CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 9: Small-set expansion

We will discuss the small-set expansion conjecture and its connection to the unique games conjecture.

We will study subexponential time algorithms to solve these problems, using the eigenspace enumeration technique and a higher eigenvalue bound for small-set expansion.

Small-set expansion conjecture and unique games conjecture

Let $\phi_\delta(G) := \min_{S : |S| \leq \delta |V|} \phi(S)$ be the δ-small-set expansion of G.

We are interested in finding a good approximation algorithm for δ-small-set expansion.

One motivation for this problem is its potential applications in finding small clusters as we discussed last time.

Another motivation to study this problem is from the small-set expansion conjecture by Raghavendra and Steurer [RS10], which roughly says that the performance ratio of any polynomial time algorithm that approximates $\phi_\delta(G)$ must depend on δ or n.

More formally, the conjecture states that for any ε, there exists δ such that it is NP-hard to distinguish the following two cases:

1. There exists $S \subseteq V$ with $\phi(S) \leq \varepsilon$ and $|S| \leq \delta n$.
2. $\phi(S) \geq 1-\varepsilon$ for every set $S \subseteq V$ with $|S| \leq \delta n$.

If this conjecture is true, then the unique games conjecture (which we will define next) is true, and this would imply optimal inapproximability results for a few problems, including the 0.878-approximation for max cut and the 2-approximation for minimum vertex cover.

If the conjecture is false, then probably the unique games conjecture is false (no formal proof for this direction), and hopefully the techniques in disproving it would lead to improved approximation algorithms.

Note that the local graph partitioning algorithms discussed last time are bicriteria approximation algorithms for the small-set expansion problem, but they could not distinguish the two cases because there is a $\sqrt{\log n}$ term in the approximation ratio, and so $\sqrt{\varepsilon \log n} \geq 1-\varepsilon$ for large enough n.

There are also SDP-based approximation algorithms for the small-set expansion problem with stronger approximation guarantees, but their approximation ratios still depend on n or δ (in the form of $\sqrt{\log n}$ or $\sqrt{\log \frac{1}{\delta}}$).

The small-set expansion conjecture is still wide open.

The uniques game conjecture can be stated as follows.

Consider the problem of linear equations over two variables modulo R, i.e. $X_3 + X_5 \equiv 2 \pmod{R}$, etc.

The conjecture by Khot [K02] states that for any ε there exists R s.t. it is NP-hard to distinguish the following:

① There exists a solution satisfying $(1-\varepsilon)$-fraction of equations.

② No solutions satisfying more than ε-fration of equations.

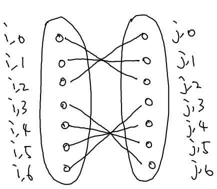
Informally, the conjecture is saying that the problem becomes harder when R is larger.

To get an idea of the connection between small-set expansion and unique games, we consider the "label-extended graph" of a unique games instance.

In an "ordinary" graph of unique games, there is a vertex for each variable, and there is an edge between two vertices if there is a constraint between the two variables.

In the label extended graph, each variable has R vertices, one for each label - from 0 to $R-1$.

For a constraint, say $X_i + X_j \equiv 2 \pmod 7$, we add a perfect matching between the two sets of R vertices (in the example, $R=7$), where an edge corresponds to a valid way to satisfy the constraint.



("unique games" means for each value of one variable, there is a unique value for the other variable to satisfy the constraint, and so there must be a perfect matching between the two sets of R vertices)

Suppose there is an assignment satisfying $\geq 1-\varepsilon$ fraction of constraint (say $X_1 = 2$, $X_2 = 5$, $X_3 = 4$), it corresponds to a subset of $\frac{1}{R}$-fraction of vertices in the label-extended graph $((1,2),(2,5),(3,4))$, and note the important point that the conductance of this set is $\leq \varepsilon$.

So, if there is a good assignment, then there is a small non-expanding set.

The other direction is not necessarily true, since a small non-expanding set may not correspond to an assignment, as the set may contain more than one vertices (or none) in a "cloud".

However, in many "hard" instances (note that, of course, there is no known NP-hard instances), finding a small sparse cut is useful in finding a good assignment.

Intuitively, the above discussion suggests that there is a reduction from unique games to small-set expansion.

Raghavendra and Steurer [RS10] proved the opposite direction, by showing a reduction from small-set expansion to unique games. thus proving that the small-set expansion problem is easier.

Also, under the assumption that the ordinary graph of unique games is a small-set expander, there is a reduction from unique games to small-set expansion.

So, these two problems are considered to be closely related.

Subexponential time algorithm

The main goal today is to present the subexponential time algorithms by Arora, Barak, Steurer [ABS 10] to distinguish the two cases.

We will study the small-set expansion algorithm in details, and mention how it can be extended to unique games.

The approach is by looking at the Laplacian spectrum of the graph.

If there are many small eigenvalues, we will show that random walks can be used to find a small sparse cut in polynomial time.

If there are not too many small eigenvalues, then we will use the subspace enumeration method first proposed by Kolla to find a small sparse in subexponential time.

These can be combined to obtain a subexponential time algorithm to distinguish the two cases in the small-set expansion conjecture.

To distinguish the two cases in the unique games conjecture, the higher eigenvalue bound is used recursively to break the graph into pieces, where each piece has few small eigenvalues, and the total number of deleted edges in bounded. Then, each piece is solved by the enumeration method.

We assume the graph is d-regular once again.

Higher eigenvalues and small-set expansion

The key technical contribution in [ABS10] is to show that if there are many small eigenvalues, then there is a small sparse cut.

We have seen the higher-order Cheeger's inequality, which says that there are $\Omega(k)$ disjoint sparse cuts with conductance $O(\sqrt{\lambda_k \log k})$, and thus there is such a sparse cut with $O(\frac{n}{k})$ vertices.

The $\sqrt{\log k}$ term is unavoidable (may see why later), and this term is preventing us to use higher-order Cheeger's inequality for disproving the small-set expansion conjecture.

Actually, the higher-order Cheeger's Inequality was inspired by the result in [ABS 10], and so was the spectral analysis of local graph partitioning we seen last time.

We will show that for large enough k (e.g. $k = n^{0.1}$), there is a set of size $\approx \frac{n}{k}$ with conductance $\approx \sqrt{\lambda_k}$.

To do this, we use the results proved last time, which were developed in [ABS10].

Lemma 1 $\quad R(W^t x_i) \leq 2 - 2\|W^t x_i\|_2^{\frac{1}{t}}$, where $R(x) = \frac{x^T \mathcal{L} x}{x^T x}$.

Lemma 2 $\quad$ If $\|W^t x_i\|_2^2 \geq \frac{1}{N}$, then there is a set $S \subseteq V$ with $|S| = O(N)$ and $\phi(S) = O\left(\sqrt{R(W^t x_i)}\right)$.

The first lemma is just the same Lemma 1 in L08, while the second lemma is a combination of Lemma 2 in L08 (with $x = W^t x_i$ and $\delta n = N$) and Cheeger's rounding.

Last time, we proved a lower bound on $\|W^t x_i\|$ using staying probability.

Today, we prove a lower bound on $\|W^t x_i\|$ using higher eigenvalues.

Suppose $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k \leq \lambda$.

Recall that $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A} = \frac{1}{2}I + \frac{1}{2}(I - \mathcal{L}) = I - \frac{1}{2}\mathcal{L}$. Call the eigenvalues of W $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$.

Our assumption $\lambda_k \leq \lambda$ translates to $\alpha_k \geq 1 - \frac{\lambda}{2}$.

Note that
$$\sum_{i=1}^{n} \|W^t x_i\|_2^2 = \sum_{i=1}^{n} x_i^T W^{2t} x_i = \mathrm{Tr}(W^{2t})$$
$$= \sum_{i=1}^{n} \alpha_i^{2t} \quad \text{(since trace = sum of eigenvalues)}$$
$$\geq k\left(1 - \frac{\lambda}{2}\right)^{2t} \quad \text{(by our assumption } \lambda_k \geq 1 - \tfrac{\lambda}{2}\text{)}.$$

Therefore, there exists a vertex i with $\|W^t x_i\|_2^2 \geq \frac{k}{n}\left(1 - \frac{\lambda}{2}\right)^{2t}$ by averaging.

If $\|W^t x_i\|_2^2 \geq \frac{k}{n}$, then we would get a set with size $O(\frac{n}{k})$, and this would be ideal.

We can't quite do that, but we try to get something close, so that the size would be $O(\frac{n}{k^{1-c}})$ for some absolute constant c by applying Lemma 2.

For simplicity, we just set t so that $\|W^t x_i\|_2^2 \geq \frac{\sqrt{k}}{n}$, so that the output set is of size $O(\frac{n}{\sqrt{k}})$.

For this, set $t = \frac{\ln k}{2\lambda}$, so that $\|W^t x_i\|_2^2 \geq \frac{k}{n}\left(1 - \frac{\lambda}{2}\right)^{2t} \approx \frac{k}{n} e^{-\lambda t} = \frac{k}{n} e^{-\frac{\ln k}{2}} = \frac{\sqrt{k}}{n}$.

By Lemma 1,
$$R(W^t x_i) \leq 2 - 2\|W^t x_i\|_2^{\frac{1}{t}} \leq 2 - 2\left(\frac{\sqrt{k}}{n}\right)^{\frac{1}{2t}} = 2 - 2\left(\frac{\sqrt{k}}{n}\right)^{\frac{\lambda}{\ln k}} \leq 2 - 2\left(\frac{1}{n}\right)^{\frac{\lambda}{\ln k}}$$
$$= 2 - 2e^{-\frac{\lambda \ln n}{\ln k}} \leq \frac{2\lambda \ln n}{\ln k} \quad (\text{using } e^{-x} \geq 1 - x)$$

Therefore, using this choice of t, by Lemma 2, we get a set S with $|S| = O(\frac{n}{\sqrt{k}})$ and $\phi(S) = O\left(\sqrt{\frac{\lambda \ln n}{\ln k}}\right)$.

and $\quad \phi(S) = O\left(\sqrt{\frac{\lambda \ln n}{\ln k}}\right)$.

For the small-set expansion conjecture, we couldn't afford to have a dependency on $\ln n$.

So, we will choose $k \geq n^{2\beta}$ for some constant β, so that $|S| = O(n^{1-\beta})$ and $\phi(S) = O\left(\sqrt{\frac{\lambda}{\beta}}\right)$.

This is what we will do, and let us record it down.

Corollary 1 For $k \geq n^{2\beta}$, there is a set S with $|S| = O(n^{1-\beta})$ and $\phi(S) = O\left(\sqrt{\frac{\lambda_k}{\beta}}\right)$.

The above corollary implies that if $\lambda_{n^{2\beta}} \leq \varepsilon$ say, then we must be in the first case (the "yes" case) of the small-set expansion conjecture.

Subspace enumeration

The above section shows that if there are many small eigenvalues, then there is always a small sparse cut, and this becomes a "trivial" case for the small-set expansion conjecture (although the proof was "non-trivial").

What if there are few small eigenvalues?

The idea, introduced by Kolla, is to simply do an "exhaustive search" in the top eigenspace, when its dimension is low enough.

Small distance to top eigenspace

Let χ_S be the characteristic vector of S normalized to norm one: $\chi_S(i) = \begin{cases} \frac{1}{\sqrt{|S|}} & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$.

We can write $\chi_S = \sum_{i=1}^{n} c_i v_i$ where v_i are the eigenvectors of $\mathcal{L}$.

Suppose S is a set with conductance $\leq \varepsilon$. Then we know that $R(\chi_S) \leq \varepsilon$.

Suppose $\lambda_{k+1} > \lambda$. Think of λ to be considerably greater than ε, say $\lambda = \sqrt{\varepsilon}$.

The claim is that χ_S has a large projection to the top eigenspace $\text{span}\{v_1, v_2, \ldots, v_k\}$.

We have
$$\begin{aligned} \varepsilon \geq \frac{\chi_S^T \mathcal{L} \chi_S}{\chi_S^T \chi_S} &= \Big(\sum_{i=1}^{n} c_i v_i\Big)^T \mathcal{L} \Big(\sum_{i=1}^{n} c_i v_i\Big) && \text{as } \|\chi_S\|_2^2 = 1 \\ &= \sum_{i=1}^{n} c_i^2 \lambda_i && \text{by orthonormal eigenvectors} \\ &\geq \sum_{i=k+1}^{n} c_i^2 \lambda. \end{aligned}$$

So, $\sum_{i=k+1}^{n} c_i^2 \leq \frac{\varepsilon}{\lambda}$.

Let $y = \sum_{i=1}^{k} c_i v_i$. Clearly $y \in \text{span}\{v_1, \ldots, v_k\}$.

Then $\|\chi_S - y\|_2^2 = \|\sum_{i=k+1}^{n} c_i v_i\|_2^2 = \sum_{i=k+1}^{n} c_i^2 \leq \frac{\varepsilon}{\lambda}$.

So, there is a vector y with $\|y\| \le 1$ in the top eigenspace with small distance to x_S.

Small conductance set from y

Next, we claim that we can find a somewhat sparse cut from such y.

Consider $S' := \{ i \mid y_i \geq \frac{1}{2\sqrt{|S|}} \}$. We argue that the symmetric difference between S and S' is small.

Notice that, for each vertex $i \in S - S'$ or $i \in S' - S$, we have $|x_S(i) - y_i| \geq \frac{1}{2\sqrt{|S|}}$. It follows from

the fact that x_S is a binary-valued vector (either $\frac{1}{\sqrt{|S|}}$ or 0) and the definition of S'.

Therefore, $\frac{\varepsilon}{\lambda} \geq \|x_S - y\|_2^2 \geq \sum_{i \in (S-S') \cup (S'-S)} (x_S(i) - y_i)^2 \geq |(S-S') \cup (S'-S)| \cdot \frac{1}{4|S|}$, and hence

$|(S-S') \cup (S'-S)| \leq \frac{4\varepsilon}{\lambda} |S|$.

Now, $\phi(S') = \frac{|\delta(S')|}{d|S'|} \leq \frac{|\delta(S)| + d \cdot |(S-S') \cup (S'-S)|}{\frac{1}{2} d |S|}$ ← the worst case is when vertices in $S'-S$ have all edges going out

← we assume that $\lambda \geq 8\varepsilon$ so that the volume changes little.

$\leq 2\phi(S) + \frac{8\varepsilon}{\lambda}$.

So, from y, we can extract a set S' with $|S'| \leq 2|S|$ and $\phi(S') = O(\frac{\varepsilon}{\lambda})$.

By choosing λ to be larger, we get a better approximation, but as we will see next it is

computationally more expensive to choose larger λ.

Eventually, we will choose λ to be ε^{1-c} for some constant $0 < c < 1$, say $\lambda = \sqrt{\varepsilon}$.

Searching over the top eigenspace

From the above discussions, there is a vector y in the eigenspace with eigenvalue at most λ,

from which we can extract a set S' with size and conductance similar to S, our target set.

The question is how to find such a vector y efficiently?

Actually, we don't know of a really good way (otherwise the small-set expansion conjecture will be disproved),

but the observation is that if the top eigenspace is of low dimension, then essentially an exhaustive

search would not be too slow.

Our task is to find $y \in \text{span}\{v_1, \ldots, v_k\}$ with $\|y\| \leq 1$ such that $\|x_S - y\|_2^2 \leq O(\frac{\varepsilon}{\lambda})$.

Epsilon net

We can not search for the entire subspace, as there are infinitely many points in the subspace.

We will discretize the search space, and find a set of representatives $u_1, u_2, \ldots, u_M$ such that

for every $v \in \text{span}\{v_1, \ldots, v_k\}$, there exists u_i such that $\|v - u_i\|_2 \leq \alpha$ for a given α.

This is called an α-net in the literature (usually called ε-net but we reserved ε for the conductance).

How many representatives we need to pick to "cover" the whole k-dimensional unit ball?

Let's pick the representatives greedily.

We start from an emptyset. We add a point p to the representative set if the distance between p and any of the current representatives is at least α. We stop when all the points are of distance at most α from some representative. By construction, the representatives form an α-net.

The crucial observation is that the representatives are of distance at least $\frac{\alpha}{2}$ to each other, as otherwise they won't be added to the set. So, the $\frac{\alpha}{2}$-balls with v_i as center are disjoint.

Each $\frac{\alpha}{2}$-ball has volume $C_d \cdot (\frac{\alpha}{2})^d$ where C_d is a constant depending on d only, and the total volume of these $\frac{\alpha}{2}$-balls are $M \cdot C_d \cdot (\frac{\alpha}{2})^d$.

On the other hand, all these balls are contained in the $(1+\frac{\alpha}{2})$-ball in d-dimension, which has volume $C_d \cdot (1+\frac{\alpha}{2})^d \leq C_d 2^d$ assuming $\alpha \leq 2$ (which is valid in our application later).

So, we have $C_d \cdot 2^d \geq M \cdot C_d \cdot (\frac{\alpha}{2})^d$, which implies that $M \leq (\frac{4}{\alpha})^d$.

<u>Searching over epsilon-net</u>

In our application, we set $\alpha = \sqrt{\frac{\varepsilon}{\lambda}}$, so that we can find a vector $y' \in \text{span}\{v_1, \dots, v_k\}$ with $\|y - y'\|_2^2 = \frac{\varepsilon}{\lambda}$, and thus $\|x_S - y'\|_2^2 = O(\frac{\varepsilon}{\lambda})$. and we can extract a set S' from y' with $|S'| \leq 2|S|$ and $\phi(S') = O(\frac{\varepsilon}{\lambda})$.

The running time is $O(M \cdot m) = O(m \cdot (\frac{4}{\alpha})^k) = O(m \cdot (\frac{\lambda}{\varepsilon})^k)$.

To summarize. we have the following result.

<u>**Lemma**</u> (subspace enumeration) Suppose $\lambda_k \leq \lambda < \lambda_{k+1}$ and there is a set S with $\phi(S) \leq \varepsilon$. The subspace enumeration algorithm can find a set S' with $|S'| \leq 2|S|$ and $\phi(S') \leq O(\frac{\varepsilon}{\lambda})$ with running time $O(m \cdot (\frac{\lambda}{\varepsilon})^k)$.

<u>**Subexponential time algorithm for small-set expansion**</u>

We now combine the higher eigenvalue bound and the subspace enumeration method to distinguish the two cases in the small-set expansion conjecture.

Suppose there is a set S with $|S| \leq \delta n$ and $\phi(S) \leq \varepsilon$, where δ is some absolute constant.

Suppose $\lambda_k \leq \lambda < \lambda_{k+1}$ where λ is a parameter that we will choose soon.

If $k \geq n^{2\beta}$ for some constant β that we will choose soon, by the higher eigenvalue bound in Corollary 1,

there is a set S' with $|S'| = O(n^{1-\beta}) << \delta n$ and $\phi(S') = O\left(\sqrt{\frac{\lambda_k}{\beta}}\right) = O\left(\sqrt{\frac{\lambda}{\beta}}\right)$.

On the other hand, if $k < n^{2\beta}$, we can run the subspace enumeration method to find a set

S' with $|S'| \leq 2|S|$ and $\phi(S') = O\left(\frac{\varepsilon}{\lambda}\right)$ in time $O\left(m \cdot \left(\frac{1}{\varepsilon}\right)^{n^{2\beta}}\right) = O\left(m \cdot \exp\left(n^{2\beta} \log\left(\frac{1}{\varepsilon}\right)\right)\right)$.

Naturally, we will choose λ and β in such a way to balance the conductance term,

such that $\frac{\varepsilon}{\lambda} = \sqrt{\frac{\lambda}{\beta}}$ or equivalently $\beta \varepsilon^2 = \lambda^3$.

We do so casually by setting $\lambda = \varepsilon^{\frac{3}{4}}$ and $\beta = \varepsilon^{\frac{1}{4}}$.

Then, we have the following theorem.

__Theorem__ Suppose there is a set S with $|S| \leq \delta n$ and $\phi(S) \leq \varepsilon$.

There is an algorithm with running time $\exp(n^{O(\varepsilon^{\frac{1}{4}})})$ that returns a set S' with $|S'| \leq 2\delta n$ and $\phi(S') = O(\varepsilon^{\frac{1}{4}})$.

__Remark:__ The size of S' can be made arbitrarily close to one, i.e. $|S'| \leq (1 + \frac{\varepsilon}{\lambda})|S| \leq (1 + \varepsilon^{\frac{1}{4}})|S|$

We can delete extra vertices until the size is $\leq \delta n$ and the conductance is still $O(\varepsilon^{\frac{1}{4}})$.

To summarize, we set $\lambda = O(\varepsilon^{\frac{3}{4}})$, if there are more than $n^{2\beta} = n^{O(\frac{1}{\varepsilon^4})}$ eigenvalues smaller than λ,

then we use the polynomial time random walks algorithm and we are done. Otherwise, we run the

subspace enumeration algorithm on the eigenspace with value $\leq \lambda$, and do a little modification as in the remark.

__Unique games conjecture via graph decomposition__ (optional)

In the following, we sketch how to extend the ideas to distinguish the two cases for unique games.

Take the ordinary graph for unique games.

We will repeatedly apply the higher eigenvalue bound to decompose the graph.

If there are $n^{2\beta}$ eigenvalues smaller than λ, then we find a set S with $|S| = O(n^{1-\beta})$ and

$\phi(S) \leq O(\frac{\varepsilon}{\lambda})$. And we cut S off the graph.

We repeat the same procedure until the remaining piece has fewer than $n^{2\beta}$ eigenvalues smaller than λ.

At that time, we have $S_1, S_2, \ldots, S_\ell$ where each piece is of size $O(n^{1-\beta})$ and the remaining piece

has few small eigenvalues.

We call this one level of decomposition.

An important point is that we only delete an $O(\frac{\varepsilon}{\lambda})$ fraction of edges in the whole graph in one level of decomposition, by the definition of conductance (comparing to the remaining edges).

We apply this decomposition procedure to each component S_i, and break it into pieces of size $O(n^{1-2\beta})$ and the remaining piece with few eigenvalues.

And then we apply the same decomposition procedure, until each piece is of size say $n^{2\beta}$ or has fewer than $n^{2\beta}$ eigenvalues.

The key point here is that we only need to execute the decomposition procedure for about $O(\frac{1}{\beta})$ levels, and thus we only delete $O(\frac{\varepsilon}{\beta\lambda})$ fraction of edges in total.

We choose β, λ such that $O(\frac{\varepsilon}{\beta\lambda})$ is just a small fraction of edges, so that we only ignore a small fraction of constraints.

Then, in each piece, we use the subspace enumeration algorithm in its label-extended graph, and we can find close to optimal solution in each piece.

Combining these solutions will be a better solution than the bad case, and so we can distinguish the two.

A lemma that we have skipped is an upper bound on the number of small eigenvalues in the label-extended graph, using such an upper bound in the ordinary graph.

Setting the parameters appropriately (lots of rooms) would get a $O(\exp(n^{R\,\mathrm{poly}(\frac{1}{\varepsilon})}))$-time algorithm that satisfies $1-\varepsilon^c$ fraction of constraints for some $c<1$ when the instance has a solution that satisfies $1-\varepsilon$ fraction of constraints.

This result has inspired further research on using SDP to obtain subexponential time algorithm for unique games, and also polynomial time approximation schemes for various problems on graphs with few small eigenvalues [BRS,GS].

References

[K02] On the power of unique 2-prover 1-round games, by Khot, 2002.

[ABS10] Subexponential algorithms for unique games and related problems, by Arora, Barak, Steurer, 2010.

[RS10] Graph expansion and the unique games conjecture, by Raghavendra, Steurer, 2010.

[BRS11] Rounding semidefinite programming hierachies using global correlation, by Barak, Raghavendra, Steurer, 2011.

[GS11] Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning..., by Guruswami, Sinop, 2011.