CS270 Algorithms & Data Structures **Spring 2008**

Lecture 24 Semidefinite Programming/Sparsest Cut : 4.17.08

Lecturer: Satish Rao *Scribe: Rao, for now*

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

24.1 Reviewing last time...

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 $d(\cdot)$.

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

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

24.2 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: Recall the best we can do. The hypercube...

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: Last time, I forgot. What else is special about the hypercube?

It obeys the triangle inequality on the squared Euclidean distance (since that is exactly the metric in the graph.)

Exercise 1: Show that any ℓ_1 metric on a set of points can be acheived by an ℓ_2^2 metric.

24.3 Projections.

Question: How does projected length relate to original Euclidean length?

Lemma 24.1 *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

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

24.3.1 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 +/- $f/\sqrt{d}$ 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 $f/\sqrt{d}$.

Then, we delete pairs of nodes (i, j) with $i \in S$ and $j \in T$ where $d(i, j)$ is less than Δ

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. This is intuitive the argument is just slightly more subtle.

Question: Last time where did we get in our analysis.

We observer that if the algorithm fails we get a set V' of points that are well-covered; for each point $i \in V'$ for δ fraction of directions u there is a $j \in V'$ where

$$d(i, j) \leq \Delta.$$

and

$$(v_i - v - j) \cdot u \geq \sigma.$$

(Recall $\sigma = f/\sqrt{d}$ for some constant f which depends on the balance c of the cut we are trying to approximate.)

We defined the set of pairs as M_u. The union of such pairs is called the matching graph.

Question: Then, we had an inductive extension of this definition. What was it?

A set S of points is k-covered if for each point $i \in S$ for $1 - \delta/2$ fraction of directions u there is a $j \in V'$ where i and j are within distance k in the matching graph

$$(v_i - v - j) \cdot u \geq k\sigma/4.$$

Question: Why is this useful?

Recall, if we can get a $\log^1 /3n$-covered set for $\Delta = 1/\log^{2/3} n$, we get a contradiction.

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 become a large constant times $\log n$ and the probability of any pair being so stretched is $1/n^c$. This contradicts the notion that there is even one node in that is covered in most directions.

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 24.2 *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 any $k \leq C\log^{1/3} n$ for some constant C.*

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 $cos\theta$?

Well, the distance on the unit sphere. So, now let's take close vectors to be those within $2\log(1/\delta)/\sqrt{d}$.

Moreover, let's assume that $|v_i - v_j| << \sigma/2\log(1/\delta)$.

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

Question: What's the conclusion?

Well, we are still 1-covered for these u'. 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.

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 grows the area by e (or shrinks the remaining area by e).

Thus, after $\log(1/\delta)$ such "growings" we get half the sphere and after another $\log(2/\delta)$ we only leave $\delta/2$ fraction of the sphere uncovered.

Question: Induction step?

Consider a node $iinS$ 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 limits us to $k = \log^{1/3} n$?

Well, in each inductive step, the size of of the k-covered set is reduced rather drastically, thus measure concentration needs to be used to boost the probabilities by much more than $1/\delta$. Indeed, by a factor of $1/\delta^k$.

Question: What is the requirement on θ then?

Well, it should be at least $\Theta(\sqrt{k}\sigma/\log(1/\delta))$.

Question: How does it affect things?

Well, the longer the vector the greater the loss on projection length.

Thus, we get a condition on the length of the vectors used (the shorter the vector the greater the angle of good directions.)

At step k, the vectors have length $\sqrt{k}$ as large. Moreover, they should remain of length at most $\Theta(1/\sqrt{k})$.

The vectors started out to be $1/\sqrt{\Delta}$ in length.

Thus, we can continue as long as $1/\sqrt{k} \geq \sqrt{k}/\sqrt{\Delta}$ ignoring constants involving $\log(1/\delta)$.

This gives that

$$k \leq 1/\sqrt{\Delta}. \tag{24.1}$$

Question: What is the probability that a vector is stretched at the kth level?

Well, again it has euclidean length $\sqrt{k/\Delta}$, and projection length $k\sigma$. (where σ is basically the expected projection length of a unit vector.)

Thus, it is at most $e^{-k/\Delta}$.

Exercise 2: Ignoring constants that depend on δ, how do we (asymptotically) set Δ to get this probability to be inverse polynomial? Hint: use equation 24.1 and the above upper bound on the probability.

Question: Are we done?

Sure, we can carry out the induction given the M_u's. But the induction leads to nonsense. Thus, the M_u's must not exist, and the algorithm could not have failed.

Question: How do we improve the analysis?

Basically, ensure that S remains large at each step. Needs some more ideas...

24.4 Embeddings.

Question: Do you recall the relationship between the linear programming sparse cut and embedding?

Sure, a general theorem by Bourgain implied the result easily. That is, Bourgains theorem was any metric space can be embedded into ℓ_1 metrics (i.e., point in R^d with distance that is the sum of the coordinate differences) with $O(\log n)$ distortion. An observation that ℓ_1 metrics basically correspond to a set of cuts (i.e., each coordinate can be represented as a set of cuts corresponding to scanning along the line) and the metric is “separable” along the coordinates.

Bourgain’s original embedding, embedded into Euclidean metrics which is perhaps more natural.

Question: Is the $O(\log n)$ required?

In general, yes. This follows from the example of an expander graph. The LP solution is a factor of $\Theta(\log n)$ away from the cut size.

Thus, Bourgain could not have done better.

Question: What kind of metric did we use above?

A metric acheivable by points in space where distances are squared Euclidean distances. An "ℓ_2" squared metric!

Question: Can this be embedded into a normal metric, say Euclidean?

Sure. Bourgain does it with distorition $O(\log n)$. Extending above technology and using many other ideas about embedding, one can embed into ℓ_2 with $O(\sqrt{\log n}\log\log n)$.

Question: So?

Well, mathemeticians were trying to understand whether ℓ_1 metrics can be embedded into ℓ_2 metrics. Since they can easily be embedded isometrically (no distance changes) into ℓ_2^2 metrics, we can conclude there is an upperbound an $O(\sqrt{\log n}\log\log n)$ upper bound. Previously, the hypercube was shown to require $\Omega(\sqrt{\log n})$ to be embedded. So, this is "close" to the final result.

Exercise 3: Show that ℓ_1 metrics can be embedded into ℓ_2^2 metrics. Hint: show that ℓ_1 metrics can be embedded into a hypercube like metric, i.e., there are only two possible values for each coordinate (also called cut metrics.)