9/16/24

Lecture 4: Smoothed analysis of tensor decompositions

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

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

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

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

$$(\hat{U} \odot \hat{V})_{:j} = \mathrm{vec}(U_{:j} \otimes V_{:j})$$

$$\boxed{\begin{array}{l}\textbf{Thm: } \text{If } k \leq 0.99 d^2, \text{ then with probability } 1 - k\exp(-\Omega(d)), \\ \qquad \sigma_{min}(\hat{U} \odot \hat{V}) \geq \Omega\left(\frac{\rho^2}{d^3}\right)\end{array}}$$

---

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

Def: Given $M \in \mathbb{R}^{n \times k}$ with columns $M_1, \ldots, M_k$, 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 $\mathrm{span}(M_1, \dots, M_{i-1}, M_{i+1}, \dots, M_k)$.

<u>Fact</u>: $\ell(M) \le \sqrt{k} \cdot \sigma_{min}(M)$

<u>Proof</u>: 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 $|u_1| \ge \frac{1}{\sqrt{k}}$. Then

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

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

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

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

So w.t.s :

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

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

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

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

1.5 Warmup w/ one vector :

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

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

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

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

Easy proof that doesn't generalize:

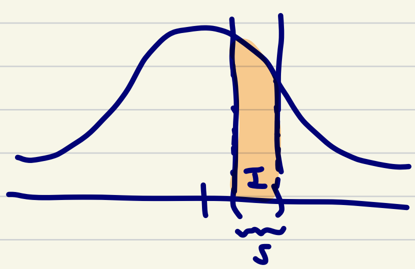
Pick any orthonormal basis $\{w_1,\dots,w_p\}$ for W, so

$$\|\Pi_W \hat{u}\| = \|(\langle w_1,\hat{u}\rangle,\dots,\langle w_p,\hat{u}\rangle)\|.$$

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

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

<u>Fact</u> [Gaussian anti-concentration]: for any interval I of length s, $\Pr_{g\sim N(0,1)}[g\in I] \le O(s)$.



So each $\langle w_i,\hat{u}\rangle$ has $\le O(1)$ chance of having magnitude $\le O(\rho/\sqrt{d})$, so

$$\Pr\left[\|\Pi_W \hat{u}\| \le O(\rho/\sqrt{d})\right] \le \Pr\left[|\langle w_i,\hat{u}\rangle| \le O(\rho/\sqrt{d})\right]$$
$$\le \exp(-\Omega(d)). \qquad \square$$

Alternative proof that will generalize:

Define "row echelon" basis for W *

$$\left.\begin{aligned} w_1 &= (1, *, *, \ldots, *) \\ w_2 &= (0, 1, *, \ldots, *) \\ w_3 &= (0, 0, 1, *, \ldots, *) \\ &\vdots \\ w_D &= (0, \ldots, 0, \underset{\substack{\uparrow \\ D^{th}\text{ entry}}}{1}, *, \ldots, *) \end{aligned}\right\} \in W$$

s.t. all $|*|$'s ≤ 1, so $\|w_i\| \le \sqrt{d}$ $\forall i$.

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

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

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

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

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

---

*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 $\frac{1}{\sqrt{d}}$ (i.e. $\frac{\rho}{d}$ vs $\frac{\rho}{\sqrt{d}}$) 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 $S \subseteq [d] \times [d]$, let $R \subset [d]$ denote row indices, i.e. $R \triangleq \{ i \text{ s.t. } \exists (i,j) \in S \}$.

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

For simplicity, let's pretend $R = \{1, \dots, D'\}$ for some $D' \leq D$ $^{\Delta}$

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

Δ N.B.: As in 1D case on prev. page, might need to permute coordinates first

Lemma:

Can construct S, basis $\{W^{(i,j)}\}_{(i,j)\in S}$ for W, and total ordering $\prec$ 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.

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

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

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

e.g. if $|R| = \sqrt{D}$ and $R_i = \{1, \ldots, \sqrt{D}\}$, and if $\prec$ were the natural ordering on $\{1, \ldots, \sqrt{D}\}$, then

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

** In general $\prec$ 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.

$$\left| \tilde{u}^T W^{(i,j)} \tilde{v} \right| \geq \Omega(\rho^2/d) \qquad (*)$$

For every $i \in R$, consider

$$e_i^T W^{(i,j)} \hat{v}.$$

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

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 $|e_i^T W^{(i,j)} \hat{v}|$.

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 (**) and (***), recalling $v^{(i)} = W^{(i,j)} \hat{v}$,

$$\left| \hat{u}^T W^{(i,j)} \hat{v} \right| \geq \Omega(\rho^2/d),$$

establishing (*). $\square$