CS-67720 Metric Embedding Theory and Its Algorithmic Applications

Lecture 12

Lecturers: Yair Bartal, Nova Fandina *Fall 2018/19*
For comments contact fandina@gmail.com

12.1 Coarse Scaling Embedding into ℓ_p Spaces

In this lecture we show that the Uniform Padding Lemma (for which we worked hard in the last lecture) implies (actually, it was exactly designed for this purpose) the coarse scaling version of Bourgain's embedding, with improved dimension. Precisely, we will prove the following theorem.

Theorem 12.1 ([1]). *For any n-point metric space X there exists an embedding $f : X \to \ell_p^{O(\log n)}$ with coarse scaling distortion $O(\log 1/\epsilon)$.*

The celebrated result of Bourgain states that any n-point metric space embeds into Euclidean space with $O(\log n)$ distortion. In [3] the authors have proved that Bourgain's embedding provides an embedding into ℓ_p with distortion $O(\log n)$, and dimension $O(\log^2 n)$. At last, Theorem 12.1 implies an embedding with average distortion $O(1)$ (check this!), and with improved dimension $O(\log n)$, while the worst case guarantee is still $O(\log n)$.

Proof of Theorem 12.1 (Worst Case Guarantee). First, we present the construction for the non-scaling version of the embedding. In the next subsection, we show how to update it to satisfy the coarse-scaling distortion guarantee.

Let X be an n-point metric space, and let $p \geq 1$. Let $\mathcal{H}$ denote the special bundle of $\Delta_i = (diam(X)/k^i)$-bounded probabilistic partitions $\mathcal{P}_i$, constructed by the uniform partition lemma (Theorem 11.1). Recall that by the lemma, each partition $\mathcal{P}_i$ satisfies the following: for all $x \in X$, for all $0 < \delta \leq 1/2$, $Pr_{\mathcal{P}}\left[B\left(x, \eta^{(\delta)}_{P(x)} \cdot \Delta_i\right) \nsubseteq P(x)\right] \leq \delta$, where $\eta^{(\delta)}_{P(x)} = \frac{\delta}{128 \cdot \log(\hat{\rho}(v_{P(x)}, \Delta_i))}$, for $\hat{\rho}(v_{P(x)}, \Delta_i) = \max\{\rho(v_{P(x)}, \Delta_i), 2\}$. Particularly, for $\delta = 1/2$, the above probability is bounded by $1/2$, and $\eta^{(\delta)}_{P(x)} = c \cdot \log(\hat{\rho}(v_{P(x)}, \Delta_i))$, for some constant $c > 0$. In what follows we set $\delta = 1/2$ and omit it from notation.

For all $x \in X$, the embedding $f(x)$ is defined by $f(x) = \frac{1}{D^{1/p}} \cdot \left(\oplus_{1 \leq t \leq D} f^{(t)}(x)\right)$, where $\oplus$ is the concatenation of a randomly and independently generated D coordinates, denoted by $f^{(t)}(x) \in \mathbb{R}$. To compute t-th coordinate for all points $x \in X$, let P_i be Δ_i-bounded partition, chosen according to $\mathcal{P}_i$, for each $\mathcal{P}_i \in \mathcal{H}$. For all clusters $C \in P_i$, let $\sigma_i(C) = 1$ with probability $1/2$, and 0, otherwise; and let $\xi_i(C) = 1$ if $\rho(v_C, \Delta_i) \geq 2$, and 0 otherwise. Note that both σ_i and ξ_i are uniform functions over a cluster. Then, for all $x \in X$, the i-th component of $f^{(t)}(x)$ is defined by: $f^{(t)}_i(x) = \sigma_i(P_i(x)) \cdot \xi_i(P_i(x)) \cdot \frac{1}{\eta_{P_i(x)}} \cdot d(x, X/P_i(x))$, where $d(x, X/P_i(x))$ is the distance from x to the outside of its cluster in partition P_i. Then, we define $f^{(t)}(x) = \sum_i f^{(t)}_i(x)$.

Note, that the dimension of the embedding f is D, which will be chosen later on (it will be set to $O(\log n)$).

We show that the following claim implies the existence of an embedding $f : X \to \ell_p$, with distortion $O(\log n)$ (yet for a coarse-scaling distortion we work a bit harder later on).

Claim 12.2. *The following two assertions hold.*

1. *$\forall\, x \neq y \in X$, with probability 1, $\|f(x) - f(y)\|_p \leq O(\log n) \cdot d(x, y)$.*
2. *$\forall\, x \neq y \in X$, for all $1 \leq t \leq D$, $Pr\left[\left|f^{(t)}(x) - f^{(t)}(y)\right| \geq \frac{1}{32} d(x, y)\right] \geq \frac{1}{8}$.*

The first assertion implies the bound on the expansion of the embedding. Let us show that the second assertion implies an existence (with positive probability) of the embedding with constant contraction for any pair $x, y \in X$.

Assume that (2) is correct. For all pairs $x \neq y \in X$, for all $1 \leq t \leq D$, consider the following indicator variable: $Z^{(t)}_{(x,y)} = 1$ iff $|f^{(t)}(x) - f^{(t)}(y)| \geq \frac{1}{32} d(x, y)$ (i.e., this variable indicates the good event of a large contribution of the t-th coordinate to $\|f(x) - f(y)\|_p$). Let $Z_{(x,y)} = \sum_{t=1}^{D} Z^{(t)}_{(x,y)}$. Therefore, $E[Z_{(x,y)}] \geq D/8$, implying

$$Pr\left[Z_{(x,y)} < D/16\right] \leq Pr\left[Z_{(x,y)} < \frac{1}{2} E[Z_{(x,y)}]\right] \leq^{Chernoff} \leq e^{-\frac{D}{8} \cdot \frac{1}{8}} \leq^{for\ D \geq 128 \ln n} \leq \frac{1}{n^2}.$$

Therefore, by the union bound, $Pr[\exists\ x \neq y \in X,\ \ s.t.\ Z_{(x,y)} < D/16] < 1/2$. Thus, with probability at least $1/2$, the embedding f is such that for all $x \neq y \in X$ the number of "good coordinates" is at least $D/16$. Therefore, there exists an embedding $f : X \to \ell_p^{O(\log n)}$, such that for all $x \neq y \in X$,

$$\|f(x) - f(y)\|_p^p = \frac{1}{D} \sum_{t=1}^{D} |f^{(t)}(x) - f^{(t)}(y)|^p \geq (1/16)(1/32 \cdot d(x, y))^p.$$

It remains to prove Claim 12.2.

Proof of Claim 12.2. **Proof of** (1)**.** We start with a short and simple lemma.

Lemma 12.3. *For all $x \in X$, for all coordinate $1 \leq t \leq D$, for each component i, let $A_{P_i(x)} = \sigma_i(P_i(x)) \cdot \xi_i(P_i(x)) \cdot \frac{1}{\eta_{P_i(x)}}$. Then, for all $x \neq y \in X$, for all coordinate t, for all component i, for all random choices over $\mathcal{P}_i$ it holds that*

$$f_i^{(t)}(x) - f_i^{(t)}(y) \leq A_{P_i(x)} \cdot d(x, y).$$

Proof. Case 1: $P_i(x) \neq P_i(y)$. In such case, $d(x, X/P_i(x)) \leq d(x, y)$, implying $f_i^{(t)}(x) \leq A_{P_i(x)} d(x, y)$. Since $f_i^{(t)}(y) \geq 0$, we get $f_i^{(t)}(x) - f_i^{(t)}(y) \leq A_{P_i(x)} \cdot d(x, y)$, as required.

Case 2: $P_i(x) = P_i(y)$. In this case we have $f_i^{(t)}(x) - f_i^{(t)}(y) = A_{P_i(x)} \cdot (d(x, X/P_i(x)) - d(y, X/P_i(x))) \leq A_{P_i(x)} d(x, y)$, since $d(x, X/P_i(x)) \leq d(x, y') \leq d(x, y) + d(y, X/P_i(x))$, where y' is the closest point to y from the outside of $P_i(x)$.

□

It follows from the lemma, that given any $x \neq y \in X$, for any coordinate t it holds

$$\frac{f^{(t)}(x) - f^{(t)}(y)}{d(x,y)} = \frac{\sum_{i=1}^{\log \Phi} f_i^{(t)}(x) - f_i^{(t)}(y)}{d(x,y)} \leq \sum_{1 \leq i \leq \log \Phi} A_{P_i(x)} \leq \sum_{1 \leq i \leq \log \Phi} \frac{\xi_i(P_i(x))}{\eta_{P_i(x)}} =$$

$$= \sum_{i,\ s.t.\ \rho(v_{P_i(x)},\Delta_i) \geq 2} \frac{1}{\eta_{P_i(x)}} \leq \sum_{i,\ s.t.\ \rho(v_{P_i(x)},\Delta_i) \geq 2} 256 \log(\rho(x, \Delta_i)) \leq \sum_{1 \leq i \leq \log \Phi} \gamma_i(x) = O(\log n),$$

where the last step we computed in the previous lecture. Therefore,

$$\frac{\|f(x) - f(y)\|_p^p}{d(x,y)^p} = \frac{1}{D} \cdot \frac{\sum_{t=1}^{D} |f^{(t)}(x) - f^{(t)}(y)|^p}{d(x,y)^p} \leq O(\log^p n).$$

Proof of (2)**.** For any $x \in X$, for each i, by the padding property of the partition $\mathcal{P}_i$, it holds that $Pr[B(x, \eta_{P_i(x)}\Delta_i) \subseteq P_i(x)] \geq 1/2$, which implies that $Pr[d(x, X/P_i(x)) \geq \eta_{P_i(x)\Delta_i}] \geq 1/2$.

For a given pair $x \neq y \in X$, consider the index i such that $d(x,y)/16 \leq \Delta_i < d(x,y)$. For such i it holds that either $\xi_i(P_i(x)) = 1$ or $\xi_i(P_i(y)) = 1$, by Lemma 11.3. We just have to ensure that $1/32\Delta_i \leq d(v_{P_i(x)}, v_{P_i(y)}) \leq 63\Delta_i$. This is surely correct, since for the i we consider, it holds that x and y are in different clusters, and the radii of the clusters are chosen to be at least $\Delta_i/4$, and at most $\Delta_i/2$.

Thus, assume w.l.g. that $\xi_i(P_i(x)) = 1$. Define $R = |\sum_{j \neq i} f_j^{(t)}(x) - f_j^{(t)}(y)|$.
If $R \leq \Delta_i/2$, then

$$Pr\left[d(x, X/P_i(x)) \geq \eta_{P_i(x)}\Delta_i,\ and\ \sigma_i(P_i(x)) = 1,\ and\ \sigma_i(P_i(y)) = 0\right] \geq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}.$$

Meaning that with probability at least 1/8 it holds that $f_i^{(t)}(x) \geq \Delta_i$, and $f_i^{(t)}(y) = 0$, namely

$$|f^{(t)}(x) - f^{(t)}(y)| \geq |f_i^{(t)}(x) - f_i^{(t)}(y)| - R \geq \Delta_i - R \geq \Delta_i/2 \geq \frac{d(x,y)}{32}.$$

If $R^{(t)} \geq \Delta_i/2$, then $Pr[\sigma_i(P_i(x)) = \sigma_i(P_i(y)) = 0] = \frac{1}{4}$, meaning that with probability $\frac{1}{4}$ it holds that $f_i^{(t)}(x) = f_i^{(t)}(y) = 0$, resulting in

$$|f^{(t)}(x) - f^{(t)}(y)| = R \geq \Delta_i/2 \geq \frac{d(x,y)}{32}.$$

Altogether, we have that with probability at least 1/8 the t-th coordinate of the embedding is at least $\frac{d(x,y)}{32}$, as stated in the claim. □

This completes the proof of the non-scaling version of the theorem. □

12.1.1 Coarse Scaling Version

For the coarse-scaling version of the embedding, we should prove the following: for all $x \neq y \in X$ that are $(\epsilon/2)$- far from each other it holds that $dist_f(x, y) \leq O\left(\log \frac{1}{\epsilon}\right)$. Given any such x and y, let $l \geq 1$ be the maximal index such that $\Delta_l \geq 16d(x, y)$. For such l we have: for all coordinate t, for all random choices over $\mathcal{P}_i$

$$\frac{\sum_{i<l} f^{(t)}(x) - f^{(t)}(y)}{d(x, y)} \leq \sum_{i<l} \gamma_i(x) = O\left(\log 1/\epsilon\right),$$

where the first inequality we showed to be correct in the previous proof, and the second inequality we showed to be correct in the last lecture, for all $\epsilon/2$-far points. Thus, it is enough to prove that $\sum_{i \geq l} f^{(t)}(x) - f^{(t)}(y) \leq O(1) \cdot d(x, y)$. For this we slightly change the definition of the embedding.

Proof of Theorem 12.1. We redefine the i-th component of the t-th coordinate of the embedding as follows

$$f_i^{(t)}(x) = \min\{A_{P_i(x)} \cdot d(x, X/P_i(x)), \Delta_i\}.$$

Therefore, we have

$$\sum_{i \geq l} f_i^{(t)}(x) - f_i^{(t)}(x) \leq \sum_{i \geq l} f_i^{(t)}(x) \leq \sum_{i \geq l} \Delta_i \leq 2\Delta_l \leq 2(16)^2 d(x, y).$$

We should confirm that all the considerations we made during the proof are still correct under the new definition. For the contraction arguments it is immediately, and for the upper bound (the only place to think is Lemma 12.3) it is enough to prove the following lemma:

Lemma 12.4. *For a Δ-bounded probabilistic partition $\mathcal{P}$, let $A_{P(x)}$ be a uniform function over a cluster $P(x)$. For any $x \in X$ and a function $f(x) = A_{P(x)} \cdot d(x, X/P(x))$, let $\hat{f}(x) = \min\{A_{P(x)} \cdot d(x, X/P(x)), \Delta\}$. Then, for all $x \neq y \in X$ it holds that $\hat{f}(x) - \hat{f}(y) \leq \min\{A_{P(x)} \cdot d(x, y), \Delta\}$.*

The proof is immediate, by considering all the cases. □

References

[1] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. STOC '06, pages 271–286, 2006.

[2] Naveen Garg, Vijay V. Vazirani, and Mihalis Yannakakis. Approximate max-flow min-(multi)cut theorems and their applications. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Theory of Computing*, STOC '93, pages 698–707, New York, NY, USA, 1993. ACM.

[3] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.