Lecture 8

A lower-bound on ℓ_2 dimensionality reduction

The main focus of this lecture is a lower bound on the dimension when doing dimensionality reduction with ϵ-distortion in ℓ_1 and ℓ_2. In particular, it will be shown that there exist graphs requiring $\Omega(\log n)$ dimensions to embed with any fixed desired distortion in Euclidean space. The main technical result is:

Theorem 1 (Alon [1]). *Let $v_1, \ldots, v_{n+1} \in \mathbb{R}^d$ and $1/\sqrt{n} \leq \epsilon < 1/3$ be given, such that $1 \leq \|v_i - v_j\| \leq 1 + \epsilon$ for all $i \neq j \in [n+1]$. Then the subspace spanned by $v_1, \ldots, v_{n+1}$ has dimension $d = \Omega\left(\frac{\log n}{\epsilon^2 \log 1/\epsilon}\right)$.*

In particular, this theorem suggests that the n-dimensional simplex Δ_n (on $n+1$ vertices) cannot be embedded with distortion $1 + \epsilon$ in fewer than $\Omega(\log n)$ dimensions.

We next proceed to the proof of the theorem. By translation we can assume that $v_{n+1} = 0$, and therefore $1 \leq \|v_i\| \leq 1 + \epsilon$ by assumption. Further set $v_i' = v_i / \|v_i\|$, and note that:

$$\left| \langle v_i', v_j' \rangle - \frac{1}{2} \right| = O(\epsilon)$$

since $\langle v_i', v_j' \rangle = \cos \angle(v_i, v_j) \leq 1/2 + \epsilon + \epsilon^2/2$ using the cosine rule ($c^2 = a^2 + b^2 - 2ab\cos\gamma$). Let us then define matrix $B \in M_{n \times n}(\mathbb{R})$ with $B = \Big[\langle v_i', v_j' \rangle\Big]_{1 \leq i,j \leq n}$, which would look like:

$$\begin{pmatrix} 1 & & & \frac{1}{2} + O(\epsilon) \\ & 1 & & \\ & & \ddots & \\ \frac{1}{2} + O(\epsilon) & & & 1 \end{pmatrix}$$

As a side note, if $\epsilon = 0$ then B would be 1 along the diagonal and 1/2 everywhere else, hence it would have full rank n. Being the Gram matrix of the v_i''s it would follow that they span a subspace of dimension n (this follows from Theorem 7.2.10 in [3]) which is also the subspace spanned by the v_i's. In other words, if no distortion were allowed, the lower bound on the dimension is n.

Continuing, let $d = \mathrm{rank}(B)$ and define $C = 2B - J$, where $J = ee^T$ is the all-ones matrix. Taking the rank operator in the definition of C, we get $|\mathrm{rank}(C) - \mathrm{rank}(B)| \leq 1$ or equivalently:

$$\mathrm{rank}(C) \leq d + 1.$$

We now prove a lower bound on the rank of a matrix for the case when the off-diagonal entries are very small; we will later apply this to C:

Lemma 2. *Consider a symmetric matrix $C \in M_{n \times n}(\mathbb{R})$ such that $C_{i,i} = 1$ for $i \in [n]$, and $C_{i,j} \leq 1/\sqrt{n}$ for $i \neq j \in [n]$. Then* $\mathrm{rank}(C) \geq n/2$.

Proof of Lemma: Since C is symmetric, all of its eigenvalues must be real (See Theorem 4.1.3 in [3]). Now, suppose $d = \text{rank}(C)$ and let $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$ be the non-zero eigenvalues of C. The invariance of the trace with respect to conjugation gives:

$$\text{Tr}(C) = \sum_{i\in[n]} C_{i,i} = n = \sum_{i\in[d]} \lambda_d. \tag{1}$$

Now observe that the non-zero eigenvalues of $C^2 = C^T C$ are exactly $\lambda_1^2, \ldots, \lambda_d^2$. Respectively:

$$\text{Tr}(C^2) = \sum_{i\in[d]} \lambda_i^2, \quad \text{but also} \tag{2}$$

$$\text{Tr}(C^2) = \sum_{i\in[n]} \sum_{j\in[n]} C_{i,j}^2 \leq n + n(n-1)\frac{1}{n} = 2n - 1 < 2n. \tag{3}$$

Now let us pretend that the λ_i's are the equally likely outcomes of a random variable Λ. Then (1), (2) and (3) tell us that:

$$\text{E}[\Lambda] = \frac{n}{d} \quad \text{and} \quad \text{E}[\Lambda^2] < \frac{2n}{d}$$

Applying that $\text{Var}[\Lambda] = \text{E}[\Lambda^2] - \text{E}[\Lambda]^2 \geq 0$, we get $d > 2n$ (this is simply the quadratic/arithmetic mean inequality). □

Lemma 3. *Suppose $A \in M_{n,n}(\mathbb{R})$ has rank d, then $F = \left(A_{i,j}^k\right)_{1\leq i,j\leq n}$ has rank at most $\binom{d+k-1}{d-1}$.*

Proof of Lemma: Let $v_1, \ldots, v_d \in \mathbb{R}^n$ be a basis for the row space of A. Then for the i-th row of A, denoted A_i, we have that:

$$A_i = \sum_{l\in[d]} \lambda_l v_l$$

for some coefficients λ_l. And moreover:

$$A_{i,j} = \sum_{l\in[d]} \lambda_l v_{l,j}$$

where $v_{l,j}$ denotes the j-th entry of v_l. For an entry of F we have:

$$\begin{aligned} F_{i,j} = A_{i,j}^k &= \left(\sum_{l\in[d]} \lambda_l v_{i,j}\right)^k \\ &= \sum_{k_1+\cdots+k_d=k} \binom{k}{k_1, \ldots, k_d} \left[\prod_{l\in[d]} \lambda_l^{k_l}\right] \left[\prod_{l\in[d]} v_{l,j}^{k_l}\right] \end{aligned}$$

From this formula, the row space of F is spanned by:

$$\left(w_{k_1,\ldots,k_d}\right)_j = \prod_{l\in[d]} v_{l,j}^{k_l}$$

i.e. there is one vector, corresponding to each partition $k_1 + \cdots + k_d = k$. The number of such partitions $\binom{k+d-1}{d-1}$ gives an upper bound on the (row) rank of F. □

Proof of Theorem: Back to the main theorem. Recall $\text{rank}(C) \leq d+1$ and $|C_{i,j}| \leq \epsilon$. Let k be an integer such that $\epsilon^k < 1/\sqrt{n}$. Consider the matrix $F = \left(C_{i,j}^k\right)_{1 \leq i,j \leq n}$, which has it that $|F_{i,j}| \leq 1/\sqrt{n}$ by construction. According to the second lemma $\text{rank}(F) \leq \binom{k+d}{d}$. According to the first lemma $\text{rank}(F) \geq n/2$. From the inequalities for $\text{rank}(F)$ we now have:

$$\frac{n}{2} \leq \binom{k+d}{d} = \binom{k+d}{k} = \frac{(k+d)!}{d!}\frac{1}{k!} \leq (k+d)^k \left(\frac{e}{k}\right)^k$$

Taking the natural logarithm on both sides and observing that $k = \frac{\ln n}{2 \ln 1/\epsilon}$ yields the result $d = \frac{\Omega(\ln n)}{\epsilon^2 \ln 1/\epsilon}$. □

A lower-bound on ℓ_1 dimensionality reduction

The result we are about to show was first discovered by Brikman and Charikar [2], but a much simpler proof is presented here due to Lee and Naor [4].

Let D_m denote the m-th level diamond graph (defined in earlier lectures). In the problem set, we prove that the shortest path metric of D_m embeds into ℓ_1 with constant distortion.

Theorem 4. *For any $1 < p \leq 2$, the distortion required to embed D_m in ℓ_p is at least $\sqrt{1+(p-1)m}$.*

Lemma 5. *Let $x \in \mathbb{R}^d$, then $d^{1/p-1}\|x\|_1 \leq \|x\|_p \leq \|x\|_1$.*

Proof of Lemma: The right inequality is trivial, so we focus on the left one. Let $x^T = (x_1, \ldots, x_d)$. Then $d^{1/p-1}\|x\|_1 \leq \|x\|_p$ is equivalent to:

$$\frac{\sum_{i \in [d]} |x_i|^p}{d} \geq \left(\frac{\sum_{i \in [d]} |x_i|}{d}\right)^p$$

This can be written as $\mathrm{E}[X^p] \geq \mathrm{E}[X]^p$, where X is a random variable uniform over $|x_1|, \ldots, |x_d|$. The latter inequality always holds due to the convexity of $f_p(t) = t^p$ for $p \geq 1$ and $t \geq 0$. □

We are going to combine the above theorem and lemma to get a lower bound on dimensionality reduction in ℓ_1. In particular, consider the following sequence of operations:

1. Embed the shortest-path metric of D_m into ℓ_1 with distortion $O(1)$ (see problem set).
2. Reduce the dimension of the resulting embedding down to d dimensions, by admitting some distortion D.
3. Apply the lemma above to argue that the same (reduced dimension) embedding viewed as an ℓ_p embedding distorts distances by at most $d^{1-1/p}$.

In summary, we have embedded D_m into ℓ_p with distortion $O(1) \cdot D \cdot d^{1-1/p}$. On the other hand, the theorem tells us that we cannot embed D_m into ℓ_p with less than $\sqrt{1+(p-1)m}$ distortion, hence we can set up the inequality:

$$O(1) \cdot D \cdot d^{1-1/p} \geq \sqrt{1+(p-1)m}.$$

Choosing $p = 1 + \frac{1}{\log d}$ yields that:

$$D^2 = \Omega\left(\frac{\log n}{\log d}\right), \quad \text{or equivalently} \quad d = n^{\Omega(1/D^2)}$$

Sketch of Proof of Theorem: Recall that for any points x, y, z, w in ℓ_2 we have the isoperimetric inequality:

$$\|x - z\|_2^2 + \|y - w\|_2^2 \leq \|x - y\|_2^2 + \|y - z\|_2^2 + \|z - w\|_2^2 + \|w - x\|_2^2.$$

Similarly in ℓ_p we have that:

$$\|x - z\|_p^2 + (p-1)\|y - w\|_p^2 \leq \|x - y\|_p^2 + \|y - z\|_p^2 + \|z - w\|_p^2 + \|w - x\|_p^2.$$

Let us now resume the notation that E_m is the set of edges of D_m and F_m is the set of anti-edges. And let f be an embedding of D_m into ℓ_p, then:

$$A = \|f(s) - f(t)\|_p^2 + (p-1) \sum_{(u,v)\in F_m \setminus \{(s,t)\}} \|f(u) - f(v)\|_p^2 \leq \sum_{(u,v)\in E_m} \|f(u) - f(v)\|_p^2 = B.$$

If f has distortion less than D, then:

$$d_{D_m}(u,v) \leq \|f(u) - f(v)\|_p \leq D \cdot d_{D_m}(u,v).$$

Then we would get that:

$$4^m\big((p-1)m + 1\big) \leq A \leq B \leq |E_m| \cdot D^2 = 4^m \cdot D^2,$$

which yields the result $D \geq \sqrt{1 + (p-1)m}$. □

References

[1] N. Alon. Problems and results in extremal combinatorics, i. *Discrete Mathematics*, 273:31–53, 2003.

[2] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[3] Robert Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

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