Bourgain’s Theorem
Computational and Metric Geometry
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,

\[ d(x, S) \leq d(x, y) + d(S, y). \]

**Definition 1.1.** Let \((X, d)\) be a metric space, \(x_0 \in X\) and \(r > 0\). The (closed) ball of radius \(r\) around \(x_0\) is

\[ B_r(x_0) = \text{Ball}_r(x_0) = \{x : d(x, x_0) \leq r\}. \]
2 Warm-up

Consider two normed spaces \((U, \| \cdot \|_U)\) and \((V, \| \cdot \|_V)\). Let \(f\) be a linear operator between \(U\) and \(V\). What is the Lipschitz norm of \(f\)? It is equal

\[
\sup_{x, y \in U, x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|_U}
\]

by linearity of \(f\)

\[
\sup_{x, y \in U, x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|_U} = \sup_{z \in U, z \neq 0} \frac{\|f(z)\|_V}{\|z\|_U}.
\]

The expression \(\sup_{z \in U, z \neq 0} \frac{\|f(z)\|_V}{\|z\|_U}\) is called the operator norm of \(f\). The above computation shows that the Lipschitz norm of a linear operator equals its operator norm.

At the previous lecture, we proved that \(\ell_q \subset \ell_p\) and \(L_p[0,s] \subset L_q[0,s]\) when \(p < q\) and \(a, b \in \mathbb{R}\). These embeddings define inclusion maps \(i_1 : \ell_p \hookrightarrow \ell_q\) and \(i_2 : L_q[0,s] \hookrightarrow L_p[0,s]\) defined by

\[
i_1(a) = a \quad \text{for every } a \in \ell_p \quad \text{and} \quad i_2(f) = f \quad \text{for every } f \in L_q[0,s].
\]

Note that even though maps \(i_1\) and \(i_2\) “do not do much”—they just map every element to itself—they are not low distortion maps!

**Exercise 3.** Compute the Lipschitz norm and distortion of map \(i_1 : \ell_p \hookrightarrow \ell_q\).

**Solution.** Consider \(a \in \ell_p\). Note that \(|a_j| \leq \|a\|_p\) and \(|a_j|q = |a_j|p \cdot |a_j|q-p \leq |a_j|p \cdot \|a\|q-p\).

Therefore,

\[
\|a\|_q^q = \sum |a_j|q \leq \sum |a_j|p \cdot \|a\|q-p = (\sum |a_j|p) \cdot \|a\|q-p = \|a\|p \cdot \|a\|q-p = \|a\|q.
\]

We get that \(\|a\|_q \leq \|a\|_p\). That is, \(\|i_1\|_{\text{Lip}} \leq 1\). On the other hand, \(\|e_1\|_p = \|e_1\|_q = 1\), where \(e_1 = (1,0,\ldots)\). We get, \(\|i_1\|_{\text{Lip}} = 1\).

Now let \(n \geq 1\). Consider \(a = (\underbrace{1,0,\ldots,0}_n,\ldots) \in \ell_p\). We have, \(\|a\|_p = n^{1/p}\) and \(\|a\|_q = n^{1/q}\). Thus

\[
\|i_1^{-1}\|_{\text{Lip}} \geq \frac{\|a\|_p}{\|a\|_q} = \frac{n^{1/p}}{n^{1/q}} = n^{1/p-1/q}.
\]

Since \(n^{1/p-1/q} \to \infty \) as \(n \to \infty\), the norm \(\|f^{-1}\|_{\text{Lip}}\) is unbounded.

**Answer:** \(\|i_1\|_{\text{Lip}} = 1\), \(i_1\) has infinite distortion.

We will need the following inequality.

**Theorem 2.1 (Lyapunov’s inequality).** Let \(1 \leq p < q = \infty\). For every random variable \(\alpha\) with finite \(q\)-th moment, we have \(\|\alpha\|_p \leq \|\alpha\|_q\).

**Proof.** The statement is obvious for \(q = \infty\) since \(|\alpha| \leq \|\alpha\|_\infty\) almost surely. Let us assume that \(q < \infty\). Let \(f(x) = x^{q/p}\) for \(x \geq 0\). Note that \(f(x)\) is a convex function. Let \(\beta = |\alpha|^p\) (\(\beta\) is a random variable). We have

\[
\|\alpha\|_q^q = \mathbb{E}[|\alpha|^q] = \mathbb{E}[|\beta|^{q/p}] = \mathbb{E}[f(|\beta|)] \geq f(\mathbb{E}[|\beta|]) = (\mathbb{E}[|\alpha|^p])^{q/p}.
\]

We conclude that \(\|\alpha\|_q \geq \|\alpha\|_p\) as required.

\[2\]
Exercise 4. Compute the Lipschitz norm and distortion of map $i_2 : L_q[0, s] \hookrightarrow L_p[0, s]$.

Proof. First, consider $f(x) = 1$, a constant function defined on $[0, s]$. We have $\|f\|_{L_p} = s^{1/p}$ and $\|f\|_{L_q} = s^{1/q}$. Therefore,

$$\|i_2\|_{\text{Lip}} \geq \frac{\|i_2(f)\|_p}{\|f\|_q} = s^{1/p-1/q}.$$ 

Now consider $f \in L_q[0, s]$. Let $\xi$ be a random variable uniformly distributed on $[0, s]$. Note that for every function $h$ on $[0, s]$, we have

$$\int_0^s h(x)dx = s \int_0^1 h(ys)dy = s \mathbb{E}\left[ h(\xi) \right].$$

Therefore, $\|f\|^p_{L_p} = s \mathbb{E}\left[ |f(\xi)|^p \right] = s \|f(\xi)\|^p_{L_p}$, and $\|f\|_{L_p} = s^{1/p} \|f(\xi)\|_{L_p}$ (here, $f(\xi)$ is a random variable). Similarly, $\|f\|_{L_q} = s^{1/q} \|f(\xi)\|_{L_q}$. By Lyapunov’s inequality for $\alpha = f(\xi)$,

$$\|f\|_{L_p} = s^{1/p} \|f(\xi)\|_{L_q} \leq s^{1/p} \|f(\xi)\|_{L_q} = s^{1/p-1/q} \left( s^{1/q} \|f(\xi)\|_{L_q} \right) = s^{1/p-1/q} \|f\|_{L_q}.$$ 

We get that $\|f\|_{\text{Lip}} \leq s^{1/p-1/q}$ and therefore $\|f\|_{\text{Lip}} = s^{1/p-1/q}$

Let $\varepsilon \in (0, 1)$. Consider $f_\varepsilon(x) = x^{1-\varepsilon}$. We compute its $p$ and $q$ norms and get that

$$\|f_\varepsilon\|_p \rightarrow \left( \frac{qs-q}{q-p} \right)^{1/p} < \infty \quad \text{as } \varepsilon \rightarrow 0,$$

$$\|f_\varepsilon\|_q \rightarrow \infty \quad \text{as } \varepsilon \rightarrow 0.$$ 

Therefore, $i_2$ has infinite distortion.

Answer: $\|i_2\|_{\text{Lip}} = s^{1/p-1/q}$, $i_2$ has infinite distortion.

3 Bourgain’s Theorem

Definition 3.1. Let $X$ be a finite metric space and $p \geq 1$. Suppose that $Z \neq \emptyset$ 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 $\xi_u$ (where $\Omega$ is the probability space and $\mu$ is the probability measure on $\Omega$). We say that $f$ is a Fréchet embedding.

Lemma 3.2. Every Fréchet embedding $f$ is non-expanding. That is, $\|f\|_{\text{Lip}} \leq 1$.

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

$$\|\xi_u - \xi_v\|_p = (\mathbb{E}[|d(u, Z) - d(v, Z)|^p])^{1/p} \leq (\mathbb{E}[|d(u, v)|^p])^{1/p} = d(u, v).$$ 

□
Remark 3.3. 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 $\hat{\xi}_u$ as $\hat{\xi}_u = d(u, Z) - d(x_0, Z)$, where $x_0$ is some point in $X$. Then the proof of Lemma 3.2 shows that $\|\hat{\xi}_u\|_p \leq d(u, x_0) < \infty$ and the map $f: u \mapsto \hat{\xi}_u$ is non-expanding.

Theorem 3.4 (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 $Z = \emptyset$.
- 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 3.2, $f$ 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 $\Delta = d(u, v)/2$. Write,

$$\|f(u) - f(v)\|_1 = \mathbb{E} \left[ |d(u, Z) - d(v, Z)| \right] = \mathbb{E} \left[ \int_{[d(u, Z), d(v, Z)]} \right] \cdot dt \quad \text{(by Fubini’s theorem)}$$

$$\geq \int_0^\Delta \mathbb{P} \left( d(u, Z) \leq t < d(v, Z) \right) dt.$$

We now prove that $\mathbb{P} \left( d(u, Z) \leq t < d(v, Z) \right)$ or $d(v, Z) \leq t < d(u, Z)$.

That will imply that $\|f(u) - f(v)\|_1 \geq \Omega(1) \cdot \Delta = \frac{\Omega(1)}{l} \cdot d(u, v)$.

We fix $t \in (0, \Delta)$. Consider balls $B_t(u)$ and $B_t(v)$. Note that 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 $E_u$ be the event that $d(u, Z) > t$ and $E_v$ be the event that $d(v, Z) \leq t$. We have,

$$\mathbb{P} \left( d(u, Z) < t < d(v, Z) \right) \geq \mathbb{P} \left( E_u \text{ and } E_v \right).$$

Note that the event $E_u$ occurs if and only if there is a point in $Z$ at distance at most $t$ from $v$; that is, when $B_t(v) \cap Z \neq \emptyset$. The event $E_u$ occurs if and only if $B_t(u) \cap Z = \emptyset$. The proof follows from these considerations.
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

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

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

We get

$$\Pr(d(u, Z) < t < d(v, Z) \text{ or } d(v, Z) < t < d(u, Z)) \geq \Pr(\mathcal{E}_u \text{ and } \mathcal{E}_v) \geq \frac{1}{l} \Pr(\mathcal{E}_u|s = s_0) \Pr(\mathcal{E}_v|s = s_0) \geq \Omega \left(\frac{1}{l}\right).$$

**Exercise 5.** The set $Z$ might be equal to $\emptyset$ 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 $x_{\infty}$ to the metric space $X$, and define $d(u, x_{\infty}) = 2 \text{diam}(X)$, where $\text{diam}(X) = \max_{u, v \in X} d(u, v)$. Then construct the set $Z$ as before, except that always add $x_{\infty}$ to $Z$. Thus we ensure that $Z \neq \emptyset$. In other words, we can define $\xi_u$ as before if $Z \neq \emptyset$, and $\xi_u = 2 \text{diam}(X)$ if $Z = \emptyset$. 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 $\ell_p^n$. We already know that if a set of $n$ points embeds in $L_p(\Omega, \mu)$ with distortion $D$ then it embeds in $\ell_p^n$ with distortion $D$. However, in fact, we need only $N = O((\log n)^2)$ dimensions: for every value of $s \in \{1, \ldots, l\}$ we make $\Theta((\log n)^2)$ 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 3.5** (Matoušek). Let $D_{n,p}$ be the smallest number $D$ such that every metric space on $n$ points embeds in $\ell_p$ with distortion at most $D_{n,p}$. Then

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