	Seminar on Algorithms and Geometry	
	Lecture 5	
	May 21, 2009	
Lecturer: Robert Krauthgamer	Scribe by: Daniel Glasner	Updated: May 22, 2009

1 Johnson-Lindenstrauss Lemma and Concentration of Measure

In the previous lecture we stated and proved the following theorem

Theorem 1 (Johnson-Lindenstrauss) *For every subset $X \subseteq \ell_2$ and every $\epsilon \geq 0$ there is an embedding $f : X \hookrightarrow \ell_2^k$ with distortion $1+\epsilon$ and dimension $k = O(\frac{1}{\epsilon^2}\log n)$.*

In this lecture we will see a sketch of an alternative proof of the theorem with an emphasis on the phenomenon of concentration of measure.

Theorem 2 *Let L be a random subspace of $\mathbb{R}^n$ of dimension k and let $f : \mathbb{R}^n \mapsto \mathbb{L}$ be an orthogonal projection onto L (here we think of L as a copy of $\mathbb{R}^k$). Then there exists a constant $c = c(n,k)$ s.t. for every $x, y \in \mathbb{R}^n$*

$$\Pr_f[1-\epsilon \leq \tfrac{\|f(x)-f(y)\|}{c\|x-y\|} \leq 1+\epsilon] \geq 1 - \tfrac{1}{n^3}$$

Sketch of Proof L is chosen by picking k orthogonal vectors from $\mathbb{S}^{n-1}$ s.t. each vector has a uniform distribution over $\mathbb{S}^{n-1}$. Alternatively choose a random rotation $U : \mathbb{R}^n \mapsto \mathbb{R}^n$ and let $L = U(span\{e_1, \ldots, e_k\})$ where $e_1, \ldots, e_k$ are the first k vectors of the canonical basis of $\mathbb{R}^n$. As in the previous class, it suffices to prove that for all $v \in \mathbb{S}^{n-1}$

$$\Pr_f[1-\epsilon \leq \frac{\|f(v)\|}{c} \leq 1+\epsilon] \geq 1 - \frac{1}{n^3}$$

Denote $w = (w_1, \ldots, w_n) = U^{-1}v$ and note that the projection v onto L has the same length as that of w onto $U^{-1}(L) = span\{e_1, \ldots, e_k\}$ so

$$\|f(v)\| = \|(w_1, \ldots, w_k)\| = \sqrt{w_1^2 + \ldots + w_k^2}$$

Let μ denote the uniform probability measure on $\mathbb{S}^{n-1}$ (the Haar measure), then $w = U^{-1}v$ is distributed according to μ. We now ask what is the length of the projection of w on the first k coordinates.

Theorem 3 *Let $g : \mathbb{S}^{n-1} \mapsto \mathbb{R}$ be 1-Lipschitz i.e.*

$$|g(x) - g(y)| \leq \|x - y\|$$

and let $m = m(g)$ be a median of g i.e.

$$\Pr_{x \in \mu}[g(x) \geq m] \geq \frac{1}{2} \text{ and } \Pr_{x \in \mu}[g(x) \leq m] \geq \frac{1}{2}$$

then for all $\delta > 0$

$$\Pr_{x \in \mu}[|g(x) - m| \geq \delta] \leq 4e^{-\delta^2 n/2}$$

We can use theorem 3 to prove theorem 2. It can be easily verified that $g(w) = \|(w_1, \ldots, w_k)\|$ is 1-Lipschitz. Applying theorem 3 gives us the following bound

$$\Pr_w[m - \delta \leq g(w) \leq m + \delta] \geq 1 - 4e^{-\delta^2 n/2}$$

We now choose $c = c(n, k) = m$ and $\delta = \epsilon m$ to get

$$\Pr_f[1 - \epsilon \leq \frac{\|f(v)\|}{c} \leq 1 + \epsilon] =$$
$$\Pr_w[m(1 - \epsilon) \leq g(w) \leq m(1 + \epsilon)] \geq 1 - 4e^{-\epsilon^2 m^2 n/2}$$

All that is left is to lower bound m. We observe that

$$\mathbb{E}_{w \in \mu}[g(w)^2] = \mathbb{E}[w_1^2 + \ldots + w_k^2] = k\mathbb{E}[w_1^2] = \frac{k}{n}\mathbb{E}[w_1^2 + \ldots + w_k^2] = \frac{k}{n}$$

where we have used the symmetry of the coordinates, all of the coordinates are identically distributed. We can use this to lower bound the median. Consider a parameter $t > 0$ we partition the integration into two parts $[0, (m + t)]$ and $[(m + t), 1]$ and upper bound each part

$$\begin{aligned}
\frac{k}{n} &= \mathbb{E}[g(w)^2] \\
&\leq \Pr[g(w)^2 \geq (m + t)^2] \cdot 1 + \Pr[g(w)^2 \leq (m + t)^2] \cdot (m + t)^2 \\
&\leq 4e^{-t^2 n/2} + (m + t)^2
\end{aligned}$$

By choosing $t = \sqrt{\frac{k}{5n}}$ we get $m \geq \Omega(\sqrt{\frac{k}{n}})$. Finally

$$\Pr_f[1 - \epsilon \leq \frac{\|f(v)\|}{c} \leq 1 + \epsilon] \geq 1 - e^{-\epsilon^2 k c'} \geq 1 - \frac{1}{n^3}$$

if $k \geq 100 \frac{1}{\epsilon^2} \log n$. ■

How can we prove theorem 3? The following isoperimetric inequalities provide an answer.

Theorem 4 (Paul Levy 1951) *Let $A \subseteq \mathbb{S}^{n-1}$ be a measurable set and let $B \subseteq \mathbb{S}^{n-1}$ be a cap with $\mu(A) = \mu(B)$. Then for all $\epsilon > 0$, $\mu(A_\epsilon) \geq \mu(B_\epsilon)$. Here*

$$A_\epsilon = \{x \in \mathbb{S}^{n-1} : \ d(x, A) \leq \epsilon\}$$

and similarly for B_ϵ.

Remark: Here distance is Euclidean, i.e. measured according to ℓ_2-norm, but similar theorems can be proved for goedesic distance on the sphere.

Using Theorem 4 plus estimates on the volume of a spherical cap, one can obtain the following bound on the measure of A_ϵ. Such a bound can also be proved directly via the Brunn-Minkowski Theorem.

Theorem 5 *Let $A \subseteq \mathbb{S}^{n-1}$ be a measurable set with $\mu(A) \geq \frac{1}{2}$ then for all $\epsilon > 0$*

$$\mu(A_\epsilon) \geq 1 - 2e^{-\epsilon^2 n/2}$$

Consider cutting the sphere by a hyperplane passing through the origin. The sets on both sides have a measure of exactly $\frac{1}{2}$. Applying the inequality to each one of these sets we see that almost all of the measure is concentrated on a thin strip around the equator. The total amount of measure outside is an exponentially small function of the dimension.

We now sketch the proof of theorem 3 using theorem 5.

Sketch of Proof Apply theorem 5 to $A^- = \{x \in \mathbb{S}^{n-1} : \ g(x) \leq m\}$ and obtain a lower bound on $\mu(A^-_\epsilon)$. Then do similarly for $A^+ = \{x \in \mathbb{S}^{n-1} : \ g(x) \geq m\}$. ■