \forall t \geq 0,$$

and

$$D_j P_t f(x) = \frac{1}{\sqrt{e^{2t}-1}} \mathbb{E}[\delta_j(t) f(x\zeta(t))] \quad \forall t > 0.$$

Proof. Let's denote $\mathbb{E}[f(x\zeta(t))]$ as $Q_t f$. By definition we have:

$$Q_t f = \sum_{\zeta \in \{-1,1\}^n} \prod_{i=1}^{n} \frac{1+e^{-t}\zeta_i}{2} f(x\zeta).$$

Next we want to calculate $D_j Q_t f$. Indeed

$$Q_t f(x_1, \cdots, -x_j, \cdots, x_n) = \sum_{\zeta \in \{-1,1\}^n} \frac{1-e^{-t}\zeta_j}{1+e^{-t}\zeta_j} \prod_{i=1}^{n} \frac{1+e^{-t}\zeta_i}{2} f(x\zeta).$$

Notice that

$$1 - \frac{1-e^{-t}\zeta_j}{1+e^{-t}\zeta_j} = \frac{e^{-t}\zeta_j}{1+e^{-t}\zeta_j} = \frac{e^{-t}}{1-e^{-2t}}(\zeta_j - e^{-t}).$$

Taking the difference we get

$$D_j Q_t f = \frac{1}{\sqrt{e^{2t}-1}} \mathbb{E}[\delta_j(t) f(x\zeta(t))].$$

To show that $Q_t f$ is the same $P_t f$, we show that it is the solution to the Kolmogorov equation with the same initial condition as f. Then the uniqueness of the solution implies the equality. It is easy to see that $Q_0 f = f$,

$$\frac{d}{dt} Q_t f = -\sum_{j=1}^{n} \sum_{\zeta \in \{-1,1\}^n} \prod_{i=1}^{n} \frac{1+e^{-t}\zeta_i}{2} \frac{e^{-t}\zeta_j}{1+e^{-t}\zeta_j} f(x\zeta) = \Delta Q_t f(x).$$

□

Now we use the convex conjugate of Φ, $\Phi^* : E^* \to (-\infty, \infty]$ (we can assume finite dimension on E as we are only working with the span of the image of the hypercube). We write

$$\Phi(x) = \sup_{z \in E^*} \{< z, x > -\Phi^*(z)\}$$

Then the LHS becomes:

$$\mathbb{E}[\Phi(f(\epsilon) - \mathbb{E}f(\epsilon))] = \sup_{g:\{-1,1\}^n \to E^*} \{\mathbb{E}[< g(\epsilon), f(\epsilon) - \mathbb{E}f(\epsilon) >] - \mathbb{E}[\Phi^*(g(\epsilon))]\}.$$

Using what we have from our lemma, $P_0 f = f$ and $\lim_{t\infty} P_t f = \mathbb{E}f(\epsilon)$, we can rewrite

$$\mathbb{E}[f(\epsilon)] - f(\epsilon) = \int_0^\infty \frac{d}{dt} P_t f(\epsilon) dt.$$

Substituting back,

$$\begin{aligned} \mathbb{E}[< g(\epsilon), f(\epsilon) - \mathbb{E}[f(\epsilon)] >] &= -\int_0^\infty \mathbb{E}[< g(\epsilon), \Delta P_t f(\epsilon) >] dt \\ &= -\int_0^\infty \mathbb{E} \sum_{j=1}^{n} [< D_j P_t g(\epsilon), D_j f(\epsilon) >] dt. \end{aligned}$$

Here we use the fact that Laplacian is self-adjoint (because it is symmetric) and it commutes with P_t. The last line can be justified by a simple calculation but in non-discrete case it can be seen as an integration by part. Then we can again use our lemma to substitute $D_j P_t g(\epsilon)$.

$$\sum_{j=1}^{n} \mathbb{E}[< D_j P_t g(\epsilon), D_j f(\epsilon) >] = \frac{1}{\sqrt{e^{2t}-1}} \sum_{j=1}^{n} \mathbb{E}[< \delta_j g(\epsilon\zeta(t)), D_j f(\epsilon) >].$$

Using the fact that ϵ and $\zeta\epsilon$ have the same uniform distribution, we can rewrite the original equation as

$$\begin{aligned}
&\mathbb{E}[< g(\epsilon), f(\epsilon) - \mathbb{E}[f(\epsilon)] >] - \mathbb{E}[\Phi^*(g(\epsilon))] \\
&= \int_0^\infty \left[\mathbb{E}[< g(\epsilon\zeta(t)), \frac{\pi}{2} \sum_{j=1}^{n} \delta_j(t) D_j f(\epsilon) >] - \Phi^*(g(\epsilon\zeta(t))) \right] \mu(dt) \\
&\leq \int_0^\infty \mathbb{E}[\Phi(\frac{\pi}{2} \sum \delta_j(t) D_j f(\epsilon))] \mu(dt).
\end{aligned}$$

The last line comes from the dual definition of Φ. □

4 Conclusion

Finally using $\Phi(x) = ||x||^p$ in the last theorem yields:

$$\mathbb{E}(||\frac{f(\epsilon) - f(-\epsilon)}{2}||^p) \leq \int \mathbb{E}(||\frac{\pi}{2} \sum_{j=1}^{n} \delta_j(t) D_j f(\epsilon)||^p) \mu(dt).$$

Using Jensen we can write:

$$\mathbb{E}(||\sum_{j=1}^{n} \delta_j(t) D_j f(\epsilon)||^p) \leq \mathbb{E}(||\sum_{j=1}^{n} \frac{\zeta_j(t) - \zeta_j'(t)}{\sqrt{\text{Var}\zeta_j(t)}} D_j f(\epsilon)||^p).$$

If we assume we are of Radermacher type p then we have

$$RHS \leq c \sum_{j=1}^{n} \mathbb{E}|\frac{\zeta_j(t) - \zeta_j'(t)}{\sqrt{\text{Var}\zeta_j(t)}}|^p \mathbb{E}(||D_j f(\epsilon)||^p).$$

Using again Jensen we get:

$$\mathbb{E}|\frac{\zeta_j(t) - \zeta_j'(t)}{\sqrt{\text{Var}\zeta_j(t)}}|^p \leq 2^{p/2}$$

which completes the proof. Thus Radermacher type is equivalent to Enflo type.

References

[1] Giles Pisier *Probabilistic methods in the geometry of Banach spaces*
[2] Djalil Chafaï notes de cours *Phénomènes de grande dimension*
[3] Paata Ivanisvili, Ramon van Handel, and Alexander Volberg *Rademacher type and Enflo type coincide*
[4] Assaf Naor *An introduction to the Ribe program*