Tensor decompositions, sum-of-squares proofs, and spectral algorithms

David Steurer
Cornell

Sam B. Hopkins
Cornell

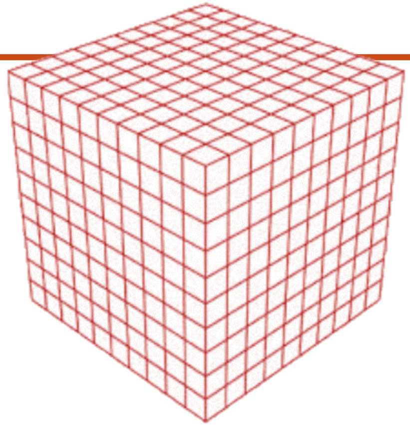
Tengyu Ma
Princeton

Tselil Schramm
Berkeley

Jonathan Shi
Cornell

Quarterly Theory Workshop, Northwestern, May 2016

tensor multi-index array of numbers (typically ≥ 3 indices/modes)

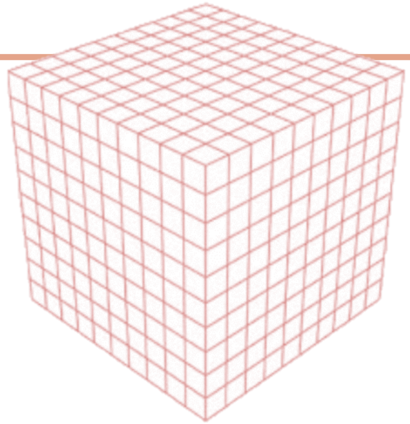


$$T = \sum_{i,j,k \in [d]} T_{ijk} \cdot e_i \otimes e_j \otimes e_k \in \left(\mathbb{R}^d\right)^{\otimes 3}$$

$$a \otimes b \otimes c = \sum_{i,j,k \in [d]} \langle a, e_i \rangle \langle b, e_j \rangle \langle c, e_k \rangle \cdot e_i \otimes e_j \otimes e_k$$

standard basis $e_1, \dots, e_d \in \mathbb{R}^d$

tensor multi-index array of numbers (typically ≥ 3 indices/modes)



$$T = \sum_{i,j,k \in [d]} T_{ijk} \cdot e_i \otimes e_j \otimes e_k \in \left(\mathbb{R}^d\right)^{\otimes 3}$$

$$a \otimes b \otimes c = \sum_{i,j,k \in [d]} \langle a, e_i \rangle \langle b, e_j \rangle \langle c, e_k \rangle \cdot e_i \otimes e_j \otimes e_k$$

standard basis $e_1, \dots, e_d \in \mathbb{R}^d$

natural shape of data

moments of multivariate distributions $T = \mathbb{E}_{x \sim D} x^{\otimes 3}$

coefficients of multivariate polynomials $T = \sum_{ijk} T_{ijk} \cdot x_i x_j x_k$

states of composite quantum systems $|\psi\rangle \in A \otimes B \otimes C$

"deep learning" frameworks: torch / theano / tensorflow

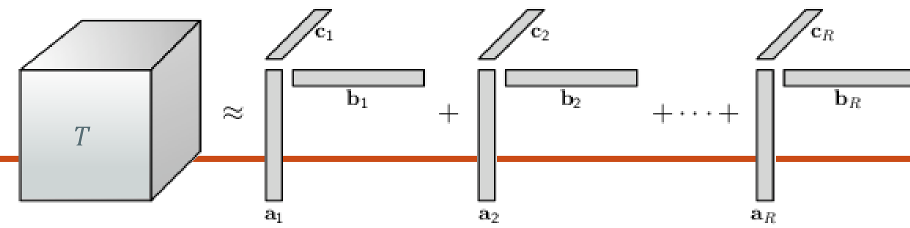
"tensors are the new matrices" tie together wide range of disciplines

"algorithms for the tensor age" hope to repeat success for matrices

tensor decomposition *(tensor rank)*

given 3-tensor T, find as few vectors $\{a_i, b_i, c_i\}_{i\in[r]}$ as possible such that

$$T = \sum_{i=1}^{r} a_i \otimes b_i \otimes c_i$$



intuition: explain data in simplest way possible

$$M = AB^\top$$
$$\Leftrightarrow M = (AU)(BU^{-1})^\top$$

key advantage over matrix rank/factorization

matrix factorization suffers from “rotation problem”

in contrast: tensor decomposition often *unique*

key challenge

tensor decomposition is NP-hard in worst case

→ cannot hope for same theory as for matrices

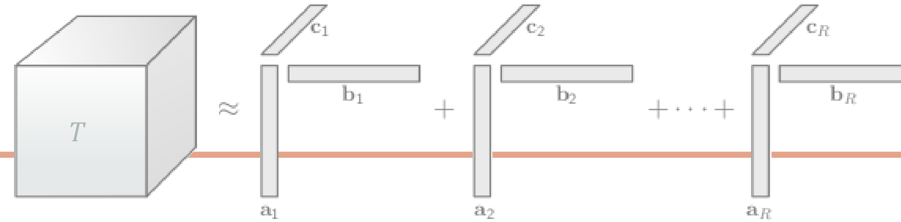
but: can still hope for algorithms with strong provable guarantees

tractability appears to go hand in hand with uniqueness

tensor decomposition (tensor rank)

given 3-tensor T, find as few vectors $\{a_i, b_i, c_i\}_{i\in[r]}$ as possible such that

$$T = \sum_{i=1}^{r} a_i \otimes b_i \otimes c_i$$



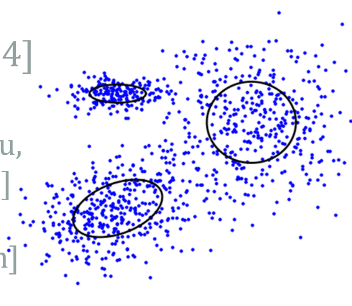
poly-time & practical unsupervised learning via tensor decomposition

blind-source separation, independent component analysis [Leurgans; Lathauwer, Castaing, Cardoso'07]

Gaussian mixtures [Bhaskara-Charikar-Moitra-Vijayaraghavan'14]

topic modelling *(latent Dirichlet allocation)* [Anandkumar, Ge, Hsu, Kakade, Telgarsky'14]

phylogenetic tree / hidden Markov model [Chang'96; Mossel, Roch]



moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k \coloneqq \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

(reformulation of tensor decomposition problem)

under what conditions on the vectors and k can we solve this problem efficiently and robustly?

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

linearly independent vectors (thus, $n \le d$)

wlog $a_1, \dots, a_n$ orthonormal (apply linear transformation $\frac{1}{\sqrt{n}} \mathcal{M}_2^{-1/2}$)

spectral algorithm for $k = 3$ *(matrix diagonalization)*

[Jennrich via Harshman'70; Leurgans-Ross-Abel'93; rediscovered many times]

key challenge: decompose <u>overcomplete</u> tensors, i.e., rank ≫ dimension

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

random unit vectors for rank $n \gg d$ and $k = 3$ moments

	largest rank n	*running time*	
spectral algorithm	$C \cdot d$	$2^{C^2} \cdot d^3$	[Anandkumar-Ge-Janzamin'15]
tensor power iteration	$d^{1.5}$	(only local convergence)	[Anandkumar-Ge-Janzamin'15]
sum-of-squares	$d^{1.5}$	$d^{\log d}$	[Ge-Ma'15 analysis of Barak-Kelner-S.'15 algorithm]

∃ poly-time algorithm for rank $n = d^{1.01}$?

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

random unit vectors for rank $n \gg d$ and $k = 3$ moments

	largest rank n	*running time*	
spectral algorithm	$C \cdot d$	$2^{C^2} \cdot d^3$	[Anandkumar-Ge-Janzamin'15]
tensor power iteration	$d^{1.5}$	(only local convergence)	[Anandkumar-Ge-Janzamin'15]
sum-of-squares	$d^{1.5}$	$d^{\log d}$	[Ge-Ma'15 analysis of Barak-Kelner-S.'15 algorithm]

this talk:

sum-of-squares	$d^{1.5}$	$d^{O(1)}$	[Ma-Shi-S'16+]
SOS-flavored spectral	$d^{1.33}$	$d^{1+\omega} \leq d^{3.33}$	[Hopkins-Schramm-Shi-S'16]

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

smoothed unit vectors *(Spielman–Teng smoothed analysis framework)*

assume each vector is independently perturbed by $n^{-O(1)}$ norm Gaussian

poly-time algorithm for $k = 4$ up to rank $n \le d^2$

combines large linear system and spectral algorithm *(FOOBI)*

[Lathauwer, Castaing, Cardoso'07]

assumes exact input; not known to tolerate $n^{-O(1)}$ error

poly-time algorithm for $k = 5$ up to rank $n \le d^2$

spectral algorithm; tolerates $n^{-O(1)}$ error

[Bhaskara-Charikar-Moitra-Vijayaraghavan'14]

this talk: same guarantees as *FOOBI* but tolerate $n^{-O(1)}$ error
based on sum-of-squares

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

general unit vectors

for simplicity: isotropic position $\sum_{i=1}^{n} a_i a_i^\top = \frac{n}{d} \mathrm{Id}$

quasi-poly time algorithm with accuracy ε for $k \geq \varepsilon^{-1} \log\left(\frac{n}{d}\right)$ [Barak-Kelner-S.'15]
based on sum-of-squares

this talk: poly-time algorithm (in size of input) with same recovery guarantees

corollary: overcomplete dictionary learning with constant relative sparsity and constant accuracy in polynomial time

previous best: either sparsity $n^{-\Omega(1)}$ or time $n^{O(\log n)}$ [Barak-Kelner-S.'15]

moment problem for multivariate discrete distributions

hidden: set of vectors $a_1, \dots, a_n \in \mathbb{R}^d$
given: low-degree moments $\mathcal{M}_1, \dots, \mathcal{M}_k$ of uniform distribution over $a_1, \dots, a_n$
find: set of vectors $\approx \{a_1, \dots, a_n\}$

$$\mathcal{M}_k := \frac{1}{n} \sum_{i=1}^{n} a_i^{\otimes k}$$

Jennrich's algorithm on 3rd moments

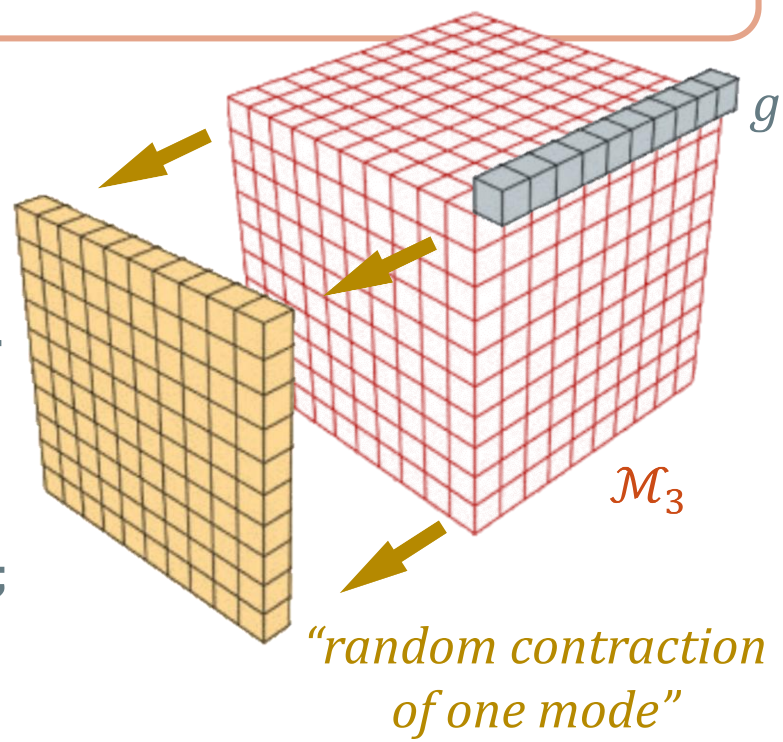
assume $\{a_1, \dots, a_n\}$ orthonormal

let $g \sim \mathcal{N}(0, \mathrm{Id}_d)$ be standard Gaussian vector

then, $(\mathrm{Id} \otimes \mathrm{Id} \otimes g^\top)\mathcal{M}_3 = \frac{1}{n}\sum_i \langle g, a_i \rangle \cdot a_i a_i^\top$

→ every a_i is eigenvector with value $\langle g, a_i \rangle / n$; w.h.p. all eigenvalues distinct

→ *eigendecomposition* recovers $a_1, \dots, a_n$



challenge: what can we do when $n \gg d$ (overcomplete case)?

*approach for **random overcomplete** 3-tensors*

let $a_1, \dots, a_n$ be random unit vectors for $n \ll d^{1.5}$

?

6^{th} moments of $a_1, \dots, a_n$
$= 3^{\text{rd}}$ moments of $a_1^{\otimes 2}, \dots, a_n^{\otimes 2}$

$\mathcal{M}_3 = \sum_i a_i^{\otimes 3}$ → **magic box** → $\mathcal{M}_6 = \sum_i a_i^{\otimes 6}$

(lifts 3^{rd} moments to 6^{th} moments)

→ **Jennrich's algorithm** → $\{a_1^{\otimes 2}, \dots, a_n^{\otimes 2}\}$

($a_1^{\otimes 2}, \dots, a_n^{\otimes 2}$ orthogonal enough for Jennrich's algorithm)

*approach for **random overcomplete** 3-tensors*

let $a_1, \dots, a_n$ be random unit vectors for $n \ll d^{1.5}$

"ideal implementation"

(ignore efficiency for now)

$\mathcal{M}_3 = \sum_i a_i^{\otimes 3}$ → ~~**magic box**~~ → $\cancel{\mathcal{M}_6 = \sum_i a_i^{\otimes 6}}$ $\mathbb{E}_{u\sim D} u^{\otimes 6}$ → **Jennrich's algorithm** → $\{a_1^{\otimes 2}, \dots, a_n^{\otimes 2}\}$

find distribution D over unit sphere subject to $\mathbb{E}_{u\sim D} u^{\otimes 3} = \mathcal{M}_3$

($a_1^{\otimes 2}, \dots, a_n^{\otimes 2}$ orthogonal enough for Jennrich's algorithm)

*approach for **random overcomplete** 3-tensors*

let $a_1, \dots, a_n$ be random unit vecto

"ideal implementation"

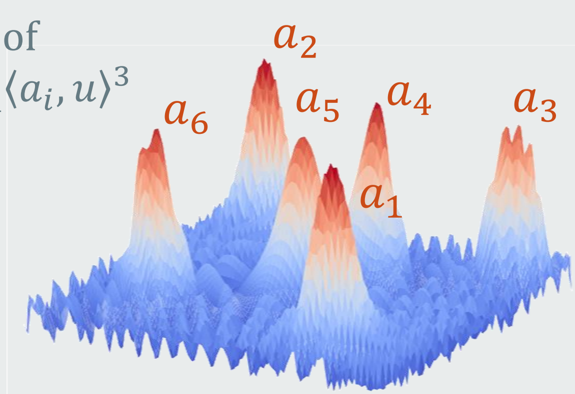
(ignore efficiency for now)

$\mathcal{M}_3 = \sum_i a_i^{\otimes 3} \rightarrow$ ~~**magic box**~~ $\rightarrow$

find distribution D over unit sphere subject to $\mathbb{E}_{u \sim D} u^{\otimes 3} = \mathcal{M}_3$

claim: $\mathbb{E}_{D(u)} u^{\otimes 6} \approx \mathcal{M}_6$

plot of $\sum_{i=1}^n \langle a_i, u \rangle^3$



proof:

$$\langle \mathcal{M}_3, \mathbb{E}_{D(u)} u^{\otimes 3} \rangle = \frac{1}{n} \mathbb{E}_{D(u)} \textstyle\sum_{i=1}^n \langle a_i, u \rangle^3$$

$$\langle \mathcal{M}_3, \mathcal{M}_3 \rangle = \frac{1}{n^2} \textstyle\sum_{i,j \in [n]} \langle a_i, a_j \rangle^3 = \frac{1 \pm o(1)}{n}$$

$$\rightarrow \mathbb{E}_{D(u)} \textstyle\sum_{i=1}^n \langle a_i, u \rangle^3 = 1 \pm o(1) \quad (*)$$

crucially: with high prob. over $a_1, \dots, a_n$,

$$\forall u.\ \textstyle\sum_{i=1}^n \langle a_i, u \rangle^3 = \max_{i \in [n]} \langle a_i, u \rangle^3 \pm o(1)$$

therefore, $(*)$ implies

$$\Pr_{D(u)} \left\{ \max_i \langle a_i, u \rangle \geq 1 - o(1) \right\} \geq 1 - o(1)$$

$$\rightarrow \mathbb{E}_{D(u)} \textstyle\sum_{i=1}^n \langle a_i, u \rangle^6 = 1 \pm o(1)$$

...

$$\rightarrow \left\| \mathcal{M}_6 - \mathbb{E}_{D(u)} u^{\otimes 6} \right\| \leq o(1) \cdot \|\mathcal{M}_6\|$$

approach for random overcomplete 3-tensors

let $a_1, \dots, a_n$ be random unit vectors for $n \ll d^{1.5}$

"ideal implementation"

(ignore efficiency for now)

$\mathcal{M}_3 = \sum_i a_i^{\otimes 3}$ → ~~**magic box**~~ → $\cancel{\mathcal{M}_6 = \sum_i a_i^{\otimes 6}}$ $\mathbb{E}_{u\sim D} u^{\otimes 6}$ → **Jennrich's algorithm** → $\{a_1^{\otimes 2}, \dots, a_n^{\otimes 2}\}$

find distribution D over unit sphere
subject to $\mathbb{E}_{u\sim D} u^{\otimes 3} = \mathcal{M}_3$

claim: $\mathbb{E}_{D(u)} u^{\otimes 6} \approx \mathcal{M}_6$

($a_1^{\otimes 2}, \dots, a_n^{\otimes 2}$ orthogonal enough
for Jennrich's algorithm)

two remaining questions:

1. how to find D efficiently? relax search to *sum-of-squares pseudo-distributions*

2. can Jennrich tolerate this kind of error? ***no, error is too large!***

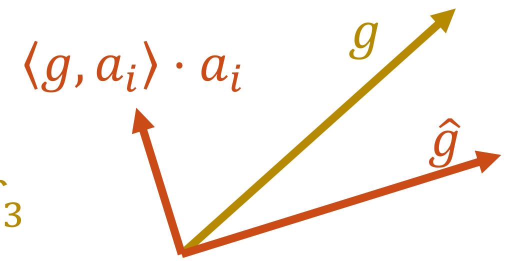
→ add *maximum entropy constraint* $\left\| \mathbb{E}_{u\sim D} u^{\otimes 4} \right\|_{\text{spectral}} \le \frac{1+o(1)}{n}$

robust analysis of Jennrich's algorithm

$a_1, \dots, a_n \in \mathbb{R}^d$ orthonormal; moments $\mathcal{M}_k = \frac{1}{n}\sum_{i=1}^n a_i^{\otimes k}$

distribution D over sphere; moments $\widetilde{\mathcal{M}}_k = \mathbb{E}_{D(u)} u^{\otimes k}$

$\boxed{\begin{array}{l} \text{suppose } \left\| \widetilde{\mathcal{M}}_3 - \mathcal{M}_3 \right\|_F \le o(1) \cdot \|\mathcal{M}_3\|_F \text{ and } \left\|\widetilde{\mathcal{M}}_2\right\|_{\text{spectral}} \le O(1)/n. \\ \text{then, for most } i \in [n] \text{, with probability } \frac{1}{n^{O(1)}} \text{ over the choice } g \sim \mathcal{N}(0, \mathrm{Id}_{d^2}), \\ \qquad (\mathrm{Id}_d \otimes \mathrm{Id}_d \otimes g^\top)\widetilde{\mathcal{M}}_3 \text{ has top eigenvector} \approx a_i \end{array}}$



$$(\mathrm{Id}_d \otimes \mathrm{Id}_d \otimes g^\top)\widetilde{\mathcal{M}}_3$$

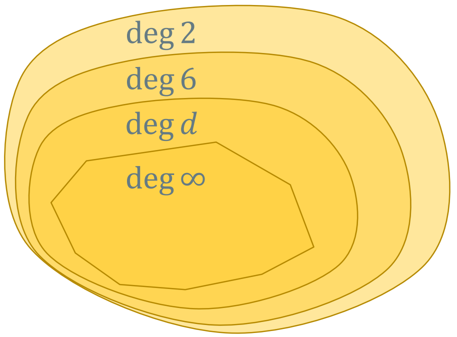
$$= \langle g, a_i \rangle \underbrace{(\mathrm{Id}_d \otimes \mathrm{Id}_d \otimes a_i^\top)\widetilde{\mathcal{M}}_3}_{\approx \frac{1}{n} a_i a_i^\top} + \underbrace{(\mathrm{Id}_d \otimes \mathrm{Id}_d \otimes \hat{g}^\top)\widetilde{\mathcal{M}}_3}_{\|\cdot\|_{\text{spectral}} \le \frac{\sqrt{\log d}}{n}}$$

overwhelms noise with

probability $e^{-(\sqrt{\log d})^2} \ge d^{-O(1)}$

□

probability theory meets complexity theory



- *low-complexity events* always have *nonnegative probability*
- *high-complexity events* may have *negative probability*

degree-k pseudo-distribution over unit sphere $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d$

- finitely supported function $D\colon \mathbb{S}^{d-1} \to \mathbb{R}$
- $\sum_u D(u) = 1$ (sum is only over support of D)
- $\sum_u D(u) \cdot f(u)^2 \geq 0$ for every $f\colon \mathbb{S}^{d-1} \to \mathbb{R}$ with $\deg f \leq k/2$

notation: $\tilde{\mathbb{E}}_{D(u)} f(u) \overset{\text{def}}{=} \sum_u D(u) \cdot f(u)$ — ***pseudo-expectation*** of f under D

efficiency of pseudo-distributions [Shor, Parrilo, Lasserre]

set of degree-k pseudo-moments has $d^{O(k)}$-time separation oracle;

key step: check k^{th} pseudo-moment satisfies $\tilde{\mathbb{E}}_{D(u)} u^{\otimes k/2} \left(u^{\otimes k/2}\right)^{\top} \succcurlyeq 0$

generalizes best known poly-time algorithms for wide range of problems

degree-k pseudo-distribution over unit sphere $\mathbb{S}^{d-1} \subseteq \mathbb{R}^d$

- finitely supported function $D\colon \mathbb{S}^{d-1} \to \mathbb{R}$
- $\sum_u D(u) = 1$ (sum is only over support of D)
- $\sum_u D(u) \cdot f(u)^2 \geq 0$ for every $f\colon \mathbb{S}^{d-1} \to \mathbb{R}$ with $\deg f \leq k$

notation: $\widetilde{\mathbb{E}}_{D(u)} f(u) \stackrel{\text{def}}{=} \sum_u D(u) \cdot f(u)$

degree-k sum-of-squares proof *of* $\forall u \in \mathbb{S}^{d-1}.\, f(u) \geq g(u)$

functions $h_1, \ldots, h_r$ with $\deg h_1, \ldots, \deg h_r \leq k/2$

$$\forall u \in \mathbb{S}^{d-1}. \qquad f(u) - g(u) = h_1(u)^2 + \cdots + h_r(u)^2$$

duality: pseudo-distributions vs sum-of-squares proofs

if $f \geq g$ has degree-k sos proof, then $\widetilde{\mathbb{E}}_{D(u)} f(u) \geq \widetilde{\mathbb{E}}_{D(u)} g(u)$ for every degree-k pseudo-distribution D

lifting moments higher via sum-of-squares

$\mathcal{M}_3 = \frac{1}{n}\sum_{i=1}^{n} a_i^{\otimes 3}$ for random unit vectors $a_1, \ldots, a_n \in \mathbb{R}^d$ and $n \ll d^{1.5}$

theorem: w.h.p. over $a_1, \ldots, a_n$, every degree-12 pseudo-distribution D with $\widetilde{\mathbb{E}}_{D(u)} u^{\otimes 3} = \mathcal{M}_3$ satisfies $\left\|\widetilde{\mathbb{E}}_{D(u)} u^{\otimes 6} - \mathcal{M}_6\right\| \leq o(1)\|\mathcal{M}_6\|$.

lifting moments higher via sum-of-squares

$\mathcal{M}_3 = \frac{1}{n}\sum_{i=1}^{n} a_i^{\otimes 3}$ for random unit vectors $a_1, \dots, a_n \in \mathbb{R}^d$ and $n \ll d^{1.5}$

theorem: w.h.p. over $a_1, \dots, a_n$, every degree-12 pseudo-distribution D with $\widetilde{\mathbb{E}}_{D(u)} u^{\otimes 3} = \mathcal{M}_3$ satisfies $\left\|\widetilde{\mathbb{E}}_{D(u)} u^{\otimes 6} - \mathcal{M}_6\right\| \leq o(1)\|\mathcal{M}_6\|$.

enough to show (same as in previous proof for probability distributions):

$$\widetilde{\mathbb{E}}_{D(u)} \sum_{i=1}^{n} \langle a_i, u\rangle^3 \geq 1 - o(1) \quad \Rightarrow \quad \widetilde{\mathbb{E}}_{D(u)} \sum_{i=1}^{n} \langle a_i, u\rangle^6 \geq 1 - o(1)$$

w.h.p. over $a_1, \dots, a_n$, the following inequality has degree-12 sos proof

$$\forall u \in \mathbb{S}^{d-1}.\ \sum_{i=1}^{n} \langle a_i, u\rangle^3 \leq \frac{3}{4} + \frac{1}{4}\sum_{i=1}^{n} \langle a_i, u\rangle^6 + o(1)$$

[Ge–Ma'15]

key ingredient is to bound spectral norm of random matrix polynomial

$$\left\|\sum_{i\neq j} \langle a_i, a_j\rangle \cdot a_i a_i^\top \otimes a_j a_j^\top\right\| \leq o(1)$$

prevailing wisdom: ***sum-of-squares is “strictly theoretical”***

sum-of-squares algorithms have nothing to do with practical ones

at odds with tenet of computational complexity that polynomial-time is a good model for practical algorithms

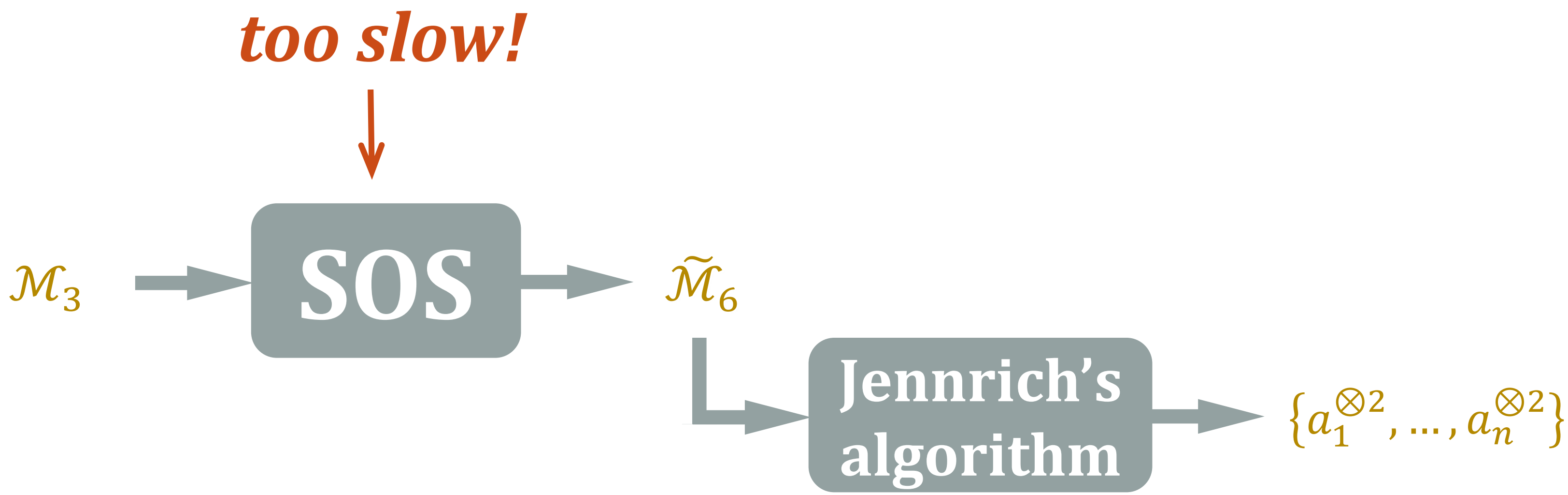
next:

algorithm to decompose random overcomplete 3-tensor
with **close to linear running time** (in size of input)
and guarantees close to those of sum-of-squares

general recipe for new kinds of **fast spectral algorithms inspired by SOS**

*approach for **fast decomposition** of overcomplete 3-tensor*

random unit vectors $a_1, \dots, a_n \in \mathbb{R}^d$ with $n \ll d^{1.5}$; moments $\mathcal{M}_k = \frac{1}{n}\sum_{i=1}^{n} a_i^{\otimes k}$



*approach for **fast decomposition** of overcomplete 3-tensor*

random unit vectors $a_1, \dots, a_n \in \mathbb{R}^d$ with $n \ll d^{1.5}$; moments $\mathcal{M}_k = \frac{1}{n}\sum_{i=1}^n a_i^{\otimes k}$

direct algorithm to fool Jennrich

$$\sum_{ijk} \langle a_i, a_k\rangle\langle a_j, a_k\rangle \cdot (a_i \otimes a_j) \otimes (a_i \otimes a_j) \otimes a_k$$

$$= \sum_{i=1}^n a_i^{\otimes 5} + E$$ with E "random-like"

and $\|E\|_F^2 \leq \frac{n^3}{d^2}$

(huge Frobenius norm)

$\mathcal{M}_3 \to$ ~~SOS~~ $\to \cancel{\widetilde{\mathcal{M}}_6}$

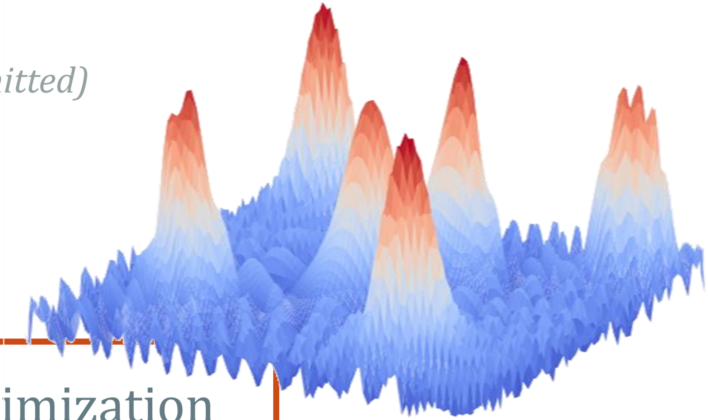
Jennrich's algorithm $\to \{a_1^{\otimes 2}, \dots, a_n^{\otimes 2}\}$

claim: if $n \ll d^{1.33}$ *then E contributes negligible spectral error for Jennrich*

input to Jennrich has "size" d^5 (computing it takes naively $O(d^6)$ time)

exploit tensor structure to implement Jennrich in time $O(d^{1+\omega}) \leq O(d^{3.3\dots})$

meta result* *(* some technical conditions omitted)*



sum-of-squares method (based on semidefinite programming) [Shor, Parrilo, Lasserre]

efficient algorithm to solve polynomial optimization problems that have ***only few global optima***

running time poly(#solutions)

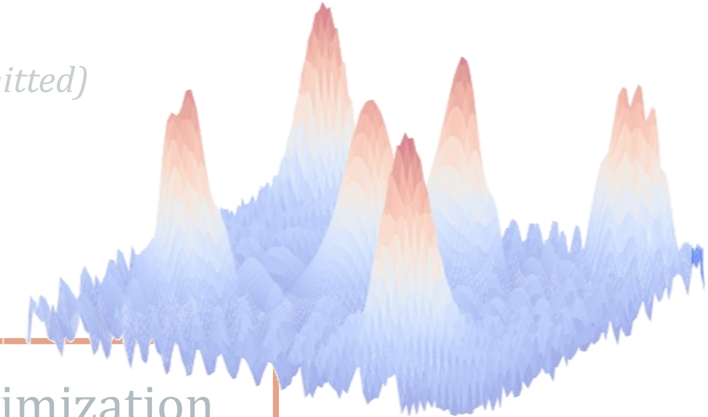
also need ***short sum-of-squares certificate*** for this fact

previous work: running time $n^{O(\log \#\text{solutions})}$
(quasi-poly time for poly #solutions)

[Barak-Kelner-S STOC'15]

 # bad local optima
can be exponential
→ local-search algorithms fail

*meta result** *(* some technical conditions omitted)*



sum-of-squares method (based on semidefinite programming) [Shor, Parrilo, Lasserre]

efficient algorithm to solve polynomial optimization problems that have ***only few global optima***

running time poly(#solutions)

also need ***short sum-of-squares certificate*** for this fact

applications: *unsupervised learning problems tend to have this property*

identifiability: data uniquely determines parameters of model

our work: notion of ***constructive identifiability proofs*** that implies *efficient inference algorithms*

conclusions

tensor decomposition / polynomial optimization via sum-of-squares

sum-of-squares proof for approximate uniqueness (identifiability)

use Jennrich's algorithm (small spectral gaps) as rounding algorithm

fast spectral algorithms via sum-of-squares

fool rounding algorithm by low-degree matrix polynomial of input

exploit tensor structure for fast algebraic operations

conclusions

tensor decomposition / polynomial optimization via sum-of-squares

sum-of-squares proof for approximate uniqueness (identifiability)

use Jennrich's algorithm (small spectral gaps) as rounding algorithm

fast spectral algorithms via sum-of-squares

fool rounding algorithm by low-degree matrix polynomial of input

exploit tensor structure for fast algebraic operations

questions

thank you very much!

random 3-tensors beyond rank $d^{1.5}$?

lower bounds? hard to distinguish from completely random 3-tensors?

smoothed analysis for overcomplete 3-tensors?

strong bounds known for 4-tensors [Lathauwer, Castaing, Cardoso'07]