Metric and Normed Spaces II, Bourgain's Theorem

Geometric Methods in Computer Science

Instructor: Yury Makarychev

1 Notation

Given a metric space (X, d) and $S \subset X$, the distance from $x \in X$ to S equals

$$d(x, S) = \inf_{s \in S} d(x, s).$$

The distance between two sets $S_1, S_2 \subset X$ equals

$$d(S_1, S_2) = \inf_{s_1 \in S_1, s_2 \in S_2} d(s_1, s_2).$$

Exercise 1. *Show that distances between sets do not necessarily satisfy the triangle inequality. That is, it is possible that $d(S_1, S_2) + d(S_2, S_3) > d(S_1, S_3)$ for some sets S_1, S_2 and S_3.*

Exercise 2. *Prove that $d(x, y) \geq d(S, x) - d(S, y)$ and thus $d(x, y) \geq |d(S, x) - d(S, y)|$.*

Proof. Fix $\varepsilon > 0$. Let $y' \in S$ be such that $d(y', y) \leq d(S, y) + \varepsilon$ (if S is a finite set, there is $y' \in S$ s.t. $d(y, y') = d(S, y)$). Then

$$d(x, S) \leq d(x, y') \leq d(x, y) + d(y, y') \leq d(x, y) + d(S, y) + \varepsilon.$$

We proved that $d(x, S) \leq d(x, y) + d(S, y) + \varepsilon$ for every $\varepsilon > 0$. Therefore,

$$d(x, S) \leq d(x, y) + d(S, y).$$

□

2 Bourgain's Theorem

Definition 2.1. *Let X be a finite metric space and $p \geq 1$. Suppose that $Z \neq \varnothing$ is a random subset of X (chosen according to some probability distribution). For every $u \in X$, define random variable $\xi_u = d(u, Z) = \min_{z \in Z} d(u, z)$. Consider the map f from X to the space of random variables $L_p(\Omega, \mu)$ that sends u to ξ_u (where Ω is the probability space and μ is the probability measure on Ω). We say that f is a Fréchet embedding.*

Lemma 2.2. *Every Fréchet embedding f is non-expanding. That is, $\|f\|_{Lip} \leq 1$.*

Proof. Consider a Fréchet embedding that sends u to $\xi_u = d(u, Z)$. For every $u, v \in X$, we have

$$\|\xi_u - \xi_v\|_p = (\mathbb{E}\left[|d(u,Z) - d(v,Z)|^p\right])^{1/p} \overset{\text{by Exercise 2}}{\leq} (\mathbb{E}\left[|d(u,v)|^p\right])^{1/p} = d(u,v).$$

□

Remark 2.3. *If X is infinite, then the random variable $\xi_u = d(u, Z)$ does not necessarily belong to $L_p(\Omega, \mu)$ (its p-norm might be infinite). However, we can define $\tilde{\xi}_u$ as $\tilde{\xi}_u = d(u, Z) - d(x_0, Z)$, where x_0 is some point in X. Then the proof of Lemma 2.2 shows that $\|\tilde{\xi}_u\|_p \leq d(u, x_0) < \infty$ and the map $f : u \mapsto \tilde{\xi}_u$ is non-expanding.*

Theorem 2.4 (Bourgain's Theorem)**.** *Every metric space X on n points embeds into $L_p(X, \mu)$ with distortion $O(\log n)$ (for every $p \geq 1$). That is, $c_p(X) = O(\log n)$.*

Proof. Let $l = \lceil \log_2 n \rceil + 1$. Construct a random set Z as follows.

- Choose s uniformly at random from $\{1, \ldots, l\}$.
- Initially, let $Z = \varnothing$.
- Add every point of X to Z with probability $1/2^s$, independently.

Now let f be the Fréchet embedding that maps $u \in X$ to random variable $\xi_u = d(Z, u)$. By Lemma 2.2, f is non-expanding. We are going to prove that for every u and v,

$$\|f(u) - f(v)\|_p \geq \frac{c}{l} \cdot d(u, v),$$

for some absolute constant c. Note that it is sufficient to prove this statement for $p = 1$, since by Lyapunov's inequality $\|f(u) - f(v)\|_p \geq \|f(u) - f(v)\|_1$.

Consider two points u and v. Let $\Delta = d(u,v)/2$. Define interval I_Z as follows: $I_Z = [d(u,Z), d(v,Z)]$ if $d(u,Z) \leq d(v,Z)$, and $I_Z = [d(v,Z), d(u,Z)]$ if $d(v,Z) < d(u,Z)$. That is, I_Z is the interval between $d(u,Z)$ and $d(v,Z)$. Denote the length of I_Z by $|I_Z|$. Let $\mathbf{1}_{I_Z}$ be the indicator function of I_Z. Write,

$$|d(u,Z) - d(v,Z)| = |I_Z| = \int_{I_Z} 1\, dt = \int_0^\infty \mathbf{1}_{I_Z}(t) dt.$$

Then,

$$\begin{aligned}\|f(u) - f(v)\|_1 &= \mathbb{E}\left[|d(u,Z) - d(v,Z)|\right] = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{I_Z}(t) dt\right] \\ \left(\substack{\text{by Fubini's} \\ \text{theorem}}\right) &= \int_0^\infty \mathbb{E}\left[\mathbf{1}_{I_Z}(t)\right] dt = \int_0^\infty \Pr\left(t \in I_Z\right) dt \geq \int_0^\Delta \Pr\left(t \in I_Z\right) dt.\end{aligned}$$

We now prove that $\Pr(t \in I_Z) \geq \frac{\Omega(1)}{l}$ if $t \in (0, \Delta)$. That will imply that $\|f(u) - f(v)\|_1 \geq \frac{\Omega(1)}{l} \cdot \Delta = \frac{\Omega(1)}{l} \cdot d(u, v)$.

Fix $t \in (0, \Delta)$. Consider balls $B_t(u)$ and $B_t(v)$. They are disjoint since $2t < 2\Delta = d(u, v)$. Assume without loss of generality that $|B_t(u)| \leq |B_t(v)|$. Denote $m = |B_t(u)|$. Let $s_0 = \lfloor \log_2 m \rfloor + 1$. Then $m < 2^{s_0} \leq 2m$. Let $\mathcal{E}_u$ be the event that $d(u, Z) > t$, and $\mathcal{E}_v$ be the event that $d(v, Z) \leq t$. We have,

$$\begin{aligned}\Pr(t \in I_Z) &= \Pr(d(u, Z) \leq t \leq d(v, Z) \text{ or } d(v, Z) \leq t \leq d(u, Z)) \\ &\geq \Pr(d(v, Z) \leq t < d(u, Z)) = \Pr(\mathcal{E}_u \text{ and } \mathcal{E}_v).\end{aligned}$$

Event $\mathcal{E}_v$ occurs if and only if there is a point in Z at distance at most t from v; that is, when $B_t(v) \cap Z \neq \varnothing$. Event $\mathcal{E}_u$ occurs if and only if $B_t(u) \cap Z = \varnothing$.

Consider the event $s = s_0$. It happens with probability $1/l$. Conditioned on this event, events $\mathcal{E}_u$ and $\mathcal{E}_v$ are independent (since $B_t(u)$ and $B_t(v)$ are disjoint) and

$$\Pr(\mathcal{E}_u | s = s_0) = \prod_{w \in B_t(u)} \Pr(w \notin Z | s = s_0) = \prod_{w \in B_t(u)} \left(1 - \frac{1}{2^{s_0}}\right) = \left(1 - \frac{1}{2^{s_0}}\right)^m \geq \frac{1}{e}.$$

$$\begin{aligned}\Pr(\mathcal{E}_v | s = s_0) &= 1 - \prod_{w \in B_t(v)} \Pr(w \notin Z | s = s_0) = 1 - \prod_{w \in B_t(v)} \left(1 - \frac{1}{2^{s_0}}\right) \geq 1 - \left(1 - \frac{1}{2^{s_0}}\right)^m \\ &\geq 1 - \frac{1}{e^{1/2}}.\end{aligned}$$

We get

$$\begin{aligned}\Pr(t \in I_Z) &\geq \Pr(\mathcal{E}_u \text{ and } \mathcal{E}_v) \geq \Pr(s = s_0) \Pr(\mathcal{E}_u \text{ and } \mathcal{E}_v | s = s_0) \\ &\geq \frac{1}{l} \Pr(\mathcal{E}_u | s = s_0) \Pr(\mathcal{E}_v | s = s_0) \geq \Omega\left(\frac{1}{l}\right).\end{aligned}$$

□

Exercise 3. *The set Z might be equal to $\varnothing$ in our proof, then random variables $\xi_u = d(u, Z)$ are not well defined. Show how to fix this problem.*

Proof. There are many ways to fix this problem. For instance, we can add an extra point x_∞ to the metric space X, and define $d(u, x_\infty) = 2\,\mathrm{diam}(X)$, where $\mathrm{diam}(X) = \max_{u,v \in X} d(u, v)$. Then construct the set Z as before, except that always add x_∞ to Z. Thus we ensure that $Z \neq \varnothing$. In other words, we can define ξ_u as before if $Z \neq \varnothing$, and $\xi_u = 2\,\mathrm{diam}(X)$ if $Z = \varnothing$. The rest of the proof goes through without any other changes. □

The proof of Bourgain's theorem provides an efficient randomized procedure for generating set Z. As presented here, this procedure gives an embedding only in $L_p(\Omega, \mu)$ and not in ℓ_p^N. We already know that if a set of n points embeds in $L_p(\Omega, \mu)$ with distortion D then it embeds in $\ell_p^{\binom{n}{2}}$ with distortion D. However, in fact, we need only $N = O((\log n)^2)$

dimensions: for every value of $s \in \{1, \dots, l\}$ we make $\Theta(\log n)$ samples of the set Z. Then the total number of samples equals $\Theta((\log n)^2)$. Using the Chernoff bound, it is easy to show that the distortion of the obtained embedding is $O(\log n)$ w.h.p.

Fact 2.5 (Matoušek)**.** *Let $D_{n,p}$ be the smallest number D such that every metric space on n points embeds in ℓ_p with distortion at most $D_{n,p}$. Then*

$$D_{n,p} = \Theta\left(\frac{\log n}{p}\right).$$