Lecture 14

In which we begin to discuss the Arora-Rao-Vazirani rounding procedure.

Recall that, in a graph $G = (V, E)$ with adjacency matrix A, then ARV relaxation of the sparsest cut problem is the following semidefinite program.

$$\begin{array}{lll}
\text{minimize} & \frac{1}{2|E|}\sum_{u,v} A_{u,v}||\mathbf{x}_u - \mathbf{x}_v||^2 & \\
\text{subject to} & & \\
& \sum_{u,v} ||\mathbf{x}_u - \mathbf{x}_v||^2 = |V|^2 & \\
& ||\mathbf{x}_u - \mathbf{x}_v||^2 \leq ||\mathbf{x}_u - \mathbf{x}_w||^2 + ||\mathbf{x}_w - \mathbf{x}_v||^2 & \forall u, v, w \in V \\
& \mathbf{x}_u \in \mathbb{R}^d & \forall u \in V
\end{array}$$

If we denote by $ARV(G)$ the optimum of the relaxation, then we claimed that

$$ARV(G) \leq \phi(G) \leq O(\sqrt{\log |V|}) \cdot ARV(G)$$

where the first inequality follows from the fact that $ARV(G)$ is a relaxation of $\phi(G)$, and the second inequality is the result whose proof we begin to discuss today.

1 Rounding the Arora-Rao-Vazirani Relaxation

Given the equivalence between the sparsest cut problem and the "ℓ_1 relaxation" of sparsest cut, it will be enough to prove the following result.

Theorem 1 (Rounding of ARV) *Let G be a graph, A its adjacency matrix, and $\{\mathbf{x}_v\}_{v\in V}$ be a feasible solution to the ARV relaxation.*

Then there is a mapping $f : V \to \mathbb{R}$ such that

$$\frac{\sum_{u,v} A_{u,v}|f(u) - f(v)|}{\sum_{u,v} |f(u) - f(v)|} \leq O(\sqrt{\log |V|}) \cdot \frac{\sum_{u,v} A_{u,v}||\mathbf{x}_u - \mathbf{x}_v||^2}{\sum_{u,v} ||\mathbf{x}_u - \mathbf{x}_v||^2}$$

As in the rounding of the Leighton-Rao relaxation via Bourgain's theorem, we will identify a set $S \subseteq V$, and define

$$f_S(v) := \min_{s \in S} ||\mathbf{x}_s - \mathbf{x}_v||^2 \tag{1}$$

Recall that, as we saw in the proof of Bourgain's embedding theorem, no matter how we choose the set S we have

$$|f_S(u) - f_S(v)| \leq ||\mathbf{x}_u - \mathbf{x}_v||^2 \tag{2}$$

where we are not using any facts about $||\cdot - \cdot||^2$ other than the fact that, for solutions of the ARV relaxation, it is a distance function that obeys the triangle inequality.

This means that, in order to prove the theorem, we just have to find a set $S \subseteq V$ such that

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{1}{O(\sqrt{\log |V|})} \cdot \sum_{u,v} ||\mathbf{x}_u - \mathbf{x}_v||^2 \tag{3}$$

and this is a considerable simplification because the above expression is completely independent of the graph! The remaining problem is purely one about geometry.

Recall that if we have a set of vectors $\{\mathbf{x}_v\}_{v \in V}$ such that the distance function $d(u,v) := ||\mathbf{x}_u - \mathbf{x}_v||^2$ satisfies the triangle inequality, then we say that $d(\cdot,\cdot)$ is a (semi-)metric of *negative type*.

After these preliminaries observations, our goal is to prove the following theorem.

Theorem 2 (Rounding of ARV – Revisited) *If $d(\cdot,\cdot)$ is a semimetric of negative type over a set V, then there is a set S such that if we define*

$$f_S(v) := \min_{s \in S} \{d(s,v)\}$$

we have

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{1}{O(\sqrt{\log |V|})} \cdot \sum_{u,v} d(u,v)$$

Furthermore, the set S can be found in randomized polynomial time with high probability given a set of vector $\{\mathbf{x}_v\}_{v \in V}$ such that $d(u,v) = ||\mathbf{x}_u - \mathbf{x}_v||^2$.

Since the statement is scale-invariant, we can restrict ourselves, with no loss of generality, to the case $\sum_{u,v} d(u,v) = |V|^2$.

Remark 3 Let us discuss some intuition before continuing with the proof.

As our experience proving Bourgain's embedding theorem shows us, it is rather difficult to pick sets such that $|f_S(u) - f_S(v)|$ is not much smaller than $d(u,v)$. Here we have a somewhat simpler case to solve because we are not trying to preserve all distances, but only the average pairwise distance. A simple observation is that if we find a set S which contains $\Omega(|V|)$ elements and such that $\Omega(|V|)$ elements of V are at distance $\Omega(\delta)$ from S, then we immediately get $\sum_{u,v} |f_S(u) - f_S(v)| \geq \Omega(\delta |V|^2)$, because there will be $\Omega(|V|^2)$ pairs u,v such that $f_S(u) = 0$ and $f_S(v) \geq \delta$. In particular, if we could find such a set with $\delta = 1/O(\sqrt{\log |V|})$ then we would be done. Unfortunately this is too much to ask for in general, because we always have $|f_S(u) - f_S(v)| \leq d(u,v)$, which means that if we want $\sum_{u,v} |f_S(u) - f_S(v)|$ to have $\Omega(V^2)$ noticeably large terms we must also have that $d(u,v)$ is noticeably large for $\Omega(|V|^2)$ pairs of points, which is not always true.

There is, however, the following argument, which goes back to Leighton and Rao: either there are $\Omega(|V|)$ points concentrated in a ball whose radius is a quarter (say) of the average pairwise distance, and then we can use that ball to get an ℓ_1 mapping with only *constant* error; or there are $\Omega(|V|)$ points in a ball of radius twice the average pairwise distance, such that the pairwise distances of the points in the ball account for a constant fraction of all pairwise distances. In particular, the sum of pairwise distances includes $\Omega(|V|^2)$ terms which are $\Omega(1)$.

After we do this reduction and some scaling, we are left with the task of proving the following theorem: suppose we are given an n-point negative type metric in which the points are contained in a ball of radius 1 and are such that the sum of pairwise distances is $\Omega(n^2)$; then there is a subset S of size $\Omega(n)$ such that there are $\Omega(n)$ points whose distance from the set is $1/O(\sqrt{\log n})$. This theorem is the main result of the Arora-Rao-Vazirani paper. (Strictly speaking, this form of the theorem was proved later by Lee – Arora, Rao and Vazirani had a slightly weaker formulation.)

We begin by considering the case in which a constant fraction of the points are concentrated in a small ball.

Definition 4 (Ball) *For a point $z \in V$ and a radius $r > 0$, the ball of radius r and center z is the set*

$$B(z,r) := \{v : d(z,v) \leq r\}$$

Lemma 5 *For every vertex z, if we define $S := B(z, 1/4)$, then*

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{|S|}{2|V|} \sum_{u,v} d(u,v)$$

PROOF: Our first calculation is to show that the typical value of $f_S(u)$ is rather large. We note that for every two vertices u and v, if we call a a closest vertex in S to u, and b a closest vertex to v in S, we have

$$\begin{aligned} d(u,v) &\leq d(u,a) + d(a,z) + d(z,b) + d(b,v) \\ &\leq f_S(u) + f_S(v) + \frac{1}{2} \end{aligned}$$

and so

$$|V|^2 = \sum_{u,v} d(u,v) \leq 2|V| \cdot \sum_v f_S(v) + \frac{|V|^2}{2}$$

that is,

$$\sum_v f_S(v) \geq \frac{|V|}{2}$$

Now we can get a lower bound to the sum of ℓ_1 distances given by the embedding $f_S(\cdot)$.

$$\begin{aligned} &\sum_{u,v} |f_S(u) - f_S(v)| \\ &\geq \sum_{u \in S, v \in V} |f_S(v)| \\ &= |S| \sum_v f_S(v) \\ &\geq \frac{1}{2}|S| \cdot |V| \end{aligned}$$

□

This means that if there is a vertex z such that $|B(z,1/4)| = \Omega(|V|)$, or even $|B(z,1/4)| = \Omega(|V|/\sqrt{\log |V|})$, then we are done.

Otherwise, we will find a set of $\Omega(|V|)$ vertices such that their average pairwise distances are within a constant factor of their maximum pairwise distances, and then we will work on finding an embedding for such a set of points. (The condition that the average distance is a constant fraction of the maximal distance will be very helpful in subsequent calculations.)

Lemma 6 *Suppose that for every vertex z we have $|B(z,1/4)| \leq |V|/4$. Then there is a vertex w such that, if we set $S = B(w,2)$, we have*

- $|S| \geq \frac{1}{2} \cdot |V|$
- $\sum_{u,v \in S} d(u,v) \geq \frac{1}{8}|S|^2$

PROOF: Let w be a vertex that maximizes $|B(w,2)|$; then $|B(w,2)| \geq |V|/2$, because if we had $|B(u,2)| < |V|/2$ for every vertex u, then we would have

$$\sum_{u,v} d(u,v) > \sum_{u} 2 \cdot (|V - B(u,2)|) > |V|^2$$

Regarding the sum of pairwise distances of elements of S, we have

$$\sum_{u,v \in S} d(u,v) > \sum_{u \in S} \frac{1}{4}(|S - B(u,1/4)|) \geq |S| \cdot \frac{1}{4} \cdot \frac{|S|}{2}$$

□

The proof of the main theorem now reduces to proving the following geometric fact.

Theorem 7 *Let d be a negative-type metric over a set V such that the points are contained in a unit ball and have constant average distance, that is,*

- *there is a vertex z such that $d(v,z) \leq 1$ for every $v \in V$*
- $\sum_{u,v \in V} d(u,v) \geq c \cdot |V|^2$

Then there are sets $S, T \subseteq V$ such that

- $|S|, |T| \geq \Omega(|V|)$;
- *for every $u \in S$ and every $v \in S$, $d(u,v) \geq 1/O(\sqrt{\log |V|})$*

where the multiplicative factors hidden in the $O(\cdot)$ and $\Omega(\cdot)$ notations depend only on c.