Rounding Techniques in Approximation Algorithms

Lecture 23: Global Correlation Rounding with SoS

Lecturer: Nathan Klein

1 Global Correlation Rounding

Let $\mu : \{-1,1\}^n \to \mathbb{R}$ be a degree $d \geq 2$ pseudodistribution. Then, we can compute:

$$\tilde{\text{Cov}}(x_i, x_j) = \tilde{\mathbb{E}}_\mu[x_i x_j] - \tilde{\mathbb{E}}_\mu[x_i]\tilde{\mathbb{E}}_\mu[x_j]$$

Define the global correlation as

$$C_\mu = \sum_{i\in[n]} \sum_{j\in[n]} |\tilde{\text{Cov}}(x_i, x_j)|$$

This lecture will have two parts:

1. First, we will show that as long as C_μ is small (at most $\epsilon\delta n^2$), we can obtain a $(1-\delta)$ approximation for max cut on dense graphs with at least ϵn^2 edges.
2. Second, we will show that using $f(\delta)$ rounds of SoS and a simple rounding step we can reduce C_μ to $\epsilon\delta n^2$.

Therefore, we can obtain an arbitrarily good approximation (a PTAS) for max cut on graphs with at least ϵn edges for any constant $\epsilon > 0$.

1.1 Uncorrelated Instances of Max Cut on Dense Graphs

It's not difficult to see that if the global correlation is small, we can define an independent distribution over the cube that approximates μ up to small error (in a dense graph).

As in the last lectures, let p_M be the max cut polynomial $\sum_{\{u,v\}\in E} \frac{1}{2}(1 - x_u x_v)$.

Lemma 1.1. *Let μ be a degree $d \geq 2$ pseudodistribution. Then, there is a distribution $\nu : \{-1,1\}^n \to \mathbb{R}$ with:*

$$\mathbb{E}_\nu[p_M] \geq \tilde{\mathbb{E}}_\mu[p_M] - \frac{1}{2}C_\mu$$

Proof. Let ν be the distribution which independently sets each index i to 1 with probability $\frac{1}{2}(1 + \mathbb{E}[x_i])$ and -1 otherwise, so that $\mathbb{E}[x_i] = \tilde{\mathbb{E}}_\mu[x_i]$, where we write $\mathbb{E}$ to indicate the expectation over ν. Now, $\mathbb{E}[x_i x_j] = \mathbb{E}[x_i]\mathbb{E}[x_j] = \tilde{\mathbb{E}}_\mu[x_i]\tilde{\mathbb{E}}_\mu[x_j]$ by independence. So,

$$\tilde{\text{Cov}}(x_i, x_j) = \tilde{\mathbb{E}}_\mu[x_i x_j] - \mathbb{E}[x_i]\mathbb{E}[x_j]$$

This exactly measures the difference of the contribution of an edge i, j to the objective function:

$$\mathbb{E}[p_M] - \tilde{\mathbb{E}}_\mu[p_M] = \frac{1}{2}\sum_{\{u,v\}\in E} \tilde{\mathbb{E}}_\mu[x_u x_v] - \mathbb{E}[x_u]\mathbb{E}[x_v] = \frac{1}{2}\sum_{\{u,v\}\in E} \tilde{\text{Cov}}(x_u, x_v)$$

The claim follows since this is at least $-\frac{1}{2}\sum_{\{u,v\}\in E} |\tilde{\text{Cov}}(x_u, x_v)| \geq -\frac{1}{2}C_\mu$. $\square$

This inequality looks far from tight, since we are only summing over edges whereas C_μ looks over all pairs i, j. However, it turns out that C_μ is the object we can control using SoS. When the graph is sufficiently dense, this inequality is reasonable and leads to a $1-\delta$ approximation for max cut.

Corollary 1.2. *Let G be a graph with at least ϵn edges and suppose $C_\mu \leq \epsilon \delta n^2$. Then given a degree $d \geq 2$ pseudodistribution with $\mathbb{E}[p_M] \geq |E|/2$ we can find a true distribution with $\mathbb{E}[p_M] \geq (1-\delta)\tilde{\mathbb{E}}_\mu[p_M]$.*

Proof. Using the above inequality, we lose

$$\frac{1}{2}C_\mu \leq \frac{1}{2}\epsilon \delta n^2 \leq \frac{1}{2}\delta |E| \leq \delta \tilde{\mathbb{E}}_\mu[p_M]$$

compared to $\tilde{\mathbb{E}}_\mu[p_M]$. □

1.2 Intuition and Conditioning

Amazingly, we can use SoS to find pseudodistributions with small global correlation. The idea is to begin with a high degree pseudodistribution and condition a constant sized set of variables to integer values to reduce the covariance.

It's not unreasonable that conditioning things to integral values reduces the covariance: once all values have been conditioned to be integral, the covariance is of course 0. However, it is not clear why we only need constantly many conditionings. The best intuition for this, in my view, is as follows: suppose that C_μ is large, at least $\Omega(n^2)$. Then, *most pairs of variables* are strongly related, as their covariance is a constant. So by conditioning on just one variable, we may reasonably expect to make progress on a "global" scale.

To get off the ground with this approach, we first need to prove that conditioning on a variable to be -1 or 1 results in a new pseudodistribution (of lower degree).

Let μ be a degree d pseudodistribution. Let $\mu_{|x_i=1}$ be the pseudodistribution conditioned on $x_i = 1$. This means we should set the probability of all $x \in \{-1,1\}^n$ with $x_i = -1$ to 0 and rescale the remaining probabilities by $\frac{1}{\mathbb{P}[x_i=1]}$. Of course, this probability has not quite been defined, since we are dealing with a pseudodistribution. But the natural thing to do would be to let

$$\mathbb{P}[x_i = 1] = \sum_{x \in \{-1,1\}^n} \mu(x)\mathbb{I}\{x_i = 1\} = \frac{1}{2}\tilde{\mathbb{E}}_\mu[1 + x_i],$$

where this is easy to see since when $x_i = -1$ this expression is 0 and otherwise it is 2.

Lemma 1.3. *$\nu = \mu_{|x_i=1}$ is a degree $d-2$ pseudodistribution (and so is $\mu_{|x_i=-1}$), so long as $\mathbb{P}[x_i = 1] > 0$ (respectively, $\mathbb{P}[x_i = -1] > 0$).*

Proof. We will prove that ν obeys the two properties of pseudodistributions. First,

$$\tilde{\mathbb{E}}_\nu[1] = \frac{1}{\frac{1}{2}\tilde{\mathbb{E}}_\mu[1 + x_i]} \sum_{x \in \{-1,1\}^n} \mu(x)\mathbb{I}\{x_i = 1\} = 1$$

Second, consider any degree $\frac{d}{2} - 1$ polynomial g. Then, using that the scaling term is strictly greater than 0 and at least 1:

$$\tilde{\mathbb{E}}_\nu[g^2] \geq \sum_{x \in \{-1,1\}^n} \mu(x)\mathbb{I}\{x_i = 1\} g^2(x)$$

but this is equivalent to the expectation of $(fg)^2$ where f is the polynomial $\frac{1}{2}(1+x_i)$, i.e. an indicator function that x_i is 1. So, this is simply $\tilde{\mathbb{E}}_\mu[(fg)^2]$ which must be positive since this polynomial has maximum degree $\frac{d}{2}$.

A similar proof holds for the case $\mu_{|x_i=-1}$. □

And of course by repeatedly applying this, we can condition $k < \frac{d}{2}$ variables to be -1 or 1 and still obtain a pseudodistribution of degree $d-2k$.

1.3 The Law of Total Variance

Before we prove our main lemma, we need to review the law of total variance and show how it relates to covariance. Given two random variables X, Y, the law of total variance is:

$$\mathrm{Var}(X) = \mathbb{E}_Y\left[\mathrm{Var}(X \mid Y)\right] + \mathrm{Var}_Y(\mathbb{E}\left[X \mid Y\right])$$

The first term on the righthand side measures the expected variance of X after Y is fixed. If X and Y are highly correlated, this should be close to 0, otherwise it will be similar to the variance of X. The second term measures how much the expected value of X fluctuates given Y. If X and Y are highly correlated, this is high: the expected value of X will change a lot as Y changes. If they are independent, this is 0 as in this case $\mathbb{E}\left[X\right]$ is just a constant. Statisticians call the first term the "unexplained" part of the variance of X after knowing Y and the second term the "explained" part.

We will skip proving this in lecture, but it is a good exercise and doing it helps cement the above intuition.

For us, it is very useful to think of the "explained part" the *expected reduction in variance after conditioning on* Y. Indeed, just rewriting the equation,

$$\mathbb{E}_Y\left[\mathrm{Var}(X \mid Y)\right] = \mathrm{Var}(X) - \mathrm{Var}_Y(\mathbb{E}\left[X \mid Y\right])$$

the lefthand side is the expected variance after conditioning. For our purposes here, what we would like is to condition on variables Y for which $\mathrm{Var}_Y(\mathbb{E}\left[X \mid Y\right])$ is large.

Let's understand this term a bit better.

Lemma 1.4.

$$\mathrm{Var}_Y(\mathbb{E}\left[X \mid Y\right]) \geq \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}$$

Proof. First, we use that variance and covariance are shift-invariant. So, to simplify calculations we may assume that $\mathbb{E}\left[X\right] = \mathbb{E}\left[Y\right] = 0$. Thus, this is equivalent to:

$$\mathbb{E}_Y[\mathbb{E}\left[X \mid Y\right]^2] - \mathbb{E}_Y[\mathbb{E}\left[X \mid Y\right]]^2 \geq \frac{\mathbb{E}\left[XY\right]^2}{\mathbb{E}\left[Y^2\right]}$$

But notice that $\mathbb{E}_Y[\mathbb{E}\left[X \mid Y\right]] = \mathbb{E}\left[X\right] = 0$. Now:

$$\mathbb{E}\left[XY\right]^2 = \mathbb{E}_Y[\mathbb{E}\left[XY \mid Y\right]]^2 = \mathbb{E}_Y[Y \cdot \mathbb{E}\left[X \mid Y\right]]^2 \leq \mathbb{E}\left[Y^2\right]\mathbb{E}_Y[\mathbb{E}\left[X \mid Y\right]^2]$$

where in the last inequality we used Cauchy-Shwarz.[1] □

Thus, for $Y \in \{-1, 1\}$, we have $\mathrm{Var}_Y(\mathbb{E}\left[X \mid Y\right]) \geq \mathrm{Cov}(X,Y)^2$.

[1] If you haven't seen this inequality used in this way before, we apply it to the vectors $(\sqrt{\mathbb{P}\left[Y=a_1\right]}a_k, \ldots, \sqrt{\mathbb{P}\left[Y=a_k\right]}a_k)$ and $(\sqrt{\mathbb{P}\left[Y=a_1\right]}\mathbb{E}\left[X \mid Y=a_1\right], \ldots, \sqrt{\mathbb{P}\left[Y=a_k\right]}\mathbb{E}\left[X \mid Y=a_k\right])$.

1.4 Global Correlation Rounding

In this section, all variances and covariances will be with respect to the pseudoexpectation operator. Define $V(x) = \sum_{i \in V} \mathrm{Var}(X_i)$. At the beginning of the algorithm, this quantity is at most n.

We will iterate conditionings and fall into one of two cases:

1. $C_\mu \leq \epsilon \delta n^2$, i.e. $\mathbb{E}_{i,j}\left[|\mathrm{Cov}(x_i, x_j)|\right] \leq \epsilon\delta$. Then we are done by independent rounding.
2. Otherwise, $\mathbb{E}_{i,j}\left[|\mathrm{Cov}(x_i, x_j)|\right] \geq \epsilon\delta$. By Lemma 1.4, the expected change in $V(x)$ after the new conditioning is at least n times $\mathrm{Cov}(x_i, x_j)^2 \geq n\epsilon^2\delta^2$. Choose an index i which reduces the change in $V(x)$ by at least this much and condition on it. (We can find this index as all quantities are polynomial time computable using pseudoexpectations).

So, this process can repeat at most $\frac{1}{\epsilon^2\delta^2}$ many times and eventually we will fall into case 1. Thus it is sufficient to let $d = 2 + \frac{1}{\epsilon^2\delta^2}$ and gives us a $1 - \delta$ approximation for any constant $\delta > 0$ in time $n^{O(\frac{1}{\epsilon^2\delta^2})}$.