

Notes for Seminar on High Dimensional Statistics

Geometrical Concentration

Beining Wu

December 15, 2021

Contents

1	Concentration via Lipshitz Continuous Functions	1
1.1	Lipshitz Continuity	1
1.2	From Isoperimetric Inequality to Concentration	2
1.3	Key Tool: Isoperimetric Inequality	2
1.4	A Powerful Technique: Blow Up	2
1.5	Proof of the Main Theorem	3
2	Concentration on Generalized Geometrical Spaces	4
2.1	Gaussian Concentration	4
2.2	Pushing Forward, Continuous Balls, Cubes	5
2.3	Hamming Cube, Symmetric Group	5
3	Application: Johnson Lindenstrauss Lemma	6
3.1	Grassmannian Manifolds	6
3.2	Main Theorem	6

Introduction

This document is based on a recent presentation by me on the high dimensional statistics seminar. The content is mainly selected from the Vershynin's High Dimensional Probability, Chapter 5, Section 1 - 3. The basic tools we will involve is the isoperimetric inequality and blow up technique. As a consequence, we can derive concentration properties on various **spaces of random elements with geometric structures**. Finally, we use these concentration results to obtain the Johnson-Lindenstrauss lemma.

§1 Concentration via Lipshitz Continuous Functions

1.1 Lipshitz Continuity

We begin with the definition of the Lipshitz continuous functions. Intuitively, Lipshitz conditions provide stronger restraints on the variation of functions, which makes concentration possible.

Definition 1 (Lipshitz Continuity). Let $f : (X, d_X) \rightarrow (Y, d_Y)$ be mappings between metric spaces. It's Lipshitz continuous if there exists a $L > 0$, such that:

$$d_Y(f(u), f(v)) \leq L d_X(u, v), \quad \forall u, v \in X.$$

We also define $\|f\|_{\text{Lip}}$ to be the minimal L that the above inequality is true.

The Lipshitz condition controls the near points from being blown away after mapping. Clearly, Lipshitz continuous functions are uniformly continuous. Functions with bounded derivatives, or gradients are automatically Lipshitz continuous, by virtue of the mean value theorem. Following examples are worthy of note.

Example. Let $f : [-1, 1] \rightarrow \mathbb{R}$, $x \mapsto \sqrt{1-x^2}$. Then the function is a continuous function on a closed interval, but not Lipshitz continuous. Notice that the derivative explode near -1 and 1 .

Example. Let $f : [-1, 1] \rightarrow \mathbb{R}$, $x \mapsto |x|$. Then the function is clearly Lipshitz continuous, with Lipshitz norm less than 1. But the function is not everywhere differentiable.

Example. Consider the bounded linear functional on \mathbb{R}^n with the canonical inner product

$$f(x) = \langle \theta, x \rangle$$

where $\theta \in \mathbb{R}^n$. Then it's Lipschitz continuous, due to the Cauchy inequality.

$$|f(x) - f(y)| \leq \|\theta\| \cdot \|x - y\|$$

This implies that $\|f\|_{\text{Lip}} \leq \|\theta\|$. On the other hand, let $y = \theta, x = 0$ we can infer that $\|f\|_{\text{Lip}} \geq \|\theta\|$. Hence $\|f\|_{\text{Lip}} = \|\theta\|$.

Example. Bounded linear operators $L(\mathbb{R}^m; \mathbb{R}^n)$. It's easy to verify that it's Lipschitz continuous with Lipschitz norm $\|A\|$, the operator norm.

1.2 From Isoperimetric Inequality to Concentration

To begin with, we present our main theorem in this section here as an appetizer.

Theorem 1 (Spherical Concentration). Let $X \sim \text{Uniform}(\sqrt{n}\mathbb{S})$, and $f : \sqrt{n}\mathbb{S} \rightarrow \mathbb{R}$ be a Lipschitz continuous function with respect to the Euclidean norm on \mathbb{R}^n , then:

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi^2} \leq C \|f\|_{\text{Lip}}. \quad (1)$$

It's equivalent to say that

$$\mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \geq t) \leq 2 \exp\{-ct^2/\|f\|_{\text{Lip}}\}. \quad (2)$$

How do we prove this? First, let's recall that, if f is a linear functional, which is automatically Lipschitz, then the result is obviously true, by the definition of sub-Gaussian random vector. An intuitive approach to the nonlinear case is linear approximation directly. But this is not how we do here. Alternately, we compare the sub level sets of each functions. This is where the isoperimetric inequality come in. As a matter of fact, the sub level sets of the linear functionals on the sphere is exactly spherical caps, the minimizer in the isoperimetric inequality. We use a diagram to represent this procedure.

$$\begin{array}{ccccccc} f(X) \geq M & \xrightarrow{\quad} & A \subset \sqrt{n}\mathbb{S}^{n-1} & \xrightarrow{\text{iso. ineq.}} & H \subset \sqrt{n}\mathbb{S}^{n-1} & & \\ \downarrow & & \downarrow \text{extension} & & \downarrow & & \\ f(X) \geq M+t & \xrightarrow{\quad} & A_t & \xrightarrow{\text{Blow up}} & H_t & \xrightarrow{\quad} & X_1 \geq M+t \end{array}$$

1.3 Key Tool: Isoperimetric Inequality

We've already encountered the toy version of this inequality in the first year analysis course in the following sense.

Theorem 2 (Isoperimetric Inequality, Toy Version). Among all the closed curves with fixed area of the region it encloses, the circles minimize the circumference.

We have a generalized version, in \mathbb{R}^n , which gives a similar result.

Theorem 3 (Isoperimetric Inequality, \mathbb{R}^n Generalization). Among all the subsets in \mathbb{R}^n with fixed interior volume, the Euclidean balls minimize the surface area, also the volume of the ε -extension. Here the extension of a given set is defined as

$$A_\varepsilon = \{y \in \mathbb{R}^n : \exists x \in A, d(x, y) \leq \varepsilon\}.$$

However, what we actually need here is not this version, but a special case which takes \mathbb{S}^{n-1} as the ambient space.

Theorem 4 (Isoperimetric Inequality, \mathbb{S}^{n-1} case). Among all the subsets $A \subset \mathbb{S}^{n-1}$, with fixed spherical area $\sigma(A)$, the spherical caps minimize the area of the ε -extension and the circumference.

Remark. Recall the definition of the spherical caps. A subset $A \subset \mathbb{S}^{n-1}$ is a **spherical cap** if, for some $\theta \in \mathbb{R}^n$ and $a \in \mathbb{R}$,

$$A = \{x \in \mathbb{S}^{n-1} : \langle x, \theta \rangle \leq a\}$$

1.4 A Powerful Technique: Blow Up

The result here actually describe a counter intuitive phenomenon that, if a set $A \subset \mathbb{S}^{n-1}$ takes over half of area, then A_ε would grow exponentially to make up the whole sphere.

Lemma 1 (Blow Up). Let $A \subset \sqrt{n}\mathbb{S}^{n-1}$, and σ is the normalized surface area measure on $\sqrt{n}\mathbb{S}^{n-1}$, i.e., with a unified total measure. Now assume that $\sigma(A) \geq 1/2$, then we have

$$\sigma(A_t) \geq 1 - 2 \exp\{-ct^2\}.$$

This is actually the first place where the concentration comes in. Once this result is build, the main result would be easily derived.

Proof. Define the special half cap as

$$H = \{x \in \sqrt{n}\mathbb{S}^{n-1} : x_1 \leq 0\}.$$

By assumption $\sigma(A) \geq 1/2 = \sigma(H)$, the isoperimetric inequality implies that there exists a cap C with $\sigma(C) = \sigma(A)$ such that

$$\sigma(A_t) \geq \sigma(C_t).$$

Both C and H are spherical caps, and C has an area no less than the H . Naturally we have

$$\sigma(A_t) \geq \sigma(H_t).$$

Hence we shall turn to bound $\sigma(H_t)$. This is where the randomness comes in, let $X \sim \text{Uniform}(\sqrt{n}\mathbb{S}^{n-1})$. Vershynin's theorem 3.4.6 asserts that X is of bounded sub-Gaussian norm. It follows that

$$\sigma(H_t) = \mathbb{P}(X \in H_t).$$

To make H_t more tractable, we introduce an ancillary result.

Claim. The spherical cap $C = \{x \in \sqrt{n}\mathbb{S}^{n-1} : x_1 \leq t/\sqrt{2}\} \subset H_t$. Indeed, for $x \in C$, let

$$\tilde{x} = (0, \frac{\sqrt{nx_2}}{\sqrt{n-x_1^2}}, \frac{\sqrt{nx_3}}{\sqrt{n-x_1^2}}, \dots, \frac{\sqrt{nx_n}}{\sqrt{n-x_1^2}})$$

Then $\tilde{x} \in \sqrt{n}\mathbb{S}^{n-1}$. The square of the distance between x and \tilde{x} is

$$\|x - \tilde{x}\|^2 = x_1^2 + (1 - \frac{\sqrt{n}}{\sqrt{n-x_1^2}})^2(n-x_1^2) = 2n - 2\sqrt{n^2 - nx_1^2} \leq t^2$$

This is true when t is sufficiently small, so is the inclusion.

Return to the proof, we have

$$\sigma(H_t) \geq \mathbb{P}(X_1 \leq t/\sqrt{2}) \geq 1 - 2\exp\{-ct^2\}$$

We have used the definition of the sub-Gaussian random vector. ■

1.5 Proof of the Main Theorem

With the tools in hand, we can now prove the main theorem. There is one key step left. What we want to prove is the concentration on mean. But the mean is not something easy to track in terms of area. Another key observation lies in the condition of blowing up, which requires the normalized area to exceed a half. This 1/2 critical value reminds us of the median. Recall the Wainwright's Exercise 2.14, which asserts that the concentration on mean is equivalent to concentration on median, with constant multipliers on the sub-Gaussian norm.

Proof. Without loss of generality we can assume that $\|f\|_{\text{Lip}} = 1$. Let M be the median of $f(X)$, i.e.

$$\mathbb{P}(f(X) \geq M) \geq \frac{1}{2}, \quad \mathbb{P}(f(X) \leq M) \geq \frac{1}{2}.$$

The corresponding sets on the sphere is $A = \{x \in \sqrt{n}\mathbb{S}^{n-1} : f(x) \leq M\}$, Lemma 1 implies that

$$\mathbb{P}(X \in A_t) \geq 1 - 2\exp\{-ct^2\}.$$

We claim that

$$\mathbb{P}(X \in A_t) \leq \mathbb{P}(f(X) \in M+t).$$

Indeed, if $x \in A_t$, then there exists $y \in A$ such that $\|y - x\| \leq t$ by the definition of extension. With the Lipschitz continuity, we now have

$$f(x) \leq f(y) + \|y - x\| \leq M + t.$$

Then the result follows easily due to the monotonicity of measure.

Combine all the results above, we now have

$$\mathbb{P}(f(X) \leq M+t) \geq 1 - 2\exp\{-ct^2\}.$$

Which is equivalent to say that $f(X) - M$ has bounded sub-Gaussian norm, so does the $f(X) - \mathbb{E}[f(X)]$. We're done. ■

§2 Concentration on Generalized Geometrical Spaces

In this section, we simply put some introductory results. We've seen the power of the blow up technique and isoperimetric inequality, which validate the concentration. Can we extend this procedures to generalized spaces? The answer is true.

2.1 Gaussian Concentration

We have proved results when f is the spherical Lipschitz function and X follows the uniform distribution on sphere. Another frequently encountered multidimensional distribution is the normal $N(0, I_n)$. Similarly, we define the Gaussian push-forward measure on \mathbb{R}^n , as the normalized volume in \mathbb{R}^n .

$$\gamma_n(A) = \int_A \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \|x\|^2\right\} dx.$$

Based on this volume, we'd like to develop similar tools.

Theorem 5 (Gaussian Isoperimetric Inequality). Among all $A \subset \mathbb{R}^n$ with fixed Gaussian measure $\gamma_n(A)$. Half spaces H minimize the volume of ε -extension A_ε .

Remark. Half spaces in \mathbb{R}^n are defined as:

$$H := \{x \in \mathbb{R}^n : \langle x, \theta \rangle \leq a, \text{ for some } \theta \in \mathbb{R}^n \text{ and } a \in \mathbb{R}\}$$

With this Gaussian isoperimetric inequality, we can present our main theorem.

Theorem 6. Let $X \sim N(0, I_n)$, and function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz continuous. Then,

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq C \|f\|_{\text{Lip}}.$$

Proof. We still begin with concentration on median, which proves convenience using our tools in hands. Define M to be the median of $f(X)$, i.e.,

$$\mathbb{P}(f(X) \leq M) \geq \frac{1}{2}, \quad \mathbb{P}(f(X) \geq M) \geq \frac{1}{2}.$$

And also let $A := \{x \in \mathbb{R}^n : f(x) \leq M\}$. By definition we have $\gamma_n(A) \geq 1/2$. Denote the t -extension of A as A_t , then the isoperimetric inequality gives

$$\gamma_n(A_t) \geq \gamma_n(H_t).$$

Where the half space is chosen specially as $H = \{x \in \mathbb{R}^n : x_1 \leq 0\}$, and by definition we have $H_t = \{x \in \mathbb{R}^n : x_1 \leq t\}$. Simple calculation gives

$$\gamma_n(H_t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} dx = \Phi(t),$$

where $\Phi(t)$ is the c.d.f. of $N(0, 1)$.

If $x \in A_t$, then there exists $x \in A$ such that $\|x - y\| \leq t$. Lipschitz continuity implies that

$$f(x) \leq f(y) + \|f\|_{\text{Lip}} \|x - y\| = f(y) + \|f\|_{\text{Lip}} t.$$

Combine all the reasoning above we have

$$\mathbb{P}(f(X) \leq M + \|f\|_{\text{Lip}} t) \geq \Phi(t).$$

This easily gives the bound of sub-Gaussian norm, because

$$\mathbb{P}(|f(X) - M| \geq \|f\|_{\text{Lip}} t) \leq 2(1 - \Phi(t)) \leq 2 \exp\{-t^2/2\}.$$

There are two simple examples which we have encountered. If $f(x) = \|x\|$, then the result is exactly **concentration on norm**. Another case is $f(x) = \langle x, \theta \rangle$, the linear functional, which originates in the definition of multi-dimensional sub-Gaussian vector.

2.2 Pushing Forward, Continuous Balls, Cubes

Gaussian is the most frequent distribution we will meet in use. A simple technique, **pushing forward** could bring the concentration results of known distribution to new distribution. Following results on continuous cube is heuristic.

Theorem 7. Let $X \sim \text{Uniform}[0, 1]^n$, function $f : [0, 1]^n \rightarrow \mathbb{R}$ be Lipschitz, then

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq C \|f\|_{\text{Lip}}.$$

Proof. Before proving this theorem, let's make a simple observation. If $X \sim N(0, I_n)$ and $\phi(x) = (\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n))$. Then $\phi(X) \sim \text{Uniform}[0, 1]^n$. Just need to notice that Z_i are i.i.d., and $\Phi(Z_1) \sim \text{Uniform}[0, 1]$.

Another observation is that $\phi(x)$ is Lipschitz, with bounded gradient

$$\|\nabla \phi(x)\| \leq n/2\pi.$$

The composition of Lipschitz continuous functions are Lipschitz continuous, and moreover

$$\|f \circ \phi\|_{\text{Lip}} \leq \|f\|_{\text{Lip}} \cdot \|\phi\|_{\text{Lip}}.$$

Gaussian concentration implies:

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} = \|f \circ \phi(Z) - \mathbb{E}[f \circ \phi(Z)]\|_{\psi_2} \leq C \|f\|_{\text{Lip}} \cdot \|\phi\|_{\text{Lip}}.$$

■

Next, we will push forward to the unit ball. We want to find a function $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, such that $g(Z) \sim \text{Uniform}(\sqrt{n}B)$. Define function $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

$$g(x) = \sqrt{n} \frac{F_n(\|x\|^2)^{1/n}}{\|x\|} x,$$

where F_n is the χ_n^2 c.d.f. Clearly it's isotropic. Notice that $\|g(x)\|^n = n^{n/2} F_n(\|x\|^2) \sim \text{Uniform}[0, n^{n/2}]$, ball uniformity is obvious. Now we want to prove that it's Lipschitz continuous. Consider the function $G : \mathbb{R} \rightarrow \mathbb{R}$,

$$G(x) = F_n(|x|^2)^{1/n} / |x|.$$

We will check that it's Lipschitz continuous. First check its continuity at 0. Indeed, with the L'Hopital's rule,

$$\lim_{x \rightarrow 0^+} F_n(x^2)/x^n = \lim_{x \rightarrow 0^+} 2x f_n(x^2)/n x^{n-1} = \lim_{x \rightarrow 0^+} c x^{1+2(\frac{n}{2}-1)}/x^{n-1} = c$$

Hence we can define $g(0) = \lim_{x \rightarrow 0} g(x)$ so that it's continuous. By taking derivative,

$$G'(x) = \frac{2}{n} F_n(x^2)^{1/n-1} f_n(x^2) - F_n(x^2)^{1/n} / x^2.$$

It's easy to check that both terms are bounded when x is far away from 0. When $x \rightarrow 0$

2.3 Hamming Cube, Symmetric Group

This part would involve some discrete geometric space, on which we can consider the distance.

Hamming Cube This is the lattice structure on the high dimensional cube. Consider the set $C = \{0, 1\}^n$. Define the metric as

$$d(x, y) = \frac{1}{n} |\{i : x_i \neq y_i\}|.$$

Define the standard measure as

$$\mathbb{P}(A) = \frac{|A|}{2^n}.$$

In this setting, similar results like isoperimetric inequality are also available. Hence we can derive similar concentration result.

Theorem 8 (Concentration on Lattice). Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$, and $X \sim \text{Uniform}\{0, 1\}^n$, then f is automatically Lipschitz continuous. The following result is true.

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq \|f\|_{\text{Lip}} / \sqrt{n}.$$

Symmetric Group Let S_n be the symmetric group of order n . Define the metric

$$d(\pi, \rho) = \frac{1}{n} |\{i : \pi(i) \neq \rho(i)\}|.$$

Also define the measure

$$\mathbb{P}(A) = \frac{|A|}{n!}.$$

Then we have

Theorem 9 (Concentration on Symmetric Group). Let $f : S_n \rightarrow \mathbb{R}$, and $X \sim \text{Uniform}S_n$. Then f is automatically Lipschitz. We have

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq C \|f\|_{\psi_2} / \sqrt{n}.$$

§3 Application: Johnson Lindenstrauss Lemma

We always have to do data reduction in practice. A usual way of doing so is via projection. Assume that we have N data points in \mathbb{R}^n with n large. What we want to do is to choose a m dimensional subspace E with $m \ll n$, on which we can project the data nicely. It's hard to find the exact subspace and finding an appropriate dimension m is fair enough. The Johnson-Lindenstrauss lemma asserts that we can choose m such that, if we project the data to a randomly selected m -dimensional subspace, then the geometric structure of the data would be nicely preserved.

3.1 Grassmannian Manifolds

How do we characterize the randomness of "selecting the subspace"? And how do we use the concentration results that have been put before? This is where the Grassmannian steps in.

Definition 2 (Grassmannian). Let $G_{n,m}$ be the collection of all the m -dimensional subspace in \mathbb{R}^n . Equip it with the metric

$$d(E, F) := \|P_E - P_F\|.$$

Here P_E is the associated projection operator of the subspace E , and $\|\cdot\|$ is the operator norm. Define \mathbb{P} as the unique Haar measure on it.¹ Let $E \sim \text{Unifrom}(G_{n,m})$. We say that E is a random subspace in $G_{n,m}$.

We give our concentration result in this space.

Theorem 10 (Concentration on Grassmannian). Let $X \sim \text{Uniform}(G_{n,m})$, and Lipschitz function $f : G_{n,m} \rightarrow \mathbb{R}$, then

$$\|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} \leq C \|f\|_{\text{Lip}} / \sqrt{n}.$$

3.2 Main Theorem

Theorem 11 (Johnson Lindenstrauss). Let $X \subset \mathbb{R}^n$ be a finite data set, which means $|X| = N < \infty$. For fixed $\varepsilon > 0$, if we choose

$$m \geq C/\varepsilon^2 \log N.$$

Then with probability $1 - 2 \exp\{-c\varepsilon^2 m\}$, the random selected m dimensional subspace $E \in G_{n,m}$, normalized random projection $Q = \sqrt{n/m} P_E$ is an ε -approximate isometry on X , i.e.,

$$(1 - \varepsilon) \|x - y\|_2 \leq \|Qx - Qy\|_2 \leq (1 + \varepsilon) \|x - y\|$$

is true for all $x, y \in X$.

A key observation in proving this theorem lies in the finiteness of the data set. Once we've proved similar result for the fixed vector $x, y \in X$, the general result would be easy to obtain.

Lemma 2 (Random Projection of Fixed Vector). Let $E \sim \text{Unifrom}(G_{n,m})$, and P_E is the associated random projection. Let $z \in \mathbb{R}^n$ be fixed, then,

1. The L^2 norm $\mathbb{E}[\|P_E z\|_2^2]^{1/2} = \sqrt{m/n} \|z\|_2$.
2. With probability $1 - 2 \exp\{-c\varepsilon^2 m\}$, the following approximate isometry is true.

$$(1 - \varepsilon) \sqrt{m/n} \|z\| \leq \|P_E z\|_2 \leq (1 + \varepsilon) \|z\|.$$

¹Unique rotational invariance measure.

Proof. Without loss of generality, let $\|z\| = 1$. Note that, if we set $Z \sim \text{Uniform}(\mathbb{S}^{n-1})$. Then,

$$\|P_E z\| \sim \|P_{E_0} Z\|,$$

i.e. identical in distribution. Here E_0 is any fixed m dimensional subspace. Hence we choose

$$E_0 = \text{span}\{e_1, e_2, \dots, e_m\}.$$

And consequently,

$$\mathbb{E}[\|P_{E_0} Z\|_2^2] = \mathbb{E}[Z_1^2 + \dots + Z_m^2] = \frac{m}{n} \mathbb{E}[Z_1^2 + \dots + Z_n^2] = \frac{m}{n}.$$

Now we turn to the second part on concentration. Let

$$f(x) = \|P_{E_0} x\|_2.$$

Then f is clearly 1-Lipshitz, as composition of 1-Lipshitz continuous functions. Recall an exercise that, concentration on mean implies concentration on L^p norm, we have

$$\mathbb{P}\left(\left|\|PX\| - \sqrt{\frac{m}{n}}\right| > t\right) \leq 2 \exp\{-cnt^2\}.$$

Take $t = \varepsilon \sqrt{\frac{m}{n}}$, then with probability $1 - 2 \exp\{-cm\varepsilon^2\}$, we have

$$(1 - \varepsilon) \sqrt{\frac{m}{n}} \|z\| \leq \|Pz\| \leq (1 + \varepsilon) \sqrt{\frac{m}{n}} \|z\|.$$

■

Proof of the Main Theorem. Now we begin our formal proof of Johnson-Lindenstrauss lemma. Let $A := X - X = \{x - y : x, y \in X\}$, then for any fixed $z \in A$, with probability $1 - 2 \exp\{-2cm\varepsilon^2\}$, we have

$$(1 - \varepsilon) \sqrt{\frac{m}{n}} \|z\| \leq \|Pz\| \leq (1 + \varepsilon) \sqrt{\frac{m}{n}} \|z\|.$$

Denote the event as E_z , then $\mathbb{P}(E_z^c) \leq 2 \exp\{-cm\varepsilon^2\}$. By the monotonicity of probability measure, we have

$$\begin{aligned} \mathbb{P}\left(\bigcap_{z \in A} E_z\right) &= 1 - \mathbb{P}\left(\bigcap_{z \in A} E_z^c\right) \\ &\geq 1 - 2N^2 \exp\{-cm\varepsilon^2\}. \end{aligned}$$

By condition $m \geq (C/\varepsilon^2) \log N$, plug in to get

$$\mathbb{P}\left(\bigcap_{z \in A} E_z\right) \geq 1 - 2 \exp\left\{-\frac{1}{2} cm\varepsilon^2\right\}.$$

Which means, with probability $1 - 2 \exp\{-cm\varepsilon^2/2\}$, we have ε -approximate isometry,

$$(1 - \varepsilon) \|x - y\| \leq \|Qx - Qy\| \leq (1 + \varepsilon) \|x - y\|, \forall x, y \in X.$$

■

Remark. For concentration on L^p norm, just to note that

$$\begin{aligned} \|f(X) - \mathbb{E}[f(X)]\|_{L^p} &\leq \|f(X) - \mathbb{E}[f(X)]\|_{\psi_2} + \|\mathbb{E}[f(X)] - f(X)\|_{L^p} \\ &\leq C_1 \|f\|_{\text{Lip}} + C_2 \|f(X)\|_{L^p} \quad (\text{Constants are sub-Gaussian}) \\ &\leq (C_1 + C_2' \sqrt{p}) \|f\|_{\text{Lip}}. \quad (\text{Equivalence of sub-Gaussianity}) \end{aligned}$$