Non-Asymptotic Theory of Random Matrices
Lecture 4: Dimension Reduction

Date: January 16, 2007

Lecturer: Roman Vershynin   Scribe: Matthew Herman

1 Introduction

Consider the set $X = \{n \text{ points in } \mathbb{R}^N\}$ where both n and N are large. Our goal will be to reduce N so as to represent X in a space of dimension $k \ll N$. One motivation for this comes from computer science where X is a data structure.

Observation 1.

1. *We can always take $N = n$, since dim(span(X)) $= n$ and $X \subset$ span(X).*
2. *For $N < n$ we may lose linear independence.*

Goal: We want to construct a map $T : \mathbb{R}^N \to \mathbb{R}^k$, $k << N$ such that pairwise distances in X are *approximately* preserved. That is, for all $x, y \in X$ there exists $\varepsilon > 0$ such that

$$(1-\varepsilon)\|x-y\|_2 \;\le\; \|Tx - Ty\|_2 \;\le\; (1+\varepsilon)\|x-y\|_2.$$

Such a map is called an ε-embedding of our data structure X into $\mathbb{R}^k$ (more precisely into ℓ_2^k). For us, T will be the linear map associated with a random matrix.

Theorem 2 (Johnson-Lindenstrauss (J-L) Flattening Lemma [3])**.** *Let X be an n-point set in a Hilbert space, and suppose $\varepsilon \in (0,1)$. Then there exists an ε-embedding of X into ℓ_2^k, for some $k \le C\varepsilon^{-2}\log n$ and some constant $C > 0$.*

Example 3. *To represent n points in $\mathbb{R}^n$ in a computer we need to store n^2 numbers. However, by the J-L Flattening Lemma we can store only $O(n\log n)$ numbers and still reconstruct all distances within ε-error.*

All known embeddings satisfying Theorem 2 are given by random matrices T such as

- Gaussian
- Bernoulli (Achlioptas [1])
- Orthogonal projections (J-L)

1.1 Random Projections in $\mathbb{R}^n$

- Random rotations (= random orthogonal matrices)
- Orthogonal group $O_n = \{\text{rotations in } \mathbb{R}^n\} = \{\text{orthogonal } n \times n \text{ matrices}\}$

Here, O_n is a probability space with probability measure called "Haar Measure" (from topological group theory).

1.2 Haar Measure

Suppose that M is a compact metric space (such as a sphere in $\mathbb{R}^n$), and that G is a group of isometries of M.

Theorem 4 (Haar Measure)**.**

1. *There exists a Borel probability measure μ on M which is invariant under G. That is, $\mu(S) = \mu(gS)$ for all $g \in G$ and $S \subset M$.*
2. *If G is transitive, then the Haar measure is unique. Here, "transitive" means for all $x, y \in X$, there exists $g \in G$ such that $gx = y$.*

Proof: See §1 of Milman-Schechtmann [7] for a simple 2-page proof.

Example 5. *Let $M = S^{n-1}, G = O_n$. Then μ = usual Lebesgue measure.*

Example 6. *Let $M = G = O_n$. Here the metric on a rotation comes from the Hilbert-Schmidt norm. In this sense $n \times n$ matrices can be viewed as "vectors" in $\mathbb{R}^{n^2}$, and the Hilbert-Schmidt norm provides a Euclidean distance between these matrices. This gives us a Haar measure on O_n.*

1.3 Random rotations, random orthogonal matrices

Fact 7 (How to compute Haar Measure)**.** *For a given $z \in S^{n-1}$ and for a given random rotation (with respect to Haar measure) $U \in O_n$, Uz is a random vector which is uniformly distributed on S^{n-1}.*

Proof. We want to show for all $A \subset S^{n-1}$ that we have $\mathbb{P}(Uz \in A) = \sigma(A)$, where $\mathbb{P}$ denotes probability (with respect to Haar measure) and σ is the usual uniform measure. Note that both sides of this equation define a probability measure on S^{n-1} which are rotationally invariant. The uniqueness part of the Haar Measure theorem implies that they coincide. □

Definition 8.

- *Random k-subspace of $\mathbb{R}^n$ = random rotation of $\mathbb{R}^k \subset \mathbb{R}^n$, denoted by $U^*(\mathbb{R}^k)$*
- *Random Projection = orthogonal projection onto a random k-subspace, denoted by $U^* P_k U$, where $P_k : \mathbb{R}^n \to \mathbb{R}^k$ is a projection (i.e., $P_k U$ = the first k rows of matrix U)*

Proof of J-L Lemma (for ε-embedding given by random projections)
Layout of proof: First we will show that J-L holds for one particular fixed pair $x, y \in X$. Then we will show that it holds collectively for all possible pairs in X.

Step 1. Fix an arbitrary pair $x, y \in X$, and let $z = x - y$.

Lemma 9 ("Norm" of a random projection)**.** *Let $z \in \mathbb{R}^n$ be fixed, and let P be a random projection in $\mathbb{R}^n$ onto a k-subspace. Then*

1. *$(\mathbb{E}\|Pz\|_2^2)^{1/2} = \sqrt{k/n}$, where $\mathbb{E}$ denotes expectation*
2. *For $\varepsilon > 0$ we have that*

$$(1-\varepsilon)\sqrt{k/n} \;\le\; \|Pz\|_2 \;\le\; (1+\varepsilon)\sqrt{k/n}$$

holds with probability $1 - 2e^{-k\varepsilon^2/2}$

Proof. 1. Note that the statement "Projecting a fixed z onto a random subspace" is equivalent to the statement "Projecting a random z onto a fixed subspace." Then with random rotation U we have

$$\|Pz\|_2 = \|U^* P_k U z\|_2 = \|P_k U z\|_2 = \|P_k x\|_2$$

where x is a random vector distributed uniformly on S^{n-1}. Then

$$\mathbb{E}\,\|P_k x\|_2^2 \;=\; \mathbb{E}\sum_{j=1}^{k} x_j^2 \;=\; \sum_{j=1}^{k}\mathbb{E}\,x_j^2 \;=\; k\,\mathbb{E}\,x_1^2 \;=\; k/n.$$

2. For the next part of Lemma 9 we want to show

$$\begin{aligned}\text{Failure Probability} \;&\le\; \mathbb{P}\Big(\Big|\|Pz\|_2 - \sqrt{k/n}\Big| > \varepsilon\sqrt{k/n}\Big) \\ &=\; \sigma\Big(\{x \in S^{n-1} : \Big|\|P_k x\|_2 - \sqrt{k/n}\Big| \;>\; \varepsilon\sqrt{k/n}\}\Big).\end{aligned}$$

Let $f(x) = \|P_k x\|_2$, and recognize that $f : S^{n-1} \to \mathbb{R}^n$ is a 1-Lipschitz function. From Part 1 of this lemma we have that

$$\Big(\mathbb{E}f^2\Big)^{1/2} = \sqrt{k/n}.$$

Let $p \;=\; \sigma(\{x \in S^{n-1} : |f(x) - (\mathbb{E}f^2)^{1/2}| \;>\; \varepsilon\sqrt{k/n}\})$. Then

$$p \;\le\; 2e^{-n(\varepsilon\sqrt{k/n}\,)^2/2} \;=\; 2e^{-k\varepsilon^2/2}$$

where we used the results of Concentration of Measure discussed in §2 of Lecture 3. □

We next normalize our embedding as $T := \sqrt{n/k}\,P$. By Lemma 9, we have for any *fixed* $z \in X$ that

$$(1-\varepsilon)\|z\|_2 \;\le\; \|Tz\|_2 \;\le\; (1+\varepsilon)\|z\|_2 \tag{1}$$

holds with high probability. Thus for any fixed $x, y \in X$ and putting $z = x - y$ we have that the inequality (1) holds with probability $1 - 2e^{-k\varepsilon^2/2}$.

Now we consider the *collective ensemble.* Notice that there are no more than n^2 pairs of $x, y \in X$. Then taking the union over all pairs we have the probability that (1) fails for some pair is less than or equal to

$$n^2 2e^{-k\varepsilon^2/2}$$

which is less than 1 if $k \ge C\varepsilon^{-2}\log n$. Hence with positive probability (1) holds with for all pairs. (*End of proof for the J-L Flattening Lemma.*) ■

Example 10 (Argument that $k \geq C(\varepsilon)\log n$ using volume point of view). *Suppose for all $x, y \in X$ that*

$$1/2 \;\leq\; \|x-y\|_2 \;\leq\; 1.$$

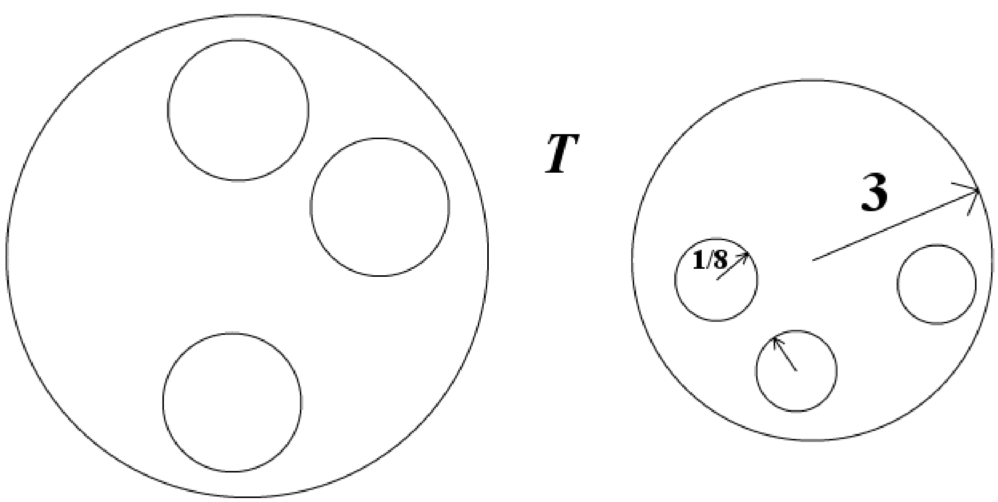
Now suppose that $\varepsilon = 1/10$. Then by the J-L Flattening Lemma we can claim that

$$1/4 \;\leq\; (1-\varepsilon)/2 \;\leq\; \|Tx - Ty\|_2 \;\leq\; (1+\varepsilon) \;\leq\; 2.$$

This tell us the following:

- *The image of X under T is contained in a ball of radius 2 (i.e., $T(X) \subset 2B_2^k$)*
- *There are n disjoint $\frac{1}{8}$-balls centered at points in $T(X) \subset (2+\frac{1}{8})B_2^k \subset 3B_2^k$ if*

$$\begin{aligned} n \cdot \mathsf{Vol}(\tfrac{1}{8}B_2^k) &\leq \mathsf{Vol}(3B_2^k) \\ n \cdot (\tfrac{1}{8})^k \,\mathsf{Vol}(B_2^k) &\leq 3^k \,\mathsf{Vol}(B_2^k) \\ n &\leq 24^k \\ \log n &\lesssim k. \end{aligned}$$



The last example showed that the $\log n$ factor in the J-L Flattening Lemma is sharp. The sharpness of the ε^{-2} factor is due to Alon and can be found in [6].

2 Related Work

Is there a version for

- ℓ_1? No: Lee and Naor [4], Brinkman and Charikar [2].
- ℓ_∞? No.
- ℓ_p (for $p \neq 1, 2, \infty$)? Still open.

3 Related Readings

- Matousek [6] and [5]
- Vempala [8]

References

[1] D. Achlioptas. Database-friendly random projections. *Proc. 20th Annual Symposium on Principles of Database Systems*, 26:274–281, 2001.

[2] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. *Journal of ACM*, 52(5):766–788, 2005.

[3] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[4] J. R. Lee and A. Naor. Embedding the diamond graph in l_p and dimension reduction in l_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004.

[5] J. Matousek. Website on open problems on embeddings of metric spaces. http://kam.mff.cuni.cz/˜matousek/.

[6] J. Matousek. *Lectures on Discrete Geometry.* Springer, Berlin-Heidelberg New York, 2002.

[7] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1986.

[8] Santosh S. Vempala. *The random projection method*, volume 65 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science.* American Mathematical Society, Providence, 2004.