Rounding Sum of Squares Relaxations

Boaz Barak – Microsoft Research

Joint work with Jonathan Kelner (MIT) and David Steurer (Cornell)

workshop on semidefinite programming and graph algorithms

February 10-14, 2014

This talk is about

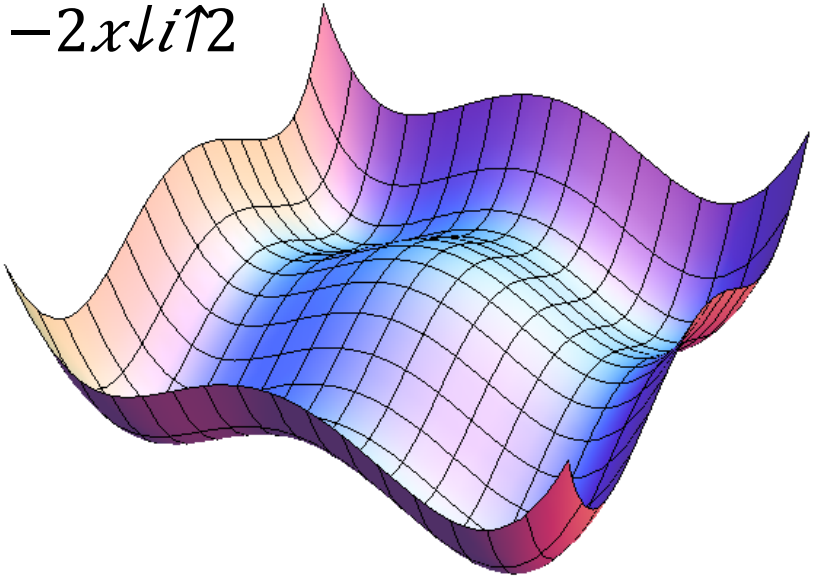
- Semi-definite programming , SOS/Positivstellensatz method
- Proof complexity
- The Unique Games Conjecture
- Graph partitioning, small set expansion
- Machine Learning
- *Cryptography.. (in spirit).*

Sum-of-Squares (SOS) Algorithm

[Shor'87,Parillo '00, Nesterov '00, Lasserre '01]

Motivation: Sometimes a polynomial P can have exponentially many local minima…

E.g. $P(x) = n + \sum_{i=1}^{n} x_i^4 - 2x_i^2$



… but there is still a short proof that $P \geq 0$

E.g. $P(x) = \sum_{i=1}^{n} (x_i^2 - 1)^2$

… and this proof can be found efficiently via semidefinite programming (SDP)

SOS Algorithm:

For low degree $P, P\downarrow 1, \ldots, P\downarrow k$ we consider the program $\mathcal{P}$:

$$\max_{x \in \mathbb{R}^n} P(x) \quad s.t.$$

$$P\downarrow 1\,(x) = \cdots = P\downarrow k\,(x) = 0$$

SOS Proof that $\mathcal{P} < v$: Polynomials $Q\downarrow 1, \ldots, Q\downarrow k$ and SOS polys S, S' s.t.

$$(v - P)S = \sum P\downarrow i\, Q\downarrow i + S' + 1$$

Positivstellensatz: All true bounds have SOS proof. [Artin '27, Krivine '64, Stengle '74]

Degree of proof: max degree of $Q\downarrow 1, \ldots, Q\downarrow k, S, S'$ [Gregoriev-Vorobjov'99]

Theorem: [Shor '87, Parillo '00, Nesterov '00, Lasserre '01]

1) A proof of degree d can be found in $n^{O(d)}$ time.
2) Can find in $n^{O(d)}$ time the min v s.t. ∃ degree d proof that $\mathcal{P} < v$

SOS Algorithm:

Program $\mathcal{P}$:

For low degree $P, P\downarrow 1, \ldots, P\downarrow k$ we consider the program $\mathcal{P}$:

$$\max_{x \in \mathbb{R}^n} P(x) \quad s.t.$$

$$P\downarrow 1(x) = \cdots = P\downarrow k(x) = 0$$

Can optimize in $n^{O(d)}$ time over programs with degree d proofs.

SOS Proof that $\mathcal{P} < v$: Polynomials $Q\downarrow 1, \ldots, Q\downarrow k$ and SOS polys S, S' s.t.

$$(v - P)S = \sum P\downarrow i \, Q\downarrow i + S' + 1$$

Positivstellensatz: All true bounds have SOS proof. [Artin '27, Krivine '64, Stengle '74]

Degree of proof: max degree of $Q\downarrow 1, \ldots, Q\downarrow k, S, S'$ [Gregoriev-Vorobjov'99]

Theorem: [Shor '87, Parillo '00, Nesterov '00, Lasserre '01]

1) A proof of degree d can be found in $n^{O(d)}$ time.
2) Can find in $n^{O(d)}$ time the min v s.t. ∃ degree d proof that $\mathcal{P} < v$

Program $\mathcal{P}$: $\max_{x \in \mathbb{R}^n} P(x)$ s.t. $P_1(x) = \cdots = P_k(x) = 0$

SOS Proof that $\mathcal{P} \vdash_d P \geq 0$ means $P = \sum P_i Q_i + S^2$ $+1$

Can optimize in $n^{O(d)}$ time over programs with degree d proofs.

Can't hope for $d \ll n$ *always:* Captures SAT, CLIQUE, 3COL, MAX-CUT, etc...

But maybe $d \ll n$ *often?* Essentially only one (robust) lower bound showing $d \geq \Omega(n)$ [Grigoriev '01]

This talk: General method to analyze the SOS algorithm. [B-Kelner-Steurer'13]

Applications:

- Optimizing polynomials w/ non-negative coefficients over sphere.
- Algorithms for quantum separability problem [Brandao-Harrow'13]
- Sparse coding: learning dictionaries beyond the $\sqrt{n}$ barrier.
- Finding sparse vectors in subspaces.
- Approach to refute the Unique Games Conjecture.

Program $\mathcal{P}$: $\max_{x \in \mathbb{R}^n} P(x)$ s.t. $P_1(x) = \cdots = P_k(x) = 0$

SOS Proof that $\mathcal{P} \vdash_d P \geq 0$: $P = \sum P_i Q_i + S^2 + 1$

Can optimize in $n^{O(d)}$ time over programs with degree d proofs.

Can't hope for $d \ll n$ *always:* Captures SAT, CLIQUE, 3COL, MAX-CUT, etc...

But maybe $d \ll n$ *often?* Essentially only one (robust) lower bound showing $d \geq \Omega(n)$ [Grigoriev '01]

This talk: General method to analyze the SOS algorithm. [B-Kelner-Steurer'13]

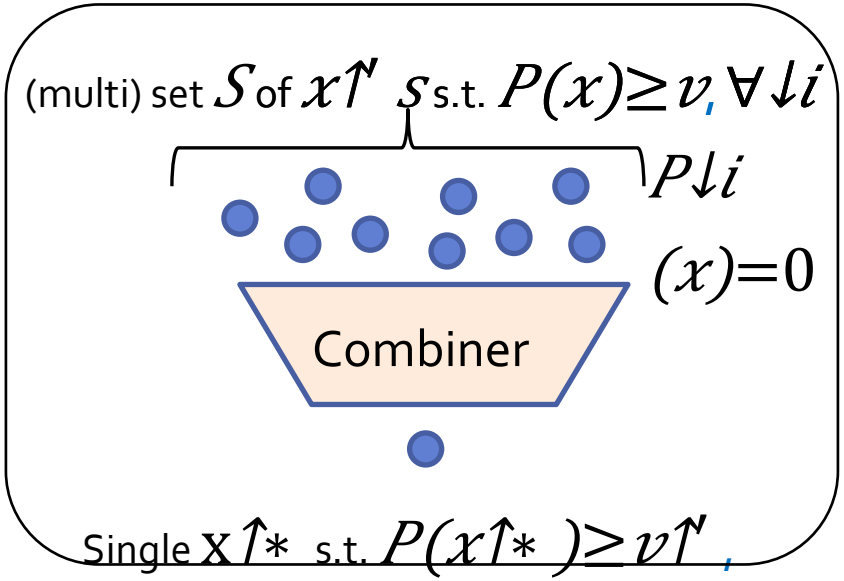
Applications:

- Optimizing polynomials w/ non-negative coefficients over sphere.
- Algorithms for quantum separability problem [Brandao-Harrow'13]
- Sparse coding: learning dictionaries beyond the $\sqrt{n}$ barrier.
- Finding sparse vectors in subspaces.
- Approach to refute the Unique Games Conjecture.

Program $\mathcal{P}$: $\max_{x \in \mathbb{R}^n} P(x)$ s.t. $P_1(x) = \cdots = P_k(x) = 0$

Finding x is hard. We consider easier problem:

Given many x's maximizing $\mathcal{P}$, find a single x^* with value close to maximum.



"Finding a needle in a needle-stack"

Non-trivial combiner:

Only depends on low degree marginals of S

$$\{ \mathbb{E}_{x \sim S} \; x_{i_1} \cdots x_{i_k} \}_{i_1, \ldots, i_k \in [n]}$$

[B-Kelner-Steurer'13]: Transform "simple" non-trivial combiners to algorithm for original problem.

Idea in a nutshell: Simple combiners will output a solution even when fed *"fake marginals"*.

Pseudoexpectations (aka "Fake Marginals")

Def: [Lasserre '01] *Degree d pseudoexpectation* is operator mapping any degree $\leq d$ poly P into a number $\mathbb{E}\ P(X)$ satisfying:

- Normalization: $\mathbb{E}\ 1=1$
- Linearity: $\mathbb{E}\ [aP(X)+bQ(X)]=a\mathbb{E}\ P(X)+b\mathbb{E}\ Q(X) \quad \forall P,Q$ of deg $\leq d$
- Positivity: $\mathbb{E}\ P^2(X)\geq 0 \quad \forall P$ of deg $\leq d/2$

Fundamental Fact: $\exists$ deg d SOS proof for $P>0 \Leftrightarrow$ $\mathbb{E}\ P(X)>0$ for any deg $O(d)$ pseudoexpectation operator

Take home message:

- Pseudoexpectation "looks like" real expectation to low degree polynomials.
- Can efficiently find pseudoexpectation matching any polynomial constraints.
- Proofs about real random vars can often be "lifted" to pseudoexpectation.

Idea in a nutshell: Simple combiners will output a solution even when fed *"fake marginals"*.

Pseudoexpectation (aka "Fake Marginals")

Problem: Given low degree $P, P_1, \ldots, P_k : \mathbb{R}^n \to \mathbb{R}$ maximize $P(x)$ s.t. $\forall i$ $P_i(x) \geq 0$

Def: [Lasserre '01] *Degree d pseudoexpectation* is operator mapping any degree $\leq d$ poly P into a number $\tilde{\mathbb{E}} P(X)$ satisfying:

- Normalization: $\tilde{\mathbb{E}}\, 1 = 1$
- Linearity: $\tilde{\mathbb{E}}\,[aP(X) + bQ(X)] = a\tilde{\mathbb{E}}\, P(X) + b\tilde{\mathbb{E}}\, Q(X) \quad \forall P, Q$ of deg $\leq d$
- Positivity: $\tilde{\mathbb{E}}\, P^2(X) \geq 0 \quad \forall P$ of deg $\leq d/2$

Fundamental Fact: $\exists$ deg d SOS proof for $P > 0 \Leftrightarrow$ $\tilde{\mathbb{E}}\, P(X) > 0$ for any deg $O(d)$ pseudoexpectation operator

Take home message:

- Pseudoexpectation "looks like" real expectation to low degree polynomials.
- Can efficiently find pseudoexpectation matching any polynomial constraints.
- Proofs about real random vars can often be "lifted" to pseudoexpectation.

Idea in a nutshell: Simple combiners will output a solution even when fed *"fake marginals"*.

Combining ⇒ Rounding

Problem: Given low degree $P, P_1, \ldots, P_k : \mathbb{R}^n \to \mathbb{R}$ maximize $P(x)$ s.t. $\forall i$ $P_i(x)=0$

[B-Kelner-Steurer'13]: Transform "simple" non-trivial combiners to algorithm for original problem.

Non-trivial combiner: Alg $\mathcal{C}$ with

INPUT: $\{ \mathbb{E} X_{i_1} \cdots X_{i_k} \}_{i_1 \ldots i_k \in [n]}$, X r.v. over $\mathbb{R}^n$ s.t. $\mathbb{E}(P(x)-v)^2 \geq 0$, $\forall i$ $\mathbb{E} P_i(X)^2 \leq v$

OUTPUT: $x^* \in \mathbb{R}^n$ s.t. $P_i(x^*) \geq -\varepsilon$, $\forall i$ $P_i(x^*)=0$

Crucial Observation: If proof that x^* is good solution is in SOS framework, then it holds even if $\mathcal{C}$ is fed with a pseudoexpectation.

Corollary: In this case, we can find x^* efficiently:

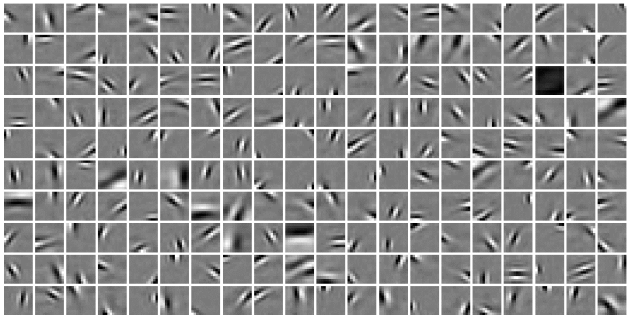
- Use SOS PSD to find pseudoexpectation matching input conditions.
- Use $\mathcal{C}$ to *round* the PSD solution into an actual solution x^*

Example Application: Dictionary Learning / Sparse Coding

Let $a\uparrow 1 ,\ldots, a\uparrow m \in \mathbb{R}\uparrow n$ set of vectors.

Goal: Given examples of form $y = \sum W\downarrow i \, a\uparrow i$, where $\mu = \Pr[W\downarrow i \neq 0] \ll 1$ recover $a\uparrow 1 \ldots a\uparrow m \downarrow$

Find the "right" representation of observed data [Olhausen-Field '96]



LOTS of work: important primitive in Machine Learning, Vision, Neuroscience…

Previous best (rigorous) results: $\mu \ll 1/\sqrt{n}$

[Spielman-Wang-Wright '12, Arora-Moitra-Ge '13, Agrawal-Anandkumar-Jain-Netrapalli-Tandon '13]

We show: $\mu \ll 1$ is sufficient* (even in non-independent, overcomplete case)

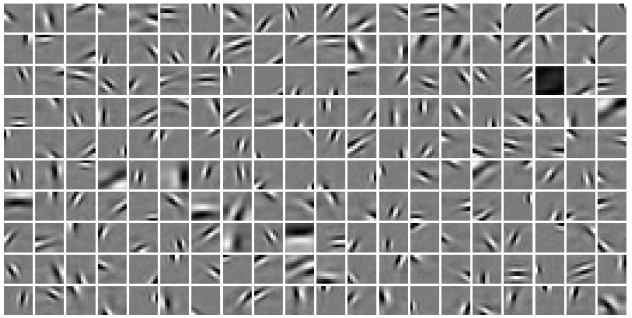
* In quasi-poly time, we show $\mu \ll n\uparrow -c$ is sufficient in poly

Example Application: Dictionary Learning / Sparse Coding

Let $a\uparrow 1 ,\ldots, a\uparrow m \in \mathbb{R}\uparrow n$ set of vectors.

Goal: Given examples of form $y = \sum W\downarrow i \, a\uparrow i$, where $\mu = \Pr[W\downarrow i \neq 0] \ll 1$ recover $a\uparrow 1 \ldots a\uparrow m \downarrow$

Find the "right" representation of observed data [Olhausen-Field '96]



LOTS of work: important primitive in Machine Learning, Vision, Neuroscience,…

Previous best (rigorous) results: $\mu \ll 1/\sqrt{n}$

[Spielman-Wang-Wright '12, Arora-Moitra-Ge '13, Agrawal-Anandkumar-Jain-Netrapalli-Tandon '13]

We show: $\mu \ll 1$ is sufficient* (even in non-independent, overcomplete case)

* In quasipoly time, we show $\mu \ll n\uparrow -c$ is sufficient in poly

Let $a\uparrow 1, \ldots, a\uparrow m \in \mathbb{R}\uparrow n$ set of vectors.

Goal: Given examples of form $y = \sum W\downarrow i\, a\uparrow i$, where $\mu = \Pr[W\downarrow i \neq 0] \ll 1$ recover $a\uparrow 1 \ldots a\uparrow m \downarrow$

Achieve in 3 steps:

(1) Find a program $\mathcal{P}$ s.t. every x maximizing $\mathcal{P}$ is close to one of $a\uparrow i$'s

(2) Give combining alg $\mathcal{C}$ taking moments of dist X over maximizers into a vector $x\uparrow *$ close to one of $a\uparrow i$'s.

(3) Show that arguments in (1) and (2) fall under the

Result generalizes to overcomplete, non independent case.

Step 1. For simplicity, assume $m = n$, $a\uparrow i$'s orthonormal basis, $W\downarrow i$ i.i.d. random vars over $\{0, \pm 1\}$ s.t. $\Pr[W\downarrow i = +1] = \Pr[W\downarrow i = -1] = \mu/2$

Consider the polynomial $P(x) = \mathbb{E}\langle y, x\rangle\uparrow 4 = \mathbb{E}(\sum W\downarrow i \langle a\uparrow i, x\rangle)\uparrow 4$ (can approximate P arbitrarily well from examples)

Opening parenthesis we get $P(x) \leq \mu \sum \langle a\uparrow i, x\rangle\uparrow 4 + 2\mu\uparrow 2 (\sum \langle a\uparrow i, x\rangle\uparrow 2)\uparrow 2 = \mu \sum \langle a\uparrow i, x\rangle\uparrow 4 + o(\mu)\|x\|\uparrow 4$

Corollary: x unit, $P(x) \geq \mu - o(\mu) \Rightarrow \max \langle a\uparrow i, x\rangle\uparrow 2 \geq 1 - o(1)$

Establishes (1) !

Let $a\uparrow 1, \ldots, a\uparrow m \in \mathbb{R}\uparrow n$ set of vectors.

Goal: Given examples of form $y = \sum W\downarrow i\, a\uparrow i$, where $\mu = \Pr[W\downarrow i \neq 0] \ll 1$ recover $a\uparrow 1 \ldots a\uparrow m \downarrow$

Achieve in 3 steps:

(1) Find a program $\mathcal{P}$ s.t. every x maximizing $\mathcal{P}$ is close to one of $a\uparrow i$'s

(2) Give combining alg $\mathcal{C}$ taking moments of dist X over maximizers into a vector $x\uparrow *$ close to one of $a\uparrow i$'s.

(3) Show that arguments in (1) and (2) fall under the

Result generalizes to overcomplete, non independent case.

Step 1. For simplicity, assume $m = n$, $a\uparrow i$'s orthonormal basis, $W\downarrow i$ i.i.d. random vars over $\{0, \pm 1\}$ s.t. $\Pr[W\downarrow i = +1] = \Pr[W\downarrow i = -1] = \mu/2$

Consider the polynomial $P(x) = \mathbb{E}\langle y, x\rangle\uparrow 4 = \mathbb{E}(\sum W\downarrow i \langle a\uparrow i, x\rangle)\uparrow 4$ (can approximate P arbitrarily well from examples)

Opening parenthesis we get $P(x) \leq \mu \sum \langle a\uparrow i, x\rangle\uparrow 4 + 2\mu\uparrow 2 (\sum \langle a\uparrow i, x\rangle\uparrow 2)\uparrow 2 = \mu \sum \langle a\uparrow i, x\rangle\uparrow 4 + o(\mu)\|x\|\uparrow 4$

Corollary: x unit, $P(x) \geq \mu - o(\mu) \Rightarrow \max \langle a\uparrow i, x\rangle\uparrow 2 \geq 1 - o(1)$

Establishes (1) !

Let $a\uparrow 1 ,..,a\uparrow m \in \mathbb{R}\uparrow n$ set of vectors.

Goal: Given examples of form $y=\sum W\downarrow i\, a\uparrow i$, where $\mu=\Pr[W\downarrow i \neq 0] \ll 1$ recover $a\uparrow 1 \ldots a\uparrow m \downarrow$

Achieve in 3 steps:

(1) Find a program $\mathcal{P}$ s.t. every x maximizing $\mathcal{P}$ is close to one of $a\uparrow i$'s

(2) Give combining alg $\mathcal{C}$ taking mom… nto a vector $x\uparrow *$ close to one of $a\uparrow i$'s.

Slightly tedious but straightforward computations.

(3) Show that arguments in (1) and (2) fall under the SOS framework.

Step 2. Let X be dist over unit vectors s.t. every $x \in Supp(X)$ satisfies $\langle a\uparrow i , x\rangle \uparrow 2 \geq 1-o(1)$ for some i

Pick set $U=\{u\uparrow 1 \ldots u\uparrow \ell\}$ of random (std gaussian) vectors.

Let M be matrix s.t. $M\downarrow i,j = \mathbb{E} f\downarrow U$ for $f\downarrow U(x)=\prod u\in U\uparrow$ Note that $\mathbb{E} f\downarrow U(x)=$

Our combining algorithm outputs the top e-vec of $\mathbb{E}_X X\downarrow i X\downarrow j \langle Mx\rangle \uparrow 2$

Happens w $\exp(-O(\ell))$ prob

Suppose that $\Pr[X \text{ close to } a\uparrow 1] \geq 1/n$ and for every $t\in[\ell]$, $\langle u\uparrow t , a\uparrow 1 \rangle \uparrow 2 \geq 2$

Then if $\ell \gg \log n$ then (up to scaling) $M \approx (a\uparrow 1)\uparrow \otimes 2$ and we'll

Establishes (2) !

A personal overview of the Unique Games Conjecture

Unique Games Conjecture: UG/SSE problem is NP-hard. [Khot'02,Raghavendra-Steurer'08]

reasons to believe

~~"Standard crypto heuristic":~~
~~Tried to solve it and couldn't.~~

SOS proof system

~~Very~~ clean picture of complexity landscape: simple algorithms are optimal
[Khot'02…Raghavendra'08….]

~~Simple poly algorithms can't refute it~~
[Khot-Vishnoi'04]

~~Simple subexp' algorithms~~ can't refute it
[B-Gopalan-Håstad-Meka-Raghavendra-Steurer'12]

reasons to suspect

~~Random instances are easy via simple algorithm~~
[Arora-Khot-Kolla-Steurer-Tulsiani-Vishnoi'05]

~~Quasipoly algo on KV instance~~
[Kolla '10]

~~Subexponential algorithm~~
[Arora-B-Steurer '10]

SOS solves all candidate hard instances
[B-Brandao-Harrow-Kelner-Steurer-Zhou '12]

SOS useful for sparse vector problem
Candidate algorithm for search problem
[B-Kelner-Steurer '13]

Conclusions

- Sum of Squares is a powerful algorithmic framework that can yield strong results for the right problems.

  (contrast with previous results on SDP/LP hierarchies, showing lower bounds when using either wrong hierarchy or wrong problem.)

- "Combiner" view allows to focus on the features of the problem rather than details of relaxation.

- SOS seems particularly useful for problems with some geometric structure, includes several problems related to unique games and machine learning.

- Still have only rudimentary understanding when SOS works or not.

- Other proof complexity $\leftrightarrow$ approximation algorithms connections?

Other Results

Sparse vector problem:

Recover μ-sparse vector in d-dimensional subspace given arbitrary basis.

(motivation: machine learning, optimization , [Demanet-Hand 13]
worst-case variant is algorithmic bottleneck in UG/SSE alg [Arora-B-Steurer'10])

Random case: Recovery for any $\mu \ll 1$

(Improving on $\mu \ll 1/\sqrt{d}$ [Demanet-Hand '13])

Worst case: Recovery* for $\mu \ll 1/d^{1/3}$

[Brandao-Harrow'12]: Using our techniques, find separable quantum state maximizing a "local operations classical communication" ($LOCC$) measurement.