Lecture 4: Smoothed analysis of tensor decompositions

$U, V \in \mathbb{R}^{d \times k}$ arbitrary

$$\tilde{U} \triangleq U + \frac{\rho}{\sqrt{d}} \cdot N(0, Id)^{d \otimes k}$$

$$\tilde{V} \triangleq V + \frac{\rho}{\sqrt{d}} \cdot N(0, Id)^{d \otimes k}$$

Def: $\tilde{U} \odot \tilde{V} \in \mathbb{R}^{d^2 \times k}$ is the Khatri-Rao product:

$$\left(\tilde{U} \odot \tilde{V} \right)_{:j} = vec\left(U_{:j} \otimes V_{:j} \right)$$

Thm: If $k \leq 0.99 d^2$, then with probability $1 - k \exp(-\Omega(d))$,

$$\sigma_{min}\left(\tilde{U} \odot \tilde{V} \right) \geq \Omega\left(\frac{\rho^2}{d^3} \right)$$

① : Reduce to "leave-one-out distance"

Def: Given $M \in \mathbb{R}^{n \times k}$ with columns $M_1, \dots, M_k$, the leave-one-out distance $\ell(M)$ is given by

$$\ell(M) \triangleq \min_{i \in [k]} \| \Pi_i^\perp M_i \|,$$

where $\Pi_i^\perp$ is projection to orthogonal complement of $\mathrm{span}(M_1, ..., M_{i-1}, M_{i+1}, ..., M_k)$.

__Fact__: $\ell(M) \leq \sqrt{k} \cdot \sigma_{min}(M)$

__Proof__: Let $u \in S^{k-1}$ be min. sing. vector so $\|Mu\| = \sigma_{min}(M)$. WLOG suppose u_1 is entry w/ largest magnitude, so $|u_1| \geq \frac{1}{\sqrt{k}}$. Then

$$\ell(M) \leq \|\Pi_1^\perp M_1\| = \inf_{v \in \mathrm{span}(M_2, ..., M_k)} \|M_1 - v\|$$

$$\leq \|M_1 + \underbrace{\sum_{i>1} \frac{u_i}{u_1} M_i}_{\in \mathrm{span}(M_2, ..., M_k)}\| = \frac{1}{|u_1|} \cdot \|Mu\| \leq \sigma_{min}(M) \cdot \sqrt{k}. \quad \square$$

So suffices to lower bound $\ell(\breve{U} \odot \breve{V})$. In fact, will lower bound $\|\Pi(\breve{U} \odot \breve{V})_i\|$ for Π a projector to an arbitrary subspace of dimension $\geq d^2 - k + 1 \geq 0.01 d^2$, for every $i \in [k]$.

Then can just union bound over $i \in [k]$, taking $\Pi = \Pi_i^\perp$ (note: $\Pi_i^\perp$ and $(\breve{U} \odot \breve{V})_i$ random but their randomness is independent)

So   w.t.s :

$$U_{:j} \quad V_{:\hat{s}}$$

Lemma : For $u, v \in \mathbb{R}^d$ arbitrary , $W \subset \mathbb{R}^{d^2}$ arbitrary subspace of dimension $D \geq 0.01 d^2$, if

$$\hat{u} = u + \frac{\rho}{\sqrt{d}} N(0, Id) \quad \text{and} \quad \tilde{v} = v + \frac{\rho}{\sqrt{d}} N(0, Id) \quad, \text{then}$$

$$\| \Pi_W (\hat{u} \otimes \tilde{v}) \| \geq \Omega(\rho^2 / d^2)$$

w.p. at least $1 - \exp(-\Omega(d))$.

(1.5) Warmup w/ one vector :

Baby lemma : For $u \in \mathbb{R}^d$ be arbitrary, $W \subset \mathbb{R}^d$ arbitrary subspace of dimension $D \geq 0.01 d$, if

$$\hat{u} = u + \frac{\rho}{\sqrt{d}} N(0, Id) \quad, \text{then}$$

$$\| \Pi_W \hat{u} \| \geq \Omega(\rho / \sqrt{d})$$

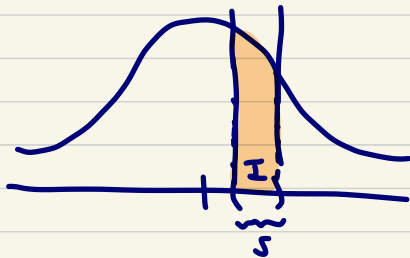w.p. at least $1 - \exp(-\Omega(d))$.

Easy proof that doesn't generalize:

Pick any orthonormal basis $\{w_1,...,w_D\}$ for W, so

$$\| \Pi_W \ddot{u} \| = \| (\langle w_1, \ddot{u} \rangle, ..., \langle w_D, \ddot{u} \rangle) \|.$$

note $\langle w_i, \ddot{u} \rangle = \underbrace{\langle w_i', u \rangle}_{\text{arbitrary}} + \frac{\rho}{\sqrt{d}} N(0,1)$, so

$\langle w_i, \ddot{u} \rangle$ is Gaussian w/ variance $\frac{\rho^2}{d}$.

__Fact__ [Gaussian anti-concentration]: for any

interval I of length s, $\Pr\limits_{g \sim N(0,1)} [g \in I] \leq O(s)$.

So each $\langle w_i, \ddot{u} \rangle$ has $\leq O(1)$ chance of

having magnitude $\leq O(\rho/\sqrt{d})$, so

$$\Pr\left[\| \Pi_W \ddot{u} \| \leq O(\rho/\sqrt{d}) \right] \leq \Pr\left[\langle w_i, \ddot{u} \rangle \leq O(\rho/\sqrt{d}) \right]$$

$$\leq \exp(-\Omega(d)). \qquad \square$$

Alternative proof that will generalize:

Define "row echelon" basis for $W^{\&}$

$$w_1 = (1, \&, *, \ldots, \&)$$
$$w_2 = (0, 1, \&, \ldots, \&)$$
$$w_3 = (0, 0, 1, \&, \ldots, \&)$$
$$\vdots$$
$$w_D = (0, \ldots, 0, 1, \&, \ldots \&)$$

$\Big\}$ $\in W$

$\uparrow$
D^{th} entry

s.t. all $|\&|$'s ≤ 1, so $\|w_i\| \leq \sqrt{\lambda}$ $\forall i$.

Consider "revealing" $\langle w_i, \ddot{u} \rangle$'s in reverse order.

$$\langle w_i, \ddot{u} \rangle = \langle w_i, u \rangle + \frac{\rho}{\sqrt{\lambda}} g_i + \frac{\rho}{\sqrt{\lambda}} \sum_{j>i} (w_i)_j \, g_j \, ,$$

So conditioned on $\langle w_D, \ddot{u} \rangle, \langle w_{D-1}, \ddot{u} \rangle, \ldots, \langle w_{i+1}, \ddot{u} \rangle$, there is still some ==“leftover” randomness== in conditional distribution over $\langle w_i, \ddot{u} \rangle$. By Gaussian anti-conc., conditional prob that $|\langle w_i, \ddot{u} \rangle| \leq O(\rho/\sqrt{\lambda})$ is $\leq O(1)$, so this holds for all i with probability $\exp(-\Omega(\lambda))$.

Recall $\|w_i\| \leq \sqrt{\lambda}$, so for $\hat{w}_i \triangleq \frac{w_i}{\|w_i\|}$,

$$\|\Pi_W \ddot{u}\| \geq |\langle \hat{w}_i, \ddot{u} \rangle| = \frac{1}{\|w_i\|} \cdot |\langle w_i, \ddot{u} \rangle| \geq O(\rho/\lambda).$$

N.B.: We are assuming for every $i \in (D]$, exists $w \in W$ s.t. $w_i \neq 0$. This is WLOG: if not, permute coordinates.

Note, we lost a factor of $\frac{1}{\sqrt{2}}$

(i.e. $\frac{\rho}{d}$ vs $\frac{\rho}{\sqrt{2}}$) in this alternative

proof, but what we gain is that this

alternative proof generalizes.

Back to (non-baby) Lemma:

Need "2D" version of row echelon basis, and

"ordering" specifying how we condition on randomness.

Given $J \subseteq [d] \times [d]$, let $R \subset [d]$ denote row indices,

i.e. $R \triangleq \{ i \text{ st. } \exists (i,j) \in S \}$.

Also $R_i \triangleq \{ j : (i,j) \in S \}$

For simplicity, let's pretend $R = \{ 1, \ldots, D' \}$ for some

$D' \leq D$ ▲

   ↑

  recall $D = \Theta(d^2)$

Lemma:

Can construct S, basis $\{W^{(i,j)}\}_{(i,j) \in S}$ for W, and total ordering $<$ on each R_i s.t. :

1). $W^{(i,j)}$ has max entry 1 in entry (i,j)

2) $W^{(i,j)}$ has zero for every entry (i',j') s.t.

$$i' < i \quad \text{or} \quad j < j$$

3). $|R| \geq \Omega(\delta)$ and $|R_i| \geq \Omega(\delta) \; \forall \; i \in R$.

Pf: omitted, see [Anari et al. '18]

e.g. if $|R| = \sqrt{D}$ and $R_i = \{1, \dots, \sqrt{D}\}$, and

if $<$ were the natural ordering on $\{1, \dots, \sqrt{D}\}$, then

$$W^{(2,3)} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & * & \cdots & * \\ 0 & 0 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & * & * & \cdots & * \end{pmatrix}$$

✱✱ In general $<$ for different rows might not be related ✱✱

To use this to complete the proof, note we w.t.s there is some $(i,j) \in S$ s.t.

$$\left| \tilde{u}^T W^{(i,j)} \tilde{v} \right| \geq \Omega(\rho^2/\delta) \tag{✱}$$

For every $i \in R$, consider
$$e_i^T W^{(i,j)} \hat{v}.$$

Note $\left\{ W^{(i,j)T} e_i \right\}_{j \in R_i}$ are in "row echelon"
form, because $\left(W^{(i,j)T} e_i \right)_j = 1$, and
for all $j' < j$, $\left(W^{(i,j)T} e_i \right)_{j'} = 0$.

So by 1D argument, w/ all but $\exp(-\Omega(d))$ prob.,
$\exists j \in R_i$ s.t.
$$\left| e_i^T W^{(i,j)} \hat{v} \right| \geq \Omega(\rho / \sqrt{d}). \qquad (**)$$

Let $v^{(i)} \triangleq W^{(i,j)} \hat{v}$ for such j maximizing $\left| e_i^T W^{(i,j)} \hat{v} \right|$.

Note $\left\{ \frac{v^{(i)}}{\| v^{(i)} \|_\infty} \right\}$ is in row echelon form, so
by again applying 1D argument, w/ all but
$\exp(-\Omega(d))$ prob., $\exists i \in R$ s.t.
$$\left| \left\langle \frac{v^{(i)}}{\| v^{(i)} \|_\infty}, \hat{u} \right\rangle \right| \geq \Omega(\rho / \sqrt{d}) \qquad (***)$$

So by $(\dagger\dagger)$ and $(\dagger\dagger\dagger)$, recalling $V^{(i)} = W^{(i,j)} \tilde{v}$,

$$\left| \tilde{u}^{\top} W^{(i,j)} \tilde{v} \right| \geq \Omega(\rho^2/\lambda),$$

establishing $(*)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$