Metric Embedddings

Lecture notes by Konstantin Makarychev

1 Bourgain's Theorem

Consider a vector x in an m dimensional space $\mathbb{R}^m$. The ℓ_p norm of x is defined as follows

$$\|x\|_p = \big(\sum_{i=1}^{m} |x_i|^p\big)^{1/p}$$

for $p \geq 1$. Particularly, for $p = 1$, we have

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|.$$

The ℓ_1 distance between two vectors x and y equals $\|x - y\|_1$:

$$\|x - y\|_1 = \sum_{i=1}^{m} |x_i - y_i|.$$

We say that a metric space (V, d) embeds into ℓ_1 with distortion D if there exists a map $\varphi : V \to \mathbb{R}^m$ (for some m) and positive numbers α and β such that $\beta/\alpha \leq D$ and for every $u, v \in V$:

$$\alpha\, d(u, v) \leq \|\varphi(u) - \varphi(v)\|_1 \leq \beta\, d(u, v).$$

Note that by rescaling vectors $\varphi(u)$ we may always assume that either α or β equals 1.

Theorem 1.1 (Bourgain)**.** *Every finite metric space embeds into ℓ_1 with distortion $D = O(\log n)$, where $n = |V|$.*

1.1 Frechet Embeddings

To prove Bourgain's theorem we need the notion of Frechet embeddings.

Fix a metric space (V, d) and consider a $2^n - 1$ dimensional space indexed by non-empty subsets $Z \subset V$. In this space, the coordinates of any point x are numbers x_Z for all non-empty $Z \subset V$. We denote the space by X. Now, for any probabilistic distribution $\mathcal{D}$ of subsets $Z \subset V$, we define an embedding φ to X as follows:

$$\varphi(u)_Z = \Pr_{\mathcal{D}}(Z) \cdot d(u, Z).$$

Here, $d(u, Z)$ denotes the distance from u to the set Z, which equals (by definition)

$$d(u, Z) = \min_{z \in Z} d(u, z).$$

Lemma 1.2. *Frechet embeddings do not increase distances: for every Frechet embedding φ and every u and v in V,*

$$\|\varphi(u) - \varphi(v)\|_1 \leq d(u, v).$$

Let us first verify that the distance to a given set satisfies the triangle inequality.

Lemma 1.3. *For every $u, v \in V$, and every set $Z \subset V$,*

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

Proof. Assume without loss of generality that $d(u, Z) \geq d(v, Z)$. By the definition $d(v, Z) = d(v, z^*)$ for some $z^* \in Z$. Observe that

$$d(u, Z) = \min_{z \in Z} d(u, z) \leq d(u, z^*).$$

Thus,

$$|d(u, Z) - d(v, Z)| = d(u, Z) - d(v, Z) \leq d(u, z^*) - d(v, z^*) \leq d(u, v).$$

□

Proof of Lemma 1.2. Pick arbitrary vertices u and v. The distance between $\varphi(u)$ and $\varphi(v)$ equals:

$$\|\varphi(u) - \varphi(v)\|_1 = \sum_{Z \subset V : Z \neq \varnothing} \Pr_{\mathcal{D}}(Z)\ |d(u, Z) - d(v, Z)|.$$

By Lemma 1.3, $|d(u, Z) - d(v, Z)| \leq d(u, v)$. Hence,

$$\|\varphi(u) - \varphi(v)\|_1 \leq \sum_{Z \subset V : Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) d(u, v) \leq d(u, v),$$

since $\sum_{Z \subset V} \Pr_{\mathcal{D}}(Z) = 1$. □

To prove Bourgain's theorem, we need to find a probabilistic distribution $\mathcal{D}$ such that

$$\|\varphi(u) - \varphi(v)\|_1 = \sum_{Z \subset V : Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) |d(u, Z) - d(v, Z)| \geq \frac{d(u, v)}{O(\log n)}.$$

Then, we will have

$$\frac{d(u, v)}{O(\log n)} \leq \|\varphi(u) - \varphi(v)\|_1 \leq d(u, v).$$

We define $\mathcal{D}$ on $Z \subset V$ via the following randomized algorithm:

1. Pick a random number $t \in \{1, \dots, \lceil \ln n \rceil\}$. Let $p = 1 - \exp(-e^{-t})$.
2. Pick each element v in Z independently at random with probability p, and add it to the set Z.
3. Return the set Z.

Lemma 1.4. *Consider two disjoint non-empty sets $A \subset V$ and $B \subset V$. Then,*

$$\Pr_{\mathcal{D}}(Z \cap A = \varnothing;\ Z \cap B \neq \varnothing) + \Pr_{\mathcal{D}}(Z \cap A \neq \varnothing;\ Z \cap B = \varnothing) \geq \frac{1}{10 \lceil \ln n \rceil}.$$

Proof. Assume without loss of generality that $|A| \geq |B|$. We estimate the probability of the event

$$\{Z \cap A \neq \varnothing; \text{ and } Z \cap B = \varnothing\}.$$

Write,

$$\begin{aligned}
\Pr_{\mathcal{D}}(Z \cap A \neq \varnothing;\ Z \cap B = \varnothing) &= \frac{1}{\lceil \ln n \rceil} \sum_{i=1}^{\lceil \ln n \rceil} \Pr_{\mathcal{D}}(Z \cap A \neq \varnothing;\ Z \cap B = \varnothing \quad | \quad t = i) \\
&= \frac{1}{\lceil \ln n \rceil} \sum_{i=1}^{\lceil \ln n \rceil} \Pr_{\mathcal{D}}(Z \cap A \neq \varnothing \quad | \quad t = i) \cdot \Pr_{\mathcal{D}}(Z \cap B = \varnothing \quad | \quad t = i).
\end{aligned}$$

Here we used that the sets A and B are disjoint, and hence the events $\{Z \cap A \neq \varnothing\}$ and $\{Z \cap B = \varnothing\}$ are independent for any fixed t. Now, fix $i \in \{1, \ldots, \lceil \ln n \rceil\}$.

$$\begin{aligned}
\Pr_{\mathcal{D}}(Z \cap B = \varnothing \quad | \quad t = i) &= \prod_{v \in B} \Pr_{\mathcal{D}}(v \text{ is not chosen in } Z \quad | \quad t = i) \\
&= \prod_{v \in B} (1 - (1 - \exp(-e^{-i}))) \\
&= \prod_{v \in B} \exp(-e^{-i}) = \exp(-|B| e^{-i}) = \exp(-e^{\ln |B| - i}).
\end{aligned}$$

Similarly,

$$\Pr_{\mathcal{D}}(Z \cap A \neq \varnothing \quad | \quad t = i) = 1 - \exp(-e^{\ln |A| - i}).$$

Therefore,

$$\begin{aligned}
\Pr_{\mathcal{D}}(Z \cap A \neq \varnothing;\ Z \cap B = \varnothing) &= \frac{1}{\lceil \ln n \rceil} \sum_{i=1}^{\lceil \ln n \rceil} \Pr_{\mathcal{D}}(Z \cap A \neq \varnothing \quad | \quad t = i) \cdot \Pr_{\mathcal{D}}(Z \cap B = \varnothing \quad | \quad t = i) \\
&= \frac{1}{\lceil \ln n \rceil} \sum_{i=1}^{\lceil \ln n \rceil} (1 - \exp(-e^{\ln |A| - i})) \cdot \exp(-e^{\ln |B| - i}).
\end{aligned}$$

Note that all terms in the sum are positive, and for $i = \lceil \ln |A| \rceil$,

$$(1 - \exp(-e^{\ln |A| - i})) \cdot \exp(-e^{\ln |B| - i}) \geq (1 - \exp(-e^{-1})) \exp(-e^{0}) \geq \frac{1}{10}.$$

The last inequality can be checked numerically. Therefore,

$$\Pr_{\mathcal{D}}(Z \cap A \neq \varnothing;\ Z \cap B = \varnothing) \geq \frac{1}{10 \lceil \ln n \rceil}.$$

This completes the proof. $\square$

We are now ready to finish the proof of Bourgain's theorem. Recall, that it remains to show that $\|\varphi(u) - \varphi(v)\|_1 \geq d(u, v)/O(\log n)$ for every $u, v \in V$. Write,

$$\|\varphi(u) - \varphi(v)\|_1 = \sum_{Z \subset V : Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) |d(u, Z) - d(v, Z)|.$$

Observe that for every nonnegative numbers a and b,

$$|a-b| = \int_0^\infty I(a \le r \le b) + I(b \le r \le a) dr,$$

where $I(a \le r \le b)$ and $I(b \le r \le a)$ are the indicator functions of the events $\{a \le r \le b\}$ and $\{b \le r \le a\}$ respectively. Using this observation for $a = d(u,Z)$ and $b = d(v,Z)$, we get

$$|d(u,Z) - d(v,Z)| = \int_0^\infty I(d(u,Z) \le r \le d(v,Z)) + I(d(v,Z) \le r \le d(u,Z)) dr.$$

Consequently,

$$\begin{aligned}
\|\varphi(u) - \varphi(v)\|_1 &= \sum_{Z \subset V: Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) |d(u,Z) - d(v,Z)| \qquad (1)\\
&= \sum_{Z \subset V: Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) \int_0^\infty I(d(u,Z) \le r \le d(v,Z)) + I(d(v,Z) \le r \le d(u,Z)) dr\\
&\int_0^\infty \sum_{Z \subset V: Z \neq \varnothing} \Pr_{\mathcal{D}}(Z) \big(I(d(u,Z) \le r \le d(v,Z)) + I(d(v,Z) \le r \le d(u,Z))\big) dr\\
&= \int_0^\infty \Pr_{Z \sim \mathcal{D}}(d(u,Z) \le r \le d(v,Z)) + \Pr_{Z \sim \mathcal{D}}(d(v,Z) \le r \le d(u,Z))\big) dr.
\end{aligned}$$

Let $\mathrm{Ball}(u,r)$ be the ball of radius r around u in the metric space V i.e.,

$$\mathrm{Ball}(u,r) = \{w \in V : d(u,w) \le r\}.$$

Then,

$$\begin{aligned}
\Pr_{Z \sim \mathcal{D}}(d(u,Z) \le r \le d(v,Z)) &= \Pr_{Z \sim \mathcal{D}}(\mathrm{Ball}(u,r) \cap Z \neq \varnothing;\ \mathrm{Ball}(v,r) \cap Z = \varnothing);\\
\Pr_{Z \sim \mathcal{D}}(d(v,Z) \le r \le d(u,Z)) &= \Pr_{Z \sim \mathcal{D}}(\mathrm{Ball}(v,r) \cap Z \neq \varnothing;\ \mathrm{Ball}(u,r) \cap Z = \varnothing);
\end{aligned}$$

Note that the balls $\mathrm{Ball}(u,r)$ and $\mathrm{Ball}(v,r)$ are disjoint if $r < d(u,v)/2$ and are non-empty for every $r \ge 0$ (simply because $u \in \mathrm{Ball}(u,r)$ and $v \in \mathrm{Ball}(v,r)$). Thus, by Lemma 1.4, for $r \le d(u,v)/2$,

$$\Pr_{Z \sim \mathcal{D}}(d(u,Z) \le r \le d(v,Z)) + \Pr_{Z \sim \mathcal{D}}(d(v,Z) \le r \le d(u,Z)) \ge \frac{1}{10\lceil \ln n \rceil}.$$

Plugging this bound in (1), we get

$$\|\varphi(u) - \varphi(v)\|_1 \ge \frac{d(u,v)}{10\lceil \ln n \rceil}.$$

This concludes the proof of Bourgain's theorem.

2 ℓ_1 as a positive combination of cut metrics

We now show that every ℓ_1 metric can be represented as a positive combination of cut metrics and, conversely, every positive combination of cut metrics can be isometrically embedded into ℓ_1. Recall, that for every subset $S \subset V$ the cut metric is defined as follows:

$$\delta_S(x,y) = \begin{cases} 1, & \text{if } u \in S,\ v \notin S \text{ or } u \notin S,\ v \in S;\\ 0, & \text{if } u \in S,\ v \in S \text{ or } u \notin S,\ v \notin S. \end{cases}$$

Theorem 2.1. *If the distance function d of a metric space (V, d) can be represented as*

$$d(u, v) = \sum_{S \subset V} \lambda_S \delta_S(u, v)$$

for some nonnegative numbers λ_S (here δ_S are cut metrics), then (V, d) can be isometrically embedded into ℓ_1.

Proof. We define an embedding φ from (V, d) into $2^{|V|}$ dimensional ℓ_1 space. We index the coordinates of the host space by subsets $S \subset V$. Let

$$\varphi(u)_S = \begin{cases} \lambda_S, & \text{if } u \in S; \\ 0, & \text{if } u \notin S. \end{cases}$$

Then,

$$\|\varphi(u) - \varphi(v)\|_1 = \sum_{S \subset V} |\varphi(u)_S - \varphi(v)_S|.$$

Observe that $|\varphi(u)_S - \varphi(v)_S| = \lambda_S \delta_S(u, v)$. To verify this equality just consider four cases $u \in S$; $v \in S$, $u \in S$; $v \notin S$, $u \notin S$; $v \in S$, $u \notin S$; $v \notin S$. Thus,

$$\|\varphi(u) - \varphi(v)\|_1 = \sum_{S \subset V} \lambda_S \delta_S(u, v) = d(u, v).$$

□

Theorem 2.2. *For every finite subset X of $\mathbb{R}^m$, there exists nonnegative numbers λ_S (where $S \subset X$) such that for all $x, y \in X$*

$$\|x - y\|_1 = \sum_{S \subset X} \delta_S(x, y).$$

Proof. Let $M = \max_{\substack{x \in X \\ 1 \le i \le d}} |x_i|$. We chose M in such a way that all coordinates x_i lied in the range $[-M, M]$. Consider the following random process: Pick a random $i \in \{1, \ldots, d\}$. Pick a random $t \in [-M, M]$. Output the set $U = \{x \in X : x_i \le t\}$. This procedure defines a probabilistic distribution on the set of all subsets of X. We now let

$$\lambda_S = \Pr(U = S)$$

and $d(x, y) = \sum_{S \subset X} \lambda_S \delta_S(x, y)$. Then, we have

$$\begin{aligned} d(x, y) &= \sum_{S \subset X} \lambda_S \delta_S(x, y) = \sum_{S \subset X} \Pr(U = S) \delta_S(x, y) \\ &= \mathbb{E}_U[\delta_U(x, y)] \\ &= \Pr(x \in U; y \notin U) + \Pr(x \notin U; y \in U) \\ &= \Pr(x_i \le t < y_i) + \Pr(y_i \le t < x_i) \\ &= \frac{1}{m} \sum_{i=1}^{m} \frac{|x_i - y_i|}{M} \\ &= \frac{\|x - y\|_1}{Mm}. \end{aligned}$$

We now rescale λ by Mm: we let $\lambda'_S = mM\lambda_S$. For $d'(x, y) = \sum_{S \subset X} \lambda'_S \delta_S(x, y)$, we have $d'(x, y) = \|x - y\|_1$. This finishes the proof. □