COMS W6998-5: Algorithms through Geometric Lens (Fall'17) Nov 13, 2017

Lecture 19 – Padded Decomposition of Metric Spaces

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribes: *Sebastian Cueva-Caro*

1 Introduction

Last time we proved **Bourgain's Theorem** which states that any n-point metric space embeds into $\left(\mathbb{R}^d, l_1\right)$, where $d = O(\log^2 n)$, with distortion $D = O(\log n)$. This is a hard lower bound for the distortion as there are some metrics that required it. The example we looked in class was the shortest path metric of an expander graph.

Today we will continue working with a general finite metric space and develop one primitive which is not hard to establish but is extremely useful. We will then see some of its many applications, extending to next lecture. This primitive is called the **Padded Decomposition**.

2 Padded Decomposition

Definition 1. *Given a (X, d_X)-metric, where $|X| = n$, and a parameter $\Delta > 0$ (the distance scale), a padded decomposition is a random partition $\mathcal{P}$ such that*

1. *For every part of $\mathcal{P}$, $Pr[diameter \ \leq \Delta] = 1$*
2. *$\forall t \leq \Delta/8$ and $\forall x \in X$, $Pr[B(x,t) \subseteq \mathcal{P}(x)]$ is decent*
3. *$\mathcal{P}$ is algorithmically "nice".*

For any x, its partition does not cut the ball around it. Of course this is dependent on the radius t. For a very small t, we hope that this probability is very close to 1. Later we will write a quantitative bound and prove it. This definition is very similar to LSH, only that these properties are stronger than LSH's , however, LSH holds for more general metric spaces (i.e. not finite).

Now we will explore one application, an approximation algorithm, somewhat in spirit of the sparsest cut.

3 The Multicut Problem

Definition 2. *Given a graph $G = (V, E)$ and a set of pairs $(s_1, t_1), \ldots, (s_k, t_k)$, the Multicut problem asks for the minimum number of edges which separate all those pairs when removed.*

This is an NP-hard problem in this generality, so we want to find the best approximation we can get. Assuming that we have a padded decomposition primitive, there is a very simple algorithm.

Theorem 3. *Given a padded decomposition, there exist a $O(\log n)$ approximation algorithm for the multicut problem.*

Unlike sparse cut, this is the best result we know. Historically, the algorithm was obtained using the same relaxation that we will use without the use of padded decomposition, making the analysis much harder.

3.1 LP-Relaxation

Like the sparsest cut, the relaxation is computing a certain metric on the vertices. The relaxation goes as follows: For some semimetric d on V, we require $\forall i, d(s_i, t_i) = 1$. The the objective function is

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

with

$$d^* = \arg\min \sum_{(u,v)\in E} d(u, v)$$

Clearly, this can be solved in polynomial time using linear programming. Thus, we have the following claim:

Claim 4. $OPT \geq OPT^*$

which works in full generality, implying that this is a relaxation. Thus, by allowing a general class of metrics, we decrease our solution. We will show how to construct it from the metric, but first, we need to show:

Theorem 5. *There exists a multicut of the value of at most* $OPT^* \cdot O(\log n)$

Proof. Sample a padded decomposition of d^* (optimal) with $\Delta = 0.9$ (needs to be anything smaller than 1) and output it.

We want to claim that a specific edge will not be cut too often. The intuition here is that if the distance is sufficiently small, then we can just invoke the padded decomposition property and say that the whole ball centered in u with the corresponding radius containing v will not be cut. Thus, if the ball is not cut, then (u, v) is not either. We will later prove that $Pr[(u, v) \text{ is cut}] \leq O(d^*(u, v) \log n)$. Now we can analyze this construction by computing its expected cost.

$$\begin{aligned}
E[\text{number of edges cut}] &= \sum_{(u,v)\in E} Pr[(u, v) \text{ is cut}] \\
&\leq \sum_{(u,v)\in E} O(d^*(u, v) \log n) \\
&= O(\log n) OPT^*
\end{aligned}$$

□

For large bounds this is a trivial statement because probability can't be greater than 1. This bound becomes interesting only for very small distances less than $1/O(\log n)$.

4 Construction

We now look into the construction of a padded decomposition by performing a form of ball carving. We take random points in our metric space and consider balls of some random radius.

1. Choose R to be a uniform on $\left[\frac{\Delta}{4}, \frac{\Delta}{2}\right]$
2. Sample $x_1 \in X$ uniformly and declare $B(x_1, R)$ to be a part of our partition $\mathcal{P}$
3. Sample $x_2 \in X$ uniformly and declare a new part $B(x_2, R) \setminus B(x_1, R)$.

We repeat these steps until we cover everything. The diameter of every part will be at most Δ by our choice of sample space. It is also clear that this is algorithmically "nice" as sampling would take $O(m \log n)$ iterations.

Now we will try to understand the probability. Suppose that we fix R and have a $x \in X$. Then, we claim:

Claim 6. $Pr[B(x,t) \subseteq \mathcal{P}(x)] \geq \frac{|B(x,R-t)|}{|B(x,R+t)|}$

Which follows from the assumption that if a point is within the $R-t$ ball, then it will always capture the ball of radius t. Our bound is the conditional probability that a point is inside the $R-t$ ball given it is inside the larger one.

To show our assumption correct, we rely on the following claims:

Claim 7. *If $x_i \notin B(x, R+t)$, then we can discard it as it will not affect the event.*

Claim 8. *If $x_i \notin B(x, R-t)$, then $B(x,t) \subseteq \mathcal{P}(x)$.*

The fact that we can say this for a fixed R explains why it does not work. A priori, the ratio could be anything which would lead to terrible bounds. This is why we need randomness; as we will show, on average for a varying R, the bound is much better.

Lemma 9. *For a randomly sampled $R \in \left[\frac{\Delta}{4}, \frac{\Delta}{2}\right]$, we have*

$$\begin{aligned} Pr[B(x,t) \subseteq \mathcal{P}(x)] &\geq E\left[\frac{|B(x,R-t)|}{|B(x,R+t)|}\right] \\ &\geq \left(\frac{|B(x,\Delta/8)|}{|B(x,\Delta)|}\right)^{8t/\Delta} \end{aligned}$$

Before we prove the lemma, we want to note the bounds of this probability P for several cases

- For any t, $P \geq n^{-8t/\Delta}$
- If $t \sim \epsilon\Delta$, $P \geq n^{-O(\epsilon)}$
- If $t \sim \frac{\Delta}{\log n}$, $P \geq 0.99$
- If $t << \frac{\Delta}{\log n}$, $P \geq 1 - O(\frac{t \log n}{\Delta})$.

Back to the proof. We will first introduce a bit of notation needed.

Definition 10. *Let $h(s) = \ln |B(x, s)|$ be a non-decreasing hashing function.*

Proof. We can use h to rewrite the expectation of the ratio of the two balls, and then use straight-forward calculus operations and **Jensen's inequality** to find a lower bond.

$$
\begin{aligned}
E_R\left[\frac{|B(x,R-t)|}{|B(x,R+t)|}\right] &= E_R\left[\exp(h(R-t) - h(R+t)\right] \\
&\geq \exp(E_R\left[h(R-t) - h(R+t\right]) \\
&= \exp\left(\frac{4}{\Delta}\int_{\Delta/4}^{\Delta/2}(h(R-t) - h(R+t)dR\right) \\
&= \exp\left(\frac{4}{\Delta}\int_{\Delta/4-t}^{\Delta/4+t} h(R)dR - \int_{\Delta/2-t}^{\Delta/2+t} h(R)dR\right) \\
&\geq \exp\left(\frac{8t}{\Delta}(h(\Delta/4-t) - h(\Delta/2+t))\right) \\
&= \left(\frac{|B(x,\Delta/4-t)|}{|B(x,\Delta/2+t)|}\right)^{8t/\Delta} \\
&\geq \left(\frac{|B(x,\Delta/8)|}{|B(x,\Delta)|}\right)^{8t/\Delta}
\end{aligned}
$$

□

5 Next Week

Next lecture we will look at other applications of padded decomposition and how it can affect Bourgain's results.

5.1 Strengthening of Bourgain

We can use padded decomposition for a better proof of Bourgain's theorem that gives a stronger statement. Remember that Bourgain's gave us a embedding in l_1. We can think of it not as a deterministic embedding but as a random embedding into a line given by a map $f : X \to (\mathbb{R}^d, l_1)$. We can reformulate Bourgain's theorem to show:

Corollary 11. *For all n-point X, there exists a random $f : X \to \mathbb{R}$ such that $\forall x_1, x_2 \in X$,*

- $\forall f$, $|f(x_1) - f(x_2)| \leq O(\log n) \cdot d_X(x_1, x_2)$
- $E_f|f(x_1) - f(x_2)| \geq d_X(x_1, x_2)$

Again, this does not improve on Bourgain's. A real strengthening is given by a similar theorem were we approximate a metric by trees.

Theorem 12 (Baltar, Fakcharoenphol, Rao, Talwar)**.** *For all n-point X, there exists a random $f : X \to \mathbb{T}_f$, where T_f is a tree, such that $\forall x_1, x_2 \in X$,*

- $E[d_T(f(x_1), f(x_2))] \leq O(\log n) \cdot d_X(x_1, x_2)$

- $Pr[d_T(f(x_1), f(x_2))] \geq d_X(x_1, x_2)$

This embedding gives us Bourgain's for free because if we have a tree, we can always decompose it into a combination of paths with distortion $O(\log n)$.

Finally, we will also look at a variant of Bourgain embedding for subsets:

Theorem 13. *For all n-point X, $\forall \epsilon > 0, \exists X' \subset X$ with $|X'| = n^{1-\epsilon}$, $\exists f : X' \to (\mathbb{R}, l_1)$ such that the distortion of f is $O(1/\epsilon)$.*