College of Computer & Information Science
Northeastern University
CS G399: Algorithmic Power Tools I

Fall 2007
30 November 2007
Scribe: Stefan Savev

Lecture Outline: In this lecture we present the Master Structure Theorem which is the core of the Arora-Rao-Vazirani (ARV) approach to the sparsest cut problem [1]. Their approach achieves $O(\sqrt{\log n})$ approximation ratio. At the heart of the approach is a phenomenon known as measure concentration. Our coverage here is primarily based on the excellent lecture notes taken in Avner Magen's class at University of Toronto [2].

Review: The sparsest cut problem could be written in the following way:

$$\begin{aligned} & \min_S \frac{|E(S,\overline{S})|}{|S||\overline{S}|} \\ = & \min_S \frac{\sum_{i,j \epsilon E} d_S(i,j)}{\sum_{i,j} d_S(i,j)} \\ = & \min_{l_1} \frac{\sum_{i,j \epsilon E} d_{l_1}(i,j)}{\sum_{i,j} d_{l_1}(i,j)} \\ = & \min_{v_i \epsilon \{-1,+1\}} \frac{\sum_{i,j \epsilon E} |v_i - v_j|^2}{\sum_{i,j} |v_i - v_j|^2} \end{aligned}$$

$d_S(i,j)$ is the cut-metric corresponding to the cut $(S, \bar{S})$. The integers v_i are relaxed to be vectors in $\Re^d$ and the following SDP is obtained. (The dimension of the vectors d is obtained as part of the SDP solution.)

$$\begin{aligned} & \min_{v_i \epsilon \Re^d} \sum_{i,j \epsilon E} \| v_i - v_j \|^2 \\ s.t. \quad & \sum_{i,j} \| v_i - v_j \|^2 = n^2 \\ \forall i,j,k \quad & \| v_i - v_j \|^2 + \| v_j - v_k \|^2 \geq \| v_i - v_k \|^2 \end{aligned}$$

The cut metric $d_S(i,j)$ is relaxed to l_2^2 where l_2^2 is defined for two vectors v_i and v_j as $d_{l_2^2}(i,j) = \| v_i - v_j \|^2$. The reason for this particular relaxation is to obtain an SDP. The triangle inequality is not in the original problem, but is introduced as constraint in the SDP to make l_2^2 a metric. This constraint has the geometric interpretation that the angle between two points is not obtuse. Previously $d_S(i,j)$ was relaxed to any metric and a linear program was obtained. Then the solution (points in some metric space) was embedded into l_1 metric space relying on a theorem of Bourgain (for embedding any metric) into l_1 with $\log n$ distortion. In that case we used the standard definition of distortion (the product of maximum expansion and maximum contraction).

ARV embeds l_2^2 into l_1 with $O(\sqrt{\log n})$ AVERAGE distortion, and therefore achieves better approximation factor.

An embedding $d \to d'$ is non-expanding if for every (i,j) $d'(i,j) \leq d(i,j)$).

Average distortion of a non-expanding embedding $d \to d'$ is defined as $\frac{\sum_{i,j} d'(i,j)}{\sum_{i,j} d(i,j)}$.

Formally we have to use the following lemma:
Lemma: If the l_2^2 metric obtained by solving the SDP embeds in l_1 with average distortion D, then the approximation ration is D.

Note: By Frechet embedding with respect to a set S, we mean the distance from x to a set S (the distance to the closest point in S using l_2^2 metric). In this way each point x is mapped to one-dimensional point $d_{l_2^2}(x, S)$. The Frechet embedding is non-expanding meaning that $d_{l_1}(x, y) = |d_{l_2^2}(x, S) - d_{l_2^2}(y, S)| < d_{l_2^2}(x, y)$.

Now we state the Master Structure Theorem, both intuitively and formally:

Master Structure Theorem: If we spread points on a unit ball such that no two points subtend an obtuse angle, then we can find well-separated sets.

Formal statement:
Let $v_1, v_2, \ldots, v_n$ are unit vectors on B^d (by B^d we denote a unit sphere in d-dimensions) which satisfy $\frac{1}{n^2} \parallel v_i - v_j \parallel^2 = 1$ and the triangle inequality , then there exist sets S and T each of size $\Omega(n)$ and $d_{l_2}(S, T) = \Omega(\frac{1}{\sqrt{\log n}})$. The sets T and S are disjoint but their union does not contain all the points.

Before the proof of MST we show how to apply it.
Application of MST:
We can have the following two cases:

- there is super dense ball. Then we will use the Frechet embedding of the points w.r.t. the dense set.
- otherwise, there is a semi-dense ball. This is the core of the theorem. Here we find the sets S and T.

Case 1: Superdense ball
A super-dense ball is defined as one that has radius $\frac{1}{4}$ and contains at least $\frac{n}{4}$ points. Assume there is a super dense ball $B(p, \frac{1}{4})$ (i.e. $|B(p, \frac{1}{4})| > \frac{n}{4}$) We can embed the solution into l_1 using Frechet embedding w.r.t. the dense ball. In other words for each point v_i in the SDP solution, we define a new distance $d'(v_i, B)$ equal to the distance from point x to the closest point on the ball (using l_2^2 metric). Between two vertices i and j we use the distance $d'(i, j) = |d'(v_i, B) - d'(v_j, B)|$. We can show that $\sum_{i,j} d'(i, j) \geq \frac{1}{16} \sum_{i,j} d_{l_2^2}(i, j)$ which shows that the embedding average distortion is constant. Therefore in that case we have a constant approximation factor.
Details follow:

$$\begin{array}{rl} \sum_{i,j} d'(i,j) \geq & \\ \sum_{i \notin B, j \in B} d'(i,j) \geq & \textit{we are taking a smaller set} \\ |B| \sum_{i,j} d_{l_2^2}(i, B) & \textit{distance from point outside is bounded by distance to the ball} \end{array}$$

$|B| \geq \frac{n}{4}$ by assumption (ball is dense). We need to show that $\sum_{i,j} d_{l_2^2}(i, B) \geq \Omega(n)$. It follows from the inequality below.

$$\begin{array}{rl}
n^2 = \sum_{i,j} d_{l_2^2}(i,j) \le & \textit{constraint in SDP} \\
\sum_{i,j}\left(d_{l_2^2}(i,p) + d_{l_2^2}(p,j)\right) = & \textit{triangle ineq. to center of ball } p \\
2n\sum_{i,j} d_{l_2^2}(i,p) \le & \textit{symmetry of distance} \\
2n\left(\sum_{i,j}(d_{l_2^2}(i,B) + \frac{1}{4})\right) = & \textit{triangle ineq. to closest point on ball, radius of ball is } \dfrac{1}{4} \\
\frac{n^2}{2} + 2n\sum_{i,j} d_{l_2^2}(i,B) &
\end{array}$$

Case 2: Semi-dense ball
In that case there is no ball of radius $\frac{1}{4}$ which contains at least $\frac{n}{4}$ elements. This condition forces the points to be well spread-out.

From the SDP we have that $\frac{1}{n^2}\sum_{i,j} d_{l_2^2}(i,j) = 1$. Then there exists a point p such that $\frac{1}{n}d_{l_2^2}(p,i) \le 1$ (average distance to p smaller than 1). Let $B(p,2)$ be a ball around point p with radius 2. We claim that $|B(p,2)| \ge \frac{n}{2}$. Otherwise, the average distance to the center of the ball p will be greater than 1. We also claim that the average distance between points in $B(p,2)$ is constant. Formally we need to show that

$$\frac{1}{\textit{number of pairs in the ball}}\sum_{i,j} d_{l_2^2}(i,j) \ge \frac{1}{16} \in \Omega(1)$$

We can show that by considering that around each point v_i in the ball there a few points (at most $\frac{n}{4}$) points that are very close (at most $\frac{1}{4}$ distance away). This is because of the non-existence of a dense ball. We can conclude that at least $\frac{n}{2} - \frac{n}{4}$ points are at a distance greater than $\frac{1}{4}$. In other words, the average distance constraint forces us to have a large group of points in a large ball. The non-existence of semi-dense ball forces points to be well-spread out. So, we have a large number of points (at least $\frac{n}{2}$) that are at constant distance away from each other.
We want to apply the MST Theorem to the points in the ball $B(p,2)$. But to match the condition of the theorem that vectors are of unit length, we need to scale the vectors v_i by $\sqrt{2}$. Notice that we already match the other conditions of MST because we have $\frac{n}{2}$ points and the average distance is $\Omega(1)$. Application of MST gives us the sets S and T. Next the set S is used to construct a Frechet embeddng w.r.t. S, i.e. every point in the solution is mapped to one-dimensional point $d'(i,S) = d_{l_2^2}(i,S)$. This embedding is not-expanding. We need to show that the average distortion is $O(\sqrt{\log n})$. This is done in the following way:

$$\begin{array}{rl}
\sum_{i,j} d'(i,j) \ge \sum_{i\in S, j\in T} d_{l_2^2}(i,j) \ge & \textit{use distances only between points } S \textit{ and } T \\
|S||T|d_{l_2^2}(S,T) = & \textit{bound all distances by shortest distance} \\
\Omega(\frac{n^2}{\sqrt{\log n}}) = & \textit{use MST} \\
\Omega(\frac{1}{\sqrt{\log n}})\sum_{i,j} d_{l_2^2}(i,j) &
\end{array}$$

1 Proof of MST:

For simplicity we assume that the points are on the surface of a unit sphere (can be fixed)

We have n points such that

- the average distances in $\Omega(1)$
- no two points subtend an obtuse angle
- all points lie on an d-dimensional sphere.

We pick a random unit vector which we call u. Then we take a hyperplane orthogonal to it that passes through the origin. We call the points on one side of this hyperplane the set $\hat{S}$ and on the other side - the set $\hat{T}$. We also take two other hyperplanes on opposite sides and each is at a distance $\frac{\sigma}{\sqrt{d}}$ from the hyperplane passing through the origin (see the picture). σ is a constant and d is the number of dimensions of the sphere. $\sqrt{d}$ is related to the standard deviation of the Gaussian in d-dimensions. First we will argue that $|\hat{S}| = |\hat{T}| = \Omega(n)$ (in expectation over u). Then we address the issue about the sizes of S and T. The sizes of S and T are small, for example $0.05n$ (then of course $Mid_s \cup Mid_t$ is large, e.g. $0.95n$). The constant 0.05 depends on σ. However, S and T depend linearly on n (this is what is important for the theorem). We throw away the points in $Mid_s \cup Mid_t$ because they are very close. We want the sets S and T to be well-separated, meaning $d_2^2(S,T) = min_{i \in S, j \in T} d_{l_2^2}(i,j)$. In order to ensure this it is not enough to simply throw away the points in $Mid_s \cup Mid_t$. We need to consider all pairs of points i, j, such that $i \in S$, $j \in T$ and $d_{l_2^2} \leq l$, where l is constant (it will turn out to be equal to $\frac{1}{\sqrt{\log n}}$). Call this set K. One option is to throw away all points in those pairs but this will force S and T to become small in terms of n. So, we want to throw as few points as possible from S and T so that no pair of points from S and T is close. To that end we consider minimum vertex cover for the bipartite graph constructed from the set of points K (see picture). For bipartite graphs vertex cover is the same as maximum matching. The points from S are on one partition and the points in T are in the other. We put an edge between points $i \in S$ and $j \in T$ if $d_{l_2^2} \leq l$. We create a minimum vertex cover for this bipartite graph and remove the points in the cover. We have to show that the expected size of the vertex cover is small, where the expectation is over the vector u.
Our goals are to show the following:

- the points roughly split in half, i.e. $|S| = |T| \in \Omega(n)$.
- there are not too many points in the middle, i.e. in $Mid_S \cup Mid_T$ (left as an homework exercise).
- the points in S and T that were removed in the vertex cover step are not many.

To show $|S| = |T| \in \Omega(n)$, we first show that $|\hat{S}| = |\hat{T}| \in \Omega(n)$. The result for S will be implied by the size of $\hat{S}$ and the size of $Mid_S \cup Mid_T$. To show that the points are split in half we consider the probability that two points which subtend angle θ are split by hyperplane passing through the origin. This is the same as the probability that two points lying on circle are split by a line passing throught the origin. From geometrical considerations this proability is $\frac{\theta}{\pi}$.

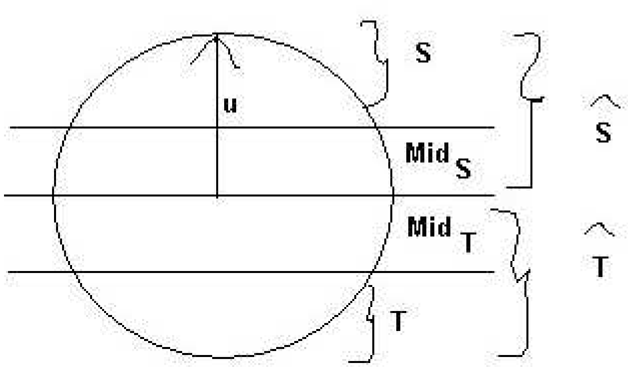
$$
\begin{array}{ll}
\Pr(\textit{points } i \textit{ and } j \textit{ are split by a random hyperplane}) = \frac{\theta}{\pi} \geq & \textit{geometrical considerations} \\
\frac{.878}{4} 2(1 - cos(\theta)) & \textit{from Max Cut Lecture} \\
\frac{.878}{4} \parallel v_i - v_j \parallel^2 & v_i \textit{ are unit vectors}
\end{array}
$$

Taking expectation we have that $E(\textit{split pairs}) = \sum_{i,j} Pr(i \textit{ and } j \textit{ are split}) \geq \frac{.878}{4} \sum_{i,j} \parallel v_i - v_j \parallel^2 = \Omega(n^2)$. Since $|\hat{S}| = |\hat{T}|$ and $|\hat{S}||\hat{T}| = \Omega(n^2)$ then $|\hat{S}| = |\hat{T}| = \Omega(n)$.

The difficult part of the proof is to show that the number of points that were removed from the set K is small. The proof of by contraction. We assume that the number of points that were removed is large ($\Omega(n)$). Then we show that if that is the case every point has a high chance to be picked in the matching (we speak of chance because of the vector u). But if every point has a high chance, the complete graph of the original n points has a dense part, which we call the core. In order for such a core to exist, by the so called Big Core Theorem, we need at least an exponential number of points in terms of σ and $\frac{1}{l}$. This will imply a contradiction. Here is the detailed proof of the contradiction. The vector u is chosen randomly, but after that our construction is deterministic. That is why with each u we can associate a matching M_u (for bipartite graphs matching is the same as vertex cover). The matching M_u is simply the set of edges that are picked to be removed from the set K. We are interested in $E(M_u) = \sum_{i,j} Pr(\textit{edge } (i,j) \textit{ is in the matching}) = \frac{1}{2}\sum_i(\sum_j p_{ij})$, where $p_{ij} =$ *prob. that edge (i,j) is picked in matching.* We will call $(\sum_j p_{i,j})$ the degree of vertex i, $deg(i)$, for the following reason. Imagine the complete graph with vertices the original points and edge weights equal to $Pr(\textit{edge } (i,j) \textit{ is in the matching})$. Then if for each vertex we sum the weights on the edges incident with this vertex we obtain what we called the degree of the vertex.

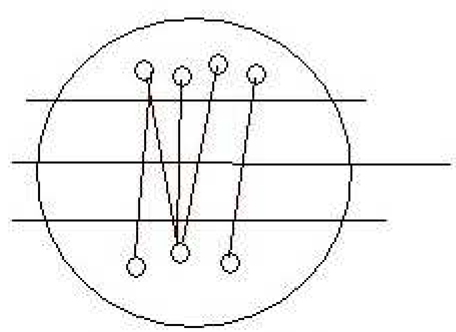
Let us assume for the sake of contradiction that $E(M_u) = \sum_i deg(i) = \Omega(n)$. Since the expectation is over all pairs (i,j), the average edgree of a vertex will be $\Omega(1)$. Let us call this average degree *avg*. Now imagine the complete graph with vertices the original points. We apply the following procedure. While there is a vertex whose degree (as defined earlier) is $< \frac{avg}{2}$ we remove that vertex. We claim that when the procedure terminates there are some vertices left. Each of the remaining vertices has degree $\geq \frac{avg}{2}$. We then use the Big Core theorem to obtain a contradiction. The Big Core theorem will imply that for such a configuration to exist one must have exponentially many points.
Finally we state the Big Core Theorem without proof.

The Big Core Theorem: We are given a set of points C (called the core) on a unit sphere and a matching of the points such that $p_x = Prob(\textit{ point } x \textit{ is in the matching}) \geq \delta$. Then the size of the set C (the core) is $\geq \exp(\Omega(\frac{\sigma^6}{l^4 \log \frac{1}{\delta}}))$.



MST Theorem Illustration

Picture 1



Removing points using matching (vertex cover).

Picture 2

References

[1] Sanjeev Arora, Satish Rao and Umesh Vazirani. *Expander Flows, Geometric Embeddings and Graph Partitioning.* April 2007.

[2] Avner Magen, Lecure Notes on Metric Embeddings, Lecture 10 *http://www.cs.toronto.edu/ avner/teaching/2414/LN10.pdf*