SUM-OF-SQUARES method and approximation algorithms I

David Steurer
Cornell

Cargese Workshop, 2014

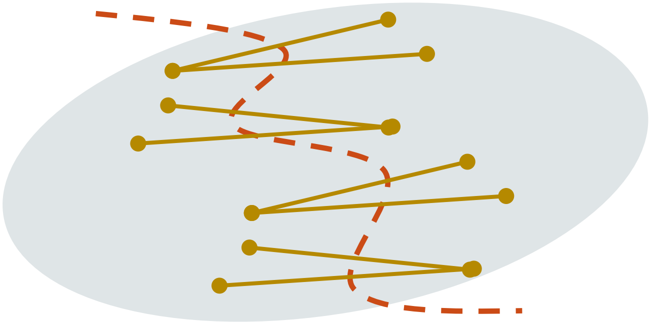
meta-task

encoded as low-degree polynomial in $\mathbb{R}[x]$

example: $f(x) = \sum_{i,j\in[n]} w_{ij} \cdot (x_i - x_j)^2$

given: functions $f_1, \ldots, f_m \colon \{\pm 1\}^n \to \mathbb{R}$

find: solution $x \in \{\pm 1\}^n$ to $\{f_1 = 0, \ldots, f_m = 0\}$



Laplacian $L_G = \frac{1}{|E(G)|} \sum_{ij\in E(G)} \frac{1}{4} (x_i - x_j)^2$

examples: combinatorial optimization problem on graph G

MAX CUT: $\{L_G = 1 - \varepsilon\}$ over $\{\pm 1\}^n$

MAX BISECTION: $\{L_G = 1 - \varepsilon, \sum_i x_i = 0\}$ over $\{\pm 1\}^n$

where $1 - \varepsilon$ is guess for optimum value

goal: develop SDP-based algorithms with provable guarantees in terms of complexity and approximation

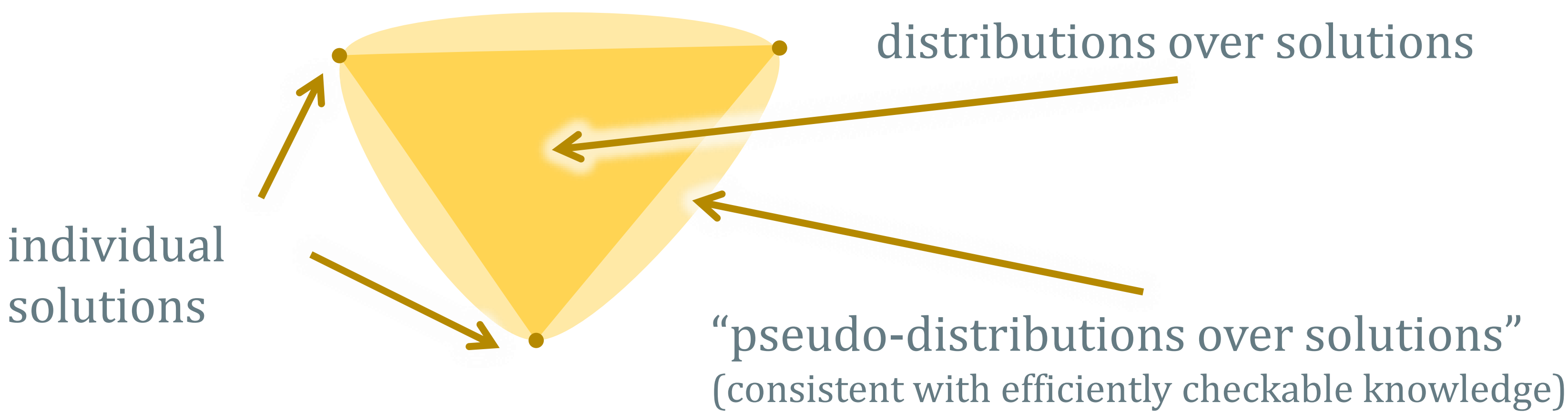
(“on the edge intractability” → need strongest possible relaxations) *meta-task*

given: functions $f_1, \dots, f_m\colon \{\pm 1\}^n \to \mathbb{R}$
find: solution $x \in \{\pm 1\}^n$ to $\{f_1 = 0, \dots, f_m = 0\}$

goal: develop SDP-based algorithms with provable guarantees in terms of complexity and approximation

price of convexity: individual solutions → distributions over solutions

price of tractability: can only enforce "efficiently checkable knowledge" about solutions



examples
uniform distribution: $D = 2^{-n}$
fixed 2-bit parity: $D(x) = (1 + x_1 x_2)/2^n$

distribution D over $\{\pm 1\}^n$

function $D: \{\pm 1\}^n \to \mathbb{R}$

function values is exponential
→ need careful representation

non-negativity: $D(x) \geq 0$ for all $x \in \{\pm 1\}^n$

normalization: $\sum_{x \in \{\pm 1\}} D(x) = 1$

independent inequalities is exponential
→ not efficiently checkable

distribution D satisfies $\{f_1 = 0, \ldots, f_m = 0\}$ for some $f_i: \{\pm 1\}^n \to \mathbb{R}$

$\mathbb{E}_D f_1^2 + \cdots + f_m^2 = 0$ (*equivalently:* $\mathbb{P}_D\{\forall i.\, f_i \neq 0\} = 0$)

convex: D, D' satisfy conditions
→ $(D + D')/2$ satisfies conditions

examples
fixed 2-bit parity distribution satisfies $\{x_1 x_2 = 1\}$
uniform distribution does not satisfy $\{f = 0\}$ for any $f \neq 0$

deg.-d pseudo-distribution D

~~distribution D~~ over $\{\pm 1\}^n$

convenient notation: $\tilde{\mathbb{E}}_D f \coloneqq \sum_x D(x) f(x)$
"**pseudo-expectation** of f under D"

function $D: \{\pm 1\}^n \to \mathbb{R}$

non-negativity: ~~$D(x) \geq 0$ for all $x \in \{\pm 1\}^n$~~

$\sum_{x \in \{\pm 1\}^n} D(x)\, f(x)^2 \geq 0$ for every deg.-$d/2$ polynomial f

normalization: $\sum_{x \in \{\pm 1\}} D(x) = 1$

$\underbrace{\text{pseudo-}}$distribution D satisfies $\{f_1 = 0, \ldots, f_m = 0\}$ for some $f_i: \{\pm 1\}^n \to \mathbb{R}$

$\tilde{\mathbb{E}}_D\ \cancel{\mathbb{E}_D}\, f_1^2 + \cdots + f_m^2 = 0$ (~~*equivalently:* $\mathbb{P}_D\{\forall i.\ f_i \neq 0\} = 0$~~)

deg.-$2n$ pseudo-distributions are actual distributions
(point-indicators $\mathbf{1}_{\{x\}}$ have deg. n → $D(x) = \tilde{\mathbb{E}}_D \mathbf{1}_{\{x\}}^2 \geq 0$)

deg.-d pseudo-distr. $D\colon \{\pm 1\}^n \to \mathbb{R}$

notation: $\widetilde{\mathbb{E}}_D f \coloneqq \sum_x D(x) f(x)$, "**pseudo-expectation** of f under D"

non-negativity: $\widetilde{\mathbb{E}}_D f^2 \geq 0$ for every deg.-$d/2$ poly. f

normalization: $\widetilde{\mathbb{E}}_D 1 = 1$

pseudo-distr. D satisfies $\{f_1 = 0, \ldots, f_m = 0\}$ for some $f_i\colon \{\pm 1\}^n \to \mathbb{R}$

$\widetilde{\mathbb{E}}_D f_1^2 + \cdots + f_m^2 = 0$ (*equivalently:* $\widetilde{\mathbb{E}}_D f_i \cdot g = 0$ whenever $\deg g \leq d - \deg f_i$)

deg.-d pseudo-distr. $D\colon \{\pm 1\}^n \to \mathbb{R}$

notation: $\widetilde{\mathbb{E}}_D f := \sum_x D(x) f(x)$, "**pseudo-expectation** of f under D"

non-negativity: $\widetilde{\mathbb{E}}_D f^2 \geq 0$ for every deg.-$d/2$ poly. f

normalization: $\widetilde{\mathbb{E}}_D 1 = 1$

pseudo-distr. D satisfies $\{f_1 = 0, \ldots, f_m = 0\}$ for some $f_i\colon \{\pm 1\}^n \to \mathbb{R}$

$\widetilde{\mathbb{E}}_D f_1^2 + \cdots + f_m^2 = 0$ (*equivalently:* $\widetilde{\mathbb{E}}_D f_i \cdot g = 0$ whenever $\deg g \leq d - \deg f_i$)

claim: can compute such D in time $n^{O(d)}$ if it exists (otherwise, certify that no solution to original problem exists) [Shor, Parrilo, Lasserre]

(can assume D is deg.-d polynomial → separation problem $\min_f \widetilde{\mathbb{E}}_D f^2$ is n^d-dim. eigenvalue prob. → $n^{O(d)}$-time via grad. descent / ellipsoid method)

deg.-d pseudo-distr. $D\colon \{\pm 1\}^n \to \mathbb{R}$

notation: $\widetilde{\mathbb{E}}_D f := \sum_x D(x) f(x)$, "**pseudo-expectation** of f under D"

non-negativity: $\widetilde{\mathbb{E}}_D f^2 \geq 0$ for every deg.-$d/2$ poly. f

normalization: $\widetilde{\mathbb{E}}_D 1 = 1$

pseudo-distr. D satisfies $\{f_1 = 0, \dots, f_m = 0\}$ for some $f_i\colon \{\pm 1\}^n \to \mathbb{R}$

$\widetilde{\mathbb{E}}_D f_1^2 + \cdots + f_m^2 = 0$ (*equivalently:* $\widetilde{\mathbb{E}}_D f_i \cdot g = 0$ whenever $\deg g \leq d - \deg f_i$)

surprising property: $\widetilde{\mathbb{E}}_D f \geq 0$ for many* low-degree polynomials f
such that $\{f \geq 0\}$ follows from $\{f_1 = 0, \dots, f_m = 0\}$ by "explicit proof"

soon: examples of such properties and how to exploit them

deg.-d pseudo-distr. $D\colon \{\pm 1\}^n \to \mathbb{R}$

notation: $\widetilde{\mathbb{E}}_D f \coloneqq \sum_x D(x) f(x)$, "**pseudo-expectation** of f under D"

non-negativity: $\widetilde{\mathbb{E}}_D f^2 \geq 0$ for every deg.-$d/2$ poly. f

normalization: $\widetilde{\mathbb{E}}_D 1 = 1$

pseudo-distr. D satisfies $\{f_1 = 0, \ldots, f_m = 0\}$ for some $f_i\colon \{\pm 1\}^n \to \mathbb{R}$

$\widetilde{\mathbb{E}}_D f_1^2 + \cdots + f_m^2 = 0$ (*equival*

$n^{o(d)}$-time algorithms cannot* distinguish between deg.-d pseudo-distributions and deg.-d part of actual distr.'s

surprising property: $\widetilde{\mathbb{E}}_D f \geq 0$ for n
such that $\{f \geq 0\}$ follows from $\{f_1 = 0, \ldots, f_m = 0\}$ by "explicit proof"

soon: examples of such properties and how to exploit them

deg.-d part of actual distr. over optimal solutions

pseudo-distr. over optimal solutions → *efficient algorithm* → approximate solution (to original problem)

emerging algorithm-design paradigm:
analyze algorithm pretending that underlying actual distribution exists;
verify only afterwards that low-deg. pseudo-distr.'s satisfy required properties

***dual view** (sum-of-squares proof system)*

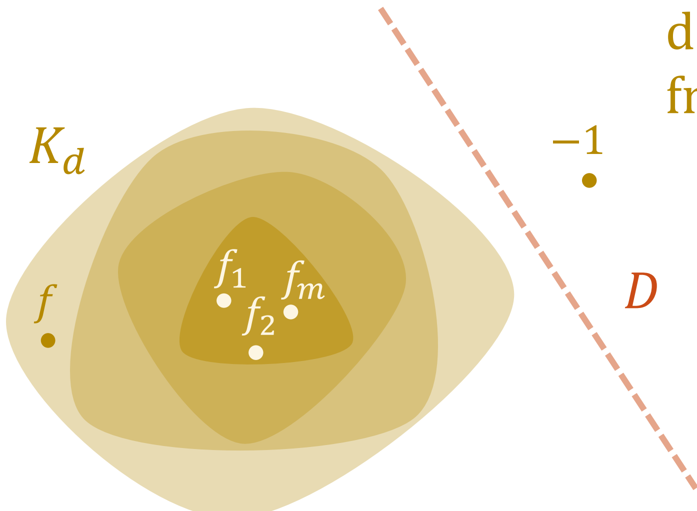
either

$\exists$ deg.-d pseudo-distribution D over $\{\pm 1\}^n$ satisfying $\{f_1 = 0, \dots, f_m = 0\}$

or

$\exists\, g_1, \dots, g_m$ and $h_1, \dots, h_k$ such that $\sum_i f_i \cdot g_i + \sum_j h_j^2 = -1$ over $\{\pm 1\}^n$
and $\deg f_i + \deg g_i \le d$ and $\deg h_i \le d/2$

derivation of unsatisfiable constraint $\{-1 \ge 0\}$
from $\{f_1 = 0, \dots, f_m = 0\}$ over $\{\pm 1\}^n$



$K_d = \{f = \sum_i f_i \cdot g_i + \sum_j h_j^2\}$

if $-1 \notin K_d$ then $\exists$ separating hyperplane D
with $\widetilde{\mathbb{E}}_D\, {-1} = -1$ and $\widetilde{\mathbb{E}}_D f \ge 0$ for all $f \in K_d$

pseudo-distribution satisfies all local properties of $\{\pm 1\}^n$

example: triangle inequalities over $\{\pm 1\}^n$

$$\widetilde{\mathbb{E}}_D \left(x_i - x_j\right)^2 + \left(x_j - x_k\right)^2 - (x_i - x_k)^2 \geq 0$$

claim

suppose $f \geq 0$ is $d/2$-junta over $\{\pm 1\}^n$ (depends on $\leq d/2$ coordinates)

then, $\widetilde{\mathbb{E}}_D f \geq 0$

proof: $\sqrt{f}$ has degree $\leq d/2$ → $\widetilde{\mathbb{E}}_D f = \widetilde{\mathbb{E}}_D \left(\sqrt{f}\right)^2 \geq 0$

corollary

for any set S of $\leq d$ coordinates, marginal $D' = \{x_S\}_D$ is actual distribution

$$D'(x_S) = \sum_{x_{[n]\setminus S}} D\left(x_S, x_{[n]\setminus S}\right) = \widetilde{\mathbb{E}}_D \underbrace{\mathbf{1}_{\{x_S\}}}_{d\text{-junta}} \geq 0$$

(also captured by LP methods, e.g., Sherali–Adams hierarchies …)

conditioning pseudo-distributions

claim

$\forall i \in [n], \sigma \in \{\pm 1\}.\ D' = \left\{x \mid x_j = \sigma\right\}_D$ is deg.-$(d-2)$ pseudo-distr.

proof

$$D'(x) = \frac{1}{\mathbb{P}_D\{x_j=\sigma\}} D(x) \cdot 1_{\{x_j=\sigma\}}$$

$$\rightarrow \widetilde{\mathbb{E}}_{D'} f^2 \propto \widetilde{\mathbb{E}}_D 1_{\{x_j=\sigma\}} f^2 = \widetilde{\mathbb{E}}_D \left(1_{\{x_j=\sigma\}} f\right)^2 \geq 0$$

$\deg f \leq (d-2)/2$

$\deg \mathbf{1}_{\{x_j=\sigma\}} f \leq d/2$

(also captured by LP methods, e.g., Sherali–Adams hierarchies …)

pseudo-covariances are covariances of distributions over $\mathbb{R}^n$

claim

there exists a (Gaussian) distr. $\{\xi\}$ over $\mathbb{R}^n$ such that

$$\tilde{\mathbb{E}}_D x = \mathbb{E}\,\xi \quad \text{and} \quad \tilde{\mathbb{E}}_D x x^T = \mathbb{E}\,\xi\xi^T$$

consequence: $\tilde{\mathbb{E}}_D q = \mathbb{E}_{\{\xi\}} q$
for every q of deg. 2

proof

let $\mu = \tilde{\mathbb{E}}_D x$ and $M = \tilde{\mathbb{E}}_D (x-\mu)(x-\mu)^T$

choose $\{\xi\}$ to be Gaussian with mean μ and covariance M

matrix M p.s.d. because $v^T M v = \tilde{\mathbb{E}}_D (v^T x)^2 \geq 0$ for all $v \in \mathbb{R}^n$

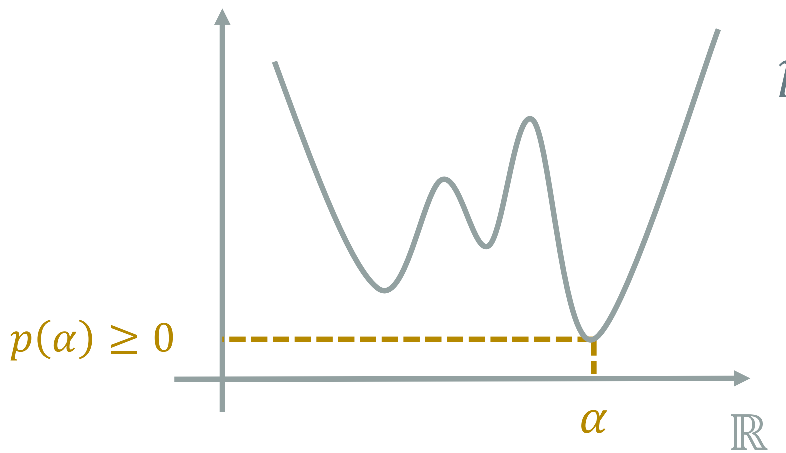
square of linear form

pseudo-distr.'s satisfy (compositions of) low-deg. univariate properties

claim
for every univariate $p \geq 0$ over $\mathbb{R}$ and every n-variate polynomial q with $\deg p \cdot \deg q \leq d$,

$$\tilde{\mathbb{E}}_D p\big(q(x)\big) \geq 0$$

useful class of non-local higher-deg. inequalities



enough to show: p is sum of squares

proof by induction on $\deg p$

choose: minimizer α of p

then: p= $p(\alpha) + (x-\alpha)^2 \cdot p'$ for some polynomial P' with $\deg p' < \deg p$

squares

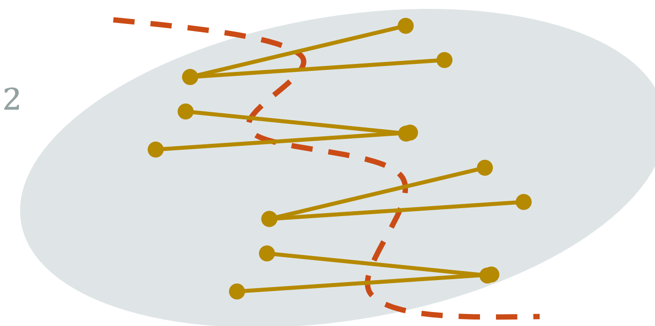
sum of squares by ind. hyp.

MAX CUT

given: deg.-d pseudo-distr. D over $\{\pm 1\}^n$, satisfies $\{L_G = 1 - \varepsilon\}$

goal: find $y \in \{\pm 1\}^n$ with $L_G(y) \geq 1 - O(\sqrt{\varepsilon})$ [Goeman-Williamson]

$$L_G = \frac{1}{|E(G)|} \sum_{ij \in E(G)} \frac{1}{4}\left(x_i - x_j\right)^2$$



algorithm

sample from Gaussian distr. $\{\xi\}$ over $\mathbb{R}^n$ with $\mathbb{E}\, \xi\xi^T = \widetilde{\mathbb{E}}_D\, xx^T$

output $y = \operatorname{sgn} \xi$

analysis

claim: $\mathbb{P}_D\{x_i \neq x_j\} = 1 - \eta \Rightarrow \mathbb{P}\{y_i \neq y_j\} \geq 1 - O(\sqrt{\eta})$

proof: $\{\xi_i, \xi_j\}$ satisfies $-\mathbb{E}\, \xi_i \xi_j = -\widetilde{\mathbb{E}}_D x_i x_j = 1 - O(\eta)$ and $\mathbb{E}\xi_i^2 = \mathbb{E}\xi_j^2 = 1$
$\rightarrow$ (tedious calculation) $\rightarrow$ $\mathbb{P}\{\operatorname{sgn} \xi_i \neq \operatorname{sgn} \xi_j\} \geq 1 - O(\sqrt{\eta})$

low global correlation in (pseudo-)distributions

[Barak-Raghavendra-S., Raghavendra-Tan]

claim

$\forall r.\ \exists$ deg.-$(d-2r)$ pseudo-distribution D', obtained by conditioning D,

$$\operatorname{Avg}_{i,j\in[n]} I_{D'}(x_i,x_j) \le 1/r$$

mutual information: $I(x,y) = H(x) - H(x|y)$

proof

potential $\operatorname{Avg}_{i\in[n]} H(x_i)$; greedily condition on variables to maximize potential decrease until global correlation is low

how often do we need to condition?

$$\text{potential decrease} \ge \operatorname{Avg}_{i\in[n]} H(x_i) - \operatorname{Avg}_{j\in[n]}\operatorname{Avg}_{i\in[n]} H(x_i \mid x_j)$$
$$= \operatorname{Avg}_{i,j\in[n]} I_{D'}(x_i,x_j)$$

→ only need to condition $\le r$ times

MAX BISECTION $d = 1/\varepsilon^{O(1)}$ [Raghavendra-Tan]

given: deg.-d pseudo-distr. D over $\{\pm 1\}^n$, satisfies $\{L_G = 1 - \varepsilon, \sum_i x_i = 0\}$

goal: find $y \in \{\pm 1\}^n$ with $L_G(y) \geq 1 - O(\sqrt{\varepsilon})$ and $\sum_i y_i = 0$

algorithm

let D' be conditioning of D with global correlation $\leq \varepsilon^{O(1)}$
sample Gaussian $\{\xi\}$ with same deg.-2 moments as D'
output y with $y_i = \operatorname{sgn}(\xi_i - t_i)$ (choose $t_i \in \mathbb{R}$ so that $\mathbb{E}\, y_i = \tilde{\mathbb{E}}_D x_i$)

analysis

($t_i = 0$ is worst case → same analysis as MAX CUT)

almost as before: $\mathbb{P}_{D'}\{x_i \neq x_j\} \geq 1 - \eta \Rightarrow \mathbb{P}\{y_i \neq y_j\} \geq 1 - O(\sqrt{\eta})$

new: $I(x_i, x_j) \leq \varepsilon^{O(1)} \Rightarrow \mathbb{E} y_i y_j = \tilde{\mathbb{E}}\, x_i x_j \pm \varepsilon^{O(1)}$

$\tilde{\mathbb{E}}(\sum_i x_i)^2 = 0$

→ $\mathbb{E}|\sum_i y_i| \leq (\mathbb{E}(\sum_i y_i)^2)^{1/2} = \left(\tilde{\mathbb{E}}(\sum_i x_i)^2\right)^{1/2} + \varepsilon^{O(1)} \cdot n = \varepsilon^{O(1)} \cdot n$
→ get bisection y' from y by correcting $\varepsilon^{O(1)}$ fraction of vertices

SUM-OF-SQUARES method and approximation algorithms II

David Steurer
Cornell

Cargese Workshop, 2014

sparse vector

given:	linear subspace $U \subseteq \mathbb{R}^n$ (represented by some basis), parameter $k \in [n]$
promise:	$\exists v_0 \in U$ such that v_0 is k-sparse (and $v_0 \in \{0, \pm 1\}^n$)
goal:	find k-sparse vector $v \in U$

efficient approximation algorithm for $k = \Omega(n)$ would be major step toward refuting Khot's Unique Games Conjecture and improved guarantees for MAX CUT, VERTEX COVER, …

planted / average-case version (benchmark for unsupervised learning tasks)
subspace U spanned by $d - 1$ random vectors and some k-sparse vector v_0

previous best algorithms only work for very sparse vectors $\frac{k}{n} \leq 1/\sqrt{d}$

[Spielman-Wang-Wright, Demanet-Hand]

here: deg.-4 pseudo-distributions work for $\frac{k}{n} = \Omega(n)$ up to $d \leq O(\sqrt{n})$

[Barak-Kelner-S.]

analytical proxy for sparsity

(tight if $v \in \{0, \pm 1\}^n$)

if vector v is k-sparse then $\frac{\|v\|_\infty}{\|v\|_1} \geq \frac{1}{k}$, $\frac{\|v\|_2^2}{\|v\|_1^2} \geq \frac{1}{k}$, and $\frac{\|v\|_4^4}{\|v\|_2^4} \geq \frac{1}{k}$

limitations of ℓ_∞/ℓ_1 (previous best algorithm; exact via linear programming)

v = sum of d random ± 1 vectors with same first coordinate

$\|v\|_\infty \geq d$, $\|v\|_1 \leq d + n\sqrt{d}$ → ratio $\approx \frac{\sqrt{d}}{n}$

→ ℓ_∞/ℓ_1 algorithm fails for $\frac{k}{n} \geq \frac{1}{\sqrt{d}}$

limitations of std. SDP relaxation for ℓ_2/ℓ_1 (best proxy for sparsity)

"ideal object": distribution D over ℓ_2 unit sphere of subspace U

ℓ_1-constraint: $\mathbb{E}_D \|v\|_1^2 \leq k$

tractable relaxation: $\sum_{i,j} \left|\mathbb{E}_D v_i v_j\right| \leq k$

not a low-deg. polynomial in v

→ unclear how to represent

(also NP-hard in worst-case)

[d'Aspremont-El Ghaoui-Jordan-Lanckriet]

but: for uniform distr. D over ℓ_2 sphere of d-dim. rand. subspace

$\sum_{i,j} \left|\mathbb{E}_D v_i v_j\right| \approx \frac{n}{\sqrt{d}}$ → same limitation as ℓ_∞/ℓ_1

degree-d SOS relaxation for ℓ_4/ℓ_2

deg.-d pseudo-distr. $D\colon \{v \in U;\ \|v\|_2 = 1\} \to \mathbb{R}$ over unit ℓ_2-sphere of U

notation: $\tilde{\mathbb{E}}_D f \coloneqq \int_{\left\{ \substack{v\in U;\\ \|v\|=1} \right\}} D \cdot f$ (only consider polynomials → easy to integrate)

non-negativity: $\tilde{\mathbb{E}}_D h(v)^2 \geq 0$ for every h of deg. $\leq d/2$

normalization: $\tilde{\mathbb{E}}_D 1 = 1$

pseudo-distribution satisfies $\{\|v\|_4^4 = 1/k\}$

orthogonality: $\tilde{\mathbb{E}}_D \left(\|v\|_4^4 - \frac{1}{k}\right) \cdot g(v) = 0$ for every g of deg. $\leq d - 4$

degree-d SOS relaxation for ℓ_4/ℓ_2

deg.-d pseudo-distr. $D\colon \{v \in U;\, \|v\|_2 = 1\} \to \mathbb{R}$ over unit ℓ_2-sphere of U

notation: $\tilde{\mathbb{E}}_D f \coloneqq \int_{\left\{\substack{v\in U;\\ \|v\|=1}\right\}} D \cdot f$ (only consider polynomials → easy to integrate)

non-negativity: $\tilde{\mathbb{E}}_D h(v)^2 \geq 0$ for every h of deg. $\leq d/2$

normalization: $\tilde{\mathbb{E}}_D 1 = 1$

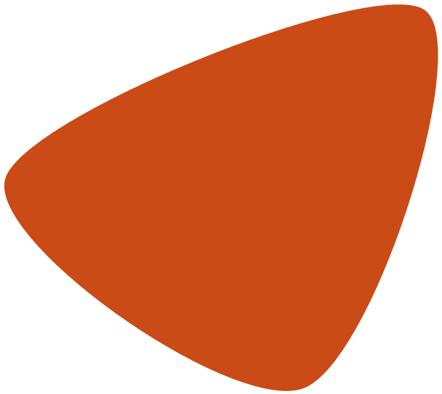
pseudo-distribution satisfies $\{\|v\|_4^4 = 1/k\}$

orthogonality: $\tilde{\mathbb{E}}_D \left(\|v\|_4^4 - \frac{1}{k}\right) \cdot g(v) = 0$ for every g of deg. $\leq d - 4$

how to find pseudo-distributions?

set of deg.-d pseudo-distributions
= convex set with $n^{O(d)}$-time separation oracle

separation problem
given: function D (represented as deg.-d polynomial)
check: quadratic form $f \mapsto \tilde{\mathbb{E}}_D f^2$ is p.s.d. or output violated constraint $\tilde{\mathbb{E}}_D f^2 < 0$



degree-d SOS relaxation for ℓ_4/ℓ_2

deg.-d pseudo-distr. $D\colon \{v \in U;\, \|v\|_2 = 1\} \to \mathbb{R}$ over unit ℓ_2-sphere of U

notation: $\tilde{\mathbb{E}}_D f \coloneqq \int_{\left\{ \substack{v \in U; \\ \|v\|=1} \right\}} D \cdot f$ (only consider polynomials → easy to integrate)

non-negativity: $\tilde{\mathbb{E}}_D h(v)^2 \geq 0$ for every h of deg. $\leq d/2$

normalization: $\tilde{\mathbb{E}}_D 1 = 1$

pseudo-distribution satisfies $\{\|v\|_4^4 = 1/k\}$

orthogonality: $\tilde{\mathbb{E}}_D \left(\|v\|_4^4 - \frac{1}{k} \right) \cdot g(v) = 0$ for every g of deg. $\leq d - 4$

how to use pseudo-distributions?

rule of thumb: set of deg.-d pseudo-moments $\left\{ \tilde{\mathbb{E}}_D f \mid \deg f \leq d \right\}$ difficult*
to distinguish / separate from deg.-d moments of actual distr. of solutions

also: values $\left\{ \tilde{\mathbb{E}}_D f \mid \deg f > d \right\}$ do not carry additional information
→ no need to look at them

(* unless you invest $n^{\Omega(d)}$ time to distinguish)

degree-d SOS relaxation for ℓ_4/ℓ_2

deg.-d pseudo-distr. $D\colon \{v \in U; \|v\|_2 = 1\} \to \mathbb{R}$ over unit ℓ_2-sphere of U

notation: $\tilde{\mathbb{E}}_D f \coloneqq \int_{\left\{ v\in U; \atop \|v\|=1 \right\}} D \cdot f$ (only consider polynomials → easy to integrate)

non-negativity: $\tilde{\mathbb{E}}_D h(v)^2 \geq 0$ for every h of deg. $\leq d/2$

normalization: $\tilde{\mathbb{E}}_D 1 = 1$

pseudo-distribution satisfies $\{\|v\|_4^4 = 1/k\}$

orthogonality: $\tilde{\mathbb{E}}_D \left(\|v\|_4^4 - \frac{1}{k}\right) \cdot g(v) = 0$ for every g of deg. $\leq d - 4$

dual view *(SOS certificates)*

$\left(\|v\|_4^4 - \frac{1}{k}\right) \cdot g + \sum_j h_j^2 = -1$ over $\{v \in U; \|v\|_2 = 1\}$
for some g of deg. $\leq d - 4$ and $\{h_j\}$ of deg. $\leq d/2$

$\Leftrightarrow$ no deg.-d pseudo-distr. exists (→ no solution exists)

for approximation algorithms: need pseudo-distr. to extract approx. solution
(hard to exploit non-existence of SOS certificate directly)

general properties of pseudo-distributions

let $D = \{u, v\}$ be a deg.-4 pseudo-distribution over $\mathbb{R}^n \times \mathbb{R}^n$

following inequalities hold as expected (same as for distributions)

Cauchy–Schwarz inequality

$$\widetilde{\mathbb{E}}_D \langle u, v\rangle \leq \left(\widetilde{\mathbb{E}}_D \|u\|^2\right)^{1/2} \left(\widetilde{\mathbb{E}}_D \|v\|^2\right)^{1/2}$$

Hölder's inequality

$$\widetilde{\mathbb{E}}_D \sum_i u_i^3 \cdot v_i \leq \left(\widetilde{\mathbb{E}}_D \|u\|_4^4\right)^{3/4} \left(\widetilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

ℓ_4-triangle inequality

$$\widetilde{\mathbb{E}}_D \|u + v\|_4^4 \leq \left(\widetilde{\mathbb{E}}_D \|u\|_4^4\right)^{1/4} + \left(\widetilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

claim

let $U' \subseteq \mathbb{R}^n$ be a random d-dim. subspace with $d \ll \sqrt{n}$

let P' be the orthogonal projector into U'

then w.h.p, $\|P'v\|_4^4 = \frac{O(1)}{n}\|v\|_2^4 - \sum_j h_j(v)^2$ over $v \in \mathbb{R}^n$ for h_j's of deg. 4

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

(SOS certificate for classical inequality $\|P'v\|_4^4 \leq \frac{O(1)}{n}\|v\|_2^4$)

proof sketch

basis change: let $x = B^T v$ where B's columns are orthonormal basis of U (so that $P' = BB^T$) → $\|P'v\|_4^4 = \frac{1}{n^2}\sum_i \langle b_i, x\rangle^4$ with $b_1, \dots, b_n$ close to i.i.d. standard Gaussian vectors (so that $\mathbb{E}_b\langle b, x\rangle^2 = \|x\|_2^2$ and $\mathbb{E}_b\langle b, x\rangle^4 = 3 \cdot \|x\|_2^4$)

enough to show: $\frac{1}{n}\sum_{i=1}^n \langle b_i, x\rangle^4 = O(1) \cdot \mathbb{E}_b\langle b, x\rangle^4 - \sum_j h_j'(x)^2$

reduce to deg. 2: $\frac{1}{n}\sum_{i=1}^n \langle b_i^{\otimes 2}, y\rangle^2 \leq O(1) \cdot \mathbb{E}_b\langle b^{\otimes 2}, y\rangle^2$ $(y = x^{\otimes 2})$

→ use concentration inequalities for quadratic forms (aka matrices)

approximation algorithm for planted sparse vector

given: some basis of subspace $U = \operatorname{span} U' \cup \{v_0\} \subseteq \mathbb{R}^n$,
where $U' \subseteq \mathbb{R}^n$ random d-dim. subspace,
and $v_0 \in \mathbb{R}^n$ with $v_0 \perp U'$, $\|v_0\|_4^4 = \frac{1}{k}$, and $\|v_0\|_2^4 = 1$ (e.g., k-sparse)

goal: find unit vector w with $\langle w, v_0\rangle^2 \geq 1 - O(k/n)^{1/4}$

algorithm

compute deg.-4 pseudo-distr. $D = \{v\}$ over unit ball of U satisfying $\left\{\|v\|_4^4 = \frac{1}{k}\right\}$

sample Gaussian distr. $\{w\}$ with $\mathbb{E}\, ww^T = \tilde{\mathbb{E}}_D vv^T$ and renormalize

analysis

claim: $\tilde{\mathbb{E}}_D \langle v, v_0\rangle^2 \geq 1 - O(k/n)^{1/4}$ ($\rightarrow$ Gaussian $\{w\}$ almost 1-dim.)

approximation algorithm for planted sparse vector

given: some basis of subspace $U = \operatorname{span} U' \cup \{v_0\} \subseteq \mathbb{R}^n$,
where $U' \subseteq \mathbb{R}^n$ random d-dim. subspace,
and $v_0 \in \mathbb{R}^n$ with $v_0 \perp U'$, $\|v_0\|_4^4 = \frac{1}{k}$, and $\|v_0\|_2^4 = 1$ (e.g., k-sparse)

goal: find unit vector w with $\langle w, v_0\rangle^2 \geq 1 - O(k/n)^{1/4}$

analysis

claim: $\widetilde{\mathbb{E}}_D \langle v, v_0\rangle^2 \geq 1 - O(k/n)^{1/4}$ (→ Gaussian $\{w\}$ almost 1-dim.)

$$
\begin{aligned}
\frac{1}{k^{1/4}} &= \left(\widetilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4} && (D \text{ satisfies } \{\|v\|_4^4 = 1/k\}) \\
&= \left(\widetilde{\mathbb{E}}_D \|\langle v, v_0\rangle v_0 + P'v\|_4^4\right)^{1/4} && (\text{same function}) \\
&\leq \left(\widetilde{\mathbb{E}}_D \|\langle v, v_0\rangle v_0\|_4^4\right)^{1/4} + \left(\widetilde{\mathbb{E}}_D \|P'v\|_4^4\right)^{1/4} && (\ell_4\text{-triangle inequ.}) \\
&\leq \frac{1}{k^{1/4}} \cdot \left(\widetilde{\mathbb{E}}_D \langle v, v_0\rangle^4\right)^{1/4} + \frac{O(1)}{n^{1/4}} && (\text{SOS cert. for } U')
\end{aligned}
$$

→ $\widetilde{\mathbb{E}}_D \langle v, v_0\rangle^4 \geq 1 - O(k/n)^{1/4}$

→ $\widetilde{\mathbb{E}}_D \langle v, v_0\rangle^2 \geq 1 - O(k/n)^{1/4}$ (because $\langle v, v_0\rangle^4 = (1 - \|P'v\|_2^2)\langle v, v_0\rangle^2$)

general properties of pseudo-distributions

let $D = \{u, v\}$ be a deg.-4 pseudo-distribution over $\mathbb{R}^n \times \mathbb{R}^n$

following inequalities hold as expected (same as for distributions)

Cauchy–Schwarz inequality

$$\tilde{\mathbb{E}}_D \langle u, v \rangle \leq \left(\tilde{\mathbb{E}}_D \|u\|^2\right)^{1/2} \left(\tilde{\mathbb{E}}_D \|v\|^2\right)^{1/2}$$

Hölder's inequality

$$\tilde{\mathbb{E}}_D \sum_i u_i^3 \cdot v_i \leq \left(\tilde{\mathbb{E}}_D \|u\|_4^4\right)^{3/4} \left(\tilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

ℓ_4-triangle inequality

$$\tilde{\mathbb{E}}_D \|u + v\|_4^4 \leq \left(\tilde{\mathbb{E}}_D \|u\|_4^4\right)^{1/4} + \left(\tilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

products of pseudo-distributions

claim

suppose $D, D' \colon \Omega \to \mathbb{R}$ is deg.-d pseudo-distr. over Ω

then, $D \otimes D' \colon \Omega \times \Omega \to \mathbb{R}$ is deg.-d pseudo-distr. over $\Omega \times \Omega$

proof

tensor products of positive semidefinite matrices
are positive semidefinite

let $D = \{u, v\}$ be a deg.-2 pseudo-distribution over $\mathbb{R}^n \times \mathbb{R}^n$

Cauchy–Schwarz inequality

$$\widetilde{\mathbb{E}}_D \langle u, v\rangle \leq \left(\widetilde{\mathbb{E}}_D \|u\|_2^2\right)^{1/2} \left(\widetilde{\mathbb{E}}_D \|v\|_2^2\right)^{1/2}$$

proof

$\left(\widetilde{\mathbb{E}}_D \langle u, v\rangle\right)^2$

$= \widetilde{\mathbb{E}}_{D \otimes D} \langle u, v\rangle \langle u', v'\rangle$ $\quad$ ($D \otimes D'$ is product pseudo-distr.)

$= \widetilde{\mathbb{E}}_{D \otimes D} \sum_{ij} u_i v_i u'_j v'_j$

$\leq \frac{1}{2} \widetilde{\mathbb{E}}_{D \otimes D} \sum_{ij} u_i^2 \left(v'_j\right)^2 + \sum_{ij} \left(u'_j\right)^2 v_i^2$ $\quad$ ($2ab = a^2 + b^2 - (a-b)^2$)

$= \frac{1}{2} \widetilde{\mathbb{E}}_{D \otimes D} \|u\|_2^2 \|v'\|_2^2 + \|u'\|_2^2 \|v\|_2^2$

$= \widetilde{\mathbb{E}}_D \|u\|_2^2 \cdot \widetilde{\mathbb{E}}_D \|v\|_2^2$ $\quad$ ($D \otimes D'$ is product pseudo-distr.)

let $D = \{u, v\}$ be a deg.-4 pseudo-distribution over $\mathbb{R}^n \times \mathbb{R}^n$

Hölder's inequality

$$\tilde{\mathbb{E}}_D \sum_i u_i^3 \cdot v_i \le \left(\tilde{\mathbb{E}}_D \|u\|_4^4\right)^{3/4} \left(\tilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

proof

$$\begin{aligned}
&\tilde{\mathbb{E}}_D \textstyle\sum_i u_i^3 \cdot v_i \\
&\le \left(\tilde{\mathbb{E}}_D \textstyle\sum_i u_i^4\right)^{1/2} \cdot \left(\tilde{\mathbb{E}}_D \textstyle\sum_i u_i^2 \cdot v_i^2\right)^{1/2} && \text{(Cauchy-Schwarz)} \\
&\le \left(\tilde{\mathbb{E}}_D \textstyle\sum_i u_i^4\right)^{1/2} \cdot \left(\tilde{\mathbb{E}}_D \textstyle\sum_i u_i^4 \cdot \tilde{\mathbb{E}}_D \textstyle\sum_i v_i^4\right)^{1/4} && \text{(Cauchy-Schwarz)}
\end{aligned}$$

we also used:

$\{u, v\}$ deg-4 pseudo-distr. → $\{u \otimes u, u \otimes v\}$ deg.-2 pseudo-distr.

(every deg.-2 poly. in $\{u \otimes u, u \otimes v\}$ is deg.-4 poly. in $\{u, v\}$)

let $D = \{u, v\}$ be a deg.-4 pseudo-distribution over $\mathbb{R}^n \times \mathbb{R}^n$

ℓ_4-triangle inequality

$$\left(\tilde{\mathbb{E}}_D \|u + v\|_4^4\right)^{1/4} \leq \left(\tilde{\mathbb{E}}_D \|u\|_4^4\right)^{1/4} + \left(\tilde{\mathbb{E}}_D \|v\|_4^4\right)^{1/4}$$

proof

expand $\|u + v\|_4^4$ in terms of $\sum_i u_i^4, \sum_i u_i^3 v_i, \sum_i u_i^2 v_i^2, \sum_i u_i v_i^3, \sum_i v_i^4$

bound pseudo-expect. of "mixed terms" using Cauchy-Schwarz / Hölder

check that total is equal to right-hand side

tensor decomposition

for simplicity: orthonormal and $m = n$

given: tensor $T \approx \sum_i a_i^{\otimes 4}$ (in spectral norm) for nice $a_1, \dots, a_m \in \mathbb{R}^n$
goal: find set of vectors $B \approx \{\pm a_1, \dots, \pm a_m\}$

approach

show "uniqueness": $\sum_i a_i^{\otimes 4} \approx \sum_i b_i^{\otimes 4} \Rightarrow \{\pm a_1, \dots, \pm a_m\} \approx \{\pm b_1, \dots, \pm b_m\}$

show that uniqueness proof translates to SOS certificate

→ any pseudo-distribution over decomposition is "concentrated" on unique decomposition $\{\pm a_1, \dots, \pm a_m\}$

→ recover decomposition by reweighing pseudo-distribution by $\log n$ degree polynomial (approximation to δ function)