Seminar on Algorithms and Geometry

Lecture 4

May 7, 2009

Lecturer: Robert Krauthgamer Scribe by: Mica Arie-Nachimson Updated: May 17, 2009

1 Dimension reduction in l_2

Theorem 1 (Johnson-Lindenstrauss, 1984) *For every n-point subset $X \subseteq l_2$ and every $0 < \epsilon < 1$, there is an embedding $f : X \to l_2^k$ with distortion $1+\epsilon$ and dimension $k = O(\frac{1}{\epsilon^2}\log n)$.*

Remark

- Proof gives randomized algorithm.
- Many algorithmic applications.
- Naive approach: isometric embedding with dimension $n-1$. The J-L theorem is gives exponential improvement in the dimension, which is logarithmic in n, at the expense of arbitrarily small distortion.
- Often, it is important that f is random (and/or a linear transformation), not depending on X.

Proof Idea Choose a linear transformation $f : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ at random, independently of X. We will optimize k at the end. Different ways to choose f:

1. Choose at random a linear subspace L of dimension k, and project every $x \in X$ onto L.

   Definition 2 (Projection) *Projection of $x \in X$ onto subspace L is the (unique) point in L closest to x. An equivalent definition is the (unique) point where a line going through x and orthogonal to L intersects L.*

   Observation 3 (Random subspace) *Let $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ be the unit sphere. Choose unit-length orthogonal vectors $b_1, ..., b_k \in \mathbb{R}^n$, where each b_j has "uniform" distribution on the unit sphere. Then:*

   $$L = \text{span}\{b_1, ..., b_k\} \quad \Rightarrow \quad f(x) = (\langle b_1, x\rangle, ..., \langle b_k, x\rangle) \in \mathbb{R}^k.$$

2. Choose $b_1, ..., b_k \in S^{n-1}$ independently at random and again let $f(x) = (\langle b_1, x\rangle, ..., \langle b_k, x\rangle)$.

3. Choose each coordinate of each b_i independently at random:
   - To be a standard Gaussian $N(0,1)$.
   - Uniformly from $\{\pm 1\}$ or from $\{1,-1,0\}$.

   This (multidimensional) Gaussian distribution has a very useful property: it is rotation invariant, i.e. applying any fixed rotation in $\mathbb{R}^n$ to this distribution produces a the same Gaussian distribution. The "uniform" distribution on the sphere which has the same property, is formally called the Haar measure, and is essentially unique. (Proving the rotation invariance is beyond the scope of the class.) It follows, for instance, that sampling a vector in $\mathbb{R}^n$ from the Gaussian distribution and normalizing it to be unit length, produces a uniformly random vector on the sphere. Another immediate consequence is the following useful fact, which says that "measuring" the Gaussian vector along any fixed direction produces a Guassian distribution.

   Fact 4 *Let $G_1, ...G_n$ be independent random variables with Gaussian distribution $N(0,1)$, and let $v \in \mathbb{R}^n$. Then $\tilde{G} = \sum_{i=1}^n v_i G_i = \langle (G_1, ..., G_n), v\rangle$ has Gaussian distribution $N(0, \sigma^2 = \sum_{i=1}^n v_i^2)$.*

■

We are going to prove variant 3 shown in the proof idea.

Theorem 5 *Let $B_{n\times k}$ be a matrix whose entries are i.i.d. $N(0,1)$, and let $f : \mathbb{R}^n \to \mathbb{R}^k$ be given by $f(x) = \frac{Bx}{\sqrt{k}}$. Then with high probability (say $\geq \frac{1}{2}$ or $\geq 1 - \frac{1}{n}$), for all $x \neq y \in X$ we have $1 - \epsilon \leq \frac{\|f(x)-f(y)\|}{\|x-y\|} \leq 1 + \epsilon$.*

Proof It suffices to prove that for all $v \in \mathbb{R}^n$,

$$\Pr_f[1 - \epsilon \leq \tfrac{\|f(v)\|}{\|v\|} \leq 1 + \epsilon] > 1 - \tfrac{1}{n^3}. \tag{1}$$

This is enough because applying (1) to $v = x - y$ for $x, y \in \mathbb{R}^n$, we get $f(x,y) = f(x) - f(y)$, and by a union bound over $\binom{n}{2}$ pairs $x \neq y \in X$ the theorem follows.

In fact, it suffices to prove (1) for unit length v, since f is a linear transformation and $f\left(\frac{v}{\|v\|}\right) = \frac{f(v)}{\|v\|}$ for all v.

Assume $\|v\| = 1$. Each coordinate of $Bv = (\langle b_1, v\rangle, ..., \langle b_k, v\rangle)$ has distribution $N(0, \sigma^2 = \|v\|^2 = 1)$ by Fact 4, and they are clearly independent. Denoting $g_i = \langle b_i, v\rangle$, we have:

$$\mathbb{E}[\|f(v)\|^2] = \mathbb{E}\left[\frac{\|Bv\|^2}{k}\right] = \frac{1}{k}\mathbb{E}[g_1^2 + ... + g_k^2] = 1, \tag{2}$$

so $\mathbb{E}[\|Bv\|^2] = k$. We will bound $\Pr[\|f(v)\|^2 \geq (1+\epsilon)^2] = \Pr[\|Bv\|^2 \geq k(1+\epsilon)^2]$. A similar argument works for the event $\|Bv\|^2 \leq k(1+\epsilon)^2$.

Define the shorthand $\alpha = k(1+\epsilon)^2$, and let $S > 0$ be chosen later. Then by Markov's inequality and the independence of g_j,

$$\Pr[\|Bv\|^2 \geq \alpha] = Pr[S \cdot \|Bv\|^2 \geq S\alpha] = Pr[e^{S\|Bv\|^2} \geq e^{S\alpha}] \leq \frac{\mathbb{E}[e^{S\|Bv\|^2}]}{e^{S\alpha}}$$
$$= e^{-S\alpha} \cdot \mathbb{E}[e^{S\sum_{j=1}^{k} g_j^2}] = e^{-S\alpha} \cdot \mathbb{E}[\prod_j e^{Sg_j^2}] = e^{-S\alpha} \cdot \prod_j \mathbb{E}[e^{Sg_j^2}]$$

A direct computation of the integrals shows that

$$\mathbb{E}[e^{Sg_j^2}] = \frac{1}{\sqrt{1-2S}}.$$

Plugging this in, with the choice of S such that $1 - 2S = \frac{k}{\alpha} = (1+\epsilon)^{-2}$, and using (by Taylor expansion) $\ln(1+x) \leq x - \frac{x^2}{2} + O(x^3)$, we have:

$$\begin{aligned}
\Pr[\|Bv\|^2 \geq \alpha] &\leq e^{-S\alpha}(1-2S)^{-\frac{k}{2}} = e^{-\alpha(\frac{1}{2}-\frac{k}{2\alpha})}(\tfrac{k}{\alpha})^{-\frac{k}{2}} \\
&= e^{-\frac{\alpha}{2}+\frac{k}{2}}(1+\epsilon)^{\frac{2k}{2}} = e^{-\frac{\alpha}{2}+\frac{k}{2}+\frac{2k}{2}\ln(1+\epsilon)} \\
&\leq e^{-\frac{k}{2}(2\epsilon+\epsilon^2)+\frac{2k}{2}(\epsilon-\frac{\epsilon^2}{2}+O(\epsilon^3))} \leq e^{-\epsilon^2 k + O(k\epsilon^3)} \\
&\leq e^{-\frac{1}{2}\epsilon^2 k} \leq e^{-3\ln n} = \frac{1}{n^3}.
\end{aligned}$$

We conclude that $\Pr[\|f(v)\| > 1+\epsilon] = \Pr\left[\frac{\|Bv\|}{k} \geq 1+\epsilon\right] = \Pr[\|Bv\|^2 \geq \alpha] \leq \frac{1}{n^3}$. ■