Tensor Decomposition via SOS

$$T = \sum_i \underset{\text{"1st mode"}}{a_i} \otimes b_i \otimes c_i \in \mathbb{R}^{d \times d \times d}$$

"order 3"
$\underbrace{\hspace{6em}}$
"dimension d"

$a_1 \ldots a_r$
$b_1 \ldots b_r$
$c_1 \ldots c_r$
$\Big\} \in \mathbb{R}^d$

$\{$ "Symmetric" if $a_i = b_i = c_i$ $\}$

$$T(f, g, h) = \sum_i a_i(f) \cdot b_i(g) \cdot c_i(h)$$

$f, g, h \in [d]$

Rank of a tensor T:

$$a_i \otimes b_i \otimes c_i \overset{\text{def}}{\equiv} \text{rank } 1 \text{ tensor}$$

T is rank r if r is the min integer so that T is a sum of at most r rank 1 tensors.

2nd order tensors $\cong$ matrices

$M \in \mathbb{R}^{d \times d}$ has rank $\leq r$ if it can be written as a sum of r rank 1 matrices.

$$\underbrace{a_i b_i^T}$$

Two problems: extremely important primitives in algo num lin algebra.

1) Given $M \in \mathbb{R}^{d \times d}$, find its rank.

2) Given M of rank r, find an explicit rank 1 decomposition of M.

① & ② are polynomial time solvable.

Q: $M = \sum_{i=1}^{r} a_i b_i^T$

Can you calculate $\{a_i, b_i\}_{i=1}^{r}$

Psychometrics: 1904 Charles Spearman

Obs^n: Rank decompositions are not unique in general...

Obs^n: tensor decompositions (order ≥ 3) are unique under some mild conditions.

Problem: Given a rank r tensor $T = \sum_{i=1} a_i \otimes b_i \otimes c_i$

find $a_1 - \!\!-\!\!- a_r$
$b_1 -\!\!-\!\!- b_r$
$c_1 -\!\!-\!\!- c_r$

Q: Given a matrix M : Symmetric has rank $r > 1$.

Obs^n, "EVD" : eigenvalue decomposition is unique if all eigenvalues are distinct.

Obs^n: $M = \sum_{i=1}^{r} \lambda_i \cdot a_i a_i^T$, $\{a_1 \ldots, a_r\}$ are orthonormal, $\lambda_i \neq \lambda_j \;\forall i \neq j$.

then there's a poly algo to compute (arbitrarily good approx.) to $a_1 \ldots a_r$. up to sign.

Desiderata for Algorithms :

Algorithms to be resilient to noise.

1. handle noise that's random/indep

evident.

(2) perturbation bounded in appropriate norms $M' = M + E$, $\|E\| \ll$ small.

(3) outlier-robustness (not our focus).

Algorithms for Tensor Decomposition.

Lemma (Jennrich's Algorithm):

Let $\{a_1 \dots a_r\}$ be orthonormal. $\|a_i\|_2 = 1$, $\langle a_i, a_j \rangle = 0$ $\forall i \neq j$

Then, given $T = \sum_i a_i^{\otimes 3}$, there's a poly time algo to recover $\{a_1 \dots a_r\}$ up to permutation.

Proof: "Random Contraction"

Let $g \sim N(0, I)$.

Suppose you get $M_g = \sum_i \langle a_i, g \rangle \cdot a_i a_i^T$

then you're done.

Then ...

$$\boxed{\begin{aligned}
&\mathbb{E}\langle g, \underline{a_i - a_j}\rangle^2 = \mathbb{E}(\langle g, a_i\rangle - \langle g, a_j\rangle)^2 \\
&\qquad = 2 \\
&\Pr\left[|\langle g, a_i - a_j\rangle| \le \frac{\delta}{d^2}\right] \le \frac{\sqrt{2}\,\delta}{d^2} \\
&h\text{: gaussian: } \Pr\left[|h| \le \delta \cdot \sqrt{\mathbb{E}h^2}\right] \le \delta \leftarrow \\
&\Rightarrow \Pr\left[\min_{i \ne j} |\langle g, a_i - a_j\rangle| \ge \frac{\delta}{d^2}\right] \le \underline{1 - \sqrt{2}\,\delta}
\end{aligned}}$$

$\underline{\text{Obs}^n}$: Suppose $T = a^{\otimes 3}$

$$M_g(s,t) \stackrel{\text{def}}{=} \underbrace{\sum_{u=1}^{d} T(s,t,u) \cdot g_u}_{= \langle T(s,t,\cdot), g\rangle}$$

linear operation on tensor

For every s,t, $T(s,t,\cdot) \in \mathbb{R}^d$

$$\sum_u T(s,t,u) \cdot g_u = \sum_u a_s \cdot a_t \cdot a_u \cdot g_u$$

$$= \langle a, g\rangle \cdot a_s \cdot a_t$$

$$M_g = \langle a, g\rangle \cdot a a^T$$

Cor: $T = \sum_i a_i^{\otimes 3}$

$$\langle T(s,t,\cdot), g\rangle = \sum_i a_i(s)\cdot a_i(t)\cdot\langle a_i, g\rangle$$

$$M_g = \sum_i a_i a_i^T \langle a_i, g\rangle.$$

<u>Lemma:</u> There's a poly time algo that takes input $T = \sum_{i=1}^{r} a_i^{\otimes 3}$ where $\{a_1 \ldots a_r\}$ are linearly independent and <u>outputs</u> $\{\hat{a}_1, \ldots, \hat{a}_r\}$ s.t. $\exists$ $\pi: [r] \to [r]$ s.t.

$$\min\{\|\hat{a}_i - a_{\pi(i)}\|, \|\hat{a}_i + a_{\pi(i)}\|\} \le \varepsilon.$$

<u>Proof:</u> Idea is to take 2 random contractions.

g, g'

$$M_g = \sum_i \underline{\langle a_i, g\rangle} \underline{a_i a_i^T}$$

$$M_{g'} = \sum_i \underline{\langle a_i, g'\rangle} \underline{a_i a_i^T}$$

$$U = \begin{pmatrix} a_1 & \text{---} & a_d \end{pmatrix}$$

$$M_g = \underbrace{U \cdot D_g} \cdot U^T \quad , \quad M_{g'} = U \cdot D_{g'} \cdot U^T$$

$$D = \begin{pmatrix} \langle a_i, g \rangle \end{pmatrix}$$

$$M_g^{-1} \cdot M_{g'} = (U^T)^{-1} \cdot D_g^{-1} \cdot \underline{U^{-1} \cdot U} \cdot D_{g'} \cdot U^T$$

$$= \underbrace{(U^T)^{-1} \cdot \underbrace{D_g^{-1} \cdot D_{g'}} \cdot U^T}_{\text{eigenvalue decomposition with distinct eigenvalues!}}$$

$$\boxed{\frac{\langle a_i, g' \rangle}{\langle a_i, g \rangle}}$$

"Undercomplete Tensor" if $\underline{r \leq d}$.

"Overcomplete tensor" if $r > d$.

---

① 3^{rd} order tensors in d-dim have rank at most $O(d^2)$.

<u>Pf sketch:</u>

$$d\left\{\underbrace{\boxed{\quad T(i,j,k) \quad}}_{d^2}\right. \in \mathbb{R}^{d \times d^2}$$

<u>"flattenings"</u>

$$= \sum_{i=1}^{d} u_i \cdot v_i^T.$$

$$v_i = \sum_j y_j \cdot z_j^T$$

② <u>random</u> tensors have rank $\Omega(d^2)$..

<u>Facts:</u> 1) Computing tensor rank is NP-hard

"Most tensor problems are NP-hard"

<u>Hillar-Lim '08</u> ??

The SoS approach to tensor decomposition.

Key idea: relationship to polynomial maximization over $\{x : \|x\|_2^2 \le 1\}$.

Lemma: Let $a_1 — a_r \in \mathbb{R}^d$, orthonormal.

$$p(x) = \langle T, x^{\otimes 3} \rangle$$

hom deg 3 polynomial

$$= \underline{\textstyle\sum_{i,j,k} T_{i,j,k} \cdot \underbrace{x_i x_j x_k}.}$$

Then: $\left\{ z : \underset{\substack{\| \\ p(z)}}{\langle T, z^{\otimes 3} \rangle} \geq \max_{x : \|x\|_2^2 \le 1} \langle T, x^{\otimes 3} \rangle \right\}$

$$= \{ a_1, \dots, a_r \}.$$

Pf: $a_1 — a_r$ $\qquad \underline{r = d}$

$$x = \sum_{i=1}^{r} \alpha_i \cdot a_i$$

$$1 = \|x\|_2^2, \qquad \sum_{i=1}^{r} \alpha_i^2 = 1$$

Then, $p(x) = \langle T, x^{\otimes 3} \rangle$

$$= \sum_i \langle a_i^{\otimes 3}, x^{\otimes 3} \rangle$$

$$= \sum_i \langle a_i, x \rangle^3$$

$$\sum_{s,t,u} a_i(s)\, a_i(t)\, a_i(u) \cdot x(s) \cdot x(t)\, x(u)$$

$$= \left(\sum_s a_i(s) \cdot x(s) \right)^3$$

$$= \langle a_i, x \rangle^3$$

$$= \sum_i \alpha_i^3 = \sum_i \alpha_i \cdot \alpha_i^2$$

$$\leq \left[\max_i \alpha_i \right] \sum_i \alpha_i^2$$

$$= \max_i \alpha_i$$

① $p(a_i) = 1 \quad \forall i$

② $\max_{x : \|x\|_2^2 \leq 1} p(x) \leq 1$ $= 1$

③ if $x \notin \{ a_1 — a_r \}$

then $\max_i \alpha_i < 1 \rightarrow p(x) < 1.$

Idea for using SoS for tensor decomp

find the maximizers of p!

Focus on certifying purported sol^{n}s

x :
$$\begin{cases} p(x) \geq 1 . \\ \|x\|_2^2 \leq 1 . \end{cases}$$
"test". ✓

Soundness: If x passes the test then it must be a true solⁿ.

real world proof ≡ Correspondence lemma above.

how to phrase soundness as poly inequality?

$\longrightarrow \prod_i \underbrace{\|x - a_i\|_2^2} = 0$ deg r equation!

$\longrightarrow \sum_i \langle a_i, x \rangle = 1 \leftarrow$

$\longrightarrow \sum_i \langle a_i, x \rangle^2 = 1$ $\quad \forall x : \|x\|_2^2 = 1$

$$\sum_i \langle a_i, x \rangle^t = 1$$
$$\|$$
$$\sum_{i=1}^{r} \alpha_i^t = 1 \Rightarrow \exists i : \alpha_i^t \geq 1/r$$
$$\Downarrow$$
$$|\alpha_i| \geq e^{-\frac{\ln r}{t}}$$
$$|\alpha_i| \geq e^{-\delta}$$

If $t \geq \frac{\ln r}{\ldots}$

$$\delta$$

$$\sim 1-\delta$$

Poly inequality version of soundness: $\sum_i \langle a_i, x\rangle^t \geq 1$

(for some $t \sim \frac{\ln r}{\delta}$)

target accuracy

<u>Lemma</u>: $\mathcal{A} = \{ \|x\|_2^2 \leq 1,\ \sum_{i=1}^{r} \langle a_i, x\rangle^3 \geq 1-\delta \}$

$$\mathcal{A} \vdash^{6}_{x} \{ \underline{\textstyle\sum_i \langle a_i, x\rangle}^4 \geq 1 - 2\delta \}$$

<u>Proof</u>:

$$\mathcal{A} \vdash^{6}_{x} \Big\{ \Big(\sum_{i=1}^{r} \langle a_i, x\rangle^3\Big)^2 \leq \Big(\sum_{i=1}^{r} \langle a_i, x\rangle^4\Big) \cdot \Big(\sum_i \langle a_i, x\rangle^2\Big) \Big\}$$

$$\sum_i \langle a_i, x\rangle^2 \leq (1+\varepsilon) \cdot \|x\|_2^2 \qquad = \Big(\sum_{i=1}^{r} \langle a_i, x\rangle^4\Big) \cdot \|x\|_2^2$$

$$\leq \sum_{i=1}^{r} \langle a_i, x\rangle^4 \Big\}$$

$$\mathcal{A} \vdash^{6}_{x} \Big\{ (1-\delta)^2 \leq \sum_{i=1}^{r} \langle a_i, x\rangle^4 \Big\}$$

$\square$

"Soundness"

Lemma: Let $k \in \mathbb{N}$ be even, $k > 4$

$$\Delta \vdash^{2k}_{x} \left\{ \sum_{i=1}^{r} \langle a_i, x \rangle^k \geq 1 - O(k\delta) \right\}$$

Lemma: $v_1 \text{---} v_r$ indeterminates.

Then, $\forall\, s \in \mathbb{N}$

$$\vdash^{4s}_{v_1 \ldots v_r} \left\{ \left(\sum_i v_i^4\right)^s \leq \left(\sum_{i=1}^{r} v_i^2\right)^s \cdot \left(\sum_{i=1}^{r} v_i^{2s}\right) \right\}$$

$$\|v\|_4^{4s} \leq \|v\|_2^{2s} \cdot \|v\|_{2s}^{2s}$$

Pf using lemma: $v_i = \langle a_i, x \rangle \quad \forall i$

$$\boxed{s = \frac{k}{2}}$$

$$\Delta \vdash^{2k}_{x} \left\{ \left(\sum_i \langle a_i, x \rangle^4\right)^{\frac{k}{2}} \leq \underbrace{\left(\sum_i^r \langle a_i, x \rangle^2\right)^{k/2}} \cdot \left(\sum_{i=1}^{r} \langle a_i, x \rangle^k\right) \right.$$

$$= \|x\|_2^k \cdot \sum_{i=1}^{r} \langle a_i, x \rangle^k \Big\}$$

$$\leq \underbrace{\sum_{i=1}^{r} \langle a_i, x \rangle^k} \Big]$$

$$\Delta \vdash^{2k} \Big\{ \ldots^{k} \ldots (1 - 2\delta)^{k/2} \Big\}$$

$\alpha \mid \frac{\cdot}{x} \left\{ \sum_{i=1} \langle a_i, x \rangle \geq (1-2\delta) \right\}$

$\geq 1 - k\delta \}$

Pf of Lemma on Vector norms.

$$\left(\sum_i v_i^4\right)^s \leq \left(\sum_i v_i^2\right)^{2s} \cdot \left(\sum_i v_i^{2s}\right)$$

$$\sum_{\substack{|\alpha|=s \\ \alpha \in [n]}} \binom{s}{\alpha} v^{4\alpha} \leq \left(\sum_{|\alpha|=s} \binom{s}{\alpha} \cdot v^{2\alpha}\right) \cdot \left(\sum_i v_i^{2s}\right)$$

$$\binom{s}{\alpha} = \frac{s!}{\alpha_1! \cdots \alpha_r!}$$

enough to prove $\quad v^{4\alpha} \leq v^{2\alpha} \cdot \left(\sum_i v_i^{2s}\right)$

$\Downarrow$

enough to prove $\quad \boxed{v^{2\alpha} \leq \left(\sum_i v_i^{2s}\right)}$

$|\alpha| = 4$

$$v_1^6 \cdot v_2^2 \leq v_1^8 + v_2^8$$

$$(a_1 + \cdots + a_8)^8$$

$$\left(\frac{a_1 + \cdots + a_8}{8}\right)^8 \geq a_1 \cdots a_8$$

$$v_1^6 \cdot v_2^2 \leq \left(\frac{6}{8} \cdot v_1 + \frac{2}{8} v_2\right)^8$$

$$\overset{\text{Jensen}}{\leq} \frac{6}{8} v_1^8 + \frac{2}{8} v_2^8$$

Noise Resilience / Generalization

Lemma: Suppose $\|\sum_i a_i a_i^T\|_2 \leq 1+\varepsilon$ a_i are unit vectors.

Suppose $T = \sum_i a_i^{\otimes 3} + E$.

$$\|E\|_{\underline{\{1\},\{2,3\}}} \leq \delta.$$

Then, $\mathcal{A} = \{\|x\|_2^2 \leq 1,\ \langle T, x^{\otimes 3}\rangle \geq 1-\varepsilon-\delta\}$

$$\mathcal{A} \vdash_{x}^{2k} \left\{ \sum_i \langle a_i, x\rangle^k \geq 1 - O(k\delta) - O(k\varepsilon) \right\}$$

Proof: Similar style proof

$\langle T, x^{\otimes 3}\rangle \geq 1-\delta-\varepsilon$

$= \sum_i \langle a_i, x\rangle^3 + \langle E, x^{\otimes 3}\rangle$

$\geq \sum_i \langle a_i, x\rangle^3 - \delta$

For $x = a_1$

$\|a_1\|^3 + \underbrace{\sum_{i\neq 1}\langle a_i, a_1\rangle^3} - \delta$

$\sum_i \langle a_i, a_1\rangle^3$

$\leq \sum_i \langle a_i, a_1\rangle^2 \leftarrow$

$= a_1^T\left(\sum_i a_i a_i^T\right)a_1$

$\leq (1+\varepsilon)\cdot\|a_1\|^2 = 1+\varepsilon$

$\Rightarrow \sum_{i\neq 1}\langle a_i, a_1\rangle^2 \leq \varepsilon.$

Proof of completeness.

<u>Algorithm / Rounding</u>

Let μ be a p.d. of deg $2k$. satisfy $\mathcal{A}$.

<u>Then.</u> $\underline{\tilde{\mathbb{E}}_{\mu} \sum_{i=1}^{r}\langle a_i, x\rangle^k} \geq 1 - O(k\delta)$

$O(k\delta) < \frac{1}{2}$

therefore $\exists i \quad \tilde{\mathbb{E}}\langle a_i, x\rangle^k \geq 1 - O(k\delta)$

μ

$\geq \frac{1}{2r}$

Note: if all points in supp(μ) are such that $\langle a_i, x\rangle \lesssim 1-\beta \sim \delta$, $k \sim \frac{1}{\delta}\log r$

$\mathbb{E}_\mu \langle a_i, x\rangle^k \leq (1-\beta)^k$

Goal: round μ to a single solution.

Lemma: Let μ be a p.d. on X satisfying

$\tilde{\mathbb{E}}_\mu \langle a, x\rangle^k \geq e^{-\delta k} \sim 1/2r$ for $k \sim \frac{\log r}{\delta}$ & $\{\|x\|_2^2 \leq 1\}$

Then, there's a $n^{O(k)}$ time algo that succeeds w.p. $\boxed{2^{-k/\mathrm{poly}(\varepsilon)}}$ that outputs a' s.t. $\langle a, a'\rangle \geq e^{-\delta} \sim 1-\delta$.

a' of unit norm

Note: If $\boxed{\tilde{\mathbb{E}}\langle a,x\rangle^2 \geq e^{-\delta} \sim 1-\delta}$

then can round easily.

$$\tilde{\mathbb{E}}\|x\|_2^2 \le 1$$

$$1 \geq tr(\tilde{\mathbb{E}}\, xx^T) = \tilde{\mathbb{E}}\sum_i x_i^2 \le 1$$

$$a^T(\tilde{\mathbb{E}}\, xx^T)\,a \geq 1-\delta$$

λ_1 —— $\lambda_d \geq 0$

$\sum_i \lambda_i \le 1$

$\lambda_1 \geq 1-\delta$

So top eigenvector of $\tilde{\mathbb{E}}\, xx^T$ works.

Rounding idea:

Reweightings.

$$\mu,\ p : \mu'(x) = \mu(x)p(x)$$

$$\tilde{\mathbb{E}}_{\mu'} f = \frac{\tilde{\mathbb{E}}\, f\cdot p}{\tilde{\mathbb{E}}\, p} \checkmark$$

Claim: $\mu : \{\|x\|_2^2 \le 1\}$, $\tilde{\mathbb{E}}_\mu \langle a,x\rangle^k \geq e^{-\delta k}$

then there's a $k-2$ deg SoS reweighting of μ

s.t. $\tilde{\mathbb{E}}_{\mu}\langle a,x\rangle^2 \geq e^{-\delta}$ ✓

Pf: $1, \langle a,x\rangle^2, \langle a,x\rangle^4, \ldots, \langle a,x\rangle^{k-2}$

$$\tilde{\mathbb{E}}_{\mu}\langle a,x\rangle^2, \quad \frac{\tilde{\mathbb{E}}_{\mu'}\langle a,x\rangle^4}{\tilde{\mathbb{E}}_{\mu}\langle a,x\rangle^2}$$

$$\vdots$$

$$\tilde{\mathbb{E}}\langle a,x\rangle^k / \tilde{\mathbb{E}}\langle a,x\rangle^{k-2}$$

Suppose $\tilde{\mathbb{E}}_{\mu'}\langle a,x\rangle^2 < e^{-\delta}$

$$\tilde{\mathbb{E}}\langle a,x\rangle^k \geq e^{-k\cdot\delta}$$

$$\tilde{\mathbb{E}}\langle a,x\rangle^2 < e^{-\delta}$$

$$\tilde{\mathbb{E}}\langle a,x\rangle^k \qquad e^{-(k-1)\delta}$$

$$\frac{\cdots}{\tilde{\mathbb{E}}\langle a, x\rangle^2} \geq e^{\cdots}$$

$$\underline{\exists i:} \quad \frac{\tilde{\mathbb{E}}\langle a, x\rangle^{i+2}}{\tilde{\mathbb{E}}\langle a, x\rangle^{i}} \geq e^{-\delta}$$

use $\langle a, x\rangle^i$ as reweighting.

<u>Actual idea:</u> pick g_1 —— $g_{\frac{k}{2}-1}$

reweight with $\langle g_1, x\rangle^2$

$\langle g_1, x\rangle^2 \cdot \langle g_2, x\rangle^2$

$|\langle g_1, a\rangle| \sim \sqrt{\log k}$

$\vdots$

$\langle g_1, x\rangle^2 \cdot$ —— $\langle g_{\frac{k}{2}-1}, x\rangle^2$

Quasipoly algo to recover at least one ... overcomplete 3rd

Component of the ... order tensor

$$\left\|\sum_i a_i a_i^T\right\|_2 \le 1+\varepsilon$$

$a_1 — a_r$, of unit norm

Overcompleteness: $d^{1.5}/\log d$ ✓

polynomial time!

Idea: use Jennrich's algo as rounding algo!

$a_1 —— a_r \qquad r \gg d$

$a_1^{\otimes 2}, \dots a_r^{\otimes 2}$: d^2 dim

Jennrich's algo: $\sum_i a_i^{\otimes 6}$ ← you don't

$$= \sum_i \left(a_i^{\otimes 2}\right)^{\otimes 3}$$

SoS: $\sum \langle a_i, x \rangle^{6} \geq 1-\delta$

$\boxed{\sum_i \langle a_i, x \rangle^{6} \geq 1 - O(\delta).}$

$\tilde{\mathbb{E}}_{\mu}\, x^{\otimes 6} \sim \sum_i a_i^{\otimes 6}$

$\mathbb{E}_{\mu}$

1. Polynomial Time Tensor Decomposition using SoS

   Ma-Shi-Steurer '15

2. Dictionary Learning & Tensor Dec using SoS

   Barak-Kelner-Steurer '14.