

Lecture 23 Semidefinite Programming/Sparsest Cut : 4.15.08

Lecturer: Satish Rao

Scribe: Rao, for now

Disclaimer: *These are rough notes, with some exercises.*

Again, recall the sparsest cut problem from the beginning of class. How do we find a provably optimal cut?

Question: How about starting with the spectral method?

Sure, recall Cheeger's inequality. That is,

$$\Theta(n\Phi^2/d) \leq \rho \leq \Phi,$$

or

$$\Theta(\alpha^2/d) \leq n\rho \leq \alpha,$$

where the Φ on the left could be found using the "spectral" algorithm (cut along sorted line.)

In the case α is approximately $\Theta(1)$. That is an expander graph. This leads to a constant factor approximation. When the graph is far from being an expander the approximation ratio is essentially the expansion. Thus, for graphs with good cuts, the algorithm could be quite far from being optimal.

Question: Can we do better? Linear programming?

Sure, in some cases. This gives an $O(\log n)$ approximation algorithm. Again, the approach was to solve the following linear program and produce a cut from its solution.

$$\max \sum_{i,j} d(i,j).$$

$$\sum_e d(e) = 1.$$

$d(i,j)$ is the distance between i and j under $d(\cdot)$.

The last constraint can be satisfied using triangle inequalities constraints which are linear.

Question: Intuitively?

Use a fixed amount of distance and make as high an average diameter as possible.

Question: How did we round?

Essentially, we showed that boundaries in a shortest path computation (breadth first search in a rounded graph) works. This uses a very easy lemma which basically says there is a "sparse" cut in high diameter graphs, basically inversely proportional to the diameter.

Question: Really? It's that easy?

No. We had to possibly make several breadth first cuts. The problem being that the "sparseness" in the "easy lemma" was with respect to the weight in the LP solution rather than the original nodes. Repeating and using an averaging argument fixed this right up.

Question: Can we do better?

Sure. Let's try semidefinite programming which by the way combines the two approaches.

23.1 Semidefinite programming.

Recall that one view of semidefinite programming is that it is choosing vectors v_1, \dots, v_n with linear inequalities and optimization function on the dot products of the vectors.

Question: Sparsest Cut Formulation?

$$\max_{i,j} d(i, j)$$

$$\sum_e d(e) = 1$$

$d(i, j)$ is at most the distance between i and j under $d(\cdot)$.

$$d(i, j) = |v_i - v_j|^2.$$

Question: Does it look familiar?

Indeed, this is just the linear program with the added condition that $d(i, j) = |v_i - v_j|^2 = v_i^2 + v_j^2 - 2v_i v_j$.

Question: Does it look familiar?

Or this is just the spectral method, (the Raleigh quotient) with the triangle inequality condition (and in higher dimension.)

23.2 Balanced partitioning.

To get the idea across, we are going to try to find a balanced partition (this avoids some uninteresting complexities, e.g., having to repeatedly cut, in the approximation algorithm.)

Question: Balanced partition?

We wish to find a cut where $|S|$ (the small side) has size at least cn and the number of edges is minimized.

Question: Semidefinite Program?

$$\min_e d(e)$$

$$\sum_{i,j} d(i,j) \geq 4c(1-c)n^2$$

$d(i,j)$ is at most the distance between i and j under $d(\cdot)$.

$$d(i,j) = |v_i - v_j|^2.$$

$$\|v_i\|^2 = 1.$$

Question: In words?

We place points on a sphere and try to get the average pair to be at least a certain distance while minimizing the squared euclidean length.

Question: Relaxation?

Sure, given a cut (S, \bar{S}) place all the nodes in S at a single point, and all the nodes in \bar{S} at the antipodal point which is 4 away. The sum over pairs of distances is $4|S||\bar{S}| = c(1-c)n^2$. The triangle inequalities hold (no triangle at all.) Thus, the solution is feasible.

The cost is 4 times the number of edges. Thus, the solution of the SDP over 4 is a lower bound on the balanced cut size.

Question: Goal?

Find a cut which is almost as balanced, i.e., $\Omega(n)$ nodes on each side, and with approximately as few edges. This is a “bicriteria” approximation, in that we relax on two fronts.

23.3 A Subgoal: Well-separated set.

Question: Subgoal?

We wish to find large sets that are far apart. Then we can use a diameter argument (as in the LP rounding algorithm) to find a small cut.

Question: How does that work?

If we can find two sets that each contain $\Omega(n)$ nodes that are far apart, say Δ .

Then, we know there is a cut between them of size $\sum_e d(e)/\Delta$ between them. Since $\sum_e d(e)$ is LP optimum, we have found a cut that separates $\Omega(n)$ nodes with cost that is within a factor of $1/\Delta$ of the SDP relaxation of the “balanced” cut problem.

23.4 Finding well-separated sets.

Question: Rounding?

Recall, Goemans and Williamson. Project onto line?

Question: Sure, is it as simple?

Well, no. They only need to reason over a single edge, since they simply wished to maximize the number of edges cut. Here, we need to deal with minimizing number of edges cut while maintaining a larger number of separated pairs.

23.4.1 Projections.

Question: Shall we learn a bit more about projections?

They behave like gaussians. In particular we have the following lemma.

Lemma 23.1 *Consider projecting a length v vector in d -dimensional space onto a random direction u . The expected length, $u \cdot v$ is $\mu = |v|/\sqrt{d}$. Moreover,*

$$Pr[|v| \geq \mu t] \leq e^{-t^2},$$

and

$$Pr[|v| \leq \epsilon t] \leq \epsilon.$$

Question: Why?

Let's look at the surface area of a d -dimensional unit sphere as one moves away from a great circle. How does the area fall away at distance δ ? The surface area of the remaining portion is less than the surface area of a sphere of the appropriate radius, i.e., using pythagoros's theorem, $r^2 = 1 - \delta^2$. That is, the surface area is at most $(1 - \delta^2)^{(d-1)/2}$ fraction of the total surface area.

Question: So?

Let $\delta = t/\sqrt{d}$, we get

$$(1 - t^2/d)^{(d-1)/2}.$$

And, once again, we see our old friend e , the base of the natural log, and get that the area of sphere (and the probability of a projection of a point being far away from the center is at most)

$$\approx e^{-t^2/2}.$$

Question: Centering and scaling, we get a part of the lemma. The decreasing probability of being long. The rest?

Another day, perhaps.

23.4.2 The well separated algorithm.

Question: So, let's project. Form partition?

One side of zero and the other, perhaps?

Not far apart, perhaps, at least $\pm f/\sqrt{d}$ apart in projection.

That is, we project onto a random direction, and choose S to be the set of nodes that are projected to less than f/\sqrt{d} and T to be the set of nodes that are projected more than f/\sqrt{d} .

Question: Each side has $\Omega(n)$ nodes?

Sure. The probability that a node falls within f is at most f . Thus, with reasonable probability most nodes fall outside and on either side.

Question: This is fishy?

One has to be slightly careful here about just reasoning with expectation. Also, the notion that points fall on either side must be argued.

First, we can choose the median point instead of the zero point, and take offsets on either side. Since the total expected distance is large (and moreover, large distances die away exponentially) the typical pair is separated along the line, and the two sets must be reasonably large.

Question: Is every point in S far from every point in T ?

Sort of. The lemma says that if there is a pair (i, j) of length $f/2\sqrt{\ln n}$, that the probability that i is in S and j is in T (i.e., they are projected farther apart than f/\sqrt{d}) is at most $e^{-t^2/2}$ where

$$t = f/2\sqrt{\ln n}/(f) = \sqrt{2 \ln n}.$$

That is, it is less than $1/n^2$.

Question: Are we done?

Sort of. We have produced two sets of $\Omega(n)$ nodes that are $1/\sqrt{\ln n}$ apart in euclidean length. Well, actually $d(i, j)$ is the squared euclidean length. So, we actually have only produced two sets that are

$$\Omega(1/\log n)$$

apart.

Question: Did we get something?

Sure, an $O(\log n)$ approximation algorithm.

23.4.3 The better well separated algorithm.

Question: Fix me up chappie?

Sure, let's do the same algorithm. Find S and T . Then, we repeatedly delete pairs $i \in S$ and $j \in T$ where $d(i, j) \leq g/\sqrt{\log n}$ where g is some constant.

Now, the sets are indeed $\Delta = \Omega(g/\sqrt{\log n})$ separated!

Question: Are we done?

Well, perhaps overdone. We might have deleted most of S and T !

Question: No we didnt!!

Why not? (The question and answers are switched but so be it.) Well, to prove that S and T remain large is not so easy. But we will endeavor to get there..

Question: If we do delete most of S and T in most directions, what have we built?

Well, this basically means that there are many pairs (i.e. $\Omega(1/\sqrt{n})$) which are stretched by a factor of $\ln^{1/4} n$. For any one pair this should happen with probability at most $e^{-\sqrt{\ln n}}$ (ignoring constants.) The notion that there $\Omega(n)$ such pairs seems unlikely.

On the other hand there is no independence here, thus it is difficult to argue that it doesn't happen.

Indeed, there is an example, where it basically does happen.

Question: What is the example?

Consider a $-1, +1$ hypercube in $d = \log n$ dimensions. If we scale by $1/\sqrt{\log n}$ the points are on the unit sphere.

We can think of a random projection in this case as a fixed easy to think about projection that sums the entries (or counts the number of $+1/-1$ ones.) Now, the sets S and T are those with around \sqrt{d} fewer or greater than $d/2, +1$'s.

Clearly, there are $\Omega(n)$ pairs that differ by around \sqrt{d} coordinates "crossing" between S and T . That is, the distance in the $1/\sqrt{\log n}$ on the sphere, yet they are projected $\Omega(1/\sqrt{d})$ apart.

Indeed, the hypercube shows that there are no well-separated sets with separation better than $\Delta = O(1/\sqrt{\log n})$.

Question: Can we argue not much worse happens?

It has been done. Today, we will set the parameter at $\Delta = \Theta(1/\ln^{2/3} n)$ and show that the project and delete does produce a Δ separated sets with constant probability.

Question: If not, what is the structure?

We get that that the following statement.

Matching Statement: For more than $1/2$ the directions u , there is a set, M_u , of $\Omega(n)$ disjoint pairs where for each pair $(i, j) \in M_u$,

$$d(i, j) \leq g/\log^{2/3} n,$$

and

$$(v_i - v_j) \cdot u \geq f/\sqrt{d}.$$

We reiterate that M_u is a matching.

Question: From a node's point of view?

For most nodes i , for a constant fraction of directions u , the node i has a mate j in matching M_u .

By deleting nodes where this does not happen, we can find a set V' of nodes with the following property.

Node Matching Statement: By deleting nodes where this does not happen, we can find a V' of nodes of size $\Omega(n)$. For all nodes in $i \in V'$, for a δ fraction of directions u , the node i has a mate $j \in V'$ in matching M_u .

Exercise 1: Show that Matching Statement implies the Node Matching Statement. Hint: set δ to be small with respect to constants involved in lower bound the $|M_u|$, and repeatedly delete nodes from V' where the conditions of the Node Matching Statement do not hold. When you delete a node you delete some small "measure" of node,direction pairs. Since there is a large "measure" of these, one cannot delete too many nodes. Play with this a bit.

Question: What?

All the nodes in V' are well covered in many δ fraction of directions. Meaning for most directions there is a close node that is far away in that direction.

Question: What now?

Induction, of course!!

Question: Well..how?

We define the graph induced by the pairs M_u as the matching graph. We define the following coverage concept.

Definition 23.2 *A subset of S nodes is said to be k -covered if for $1 - \delta/2$ directions each node i has a partner j_u in V' such that i is within distance k in the matching graph and*

$$|v_i - v_{j_u}| \cdot u \geq k\sigma/4,$$

where $\sigma = f/\sqrt{d}$ where f is from the projection algorithm and the definition of M_u .

The set V' is sort of 1-covered except that the nodes are only covered in δ fraction of the directions rather than $1 - \delta/2$ of the directions. We will fix this later.

Question: What can we do with this definition?

Well, if we can find a set S that is $k = \log^{1/3} n$ -covered. Then, we have that for most directions each node i in S has a mate j which is projected at least $\Omega(k\sigma)$ away. Moreover, the distance to that mate is k hops in the matching graph and thus $d(i, j) \leq k\Delta$. The Euclidean length is then $\sqrt{k}\sqrt{\Delta}$. The stretch of this projection is thus $\sqrt{k}/\sqrt{\Delta} = \Omega(\sqrt{\log n})$.

By choosing constants, we can make this occur with probability less than $1/n^2$. This contradicts the notion that such an S exists.

Question: What now?

We will try to prove such an S does exist given that the set of matchings M_u exists. This contradicts the existence of the set of matchings and the failure of the project and delete algorithm.

Question: What is the induction?

Lemma 23.3 *Given the set M_u , and a k -covered set $S \subset V'$ there is a set S' of size $\delta|S|$ that is $k+1$ -covered, for $k = O(\log^{1/3} n)$.*

Question: The base case?

The set V' is almost 1-covered except that it is not covered in most $(1 - \delta/2)$ directions. It is only covered in δ directions.

Question: How to extend probability of coverage?

Well, consider that node i is covered by j in direction u . The projection length remains large for directions u' that don't differ from u by much. The loss is at most $|v_i - v_j|\sin\theta$.

Question: For small values of θ what is $\sin\theta$?

Well, the distance on the unit sphere. So, now let's take close vectors to be those within $\theta \approx 2\sqrt{\log(2/\delta)}/\sqrt{d}$.

Moreover, let's ensure that $|v_i - v_j| \ll \sigma/2\sqrt{\log(2/\delta)}$. Initially, this holds by choosing Δ to be a sufficiently small and holding the number of iterations down.

Here, the loss in projection length is much less than σ/\sqrt{d} .

Question: What's the conclusion?

Well, we are still 1-covered for these u' directions that are within the θ defined above.

This is from the fact that 1-covering had a bit less of a requirement on the projection lengths than the matchings M_u provided.

Question: How large of a set are the u' ?

Well, the worst set of u is when they are all concentrated on a polar cap of area δ . This is fancy theorem called "measure concentration." Something akin to the notion that the minimal surface area shape is the sphere.

Now, from the gaussian nature of the surface area it is easy to see that with each addition $1/\sqrt{d}$ distance from this cap along the sphere one drastically increases the area (basically the area is proportional to $e^{-(t-1)^2/2}$ versus $e^{-(t^2/2)}$. For example, always by a factor of around e^{-t} . Thus, one quickly gets to half the ball. After that, one similarly decreases the remaining area with each additional $1/\sqrt{d}$ that one grows.

So, the set of the u' is a $1 - \delta/2$ fraction of all the directions.

Question: On a side note, why is this not so intuitive?

It's just in high dimensions, almost all the area is at the middle of a sphere. Difficult for me as you to visualize in two or even three dimensions.

Question: Induction step?

Consider a node inS that is k covered. Consider a direction u . With probability $1 - \delta/2$ it has a mate j with projection length $k\sigma/4$ in direction u . Moreover, with probability δ it has a mate j' in direction $-u$ with projection length σ .

Thus, the pair (j', j) has projection length $k\sigma/4 + \sigma$ along u .

That is, the pair is (sort of) $k + 1$ and more covered for u .

For each node, we produce pairs for $\delta/2$ fraction of the directions.

Moreover, for each j' it only participates in one pair for each direction! (This is not true of the j 's.) Here we use the property of the matching cover!! VERY IMPORTANT POINT.)

Question: We donate the pair to the j' . What's interesting about the j' ?

It only gets one donation for each direction. The j' 's come from V' nodes. Thus, the typical node in V' gets $\delta|S|/2/|V'|$ donations.

Thus, there are at least $\delta|S|/4$ nodes in V' that are covered in $\Omega(\delta|S|/|V|)$ directions.

Let's call this S' . This is well-covered, i.e., has larger projection length (by σ) but for a small fraction of the directions.

Question: Are we done?

No, we are not covered in $1 - \delta/2$ of the directions.

Question: What do we do?

We use measure concentration to give up a bit of projection length (we added σ and only need $\sigma/4$) and use measure concentration to cover more directions as we did in the base case.

Question: Well, what is the error that limits us to $k = \log^{1/3} n$?

Well, in each inductive step, the size of the k -covered set is reduced rather drastically, thus measure concentration needs to be used to boost the probabilities. Thus, imposes limits on the vector lengths that one can work with (since the longer the vector the greater the loss in projection when adding directions on the sphere.)

If one works out the calculations, we can run the induction up to $k = \theta(\log^{1/3} n)$.

Question: What now?

This reduction can be combated using another idea or two ... but alas that is for another time or place.