10/1/23

Lecture 8: SoS for tensor decomposition (II)

Recall setup:

$$T = \sum_{i=1}^{d} u_i^{\otimes 3} + E$$

for $u_1, \dots, u_k$ orthonormal, and $\|E\|_{sos_6} \ll 1$.

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

over deg-6 pseudo-exp's over x s.t.

1). $\hat{\mathbb{E}}$ satisfies constraint $\{\|x\|^2 = 1\}$

entropy maximization

2). $\|\hat{\mathbb{E}}[x x^T]\|_{op} \le \frac{1}{d}$

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

4) $\|\hat{\mathbb{E}}[(x \otimes x \otimes x)(x \otimes x \otimes x)^T]\|_{op} \le \frac{1}{d}$ $\quad \hat{\mathbb{E}}[x^{\otimes 4}]$

Then run Jennrich's many times on $\hat{T} \triangleq \cancel{\hat{\mathbb{E}}[x^{\otimes 3}]}$,

i.e. take top eigenvector of $\hat{T}(\underbrace{g}_{\mathbb{R}^{d^2}}, :, :)$ for many random g's.

End of last lecture:

Lemma: For optimal $\hat{\mathbb{E}}$, for $1 - o(1)$ fraction of i's,

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

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 $g \sim N(0, \mathrm{Id}_{d^2})$ 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 $\gamma^\perp \sim N(0, \mathrm{Id}_{d^2} - (a \otimes a)(a \otimes a)^T)$

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

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

We will come back to this when we discuss the overcomplete setting..

Also,

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

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

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

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

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

---

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

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

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

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

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

independent Gaussians

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

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

with high probability.

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

$S = \hat{\mathbb{E}}\left[x^{\otimes 4}\right]$, 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 $\displaystyle\sum_{i,j} A_{ij}^2 = B B^T$

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

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 $\rightarrow$ $\displaystyle\leq \tilde{\mathbb{E}}\left[\langle x, z\rangle^2\right] \cdot \tilde{\mathbb{E}}\left[\langle x^{\otimes 3}, z'\rangle^2\right]$

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

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

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

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

This idea can also be used for <u>overcomplete</u> tensor decomposition!

$(k \ll d^{3/2})$ $\quad T = \sum_{i=1}^{k} u_i^{\otimes 3}$ for $u_1, \dots, u_k \sim S^{d-1}$ uniformly

Main conditions we needed to recover component $a \in \mathbb{R}^d$ were that

1. $\hat{\mathbb{E}}\left[\langle a, x\rangle^4\right] \ge \underbrace{\boxed{\tfrac{1}{d}}}_{(1)}(1 - o(1))$
2. $\hat{\mathbb{E}}$ satisfies entropy maximization constraints, e.g. $\left\|\hat{\mathbb{E}}\left[xx^T\right]\right\|_{op} \le \underbrace{\boxed{\tfrac{1}{d}}}_{(2)}$

It was important that (1), (2) were both $\frac{1}{d}$

Recall we needed to bound $b^T M_{a\otimes a} b$ for b orthogonal to a.

$$
\begin{aligned}
b^T M_{a\otimes a} b &= \hat{\mathbb{E}}\left[\langle x,a\rangle^2 \langle x,b\rangle^2\right] \\
&\le \hat{\mathbb{E}}\left[\langle x,a\rangle^2 \left(1-\langle x,a\rangle^2\right)\right] \\
&= \hat{\mathbb{E}}\left[\langle x,a\rangle^2\right] - \hat{\mathbb{E}}\left[\langle x,a\rangle^4\right] \\
&\le \left\|\hat{\mathbb{E}}\left[xx^T\right]\right\|_{op} - \hat{\mathbb{E}}\left[\langle x,a\rangle^4\right]
\end{aligned}
$$

<u>Issue</u>: In overcomplete setting, even if $\hat{\mathbb{E}}$ is uniform over $\{u_1,\dots,u_k\}$, and $a=u_i$,

$$
\begin{aligned}
\hat{\mathbb{E}}\left[\langle a,x\rangle^4\right] &= \frac{1}{k}\sum_{j=1}^{k} \underbrace{\langle u_i,u_j\rangle^4} \\
&= \frac{1}{k}\Big(1+\sum_{j:j\neq i} \underbrace{\langle u_i,u_j\rangle^4}_{\approx\left(\frac{1}{\sqrt{d}}\right)^4=\frac{1}{d^2}}\Big) \\
&\approx \frac{1}{k}\left(1+O\left(\frac{k}{d^2}\right)\right) \\
&= \boxed{\frac{1}{k}}\left(1+o(1)\right)
\end{aligned}
$$

yet

$$\hat{\mathbb{E}}[xx^T] = \frac{1}{k}\sum_{i=1}^{k} u_i u_i^T \approx \frac{1}{d} I_d,$$

so $$\|\hat{\mathbb{E}}[xx^T]\|_{op} = \boxed{\frac{1}{d}} \gg \frac{1}{k}$$

Intuitively, issue is that there are too few dimensions, so u_i's have tiny but non-negligible correlations, and so $\|\hat{\mathbb{E}}[xx^T]\|_{op}$ can't be small enough for previous argument to go through.

<u>Idea:</u> Lift to higher dimensions!

Suppose instead of $T = \sum_{i=1}^{k} u_i^{\otimes 3}$,

We'll use SoS to "hallucinate" this tensor

we had the tensor $$\boxed{T = \sum_{i=1}^{k} u_i^{\otimes 6} = \sum_{i=1}^{k} \Big(\underbrace{u_i^{\otimes 2}}_{\triangleq w_i}\Big)^{\otimes 3}}$$

$\{w_i\}_{i=1,\dots,k}$ are $k \ll d^{3/2}$ "random" vectors

in d^2 dimensions of norm

Claim*: If $k \ll d^2$, then

$$\left\| \frac{1}{k} \sum_i w_i w_i^T \right\|_{op} = \frac{1}{k}\left(1 + o(1)\right).$$

Pf sketch: (next page)

Fine print: (can skip on first reading)

* Not quite right as $z^T\left(\sum_j w_j w_j^T z\right)$ for

$$z \stackrel{\Delta}{=} \frac{1}{\sqrt{d}} \sum_i e_i \otimes e_i \quad \text{is} \quad \frac{1}{d} \sum_{i,i',j} \underbrace{\langle w_j, e_i\rangle^2}_{\approx \frac{1}{d}} \underbrace{\langle w_j, e_{i'}\rangle^2}_{\approx \frac{1}{d}}$$

$$\approx \frac{1}{d} \cdot d^3 \cdot \frac{1}{d^2} = 1,$$

so there is a "bad" direction along which $\sum w_i w_i^T$ and Id_{d^2} differ significantly. But can show that apart from this direction, i.e. if we instead take $W_i := \Pi(u_i \otimes u_i)$ for an appropriate projection away from z, the claim holds.

Use the following general fact:

__Lemma__ (Theorem 5.62 in Vershynin):

Let $A \in \mathbb{R}^{D \times k}$ have random norm-$\sqrt{D}$ columns that are independent with mean zero and identity covariance.

Then

$$\mathbb{E} \left\| \frac{1}{D} A^T A - Id_k \right\|_{op} \lesssim \sqrt{\frac{m \lg k}{D}}$$

where $m \triangleq \frac{1}{D} \mathbb{E} \max_i \sum_{j \neq i} \langle A_i, A_j \rangle^2$.

i.e. if columns have typical inner product $\sim \pm \rho$, then $m \approx \frac{\rho k}{D}$

Take $D = d^2$, $A_i = \sqrt{D} \cdot w_i$ so* $\frac{\sqrt{k}}{D}$

$$\frac{1}{D} A A^T = \sum w_i w_i^T.$$

Then $\rho^2 = D^2 \langle w_i, w_j \rangle^2 = D^2 \cdot \underbrace{\langle u_i, u_j \rangle^4}_{\approx \frac{1}{d^2} = \frac{1}{D}} = O(D)$

---

* These do not have mean zero and identity covariance, but after applying Π from previous page, they will, see [Hopkins-Schramm-Shi-Steurer '15, Section C.0.4].

So m above is $O(k)$, so

$$\hat{\mathbb{E}} \left\| \frac{1}{D} \sum_i A^T A - Id_k \right\|_{op} \leq \tilde{O}\left(\frac{\sqrt{k}}{d}\right) << 1$$

$$\Rightarrow \hat{\mathbb{E}} \left\| \frac{1}{D} \sum_i A^T A \right\|_{op} = 1 + o(1)$$

$$\Rightarrow \hat{\mathbb{E}} \Big\| \underbrace{\frac{1}{D} \sum_i AA^T}_{\substack{\| \\ \sum_i w_i w_i^T}} \Big\|_{op} = 1 + o(1)$$

$\square$

So we've recovered (2) above. Remains to check that lifting $u_i \to u_i^{\otimes 2}$ doesn't break (1). But let's first specify algorithm:

<u>Alg</u>:

Let $L(u) \triangleq \Pi(u \otimes u)$ "Lifting map"

Solve

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

over degree-$\boxed{12}$ pseudoexpectations s.t.

1). $\hat{\mathbb{E}}$ satisfies $\|x\|^2 = 1$

2). $\left\| \hat{\mathbb{E}}\left[L(x) L(x)^T \right] \right\|_{op} \leq \frac{1}{k}(1+o(1))$

3) $\left\| \hat{\mathbb{E}}\left[\left(L(x) \otimes L(x)\right)\left(L(x) \otimes L(x)\right)^T \right] \right\|_{op} \leq \frac{1}{k}(1+o(1))$

4) $\left\| \hat{\mathbb{E}}\left[\left(L(x)^{\otimes 3}\right)\left(L(x)^{\otimes 3}\right)^T \right] \right\|_{op} \leq \frac{1}{k}(1+o(1))$

Run Jennrich's many times on $\underbrace{\hat{T} \triangleq \hat{\mathbb{E}}\left[L(x)^{\otimes 4}\right]}$,

SoS "hallucination" of $T^* = \sum_{i=1}^{k} u_i^{\otimes 8}$

i.e. compute top eigenvector of

$$\hat{T}(\underset{\mathbb{R}^{d^4}}{\underbrace{g}}, :, :) = \hat{\mathbb{E}}\left[\langle L(x) \otimes L(x), g \rangle L(x) L(x)^T \right]$$

for various g to get vectors in $\mathbb{R}^{d^2}$.

$\boxed{\text{Guaranteed that these contain vectors of the form } L(u_i) \text{ for } 1-o(1) \text{ fraction of } i\text{'s. Can extract } u_i \text{ from } L(u_i) \text{ (we won't prove this)}}$

Main TODO:

<u>Claim</u>: For optimal $\tilde{\mathbb{E}}$, w.h.p over u_i's,

$$\tilde{\mathbb{E}}\left[\sum_i \langle w_i, x^{\otimes 2}\rangle^4\right] \geq 1 - o(1), \qquad (*)$$

so by averaging argument from last time, for $1-o(1)$ fraction of i's,

$$\tilde{\mathbb{E}}\langle w_i, x^{\otimes 2}\rangle \geq \frac{1}{k}(1-o(1)).$$

<u>Pf sketch</u>:

LHS of $(*)$ $= \tilde{\mathbb{E}}\left[\sum_i \langle u_i, x\rangle^8\right]$

We will prove "baby version" of this for intuition:

<u>Baby Claim</u>: $\tilde{\mathbb{E}}\left[\sum_i \langle u_i, x\rangle^4\right] = 1 - o(1)$.

<u>Pf</u>: will give low-degree SoS proof starting from $\sum_i \langle u_i, x\rangle^3 \geq 1 - o(1)$.

$$\left(\sum_i \langle u_i, x\rangle^3\right)^2 = \left\langle \sum_i \langle u_i,x\rangle^2 u_i \ , x \right\rangle^2$$

$$\leq \left\| \sum_i \langle u_i,x\rangle^2 u_i \right\|^2 \cdot \underbrace{\|x\|^2}_{=1}$$

Note $\left\| \sum_i \langle u_i, x\rangle^2 u_i \right\|^2$

$$= \sum_i \langle u_i, x\rangle^4 \underbrace{\|u_i\|^2}_{=1} + \sum_{i\neq j} \langle u_i,x\rangle^2 \langle u_j,x\rangle^2 \langle u_i,u_j\rangle$$

$$= \sum_i \langle u_i, x\rangle^4 + \underbrace{\boxed{\sum_{i\neq j} \langle u_i,x\rangle^2 \langle u_j,x\rangle^2 \langle u_i,u_j\rangle}}_{\text{junk}}$$

So $\sum_i \langle u_i, x\rangle^4 \geq \sum_i \langle u_i,x\rangle^3 - \text{junk}$

To bound "junk", suffices to bound

$$\left\| \sum_{i\neq j} \langle u_i, u_j\rangle (u_i \otimes u_j)(u_i \otimes u_j)^T \right\|_{op.} = o(1)$$

A lossy bound: Let $M \in \mathbb{R}^{d^2 \times \binom{k}{2}}$ have columns consisting of $u_i \otimes u_j$'s. Then

$$\sum_{i \neq j} \langle u_i, u_j \rangle (u_i \otimes u_j)(u_i \otimes u_j)^T = M \operatorname{diag}(\{\langle u_i, u_j \rangle\}_{i \neq j}) M^T,$$

which has $\|\cdot\|_{op} \leq \|M\|_{op}^2 \cdot \underbrace{\max_{i \neq j} \langle u_i, u_j \rangle}_{\lesssim \frac{1}{\sqrt{d}}}$

Note: M is a submatrix of $U \otimes U$, where $U \in \mathbb{R}^{d \times k}$ has columns consisting of u_i's. Then

$$\|M\|_{op}^2 = \|U \otimes U\|_{op}^2 = \|U\|_{op}^4 \lesssim \left(\frac{\sqrt{k}}{\sqrt{d}}\right)^4 = \frac{k^2}{d^2}$$

So $\|M \operatorname{diag}(\ldots) M^T\|_{op} \lesssim \frac{1}{\sqrt{d}} \cdot \frac{k^2}{d^2} = \frac{k^2}{d^{5/2}}$, so

junk $= o(1)$ if $\boxed{k << d^{5/4}}$

Pset 2, question #3 (ungraded) gives a proof of the "right" bound getting $\boxed{k << d^{3/2}}$.