CS 369M: Metric Embeddings and Algorithmic Applications October 4, 2018

Lecture 4: Uniform Sparsest Cut via Bourgain's Embedding

Lecturer: Moses Charikar *Scribe: Jay Mardia*

1 Introduction

In the previous lecture we studied Bourgain's embedding, which embeds any n-point metric into ℓ_1 with $O(\log n)$ distortion (and after a slight modification, with distortion $O\left(\frac{\log n}{p}\right)$ into ℓ_p) and began to set up the problem of Uniform Sparsest Cut. This problem is NP-hard to solve exactly, and assuming the Unique Games Conjecture, its generalized version is hard to approximate within any constant factor. In this lecture we see how to obtain an $O(\log n)$ approximation for uniform sparsest cut using Bourgain's embedding into ℓ_1. To do this, we rewrite uniform sparsest cut as an optimization problem over cut metrics, and then relax this optimization problem. We then show that this relaxation doesn't hurt "too much" (at most by a factor of $O(\log n)$). Thus the algorithm we use is a relax-and-round type of algorithm, where we use Bourgain's embedding in the rounding step.

We then state that the same algorithm with the same approximation guarantees works for a generalized version of uniform sparsest cut called (unsurprisingly) non-uniform sparsest cut. We end this lecture by embarking on a proof of lower bounds for distortion-dimension tradeoffs for embedding arbitrary n-point metrics into any normed space. We do not complete this proof, and will do so in the next lecture.

2 Uniform Sparsest Cut

Let $G = (V, E)$ be a graph with vertices V and edges E, with $|V| = n$. Let $S \subset V$ such that $S \neq \emptyset$ and let $\bar{S} = V \setminus S$. We define the *sparsity* of the cut $\left(S, \bar{S}\right)$ to be

$$\phi\left(G, S\right) \triangleq \frac{|E\left(S, \bar{S}\right)|}{|S||\bar{S}|}.$$

The sparsest cut is the cut $\left(S, \bar{S}\right)$ which minimizes $\phi\left(G, S\right)$.

Our strategy to solve this optimization problem is to first look at it from a different angle. We simply restate the same problem in the language of metrics to make it more amenable to tools developed in this course.

Every cut $\left(S, \bar{S}\right)$ defines a metric (d_S, V), which we call a cut metric, on the graph G as follows:

$$d_S\left(u, v\right) \triangleq \begin{cases} 1 & |S \cap \{u, v\}| = 1 \\ 0 & \text{otherwise} \end{cases}$$

We showed in the previous lecture that the uniform sparsest cut problem is exactly equivalent to minimizing the following objective function (which we denote by $P_{\text{cut metrics}}$ in these notes – this is not standard terminology):

$$\min_{d_S \text{ is a cut metric}} \frac{\sum_{\{u,v\}\in E} d_S(u,v)}{\sum_{u,v\in V} d_S(u,v)}$$

Of course, since uniform sparsest cut is a hard problem, so is solving $P_{\text{cut metrics}}$ since the two are equivalent. However, we will make two successive relaxations to the set we optimize over. The first relaxation doesn't change the optimum, and hence is hard too. The second relaxation makes the problem tractable, but its optimal solution may have a lower value than the sparsest cut (and will need to be rounded to a solution to the sparsest cut problem).

We first relax the problem to $P_{\text{conic combs of cut metrics}}$ as defined below. A conic combination (or non-negative linear combination) of cut metrics is a metric of the form $d(u,v) = \sum_S \alpha_S d_S(u,v)$, where the sum is over all possible cut metrics d_S, and the coefficients α_S are non-negative. As the name suggests, $P_{\text{conic combs of cut metrics}}$ is simply the optimization problem with the feasible set relaxed to all conic combinations of cut metrics.

$$\min_{d \text{ is a conic combination of all cut metrics}} \frac{\sum_{\{u,v\}\in E} d(u,v)}{\sum_{u,v\in V} d(u,v)}.$$

This changes the set we must optimize over from a discrete set to a continuum. Note that since every cut metric d_S is a conic combination of cut metrics, the optimal value of $P_{\text{conic combs of cut metrics}}$ is at most the optimal value of $P_{\text{cut metrics}}$. In the other direction, if we consider the optimal solution (or any feasible point) $d(u,v) = \sum_S \alpha_S d_S(u,v)$ for $P_{\text{conic combs of cut metrics}}$, we can find a cut metric d_S for which the value of the objective function is at most the value of the objective function for d. This follows from the inequality

$$\frac{a_1 + a_2 + \ldots + a_k}{b_1 + b_2 + \ldots + b_k} \geq \min_{i=1,\ldots,k} \frac{a_i}{b_i}.$$

Hence, the optimal value of $P_{\text{conic combs of cut metrics}}$ is the same as the optimal value of $P_{\text{cut metrics}}$.

However, $P_{\text{conic combs of cut metrics}}$ is still not tractable problem. We thus further relax the problem to be an optimization problem over **all** metrics. We call this P_{metrics}:

$$\min_{d \text{ is a metric}} \frac{\sum_{\{u,v\}\in E} d(u,v)}{\sum_{u,v\in V} d(u,v)}.$$

Why is P_{metrics} tractable? And if we obtain a *good solution* for P_{metrics}, how can we use it to obtain a *good-enough solution* for $P_{\text{conic combs of cut metrics}}$?

1. P_{metrics} is actually equivalent to a linear program (which we will call LP_{metrics}). We know that linear programs can be solved in polynomial time.

   (a) For every pair u, v, let x_{uv} be a variable in the LP. x_{uv} will be corresponding to $d(u,v)$. The constraints that ensure that the variables x_{uv} define a metric are simply linear constraints. These constraints are simply the non-negativity and triangle inequality constraints over all pairs, which are:

i. $x_{uv} \geq 0$ for all $u, v \in V$

ii. $x_{uv} + x_{vw} \geq x_{uw}$ for all $u, v, w \in V$

(b) Recall that the objective function of P_{metrics} is

$$\min_{d \text{ is a metric}} \frac{\sum\limits_{\{u,v\} \in E} d(u,v)}{\sum\limits_{u,v \in V} d(u,v)}.$$

This objective function is fractional, which seems like it won't let us get a linear program. However, there is a straightforward trick that lets us convert objective functions that are fractions of linear functions into a linear program, simply by noting that the objective function is invariant to scaling all the variables by the same constant and by then adding the constraint $\sum\limits_{u,v \in V} x_{uv} \geq 1$ to our set of constraints. We then let the objective of LP_{metrics} then simply be $\sum\limits_{\{u,v\} \in E} x_{uv}$. Any solution to LP_{metrics} can be scaled down to let $\sum\limits_{u,v \in V} x_{uv} = 1$ (the scaling can only decrease the value of the objective function), and hence the obtained optimal solution to the LP will be an optimal solution to P_{metrics}.

Hence we see that P_{metrics} is equivalent to a linear program LP_{metrics} which can be solved efficiently.

2. Thus we know that we can obtain an optimal solution to P_{metrics}, and this solution is an arbitrary n-point metric. How can we convert this to a conic combination of cut metrics that doesn't do too much worse on the objective function? Here we use the following fact

Fact 2.1. *The set of all conic combinations of cut metrics is equivalent to the set of all ℓ_1 metrics.*

It is easy to see that every non-negative linear combination of cut metrics is an ℓ_1 metric. For each cut metric with coefficient α_S in our non-negative linear combination, simply define a new coordinate in our ℓ_1 space. For the coordinate corresponding to each cut, set the value of all points in S to 0 and in $\bar{S}$ to α_S. Thus this is an ℓ_1 metric. The reverse implication requires some work. At a high level, an ℓ_1 metric can be thought of as a sum of line metrics. Then, each line metric can be represented as a non-negative linear combination of cut metrics.

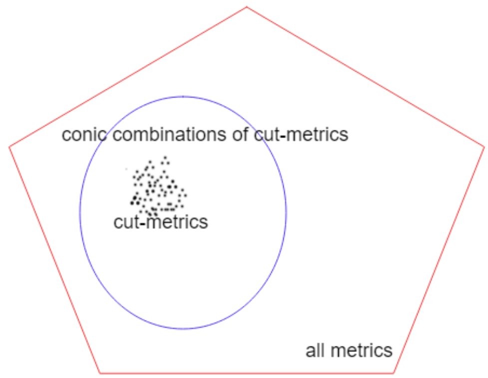
So the question now becomes "How to map an arbitrary n-point metric to an ℓ_1 metric without too much loss?". But this is exactly what Bourgain's embedding (which we studied in the last lecture) does.

Theorem 2.2 (Bourgain)**.** *There exists an embedding f from an arbitrary n-point metric (V, d_V) to ℓ_1 such that for all $u, v \in V$, the following holds*

$$d_V(u,v) \leq ||f(u) - f(v)||_1 \leq O(\log n) \cdot d_V(u,v).$$

If we apply Bourgain's embedding to the solution of P_{metrics}, the denominator of our objective can only increase (which is good for us) and the numerator of our objective can increase by at most $O(\log n)$. Hence we obtain a *good-enough solution* for $P_{\text{conic combs of cut metrics}}$ which is at most $O(\log n)$ worse than the optimal solution to the LP.

Figure 1: An illustration of the relaxations we use to approximately solve uniform sparsest cut



3 Multi-Commodity Flow and Non-Uniform Sparsest Cut: Known Upper and Lower Bounds

3.1 Multi-Commodity Flow

Much like the famed max-flow-min-cut duality which says that max flow and min cut are dual problems, we also have a dual of the linear program LP_{metrics}. Consider the following multi-commodity flow problem. We have k pairs of vertices in the (weighted) graph $G = (V, E)$ given by $\{s_i, t_i\}$ for $i \in \{1, 2, ..., k\}$. We want to maximize α such that there exists a feasible flow on G such that for all $i \in \{1, 2, ..., k\}$, the flow from s_i to t_i is at least α.

With a little thought, one can see the optimization problem above is dual to LP_{metrics}, which is the relaxation of uniform sparsest cut.

3.2 Non-Uniform Sparsest Cut

We can actually generalize the Uniform Sparsest Cut in such a way that the problem becomes much more general but can still be solved approximately in the same way by using relax and round with Bourgain's embedding like uniform sparsest cut is solved. We now present this generalization.

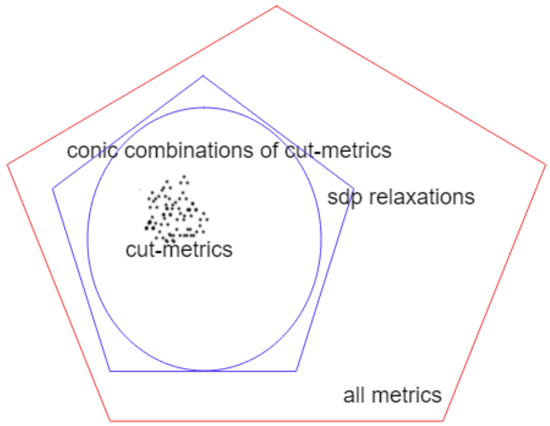
Let $G = (V, E)$ be the complete graph on $|V| = n$ points, where there are sets of edge weights: $\alpha_{uv} \geq 0$ and $\beta_{uv} \geq 0$ (for all $u, v \in V$). Then the Non-Uniform Sparsest Cut problem is that of finding $S \subseteq V$ such that

$$\min_{S \subseteq V} \frac{\sum\limits_{u \in S, v \in \bar{S}} \alpha_{uv}}{\sum\limits_{u \in S, v \in \bar{S}} \beta_{uv}}$$

It is easy to observe that this is indeed a generalization of uniform sparsest cut by letting $\alpha_{uv} = 1$ if $\{u, v\} \in E$ and 0 otherwise, and letting $\beta_{uv} = 1$ for all $u, v \in V$.

For a given optimization problem and an LP relaxation corresponding to it, we often care about something called the *Integrality Gap*. The Integrality Gap is defined as the maximum over all problem instances of the ratio between the true optimal value to the optimal value of the relaxed

Figure 2: An illustration of the tighter relaxations we can use to approximately solve sparsest cut



	Upper Bound	Lower Bound
Uniform Sparsest Cut	$O(\sqrt{\log n})$	$\Omega(\log\log n)$
Non-Uniform Sparsest Cut	$O(\sqrt{\log n}\log\log n)$	$\Omega(\sqrt{\log n})$ (only for solving using SDP)

LP:

$$\text{Integrality Gap} = \max_{\text{all problem instances } I} \frac{\text{OPT}(I)}{\text{LP}(I)}$$

Here $\text{OPT}(I)$ corresponds to the true optimal value and $\text{LP}(I)$ corresponds to the optimal value of the relaxed LP.

We have shown that the Integrality Gap of the uniform sparsest cut LP_{metrics} relaxation is $O(\log n)$. In fact, the following is true.

Fact 3.1. *The Integrality Gap of Non-Uniform Sparset Cut = Worst Case Distortion Needed to map arbitrary n-point metrics into* ℓ_1 $= O(\log n)$

This shows that the Bourgain embedding is indeed the right tool for the rounding job when trying to solve Non-Uniform Sparsest Cut by using LP relaxations.

3.3 Known Upper and Lower Bounds for Sparsest Cut

The fact above shows that if we hope to get better results for Non-Uniform Sparsest Cut, we will need to use tighter relaxations than LP relaxations. There exists a hierarchy of such tighter relaxations, for example SDP relaxations (which corresponds to ℓ_2^2 metrics) that people have studied.

We also state some known upper and lower bounds in the table above.

4 Distortion-Dimension Tradeoff Lower Bounds

We now embark on a proof of lower bounds for distortion-dimension tradeoffs for embedding arbitrary n-point metrics into any normed space. We do not complete this proof, and will do so in the next lecture.

Theorem 4.1. *For every $D < 3$, there is a constant $c_D > 0$ (which depends only on the distortion D) such that if Z is a k-dimensional normed space such that all n-point metrics embed into Z with distortion at most D, then $k \geq c_D \cdot n$.*

This theorem says that we can't achieve distortion less than 3 with dimension sublinear in the number of points. The idea behind the proof is that we will come up with a large number of n-point metrics that are different from each other, and and argue that if we find a low distortion embedding into Z for all of them, then all these embeddings must be sufficiently different from each other. We will then argue that there can't be these many embeddings into Z that are all so different if k is too small (sublinear in n). This sort of counting argument will provide us with a contradiction.

Consider n even, and $K_{\frac{n}{2},\frac{n}{2}} = (V, E)$, the complete bipartite graph on n vertices. So $|V| = n$ and $|E| = \left(\frac{n}{2}\right)^2 = m$

Now consider the set of all subgraphs of $K_{\frac{n}{2},\frac{n}{2}}$ (this set has size $2^{|E|} = 2^m$) and for every graph H in this set define an n-point metric as follows:

$$d_H(u, v) \triangleq \min\{\text{shortest path in } H \text{ between } u \text{ and } v, 3\}$$

Fact 4.2. *If H_1 and H_2 are two different subgraphs of $K_{\frac{n}{2},\frac{n}{2}}$, then d_{H_1} and d_{H_2} are different in the following sense: There exists $\{u, v\} \in E$ such that either ($d_{H_1}(u, v) = 1$ and $d_{H_2}(u, v) = 3$) or ($d_{H_1}(u, v) = 3$ and $d_{H_2}(u, v) = 1$).*

This is easy to see using the fact that there must exist an edge $\{u, v\} \in E$ which exists in exactly one of H_1 or H_2. Thus we have a set of 2^m different n-point metrics, each corresponding to a subgraph of $K_{\frac{n}{2},\frac{n}{2}}$. In particular, if for each such H_1 and H_2, we consider their respective embeddings into Z (scaled to be non-expanding), they must be different if the distortion is strictly less than 3.

Now to use any counting argument for maps into a continuous space, we must discretize the continuous space (Z in our case) using a tool called δ-net.

Definition 4.3 (δ-net)**.** A set of points $P \subseteq Z$ is called a δ-net of Z if for all $x \in Z$ we can find at least one point $y \in P$ such that $d_Z(x, y) \leq \delta$.

We can show, using a greedy construction that any k-dimensional normed space Z has a δ-net P with $|P| \leq \left(\frac{c}{\delta}\right)^k$ where c is a universal constant.

We complete the rest of the proof in the next lecture.