CSE 392: Matrix and Tensor Algorithms for Data Spring 2024

Lecture 11 — February 21, 2024

Instructor: Shashanka Ubaru *Scribe: Addie Duncan*

1 Randomized SVD by Sampling

The LinearTimeSVD algorithm is a nice way to compute the SVD efficiently. Specifically it requires only one pass over the matrix $\mathbf{A}$ making it useful in streaming contexts. The following result holds for any probability distribution on the sampling.

Proposition 1. Given $\mathbf{A} \in \mathbb{R}^{n\times d}$ and $\mathbf{H}_k$ is computed from the LinearTimeSVD algorithm, then

$$||\mathbf{A} - \mathbf{H}_k\mathbf{H}_k^T\mathbf{A}||_F^2 \leq ||\mathbf{A} - \mathbf{A}_k||_F^2 + 2\sqrt{k}||\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T||_F$$

and

$$||\mathbf{A} - \mathbf{H}_k\mathbf{H}_k^T\mathbf{A}||_2^2 \leq ||\mathbf{A} - \mathbf{A}_k||_2^2 + 2||\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T||_2$$

Proof. Begin by noticing that

$$\begin{aligned}
||\mathbf{A} - \mathbf{H}_k\mathbf{H}_k^T\mathbf{A}||_F^2 &= Tr[(\mathbf{A} - \mathbf{H}_k\mathbf{H}_k^T\mathbf{A})^T(\mathbf{A} - \mathbf{H}_k\mathbf{H}_k^T\mathbf{A})] \\
&= Tr(\mathbf{A}^T\mathbf{A} - 2\mathbf{A}^T\mathbf{H}_k\mathbf{H}_k^T\mathbf{A} + \mathbf{A}^T\mathbf{H}_k\mathbf{H}_k^T\mathbf{H}_k\mathbf{H}_k^T\mathbf{A}) \\
&= Tr(\mathbf{A}^T\mathbf{A} - \mathbf{A}^T\mathbf{H}_k\mathbf{H}_k^T\mathbf{A}) \\
&= ||\mathbf{A}||_F^2 - ||\mathbf{A}^T\mathbf{H}_k||_F^2.
\end{aligned}$$

Then we have the following relations between $||\mathbf{A}^T\mathbf{H}_k||_F^2$ and $\sum_{t=1}^k \sigma_t^2(\mathbf{C})$:

$$\begin{aligned}
\left| ||\mathbf{A}^T\mathbf{H}_k||_F^2 - \sum_{t=1}^k \sigma_t^2(\mathbf{C}) \right| &= \left| \sum_{t=1}^k (||\mathbf{A}^T\mathbf{H}_k||_2^2 - \sigma_t^2(\mathbf{C})) \right| \\
&\leq \sqrt{k} \left(\sum_{t=1}^k (||\mathbf{A}^T\mathbf{H}_t||_2^2 - \sigma_t^2(\mathbf{C}))^2 \right)^{1/2} \\
&= \sqrt{k} \left(\sum_{t=1}^k (||\mathbf{A}^T\mathbf{H}_t||_2^2 - ||\mathbf{C}^T\mathbf{H}_t||_2^2)^2 \right)^{1/2} \\
&= \sqrt{k} \left(\sum_{t=1}^k (\mathbf{H}^T(\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T)\mathbf{H})^2 \right)^{1/2} \\
&\leq \sqrt{k}(||\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T||_F)
\end{aligned}$$

and

$$\begin{aligned}
\left| \sum_{t=1}^k (\sigma_t^2(\mathbf{C}) - \sigma_t^2(\mathbf{A})) \right| &\leq \sqrt{k} \left(\sum_{t=1}^k (\sigma_t^2(\mathbf{C}) - \sigma_t^2(\mathbf{A}))^2 \right)^{1/2} \\
&= \sqrt{k}(||\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T||_F)
\end{aligned}$$

where the last equality follows by a similar argument as above. Combining we get

$$\left| ||\mathbf{A}^T\mathbf{H}_k||_F^2 - \sum_{t=1}^{k} \sigma_t^2(\mathbf{A}) \right| \leq 2\sqrt{k}(||\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T||_F).$$

■

2 Randomized SVD by Sketching

Proposition 2. Given $\mathbf{A} \in \mathbb{R}^{n\times d}$, let $\mathbf{S} \in \mathbb{R}^{m\times n}$ be a sketching matrix such that if it is a Countsketch matrix with $m = O(k^2/\epsilon)$ or SRHT with $m = O(k \log k/\epsilon)$ or Gaussian sketch with $m = O(k/\epsilon)$, then

$$||\mathbf{A} - \hat{\mathbf{A}}_k||_F \leq (1+\epsilon)||\mathbf{A} - \mathbf{A}_k||_F,$$

where $\hat{\mathbf{A}}_k$ is a rank-k approximation in row-space of $\mathbf{SA}$.

Proof. Let $\mathbf{U}_k$ be the top k left singular vectors of $\mathbf{A}$. Consider:

$$||\mathbf{U}_k(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{A}||_F^2.$$

We want to show that this is $(1+\epsilon)||\mathbf{A} - \mathbf{A}_k||_F^2$.

Remark that $\mathbf{A} - \mathbf{A}_k$ is orthogonal to $\mathbf{U}_k$. Take

$$\begin{aligned} ||\mathbf{U}_k(\mathbf{SS}_k)^\dagger\mathbf{SA} - A||_F^2 &= ||\mathbf{U}_k(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{A}_k||_F^2 + ||\mathbf{A} - \mathbf{A}_k||_F^2 \\ &= ||\mathbf{U}_k(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2 + ||\mathbf{A} - \mathbf{A}_k||_F^2 \\ &= ||(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2 + ||\mathbf{A} - \mathbf{A}_k||_F^2 \end{aligned}$$

So it suffices to show that $||(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2$ is $O(\epsilon)||\mathbf{A} - \mathbf{A}_k||_F^2$. Let $A = \mathbf{U}_k\sigma_k\mathbf{V}_k^T + \mathbf{U}_{n-k}\mathbf{\Sigma}_{r-k}\mathbf{V}_{d-k}^T$. Then

$$\begin{aligned} ||(\mathbf{SU}_k)^\dagger\mathbf{SA} - \mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2 &= ||(\mathbf{SU}_k)^\dagger\mathbf{SU}_k\sigma_k\mathbf{V}_k^T + (\mathbf{SU}_k)^\dagger\mathbf{SU}_{n-k}\mathbf{\Sigma}_{r-k}\mathbf{V}_{d-k}^T - \mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2 \\ &= ||I\sigma_k\mathbf{V}_k^T + (\mathbf{SU}_k)^\dagger\mathbf{SU}_{n-k}\mathbf{\Sigma}_{r-k}\mathbf{V}_{d-k}^T - \mathbf{\Sigma}_k\mathbf{V}_k^T||_F^2 \\ &= ||(\mathbf{SU}_k)^\dagger\mathbf{SU}_{n-k}\mathbf{\Sigma}_{r-k}\mathbf{V}_{d-k}^T||_F^2 \end{aligned}$$

and $||(\mathbf{SU}_k)^\dagger\mathbf{SU}_{n-k}\mathbf{\Sigma}_{r-k}\mathbf{V}_{d-k}^T||_F^2$ can be shown to be $O(\epsilon)||\mathbf{A} - \mathbf{A}_k||_F^2$ using the fact that $(\mathbf{SU}_k)^\dagger$ and $(\mathbf{SU}_k)^T$ have the same row space and applying the AMM property. ■