Sums of Squares Max Cut and Cheeger Lecture Notes

Aaron Geelon So

September 25, 2018

1 Max Cut

MAXCUT: given a graph, bipartition the vertices in a way that cuts as many of the edges as possible. Define maxcut(G) to be the maximum fraction of edges cut by a bipartition.

Erdös gave a randomized algorithm that achieves a 1/2-approximation. Suppose for each vertex, we assign it to one of the bipartitions according to a fair flip of the coin. Each edge will be cut in exactly half of the possible partitions, so in expectation, a random bipartition will cut half of all edges. In particular, it must cut at least a $\frac{1}{2}$maxcut(G) fraction of the edges.

Fact 1 (BS2016)**.** *No linear programming relaxation of polynomial-size can have an approximation factor smaller than* $1/2$ *factor for* MAXCUT.

In contrast, using SoS, it is possible to obtain a 0.878 approximation in polynomial time. Before proving this, we need a little bit of preliminaries. For a graph G, let L be the combinatorial graph Laplacian, $L = D - A$, and f_G be its associated quadratic form: $f_G(x) = x^T Lx$. That is:

$$f_G(x) = \sum_{(i,j)\in E(G)} (x_i - x_j)^2.$$

Each $x \in \{0,1\}^n$ corresponds to a bipartition of the vertices, and $f_G(x)$ measures the number of edges cut by the bipartition x. Thus, computing maxcut(G) is equivalent to computing:

$$\max_{x\in\{0,1\}^n} f_G(x).$$

As a quick aside, this problem is related to the eigenvalues of L. Indeed, if instead we were to optimize over the ℓ_2-sphere, we could efficiently solve this problem (the solutions being the maximum eigenvector). Still, we can use the eigenvectors to give an upper bound. Since $\mathbf{1}$ is in the kernel of L, $f_G(x) = f_G(x + C)$ for any constant C. In particular, this allows us to recenter $x \in \{0,1\}^n$ at its center of mass; let's call the recentered points x'.

Those x's where an equal number of coordinates are 0 and 1 have maximal ℓ_2-norm of the newly centered set. The value for such $\|x'\|_2^2 = n/4$. It follows that:

$$\max_{x\in\{0,1\}^n} f_G(x) \le \frac{n}{4} \cdot \lambda_{\max}(L).$$

To get a sense of the fraction of edges this is, suppose G is a d-regular graph, in which case, $\lambda_{\max}(L) = d - \lambda_{\min}(A)$, where A is the adjacency matrix, and $|E| = nd/2$. Then:

$$\max_{x\in\{0,1\}^n} f_G(x) \le \frac{n}{4}(d - \lambda_{\min}(A)) = \frac{1}{2}\left(1 - \frac{\lambda_{\min}(A)}{d}\right) \cdot |E|.$$

It's also not hard to see that since $-d \le \lambda_{\min}(A) < 0$, this bound is not trivial.

1.1 Review of SoS

The pseudo-distribution generalizes a probability distribution by relaxing the nonnegativity condition.

Definition 2 (Pseudo-distribution)**.** *A* degree d pseudo-distribution *over* $\{0,1\}^n$ *is a function* $\mu : \{0,1\}^n \to \mathbb{R}$ *such that it satisfies:*

$$\widetilde{\mathbb{E}}_{\mu} \mathbf{1} = \sum_{x \in \{0,1\}^n} \mu(x) \cdot 1 = 1, \qquad \text{(normalization)}$$

and also for all polynomials $f : \{0,1\}^n \to \mathbb{R}$ *of degree at most* $d/2$*,*

$$\widetilde{\mathbb{E}}_{\mu} f^2 \geq 0. \qquad \text{(relaxed nonnegativity)}$$

Fact 3 (SoS Duality)**.** *For every function* $f : \{0,1\}^n \to \mathbb{R}$*, and every even* $d \in \mathbb{N}$*, there exists a degree-*d *SoS certificate for nonnegativity of* f *if and only if every degree-*d *pseudo-distribution* μ *over* $\{0,1\}^n$ *satisfies* $\widetilde{\mathbb{E}}_{\mu} f \geq 0$*.*

Fact 4 (Separation Algorithm for Pseudo-Moments)**.** *Let* $d \in \mathbb{N}$ *be even. The following set of* pseudo-moments *admits a separation algorithm with running time* $n^{O(d)}$*,*

$$\mathcal{M}_d = \left\{ \widetilde{\mathbb{E}}_{\mu(x)} (1,x)^{\otimes d} : \mu \text{ is a deg.-}d \text{ pseudo-dist. over } \{0,1\}^d \right\}.$$

Corollary 5 (Optimization over polynomial objective)**.** *Let* f *be a function over* $\{0,1\}^n$*, and* v *be a vector so that* $f(x) = \langle v, (1,x)^{\otimes d} \rangle$*. There is an algorithm that can approximately minimize* $\widetilde{\mathbb{E}}_{\mu} f$ *in time* $n^{O(d)}$*.*

Proof. By definition,

$$\widetilde{\mathbb{E}}_{\mu} f = \langle v, \widetilde{\mathbb{E}}_{\mu} (1,x)^{\otimes d} \rangle.$$

Thus, it suffices to minimize over the linear function $y \mapsto \langle v, y \rangle$ over $\mathcal{M}_d$. Because $\mathcal{M}_d$ has a separation algorithm of time $n^{O(d)}$, we can use ellipsoid method to approximately minimize also in time $n^{O(d)}$. □

Fact 6 (Quadratic Sampling)**.** *For every degree-2 pseudo-distribution* μ *over* $\{0,1\}^n$*, there exists a probability distribution* ρ *over* $\mathbb{R}^n$ *with the same first two moments. That is,*

$$\widetilde{\mathbb{E}}_{\mu(x)} (1,x)^{\otimes 2} = \mathbb{E}_{x \sim \rho} (1,x)^{\otimes 2}.$$

From a lecture by David Steurer: "$n^{o(d)}$-time algorithms cannot distinguish between degree-d pseudodistributions and degree-d part of actual distributions." The overall idea:

1. Optimize over low-degree moments of pseudo-distributions (separation algorithm $\implies$ efficient)
2. Find distribution with equal low-degree moments

1.2 Using SoS for Approximation

We would like to give certificates of nonnegativity to functions of the form $c - f_G$, for this implies that

$$\text{maxcut}(G) = \max f_G \leq c.$$

However, to prove such a bound, we might potentially have to test $O(2^d)$ points $x \in \{0,1\}^d$. It turns out that when $c \geq \max f_G / 0.878$, there is an efficient algorithm to produce a certificate. Here is the main theorem:

Theorem 7 (Rounding pseudo-distributions for MAXCUT)**.** *For every graph G and degree-2 pseudo-distribution μ over the hypercube, there exists a probability distribution μ' over the hypercube such that:*

$$\mathop{\mathbb{E}}_{\mu'} f_G \geq 0.878 \cdot \mathop{\widetilde{\mathbb{E}}}_{\mu} f_G.$$

Furthermore, there exists a randomized polynomial-time algorithm that given the pseudo-distribution μ outputs a sample from μ'.

Corollary 8 (MAXCUT approximation)**.** MAXCUT *has a polynomial time 0.878-approximation algorithm.*

Proof. By Corollary 5, it is possible to efficiently maximize $\widetilde{\mathbb{E}}_\mu f_G$ to arbitrary accuracy $\epsilon > 0$ over degree-2 pseudo-distributions μ. We obtain a degree-2 pseudo-distribution such that:

$$\mathop{\widetilde{\mathbb{E}}}_{\mu} f_G > \max(f_G) - \epsilon.$$

By Theorem 7, we can efficiently sample from a distribution μ' such that:

$$E_{\mu'} f_G \geq 0.878 \cdot \max(f_G) - O(\epsilon).$$

Thus, with high probability, repeat sampling $x \sim \mu'$ will return some x such that $f_G(x) \geq 0.878 \cdot \max(f_G)$. □

Corollary 9 (Degree-2 SoS certificates for MAXCUT)**.** *For every graph G with n vertices, there exists a degree-2 SoS certificate for $c - f_G(x)$, where $c \geq \max(f_G)/0.878$.*

Proof. Since $\max(f_G) \geq \mathbb{E}_{\mu'} f_G$ for all distributions μ', Theorem 7 implies that for all degree-2 pseudo-distributions,

$$\max(f_G) \geq 0.878 \cdot \mathop{\widetilde{\mathbb{E}}}_{\mu} f_G.$$

By the SoS-duality, Fact 3, this implies that the function:

$$\frac{\max f_G}{0.878} - f_G$$

has a degree-2 SoS certificate. □

Now, the main proof.

Proof of Theorem 7. Without loss of generality, we may assume that μ satisfies:

$$\mathop{\widetilde{\mathbb{E}}}_{\mu(x)} x = \frac{1}{2} \cdot \mathbf{1}. \tag{1}$$

If it does not, then note that since $x^T L x = (\mathbf{1} - x)^T L (\mathbf{1} - x)$,

$$\mathop{\widetilde{\mathbb{E}}}_{\mu(x)} f_G = \sum_{x \in \{0,1\}^n} \mu(x) x^T L x = \sum_{x \{0,1\}^n} \mu(\mathbf{1} - x) x^T L x = \mathop{\widetilde{\mathbb{E}}}_{\mu(\mathbf{1}-x)} f_G.$$

Our analysis on f_G does not change if we instead consider $\frac{1}{2}(\mu(x) + \mu(\mathbf{1} - x))$. Furthermore, this pseudo-distribution satisfies Equation 1.

By Fact 6, the Quadratic Sampling lemma, we may efficiently sample from the Gaussian probability distribution $\xi \sim \mathcal{N}\left(\widetilde{\mathbb{E}}_\mu x, \widetilde{\mathbb{E}}_\mu x x^T\right)$ on $\mathbb{R}^n$. As each coordinate of ξ is centered at $\frac{1}{2}$, we can apply a threshold function coordinate-wise at the value $\frac{1}{2}$ to generate $x' \in \{0,1\}^n$. In this manner, the Gaussian induces a probability distribution μ' on the hypercube.

Notice three properties about μ'. Consider two vertices, $i, j \in [n]$. Then

- $\mathbb{E}_\mu (x_i')^2 = \mathbb{E}_\mu (x_i') = \frac{1}{2}$. Thus, $\mathrm{Var}[x_i'] = \frac{1}{4}$.
- $\mathrm{Cov}(x_i, x_j) = \frac{1}{4}\rho = \frac{1}{4}\left(\widetilde{\mathbb{E}}_\mu\, x_i x_j - 1\right)$.
- $(\xi_i, \xi_j) \sim \mathcal{N}\left(\frac{1}{2}\cdot \mathbf{1}, \frac{1}{4}\cdot \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$

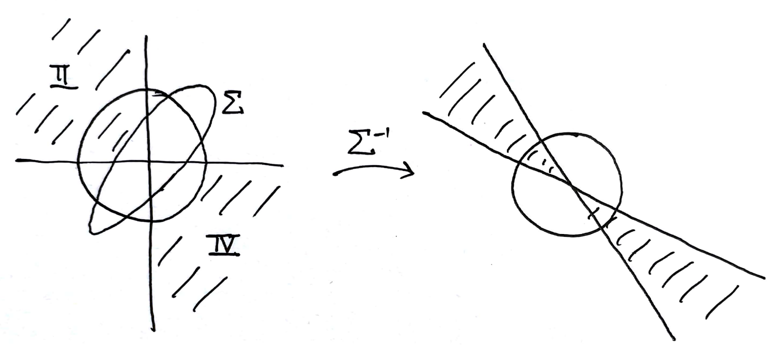
The following claim implies our result. It is completely geometric. **Claim:**

$$\mathop{\mathbb{E}}_{\mu'(x')} (x_i' - x_j')^2 \geq 0.878 \cdot \mathop{\widetilde{\mathbb{E}}}_{\mu(x)} (x_i - x_j)^2.$$

To show, this, first we claim that:

$$\mathop{\mathbb{E}}_{\mu'(x')} (x_i' - x_j')^2 = \frac{\cos^{-1}\rho}{\pi}. \tag{2}$$

This is most easily seen by the following diagram: And on the other hand, we have:



$$\mathop{\widetilde{\mathbb{E}}}_{\mu}(x_i - x_j)^2 = \mathop{\widetilde{\mathbb{E}}}_{\mu} x_i^2 + x_j^2 - 2x_i x_j = \frac{1}{2}(1-\rho). \tag{3}$$

Now, we obtain:

$$\mathop{\mathbb{E}}_{\mu'}(x_i' - x_j')^2 = \underbrace{\frac{2\cos^{-1}\rho}{(1-\rho)\pi}}_{\geq 0.878} \cdot \mathop{\widetilde{\mathbb{E}}}_{\mu}(x_i - x_j)^2.$$

Completing the proof: $\mathbb{E}_{\mu'}(x_i' - x_j')^2 \geq 0.878 \cdot \widetilde{\mathbb{E}}_\mu (x_i - x_j)^2$.

□

2 Cheeger's Inequality

Definition 10. *Let G be a d-regular graph, with vertex set $V = [n]$. For $S \subset V$, the* expansion $\phi_G(S)$ *is:*

$$\phi_G(S) = \frac{|E(S, V \setminus S)|}{\frac{d}{n}\cdot |S| \cdot |V \setminus S|}.$$

Notice that there are $nd/2$ many edges out of a possible $n(n-1)/2$ number of edges, on average, a $d/(n-1)$ fraction of possible edges are realized. As there are $|S| \cdot |V \setminus S|$ possible number of edges between a subset of size $|S|$ and its complement, the denominator gives the expected number of edges between a subset and its complement.

Definition 11. *The* expansion $\phi(G)$ *of a graph is the minimum* $\phi_G(S)$ *over all* S*. The* min expansion problem *is to find a vertex set* $S \subset V$ *that minimizes* $\phi_G(S)$*.*

While it's not obvious how to use SoS on a rational function, like ϕ_G, we have the following theorem:

Theorem 12 (Degree-2 SoS certificate for expansion)**.** *For every* d*-regular graph* G *with vertex set* $[n]$*, the following has a degree-2 SoS certificate:*

$$f_G(x) - \frac{1}{2}\phi(G)^2 \cdot \frac{d}{n}|x|(n - |x|).$$

References

[S2014] Steurer, D. *"Sum-of-Squares method and approximation algorithms I"*. Cargese Workshop, 2014. https://www.dsteurer.org/talk/cargese.pdf

[BS2016] Barak, B. Steurer, D. *Proof, beliefs, and algorithms through the lens of sum-of-squares.* 2016.