# Lower bounds on the size of semidefinite relaxations

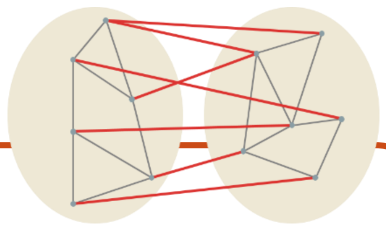
David Steurer
*Cornell*

James R. Lee
*Washington*

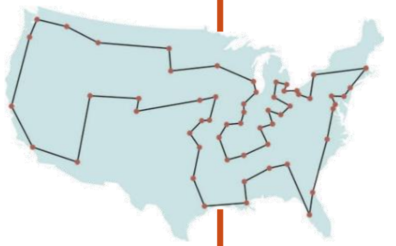
Prasad Raghavendra
*Berkeley*

Institute for Advanced Study, November 2015

# overview of results



**unconditional computational lower bounds**



**for classical combinatorial optimization problems**

- *examples:* MAX CUT, TRAVELING SALESMAN

**in restricted but powerful model of computation**

- generalizes best known algorithms
- all possible linear and ***semidefinite relaxations***
- first super-polynomial lower bound in this model

general program goes back to [Yannakakis'88] for refuting flawed P=NP proofs

**connection to optimization / convex geometry**

settle open question about *semidefinite lifts* of polytopes and *positive semidefinite rank*

[see survey Fazwi, Gouveia, Parrilo, Robinson, Thomas]

## *overview of results*

proof strategy for lower bounds

**optimal approximation algorithm in this model**

achieves best possible approximation guarantees
among all poly-time algorithms in this model

*wide-range of problems:* ***every constraint satisfaction problem***

*concrete algorithm:* ***sum-of-squares (aka Lasserre) hierarchy***

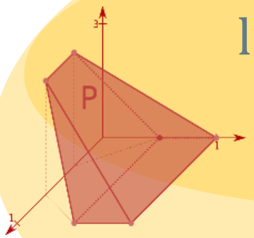
[Shor'87, Parrilo'00, Lasserre'00]

derive lower bounds for general model from known ***counterexamples (integrality gaps) for sum-of-squares algorithm***

[Grigoriev, Schoenebeck, Tulsiani, Barak-Chan-Kothari]

***mathematical programming relaxations:*** powerful general approach for approximating NP-hard optimization problems

***three flavors:***

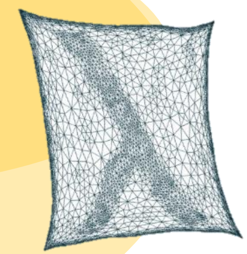



linear (**LP**)

(A)TSP

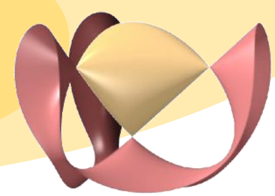
spectral

edge expanders



semidefinite (**SDP**)

MAX CUT, CSP, SPARSEST CUT



***intriguing connection to hardness reductions*** (e.g., Unique Games Conjecture)

plausibly optimal polynomial-time algorithms

***mathematical programming relaxations:*** powerful general approach for approximating NP-hard optimization problems

**Yannakakis's model** motivated by flawed P=NP proofs [Yannakakis'88]

formalizes ***intuitive notion of LP relaxations*** for problem

enough structure for ***unconditional lower bounds*** (indep. of P vs. NP)

Fiorini-Massar-Pokutta-Tiwary-de Wolf'12, Braun-Pokutta-S., Braverman-Moitra, Chan-Lee-Raghavendra-S., Rothvoß

extends to SDP relaxations (but LP lower bound techniques break down)

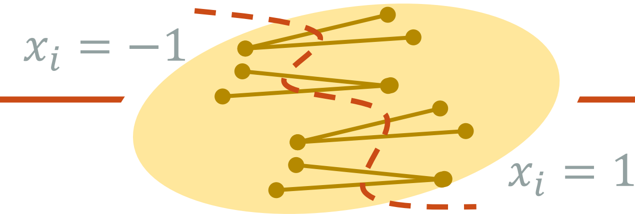
[Fiorini-Massar-Pokutta-Tiwary-de Wolf, Gouveia-Parrilo-Thomas]

**testing computational complexity conjectures**

*approximation / PCP:* UNIQUE GAMES, sliding scale conjecture

*average-case:* RANDOM 3 SAT, PLANTED CLIQUE

# *LP formulations of MAX CUT*



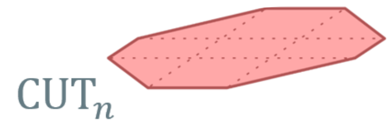

find bipartition in given $n$-vertex graph $G$
to cut as many edges as possible

maximize $f_G(x) = \sum_{ij \in E(G)} (x_i - x_j)^2/4$ over $x \in \{-1,1\}^n$ (hypercube)

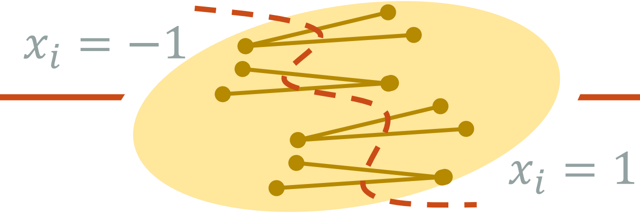
*equivalently:* maximize $\sum_{ij \in E(G)} (1 - X_{ij})/2$ over cut polytope

$\mathrm{CUT}_n$ = convex hull of $\{xx^\top \mid x \in \{-1,1\}^n\}$

#facets is exponential → *no small direct LP formulation*



# LP formulations of MAX CUT



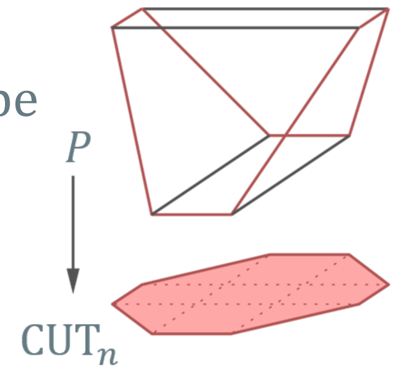

find bipartition in given $n$-vertex graph $G$
to cut as many edges as possible

maximize $f_G(x) = \sum_{ij \in E(G)} (x_i - x_j)^2/4$ over $x \in \{-1,1\}^n$ (hypercube)

*equivalently:* maximize $\sum_{ij \in E(G)} (1 - X_{ij})/2$ over cut polytope

$\mathrm{CUT}_n$ = convex hull of $\{xx^\top \mid x \in \{-1,1\}^n\}$

#facets is exponential → *no small direct LP formulation*




***general size-$n^d$ LP formulation of MAX CUT*** [Yannakakis'88]

polytope $P \subseteq \mathbb{R}^{n^d}$ defined by $\leq n^d$ linear inequalities that projects to $\mathrm{CUT}_n$

***often exponential savings:*** $\ell_1$-norm unit ball, Held-Karp TSP relaxation, LP/SDP hierarchies

***size lower bounds for LP formulations of MAX CUT?*** (implied by NP $\neq$ P/poly)

***example: poly-size LP formulation for $\ell_1$-norm ball***

$\ell_1$-unit ball

$$\left\{\begin{matrix} \sum_i |x_i| \le 1 \\ x \in \mathbb{R}^n \end{matrix}\right\} \xleftarrow{\text{project on } x \text{ variables}} \left\{\begin{matrix} -y \le x \le y \\ \sum_i y_i \le 1 \\ x, y \in \mathbb{R}^n \end{matrix}\right\}$$

$2^n$ linear inequalities

$2n+1$ linear inequalities

exponential savings

***general size-$n^d$ LP formulation of MAX CUT*** [Yannakakis'88]

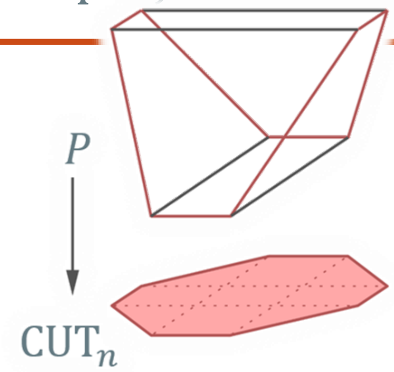
polytope $P \subseteq \mathbb{R}^{n^d}$ defined by $\le n^d$ linear inequalities that projects to $\mathrm{CUT}_n$

***often exponential savings:*** $\ell_1$-norm unit ball, Held-Karp TSP relaxation, LP/SDP hierarchies

***size lower bounds for LP formulations of MAX CUT?*** (implied by NP $\neq$ P/poly)

***general size-$n^d$ LP formulation of MAX CUT***

polytope $P \subseteq \mathbb{R}^{n^d}$ defined by $\leq n^d$ linear inequalities that projects to $\mathrm{CUT}_n$




***size lower bounds for LP formulations of MAX CUT***

**exponential lower bound:** $d \geq \widetilde{\Omega}(n)$ [Fiorini-Massar-Pokutta-Tiwary-de Wolf'12]

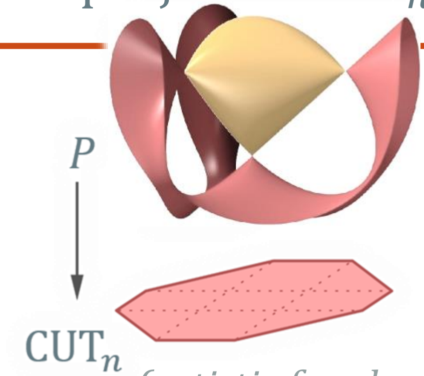
approx. ratio > ½ requires superpolynomial size [Chan-Lee-Raghavendra-S.'13]

***<u>but</u>: best known MAX CUT algorithms based on semidefinite programming***

***general size-$n^d$ ~~LP~~ SDP formulation of MAX CUT***

~~polytope $P \subseteq \mathbb{R}^{n^d}$ defined by $\le n^d$ linear inequalities~~ that projects to $\mathrm{CUT}_n$

**spectrahedron** $P \subseteq \mathbb{R}^{n^d \times n^d}$ defined by intersecting some affine linear subspace with psd cone




*(artistic freedom)*

***size lower bounds for ~~LP~~ SDP formulations of MAX CUT ?*** [Lee-Raghavendra-S.'15]

**exponential lower bound:** $d \geq \Omega(n^{0.1})$

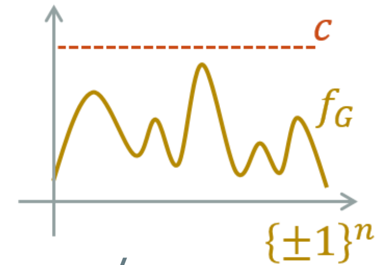
approx. ratio $> 0.99$ requires super polynomial size (match NP-hardness for CSPs)

best approx. ratio by $n^d$-size SDP no better than $O(d)$-deg. sum-of-squares

***→ sum-of-squares is optimal SDP approximation algorithm for CSPs***

***recall*** ***MAX CUT:*** maximize $f_G(x) = \sum_{ij \in E(G)} (x_i - x_j)^2/4$ over $x \in \{-1,1\}^n$

### *upper bound certificates*




algorithm with approx. guarantee must
*certify upper bounds* on objective function $f_G$

approx. ratio $\alpha \Rightarrow$ algorithm certifies $f_G \leq c$ for some $c \leq \mathrm{OPT}_G/\alpha$

*can characterize LP/SDP algorithms by their certificates*

### *certificates of deg-d sum-of-squares algorithm ($n^d$-size SDP example)*

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

equal as f'ns on hypercube

*captures best known algorithms for wide range of problems*

### *deg-1 sum-of-squares captures Goemans-Williamson MAX CUT 0.878-approx.*

for every graph $G$, $\mathrm{OPT}_G - 0.878 \cdot f_G = \sum_i g_i^2$ with $\deg g_i \leq 1$

***recall*** ***MAX CUT:*** maximize $f_G(x) = \sum_{ij \in E(G)} (x_i - x_j)^2/4$ over $x \in \{-1,1\}^n$

***certificates of deg-d sum-of-squares algorithm ($n^d$-size SDP example)***

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

equal as f'ns on hypercube

*captures best known algorithms for wide range of problems*

***deg-1 sum-of-squares captures Goemans-Williamson MAX CUT 0.878-approx.***

for every graph $G$, $\mathrm{OPT}_G - 0.878 \cdot f_G = \sum_i g_i^2$ with $\deg g_i \leq 1$

***recall*** ***MAX CUT:*** maximize $f_G(x) = \sum_{ij \in E(G)} (x_i - x_j)^2/4$ over $x \in \{-1,1\}^n$

## *connection to Unique Games Conjecture*

*best candidate algorithm to refute UGC:* deg-$\tilde{O}(1)$ sum of squares

*enough to show:* $\exists d.\ \forall G.\ \mathrm{OPT}_G - \mathbf{0.879} \cdot f_G = \sum_i g_i^2$ with $\deg g_i \le d$

## *does larger degree help?*

*yes:* if $f \ge 0$, then $f = g^2$ for some function $g$ with $\deg g \le n$ *(but $2^n$-size SDP)*

(*tight:* $(½ - \sum_i x_i)^2 - ¼ \ge 0$ has no deg-$o(n)$ s.o.s. certificate)

## *certificates of deg-d sum-of-squares algorithm ($n^d$-size SDP example)*

certify $f \ge 0$ for function $f: \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \le d$

equal as f'ns on hypercube

*captures best known algorithms for wide range of problems*

## *deg-1 sum-of-squares captures Goemans-Williamson MAX CUT 0.878-approx.*

for every graph $G$, $\mathrm{OPT}_G - 0.878 \cdot f_G = \sum_i g_i^2$ with $\deg g_i \le 1$

***where are the vectors?***

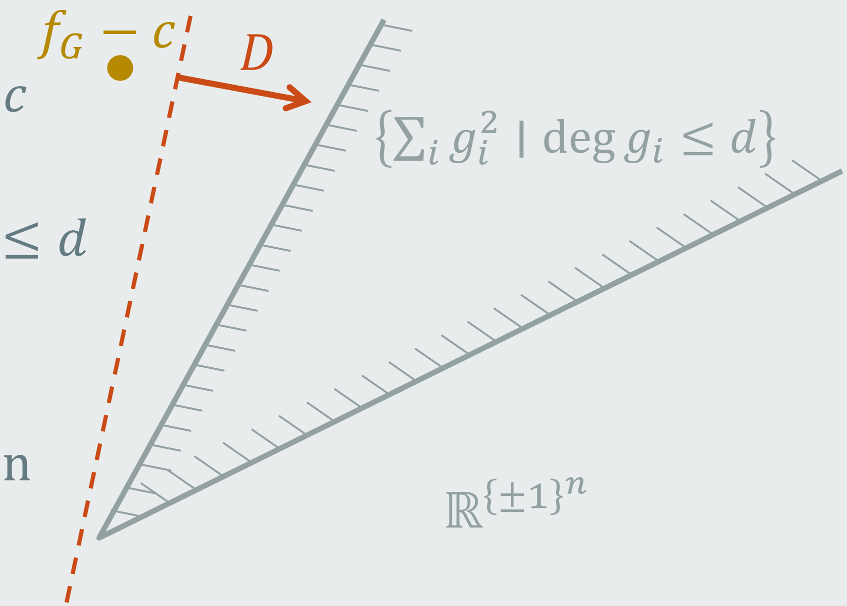
*suppose:* no deg-$d$ sos certificate for $f_G \geq c$

→ separating hyperplane $D\colon \{\pm 1\}^n \to \mathbb{R}$

$$\textstyle\sum_x D(x) \cdot g(x)^2 \geq 0 \text{ whenever } \deg g \leq d$$
$$\textstyle\sum_x D(x) \cdot 1 = 1$$
$$\textstyle\sum_x D(x) \cdot f_G(x) > c$$

→ $M = \sum_x D(x) \cdot x x^\top$ is usual SDP solution
(in particular $M \succcurlyeq 0$ and $M_{ii} = 1$)

→ $\exists$ vectors $\{v_i\}$ with $M_{ij} = \langle v_i, v_j \rangle$




***certificates of deg-d sum-of-squares algorithm ($n^d$-size SDP example)***

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

equal as f'ns on hypercube

*captures best known algorithms for wide range of problems*

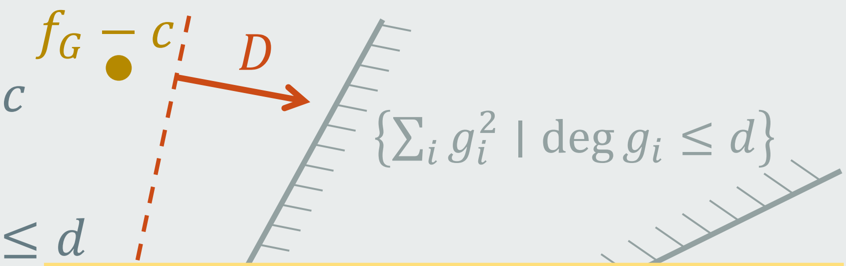
***deg-1 sum-of-squares captures Goemans-Williamson MAX CUT 0.878-approx.***

for every graph $G$, $\mathrm{OPT}_G - 0.878 \cdot f_G = \sum_i g_i^2$ with $\deg g_i \leq 1$

**where are the vectors?**

*suppose:* no deg-$d$ sos certificate for $f_G \geq c$

→ separating hyperplane $D: \{\pm 1\}^n \to \mathbb{R}$




$$\textstyle\sum_x D(x) \cdot g(x)^2 \geq 0 \text{ whenever } \deg g \leq d$$
$$\textstyle\sum_x D(x) \cdot 1 = 1$$
$$\textstyle\sum_x D(x) \cdot f_G(x) > c$$

→ $M = \sum_x D(x) \cdot xx^\top$ is usual SDP solution
(in particular $M \succcurlyeq 0$ and $M_{ii} = 1$)

→ ∃ vectors $\{v_i\}$ with $M_{ij} = \langle v_i, v_j \rangle$

$D$ behaves like *probability distribution* over MAX CUT solutions with *expected value* $> c$

→ ***pseudo-distribution:* useful way to think about LP/SDP relaxations in general**

**certificates of deg-d sum-of-squares algorithm ($n^d$-size SDP example)**

certify $f \geq 0$ for function $f: \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

equal as f'ns on hypercube

*captures best known algorithms for wide range of problems*

**deg-1 sum-of-squares captures Goemans-Williamson MAX CUT 0.878-approx.**

for every graph $G$, $\mathrm{OPT}_G - 0.878 \cdot f_G = \sum_i g_i^2$ with $\deg g_i \leq 1$

***certificates of deg-d sum-of-squares SDP algorithm***

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\, \deg g_i \leq d$

***certificates of general $n^d$-size SDP algorithm***

characterized by psd-matrix valued function $Q\colon \{\pm 1\}^n \to \mathbb{R}^{n^d \times n^d}$

certify $f \geq 0$ iff $\exists P \succcurlyeq 0.\, \forall x \in \{\pm 1\}^n.\ f(x) = \operatorname{Tr} PQ(\mathrm{x}) = \left\| \sqrt{P}\sqrt{Q(x)} \right\|_F^2$

*example:* **deg-$d$ sum-of-squares SDP algorithm,** $Q(x) = x^{\otimes d}\left(x^{\otimes d}\right)^\top$

***general SDP Q captured by deg-d sum-of-squares if*** $\deg \sqrt{Q(x)} \leq d$

***certificates of deg-d sum-of-squares SDP algorithm***

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

***certificates of general $n^d$-size SDP algorithm***

characterized by psd-matrix valued function $Q\colon \{\pm 1\}^n \to \mathbb{R}^{n^d \times n^d}$

certify $f \geq 0$ iff $\exists P \succcurlyeq 0.\ \forall x \in \{\pm 1\}^n.\ f(x) = \operatorname{Tr} PQ(\mathrm{x}) = \left\| \sqrt{P}\sqrt{Q(x)} \right\|_F^2$

*example:* **deg-$d$ sum-of-squares SDP algorithm,** $Q(x) = x^{\otimes d}\left(x^{\otimes d}\right)^\top$

***general SDP Q captured by deg-d sum-of-squares if*** $\deg \sqrt{Q(x)} \leq d$

***where does Q come from?*** *general spectrahedron:* $\mathrm{S} = \left\{ z \in \mathbb{R}^{n^d} \,\middle|\, \sum_i z_i A_i \succcurlyeq B \right\}$

***SDP relax. for MAX CUT:*** $\exists$ *cost functions* $\{y_G\}$ and *feasible solutions* $\{z_x\} \subseteq S$

with $\langle y_G, z_x \rangle = f_G(x)$ (obj. value of cut $x$ in graph $G$)

choose $Q$ with $Q(x) = B - \sum_i z_{x,i} A_i$ (slack of constraint at $z_x \in S$)

***duality:*** $\max\limits_{z \in S} \langle y_G, z \rangle \leq c$ iff $\exists P \succcurlyeq 0.\ \mathrm{c} - \langle y_G, z \rangle = \operatorname{Tr} P \cdot (B - \sum_i z_i A_i)$

$\rightarrow c - f_G(x) = \operatorname{Tr} P \cdot Q(x)$ for all $x$

***certificates of deg-d sum-of-squares SDP algorithm***

certify $f \geq 0$ for function $f\colon \{\pm 1\}^n \to \mathbb{R}$ iff $f = \sum_i g_i^2$ with $\forall i.\ \deg g_i \leq d$

***certificates of general $n^d$-size SDP algorithm***

characterized by psd-matrix valued function $Q\colon \{\pm 1\}^n \to \mathbb{R}^{n^d \times n^d}$

certify $f \geq 0$ iff $\exists P \succcurlyeq 0.\ \forall x \in \{\pm 1\}^n.\ f(x) = \operatorname{Tr} PQ(x) = \left\| \sqrt{P}\sqrt{Q(x)} \right\|_F^2$

*example:* **deg-$d$ sum-of-squares SDP algorithm,** $Q(x) = x^{\otimes d}\left(x^{\otimes d}\right)^\top$

***general SDP Q captured by deg-d sum-of-squares if*** $\deg \sqrt{Q(x)} \leq d$

***can simulate general $n^d$-size SDP algorithm by ~~deg O(d)~~ low-degree sum-of-squares***

$\forall\ n^d$-size SDP algorithm $Q$.

$\forall$ ***low-deg*** matrix-valued function $F$. $\langle F, Q \rangle \approx \langle F, Q' \rangle$

$\exists$ **low-deg** SDP algorithm $Q'$. $\deg \sqrt{Q'(x)} \approx \log n^d$ and $\cancel{Q \approx Q'}$

[Lee-Raghavendra-S.'15]

***general phenomenon***

in order to approximate an object with respect to a family of tests,
**the approximator need not be more complex than the tests**

***technical challenge***

naïve application allows us to bound $\deg Q'(x)$ but need to bound $\deg \sqrt{Q'(x)}$

*in general:* $\mathbf{deg}\sqrt{\boldsymbol{Q'}} \gg \mathbf{deg}\, \boldsymbol{Q'}$ *(at the heart of sum-of-squares counterexamples)*

*example:* **deg-*d* sum-of-squares SDP algorithm,** $Q(x) = x^{\otimes d}\left(x^{\otimes d}\right)^{\top}$

***general SDP Q captured by deg-d sum-of-squares if*** $\deg \sqrt{Q(x)} \le d$

***can simulate general $n^d$-size SDP algorithm by ~~deg O(d)~~ low-degree sum-of-squares***

$\forall\ n^d$-size SDP algorithm $Q$.

$\forall$ ***low-deg*** matrix-valued function $F$. $\langle F, Q\rangle \approx \langle F, Q'\rangle$

$\exists$ **low-deg** SDP algorithm $Q'$. $\deg \sqrt{Q'(x)} \approx \log n^d$ and $\cancel{Q \approx Q'}$

[Lee-Raghavendra-S.'15]

**_can simulate general $n^d$-size SDP algorithm by ~~deg $O(d)$~~ low-degree sum-of-squares_**

$\forall$ $n^d$-size SDP algorithm $Q$.

$\forall$ ***low-deg*** matrix-valued function $F$.

$\exists$ **low-deg** SDP algorithm $Q'$. $\deg \sqrt{Q'(x)} \approx \log n^d$ and $\cancel{Q \approx Q'}$ $\langle F, Q \rangle \approx \langle F, Q' \rangle$

***approach:*** learn "*simplest*" SDP algorithm $Q'$ that satisfies $\langle F, Q \rangle \approx \langle F, Q' \rangle$

*measure of simplicity:* quantum entropy (classical entropy of eigenvalues of $\{Q(x)\}$)

*closed-form solution:* $Q'(x) = e^{t \cdot F(x)}$ where $t$ = entropy-defect$(Q) \leq \log n^d$

→ ***matrix multiplicative weights method!***

***simple square root:*** $\sqrt{Q'(x)} = e^{t \cdot F(x)/2} \approx \sum_{k=0}^{t} \frac{1}{k!} (t \cdot F(x)/2)^k$

→ degree $\leq \deg F \cdot t$

## *summary*

**can simulate general small SDP alg. by low-degree SDP alg.**

interpret poly-size SDP algorithm as **quantum state** with high entropy

learn simplest SDP / quantum state via **matrix multiplicative weights** *(maximum entropy)*

## *open questions*

**approximation beyond CSP and relatives**

rule out 0.999-approximation for **TRAVELING SALEMAN** by poly-size LP/SDP

**strong quantitative lower bounds for approximation**

rule out 0.999-approximation for MAX CUT by $\mathbf{2^{n^{\Omega(1)}}}$**-size LP/SDP**

***latest news:*** **solved by Raghavendra-Meka-Kothari !**

**stronger quantitative lower bounds for SDP**

rule out $\mathbf{2^{n^{0.999}}}$**-size SDP** for (exact) MAX CUT

***Thank you!***