COMS W6998-5: Algorithms through Geometric Lens (Fall'17) Nov 8, 2017

Lecture 18 – Bourgain's Theorem for Metric Embeddings

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribes: *Jonathan Terry*

1 Introduction

This lecture serves as a continuation of the previous lecture, in which we introduced metric embeddings as a segue into an efficient algorithm for solving the Sparsest Cut Problem. The **Sparsest Cut Problem** attempts to partition a graph into two disjoint sets S and $\bar{S}$ such that the ratio of the number of edges shared to the product of their sizes is minimized. This optimization captures the qualitative notion that we are searching for a small yet balanced cut. Recall that the general plan of action for solving this problem was as follows:

$$G \xrightarrow{\text{LP}} \text{metric on vertices} \xrightarrow{l_1\text{-embedding}} l_1 \rightarrow \text{cut}$$

Last lecture, we defined the optimal solution to this problem as follows:

$$\min_{S\subset V} \frac{E(S,\bar{S})}{|S||\bar{S}|} = OPT$$

From there, we considered a relaxation of this NP-Hard problem in which we solved a Linear Program to obtain a **semi-metric** d^* such that the following optimization problem is solved:

$$\min_{d^*} \frac{\sum_{(u,v)\in E} d^*(u,v)}{\sum_{u,v} d^*(u,v)} = OPT^*$$

Given this new metric on vertices, the task at hand was was to prove the existence of an embedding with reasonable distortion taking $d^* \rightarrow (\mathbb{R}^d, l_1)$. What to be proven herein is **Bourgain's Theorem**, which provides guarantees the existence of such an embedding.

Theorem 1 (**Bourgain's Theorem ('85)**). *Given a metric space (X, d) where $|X| = n$, there exists a map $f : X \rightarrow (\mathbb{R}^d, l_1)$ such that the distortion $D = O(\log(n))$ and the embedded space is of dimension $d = O(\log(n))$.*

Definition 2 (**Distortion**). *Given two metric spaces (X, d_X) and (Y, d_Y) and a function $f : X \rightarrow Y$, the function is an embedding with distortion $D > 1$ if there is a $\lambda > 0$ such that for all $x_1, x_2 \in X$:*

$$\lambda d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq D d_X(x_1, x_2)$$

Observation 3. *Metric spaces are more general than normed spaces, so the definition immediately applies to metric embeddings where $d_Y = ||\cdot||_p$ with $p \geq 1$. For the purposes of this lecture, take $\lambda = 1$.*

2 Frechet Embeddings

The previous lecture ended with the introduction of Frechet embeddings, which are a crucial concept in the proof of Bourgain's Theorem.

Definition 4 (**Frechet Embeddings**). *Given a metric space* (X, d) *and some* $S \subset X$*, we define a function* $f_S : X \to \mathbb{R}$ *such that:*

$$f_S(x) = \text{dist}(x, S) = \min_{\tilde{x} \in S} d_X(x, \tilde{x})$$

Claim 5. *For any* $x_1, x_2 \in X$*, we always have* $|f_S(x_2) - f_S(x_1)| \leq d_X(x_1, x_2)$*.*

Observation 6. *Frechet embeddings have the nice property that a distance between points can only shrink. The downside is that for any two* $x_1, x_2 \in S$*, their distance becomes zero by definition. It is this issue that motivates the sampling procedure behind the proof of Bourgain's Theorem.*

With the definitions and crucial remarks out of the way, we can move on to how we actually use Frechet Embeddings to accomplish a low dimensional embedding with reasonable distortion. If we wish to use Frechet Embeddings, we must intelligently select some series of sets $S_{i,j}$ such that the distance between any two points will be somewhat preserved. In particular, we wish to select sets in a manner such that no two distinct points reside in exactly the same sets, as this would shrink their distance in the embedded space to nothing! The key to an effective construction lies in an intelligent selection of the various sets across scales. Upon selecting this series of sets, we then construct a mapping $f_S : X \to \mathbb{R}^d$ where each coordinate corresponds to distance from some point x to one of the sets $S_{i,j}$. As such, we can see that more sets reduces the probability of failure at the expense of a higher dimensional embedding space. Thus, the sample procedure for these sets S_{ij} needs to strike a sweet balance.

2.1 Intelligently Selecting Sets for Frechet Embeddings

The protocol for selecting sets from which we then construct the embedding lies in having a set of scales, and creating a certain number of sets at each scale. At each scale, we construct sets probabilistically, with the probability of set membership fixed by the scale we are operating at.

Definition 7 (**Sampling Procedure**). *Given an n-point metric space* (X, d)*, choose sets* S_{ij} *where* $1 \leq i \leq \log_2(n)$ *and* $1 \leq j \leq O(\log(n))$ *where for any point* $\Pr[x \in S_{ij}] = 2^{-i}$*. The selection is independent over* (x, i, j)*. If any constructed set* S_{ij} *is empty, discard it.*

Corollary 8 (**Sampling Procedure Distortion**). *By the construction of* S_{ij} *and claim about Frechet Embeddings, we can see that for any two* $x_1, x_2 \in X$*, we have the following relation:*

$$||f(x_2) - f(x_1)||_1 \leq O(\log^2(n)) d_X(x_1, x_2)$$

2.2 Guarantees of the Embedding

As with anything probabilistic, we hope to show that this scheme works almost all of the time: in other words, the distance between any two points in the embedded space is almost never 0. Before getting too deep into the math, the lemma which we will prove in this section is as follows:

Lemma 9. *For any two points in X and the map $f_S : X \to \mathbb{R}^{O(\log^2(n))}$, the following holds:*

$$\Pr[||f_S(x_2) - f_S(x_1)|| \geq \Omega(\log(n))d_X(x_1, x_2)] \geq 1 - \frac{1}{10n^2}$$

Before we can show this lemma, which effectively says that our procedure works well, we need to do some more legwork first. The first step in showing that for any two points, we want $|f_S(x_2) - f_S(x_1)| \geq \delta > 0$ for some certain δ. Consider two balls, $B(x_1, r)$ and $B(x_2, r + \delta)$. Intuitively, we want it such that for some S_{ij}, we hit one ball but not the other. Formally:

1. $S_{ij} \subset B(x_1, r)$
2. $S_{ij} \cap B(x_2, r + \delta) = \emptyset$

The first condition is effectively selecting a coordinate in our embedded space, and then second condition shows that so long as we have a different in ball size δ we get a non-zero along that coordinate. This intuitive idea forms the central lemma of this proof. After which

Lemma 10 (**Central Lemma**)**.** *For any two points in X and $\delta > 0$, if there exists some $r > 0$ such that*

1. $B(x_1, r) \cap B(x_2, r + \delta) = \emptyset$
2. $|B(x_1, r)| \approx |B(x_2, r + \delta)| \in [\frac{1}{10}2^i, 10 \cdot 2^i]$

then there exists a scale i^* where $1 \leq i \leq \log_2(n)$ and for all sets S_{i^*j} we have the following guarantee:

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset \wedge S_{i^*j} \cap B(x_2, r + \delta) = \emptyset] = \Omega(1)$$

Proof. To show that the above holds, lets first frame the problem in plain English, and from there the proof is quite obvious. The statement says that for some scale along which we constructed the sets, every set in that scale S_{ij} will only contain one of the points with good probability. From here, we note that the sets were constructed independently across all variables, so we can simply multiply the probability of each happening:

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset \wedge S_{i^*j} \cap B(x_2, r + \delta) = \emptyset] = \Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset]\Pr[[S_{i^*j} \cap B(x_2, r + \delta) = \emptyset]$$

From here, we simply need to calculate each product term. Note that based on the construction of the sets, it is easiest to calculate the probability that a ball is empty. Upon this realization, both terms are straightforward to calculate, with the steps following from basic combinatoric arguments.

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset] = 1 - \Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset]$$
$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset] = 1 - (1 - 2^{-i})^{|B(x_1,r)|}$$

The one trick to note is that we want the lower bound on the number of points in the set since we are trying to hit it.

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset] = 1 - (1 - 2^{-i})^{m/10}$$

By assumption, $m = 2^{-i}$, then the following holds:

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset] \geq \Omega(1)$$

Proving that the other ball has no intersection with the sampled sets follows a same paradigm, but the exponent comes from the worst case where m is the largest:

$$\Pr[S_{i^*j} \cap B(x_2, r + \delta) \neq \emptyset] = textnormal Pr[S_{i^*j} \cap B(x_2, r + \delta) = \emptyset]$$

$$\Pr[S_{i^*j} \cap B(x_2, r + \delta) \neq \emptyset] = (1 - 2^{-i})^{|B(x_1,r)|}$$

The one trick to note is that we want the lower bound on the number of points in the set since we are trying to hit it.

$$\Pr[S_{i^*j} \cap B(x_2, r + \delta) \neq \emptyset] = (1 - 2^{-i})^{10m}$$

By assumption, $m = 2^{-i}$, then the following holds:

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset] \geq \Omega(1)$$

From the independence assumption, we can therefore conclude the following from the independence assumption:

$$\Pr[S_{i^*j} \cap B(x_1, r) \neq \emptyset \wedge S_{i^*j} \cap B(x_2, r + \delta) = \emptyset] \geq \Omega(1)$$

□

Corollary 11 (Corollary of the Central Lemma)**.** *Under the conditions of the Central lemma, it immediately follows that:*

$$\Pr\left[\sum_{i=1}^{\log(n)} |f_{S_{ij}}(x_2) - f_{S_{ij}}(x_1)| \geq \Omega(\log(n))\delta\right] \geq 1 - \frac{1}{100n^2}$$

With this proof in hand, we can finally prove the guarantee that the sampling procedure gives an embedding with low distortion.

Lemma 12 (**Main Lemma**). *For any two points y_1, y_2 in X*

$$\Pr[||f_S(y_2) - f_S(y_1)|| \geq \Omega(\log(n)) d_X(y_1, y_2)] \geq 1 - \frac{1}{100n^2}$$

Proof. Let us define the r_k to be the smallest radius such that $|B(y_1, r_k)| = 2^k$ and $|B(y_2, r_k)| \geq 2^k$. Note that this radius will depend on the the points. Despite this, let us now consider a negligibly smaller $r' < r_k$ such that $|B(y_1, r')| < 2^k$. Additionally, consider r_{k-1} such that $|B(y_2, r_{k-1})| \geq 2^{k-1}$. Assume that $B(y_1, r') \cap B(y_2, r_{k-1}) = \emptyset$. From here, we can apply the Corollary of the Central Lemma to the l_1 norm:

$$||f_S(y_2) - f_S(y_1)||_1 = \sum_i \sum_j |f_{S_{ij}}(y_2) - f_{S_{ij}}(y_1)|$$

$$\sum_i \sum_j |f_{S_{ij}}(y_2) - f_{S_{ij}}(y_1)| \geq \sum_i^{\log_2(n)} \Omega(\log(n))(r_i - r_{i-1})$$

Note that the summation telescopes to give:

$$\sum_i^{\log_2(n)} |f_{S_{ij}}(y_2) - f_{S_{ij}}(y_1)| \geq \Omega(\log(n))(r_{\log_2(n)} - r_0)$$

Further note that $r_0 = 0$ since the ball of radius zero has exactly one point. Furthermore, $r_{\log_2(n)}$ is exactly $d_X(y_1, y_2)$ by the assumption that the balls were disjoint.

$$\sum_i^{\log_2(n)} |f_{S_{ij}}(y_2) - f_{S_{ij}}(y_1)| \geq \Omega(\log(n)) d_X(y_1, y_2)$$

Back substituting, we see the following:

$$||f_S(y_2) - f_S(y_1)||_1 \geq \Omega(\log(n)) d_X(y_1, y_2)$$

As we invoked the Corollary to the Central Lemma, it immediately follows that this bound occurs with the following probability:

$$\Pr\left[\sum_j |f_{S_{ij}}(y_2) - f_{S_{ij}}(y_1)| \geq \Omega(\log(n))(r_k - r_{k-1})\right] \geq 1 - \frac{1}{100n^2}$$

□

In conclusion, these series of Lemmas in tandem allow us to state a weaker version of Bourgain's Theorem, which is still quite powerful with a rich geometric interpretation:

Theorem 13. *For any n-point metric space (X, d), there exists a function $f : X \to (\mathbb{R}^{O(\log^2(n))}, l_1)$ with distortion $D = O(\log(n))$.*

3 A Different Perspective via Expander Graphs

Definition 14 (**Expander Graphs**). *Expander graphs are a class of graphs with the following properties, where the number of vertices can tend to infinity:*

1. *The expander graph $G_n = (V_n, E_n)$ is a 3-regular graph.*
2. *$G_n = (V_n, E_n)$ is well connectected.*
3. *$\phi(G_n) \geq \epsilon > 0$, which implies for any subset of vertices $S \subset V$ such that $|S| \leq \frac{n}{2}$, we have $E(S, \bar(S) \geq \Omega(|S|)$.*

Beyond the definition, Expander Graphs are a quirky class of graphs: simultaneous sparse and yet *very* well connected. The utility of expander graphs is a cornerstone of theoretical computer science, but here we are going to concentrate on how the graphs allow us to reason about embeddings. In particular, we consider the shortest cut metric on G_n. Now consider the Laplacian of the expander graph. Clearly the first eigenvalue is non-zero as G_n is connected. Now, we need to reason that the second eigenvalue is greater than some constant. To do this, one can consider a reformulation of the Cheeger Inequlaity, which will be stated without proof:

Lemma 15 (**Cheeger Relation via Expanders**). *Consider some map $f : V_n \to \mathbb{R}$, then the following holds:*

$$\mathbb{E}_{(u,v)\in V\times V}[|f(u) - f(v)|_2^2] \leq \frac{1}{\mu_2(\hat{L}(G_n))}\mathbb{E}_{(u,v)\in E}[|f(u) - f(v)|_2^2]$$

Observation 16 (**Implication of Cheeger's Inequality**). *Cheeger's inequality is implied as the ration of expectations gives a bound on the conductance of the graph $\phi(G_n)$.*

Observation 17 (**Expansion to Higher Dimensions**). *By linearity of expectation, the lemma above can be expanded to maps $f : V_n \to \mathbb{R}^d$.*

A similar theorem holds for the l_1 norm, which then gives us a Bourgain-like bound on the distortion for an embedding:

Theorem 18 (**Expander Graph Embeddings**). *For any function $f : V_n \to \mathbb{R}^d$:*

$$\mathbb{E}_{(u,v)\in V\times V}[||f(u) - f(v)||_1] \leq O(1)\mathbb{E}_{(u,v)\in E}[||f(u) - f(v)||_1]$$

Observation 19. *Recognizing that the embedding causes distortion, we can see the right hand side is upper bounded by $O(D)$.*

Claim 20 (**Expander Graphs Embed Poorly**). *Via properties of Expander Graphs, the left hand side is bounded below by $\Omega(\log(n))$. This implies that embedding the shortest path metric on expander graphs into l_1 incurs a distortion $D = \Omega(\log(n))$.*

Ultimately, this discussion may lead one to conclude that Expander Graphs simply don't embed well. While that is true, a much more positive view of this result is that graphs with good sparse cuts do tend to embed well!

3.1 Improving Bourgain

Using Expander Graphs, it is actually possible to get an improved bound on the distortion via Ramsay Embeddings. The theorem is stated without proof, although the underlying tradeoff and resultant advantages are quite natural.

Theorem 21 (**Bourgain++**). *For an n-point metric space (X, d), for every $\epsilon > 0$, there exists a subset $X' \subset X$ of size $|X'| \geq n^{1-\epsilon}$ such that X' embeds into l_1 with distortion $D = O(\frac{1}{\epsilon})$.*

4 Next Week and Further Reading

Next class, we will be discussing **Padded Decompositions of Metric Spaces**. In the mean time, the original Bourgain paper is here, where one can see the original distortion bound of $O(\log(n)/\log(\log(n)))$that Ilya briefly mentioned in class as a segue into Expander Graphs.

Bou85 On lipschitz embedding of finite metric spaces in Hilbert space. IJM, 1985.