

Lecture 3: Iterative methods for tensors

Whitening:

Given: $M = \sum \lambda_i u_i u_i^T$ for u_i 's non-orthogonal

$$T = \sum \lambda_i u_i^{\otimes 3}$$

Let $\underset{d \times k}{V} \underset{k \times k}{D} \underset{k \times d}{V}^T$ be eigendecomposition of M , define

$$W \triangleq V D^{-1/2} \in \mathbb{R}^{d \times k}$$

$$\tilde{u}_i \triangleq \lambda_i W^T u_i \in \mathbb{R}^k$$

Motivation: W transforms M into Id_k !

$$\begin{aligned} W^T M W &= D^{-1/2} \underbrace{V^T V D V^T V}_{\text{Id}_k} D^{-1/2} \\ &= D^{-1/2} D D^{-1/2} = \text{Id}_k \end{aligned}$$

$$= \sum \lambda_i (W^T u_i) (W^T u_i)^T = \sum \tilde{u}_i \tilde{u}_i^T$$

So \tilde{u}_i 's are orthonormal basis for \mathbb{R}^k .

Then consider

$$T' \triangleq T(W, W, W) \in \mathbb{R}^{k \times k \times k}$$

where $T'(x, y, z) \triangleq T(Wx, Wy, Wz)$

$$\begin{aligned} \text{Then } T'(x, y, z) &= \sum_i \lambda_i \langle Wx, u_i \rangle \langle Wy, u_i \rangle \langle Wz, u_i \rangle \\ &= \sum_i \lambda_i (x^T W^T u_i) \cdot (y^T W^T u_i) \cdot (z^T W^T u_i) \\ &= \sum_i \lambda_i (\lambda_i^{-1/2} \langle \tilde{u}_i, x \rangle) \cdot (\lambda_i^{-1/2} \langle \tilde{u}_i, y \rangle) \cdot (\lambda_i^{-1/2} \langle \tilde{u}_i, z \rangle) \\ &= \sum_i \lambda_i^{-1/2} \langle \tilde{u}_i, x \rangle \langle \tilde{u}_i, y \rangle \langle \tilde{u}_i, z \rangle \end{aligned}$$

So $T' = \sum_i \lambda_i^{-1/2} \tilde{u}_i^{\otimes 3}$

ie T' has orthogonal components.

Tensor power method for non-orthogonal components:
(without whitening) [Vatsal-Sharan '17]

$$T = \sum_{i=1}^k u_i^{\otimes 3}$$

Iterates of tensor power method (Z_t) given by:

$$\begin{aligned} Z'_t &\triangleq T(Z_{t-1}, Z_{t-1}, :) \\ &= \sum_{i=1}^k \langle Z_{t-1}, u_i \rangle^2 u_i \end{aligned}$$

$$Z_t \triangleq Z'_t / \|Z'_t\|$$

Define $a_{i,j,t} \triangleq \langle u_i, Z_t \rangle$, $\hat{a}_{i,t} \triangleq \frac{a_{i,t}}{a_{1,t}}$

$$a_{j,t} = \frac{\sum_i a_{i,t-1}^2 \overbrace{\langle u_i, u_j \rangle}^{c_{ij}}}{\|Z'_t\|} = \frac{a_{1,t-1}^2 \sum_i \hat{a}_{i,t-1}^2 c_{ij}}{\|Z'_t\|}$$

Because $\|Z'_t\|$ is fixed factor independent of j ,

$$\hat{a}_{j,t} = \frac{\sum_i \hat{a}_{i,t-1}^2 c_{ij}}{\sum_i \hat{a}_{i,t-1}^2 c_{i1}}$$

Rewriting this further to isolate $i=1$ terms,

$$\hat{\alpha}_{j,t} = \frac{c_{1,j} + \sum_{i \neq 1} \hat{a}_{i,t-1}^2 c_{i,j}}{1 + \underbrace{\sum_{i \neq 1} \hat{a}_{i,t-1}^2 c_{i,1}}_{| \cdot | \leq k c_{\max} \ll 1}}$$

$$\approx \left(c_{1,j} + \sum_{i \neq 1} \hat{a}_{i,t-1}^2 c_{i,j} \right) \left(1 - \sum_{i \neq 1} \hat{a}_{i,t-1}^2 c_{i,1} \right) \quad (1)$$

We show

$$\max_{j \neq 1} |\hat{\alpha}_{j,t}| < \beta_t \quad (\diamond)$$

for sequence (β_t) defined recursively by

$$\begin{aligned} \beta_0 &= \max_{j \neq 1} |\hat{\alpha}_{j,0}| \\ \beta_t &= c_{\max} + \beta_{t-1}^2 + 3k c_{\max} \beta_{t-1}^2 \end{aligned}$$

Provided $k c_{\max} \ll 1 - \beta_0$, can show $\beta_t \leq 1$
(proof omitted)

Pf of (\diamond) :

Note

$$\begin{aligned} \left| c_{1,j} + \sum_{i \neq 1} \hat{a}_{i,t-1}^2 c_{i,j} \right| &= \left| c_{1,j} + \hat{a}_{j,t-1}^2 + \sum_{i \neq 1, j} \hat{a}_{i,t-1}^2 c_{i,j} \right| \\ &\leq c_{\max} + \beta_{t-1}^2 + k c_{\max} \beta_{t-1}^2 \end{aligned}$$

So by (1),

$$\begin{aligned} |\hat{a}_{j,t}| &< (c_{\max} + \beta_{t-1}^2 + kc_{\max} \beta_{t-1}^2) (1 + kc_{\max} \beta_{t-1}^2) \\ &\leq c_{\max} + \beta_{t-1}^2 + 2kc \beta_{t-1}^2 < \beta_t. \quad \square \end{aligned}$$

Then suffices to analyze the recursion defining β_t (note: we have reduced keeping track of $k-1$ different quantities $\hat{a}_{2,t}, \dots, \hat{a}_{k,t}$ to just keeping track of a single quantity!).

To get intuition, consider case where A_i 's orthogonal, i.e. $c_{\max} = 0$.

Then $\beta_t = \beta_{t-1}^2$, so

$$\beta_t = \beta_0^{2^t}$$

i.e. if $\beta_0 < 1$, then $\beta_t \rightarrow 0$ at

doubly exponential rate.

For $c_{\max} > 0$ case, β_0 has to be sufficiently smaller than 1, i.e.

$$1 - \beta_0 \gg k c_{\max}$$

If $c_{\max} \ll \frac{1}{k^2}$, then

this is satisfied w.h.p by randomly initializing (proof omitted, see Lemma 1 in [Sharan-Valiant], link on course page).

Analyzing recursion for (β_t) :

- 3 stages:
- 1). $\beta_t \geq 0.1$
 - 2). $0.1 \geq \beta_t \geq \sqrt{\eta}$ $\eta \triangleq \max(c_{\max}, 1/d)$
 - 3). $\beta_t \leq \sqrt{\eta} \leftarrow \text{😊}$

Stage 1: note $c_{\max} \leq k c_{\max} \beta_{t-1}^2$

b/c $k \beta_{t-1}^2 \geq 0.1 k \geq 1$

(for k larger than some constant)

$$\text{so } \beta_t \leq (1 + 4kc_{\max})\beta_{t-1}^2,$$

and unrolling, we get

$$\begin{aligned} \beta_t &\leq (1 + 4kc_{\max})^{1+\dots+2^{t-1}} \beta_0^{2^t} \\ &= (1 + 4kc_{\max})^{2^t-1} \beta_0^{2^t} \\ &\leq \left[\beta_0 (1 + 4kc_{\max}) \right]^{2^t} \end{aligned}$$

so if $\beta_0 \leq 1 - 5kc_{\max}$ (which happens w.h.p.),

$$\leq \left(1 - \underbrace{kc_{\max}}_{\leq \frac{1}{k^2}} \right)^{2^t}$$

so we stay in this stage for $\lg k$ iterations

Stage 2: Reindex so β_0 is start of this stage

because $\beta_t \geq \sqrt{\eta} \geq \sqrt{c_{\max}}$,

$$\begin{aligned} \beta_t &= (1 + 3kc_{\max})\beta_{t-1}^2 + c_{\max} \leq \underbrace{(2 + 3kc_{\max})}_{< 1} \beta_{t-1}^2 \\ &\leq 3\beta_{t-1}^2 \end{aligned}$$

unrolling, we get

$$\begin{aligned}\beta_t &\leq 3^{2^t - 1} \cdot \beta_0^{2^t} \\ &\leq (3\beta_0)^{2^t} \\ &\leq (0.3)^{2^t}\end{aligned}$$

so we stay in this stage for $O(\lg \lg d)$ iterations.

□