COL866: Foundations of Data Science
Notes: Power Method

Let us complete the discussion on the power method. We were analysing the following (randomized) algorithm for computing the first singular value and vectors of the given $n \times d$ data matrix A. We work under the assumption that $\log\left(\frac{\sigma_1}{\sigma_2}\right) \geq \lambda$.

```
FindFirstSingular(A)
  - Sample x_0 from a spherical Gaussian with 0 mean and variance 1
  - s ← log((8d/(εδ)) log(2n/δ)) / (2λ)
  - For i = 1 to s
     - x_i ← (A^T A) x_{i-1}
  - ṽ_1 ← x_s / ||x_s||
  - σ̃_1 ← ||A ṽ_1||
  - ũ_1 ← A ṽ_1 / σ̃_1
  - return(σ̃_1, ũ_1, ṽ_1)
```

For every i we can write $\mathbf{x}_i = \sum_j \alpha_j^i \mathbf{v}_j$. Since $A^T A = \sum_i \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$, we have that $\alpha_j^i = \alpha_j^{i-1}\sigma_j^2$. This further gives $\alpha_j^s = \alpha_j^0 \sigma_j^{2s}$. So, for any i, we have

$$\frac{|\mathbf{x}_s^T \mathbf{v}_1|}{|\mathbf{x}_s^T \mathbf{v}_i|} = \frac{|\alpha_1^0|}{|\alpha_i^0|}\left(\frac{\sigma_1}{\sigma_i}\right)^{2s} \tag{1.0.1}$$

The next claim will help establish the claimed bound on the estimate.

Claim 1: If RHS of the above equation is at least d/ε, then $|\tilde{\sigma_1} - \sigma_1| \leq \varepsilon$

Proof: For any unit vector $\mathbf{x}, ||A\mathbf{x}|| \leq \sigma_1$. So, what we need to show is that $\tilde{\sigma}_1 \geq (1-\varepsilon)\sigma_1$.

We have $\mathbf{x}_s = \sum_j \alpha_j^0 \sigma_j^{2s} \mathbf{v}_j$ and $A = \sum_j \sigma_j \mathbf{u}_j \mathbf{v}_j^T$. So, we have:

$$\begin{aligned}
\tilde{\sigma}_1 = \frac{||A\mathbf{x}_s||}{||\mathbf{x}_s||} &= \frac{||\sum_j \sigma_j \alpha_j^0 \sigma_j^{2s} \mathbf{u}_j||}{||\mathbf{x}_s||} \\
&= \frac{\sqrt{\sum_j (\sigma_j \alpha_j^0 \sigma_j^{2s})^2}}{\sqrt{\sum_j (\alpha_j^0 \sigma_j^{2s})^2}} \\
&\geq \frac{\sigma_1 |\alpha_1^0| \sigma_1^{2s}}{\sqrt{\sum_j (\alpha_j^0 \sigma_j^{2s})^2}} \\
&= \frac{\sigma_1}{\left(1 + \sum_{j=2}^{d} \left(\frac{|\alpha_j^0|}{|\alpha_1^0|} \frac{\sigma_j^{2s}}{\sigma_1^{2s}}\right)^2\right)^{1/2}} \\
&\geq \frac{\sigma_1}{(1 + \varepsilon^2/d)^{1/2}} \\
&\geq \sigma_1 (1 - \frac{\varepsilon^2}{2d}) \\
&\geq \sigma_1 (1 - \varepsilon)
\end{aligned}$$

This completes the proof of the claim. ■

The condition $\frac{|\alpha_1^0|}{|\alpha_i^0|} \left(\frac{\sigma_1}{\sigma_i}\right)^{2s} \geq d/\varepsilon$ gives the following bound on s (which determines the running time of the algorithm):

$$s \geq \frac{\log\left(\frac{d}{\varepsilon} \frac{|\alpha_i^0|}{|\alpha_1^0|}\right)}{2\lambda}$$

So, an upper bound on $\frac{|\alpha_i^0|}{|\alpha_1^0|}$ helps up get a bound on s. First, we note that for all i, α_i^0 is a gaussian with 0 mean and variance 1. The next simple claim gives a lower bound on $|\alpha_1^0|$.

Claim 2: $\mathbf{Pr}[|\alpha_1^0| < \delta/4] \leq \delta/2$.

Proof: This follows from the fact that the gaussian density at 0 is at most 1 and $\mathbf{Pr}[|\alpha_1^0| < \delta/4] \leq 2(\delta/4) = \delta/2$. ■

We now try to get an upper bound on $|\alpha_i^0|$. We again use the fact that α_i^0 is a Gaussian with 0 means and variance 1. For any $t > 1$, we have:

$$\begin{aligned}
\mathbf{Pr}[|\alpha_i^0| > t] &= \frac{2}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2} dx \\
&\leq \frac{2}{\sqrt{2\pi}} \int_t^\infty \frac{x}{t} e^{-x^2/2} dx \\
&\leq \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-t^2/2}}{t} \leq e^{-t^2/2}
\end{aligned}$$

We prove the above next claim using the above inequality.

Claim 3: $\mathbf{Pr}[\forall i, |\alpha_i^0| < \sqrt{2\log(2d/\delta)}] \geq (1 - \delta/2)$.

Proof: The proof follows from a straightforward application of the previous inequality. ■

Combining claims (2) and (3), we get that with probability at least $(1-\delta)$, $\frac{|\alpha_i^0|}{|\alpha_1^0|} \leq \frac{\sqrt{2\log(2d/\delta)}}{\delta/4}$. This completes the argument that with probability at least $(1-\delta)$ the algorithm outputs an ε-accurate estimate of the first singular value.