9/27/23

Lecture 7: SoS for tensor decomposition (I)

Tensor norms:

Frobenius:

$$\|T\|_F = \Big(\sum_{i,j,k} T_{ijk}^2\Big)^{1/2}$$

Injective (tensor analogue of operator norm):

$$\|T\|_{inj} = \max_{\|x\|_2=1} |\langle T, x^{\otimes 3}\rangle|$$

(NP-hard even to approximate to within $n^{o(1)}$ factor)

Note:

$$\langle T, x^{\otimes 3}\rangle^2 = \Big(\sum_{i,j,k} T_{ijk} x_i x_j x_k\Big)^2$$

$$\leq \Big(\sum_{i,j,k} T_{ijk}^2\Big)\Big(\sum_{i,j,k} x_i^2 x_j^2 x_k^2\Big)$$

$$= \|T\|_F^2 \cdot \Big(\sum_i x_i^2\Big)^3 = \|T\|_F,$$

So $\boxed{\|T\|_{inj} \leq \|T\|_F}$.

Note: This was a degree-6 proof!

SoS norm: For $t \geq 6$ even,

$$\|T\|_{sos_t} \triangleq \max_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}} \langle T, x^{\otimes 3} \rangle$$

where $\tilde{\mathbb{E}}$ ranges over deg-t pseudo-expectations over SoS variable x satisfying $\|x\|_2^2 = 1$.

Note as $t \to \infty$, SoS norm approaches injective norm:

$$\|T\|_{sos_6} \geq \|T\|_{sos_8} \geq \cdots \geq \|T\|_{sos_t} \geq \|T\|_{inj}$$

Note: $\|T\|_F \geq \|T\|_{sos_6}$ b/c proof of $\|T\|_F \geq \|T\|_{inj}$ was deg 6 SoS proof.

Norm of reshaping: Let M be reshaping of T into $d \times d^2$ matrix, i.e.

$$M_{i,jk} = T_{ijk}.$$

$$\|T\|_{\{1\},\{2,3\}} = \|M\|_{op}$$

$$\left(\sum_{ijk} T_{ijk}\, x_i x_j x_k\right)^2 = \left(\sum_{j,k} x_j x_k \underbrace{\sum_i T_{ijk}\, x_i}\right)^2$$

$$\leq \underbrace{\left(\sum_{j,k} x_j^2 x_k^2\right)}_{=1} \cdot \sum_{j,k}\left(\sum_i T_{ijk}\, x_i\right)^2$$

$$= \|M^T x\|_2^2 \leq \|M\|_{op}^2$$

Note: This is a deg-6 SoS proof, so

$$\|T\|_{\{1\},\{2,3\}} \geq \|T\|_{sos_6}$$

Today: (Very) Noisy Orthogonal tensor decomposition

$$T = \sum_{i=1}^{d} u_i^{\otimes 3} + \underset{\substack{\uparrow \\ \text{noise}}}{E} \qquad u_1, \dots, u_d \in \mathbb{R}^d \text{ orthonormal}$$

Previously:

if every entry of E had magnitude $\frac{1}{d^c}$ for large enough constant $c > 0$, then Jennrich's succeeds.

What if E somewhat large?

<u>Example</u>: $E_{jk} \sim N\left(0, \frac{1}{d^{1.9}}\right)$

Jennrich's: sample $g \sim N(0, Id)$ and consider

$$T(g,:,:) = \underbrace{\sum_{i=1}^{d} \langle g, u_i \rangle u_i u_i^T}_{\text{signal}} + \underbrace{\sum_{i=1}^{d} g_i E_{i::}}_{\text{Noise}}$$

Every entry of $\sum_{i=1} g_i E_{i::} \approx N\left(0, \frac{1}{d^{0.9}}\right)$.

$$\| N(0, Id)^{d \times d} \|_{op} \approx \sqrt{d}, \quad \text{so}$$

$$\| \sum_i g_i E_{i::} \|_{op} \approx \sqrt{d} \cdot \sqrt{\frac{1}{d^{0.9}}} = d^{0.05} >> 1$$

Whereas

$$\| \sum_{i=1}^{d} \underbrace{\langle g, u_i \rangle}_{\approx 1} u_i u_i^T \|_{op} \approx 1$$

i.e. Noise >> Signal, so Jennrich's fails!

<u>Fact</u> (Hopkins - Schramm - Steurer '15): with high prob,

$$\| N(0, \sigma^2)^{d \times d \times d} \|_{sos_6} \leq \sigma d^{3/4} \cdot \text{polylg}(d),$$

so $\|E\|_{sos_6}$ in example above is $\lesssim \frac{1}{d^{0.2}}$.

In fact, we will show that as long as $\|E\|_{sos_6} << 1$, there is an algorithm to recover $u_1, \dots, u_d$!

Let $T = \sum_{i=1}^{d} u_i^{\otimes 3} + E$ for $\|E\|_{sos_6} = o(1)$.

Define $p_\ell(x) \triangleq \sum_{i=1}^{d} \langle u_i, x \rangle^\ell$.

$\boxed{\text{Goal: maximize } p_3}$

Alg (attempt #1):

SoS variables: x (d-dimensional)

Constraints : $\|x\|_2^2 \leq 1$

Objective :

$$\max_{\hat{\mathbb{E}}} \hat{\mathbb{E}} \langle T, x^{\otimes 3} \rangle$$

Lemma 1: Optimal $\hat{\mathbb{E}}$ for the above satisfies

$$\hat{\mathbb{E}}\left[p_3(x) \right] \geq 1 - o(1) \qquad (*)$$

Pf: Note that for pseudo-exp given by uniform dist over $\{u_1, \dots, u_k\}$, call it $\mathbb{E}[\cdot]$,

$$\tilde{\mathbb{E}}[p_3(x)] = \tilde{\mathbb{E}}\left[\sum_i \langle u_i, x\rangle^3\right] = \frac{1}{d}\sum_j \left(\sum_i \underbrace{\langle u_i, u_j\rangle^3}_{= \mathbb{1}[i=j]}\right) = \frac{1}{d}\cdot\sum_i 1 = 1,$$

so $$\tilde{\mathbb{E}}[\langle T, x^{\otimes 3}\rangle] = \tilde{\mathbb{E}}[p_3(x)] + \tilde{\mathbb{E}}[\langle E, x^{\otimes 3}\rangle]$$
$$\geq 1 - \|E\|_{sos_6} = 1 - o(1)$$

so LHS of $(*)$ $= \hat{\tilde{\mathbb{E}}}[\langle T, x^{\otimes 3}\rangle]$ $\hat{\tilde{\mathbb{E}}}[\langle E, x^{\otimes 3}\rangle]$

$$\underset{\text{(by maximality of } \hat{\tilde{\mathbb{E}}})}{\geq} 1 - o(1) - \|E\|_{sos_6}$$
$$= 1 - o(1). \qquad \square$$

Lemma 2: optimal $\hat{\tilde{\mathbb{E}}}$ also satisfies:

$$\hat{\tilde{\mathbb{E}}}[p_4(x)] \geq 1 - o(1) \qquad (**)$$

Pf:

$$1 - o(1) \leq \hat{\tilde{\mathbb{E}}}[p_3(x)]^2 = \hat{\tilde{\mathbb{E}}}\left[\sum_i \langle u_i, x\rangle^3\right]^2$$
$$\underset{\text{pseudo-exp. Cauchy Schwarz}}{\leq} \hat{\tilde{\mathbb{E}}}\left[\left(\sum_i \langle u_i, x\rangle^3\right)^2\right]$$

Note that in deg-6 SoS:

$$\left(\sum_i \langle u_i, x\rangle^3\right)^2 \le \underbrace{\left(\sum_i \langle u_i, x\rangle^2\right)}_{\|x\|^2 = 1} \cdot \left(\sum_i \langle u_i, x\rangle^4\right)$$

$$= P_4(x),$$

so $\tilde{\mathbb{E}}[P_4(x)] \ge 1 - o(1)$ $\square$

How to round $\tilde{\mathbb{E}}$ to a solution?

$\boxed{\text{What if we ran Jennrich's, but on } \tilde{\mathbb{E}}[x^{\otimes 3}] \text{ instead of on } T\text{?}}$

Issue: Suppose $\tilde{\mathbb{E}}$ is an actual distribution that places $\frac{1}{\sqrt{d}}$ mass on an arbitrary vector $w \perp u_i$ and $\frac{1 - 1/\sqrt{d}}{d}$ for $u_1, \dots, u_d$.

$$\hat{T} \triangleq \hat{\mathbb{E}}[x^{\otimes 3}] = \frac{1 - 1/\sqrt{d}}{d} \sum_{i=1}^{d} u_i^{\otimes 3} + \frac{1}{\sqrt{d}} w^{\otimes 3}$$

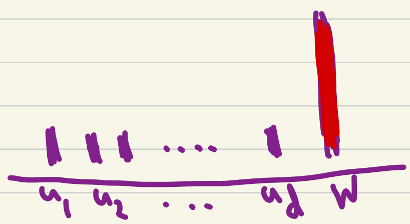
$$\hat{T}(g, :, :) = \frac{1 - 1/\sqrt{d}}{d} \sum_{i=1}^{d} \langle u_i, g \rangle u_i u_i^T + \frac{1}{\sqrt{d}} \langle w, g \rangle w w^T$$

eigenvectors are $u_1, \ldots, u_d, w$

w/ eigenvalues $\approx \Theta(\frac{1}{d}), \ldots, \Theta(\frac{1}{d}), \Theta(\frac{1}{\sqrt{d}})$

So top eigenvector completely useless! $\ddot{\frown}$

Issue: This $\hat{\mathbb{E}}$ is very "low-entropy"



Idea: add constraints that force entropy to be high

Algorithm (attempt 2):

Same SoS program as before but,

$$\max_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}}\left[\langle T, x^{\otimes 3}\rangle\right]$$

over deg-6 pseudo-exp. $\tilde{\mathbb{E}}$'s which satisfy the SoS program constraints and additionally

1) $\|\tilde{\mathbb{E}}[xx^T]\|_{op} \le 1/d$ (1)

2). $\|\tilde{\mathbb{E}}[(x\otimes x)(x\otimes x)^T]\|_{op} \le 1/d$ (2)

3). $\|\tilde{\mathbb{E}}[(x\otimes x\otimes x)(x\otimes x\otimes x)^T]\|_{op} \le 1/d$ (3)

(note: this holds for $\tilde{\mathbb{E}}$ the uniform dist over $u_1, \dots, u_d$)

Lemma 3: For optimal $\tilde{\mathbb{E}}$ in the above, for $1 - o(1)$ fraction of $i \in [d]$,

$$\tilde{\mathbb{E}}\langle u_i, x\rangle^4 \ge \frac{1}{d}(1 - o(1))$$

Pf: By Lemma 2,

$$\sum_{i=1}^{d} \tilde{\mathbb{E}}\langle u_i, x\rangle^4 = 1 - o(1),$$

Suppose for $\delta = \Omega(1)$ fraction of i's,
we have $\hat{\mathbb{E}} \langle u_i, x\rangle^4 \leq \frac{1-\delta}{d}$.

Then for some other i, by averaging,

$$\hat{\mathbb{E}} \langle u_i, x\rangle^4 \underline{> \frac{1}{d}}.$$

$$\begin{aligned} &\| \\ &(u_i \otimes u_i)^T \hat{\mathbb{E}}\left[(x\otimes x)(x \otimes x)^T\right](u_i \otimes u_i) \\ &\leq \left\| \hat{\mathbb{E}}\left[(x\otimes x)(x\otimes x)^T\right]\right\|_{op} \leq \underline{\frac{1}{d}}, \end{aligned}$$

by high entropy constraint

Contradiction! $\square$

We now prove that by running Jennrich's on

$$\hat{\mathbb{E}}\left[x^{\otimes 4}\right]$$

many times, we recover $1-o(1)$ of the components

(see [Ma-Shi-Steurer '16] for how to recurse to find the remaining $o(1)$ fraction).

Theorem (main rounding analysis):

If $a \in \mathbb{R}^d$ satisfies

$$\tilde{\mathbb{E}}^{n} \langle a, x\rangle^4 \geq \frac{1}{d}(1 - o(1)),$$

then w.p. $\geq \frac{1}{\mathrm{poly}(d)}$ over $g \sim N(0, Id_{d^2})$ the top eigenvector v of $\tilde{\mathbb{E}}^{n}[\langle x \otimes x, g\rangle xx^T]$ satisfies $\langle v, a\rangle^2 \geq 0.99$.

Pf: Write $g = \gamma\, a \otimes a + \gamma^\perp$ for $\gamma \sim N(0,1)$ and $\gamma^\perp \sim N(0, Id_{d^2} - (a\otimes a)(a \otimes a)^T)$

Denote $M_g \triangleq \tilde{\mathbb{E}}^{n}[\langle x \otimes x, g\rangle xx^T]$ so that

$$M_g = \gamma \underbrace{M_{a\otimes a}}_{\text{① will show } = \frac{1}{d}aa^T + o(\frac{1}{d})} + \underbrace{M_{\gamma^\perp}}_{\text{② will show } \lesssim \frac{\sqrt{\lg d}}{d}}$$

So if $\gamma \geq 100\sqrt{\lg d}$ (happens w.p. $\geq \frac{1}{\mathrm{poly}(d)}$), then $M_g \approx \frac{\gamma}{d}\left(aa^T + 0.01\right)$,

so top eigenvec has large constant correlation with a as desired.

①: To analyze $M_{a\otimes a} = \hat{\mathbb{E}}\left[\langle x,a\rangle^2 xx^T\right]$,

consider $b \in \mathbb{S}^{d-1}$ orthogonal to a. Then:

$$b^T M_{a\otimes a} b$$

$$= b^T\left\{\hat{\mathbb{E}}\left[\langle x,a\rangle^2 xx^T\right]\right\} b = \hat{\mathbb{E}}\left[\langle x,a\rangle^2 \langle x,b\rangle^2\right]$$

note, $\langle x,b\rangle^2 + \langle x,a\rangle^2 \le \|x\|^2 = 1$, so

$$\le \underbrace{\hat{\mathbb{E}}\left[\langle x,a\rangle^2\right]}_{\substack{= a^T \hat{\mathbb{E}}[x\otimes x] a \\ \le \|\hat{\mathbb{E}}[x\otimes x]\|_{op} \\ \le 1/d}} - \underbrace{\hat{\mathbb{E}}\left[\langle x,a\rangle^4\right]}_{\ge \frac{1}{d}(1-o(1))}$$

$$\le o(1/d).$$

Also,

$$a^T(M_{a\otimes a})a = \hat{\mathbb{E}}\left[\langle x, a\rangle^4\right] \geq \frac{1-o(1)}{d}.$$

And because $M_{a\otimes a}$ is psd,

$$|a^T M_{a\otimes a} b|^2 \leq \underbrace{\left(a^T(M_{a\otimes a})a\right)}_{\leq \frac{1}{d}}\underbrace{\left(b^T(M_{a\otimes a})b\right)}_{\leq o\left(\frac{1}{d}\right)}$$

$$\leq o(1/d),$$

so $\left\|\frac{1}{d}aa^T - M_{a\otimes a}\right\|_{op} \leq o\left(\frac{1}{d}\right)$. $\square$

---

(2): To bound $M_{\gamma^\perp} = \hat{\mathbb{E}}\left[\langle \gamma^\perp, x\otimes x\rangle xx^T\right]$,

write $\gamma^\perp = \frac{1}{2}(g_1 + g_2)$, where

$$g_1 = \gamma^\perp + \gamma'$$
$$g_2 = \gamma^\perp - \gamma', \quad \text{for } \gamma' \sim N(0, (a\otimes a)(a\otimes a)^T)$$

Note, g_1, g_2 are marginally distr'd as $N(0, Id)$.

So suffices to bound

$$M_h = \tilde{\mathbb{E}}\left(\langle h, x \otimes x\rangle x x^T\right)$$

for $h \sim N(0, Id)$.

$$M_h = \sum_{i,j=1}^{d} \underbrace{h_{ij}}_{\substack{\uparrow \\ \text{independent Gaussians}}} \cdot A_{ij} \quad \text{for} \quad A_{ij} \triangleq \tilde{\mathbb{E}}\left[x_i x_j \cdot x x^T\right]$$

By concentration of matrix Gaussian series (see. e.g. Theorem 4.1.1 in Tropp "Intro to matrix concentration inequalities"),

$$\|M_h\|_{op} \le \left\|\sum_{ij} A_{ij}^2\right\|_{op}^{1/2} \cdot O(\sqrt{\lg d}) \qquad (\star\star)$$

with high probability.

If $\quad B = \tilde{\mathbb{E}}\left[x \left(x^{\otimes 3}\right)^T\right] \in \mathbb{R}^{d \times d^3}$,

$$S = \tilde{\mathbb{E}}\left[x^{\otimes 4}\right],$$

then

$$\left(\sum_{i,j} A_{ij}^2\right)_{ab} = \sum_{c,i,j} S_{ijac} S_{ijcb}$$

$$\overset{\text{(symmetry)}}{=} \sum_{c,i,j} S_{aijc} S_{ijcb}$$

$$= B_{a:} B^T_{:b} \, ,$$

so $\sum_{i,j} A_{ij}^2 = B B^T$

and $\left\| \sum_{i,j} A_{ij}^2 \right\|_{op}^{1/2} = \|B\|_{op}$

not for any $z \in \mathbb{R}^d$, $z' \in \mathbb{R}^{d^3}$ of unit norm,

$$\left(z^T B z'\right)^2 = \tilde{\mathbb{E}} \left[\langle x, z\rangle \langle x^{\otimes 3}, z'\rangle\right]^2$$

pseudo-exp. Cauchy Schwarz $\rightarrow$ $\leq \tilde{\mathbb{E}} \left[\langle x, z\rangle^2\right] \cdot \tilde{\mathbb{E}}\left[\langle x^{\otimes 3}, z'\rangle^2\right]$

$$= \left(z^T \tilde{\mathbb{E}}\left[x x^T\right] z\right) \cdot \left(z' \tilde{\mathbb{E}}\left[x^{\otimes 3} (x^{\otimes 3})^T\right] z'\right)$$

entropy bounds $\rightarrow$ $\leq \frac{1}{d^2}$

So $\|B\|_{op} \leq 1/d$, and $(**)$ yields

$$\|M_n\| \lesssim 1/d \cdot \sqrt{\lg d}\,. \qquad \square$$