Sum of Squares: Part 2

Apoorv Vikram Singh

February 17, 2021

Recall MaxCut

- Given: $G = (V, E)$.
- Goal: Find $S \subseteq V$, such that $\left|E(S, \overline{S})\right|$ is maximized

Approximation Algorithm for MaxCut

- Algorithm: Return a random cut.
- In expectation: Algorithm cuts half the edges.
- MaxCut $\leq |E|$.
- Therefore, it is a $\frac{1}{2}$-approximation algorithm.

Can we improve the 1/2-approximation?

- Question: Is there an LP-based algorithm that achieves $(0.5+\varepsilon)$-approximation algorithm?
- Answer: There does not exist a 2^{n^δ} size LP that gets $(0.5+f(\delta))$-approximation.

- [Goemans-Williamson, 1994] Gave a 0.878-approximation algorithm for MaxCut (based on SDP).

Goal Today

- $G = (V, E)$, and let $\mathrm{Opt}(G) = \mathrm{MaxCut}(G)$.
- $f_G(\boldsymbol{x}) = \frac{1}{4}\sum_{(i,j)\in E}(\boldsymbol{x}_i - \boldsymbol{x}_j)^2$, for $x \in \{-1, 1\}^n$.
- $\max_{\boldsymbol{x}\in\{-1,1\}^n} f_G(x) = \mathrm{MaxCut}(G)$.

Theorem (0.878 Theorem)

For all G,

$$\frac{Opt(G)}{0.878} - f_G(\boldsymbol{x}),$$

has a degree-2 SoS certificate.

To prove the theorem, we will prove a "rounding" theorem.

Theorem (Rounding Theorem)

Let μ be a degree-2 pseudo-distribution on $\{-1,1\}^n$. Then, there is an actual distribution μ' such that

$$\mathop{\mathbb{E}}_{\mu'} f_G(\boldsymbol{x}) \geq 0.878\, \tilde{\mathbb{E}}_\mu f_G(\boldsymbol{x})\,.$$

- Rounding: Takes pseudo-distribution to actual distribution.

Rounding Theorem $\implies$ 0.878 Theorem

Proof.
Suppose $\frac{\mathsf{Opt}(G)}{0.878} - f_G(\boldsymbol{x})$ is not SoS_2, then,

- $\exists$ a degree-2 p.d. μ such that $\tilde{\mathbb{E}}_\mu \left(\frac{\mathsf{Opt}(G)}{0.878} - f_G(\boldsymbol{x}) \right) < 0$.
- Rearranging: $\tilde{\mathbb{E}}_\mu f_G > \frac{\mathsf{Opt}(G)}{0.878}$.
- Rounding Theorem $\implies$ $\exists$ a distribution μ', such that,

$$\mathop{\mathbb{E}}_{\mu'} f_G \geq 0.878\, \tilde{\mathbb{E}}_\mu f_G(\boldsymbol{x}) > \mathsf{Opt}(G)\,.$$

- $\mathbb{E}_{\mu'} f_G > \mathsf{Opt}(G)$, contradiction.

$\square$

Interpreting Rounding Theorem

- Suppose we have a p.d. μ, and under this p.d., $\tilde{\mathbb{E}}_\mu f_G(\boldsymbol{x}) = \mathrm{Opt}_{\mathrm{SoS}_2}$.
- We are interested in finding such cuts, or, if there are such cuts.
- Find distribution μ', such that $\mathbb{E}_{\mu'} f_G(\boldsymbol{x})$ is as large as possible.
- We won't be able to prove it is equal, but we can prove

$$\mathop{\mathbb{E}}_{\mu'} f_G(\boldsymbol{x}) \geq 0.878\, \mathrm{Opt}_{\mathrm{SoS}_2}\,.$$

- $\mu \to \mu'$ will be efficient $\implies$ algorithm to approximate MaxCut.

Proving Rounding Theorem

Ideally:

- Given p.d. μ, find distribution μ' over $\{-1,1\}^n$, such that

$$\mathop{\mathbb{E}}_{\mu'}(1,\boldsymbol{x})^{\otimes 2} = \tilde{\mathbb{E}}_{\mu}(1,\boldsymbol{x})^{\otimes 2}\,.$$

  This is called: *Generalized Moment Problem*.

- Not possible, otherwise we would have solved MaxCut exactly.

But, we can do it over $\mathbb{R}^n$

Lemma (Gaussian Sampling)

For any degree-2 p.d. μ, there exists an actual distribution over $\mathbb{R}^n$ with same first and second moments.

Proof.

For any p.d. μ of degree-2,

$$\tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top \succcurlyeq 0\,.$$

- First Moment: $\tilde{\mathbb{E}}_\mu \boldsymbol{x}$.
- Second Moment: $\tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top$.
- Sample: $\boldsymbol{g} \sim \mathcal{N}\left(\tilde{\mathbb{E}}_\mu \boldsymbol{x}, \tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top\right)$.

$\square$

Wlog $\tilde{\mathbb{E}}_\mu \boldsymbol{x} = \mathbf{0}$

- If μ was an actual distribution, then $\boldsymbol{x} \sim \mu$, and output $+\boldsymbol{x}$ or $-\boldsymbol{x}$ uniformly.
- Second Moment $\tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top$ remains unchanged.
- Mean $= \mathbf{0}$.

Look at the p.d. with mean $\mathbf{0}$ and second moment $\tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top$. The value of $\tilde{\mathbb{E}}_\mu f_G$ remains unchanged.

$$
\begin{aligned}
f_G(\boldsymbol{x}) &= \frac{1}{4} \sum_{(i,j)\in E} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \\
&= \frac{1}{4} \sum_{(i,j)\in E} (2 - 2\boldsymbol{x}_i \boldsymbol{x}_j) \\
\implies \tilde{\mathbb{E}}_\mu f_G(\boldsymbol{x}) &= \frac{1}{4} \sum_{(i,j)\in E} (2 - 2\tilde{\mathbb{E}}_\mu \boldsymbol{x}_i \boldsymbol{x}_j) \,.
\end{aligned}
$$

Efficient Algorithmic Process

$$\text{Recall:} \quad \boldsymbol{g} \sim \mathcal{N}\left(\mathbf{0}, \tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top\right) .$$

- $\mu \to \boldsymbol{g}$, such that $\tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top = \mathbb{E}\, \boldsymbol{g}\boldsymbol{g}^\top$.
- Issue: $\boldsymbol{g}$ does not have entries in $\{\pm 1\}$.

Efficient Algorithmic Process,

1. Take $\boldsymbol{g} \sim \mathcal{N}\left(\mathbf{0}, \tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top\right)$.
2. $\hat{\boldsymbol{x}}_i = \operatorname{sign}(\boldsymbol{g}_i)$, which gives that $\hat{\boldsymbol{x}} \in \{-1, 1\}^n$.

Call μ' the distribution on $\hat{\boldsymbol{x}}$.

Claim (Rounding Theorem)
$\mathbb{E}_{\mu'} f_G(\boldsymbol{x}) \geq 0.878\, \tilde{\mathbb{E}}_{\mu} f_G(\boldsymbol{x})$.

Lemma (Sheppard's Lemma)

$$\mathbb{P}\left[sign(\boldsymbol{g}_i) \neq sign(\boldsymbol{g}_j)\right] \geq \frac{2\arccos(\rho)}{\pi(1-\rho)}\ \mathbb{E}(\boldsymbol{g}_i - \boldsymbol{g}_j)^2\,,$$

for $\rho = \tilde{\mathbb{E}}_{\mu} \boldsymbol{x}_i \boldsymbol{x}_j = \mathbb{E}\, \boldsymbol{g}_i \boldsymbol{g}_j$.

Remark(s): Comparing LHS and RHS of claim with lemma.

$$\cdot \mathop{\mathbb{E}}_{\mu'} f_G(\boldsymbol{x}) = \frac{1}{4} \sum_{(i,j)\in E} \mathop{\mathbb{E}}_{\mu'} (\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j)^2 = \sum_{(i,j)\in E} \mathbb{P}\left[\mathrm{sign}(\boldsymbol{g}_i) \neq \mathrm{sign}(\boldsymbol{g}_j)\right].$$

$$\cdot \tilde{\mathbb{E}}_{\mu} f_G(\boldsymbol{x}) = \frac{1}{4} \sum_{(i,j)\in E} \tilde{\mathbb{E}}_{\mu} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 = \frac{1}{4} \sum_{(i,j)\in E} \mathbb{E}(\boldsymbol{g}_i - \boldsymbol{g}_j)^2\,.$$

Sheppard's Lemma $\implies$ Rounding Theorem

Proof.

$$\min_{\rho \in [-1,1]} \frac{2\arccos(\rho)}{\pi(1-\rho)} \geq \underbrace{\alpha_{GW}}_{=0.878\ldots}, \quad (\text{ min at } \rho = -0.69) \,.$$

This implies

$$\frac{1}{4} \mathop{\mathbb{E}}_{\mu'} (\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j)^2 \geq \alpha_{GW} \frac{1}{4} \tilde{\mathbb{E}}_{\mu} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \,,$$

$$\frac{1}{4} \sum_{(i,j)\in E} \mathop{\mathbb{E}}_{\mu'} (\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j)^2 \geq \alpha_{GW} \frac{1}{4} \sum_{(i,j)\in E} \tilde{\mathbb{E}}_{\mu} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \,.$$

$\square$

Proving Sheppard's Lemma

Proof

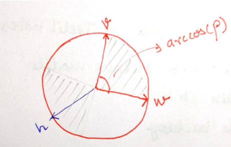
We have Gaussians $\boldsymbol{g}_i, \boldsymbol{g}_j$, such that $\mathbb{E}\, \boldsymbol{g}_i \boldsymbol{g}_j = \tilde{\mathbb{E}}_\mu \boldsymbol{x}_i \boldsymbol{x}_j = \rho$, and $\mathbb{E}\, \boldsymbol{g}_i^2 = \tilde{\mathbb{E}}_\mu \boldsymbol{x}_i^2 = 1$.
Procedure to generate such Gaussian vectors:

- Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{S}^{(2-1)}$ such that $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \rho$.
- Take $\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, I_2)$.
- $\hat{\boldsymbol{g}}_i = \langle \boldsymbol{h}, \boldsymbol{v} \rangle$, $\hat{\boldsymbol{g}}_j = \langle \boldsymbol{h}, \boldsymbol{w} \rangle$, this has same joint-distribution as $\boldsymbol{g}_i, \boldsymbol{g}_j$.

<u>We are interested in:</u>

$$\mathbb{P}\left[\operatorname{sign}(\boldsymbol{g}_i) \neq \operatorname{sign}(\boldsymbol{g}_j)\right] = \mathbb{P}\left[\operatorname{sign}(\hat{\boldsymbol{g}}_i) \neq \operatorname{sign}(\hat{\boldsymbol{g}}_j)\right] .$$

Proof Cont...



$$\begin{aligned}\mathbb{P}\left[\operatorname{sign}(\boldsymbol{g}_i) \neq \operatorname{sign}(\boldsymbol{g}_j)\right] &= \mathbb{P}\left[\operatorname{sign}(\hat{\boldsymbol{g}}_i) \neq \operatorname{sign}(\hat{\boldsymbol{g}}_j)\right] \\ &= \mathbb{P}\left[\operatorname{sign}(\langle \boldsymbol{h}, \boldsymbol{v}\rangle) \neq \operatorname{sign}(\langle \boldsymbol{h}, \boldsymbol{w}\rangle)\right] \\ &= \frac{\arccos(\rho)}{\pi}\,.\end{aligned}$$

And the other quantity

$$\frac{1}{4}\tilde{\mathbb{E}}_\mu(\boldsymbol{x}_i - \boldsymbol{x}_j)^2 = \frac{1}{4}\,\mathbb{E}(\boldsymbol{g}_i - \boldsymbol{g}_j)^2 = \frac{1}{4}\,\mathbb{E}(\hat{\boldsymbol{g}}_i - \hat{\boldsymbol{g}}_j)^2 = \frac{1}{2}(1-\rho).$$

$$\implies \mathbb{P}\left[\operatorname{sign}(\boldsymbol{g}_i) \neq \operatorname{sign}(\boldsymbol{g}_j)\right] \geq \frac{2\arccos(\rho)}{\pi(1-\rho)}\,\mathbb{E}(\boldsymbol{g}_i - \boldsymbol{g}_j)^2\,. \qquad \square$$

MaxCut Approximation Done.

Can we do better?

1. Can we do better with degree-2 SoS?: No.
2. Can we improve it with degree-4, degree-6, . . . , degree-$\log n$ SoS? Open.

How likely?

$\underline{\text{Unique Games Conjecture}} \implies (\alpha_{GW} + \varepsilon)$-approx to MaxCut is NP-Hard $\forall \varepsilon > 0$.

- Corollary: Suppose $\mathrm{Opt}(G) \geq (1 - \delta)\,|E|$, then Gaussian rounding gives $\mathbb{E}_{\mu'} f_G(\boldsymbol{x}) \geq \left(1 - \mathcal{O}\left(\sqrt{\delta}\right)\right)|E|$.

3. Is this the most optimal rounding? No (RPR^2 rounding does better in some regimes of δ).

Integrality Gaps?

What's the largest c for which degree-2 SoS certificate exists for $\frac{\mathrm{Opt}(G)}{c} - f_G(\boldsymbol{x})$?
Ans: $c = 0.878..$ is optimal.

Fact

C_n: Cycle on n vertices, n odd.
$MaxCut(C_n) = Opt(C_n) = \left(1 - \frac{1}{n}\right)|E|$.

Theorem

There is a p.d. μ of degree-2 such that

$$\tilde{\mathbb{E}}_\mu f_{C_n}(\boldsymbol{x}) = \left(1 - \mathcal{O}\left(\frac{1}{n^2}\right)\right)|E| \ .$$

Choose $n = \frac{1}{\delta}$, then $\mathrm{Opt}(C_n) = (1-\delta)\,|E|$, and
$\mathrm{Opt}_{\mathrm{SoS}_2}(C_n) \geq 1 - \mathcal{O}(\delta^2)\,|E|$.
$\implies$ Corollary for small δ is tight up to constant factors.

$$
\begin{aligned}
\text{Cycle} &= \text{``Discretized'' 2-dimn Sphere} \\
&\vdots \\
&= \text{``Discretized'' high-dimn Sphere}
\end{aligned}
$$

[Feige-Schechtman'02] Proved α_{GW} is optimal.

Proof Sketch of Theorem

MaxCut $= \max_{\boldsymbol{x} \in \{-1,1\}^n} \boldsymbol{x}^\top L_G \boldsymbol{x}$.
Relaxation $= \max_{\|\boldsymbol{x}\| = \sqrt{n}} \boldsymbol{x}^\top L_G \boldsymbol{x} = n \|L_G\|_2$.

- How to construct such a degree-2 p.d.?
- Choose a distribution on $\boldsymbol{x}$ that are in the "largest eigenspace" of L_G.
- We just need $\tilde{\mathbb{E}}_\mu (1, \boldsymbol{x})(1, \boldsymbol{x})^\top \succcurlyeq 0$, $\tilde{\mathbb{E}}_\mu \boldsymbol{x}_i^2 = 1$, $\tilde{\mathbb{E}}_\mu 1 = 1$.

1. Idea: $\lambda\max(L_G) = 1 - \mathcal{O}(1/n^2)$. It is not Boolean because maxcut is $(1 - \mathcal{O}(1/n)) |E|$. Top eigenspace is 2-dimensional with vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$.
2. set $M = \tilde{\mathbb{E}}_\mu \boldsymbol{x}\boldsymbol{x}^\top = \boldsymbol{v}_1 \boldsymbol{v}_1^\top + \boldsymbol{v}_2 \boldsymbol{v}_2^\top \succcurlyeq 0$.
3. Moreover, $\boldsymbol{v}_1 \boldsymbol{v}_1^\top + \boldsymbol{v}_2 \boldsymbol{v}_2^\top$ has diagonal entries 1.
4. Therefore, this is a valid pseudo-expectation.