CS 369M: Metric Embeddings and Algorithmic Applications September 27, 2018

Lecture 2: Embeddings into ℓ_∞

Lecturer: Moses Charikar *Scribe: Margalit Glasgow*

1 Introduction

This lecture covers some basic embedding results for embedding metric spaces into ℓ_∞. These ideas are used in other embedding results, e.g., Bourgain's theorem which provides embeddings into ℓ_2.

2 Embeddings into ℓ_∞

We begin by showing that any finite metric embeds isometrically into ℓ_∞.

Theorem 2.1. *Every n-point metric space (V, d) embeds isometrically into ℓ_∞.*

Proof. Let $V = \{v_1, \dots v_n\}$. We will map each point $v \in V$ to a vector of length n, where the ith coordinate is the distance between v and v_i. Formally, define $f_i(v_j) = d(v_i, v_j)$, and let $f(v_j) = \begin{pmatrix} f_1(v_j) \\ \cdots \\ f_n(v_j) \end{pmatrix}$.

Claim 2.2. *For any v_i, v_j, we have $\|f(v_i) - f(v_j)\|_\infty \geq d(v_i, v_j)$.*

Proof.

$$\|f(v_i) - f(v_j)\|_\infty = \max_k |f_k(v_i) - f_k(v_j)| \geq |f_i(v_i) - f_i(v_j)| = d(v_i, v_j).$$

□

Claim 2.3. *For any v_i, v_j, we have $\|f(v_i) - f(v_j)\|_\infty \leq d(v_i, v_j)$.*

Proof. By the triangle inequality, for any k,

$$|f_k(v_i) - f_k(v_j)| = |d(v_i, v_k) - d(v_k, v_j)| \leq d(v_i, v_j).$$

□

The theorem follows from the two claims. □

The embedding in the proof used n coordinates. Do we need to all n coordinates to construct an isometric embedding?

The answer is no: $n-1$ coordinates suffice. If we remove any one coordinate, it is easy to check that the first claim still holds (plugging in j instead of i if necessary). Eliminating a coordinate cannot increase the ℓ_∞ norm, so the second claim also holds.

However, we cannot use less than $n-1$ coordinates without allowing some distortion.

If we do allow for some distortion, we can drastically reduce the number of coordinate. The following theorem yields an upper bound for the number of coordinates necessary to achieve distortion D.

Theorem 2.4 (Upper Bound)**.** *Let $D = 2q - 1 \geq 3$ be an odd integer. Given an n-point metric space (V, d), there exists an embedding with distortion D of (V, d) into ℓ_∞^k with $k = O(qn^{1/q} \ln n)$.*

For example, for distortion $D = 3$, we get an embedding into $k = O(\sqrt{n} \ln n)$ dimensions. For a smaller value of D, we cannot find an embedding with a number of dimensions that is asymptotically better than needed for the isometric embedding described above.

Theorem 2.5 (Lower Bound)**.** *There is an n-point metric space (V, d) such that every embedding of (V, d) into ℓ_∞ with distortion $D < 3$ requires $\Omega(n)$ dimensions.*

For $D = 3$, there is a lower bound of $\Omega(\sqrt{n})$ dimensions, and there are similar matching lower bounds for $D = 5, 7$, and possibly more (see Corollary 15.3.4 in [2]).

We will prove the upper bound in Theorem 2.4 using a mapping called a *Fréchet embedding.*

Definition 2.6. A *Fréchet embedding* is a map $f : (V, d) \to \ell_p^k$ which consists of k functions f_i (one per coordinate), each of which has the form $f_i(v) = d(v, A_i)$, where $A_i \subset V$. The distance $d(v, A)$ is the minimum distance between a point $a \in A$ and the point v.

More generally, we will sometimes allow scaling by constants α_i in the definition of Fréchet embeddings, that is, each coordinate will be given by $f_i(v) = \alpha_i d(v, A_i)$.

Fact 2.7. *Fréchet embeddings (without scaling) are non-expanding for a single coordinate, that is, for every i,*

$$|f_i(u) - f_i(v)| \leq d(u, v).$$

This property of Fréchet embeddings follows from the triangle inequality. Since for every u, v, $d(u, A) \leq d(u, v) + d(v, A)$ (convince yourself why), we get that $|f_i(u) - f_i(v)| = |d(u, A_i) - d(v, A_i)| \leq d(u, v)$.

Note that the isometric embedding into ℓ_∞ that we saw earlier today is a Fréchet embedding. We are now ready to prove Theorem 2.4.

Proof of Theorem 2.4. Since we will use a Fréchet embedding and the target metric is ℓ_∞, the embedding will be non-expanding ($\|f(u) - f(v)\|_\infty = \max_i |f_i(u) - f_i(v)| \leq d(u, v)$). Hence in order to prove that the embedding has distortion D, it suffices to show that for some coordinate i, we have $|f_i(u) - f_i(v)| \geq \frac{1}{D} d(u, v)$.

Recall that for a point $v \in V$ and radius $r \in \mathbb{R}^{\geq 0}$, the (closed) ball $B(v, r)$ around v of radius r contains all the points in the metric space that are within distance r from v, that is, $B(v, r) = \{u \in V \mid d(u, v) \leq r\}$. The open ball is defined similarly as $B^o(v, r) = \{u \in V \mid d(u, v) < r\}$.

The intuition behind the proof is as follows. To get a difference $|f_i(u) - f_i(v)| \geq \Delta$ for coordinate i, we need to have some r such that $B^o(u, r + \Delta) \cap A_i = \emptyset$, but $B(v, r) \cap A_i \neq \emptyset$ or vice versa (replace the roles of u and v). This would imply that $d(u, A_i) \geq r + \Delta$ while $d(v, A_i) \leq r$ (or the same with the opposite roles). If we can find r such that the number of points in $B^o(u, r + \Delta)$ is not too large compared to $B(v, r)$, we can pick A_i at random and hope for it to intersect $B(v, r)$ but not $B^o(u, r + \Delta)$ (see Figure 1).

The embedding of (V, d) into ℓ_∞^k will be a Fréchet embedding constructed in the following way:

- Let $p = n^{-1/q}$.
- For $j = 1, \ldots, q$, let $p_j = \min\{1/2, p^j\}$.

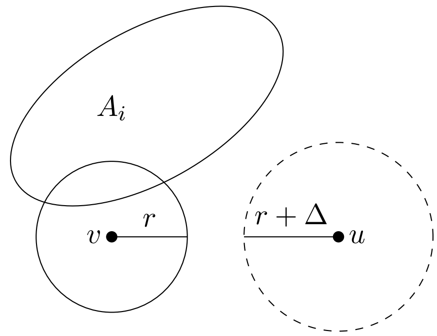

Figure 1: The random set A_i intersects $B(v, r)$ but not $B^o(u, r + \Delta)$.

- Let $m = \lceil 24 n^{1/q} \ln n \rceil$.
- Let $k = mq$ be the number of coordinates in our embedding.
- For each $i \in [m]$, $j \in [q]$, let A_{ij} be a random sample of points in V with density p_j, i.e., each point $v \in V$ is chosen to be in A_{ij} independently with probability p_j. Our mapping is the Fréchet embedding with these k sets.

We proceed to show that this randomized construction returns an embedding with distortion D with probability at least $1/2$. We say that a subset A_{ij} is *good* for (u, v) if $|d(u, A_{ij}) - d(v, A_{ij})| \geq \frac{1}{D} d(u, v)$.

Lemma 2.8. *For any pair $u, v \in V$, there exists some j such that if A_{ij} is a random sample of points from V where each point is sampled with probability p_j, then*

$$\Pr[A_{ij} \text{ is good for } (u, v)] \geq p/12.$$

Before we prove the lemma, let us first show how it can be used to complete the proof of the theorem. For any pair $u, v \in V$,

$$\Pr[\text{no } A_{ij} \text{ is good for } (u, v)] \leq \left(1 - \frac{p}{12}\right)^m \leq e^{-\frac{pm}{12}} \leq \frac{1}{n^2}.$$

By the union bound, since there are $\binom{n}{2}$ pairs (u, v), the probability that there exists a pair with no good A_{ij} is less than $\frac{1}{n^2}\binom{n}{2} \leq 1/2 < 1$. Hence, there exists some choice of sets A_{ij} for which the mapping has distortion D. The probability of success (that is, choosing an embedding with distortion D) can be amplified by choosing larger m.

Finally, we prove Lemma 2.8. Recall that we want to find a coordinate with distance at least

$d(u,v)/D$. Define $\Delta = d(u,v)/D$. Consider the following sequence of balls:

$$B_0 = B(u,0)$$
$$B_1 = B(v,\Delta)$$
$$B_2 = B(u,2\Delta)$$
$$\cdots$$
$$B_q = \begin{cases} B(u,q\Delta) & q \text{ is even} \\ B(v,q\Delta) & q \text{ is odd} \end{cases}$$

Our goal is to find two consecutive balls B_t and B_{t+1} such that the number of points in B_{t+1} is not much greater than the number in B_t. For each t, define $n_t = |B_t|$, the number of points in B_t. Recall that $D = 2q-1$, and note that for any $t \le q-1$,

$$B_t \cap B^o_{t+1} = \emptyset,$$

because $(2t+1)\Delta \le (2q-1)\Delta = d(u,v)$. (Here B^o_{t+1} denotes the corresponding open ball.)

Claim 2.9. *There exist some $t < q$ and $j \le q$ such that $n_t \ge n^{\frac{j-1}{q}}$ and $n_{t+1} \le n^{\frac{j}{q}}$.*

Proof. If there is some $t < q$ such that $n_t \ge n_{t+1}$, the claim holds for that t and the maximum j such that $n_t \ge n^{\frac{j-1}{q}}$. Otherwise, $n_0 < n_1 < \cdots < n_t$. Consider the q intervals $\left[n^{\frac{j-1}{q}}, n^{\frac{j}{q}}\right]$ for $j = 1,\ldots,q$. Since there are $q+1$ *increasing* values $n_0,\ldots,n_q$, by the pigeonhole principle, there must be some t and j such that n_t and n_{t+1} belong to the same interval $\left[n^{\frac{j-1}{q}}, n^{\frac{j}{q}}\right]$. The claim holds for those t and j. □

We are now ready to show that for the values t and j that were found in Claim 2.9, with probability at least $p/12$, the set A_{ij} intersects B_t but not B^o_{t+1}. As explained earlier, that would imply that A_{ij} is good for (u,v) (note that the difference in radius between B_t and B^o_{t+1} is Δ). Let E_1 be the event that $A_{ij} \cap B_t \ne \emptyset$. Let E_2 be the event that $A_{ij} \cap B^o_{t+1} = \emptyset$. Because each point in A_{ij} is chosen independently at random with probability p_j, we have

$$\Pr[E_1] = 1 - \Pr[A_{ij} \cap B_t = \emptyset] = 1 - (1-p_j)^{n_t} \ge 1 - e^{-p_j n_t}.$$

Plugging in $p_j = \min\{1/2, n^{-\frac{j}{q}}\}$ and $n_t \ge n^{\frac{j-1}{q}}$, we have

$$\Pr[E_1] \ge 1 - e^{-n_t \min\{1/2, n^{-\frac{j}{q}}\}} \ge \min\{1 - e^{-\frac{1}{2}n^{\frac{j-1}{q}}}, 1 - e^{-p}\} \ge \min\{1 - e^{-\frac{1}{2}}, 1 - e^{-p}\},$$

where the last inequality follows because $n^{\frac{j-1}{q}} \ge 1$. If $p \ge 1/2$, then this minimum is $1 - e^{-\frac{1}{2}} > 1/3 \ge p/3$. If $p < 1/2$, then using the first three terms of the Taylor expansion for e^x, we get $1 - e^{-p} \ge p - p^2/2 > p/3$. We conclude that in all cases,

$$\Pr[E_1] \ge p/3. \tag{1}$$

Now,

$$\Pr[E_2] \ge (1-p_j)^{n_{t+1}} \ge (1-p_j)^{n^{\frac{j}{q}}}.$$

We again split this into two cases based on the value of p_j. If $p_j > 1/2$, then $n^{\frac{j}{q}} = 1/p_j$, and hence

$$(1-p_j)^{n^{\frac{j}{q}}} = (1-p_j)^{1/p_j} \geq 1/4,$$

where the last inequality holds by checking that $(1-p_j)^{\frac{1}{p_j}}$ is decreasing in p_j and is hence minimized when $p_j = 1/2$.

If $p_j = 1/2$, then we must have $n^{\frac{j}{q}} \leq 2$, and

$$(1-p_j)^{n^{\frac{j}{q}}} \geq \left(\frac{1}{2}\right)^2 = 1/4.$$

In all cases, we have

$$\Pr[E_2] \geq 1/4. \tag{2}$$

It follows from Equation (1) and Equation (2) that

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2] \geq p/12,$$

where we used the fact that B_t and B_{t+1} are disjoint, and as a result, the events E_1 and E_2 are independent. This completes the proof of Lemma 2.8. □

3 Embeddings into ℓ_2

In the previous lecture, we mentioned Bourgain's theorem.

Theorem 3.1 (Bourgain [1])**.** *Any n-point metric embeds into ℓ_2 with distortion $O(\log n)$.*

We claim that the proof of Theorem 2.4 already provides an embedding of any n-point metric into ℓ_2 with distortion $O(\log^2 n)$.

Proposition 3.2. *Any n-point metric embeds into ℓ_2 with distortion $O(\log^2 n)$.*

Proof. Let $q = \log n$, and consider the mapping provided by Theorem 2.4. This mapping is an embedding into ℓ_∞ with dimension $k = O(\log n \cdot n^{\frac{1}{\log n}} \log n) = O(\log^2 n)$ and distortion $O(\log n)$. We use this mapping (which we denote by f), and then embed these vectors into ℓ_2^k using the identity map, that is, we consider the same points but with ℓ_2 distance instead of ℓ_∞.

Consider two points u, v in the original metric space. We have shown in the proof of Theorem 2.4 that

$$\frac{1}{2q-1} d(u,v) \leq \|f(u) - f(v)\|_\infty \leq d(u,v).$$

Now, for every $x \in \mathbb{R}^k$, $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{k}\|x\|_\infty$ (convince yourself why). Together, we get that the embedding into ℓ_2 satisfies

$$\frac{1}{2q-1} d(u,v) \leq \|f(u) - f(v)\|_2 \leq \sqrt{k} d(u,v),$$

and since $k = O(\log^2 n)$ and $2q - 1 = O(\log n)$, we conclude that f is an embedding into ℓ_2 with distortion $O(\log^2 n)$. □

3.1 Overview of the Proof of Bourgain's Theorem

We will prove Bourgain's theorem in the next lecture, but we provide a brief sketch of the analysis here.

As in Theorem 2.4, we will construct a Fréchet embedding with random sets of varying densities.

To perform the analysis, for each pair (u, v) we will grow balls around u and v with radii in $[0, d(u, v)/2]$. Let $q = \lfloor \log_2 n \rfloor + 1$ and $r_1, r_2, \ldots r_q$ be radii such that r_j is the smallest choice of radius such that both $|B(u, r_j)|, |B(v, r_j)| \geq 2^j$ (if this radius is greater than $d(u, v)/2$, we set $r_j = d(u, v)/2$). We define $r_0 = 0$ and $\Delta_j = r_j - r_{j-1}$ so that $\displaystyle\sum_{j=1}^{q} \Delta_j \geq d(u, v)/2$.

The following claim forms the core of the proof.

Claim 3.3. *If a random set A_j is sampled from V such that each point is included independently with probability $\frac{1}{2^j}$, then*

$$\Pr[|d(u, A_j) - d(v, A_j)| \geq \Delta_j] \geq 1/12.$$

Assuming this claim, we can show using the Cauchy-Schwarz inequality that if enough sets are sampled with each of the densities 2^{-j}, then with constant probability, the embedding into ℓ_2 does not contract the distance between any two points too much.

References

[1] Jean Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, Mar 1985.

[2] Jiri Matousek. *Lectures on Discrete Geometry.* Springer-Verlag New York, 2002.