CSC2414 - Metric Embeddings*
Lecture 2: Bourgain's theorem for metric embeddings

Notes taken by Periklis Papakonstantinou
Revised by Hamed Hatami

Summary: We begin by discussing a simple application of metric embeddings to the efficient computation of the ℓ_1-diameter of a finite set of points in $\mathbb{R}^k$. The main part of this lecture concerns Bourgain's theorem for embedding an arbitrary finite metric space into an ℓ_p space. In particular, this theorem states that every metric space can be embedded into an ℓ_p space of dimension $O(\log^2 n)$ with distortion $O(\log n)$. Our proof of the theorem directly implies an efficient randomized algorithm for computing such an embedding. Bourgain's theorem is of particular interest since it provides a way to represent an arbitrary metric space with only a logarithmic loss in distortion, into a nice normed space (e.g. Euclidean) which is well-understood and enjoys desirable analytic and algorithmic properties.

1 Preliminaries - terminology - notational conventions

All of the metric spaces encountered here are of finite dimension. Throughout this lecture, unless stated otherwise, the term *metric space* refers to a *finite metric space* (a metric space with finite number of points). In general, (X, d) denotes a metric space on the points of X equipped with metric d. In the previous lecture we defined the norm ℓ_p on $\mathbb{R}^k$, and to emphasis the dependency on k we denoted this by ℓ_p^k. In general ℓ_p is defined on a subset of the set of sequences $x = \{x_i\}_{i=1}^{\infty}$ of real numbers:

Definition 1.1. For $p \in [1, \infty)$, the normed space ℓ_p is defined on the sequences $x = \{x_i\}_{i=1}^{\infty}$ for which the sum $\sum_{i=1}^{\infty} |x_i|^p$ converges. For $x \in \ell_p$, the ℓ_p norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}.$$

On the other hand the normed space ℓ_∞ is defined on the sequences $x = \{x_i\}_{i=1}^{\infty}$ for

* Lecture Notes for a course given by Avner Magen, Dept. of Computer Sciecne, University of Toronto.

which $\sup_{i\geq 1}|x_i| < \infty$. For $x \in \ell_\infty$, the ℓ_∞ norm of x is defined as

$$\|x\|_\infty = \sup_{i\geq 1}|x_i|.$$

Note that $\ell_p^k \subset \ell_p$, as the elements of $\mathbb{R}^k$ can be identified with the sequences $\{x_i\}_{i=1}^\infty$ with $x_i = 0$ for $i > k$. Since both ℓ_p and ℓ_p^k induce the same norm on $\mathbb{R}^k$, we may write ℓ_p instead of ℓ_p^k when k is clear from the context. As we discussed in the previous lecture every norm induces a metric space. We use the term the *metric induced by* ℓ_p to refer to the standard metric induced by the ℓ_p norm; that is, for every $x, y \in \ell_p$, $d(x,y) = \|x-y\|_p$. Usually, we are interested in embedding a finite metric space into an infinite metric space with small distortion.

We say that a finite metric space (X,d) embeds into (Y,d') if (X,d) embeds isometrically into (Y,d'); and we write $(X,d) \overset{1}{\hookrightarrow} (Y,d')$ to denote this. In case of non-isometric embeddings we say that (X,d) is ρ-embeddable in (Y,d') meaning that there exists a ρ-embedding (i.e. an embedding with distortion ρ) of (X,d) into (Y,d') and we write $(X,d) \overset{\rho}{\hookrightarrow} (Y,d')$.

For a vector (point) $x \in \mathbb{R}^k$ we denote by x_i, $1 \leq i \leq k$ its i-th coordinate. For vectors in $\mathbb{R}^k$ we use e_j, $j = 1,\ldots,k$ to denote the vectors of the standard orthonormal basis; i.e. e_j is the vector with all coordinates are 0 except the j-th coordinate which is equal to 1.

Regarding the embeddings into $\mathbb{R}^k$ sometimes we construct each of the k coordinates of a point separately. We use $\oplus$ to denote the concatenation of the computed components of a point $x \in \mathbb{R}^k$. Let $f(i)$ be the value of the i-th coordinate of x. Then, $x = \oplus_{1\leq i\leq k} f(i) = \sum_{i=1}^k f(i)e_i$ (see section 3 for an example).

All logarithms are in base 2.

2 Application of metric embeddings in the computation of the ℓ_1-diameter

The ℓ_∞ norm is said to be the "universal ℓ_p-norm". In the previous lecture we saw that an arbitrary metric space of n points can be (isometrically) embedded into ℓ_∞^n (or into ℓ_∞^{n-1} if we are a bit more careful). Here we will use an embedding into ℓ_∞ which has only a loose resemblance with the "universality property" of ℓ_∞. The properties of ℓ_∞ (together with the properties of ℓ_1) allow us to embed a set of points in ℓ_1^k into $\ell_\infty^{2^k}$ such that this embedding enjoys many desirable algorithmic properties. Note that this is different from what we did in the previous lecture - e.g. the embedding will be into a space of dimension 2^k.

Given a finite set of points P in ℓ_p we define the diameter of P (w.r.t. ℓ_p) as

$$diam_p(P) = \max_{x,y\in P}\|x-y\|_p.$$

Given a finite set of n points $P \subset \ell_1^k$ the naive way of computing the diameter is by checking the ℓ_1 distance of every pair of the $\binom{n}{2}$ points in P. That is, we can compute the diameter with $O(kn^2)$ operations. Can we do any better? We show that we can compute the diameter using only a number of operations which is linear in n; which is of particular interest when k is small (e.g. a constant). We do this by first embedding ℓ_1 to ℓ_∞. Moreover, we will see that this embedding has other desirable algorithmic consequences.

Theorem 2.1. *If (P, d) is an ℓ_1^k metric space then we can (isometrically) embed it into $\ell_\infty^{2^k}$.*

Proof. The proof of this theorem is based on the observation that for every $x \in \mathbb{R}^k$, $\|x\|_1 = \max_{y \in \{-1,1\}^k} \langle x, y \rangle$. This is because $\|x\|_1 = \sum_{i=1}^k |x_i| = \sum_{i=1}^k sign(x_i)x_i = \max_{y \in \{-1,1\}^k} \langle x, y \rangle$, where for $a \neq 0$, $sign(a) = \frac{|a|}{a}$ and when $a = 0$, $sign(a) = 0$. By definition of ℓ_∞,

$$f(x) = \oplus_{y \in \{-1,1\}^k} \langle x, y \rangle \tag{1}$$

is an isometric embedding. □

We apply this proposition to compute the diameter in ℓ_1. We observe that we can compute the diameter of a finite metric space $P \subset \ell_\infty^{k'}$

$$\begin{aligned} diam_\infty(P) &= \max_{x,y \in P} \|x - y\|_\infty = \max_{x,y \in P} \max_{1 \leq j \leq k'} |x_j - y_j| \\ &= \max_j \max_{x,y} |x_j - y_j| = \max_j (\max_x x_j - \min_y y_j) \end{aligned}$$

That is, we can compute the diameter of $\ell_\infty^{k'}$ in time $O(k'n)$.

Therefore, the algorithm is as follows: (i) embed P into $\ell_\infty^{k'}$, where $k' = 2^k$ and then (ii) compute the diameter in $\ell_\infty^{k'}$. The time required for this computation is $O(nk2^k + n2^k) = O(nk2^k)$ which for constant k is $O(n)$.

Remark 2.2. Observe that the embedding of x (Eq. (1)) does not depend on the other points in P. This is important in applications e.g. on-line data-structures, databases etc. Contrast this "oblivious" embedding with the embedding of any metric space into ℓ_∞ (previous lecture) or with the embedding of the Bourgain's theorem (Section 4).

3 Fréchet embeddings

Let (X, d) be a finite metric space. Say that $S_1, S_2, \ldots, S_r$ are subsets of X. We abuse notation and we write $d(x, S)$ to denote $d(x, S) = \min_{y \in S} d(x, y)$. A Fréchet embedding f is of the following form:

$$f(x) = \oplus_{1 \leq i \leq r} \alpha_{S_i} d(x, S_i)$$

Usually α_{S_i}'s will be equal to 1. Here is an example of such an embedding (all α_{S_i}'s are 1). Consider the graph in Figure 1. Say that $S_1 = \{1, 2, 3\}, S_2 = \{4, 5\}$ and $S_3 = \{7, 6\}$. Then $f(1) = (0, 3, 0), f(2) = (0, 2, 1), f(3) = (0, 1, 3), f(4) = (1, 0, 2), f(5) = (2, 0, 1), f(6) = (2, 1, 0)$ and $f(7) = (1, 2, 0)$.

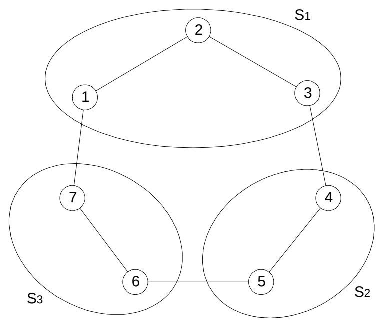

Figure 1: A cycle equipped with the shortest-path metric where each edge is of weight 1.

Lemma 3.1. *Let (X, d) be a metric space on n points. Consider the Fréchet embedding f of (X, d) into ℓ_1^r, for some sets $S_1, S_2, \ldots, S_r$ which correspond to the coordinates of the value of f. Then, $\|f(x) - f(y)\|_1 \leq rd(x, y)$.*

Proof. We wish to show that for every $S \subseteq X$, $|d(x, S) - d(y, S)| \leq d(x, y)$. Let $d(y, S) = d(y, w)$ for some $w \in S$ (by definition of $d(y, S)$ such a w exists). Also, by definition for every $w \in S$, $d(x, S) \leq d(x, w)$. Therefore, $d(x, S) - d(y, S) \leq d(x, w) - d(y, w) \leq d(x, y)$ where the last inequality follows by the triangular inequality. Let

$$f(x) = (d(x, S_1), d(x, S_2), \ldots, d(x, S_r)),$$

and

$$f(y) = (d(y, S_1), d(y, S_2), \ldots, d(y, S_r)).$$

Then, the expansion of the embedding is at most $rd(x, y)$ since $\|f(x) - f(y)\|_1 = \sum_i^r |d(x, S_i) - d(y, S_i)| \leq rd(x, y)$. ☐

4 Bourgain's theorem

We want to show the existence of an embedding with distortion $O(\log n)$ to an $\ell_p^{O(\log^2 n)}$ space for every p. The proof of theorem 4.1 implies a randomized algorithm for efficiently computing such an embedding.

Theorem 4.1 (Bourgain's theorem). *Let (X, d) be a metric space on n points. Then,*

$$(X, d) \overset{O(\log n)}{\hookrightarrow} \ell_p^{O(\log^2 n)}$$

Theorem 4.1 appears in [Bou85]. Linial et al. [LLR95] provide algorithmic applications and an explicit lower bound showing that the theorem is tight[1] (for ℓ_2). Also,

[1] The original bound on the dimension was exponential in n. In [LLR95] the dimension was reduced to $O(\log^2 n)$. Specifically for the euclidean space we can reduce the dimension to $O(\log n)$ (by applying the JL-flattening lemma) as we will see in some upcoming lecture.

Matousek [Mat96] shows that a stronger version of theorem 4.1. In particular,

$$(X,d) \overset{O(\frac{\log n}{p})}{\hookrightarrow} \ell_p^{O(\log^2 n)}$$

Note that this dependence on p also refers to the universality of ℓ_∞.

Before getting to the proof of Bourgain's theorem let us state without a proof a similar theorem for the special case of embeddings into ℓ_∞. This theorem is due to Matousek (see [Mat96] or [Mat02] pp. 386-388).

Theorem 4.2. *Let (X,d) be a metric space on n points. Then, $(X,d) \overset{D}{\hookrightarrow} \ell_\infty^{O(Dn^{2/D}\log n)}$.*

For the special case of ℓ_∞ this theorem generalizes Theorem 4.1, since for $D = \Theta(\log n)$ we get that $(X,d) \overset{O(\log n)}{\hookrightarrow} \ell_\infty^{O(\log^2 n)}$). Note that theorem 4.2 provides a trade-off between distortion D and and the dimension of the host ℓ_∞ space. It is easy to see that for every $x \in \mathbb{R}^m$, $\|x\|_\infty \le \|x\|_p \le m\|x\|_\infty$. Therefore, a corollary of Theorem 4.2 is that

$$(X,d) \overset{O(\log^3 n)}{\hookrightarrow} \ell_p^{O(\log^2 n)}$$

(which is weaker than Bourgain's theorem). Now, we get to the proof of Bourgain's theorem.

Intuition behind the proof of theorem 4.1 In Fréchet embeddings for each coordinate of the vectors we measure the distance of a point to a set. In Bourgain's theorem we will use Fréchet embeddings where the corresponding A_{ij} sets are constructed randomly by sampling independently the metric space with different probabilities 2^{-j}, $j = 1, 2, \ldots, \lceil \log n \rceil$) for many rounds $i = 1 \ldots, \Theta(\log n)$. Then, we will show that with positive probability there exists an embedding which satisfies the requirements of the theorem. Clearly the same embedding must "work well" for the distance of every pair of points in the metric space. Hence, the reason why we use different probabilities (to sample points) has to do with the "structure" of the metric space. Also, for the same probability (used to independently sample elements from the metric space) we construct several sets. All these will become clear by getting into the details of the proof. In Figure 2 we give an intuitive example. Although, we use the plane to somehow refer to the notion of distance, keep in mind that the metric space is not (necessarily) Euclidean and drawing on the plane is done just for the sake of this intuitive demonstration. Before getting to the proof let us give some more intuition regarding why we need the two extreme cases, where the sampling probability is $1/2$ and $1/n$. Consider two points x, y to be far apart in the line. In one extreme we choose elements independently with probability $\frac{1}{2}$. In this case with high probability A_{i1} will contain points close both to x and to y ("no matter" how many times we will sample with the same sampling frequency). Therefore, we expect $|d(x, A_{ij}) - d(y, A_{ij})|$ to hardly contribute to $\|f(x) - f(y)\|_1$ (where f is the Fréchet embedding we are talking about) - actually in the example in the figure the contribution is zero. In the other extreme the probability is $1/n$. In this case (if we sample with the same frequency for a sufficient number of times) with high probability we will have few points in A_{ij} which are close to x (or to y - but not both). In this case $|d(x, A_{ij}) - d(y, A_{ij})|$ is going to be close to $d(x, y)$.

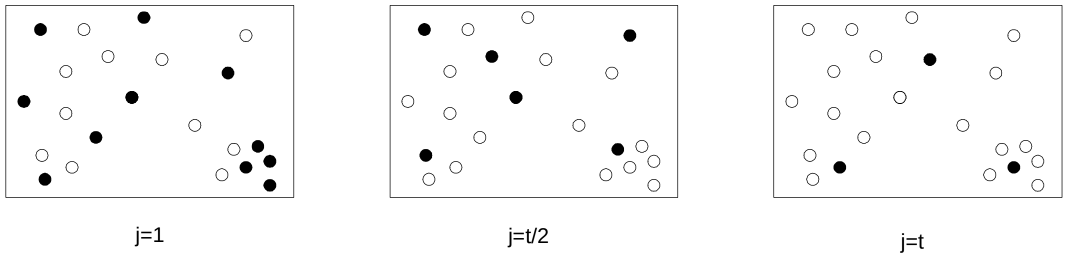

Figure 2: The black dots correspond to elements that are picked (randomly) and placed in A_{ij} when sampling for particular values of j.

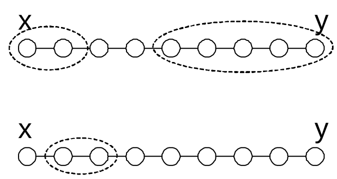

Figure 3: The interior of the dotted closed curves shows the elements that are chosen in A_{ij}.Top: sampling with probability $1/2$. Bottom: sampling with probability $1/n$.

Proof of Theorem 4.1 - Bourgain's theorem. We prove the theorem for embedding into ℓ_1 (i.e. $p = 1$). At the end of the proof we will show how to obtain the theorem for every p.

We probabilistically construct a Fréchet embedding. We will show that in the (finite) probability space there exists with positive probability an embedding that satisfies the statement of the theorem.

Given (X, d) construct $O(\log^2 n)$ sets A_{ij} as follows: For every $j = 1, \ldots, \lceil \log n \rceil$ ($t = \lceil \log n \rceil$) construct $i = 1, \ldots, 144\lceil \log n \rceil$ ($m = 144\lceil \log n \rceil$) sets[2]. Each A_{ij} set is constructed by choosing independently at random for every elements of X with probability 2^{-j}. The embedding f for every point $x \in X$ is:

$$f(x) = \oplus_{\substack{1 \leq j \leq m \\ 1 \leq i \leq t}} d(x, A_{ij}).$$

Algorithmic implication of the proof: An algorithm that computes the construction of the embedding can be implemented to work in polynomial time. Since this embedding exists with constant probability (over all embeddings constructed in this probabilistic way) we can make the probability that the desired embedding is not constructed exponentially small, by independently repeating the randomized construction for a polynomial number of times.

[2]The constants in t and m can be improved. Moreover, working through the details of the proof we observe that there is a trade-off between these constants in the dimension and the distortion of the embedding; and we can tune this trade-off.

For every $x, y \in X$ we are going to split the distance $d(x, y)$: we will find a "desirable" set of numbers $\Delta_1, \Delta_2, \ldots, \Delta_t \geq 0$ with $\sum_j \Delta_j = \frac{d(x,y)}{4}$. For such an splitting we call a set A_{ij} *good* when $|d(x, A_{ij}) - d(y, A_{ij})| \geq \Delta_j$. Our goal is to have many good sets.

Lemma 4.3 (Main lemma). *For each $1 \leq j \leq t$ and for each $x, y \in X$ there exists a set of numbers $\Delta_1, \Delta_2, \ldots, \Delta_t \geq 0$ with $\sum_j \Delta_j = \frac{d(x,y)}{4}$ such that with probability greater than $1 - \frac{1}{n^3}$, a constant fraction $c = \frac{1}{20}$ of sets A_{ij} are good.*

We defer the proof of the above lemma, and first finish the proof of the theorem (using the lemma). Let $\mathcal{E}_j^{xy}$ be the event stated in the lemma. By this lemma $Pr(\overline{\mathcal{E}_j^{xy}}) \leq 1/n^3$. Therefore, (by the union-bound)

$$Pr(\bigcup_{j,x,y} \overline{\mathcal{E}_j^{xy}}) \leq \sum_{\substack{j \\ x,y}} Pr(\overline{\mathcal{E}_j^{xy}}) = \frac{\frac{n(n-1)}{2} \log n}{n^3} < 1$$

Hence the negation of the above event occurs with positive probability which implies that there exists an embedding such that:

$$\begin{aligned} \frac{cm}{4} d(x,y) &= \sum_j cm\Delta_j \leq \sum_{i,j} |d(x, A_{ij}) - d(y, A_{ij})| \\ &= \|f(x) - f(y)\|_1 \leq mtd(x,y) \end{aligned}$$

where the last inequality follows from Lemma 3.1. This means that the distortion of the embedding is at most $(mt) \,/\, \left(\frac{cm}{4}\right) = \frac{4t}{c} = O(\log n)$.

Remark 4.4. Note that for example, $\frac{\frac{n(n-1)}{2} \log n}{n^3} < \frac{1}{10}$ for every $n > 20$. This implies that at least 90% of the embeddings have distortion $O(\log n)$. We can amplify the probability of success by repeating the experiment and after a polynomial number of repetitions we get a probability of failure which is exponentially small. That is, we have a randomized algorithm for this problem.

Proof of Lemma 4.3 - Main lemma. The idea behind the proof is to make use of a "general" argument regarding the distribution of points of a randomly constructed set and the points in the metric space. For this we introduce the notion of balls around the points (in this not-necessarily euclidean space). Our goal is to show for every two points $x, y \in X$ that both probabilities of the randomly constructed set to intersect the ball around x and to not intersect the ball around y is bounded away from zero and one.

For a point $x \in X$ we define the (closed) ball of radius $r \geq 0$ as $B(x, r) = \{y | d(x, y) \leq r\}$ and the open ball of radius $r \geq 0$ as $B^o(x, r) = \{y | d(x, y) < r\}$. Note that for a set $S \subseteq X$, when the ball around x intersects S, the radius of the ball is larger than the distance of x from S.

For a pair x, y let r_j be the minimum r such that $|B(x, r)|, |B(y, r)| \geq 2^j$. We show the existence of the splitting of the distance (as mentioned above) by defining $\Delta_j = r_j - r_{j-1}$. By definition the sum $\sum_j \Delta_j$ telescopes and thus $\sum_j \Delta_j = r_t$ and we may define r_t as $\frac{d(x,y)}{4}$ (without having the two balls intersecting each other).

Without loss of generality assume that $|B^o(y, r_j)| < 2^j$ and of course we have that $|B(x, r_{j-1})| \geq 2^{j-1}$. Intuitively, this since $B^o(y, r_j)$ is not very large with a constant probability A_{ij} does not intersect it and also since $B(x, r_{j-1})$ is not very small with a constant probability A_{ij} intersects it. More formally:

To avoid dealing with dependencies we bound from below the probability λ of the event that $d(x, A_{ij}) - d(y, A_{ij}) \geq \Delta_j$ (we denote this event by $\mathcal{E}_{ij}^{xy}$) by the easier to compute probability of the following event[3]:

$$Pr\big(A_{ij} \cap B^o(y, r_j) = \emptyset \text{ and } A_{ij} \cap B(x, r_{j-1}) \neq \emptyset\big) = constant > 0$$

Let us precisely compute this probability:

$$\begin{aligned}\lambda := \Pr[\mathcal{E}_{ij}^{xy}] &\geq Pr\big(A_{ij} \cap B^o(y, r_j) = \emptyset \text{ and } A_{ij} \cap B(x, r_{j-1}) \neq \emptyset\big) \\ &= Pr\big(A_{ij} \cap B^o(y, r_j) = \emptyset\big) Pr\big(A_{ij} \cap B(x, r_{j-1}) \neq \emptyset\big),\end{aligned}$$

since the two events are independent[4]. It suffices to lower bound the two probabilities such that their product is bounded below by a constant (i.e. away from zero and one).

$$Pr\big(A_{ij} \cap B^o(y, r_j) = \emptyset\big) = (1 - 2^{-j})^{|B^o(y, r_j)|} > (1 - 2^{-j})^{2^j} \geq \frac{1}{4}.$$

Similarly, we compute $Pr\big(A_{ij} \cap B(x, r_{j-1}) = \emptyset\big) \geq \frac{1}{4}$. Therefore, $\lambda \geq \frac{1}{16}$.

We complete the proof by applying Chernoff bound [Che52] in order to show that for each x, y there exists a constant fraction of A_{ij} sets which are good with high probability. Here is the version of Chernoff bound we will use.

Theorem 4.5. *Let $X_1, X_2, \ldots, X_m$ be binary random variables, with $E(X_i) \geq \mu$, $1 \leq i \leq m$. Let $X = \sum_i X_i$. Then,*

$$Pr(X \leq (1 - \epsilon)\mu m) \leq e^{-\epsilon^2 \mu m/3}$$

We define the indicator variable X_i to be 1 iff $\mathcal{E}_{ij}^{xy}$ happens. Therefore, $E(X_i) = Pr(X_i = 1) = \lambda \geq \frac{1}{16}$. Therefore, $Pr(X \leq (1 - \epsilon)\frac{1}{16}m) \leq e^{-\epsilon^2 m/48}$. We wish to show that for a fixed j the good event happens for a constant fraction $(1 - \epsilon)$ of the m, A_{ij} sets with probability at least $1 - 1/n^3$. For this we require:

$$\frac{1}{e^{\epsilon^2 144 \log n/48}} < \frac{1}{n^3} \Leftrightarrow \frac{\epsilon^2 144 \log n}{48} > 3\frac{\log n}{\log e} \Leftrightarrow \epsilon > \sqrt{\frac{1}{\log e}},$$

and thus there exists such a constant $0 < \epsilon < 1$. In particular, we choose $\epsilon = 1/5$ and thus $(1 - 1/5)/16 = \frac{1}{20}$ of the m A_{ij} sets are good with probability greater than $1 - 1/n^3$. □

[3] It is clear that first event is a superset of the event for which we compute the probability.

[4] The two events are independent since the two balls do not intersect and in the definition of the probability space we choose elements in A_{ij} independently; that is why in the constructed product probability space the two events are independent.

We proved the theorem only for the case of ℓ_1. Now we extend it to any ℓ_p. Surprisingly, the same embedding f works for every ℓ_p. Let $K = 144\lfloor \log n \rfloor^2$ be the number of coordinates. First note that

$$\begin{aligned} \|f(x) - f(y)\|_p &= (\sum_{i,j} |d(x, A_{ij}) - d(y, A_{ij})|^p)^{1/p} \\ &\leq (K d(x,y)^p)^{1/p} = K^{1/p} d(x,y), \end{aligned} \tag{2}$$

by Lemma 3.1. Next we need to lower bound $\|f(x) - f(y)\|_p$. This can be done by Hölder's Inequality:

Theorem 4.6 (Hölder's Inequality). *For $x, y \in \mathbb{R}^k$ and $\frac{1}{p} + \frac{1}{q}$ we have*

$$\sum |x_i y_i| \leq \|x\|_p \|y\|_q.$$

By Hölder's inequality we obtain

$$\begin{aligned} \|f(x) - f(y)\|_p K^{1/q} &\geq \sum_{i,j} |d(x, A_{ij}) - d(y, A_{ij})| \\ &= \|f(x) - f(y)\|_1 \geq \Theta(\log n) d(x,y), \end{aligned} \tag{3}$$

where in the last inequality we used Bourgain's theorem for ℓ_1 which we have already proved. Now combining (2) and (3) gives

$$K^{-1/q} \Theta(\log n) d(x,y) \leq \|f(x) - f(y)\|_p \leq K^{1/p} d(x,y)$$

which shows that the distortion is bounded by

$$K^{1/p} K^{1/q} / \Theta(\log n) = O(\log n).$$

□

References

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *COMBINAT: Combinatorica*, 15, 1995.

[Mat96] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.