IFT 6760A - Lecture 14
Johnson – Lindenstrauss theorem

Scribe(s): Abdulmonhem Alkhalil , Adel Nabli, Timothy Nest, Tayssir Doghri **Instructor:** Ioannis Mitliagkas

1 Summary

In the previous lectures we presented some dimensionality reduction algorithms such as PCA and LLE. The purpose of these methods is, given a high dimensional dataset, find a lower dimensional space to project the data into that manages to keep some of the important structural characteristics of the original dataset. For example, in the case of the PCA, we aim at preserving as much variance as possible.
One important assumption that both PCA and LLE leverage to perform "good" dimensionality reduction is that the dataset intrinsically lives in a lower dimensional subspace than the one it is embedded in.

In this lecture we will present a method based on the Johnson-Lindenstrauss theorem that doesn't require such an hypothesis to guarantee "good results". In fact, we will even see that the targeted *lower dimension* doesn't even depend on the original dimension of the data but only on an *error rate* and on the number of points that make the dataset.
The structural aspect we will aim to preserve here is the pairwise euclidean distance between the data points: we want to find a low dimensional subspace in which the distance between each pair of points is the same as in the original high dimensional space. To do that, we will iterate a random projection method.

One application of such a method is performing linear regression in high dimensional datasets: when traditional methods for solving the linear regression problem are computationally too expensive to perform in high dimension, one can use this random projection method to speed up the computation without affecting "too much" the end result.

Note: *To give some intuition about what we are doing, we did not specified the terms between quotation marks, but it is important to note that the measure of goodness of a given method is only defined with respect to its pursued objective, which is precisely what distinguishes the listed methods of dimensionality reduction from one another. Thus, it does not make much sense to directly compare those methods as they are designed to fulfill different goals.*

2 An outline of the method

In this section, we will try to explain the algorithm used to find the low dimensional projection, the formal definitions and proofs will be given in the next section. First, let's formalize our goal.

Goal 1. *For any $0 < \epsilon < 1$ and any set V of n points in $\mathbb{R}^d$, we want to find a map $f : \mathbb{R}^d \to \mathbb{R}^k$ where $k << d$ such that for any pair $\mathbf{u}, \mathbf{v}$ of V, we have:*

$$(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2 \leq \|f(\mathbf{u})-f(\mathbf{v})\|_2^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2$$

Note: *In the original formulation of the Johnson-Lindenstrauss theorem, we set k to be equal to $k = O\Big(\frac{\ln(n)}{\epsilon^2}\Big)$.*

To give a bit of an intuition on what we are looking for, let's consider the case of $d = 2$, $n = 2$ and $k = 1$:

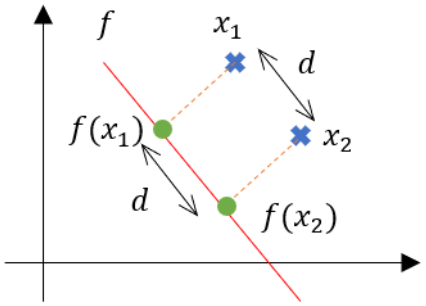

Figure 1: Example with two points in $\mathbb{R}^2$

In this example, we found a map $f : \mathbb{R}^2 \to \mathbb{R}$ such that the pairwise distance is preserved: $\|f(\mathbf{x_1}) - f(\mathbf{x_2})\|_2 = \|\mathbf{x_1} - \mathbf{x_2}\|_2$.

We want to generalize this in arbitrary dimension d with arbitrarily many points n. To do so, we will use the probabilistic method, which is defined as follows in [3]:

Definition 2 (Probabilistic method)**.** *Trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability.*

To rephrase it in our particular case, for a given ϵ and dataset V, we will fix k to be equal to a precise value, and to prove that a map f following **Goal 1** exists, we will first specify a set of maps $\mathcal{F}_k$ and then show that a randomly picked one in this set will follow **Goal 1** with a probability strictly superior to 0. Therefore, as the probability to find a "good" map is non null, there must exist one that follows the desired properties. Thus, to find one, it suffices to pick at random sufficiently many of them in $\mathcal{F}_k$.

Having understood the spirit of the method, we will now explain in more details the structure of the proof proposed in [4] that we will present in the next section:

Outline of the proof:

- **Step 1:** Define a set $\mathcal{F}_k$ of possible projections in $\mathbb{R}^k$ and pick one f in this set at random.
- **Step 2:** Show that for any pair of points $\mathbf{u}, \mathbf{v}$ of V, f preserves the squared l_2-norm with a probability $\geq 1 - \frac{2}{n^2}$
  1. First, show that **in average** f will preserve the squared distance between two points
  2. Then, using a concentration inequality, show that the probability of f being far from this average is $\leq \frac{2}{n^2}$
- **Step 3:** Using a union bound, show that the probability of f preserving **all** of the pairwise squared distances in V is greater than $\frac{1}{n}$
- **Step 4:** Deduce that sampling $O(n)$ different f from $\mathcal{F}_k$ produces a method that will eventually manage to pick with **constant probability** an f that satisfies **Goal 1**.

  Note: *It is this property that implies that f can be found in "randomized polynomial time" in the Johnson - Lindenstrauss theorem.*

3 Proof of the Johnson-Lindenstrauss theorem

We have a set V of n points in $\mathbb{R}^d$ and we fix an ϵ such that $0 < \epsilon < 1$.

3.1 Step 1: sampling a random projection

Let's take $k \in \mathbb{N}$ such that

$$k \geq \frac{4\ln(n)}{\epsilon^2/2 - \epsilon^3/3} \tag{1}$$

and let's consider $\mathcal{F}_k = \left\{ \frac{1}{\sqrt{k}}\mathbf{R} \in \mathbb{R}^{k\times d} | \ \mathbf{R}_{i,j} \overset{i.i.d}{\sim} \mathcal{N}(0,1) \right\}$. We sample $\mathbf{F}$ from $\mathcal{F}_k$ and define f as $f : \begin{array}{ccc} \mathbb{R}^d & \to & \mathbb{R}^k \\ \mathbf{x} & \mapsto & \mathbf{F}\mathbf{x} \end{array}$.

Note that we can look at $\mathbf{F}$ as $\mathbf{F} = \frac{1}{\sqrt{k}} \begin{pmatrix} -\mathbf{R}_1- \\ \vdots \\ -\mathbf{R}_k- \end{pmatrix}$ with $\forall i \in [\![1,k]\!],\ \mathbf{R}_i \sim \mathcal{N}(0, \mathbf{I}_{d\times d})$ and then $f : \mathbf{x} \mapsto \frac{1}{\sqrt{k}}\Big(\mathbf{x}^T\mathbf{R}_1, ..., \mathbf{x}^T\mathbf{R}_k\Big)^T$.

3.2 Step 2: distance preservation between each pairs of points with high probability

Lemma 3. *For any pair* $\mathbf{u}, \mathbf{v}$ *of* V, $\mathbb{E}\Big[\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2\Big] = \|\mathbf{u} - \mathbf{v}\|_2^2$

Proof. Let's re-write the expectation as:

$$\begin{aligned}\mathbb{E}\Big[\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2\Big] &= \mathbb{E}\Big[\sum_{i=1}^k \big(f(\mathbf{u})_i - f(\mathbf{v})_i\big)^2\Big] = \mathbb{E}\Big[\sum_{i=1}^k \frac{1}{k}\big(\mathbf{u}^T\mathbf{R}_i - \mathbf{v}^T\mathbf{R}_i\big)^2\Big] \\ &= \mathbb{E}\Big[\sum_{i=1}^k \frac{1}{k}(\mathbf{u}-\mathbf{v})^T\mathbf{R}_i\mathbf{R}_i^T(\mathbf{u}-\mathbf{v})\Big] = \sum_{i=1}^k \frac{1}{k}(\mathbf{u}-\mathbf{v})^T\,\mathbb{E}\Big[\mathbf{R}_i\mathbf{R}_i^T\Big](\mathbf{u}-\mathbf{v})\end{aligned}$$

But $\forall i \in [\![1,k]\!],\ \mathbf{R}_i = \begin{pmatrix} r_{i,1} \\ \vdots \\ r_{i,d} \end{pmatrix}$ with the $r_{i,j} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. Thus,

$$\mathbb{E}\Big[\mathbf{R}_i\mathbf{R}_i^T\Big] = \Big(\mathbb{E}[r_{i,m}r_{i,l}]\Big)_{l,m\in[\![1,d]\!]} = \begin{cases} \mathbb{E}[r_{i,m}^2] = \mathbb{V}[r_{i,m}] = 1 \ if\ l = m \\ \mathbb{E}[r_{i,m}]\,\mathbb{E}[r_{i,l}] = 0 \ otherwise \end{cases}$$

Then, we have:

$$\mathbb{E}\Big[\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2\Big] = \sum_{i=1}^k \frac{1}{k}(\mathbf{u}-\mathbf{v})^T\mathbf{I}_{d\times d}(\mathbf{u}-\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|_2^2$$

□

Lemma 4. *For any pair* $\mathbf{u}, \mathbf{v}$ *of* V*, we have:*

$$\mathbb{P}\Big(\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2 > (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2\Big) \leq \frac{1}{n^2} \tag{2}$$

$$\mathbb{P}\Big(\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2 < (1-\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2\Big) \leq \frac{1}{n^2} \tag{3}$$

Proof. In the end, what we want to prove is that the probability of $\frac{\|f(\mathbf{u})-f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2}$ being outside of $[1-\epsilon, 1+\epsilon]$ is smaller than $\frac{2}{n^2}$.
To do that, let's focus on (2) as showing (3) follows the same principle. First, let's notice that if we set $\mathbf{x} = \mathbf{u} - \mathbf{v} \in \mathbb{R}^d$ and use the definition of f as introduced in **Step 1**, we have:

$$\frac{\|f(\mathbf{u}) - f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2} = \frac{1}{k}\frac{\|\mathbf{R}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} = \frac{1}{k\|\mathbf{x}\|_2^2}\sum_{i=1}^k \big(\mathbf{R}_i^T\mathbf{x}\big)^2 = \frac{1}{k}\sum_{i=1}^k \Big(\mathbf{R}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\Big)^2 \tag{4}$$

But, $\forall i \in [\![1,k]\!],\ z_i := \mathbf{R}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|_2} = \sum_{j=1}^d r_{i,j}\frac{x_j}{\|\mathbf{x}\|_2}$ with $r_{i,j} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. Using the fact that a sum of i.i.d gaussian random variables is still a gaussian random variable, we find that the $\mathbf{R}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|_2} \sim \mathcal{N}\Big(0, \underbrace{\frac{1}{\|\mathbf{x}\|_2^2}\sum_{j=1}^d x_j^2}_{=1}\Big)$ and then the

$z_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. Thus, using (4), we can write that proving (2) is equivalent to prove $\mathbb{P}\Big(\sum_{i=1}^{k} z_i^2 > k(1+\epsilon)\Big) \leq \frac{1}{n^2}$. Introducing a dummy variable α, we write that, for any $\alpha > 0$:

$$\mathbb{P}\Big(\sum_{i=1}^{k} z_i^2 > k(1+\epsilon)\Big) = \mathbb{P}\Big(e^{\alpha \sum_i z_i^2} > e^{\alpha k(1+\epsilon)}\Big) \tag{5}$$

$$Markov\ inequality \rightarrow \leq \frac{\mathbb{E}\left[e^{\alpha \sum_i z_i^2}\right]}{e^{\alpha k(1+\epsilon)}} \tag{6}$$

$$the\ z_i\ are\ i.i.d \rightarrow = \frac{\mathbb{E}\left[e^{\alpha z_1^2}\right]^k}{e^{\alpha k(1+\epsilon)}} \tag{7}$$

We now recognize the **moment generating function** of χ_1^2 evaluated in α, which leads us to write [1]:

$$\mathbb{E}\left[e^{\alpha z_1^2}\right]^k = \Big(\frac{1}{\sqrt{1-2\alpha}}\Big)^k \tag{8}$$

As (8) holds as long as $\alpha < 1/2$, setting $\alpha = \frac{\epsilon}{2(1+\epsilon)}$ is possible *(this particular value of α is the one that minimizes the right side of the inequality (7))*. Then, for this α, we have:

$$\mathbb{P}\Big(\sum_{i=1}^{k} z_i^2 > k(1+\epsilon)\Big) \leq \Big(\frac{1}{1-\frac{\epsilon}{2(1+\epsilon)}}\Big)^{k/2} e^{-\frac{\epsilon}{2(1+\epsilon)}k(1+\epsilon)} = (1+\epsilon)^{k/2} e^{-\epsilon k/2} \tag{9}$$

But, we know that $\forall x \geq 0,\ \ln(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}$. Thus, we can use that in (9) to upper bound the $(1+\epsilon)^{k/2}$ term:

$$\mathbb{P}\Big(\sum_{i=1}^{k} z_i^2 > k(1+\epsilon)\Big) \leq e^{\frac{\epsilon k}{2} - \frac{k}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)} e^{-\epsilon k/2} = e^{-\frac{k}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)}$$

Finally, we can use the lower bound on k given in (1) to write:

$$\mathbb{P}\Big(\sum_{i=1}^{k} z_i^2 > k(1+\epsilon)\Big) \leq e^{-2\ln(n)} = \frac{1}{n^2}$$

□

3.3 Step 3: preservation of all the pairwise distances in V with non-zero probability

Lemma 5. $\mathbb{P}\left(\bigcap_{\mathbf{u}\neq\mathbf{v}\in V}\left\{\frac{\|f(\mathbf{u})-f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2} \in [1-\epsilon, 1+\epsilon]\right\}\right) \geq \frac{1}{n}$

Proof. As we know that in a set of n points there is $\frac{n(n-1)}{2}$ different pairs, we can write:

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{\mathbf{u}\neq\mathbf{v}\in V}\left\{\frac{\|f(\mathbf{u})-f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2} \in [1-\epsilon, 1+\epsilon]\right\}\right) &= 1 - \mathbb{P}\left(\bigcup_{\mathbf{u}\neq\mathbf{v}\in V}\left\{\frac{\|f(\mathbf{u})-f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2} \notin [1-\epsilon, 1+\epsilon]\right\}\right) \\
&\geq 1 - \sum_{\mathbf{u}\neq\mathbf{v}\in V} \mathbb{P}\left(\left\{\frac{\|f(\mathbf{u})-f(\mathbf{v})\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2} \notin [1-\epsilon, 1+\epsilon]\right\}\right) \\
using\ Lemma\ 4 \rightarrow &\geq 1 - \frac{n(n-1)}{2}\frac{2}{n^2} = \frac{1}{n}
\end{aligned}$$

□

3.4 Conclusion: the Johnson-Lindenstrauss theorem

Theorem 6 (Johnson-Lindenstrauss)**.** *For any $0 < \epsilon < 1$ and any integer n, let k be a positive integer such that*

$$k \geq \frac{4\ln(n)}{\epsilon^2/2 - \epsilon^3/3}$$

Then, for any set V of n points in $\mathbb{R}^d$, there is a map $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $\mathbf{u}, \mathbf{v} \in V$,

$$(1-\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2 \leq \|f(\mathbf{u}) - f(\mathbf{v})\|_2^2 \leq (1+\epsilon)\|\mathbf{u}-\mathbf{v}\|_2^2$$

Furthermore this map can be found in randomized polynomial time.

Note: *Alon showed in **Theorem 9.3** of [2] that all maps trying to satisfy **Goal 1** must have a target dimension k which is at least $\Omega\Big(\frac{\log(n)}{\epsilon^2 \log(\epsilon)}\Big)$, making the lower bound of the Johnson-Lindenstrauss theorem almost optimal.*

4 Application of the Johnson-Lindenstrauss theorem to regression

Once we defined Johnson-Lindenstrauss theorem, we present in this section one of its application as a dimensiality reduction tool to regression [5]. Given a $m \times p$ matrix $\mathbf{A}$ where $m \gg p$, we try to solve this linear regression problem:

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 \tag{10}$$

This problem is well studied and there are several methods to solve it exactly. But as [5] points out, the naive matrix multiplication solving the normal equation for least squares is of complexity $O(mp^2)$ and linear programs are of order at least $O(m^3)$ which is prohibitive when we are dealing with lots of data *(large m)*. To improve upon those complexities, one idea is to relax the objective : instead of finding $\mathbf{x}^*$ solution of (10), we want to find an $\mathbf{x}$ that will allow $||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$ to be *"not too far from its true minimum"*, which we could write as:

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \leq (1+\epsilon)\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2^2 \tag{11}$$

One way of doing that while reducing the complexity of the method would be to *"project the problem"* in a lower dimension space and solve it there. Let's write this new problem as:

$$\min_{\mathbf{x}} ||\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})||_2^2 \tag{12}$$

where $\mathbf{S}$ is a $k \times m$ matrix and $k \ll m$. Then, solving (12) becomes tractable if we relocated the problem to a sufficiently low dimensional one. The question is then: *How can we make sure that solving the lower dimensional problem (12) brings a solution $\mathbf{x}$ that respects our relaxed goal (11) ?* If we could find an $\mathbf{S}$ such that:

$$\forall \mathbf{x} \in \mathbb{R}^p,\ ||\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})||_2^2 = (1 \pm \epsilon)\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{13}$$

then the solution of (12) will verify the relaxed goal (11). And if such an $\mathbf{S} \in \mathbb{R}^{k \times m}$ can also verify $k \ll m$, we would be done.

To see how we could find a good $\mathbf{S}$, let's rewrite our goal (13) in a simpler way and see if it rings a bell. First, let's rename $\mathbf{u}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^p$. We have that $\mathbf{u}, \mathbf{b} \in \mathbb{R}^m$. Then, we can rewrite (13) and say that we want to find an $\mathbf{S}$ such that:

$$\forall \mathbf{x} \in \mathbb{R}^p,\ (1-\epsilon)\|\mathbf{u}(\mathbf{x}) - \mathbf{b}\|_2^2 \leq ||\mathbf{S}\mathbf{u}(\mathbf{x}) - \mathbf{S}\mathbf{b}||_2^2 \leq (1+\epsilon)\|\mathbf{u}(\mathbf{x}) - \mathbf{b}\|_2^2 \tag{14}$$

which looks suspiciously close to the goal reached by the Johnson-Lindenstrauss theorem *(here we would have that the d from **Theorem 6** be equal to m)*. The only thing that prevents us from directly set the $\mathbf{S}$ we are looking for to be equal to the linear map f found by the Johnson-Lindenstrauss theorem is that this theorem can only be applied when the set of vectors V is finite, whereas here $\mathbf{u}(\mathbf{x})$ can take infinitely many values. Let's remedy this problem.

First, if we set $\mathbf{y} := \mathbf{Ax} - \mathbf{b}$, let's notice that if the requirement (13) is met for any unit vector $\mathbf{y}$ then it is satisfied for all vector $\mathbf{y}$ simply by scaling and because $\mathbf{S}$ is a linear map. Thus let's define $\mathcal{S}$ the set we will work on:

$$\mathcal{S} = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{Ax} - \mathbf{b},\ \mathbf{x} \in \mathbb{R}^p,\ \|\mathbf{y}\|_2 = 1\} \subset S^{m-1}\ (unit\ sphere) \tag{15}$$

Then, to be able to use the Johnson-Lindenstrauss theorem, we will find a **finite set** *(so that we can apply the JL theorem on it)* $V \subset \mathcal{S}$ not too big *(remember that the JL theorem requires setting a k that depends on the size n of the set V and that we want here k being low)*, but big enough so that every vector $\mathbf{y}$ of $\mathcal{S}$ is close to at least one $\mathbf{v} \in V$ in a way such that $\|\mathbf{y} - \mathbf{v}\|_2 \leq 1/2$ *(a V verifying this property is called a $\frac{1}{2}$-net)*. We will show that:

1. **Lemma 7**: Having V a $\frac{1}{2}$-net is sufficient to extend the result of the Johnson-Lindenstrauss theorem for the map $\mathbf{S}$ found for V to all of $\mathcal{S}$.
2. **Lemma 8**: The V we had to pick is not too big.

Lemma 7. *If $\forall \mathbf{y} \in \mathcal{S}\ ,\ \exists v \in V\ s.t\ \|\mathbf{y} - \mathbf{v}\|_2 \leq \frac{1}{2}$, then $\forall \mathbf{y} \in \mathcal{S},\ \|\mathbf{Sy}\|_2^2 = (1 \pm \epsilon)$ with $\mathbf{S}$ the mapping found by the Johnson-Lindenstrauss procedure on V*

Proof. Here we suppose that $\forall \mathbf{y} \in \mathcal{S}\ ,\ \exists v \in V\ s.t\ \|\mathbf{y} - \mathbf{v}\|_2 \leq \frac{1}{2}$. Then, let's take an arbitrary $\mathbf{y} \in \mathcal{S}$ and show that the Johnson-Leidenstrauss property still holds. We have:

$$\forall \mathbf{y} \in \mathcal{S},\ \exists \mathbf{y_0} \in V\ s.t\ \mathbf{y} = \mathbf{y_0} + (\mathbf{y_0} - \mathbf{y})\ with\ \|\mathbf{y} - \mathbf{y_0}\|_2 \leq \frac{1}{2}$$

But then, $\frac{\mathbf{y}-\mathbf{y_0}}{\|\mathbf{y}-\mathbf{y_0}\|_2} \in \mathcal{S}$ and we have $\exists \mathbf{z} \in V\ s.t\ \mathbf{y} - \mathbf{y_0} = \mathbf{z} + \big((\mathbf{y_0} - \mathbf{y}) - \mathbf{z}\big)$ with:

$$\|\frac{\mathbf{y} - \mathbf{y_0}}{\|\mathbf{y} - \mathbf{y_0}\|_2} - \mathbf{z}\|_2 \leq \frac{1}{2} \Leftrightarrow \|\mathbf{y} - \mathbf{y_0} - \underbrace{\mathbf{z}\|\mathbf{y} - \mathbf{y_0}\|_2}_{:=\mathbf{y_1}}\|_2 \leq \frac{\|\mathbf{y} - \mathbf{y_0}\|_2}{2} \leq \frac{1}{4}$$

Then, by induction, we have that:

$$\forall \mathbf{y} \in \mathcal{S},\ \mathbf{y} = \mathbf{y_0} + \mathbf{y_1} + \mathbf{y_2} + ...$$

with $\forall i$, $\|\mathbf{y_i}\|_2 \leq \frac{1}{2^i}$ and y_i a scalar multiple of an element of V. As we can re-write the JL property on V as a property on the scalar product:

$$\forall \mathbf{u}, \mathbf{v} \in V, \langle \mathbf{Su}, \mathbf{Sv} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \pm \epsilon \tag{16}$$

we can then write:

$$\begin{aligned}
\forall \mathbf{y} \in \mathcal{S},\ \|\mathbf{Sy}\|_2^2 &= \|\mathbf{S}(\mathbf{y_0} + \mathbf{y_1} + \mathbf{y_2} + ...)\|_2^2 \\
&= \Big(\sum_{0 \leq i < j < \infty} \|\mathbf{Sy_i}\|_2^2 + 2\langle \mathbf{y_i}, \mathbf{y_j} \rangle\Big) \pm \epsilon \sum_{0 \leq i < j < \infty} \|\mathbf{y_i}\|_2 \|\mathbf{y_j}\|_2 \\
&= 1 \pm O(\epsilon)
\end{aligned}$$

Which is what we want to have (if we rescale ϵ properly). □

Lemma 8. *We can find a $\frac{1}{2}$-net $V \subset \mathcal{S}$ of finite size $|V| \leq 9^{p+1}$ s.t $\forall y \in \mathcal{S},\ \exists v \in V$ s.t $\|\mathbf{y} - \mathbf{v}\|_2 \leq \frac{1}{2}$*

Proof. A proof of this lemma can be found in [5] in the proof of its **Lemma 5**. □

What is interesting with this result is that $n = |V|$ doesn't depend on m but only on p. Then, as **Theorem 6** requires a dimension $k \geq \frac{4\ln(|V|)}{\epsilon^2/2 - \epsilon^3/3}$, here we could pick $k = \frac{(p+1)4\ln(9)}{\epsilon^2/2 - \epsilon^3/3}$ and solve (12) with complexity $O(k^3)$ using a linear program, at the expense of having to compute the matrix multiplication $\mathbf{SA}$. But in [5] it is said this multiplication could be done in $O(nnz(\mathbf{A}))$ with $nnz(\mathbf{A})$ the number of non-zero entries of $\mathbf{A}$.

Thus, we can say that we found a way to solve approximately (10) in a way that doesn't depend on the number of points m we are considering but only on the dimension $p << m$ which is exactly what we were looking for.

References

[1] Wikipedia page of the moment generating function. URL `https://en.wikipedia.org/wiki/Moment-generating_function`.

[2] N. Alon. Problems and results in extremal combinatoricsi. *Discrete Mathematics*, 273(1):31 – 53, 2003.

[3] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley, 2000.

[4] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[5] D. P. Woodruff. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10:1–157, Oct. 2014. ISSN 1551-305X.