1 Metric space embeddings

Let (X, d) be a finite metric space with $n = |X|$. Recall that the distance function $d : X \times X \to \mathbb{R}_+$ satisfies the axioms of a metric: For all $x, y, z \in X$:

1. $d(x, y) = 0 \iff x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leqslant d(x, z) + d(z, y)$

While properties (2) and (3) are essential for us, the implication $d(x, y) = 0 \implies x = y$ is often not particularly important. If a distance function satisfies only (2), (3), and $d(x, x) = 0$, it is commonly called a *pseudometric.*

Metric spaces arise in a variety of mathematical and scientific domains since they abstract the properties of many natural notions of "similarity" between objects. Consider, for instance, the latency between nodes in a network, the travel-distance between cities, the edit distance between genetic sequences, or various similarity measures between proteins.

Often one might first try to understand a given metric space (X, d) by trying to compare it to a well-understood space. For instance, one could think about mapping $F : X \to \mathbb{R}^k$ into a Euclidean space $\mathbb{R}^k$ equipped with the Euclidean norm $\|x\|_2 = \sqrt{x_1^2 + \cdots + x_k^2}$. One way to measure how well this mapping preserves the geometry of X is via the *bilipschitz distortion.* This is the smallest number $D > 0$ such that

$$d(x, y) \leqslant \|F(x) - F(y)\|_2 \leqslant D \cdot d(x, y) \qquad \forall x, y \in X\,. \tag{1.1}$$

Today we will prove the following result.

Theorem 1.1 (Bourgain 1985)**.** *Every n-point metric space embeds into some Euclidean space $\mathbb{R}^k$ with bilipschitz distortion D, where $D \leqslant O(\log n)$.*

We will show that this is possible $k \leqslant O((\log n)^2)$, but in the next lecture, we will see that for general reasons, one can achieve $k \leqslant O(\log n)$.

2 Distances to subsets

2.1 Fréchet's embedding

Let us first show how we can achieve the significantly worse bounds $D \leqslant \sqrt{n}$ and $k = n$. Enumerate the points $X = \{x_1, x_2, \ldots, x_n\}$ and let $F : X \to \mathbb{R}^n$ be defined by $F(x) = (F_1(x), \ldots, F_n(x))$, where

$$F_i(x) = d(x, x_i)\,.$$

First, note that every coordinate is 1-Lipschitz: For all $x, y \in X$,

$$|F_i(x) - F_i(y)| = |d(x, x_i) - d(y, x_i)| \leqslant d(x, y)\,,$$

where we have used the triangle inequality. From this, we get

$$\|F(x) - F(y)\|_2^2 = \sum_{i=1}^{n} |F_i(x) - F_i(y)|^2 \leqslant n\, d(x,y)^2\,, \tag{2.1}$$

implying that $\|F(x) - F(y)\|_2 \leqslant \sqrt{n} \cdot d(x,y)$ for all $x, y \in X$.

On the other hand, for any $x \in X$, it holds that

$$\|F(x) - F(y)\|_2 \geqslant |F_i(x) - F_i(x_i)| = d(x, x_i)\,,$$

Therefore (1.1) is satisfied with $D = \sqrt{n}$.

2.2 Bourgain's embedding

To get improved distortion, we will construct our coordinates out of distances to *subsets* instead of simply to points. For a subset $S \subseteq X$ and $x \in X$, let us define

$$d(x,S) := \min_{y \in S} d(x,y)\,.$$

First, observe that such maps are also 1-Lipschitz: The triangle inequality yields

$$d(x,S) \leqslant d(y,S) + d(x,y)\,,$$

hence

$$|d(x,S) - d(y,S)| \leqslant d(x,y) \qquad \forall x, y \in X, S \subseteq X\,. \tag{2.2}$$

For some number $m \leqslant O(\log n)$ that we will choose later, let

$$\left\{ \boldsymbol{S}_{t,j} : t = 1, 2, \ldots, \lfloor \log_2 n \rfloor, j = 1, 2, \ldots, m \right\}\,.$$

denote independent random subsets $\boldsymbol{S}_{t,j} \subseteq X$, where $\boldsymbol{S}_{t,j}$ is formed by sampling every point of X independently with probability 2^{-t}. Our embedding is

$$\begin{aligned} F(x) = \Big(& d(x, S_{1,1}), \ldots, d(x, S_{1,m}), \\ & d(x, S_{2,1}), \ldots, d(x, S_{2,m}), \\ & \cdots \\ & d(x, S_{\lfloor \log_2 n \rfloor,1}), \ldots, d(x, S_{\lfloor \log_2 n \rfloor,m}) \Big)\,. \end{aligned}$$

From (2.2), we see that

$$\|F(x) - F(y)\|_2 \leqslant \sqrt{m \lfloor \log_2 n \rfloor} \cdot d(x,y) \qquad x, y \in X \tag{2.3}$$

(just as in in (2.1)).

We move on to the lower bound. To this end, we define the open and closed balls: For $R \geqslant 0$,

$$\begin{aligned} B(x,R) &= \{ y \in X : d(x,y) \leqslant R \}\,, \\ B^\circ(x,R) &= \{ y \in X : d(x,y) < R \}\,. \end{aligned}$$

Fix $x, y \in X$ and for $t = 1, 2, \ldots, \lfloor \log_2 n \rfloor$, let r_t be the smallest radius such that

$$\max \left\{ |B(x, r_t)|, |B(y, r_t)| \geqslant 2^t \right\}\,.$$

Let t^* be the smallest value of t such that $r_t \geqslant d(x,y)/4$ and reassign $r_{t^*} = d(x,y)/4$.

Note that

$$\frac{d(x,y)}{4} = r_1 + (r_2 - r_1) + (r_3 - r_2) + \cdots + (r_{t^*} - r_{t^*-1}) . \tag{2.4}$$

We will use the sets $S_{t,j}$ to get a contribution of $r_t - r_{t-1}$ to the lower bound, and therefore (2.4) shows we will get a contribution of $\Omega(d(x,y))$.

So consider now some $t \in \{1,2,\ldots,t^*\}$. For the sake of analysis, let $r_0 = 0$. Note that, by definition of r_t, we have at least one of $|B(x,r_{t-1})| \geqslant 2^{t-1}$ or $|B(y,r_{t-1})| \geqslant 2^{t-1}$. Without loss of generality, assume that it holds for x. It also true that $B^\circ(y,r_t)| < 2^t$. Let us summarize:

$$|B(x,r_{t-1})| \geqslant 2^{t-1}$$
$$|B^\circ(y,r_t)| < 2^t .$$

Let $S_t \subseteq X$ be a random subset where every point is sampled independently with probability 2^{-t}. Consider the event

$$\mathcal{E}_t = \left\{ S_t \cap B(x,r_{t-1}) \neq \emptyset \text{ and } S_t \cap B^\circ(y,r_t) = \emptyset \right\} .$$

Notice that

$$\mathcal{E}_t \text{ occurs} \implies |d(x,S_t) - d(y,S_t)| \geqslant r_t - r_{t-1} . \tag{2.5}$$

Claim 2.1. $\mathbb{P}(\mathcal{E}_t) \geqslant \frac{1}{12}$.

Proof. Observe that $r_t \leqslant d(x,y)/4$, hence $B(x,r_t)$ and $B(y,r_t)$ are disjoint. In particular, the two events composing $\mathcal{E}_t$ are independent, and it suffices to lower bound their probabilities separately. First, note that

$$\begin{aligned}
\mathbb{P}\left(S_t \cap B(x,r_{t-1}) \neq \emptyset\right) &\geqslant 1 - \mathbb{P}\left(S_t \cap B(x,r_{t-1}) = \emptyset\right) \\
&= 1 - \left(1 - 2^{-t}\right)^{|B(x,r_{t-1})|} \\
&\geqslant 1 - \left(1 - 2^{-t}\right)^{2^{t-1}} \\
&\geqslant 1 - \frac{1}{\sqrt{e}} \geqslant \frac{1}{3} ,
\end{aligned}$$

where we have used the fact that $(1 - \frac{1}{k})^k \leqslant \frac{1}{e}$ for $k \geqslant 1$. Next, calculate

$$\mathbb{P}\left(S_t \cap B^\circ(y,r_t) = \emptyset\right) = \left(1 - 2^{-t}\right)^{|B^\circ(y,r_t)|} \geqslant \left(1 - 2^{-t}\right)^{2^t} \geqslant \frac{1}{4} ,$$

where we have used $(1 - \frac{1}{k})^k \geqslant 1/4$ for $k \geqslant 2$. □

Now let $\mathcal{E}_{t,j}$ be the event corresponding to (2.5) for the set $S_{t,j}$.

Corollary 2.2. *If $\Omega(m)$ of the events $\{\mathcal{E}_{t,j} : j = 1,\ldots,m\}$ occur, then*

$$\|F(x) - F(y)\|_2^2 \geqslant \Omega(m)(r_t - r_{t-1})^2 .$$

We can say something more: If it holds that

$$\Omega(m) \text{ of the events } \{\mathcal{E}_{t,j} : j = 1,\ldots m\} \text{ occur for every } t = 1,2,\ldots,\lfloor \log_2 n \rfloor, \tag{2.6}$$

then since the contributions come from disjoint sets of coordinates,

$$\|F(x)-F(y)\|_2^2 \geqslant \Omega(m) \sum_{t=1}^{t^*} (r_t - r_{t-1})^2 \geqslant \Omega\left(\frac{m}{t^*}\right) \left(\sum_{t=1}^{t^*} (r_t - r_{t-1})\right)^2 \geqslant \Omega\left(\frac{m}{t^*}\right) d(x,y)^2 \geqslant \Omega\left(\frac{m}{\log n}\right) d(x,y)^2 .$$

The second inequality is Cauchy-Schwarz, and the third is from (2.4).

Combining this with (2.3), our map has distortion $O(\log n)$ as long as we choose m large enough so that (2.6) holds with probability, say, $1 - 1/n^3$. That's because we can then take a union bound over all possible pairs $x, y \in X$. But since each event $\mathcal{E}_{t,j}$ occurs with probability at least $1/12$, a simple Chernoff bound shows that choosing some $m \leqslant O(\log n)$ suffices.