Rounding SUM-OF-SQUARES Relaxations

David Steurer
Cornell

joint work with

Boaz Barak
MSR

Jonathan Kelner
MIT

STOC, New York, June 2014

bird's eye view: sum-of-squares (SOS) method

conceptually simple meta algorithm studied by many disciplines [Shor, Nesterov, Parrilo, Lasserre]

applies to wide range of problems in unified way

contrast: problem-specific methods,
very general algorithmic techniques (like LP/SDP)

what do we lose? *surprisingly little*

captures best-known algorithms in most cases (even problem-specific ones)

prediction of Unique Games Conjecture (UGC): [Khot]
very restricted special case of SOS gives best-possible approximations for many problems [Khot-Regev, Khot-Kindler-Mossel-O'Donnell, …, Raghavendra, …]

what do we gain? *potentially a lot*

only few limitations known [Grigoriev, Schoenebeck]

for "all known instances," better approx. than predicted by UGC in worst-case [Barak-Brandao-Harrow-Kelner-S.-Zhou]

this talk: first general framework for proving guarantees of SOS;
better guarantees for various problems (related to small-set expansion, unsupervised learning, quantum information)

results: polynomial optimization over unit sphere

given: low-degree n-variate polynomial P with only nonneg. coefficients
find: $\|P\| \stackrel{\text{def}}{=} \max_{\|x\|=1} |P(x)|$ within error $\varepsilon \cdot \|P\|_{\text{spectral}}$

this paper: SOS runs in quasi–poly(n) time

naïve upper bound on $\|P\|$:
minimum $v \in \mathbb{R}$ s.t.
$v - P$ is sum of squares

follow-up: LOCC-1 polynomial instead of nonneg. coefficients [Brandao-Harrow'13]
→ greatly simplifies quantum breakthrough [Brandao-Christiandl-Yard'11]

open: general instead of nonneg. coefficients?
→ could solve major open problem in quantum; QMA vs QMA[2]

open: replace $\|P\|_{\text{spectral}}$ by $\|P\|$?
→ could solve small-set expansion on $\mathbb{F}_2^n$-Cayley graphs

open: remove both restrictions?
→ refute Small-Set Expansion Hypothesis—relative of UGC

results: sparse vectors in subspaces

given: d-dim. linear subspace W of $\mathbb{R}^n$ that contains k-sparse vector v_0
find: vector $v \in W$ with ℓ_4/ℓ_2-sparsity $C \cdot k$

this paper: SOS runs in poly(n) time for $C = d^{1/3}$ in worst case
——————— " ——— for $C = O(1)$ in average case

worst case: connection to small-set expansion / Unique Games Conjecture

solve small-set expansion for very small sets (beating other algorithms)

$C = O(1)$ in worst case would refute SSE hypothesis

average case: connection to learning (sparse coding & over-complete dictionaries)

previous methods: work only for sparsity $\frac{k}{n} \leq 1/\sqrt{d}$ [Spielman-Wang-Wright, Demanet-Hand]`

upcoming work: SOS learns over-complete dictionaries [Barak-Kelner-S.]

previous methods: assume strong incoherence and sparsity [Arora-Ge-Moitra Anankumar et al.]

multivariate polynomials
$P_1, \dots, P_m \in \mathbb{R}[x_1, \dots, x_n]$

system of equations
$\mathcal{E} = \{P_1 = 0, \dots, P_m = 0\}$

when is $\mathcal{E}$ unsatisfiable over $\mathbb{R}^n$?

idea: derive "obviously unsatisfiable equation" from $\mathcal{E}$

sum-of-squares (SOS) refutation of $\mathcal{E}$

vanishes on $\mathcal{E}$ → $Q_1 \cdot P_1 + \cdots + Q_m \cdot P_m$
$= 1 + R_1^2 + \cdots + R_t^2$ ← *positive over $\mathbb{R}^n$*

intuitive proof system: many common inequalities have proofs in this form, e.g., Cauchy-Schwarz, Hölder, ℓ_p-triangle inequalities

Real Nullstellensatz — *linear case:* Gaussian elimination, Farkas lemma — [Artin, Krivine, Stengle]

every polynomial system is either satisfiable over $\mathbb{R}^n$ or SOS refutable

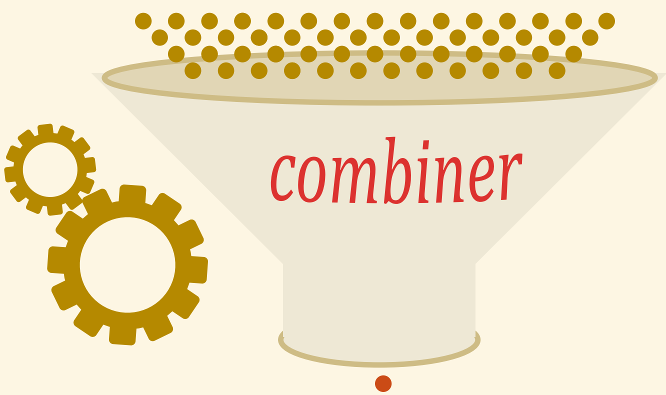
SOS method: $n^{O(k)}$-time algorithm to find SOS refutation with degrees $\leq k$ if one exists (uses SDP) [Shor, Nesterov, Parrilo, Lasserre] optimization (e.g., MAX CUT) maximize P_0 over $\{P_1 = 0, \ldots, P_m = 0\}$

v-vs-v' approximation: *given*: sat. system $\{P_0 = v, P_1 = 0, \ldots, P_m = 0\}$
find: solution to $\{P_0 = v', P_1 = 0, \ldots, P_m = 0\}$

claim: SOS reduces approximation in time $n^{O(k)}$ to "deg.-k combining"

subset $\mathcal{X}$ of solutions to $\{P_0 \geq v, P_1 = 0, \ldots, P_m = 0\}$

represented by all degree-k moments of $\mathcal{X}$, e.g., $\mathbb{E}_{\mathcal{X}} x_1 \cdots x_k$



use only properties of moments / solutions with degree-k SOS proofs

single solution to $\{P_0 \geq v', P_1 = 0, \ldots, P_m = 0\}$

pseudo-moments

"proof:" obstructions to degree-k SOS refutations *indistinguishable* from deg.-k moments *with respect to deg.-k SOS arguments*

planted sparse vector recovery

one k-sparse vector among d random vectors → arbitrary / random basis of span W of these vectors

k non-zeros

$$a_0 = \frac{1}{\sqrt{k}}(0, \dots, 0, \pm 1, \dots, \pm 1) \in \mathbb{R}^n \qquad y_0 = (\pm 1, \dots, \pm 1) \in \mathbb{R}^n$$

$$a_1 = \frac{1}{\sqrt{n}}(\pm 1, \dots, \pm 1) \qquad y_1 = (\pm 1, \dots, \pm 1)$$

$$\vdots \qquad \vdots$$

$$a_d = \frac{1}{\sqrt{n}}(\pm 1, \dots, \pm 1) \qquad y_d = (\pm 1, \dots, \pm 1)$$

goal: given $y_0, \dots, y_d$, recover vector $a^* \approx \pm a_0$ [Demanet-Hand'13]

idealized inference problem; subproblem for dictionary learning proxy for sparsity: if vector x is k-sparse then $\frac{\|x\|_\infty}{\|x\|_1} \geq \frac{1}{k}$ and $\frac{\|x\|_4^4}{\|x\|_2^4} \geq \frac{1}{k}$

previous best algorithm

find vector $x \in W$ with maximum ℓ_∞/ℓ_1 ratio
(exact using linear programming)

[Spielman-Wang-Wright, Demanet-Hand]

recovers $x \approx \pm a_0$ if and only if $\frac{k}{n} \ll 1/\sqrt{d}$

idea: use ℓ_4/ℓ_2 ratio instead

$\left\{\|x\|_4^4 = \frac{1}{k}, \|x\|_2^2 = 1, x \in W\right\}$

good: better proxy for sparsity ($d \ll \sqrt{n}$); system of polynomial equations

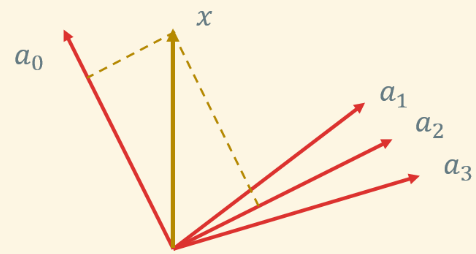
bad: NP-hard to solve exactly; somewhat hard to approximate (SSE-hard)

here: SOS works for this problem (exploit randomness in W)

combining problem: given $\mathcal{X} \subseteq \left\{ \|x\|_4^4 = \frac{1}{k}, \|x\|_2^2 = 1, x \in W \right\}$, find $x^* \approx \pm a_0$

claim: set $\mathcal{X}$ concentrated around planted vector (up to sign)

arguments used in analysis



deg.-4 SOS proofs

ℓ_4 triangle inequality

ℓ_4/ℓ_2 ratio bound for random subspaces for $d \ll \sqrt{n}$

[Barak-Brandao-Harrow-Kelner-S.-Zhou]

combiner: sample Gaussian distr. $\gamma_{\mathcal{X}}$ with same deg.-2 moments as $\mathcal{X}$

$\mathbb{E}_{\gamma_{\mathcal{X}}} x x^\top = \mathbb{E}_{\mathcal{X}} x x^\top \approx a_0 a_0^\top$ → random sample from $\gamma_{\mathcal{X}}$ is close to $\pm a_0$

deg.-2 SOS proof that $\mathrm{Cov}(X) \succcurlyeq 0$

→ *get algorithm via SOS*

conclusions

low-degree combiner: general way to make proofs into algorithms

unsupervised learning: higher-degree SOS gives better guarantees for recovering hidden structures

polynomial optimization: often easy when global optima unique (occurs naturally for recovery problems)

thank you!