A1 Thm 8

Wednesday, December 25, 2019 10:49 PM

Thm 8 Given a PSD matrix $A \in \mathbb{R}^{d\times d}$ with eigenvalues $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_d$ with normalized eigenvectors $u_1, \ldots, u_d$ we run the power iteration with momentum update A $(*)$ with a unit vector $w_0 \in \mathbb{R}^d$, then we have

$$1 - \frac{(u_1^T w_t)^2}{\|w_t\|^2} \leq \frac{1-(u_1^T w_0)^2}{(u_1^T w_0)^2} \cdot \begin{cases} 4\left(\frac{2\sqrt{\beta}}{\lambda_1 + \sqrt{\lambda_1^2 + 4\beta}}\right. \\ 4\left(\frac{\lambda_2 + \sqrt{\lambda_2^2 - 4\beta}}{\lambda_1 + \sqrt{\lambda_1^2 - 4\beta}}\right) \end{cases}$$

different from Arxiv! $\rightarrow$

$(*)$ $w_t = p_t(A) w_0 = \sum_{i=1}^{d} p_t(\lambda_i) u_i u_i^T w_0$

Denote $\alpha_i = w_0^T u_i = u_i^T w_0$, $\zeta^{(t)} = \max_i \frac{p_t^2(\lambda_i)}{p_t^2(\lambda_1)}$

Proof pt 1

Wednesday, December 25, 2019 11:09 PM

We first prove

$$(u_1^T p_t(A) w_0)^2 = (u_1^T w_t)^2 = d_1^2 p_t^2(\lambda_1) :$$

$$(u_1^T p_t(A) w_0)^2 = (u_1^T \sum_{i=1}^{d} p_t(\lambda_i) u_i u_i^T w_0)^2$$

$$= \left(\sum_{i=1}^{d} p_t(\lambda_i)(u_1^T u_i) u_i^T w_0\right)^2 \quad (\text{push in } u_1^T)$$

$$= \left(p_t(\lambda_1)(u_1^T u_1) u_1^T w_0 + p_t(\lambda_2)(u_1^T u_2) u_2^T w_0 + \dots + p_t(\lambda_d)(u_1^T u_d) u_d^T w_0\right)^2$$

$$= \left(p_t(\lambda_1)(u_1^T u_1) u_1^T w_0\right)^2 \quad (u_1^T u_j = 0 \text{ for } 1 \neq j)$$

$$= p_t^2(\lambda_1) d_1^2 \quad (u_1^T u_1 = 1)$$

$\blacksquare$

Proof pt 2

Wednesday, December 25, 2019 11:26 PM

Now,

$$1-\frac{(u_1^T w_t)^2}{\|w_t\|^2} = 1-\frac{(u_1^T p_t(A) w_0)^2}{w_0^T p_t(A)^2 w_0}$$

We prove $w_0^T p_t(A)^2 w_0 = \sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i)$:

Remember $A^T = A$

$$\begin{aligned}
w_0^T p_t(A)^2 w_0 &= w_0^T p_t(A) \cdot p_t(A) w_0 \\
&= (p_t(A)^T w_0)^T p_t(A) w_0 \\
&= (p_t(A^T) w_0)^T p_t(A) w_0 \\
&= (p_t(A) w_0)^T p_t(A) w_0 \\
&= \left(\sum_{i=1}^{d} p_t(\lambda_i) u_i u_i^T w_0\right)^T \sum_{i=1}^{d} p_t(\lambda_i) u_i u_i^T w_0 \\
&= \left(\sum_{i=1}^{d} p_t(\lambda_i) w_0^T u_i u_i^T\right) \sum_{i=1}^{d} p_t(\lambda_i) u_i u_i^T w_0 \\
&= \sum_{j=1}^{d} \sum_{i=1}^{d} p_t(\lambda_i) p_t(\lambda_j) w_0^T u_i \underbrace{(u_i^T u_j)}_{=0 \text{ for } i \neq j} u_j^T w_0 \\
&= \sum_{i=1}^{d} p_t^2(\lambda_i) (w_0^T u_i)(u_i^T w_0) \\
&= \sum p_t^2(\lambda_i) d_i^2
\end{aligned}$$

∎

Proof pt 3

Wednesday, December 25, 2019 11:48 PM

Now,

$$1-\frac{(u_1^T p_t(A) w_0)^2}{w_0^T p_t^2(A) w_0} = 1-\frac{d_1^2 p_t^2(\lambda_1)}{\sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i)}$$

$$= \frac{\sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i) - d_1^2 p_t^2(\lambda_1)}{\sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i)}$$

$$= \frac{\sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i)}{\sum_{i=1}^{d} d_i^2 p_t^2(\lambda_i)}$$

$$= \frac{\sum_{i=1}^{d} d_i^2 \frac{p_t^2(\lambda_i)}{p_t^2(\lambda_1)}}{\sum_{i=1}^{d} d_i^2 \frac{p_t^2(\lambda_i)}{p_t^2(\lambda_1)}}$$

$$= \frac{\sum_{i=2}^{d} d_i^2 \frac{p_t^2(\lambda_i)}{p_t^2(\lambda_1)}}{d_1^2 \frac{p_t^2(\lambda_1)}{p_t^2(\lambda_1)} + \underbrace{\sum_{i=2} d_i^2 \frac{p_t^2(\lambda_i)}{p_t(\lambda_1)}}_{\geq 0}} \leq \frac{\sum_{i=2}^{d} d_i^2 \delta^{(t)}}{d_1^2}$$

Proof pt 4

Thursday, December 26, 2019 12:00 AM

We bound $S^{(t)}$

Let $k = \max\{k \mid \lambda_k > 2\sqrt{\beta}\}$

We skip to cases:

$$2 \le i \le k$$

$$\left|\frac{p_t(\lambda_i)}{p_t(\lambda_1)}\right| = \frac{\left(\frac{\lambda_i - \sqrt{\lambda_i^2 - 4\beta}}{2}\right)^t + \left(\frac{\lambda_i + \sqrt{\lambda_i^2 - 4\beta}}{2}\right)^t}{\left(\frac{\lambda_1 - \sqrt{\lambda_1^2 - 4\beta}}{2}\right)^t + \left(\frac{\lambda_1 + \sqrt{\lambda_1^2 - 4\beta}}{2}\right)^t}$$

Proof pt 5

Thursday, December 26, 2019 12:15 AM

Note that

$$\left(\frac{\lambda_i - \sqrt{\lambda_i^2 - 4\beta}}{2}\right) \le \frac{\lambda_i + \sqrt{\lambda_i^2 - 4\beta}}{2} \quad \text{since } \sqrt{\lambda_i^2 - 4\beta} \ge 0$$

Then let $\Delta_i := \sqrt{\lambda_i^2 - 4\beta}$

$$\frac{\left(\frac{\lambda_i - \Delta_i}{2}\right)^+ + \left(\frac{\lambda_i + \Delta_i}{2}\right)^+}{\left(\frac{\lambda_1 - \Delta_1}{2}\right)^+ + \left(\frac{\lambda_1 + \Delta_1}{2}\right)^+} \le \frac{\left(\frac{\lambda_i - \Delta_i}{2}\right)^+ + \left(\frac{\lambda_i + \Delta_i}{2}\right)^+}{\left(\frac{\lambda_1 + \Delta_1}{2}\right)^+}$$

$$\le \frac{2\left(\frac{\lambda_i + \Delta_i}{2}\right)^+}{\left(\frac{\lambda_1 + \Delta_1}{2}\right)^+} = \frac{2\left(\lambda_i + \Delta_i\right)^+}{\left(\lambda_1 + \Delta_1\right)^+}$$

Proof pt 6

Thursday, December 26, 2019 12:41 AM

Case $i > k$ is straightforward, except note

$$\left| \frac{p_t(\lambda_i)}{p_t(\lambda_1)} \right| \leq \frac{2(\sqrt{\beta})^t}{\left(\frac{\lambda_1 - \Delta_1}{2}\right)^t + \left(\frac{\lambda_1 + \Delta_1}{2}\right)^t}$$

not $=$ as claimed at the beginning.

Proof pt 7

Thursday, December 26, 2019 12:45 AM

To finish off we show

$$\sum_{i=2}^{d} \frac{d_i^2}{d_1^2} S^{(t)} = \frac{1-(u_1^T w_0)^2}{(u_1^T w_0)^2}$$

Note that $(u_1^T w_0)^2 = \frac{(u_1 \cdot w_0)^2}{\|u_1\|^2 \|w_0\|^2} = \cos^2(\theta)$

$\underline{1} = \leftarrow$

Fact: Let V be a vector space over $\mathbb{R}$. Suppose that $\{u_1, \ldots, u_n\}$ is an orthonormal basis of V. Let θ_i be the angle between w and u_i. Then, $\cos^2\theta_1 + \cos^2\theta_2 + \ldots + \cos^2\theta_n = 1$.

So, $\sum_{i=1}^{d} d_i^2 = \sum_{i=1}^{d} \cos^2(\theta_i) = 1 \iff \sum_{i=2}^{d} d_i^2 = 1 - d_1^2 = 1 - (u_1^T w_0)^2$

$\blacksquare$