Massachusetts Institute of Technology Lecturer: Michel X. Goemans
**18.409: Topics in TCS: Embeddings of Finite Metric Spaces**
September 27, 2006 Scribe: Benjamin Rossman

---

# Lecture 6

Today we look at dimension reduction in $\ell_2$. Suppose $X$ is a metric space in $\ell_2$ of size $n$. From previous lectures, we know that $X$ embeds isometrically into $\ell_2^n$. We ask the question: for $k < n$, what is the minimal distortion $D$ needed to embed $X$ in $\ell_2^d$? We will see that there is a tradeoff between distortion and dimension. To achieve distortion close to 1, we need only logarithmic many dimensions.

**Theorem 1** (Johnson-Lindenstrauss, 1984). *For all $\varepsilon > 0$, $X$ embeds into $\ell_2^{O(\frac{1}{\varepsilon^2}\log n)}$ with distortion $1 + \varepsilon$.*

We also prove a theorem of Alon which shows that the Johnson-Lindenstrauss Lemma (as Theorem 1 is known) is tight.

**Theorem 2** (Alon [1]). *If $v_1, \ldots, v_{n+1} \in \mathbb{R}^d$ are such that $1 \leq \|v_i - v_j\| \leq 1 + \varepsilon$ for all $i \neq j$, then $d = \Omega(\frac{\log n}{\varepsilon^2 \log \frac{1}{\varepsilon}})$.*

We give two proofs of the Johnson-Lindenstrauss Lemma. The idea in both proofs is to project $X$ onto a random $k$-dimensional subspace of $\mathbb{R}^n$ where $k = O(\frac{1}{\varepsilon^2}\log n)$. The proofs differ in the way the projection is randomly chosen.

## Measure Concentration and Levy's Lemma

Let $S_{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ and let $\mu$ be the unique rotation-invariant (Haar) measure on $S_{n-1}$ such that $\mu(S_{n-1}) = 1$. For points $x, y \in S_{n-1}$, $d(x, y)$ denotes the geodesic distance between $x$ and $y$ defined by $d(x, y) = \arccos(\langle x, y\rangle)$. For a point $a \in S_{n-1}$ and $r \geq 0$, $B_a(r)$ denotes the cap of radius $r$ around $a$ defined by $B_a(r) = \{x \in S_{n-1} : d(a, x) \leq r\}$. We will need the following lemma:

**Lemma 3** (Levy's Lemma). *Let $A \subseteq S_{n-1}$ be a closed set and let $B \subseteq S_{n-1}$ be a cap such that $\mu(A) = \mu(B)$. Then, for all $t \geq 0$,*

$$\mu(\{x : d(A, x) \leq t\}) \geq \mu(\{x : d(B, x) \leq t\}).$$

*In particular, if $B = B_a(r)$ then $\mu(\{x : d(A, x) \leq t\}) \geq \mu(B_a(r + t))$.* □

We remark that Levy's Lemma also holds when $d(\cdot, \cdot)$ denotes Euclidean instead of geodesic distance.

**Lemma 4.** *Consider a function $f : S_{n-1} \to \mathbb{R}$ which is 1-Lipschitz, meaning that $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in S_{n-1}$. We define $m(f) \in \mathbb{R}$, called the median of $f$, such that $\mu(A^+) \geq \frac{1}{2}$ and $\mu(A^-) \geq \frac{1}{2}$ where $A^+ = \{x : f(x) \geq m(f)\}$ and $A^- = \{x : f(x) \leq m(f)\}$. Then*

$$\mu(\{x : |f(x) - m(f)| > \varepsilon\}) \leq (1 + o(1))e^{-\frac{\varepsilon^2 n}{2}}.$$

This lemma says that 1-Lipschitz functions are highly concentrated around the mean. Before we prove the lemma, we need a bound on $\mu(B_a(\frac{\pi}{2}-s))$. One can show (for the derivation see, for example, Barvinok [2, p. 58]) that, for any $0 \le s \le \pi/2$, $\mu(B_a(\frac{\pi}{2}-s)) \le \sqrt{\frac{\pi}{8}}e^{-\frac{s^2(n-2)}{2}}$, or since we are interested in large values of $n$ that

$$\mu\left(B_a\left(\frac{\pi}{2}-s\right)\right) \le \left(\frac{1}{2}+o(1)\right)e^{-\frac{s^2 n}{2}}.$$

*Proof.* By Levy's Lemma and the inequality above, we have

$$\mu(\{x : d(A^{\pm}, x) \ge \varepsilon\}) \ge \mu(B_a(\tfrac{\pi}{2}+\varepsilon)) \ge 1 - (\tfrac{1}{2}+o(1))e^{-\frac{\varepsilon^2 n}{2}}.$$

This implies

$$\mu(\{x : d(A^+, x) \le \varepsilon\} \cap \{x : d(A^-, x) \le \varepsilon\}) \ge 1 - (1+o(1))e^{-\frac{\varepsilon^2 n}{2}}.$$

Using the fact that $f$ is 1-Lipschitz, it is easy to see that $\{x : |f(x) - m(x)| > \varepsilon\}$ lies inside the complement of $\{x : d(A^+, x) \le \varepsilon\} \cap \{x : d(A^-, x) \le \varepsilon\}$. Therefore,

$$\mu(\{x : |f(x) - m(f)| > \varepsilon\}) \le (1+o(1))e^{-\frac{\varepsilon^2 n}{2}}.$$

□

## First Proof of Johnson-Lindenstrauss Lemma

Rather than project onto a random $k$-dimensional subspace of $\mathbb{R}^n$, we apply a random rotation of $\mathbb{R}^n$ and then project onto the first $k$ coordinates. Choose $v \in \mathbb{R}^n$ at random where the direction $\frac{v}{\|v\|} \in S_{n-1}$ is distributed with respect to $\mu$, and let $f(v) = \sqrt{\sum_{i=1}^k v_i^2}$. We argue that the value $f(v)$ is close to $\|v\|$ with high probability when $k = \Theta(\frac{1}{\varepsilon^2}\log n)$. Specifically, we show there exists a constant $c > 0$ such that

$$\Pr[c\|v\| \le f(v) \le c(1+\varepsilon)\|v\|] \ge 1 - \frac{1}{n^2}. \qquad (*)$$

Once we prove $(*)$, the Johnson-Lindenstrauss Lemma follows easily. For points $x_1, \ldots, x_n \in \mathbb{R}^n$, we let $v_{ij} = x_i - x_j$ for all $i \ne j$. Then $f(v_{ij})$ equals the distance between $x_i$ and $x_j$ after projecting onto a random $k$-dimensional subspace. Applying a union bound to inequality $(*)$, we get

$$\Pr[\forall i \ne j,\ c\|v_{ij}\| \le f(v_{ij}) \le c(1+\varepsilon)\|v_{ij}\|] \ge 1 - \frac{\binom{n}{2}}{n^2}.$$

Since $1 - \frac{\binom{n}{2}}{n^2} > 0$, there exists a projection $\mathbb{R}^n \to \mathbb{R}^k$ for which the $\ell_2$-metric space on points $x_1, \ldots, x_n$ has distortion $1+\varepsilon$.

To prove the inequality $(*)$, we invoke Lemma 4. We first note that $f$ is 1-Lipschitz. We then note that $m(f)$ is close to $\sqrt{\frac{k}{n}}$ since $E[f(v)^2] = \frac{k}{n}$; one can argue for example that $m(f) = \sqrt{\frac{k}{n}} + O(1/\sqrt{n})$ for all $k$. Lemma 4 now gives us

$$\begin{aligned}\Pr[|f(v) - m(f)| > \varepsilon m(f)] &= \mu(\{x \in S_{n-1} : |f(x) - m(f)| > \varepsilon m(f)\}) = (1+o(1))e^{-(\varepsilon m(f))^2 \frac{n}{2}} \\ &= c_0(1+o(1))e^{-\frac{\varepsilon^2 k}{2}}\end{aligned}$$

for some constant $c_0$. Since $k = \Theta(\frac{1}{\varepsilon^2}\log n)$, we have $\varepsilon^2 k = \Theta(\log n)$. Therefore, $c_0 e^{-\frac{\varepsilon^2 k}{2}} \leq \frac{1}{n^2}$ for suitably chosen constant in the expression for $k$. This proves the inequality $(*)$ where $c = m(f) \approx \sqrt{\frac{k}{n}}$. ☐

## Second Proof of Johnson-Lindenstrauss Lemma

We now give a different proof of the Johnson-Lindenstrauss Lemma due to Indyk and Motwani (1998). The elementary presentation we follow is due to Dasgupta and Gupta (2003).

Let $x_1, \ldots, x_n \in \mathbb{R}^n$. The idea is to project $X = \{x_1, \ldots, x_n\}$ onto $k$ independently generated directions. We define random vectors $r_1, \ldots, r_k \in \mathbb{R}^n$ where $r_{ij} \in N(0,1)$ are independent Gaussian random variables for all $1 \leq i \leq k$ and $1 \leq j \leq n$. Thus, $E[r_{ij}] = 0$ and $E[r_{ij}r_{ik}] = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k. \end{cases}$

We define a projection $\mathbb{R}^n \to \mathbb{R}^k$ by

$$f : x \longmapsto (\langle x, r_i \rangle)_{i=1,\ldots,k}.$$

Our goal is to show that the random embedding $f$ of $X$ into $\ell_2^k$ has distortion $1+\varepsilon$ with positive probability.

**Theorem 5.** *For $k = O(\frac{\log n}{\varepsilon^2})$ and $v \in \mathbb{R}^n$,*

$$\Pr\left[1 - \varepsilon \leq \frac{\|f(v)\|}{\sqrt{k}\|v\|} \leq 1 + \varepsilon\right] \geq 1 - \frac{1}{n^2}.$$

Once we prove Theorem 5, the J-L Lemma follows by the same argument as in the first proof.

*Proof.* We shall assume that $\|v\| = 1$, since the fraction $\frac{\|f(v)\|}{\sqrt{k}\|v\|}$ is invariant under scaling of $v$. For random variables $X_i = \langle v, r_i \rangle = \sum_{j=1}^n v_j r_{ij}$, we have

$$E[X_i] = \sum_{j=1}^n v_j E[r_{ij}] = 0,$$

$$E[X_i^2] = \left(\sum_{j=1}^n v_j^2 E[r_{ij}^2]\right) + \left(\sum_{\substack{j,k \in \{1,\ldots,n\} \\ j \neq k}} v_j v_k E[r_{ij} r_{ik}]\right) = \sum_{j=1}^n v_j^2 = \|v\|^2 = 1.$$

Therefore, $E[\|f(v)\|^2] = \sum_{i=1}^n E[X_i^2] = k$.

We now use Chernoff bounds to prove the inequalities

$$\Pr\left[\frac{\|f(v)\|}{\sqrt{k}\|v\|} \leq 1 + \varepsilon\right] \geq 1 - \frac{1}{2n^2} \qquad \text{and} \qquad \Pr\left[\frac{\|f(v)\|}{\sqrt{k}\|v\|} \geq 1 - \varepsilon\right] \geq 1 - \frac{1}{2n^2},$$

which together imply the theorem. We give the argument for the lefthand inequality only (the argument for the righthand inequality is similar). Since $\|v\| = 1$, this means we must show $\Pr[\|f(v)\|^2 \geq k(1+\varepsilon)^2] \leq \frac{1}{2n^2}$.

Let $Y$ be the random variable $\|f(v)\|^2$ and let $\alpha = k(1+\varepsilon)^2$. For every $s > 0$, we have $\Pr[Y > \alpha] = \Pr[e^{sY} > e^{s\alpha}]$. Recall Markov's inequality: $E[X \geq \beta] \leq \frac{E[x]}{\beta}$ where $X$ is a nonnegative random variable and $\beta > 0$. We apply Markov's inequality to get

$$\Pr[Y > \alpha] = \Pr[e^{sY} > e^{s\alpha}] \leq \frac{E[e^{sY}]}{e^{s\alpha}} = e^{-s\alpha} E[e^{s\sum_{i=1}^{k} X_i^2}] = e^{-s\alpha} \prod_{i=1}^{k} E[e^{sX_i^2}] \qquad (\dagger)$$

where the last equality follows from independence of the random variables $X_1, \ldots, X_k$. Each $X_i$ is Guassian with mean 0 and variance 1, so by elementary calculus

$$E[e^{sX_i^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{st^2} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(s-\frac{1}{2})t^2} dt.$$

We now apply a change of variables, letting $u^2 = (1-2s)t^2$ so that $dt = \frac{u}{t}\frac{1}{1-2s}du = \frac{1}{\sqrt{1-2s}}du$. Thus,

$$E[e^{sX_i^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(s-\frac{1}{2})t^2} dt = \frac{1}{\sqrt{1-2s}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{1-2s}}.$$

Plugging this into (†), we get

$$\Pr[Y > \alpha] = e^{-s\alpha}(1-2s)^{-\frac{k}{2}}.$$

We now choose $s = \frac{1}{2} - \frac{k}{2\alpha}$, so that $1 - 2s = \frac{k}{\alpha}$. This gives

$$e^{-s\alpha}(1-2s)^{-\frac{k}{2}} = e^{-\frac{\alpha}{2}(1-\frac{k}{\alpha})} \left(\frac{k}{\alpha}\right)^{-\frac{k}{2}} = e^{\frac{k-\alpha}{2}} \left(\frac{k}{\alpha}\right)^{-\frac{k}{2}}.$$

Recall that $\alpha = k(1+\varepsilon)^2$. Thus, we have:

$$\begin{aligned} e^{\frac{k-\alpha}{2}} \left(\frac{k}{\alpha}\right)^{-\frac{k}{2}} &= e^{-\varepsilon k - \frac{\varepsilon^2}{2}k} e^{-\frac{k}{2}\ln(\frac{k}{\alpha})} = e^{-\varepsilon k - \frac{\varepsilon^2}{2}k} e^{-\frac{k}{2}\ln(\frac{1}{(1+\varepsilon)^2})} \\ &= e^{-\varepsilon k - \frac{\varepsilon^2}{2}k} e^{k\ln(1+\varepsilon)} = e^{k(-\varepsilon - \frac{1}{2}\varepsilon^2 + \varepsilon - \frac{1}{2}\varepsilon^2 + O(\varepsilon^3))} = e^{-k\varepsilon^2 + kO(\varepsilon^3)}, \end{aligned}$$

using the taylor's expansion $\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$.

Taking $k = \Theta(\frac{\log n}{\varepsilon^2})$, we have

$$\Pr[f(v)^2 \geq k(1+\varepsilon)^2] = e^{-k\varepsilon^2 + O(k\varepsilon^3)} = O(\tfrac{1}{2n^2}).$$

□

## Alon's Theorem (to be continued)

In the next lecture, we will give a proof of Alon's theorem. For now, we give a sketch of the proof. Let $v_1, \ldots, v_{n+1} \in \mathbb{R}^d$ be such that $1 \leq \|v_i - v_j\| \leq 1 + \varepsilon$ for all $i \neq j$. The theorem states that $d = \Omega(\frac{\log n}{\varepsilon^2 \log \frac{1}{\varepsilon}})$.

Clearly, we can assume that $v_{n+1} = (0, \ldots, 0)$ by translating al vectors $v_i$. We then scale vectors $v_i$ to obtain new vectors $v_i'$ such that $\|v_i'\| = 1$. After scaling, we have $\big|\|v_i' - v_j'\| - 1\big| = O(\varepsilon)$. We now look at the symmetric matrix $B = \big(\langle v'_i, v'_j\rangle\big)_{1\le i,j\le n}$, which has the form

$$\begin{pmatrix} 1 & & & & & \\ & 1 & & [\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon] & & \\ & & 1 & & & \\ & & & 1 & & \\ [\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon] & & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

i.e., ones along the diagonal and all other entries between $\frac{1}{2} - \varepsilon$ and $\frac{1}{2} + \varepsilon$. This matrix has rank $d$. Alon's theorem is proved by establishing a lower bound on $d$ in terms of $n$ and $\varepsilon$.

## References


[1] N. Alon, Problems and results in extremal combinatorics, I, Discrete Math. 273 (2003), 31-53.

[2] A. Barvinok, Lecture Notes on Measure Concentration, available from `http://www.math.lsa.umich.edu/ barvinok/total710.pdf`.