COM S 783 / OR&IE 634 Approximation Algorithms **Fall Semester, 2003**

Lecture 11: Oct 6, 2003

Lecturer: Yuval Rabani *Scribe: Ara Hayrapetyan*

1 Bourgain's Theorem

We ended last lecture in observing that if we could approximate an arbitrary semimetric into an L_1 metric of small distortion, we would get good approximation algorithms for the sparsest cut problem. In this lecture we will do just that. We will prove *Bourgain's theorem* [1], which states that any semimetric can be approximated into an L_1 metric with distortion $O(\log n)$ and $O(\log^2 n)$ dimensions [2]. More rigorously, we want a construction that given a finite semimetric (X, d) of cardinality n and a designated subset of nodes $Y \subseteq X$ of cardinality k, finds an embedding $\varphi : X \mapsto R^m$ such that

- $\forall x, y \in X \quad ||\varphi(x) - \varphi(y)||_1 \leq d(x, y)$,
- $\forall x, y \in Y \quad ||\varphi(x) - \varphi(y)||_1 \geq \frac{d(x,y)}{c \log k}$,

where c is an absolute constant and $m = O(\log^2 k)$. From now on unless mentioned otherwise when talking about points we will refer to points from Y.

Before proceeding with the actual proof, we notice that if we have such an embedding, then by scaling all the distances of the embedding by $O(\log k)$ (or equivalently scaling all the coordinates of all the points) we get a non-contracting embedding with distortion $O(\log k)$, which is what we will actually use for the sparsest cut.

The proof of Bourgain's Theorem will be constructive. We will have an index i ranging over $1, 2, \ldots, \log k$, and for each i we will generate $\log k$ coordinates (thus $m = O(\log^2 k)$). Each coordinate will be generated randomly and independently.

The j-th i-coordinate is generated as follows. Pick uniformly at random a set $S_{i,j}$ of size $\frac{k}{2^i}$ from the set Y. Let the (i, j)-th coordinate of a point x be

$$\varphi(x)_{i,j} = d(x, S_{i,j}) \frac{1}{O(\log^2 k)}, \tag{1}$$

(the distance to the closest point scaled down by the total number of coordinates).

We first note that for all $S_{i,j}$, $|d(x, S_{i,j}) - d(y, S_{i,j})| \leq d(x, y)$ because of triangle inequality. Thus for any two points (x, y) along each coordinate (i, j) the distance will be bounded by $\frac{d(x,y)}{O(\log^2 k)}$ (since we scaled down all coordinates by $O(\log^2 k)$), thus the total L_1 distance will be bounded by $d(x, y)$. Hence the construction is non-expanding.

To show that the embedding is approximately distance preserving, consider again nodes x and y. We will use the notation $B(x, \rho)$ to denote the ball around x of radious ρ. Let $0 = \rho_0 \leq \rho_1 \leq \ldots \rho_{max}$ denote radii such that ρ_i is the minimum radius where both $B(x, \rho)$ and $B(y, \rho)$ contain at least 2^i points from Y. Since we want no two balls to contain the same point, ρ_{max} will be defined as $\frac{1}{2}d(x, y) - \epsilon$.

Consider a given ρ_i, and let x be the side that has exactly 2^i nodes in $B(x, \rho_i)$ and y has at least 2^i nodes in a ball of identical radius. Look at a particular $S_{i,j}$. Our first observation is that there is a constant probability that $S_{i,j}$ does not intersect with $B(x, \rho_i)$ and yet intersects with $B(y, \rho_{i-1})$. Indeed, since $S_{i,j}$ picks $\frac{k}{2^i}$ points, the probability that each point is picked is $\frac{1}{2^i}$, thus if these probabilities were independent then with probability $(1 - 2^{-i})^{2^i}$ we will miss the set $B(x, \rho_i)$ entirely (the ball contains exactly 2^i points). This quantity is approximately e^{-1} and similar calculations approximately (within a constant) hold for our experiment. Similarly one can show that the probability of hitting $B(y, \rho_{i-1})$ is a constant, therefore there is a constant probability of both of the events happening simultaneously.

Notice that what this implies is that

- $\varphi(y)_{i,j} \leq \frac{\rho_{i-1}}{O(\log^2 k)}$
- $\varphi(x)_{i,j} \geq \frac{\rho_i}{O(\log^2 k)}$

Therefore $|\varphi(x)_{i,j} - \varphi(y)_{i,j}| \geq \frac{1}{O(\log^2 k)}(\rho_i - \rho_{i-1})$ happens with constant probability and therefore will occur in $O(\log k)$ of the experiments that we do for a fixed i (one could use Chernoff-style bounds to formalize this - for details please see the paper).

Finally, summing over all i and all j when this event happens we obtain

$$||\varphi(x) - \varphi(y)||_1 \geq \sum_i O(\log k) \frac{1}{O(\log^2 k)} (\rho_i - \rho_{i-1}) = \frac{1}{O(\log k)} d(x, y), \tag{2}$$

since the sum of ρ_i's telescope to ρ_{max} which is $O(d(x, y))$. The proof is complete.

2 Applications of Sparsest Cuts

One of the most important features of sparsest cuts is that it allows us to find "good" (balanced) cuts which in turn often yields divide and conquer algorithms for many cut problems. In this section we will see how to find approximate balanced cuts using the sparsest cut results.

First we define the balance of the cut $(S, \bar{S})$ to be the quantity $\frac{min\{|S|, |\bar{S}|\}}{|V|}$, and we say that a cut $(S, \bar{S})$ is B-balanced if the balance of the cut is at least B. 1/2-balanced cuts will be called *bisections* and 1/3-balanced cuts will be called *separators*.

Our goal is to find a good approximate bisection (which could later often be used for recursion) and we will do this using sparsest cuts. Our algorithm will be a *bicriteria approximation* for the bisection problem, i.e. an algorithm returning a separator instead of a bisection, whose capacity is within $O(\log n)$ of the capacity of the cheapest bisection.

The algorithm is very simple. We repeatedly find a sparse cut in the graph and remove the smaller side while recursing on the larger one. The process terminates when the size of the removed component exceeds $\frac{n}{3}$. We first observe that this indeed is a separator. Indeed let $S_1, S_2, \ldots, S_t$ be the removed components. Then the cut we obtain is $(\bigcup_{i=1}^t S_i, S - \bigcup_{i=1}^t S_i)$. The left side of the cut is at least $\frac{n}{3}$ large by construction. As for the right side, observe that $S_1 \cup S_2 \cup \ldots \cup S_{t-1}$ is less that $\frac{n}{3}$, thus $S - \bigcup_{i=1}^{t-1}$ is at least $\frac{2}{3}n$, and S_t takes no more than half of this. Therefore at least $\frac{1}{3}n$ remains.

Now we need to bound the capacity. Let the optimal bisection have value α. This means that the flux of this bisection is $\frac{2\alpha}{n}$ (recall that the flux was the capacity of the cut divided by the size of the smaller side). And since sparsest cut algorithm also gives good approximation algorithm for the best flux we have that

$$\frac{c(S_1, \bar{S_1})}{|S_1|} \leq O(\log n)\frac{2\alpha}{n} \tag{3}$$

$$c(S_1, \bar{S_1}) \leq O(\log n)\frac{2\alpha}{n}|S_1| \tag{4}$$

(recall that S_1 was always the smaller side). And in general after stage i we have that the flux of the remaining portion of the graph is upper-bounded by $\alpha/(\frac{n}{2} - \sum_{j \leq i} |S_j|)$. For the last stage this flux is upper-bounded by $\frac{6\alpha}{n}$. Moreover, for each cut we thus obtain

$$c(S_i, \bar{S_i}) \leq O(\log n)\frac{6\alpha}{n}|S_i| \tag{5}$$

Summing the capacities of the different cuts (which is clearly an upper-bound on the separator capacity) we get

$$c(\bigcup S_i, S - \bigcup S_i) \leq O(\log n)\frac{6\alpha}{n}\sum |S_i| \leq O(\log n)\frac{6\alpha}{n}\frac{2n}{3} = O(\log n)\alpha. \tag{6}$$

Thus we obtain an $O(\log n)$ approximation to the capacity of the best bisection.

References

[1] Jean Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

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