Homework V

Name: Deng Qisheng   Student ID: 16307110232

Lecture Date: May 22, 2019

1. **Solution.**

Use Algorithm 1 to approximate the product of two 2×2 matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ by picking only $k = 1$ column of $\boldsymbol{A}$ and the corresponding row of $\boldsymbol{B}$.

- An example where **uniform sampling** gives the exact solution to $\boldsymbol{AB}$:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}, \ \boldsymbol{B} = \begin{bmatrix} 3 & 4 \\ 3 & 4 \end{bmatrix}$$

where

$$p_1 = \mathbb{P}[i_1 = 1] = \frac{1}{2}, \ p_2 = \mathbb{P}[i_1 = 2] = \frac{1}{2}$$

$$\boldsymbol{C} = \frac{1}{\sqrt{p_1}}\boldsymbol{A} = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2} \end{bmatrix}, \ \boldsymbol{R} = \frac{1}{\sqrt{p_1}}\boldsymbol{B} = \begin{bmatrix} 2\sqrt{2} & 4\sqrt{2} \end{bmatrix}$$

satisfy

$$\boldsymbol{CR} = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 4\sqrt{2} \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 12 & 24 \end{bmatrix} = \boldsymbol{AB}$$

- An example where **uniform sampling** doesn't return an exact solution but **length-squared sampling** returns an exact solution:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \ \boldsymbol{B} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

where **uniform sampling**

$$p_1 = \mathbb{P}[i_1 = 1] = \frac{1}{2}, \; p_2 = \mathbb{P}[i_1 = 2] = \frac{1}{2}$$

$$\boldsymbol{C} = \frac{1}{\sqrt{p_1}}\boldsymbol{A} = \begin{bmatrix} \sqrt{2} \\ 2\sqrt{2} \end{bmatrix}, \; \boldsymbol{R} = \frac{1}{\sqrt{p_1}}\boldsymbol{B} = \begin{bmatrix} \sqrt{2} & 3\sqrt{2} \end{bmatrix}$$

satisfy

$$\boldsymbol{CR} = \begin{bmatrix} \sqrt{2} \\ 2\sqrt{2} \end{bmatrix}\begin{bmatrix} \sqrt{2} & 3\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 4 & 12 \end{bmatrix} \neq \boldsymbol{AB} = \begin{bmatrix} 5 & 15 \\ 10 & 30 \end{bmatrix}$$

where **length-squared sampling**

$$p_1 = \mathbb{P}[i_1 = 1] = \frac{\sqrt{5} \times \sqrt{10}}{\sqrt{5} \times \sqrt{10} + 2\sqrt{5} \times 2\sqrt{10}} = \frac{1}{5},$$
$$p_2 = \mathbb{P}[i_1 = 2] = \frac{2\sqrt{5} \times 2\sqrt{10}}{\sqrt{5} \times \sqrt{10} + 2\sqrt{5} \times 2\sqrt{10}} = \frac{4}{5}$$

$$\boldsymbol{C} = \frac{1}{\sqrt{p_1}}\boldsymbol{A} = \begin{bmatrix} \sqrt{5} \\ 2\sqrt{5} \end{bmatrix}, \; \boldsymbol{R} = \frac{1}{\sqrt{p_1}}\boldsymbol{B} = \begin{bmatrix} \sqrt{5} & 3\sqrt{5} \end{bmatrix}$$

satisfy

$$\boldsymbol{CR} = \begin{bmatrix} \sqrt{5} \\ 2\sqrt{5} \end{bmatrix}\begin{bmatrix} \sqrt{5} & 3\sqrt{5} \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 10 & 30 \end{bmatrix} = \boldsymbol{AB}$$

- An example where neither **uniform sampling** nor **length-squared sampling** returns an exact solution:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \; \boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

where **uniform sampling**

$$p_1 = \mathbb{P}[i_1 = 1] = \frac{1}{2}, \; p_2 = \mathbb{P}[i_1 = 2] = \frac{1}{2}$$

$$\boldsymbol{C} = \frac{1}{\sqrt{p_1}}\boldsymbol{A} = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2} \end{bmatrix}, \; \boldsymbol{R} = \frac{1}{\sqrt{p_1}}\boldsymbol{B} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \end{bmatrix}$$

satisfy

$$\boldsymbol{CR} = \begin{bmatrix} \sqrt{2} \\ 3\sqrt{2} \end{bmatrix}\begin{bmatrix} \sqrt{2} & 2\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} \neq \boldsymbol{AB} = \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix}$$

where **length-squared sampling**

$$p_1 = \mathbb{P}[i_1 = 1] = \frac{\sqrt{10}\times 2\sqrt{5}}{\sqrt{10}\times 2\sqrt{5} + \sqrt{5}\times 5} = \frac{8 - 2\sqrt{10}}{3} \approx 0.558,$$

$$p_2 = \mathbb{P}[i_1 = 2] = \frac{\sqrt{5}\times 5}{\sqrt{10}\times 2\sqrt{5} + \sqrt{5}\times 5} = \frac{-5 + 2\sqrt{10}}{3} \approx 0.442$$

$$\boldsymbol{C} = \frac{1}{\sqrt{p_1}}\boldsymbol{A} \approx \begin{bmatrix}1.339\\4.016\end{bmatrix},\ \boldsymbol{R} = \frac{1}{\sqrt{p_1}}\boldsymbol{B} \approx \begin{bmatrix}1.504 & 3.008\end{bmatrix}$$

satisfy

$$\boldsymbol{CR} = \begin{bmatrix}1.339\\4.016\end{bmatrix}\begin{bmatrix}1.504 & 3.008\end{bmatrix} = \begin{bmatrix}2.014 & 4.028\\6.040 & 12.080\end{bmatrix} \neq \boldsymbol{AB} = \begin{bmatrix}7 & 10\\15 & 22\end{bmatrix}$$

- Compare the variance for the two different sampling schemes:
  - **uniform sampling**:

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{CR} - \boldsymbol{AB}\|_F^2\right] &= 2\left(\sum_{i=1}^{n} \|\boldsymbol{A}(:,i)\|_2^2\|\boldsymbol{B}(i,:)\|_2^2\right) - \|\boldsymbol{AB}\|_F^2\\ &= 2\times\left(10\times 5 + 20\times 25\right) - 858\\ &= \mathbf{242}\end{aligned}$$

  - **length-squared sampling**:

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{CR} - \boldsymbol{AB}\|_F^2\right] &= \left(\sum_{i=1}^{n} \|\boldsymbol{A}(:,i)_2\|\boldsymbol{B}(i,:)\|_2\right)^2 - \|\boldsymbol{AB}\|_F^2\\ &= \left(\sqrt{10}\times\sqrt{5} + 2\sqrt{5}\times 5\right)^2 - 858\\ &\approx \mathbf{8.228}\end{aligned}$$

Observation. Obviously, the variance for **length-squared sampling** is much smaller than the one for **uniform sampling**, which means length-squared sampling is a better method.

2. **Solution.**

Theorem. Given a matrix $\boldsymbol{A}$, construct matrix $\boldsymbol{C}$ with the Algorithm 1 with $p_i = \|\boldsymbol{A}(:,i)\|_2^2/\|\boldsymbol{A}\|_F^2$. Let ϵ be a constant obeying $0 < \epsilon < 3\|\boldsymbol{A}\|_2/\|\boldsymbol{A}\|_F$, then,

$$\mathbb{P}\left[\|\boldsymbol{CC}^\top - \boldsymbol{AA}^\top\|_2 \geq \epsilon\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F\right] \leq 4ne^{-\frac{k\epsilon^2}{2}}$$

Proof. Suppose $\boldsymbol{A} = \boldsymbol{B}^\top$, hence

$$\boldsymbol{CC}^\top - \boldsymbol{AA}^\top = \frac{1}{k}\sum_{t=1}^{k}\left(\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}} - \boldsymbol{AA}^\top\right) = \frac{1}{k}\left(\sum_{t=1}^{k}\boldsymbol{Z}_t\right)$$

note that

$$\frac{\|\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)\|_2}{p_{i_t}} = \frac{\|\boldsymbol{A}(:,i_t)\|_2^2}{p_{i_t}} = \|\boldsymbol{A}\|_F^2$$

Hence,

$$\begin{aligned}
\|\boldsymbol{Z}_t\|_2 &= \left\|\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}} - \boldsymbol{A}\boldsymbol{A}^\top\right\|_2 \\
&\leq \max\left(\left\|\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}}\right\|_2, \|\boldsymbol{A}\boldsymbol{A}^\top\|_2\right) \\
&= \max\left(\|\boldsymbol{A}\|_F^2, \|\boldsymbol{A}\|_2^2\right) \\
&\leq \|\boldsymbol{A}\|_F^2 \\
\left\|\mathbb{E}[\boldsymbol{Z}_t^2]\right\|_2 &= \left\|\mathbb{E}\Big[\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}^2}\Big] - \boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{A}^\top\right\|_2 \\
&\leq \max\left(\left\|\mathbb{E}\Big[\frac{\|\boldsymbol{A}(:,i_t)\|^2}{p_{i_t}}\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}}\Big]\right\|_2, \left\|\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{A}^\top\right\|_2\right) \\
&\leq \max\left(\left\|\sum_{t=1}^k p_{i_t}\Big(\|\boldsymbol{A}\|_F^2\frac{\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)}{p_{i_t}}\Big)\right\|_2, \left\|\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{A}^\top\right\|_2\right) \\
&\leq \max\left(\left\|\sum_{t=1}^k \Big(\|\boldsymbol{A}\|_F^2\boldsymbol{A}(:,i_t)\boldsymbol{A}^\top(i_t,:)\Big)\right\|_2, \left\|\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{A}^\top\right\|_2\right) \\
&\leq \max\left(\|\boldsymbol{A}\|_F^2\|\boldsymbol{A}\boldsymbol{A}^\top\|_2, \|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_2^2\right) \\
&\leq \max\left(\|\boldsymbol{A}\|_F^2\|\boldsymbol{A}\|_2^2, \|\boldsymbol{A}\|_F^2\|\boldsymbol{A}\|_2^2\right) \\
&\leq \|\boldsymbol{A}\|_F^2\|\boldsymbol{A}\|_2^2
\end{aligned}$$

where

$$\begin{aligned}
\|\boldsymbol{A}\boldsymbol{A}^\top\|_2 &= \|\boldsymbol{A}\|_2^2 \\
\|\boldsymbol{A}\boldsymbol{B}\|_2 &\leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|_2 \\
\|\boldsymbol{A}\|_2 &\leq \|\boldsymbol{A}\|_F
\end{aligned}$$

Apply the matrix Bernstein's inequality and the concentration on random matrices, we have

$$\mathbb{P}\Big(\|\boldsymbol{C}\boldsymbol{C}^\top - \boldsymbol{A}\boldsymbol{A}^\top\|_2 \geq \delta\Big) = \mathbb{P}\Big(\Big\|\frac{1}{k}\sum_{t=1}^k \boldsymbol{Z}_t\Big\|_2 \geq \delta\Big) = \mathbb{P}\Big(\Big\|\sum_{t=1}^k \boldsymbol{Z}_t\Big\|_2 \geq k\delta\Big)$$

and

$$\mathbb{P}\Big(\Big\|\sum_{t=1}^k \boldsymbol{Z}_t\Big\|_2 \geq k\delta\Big) \leq 4ne^{-\frac{k^2\delta^2}{\nu^2 + \frac{1}{3}Bk\delta}}$$

Taking

$$\begin{aligned}
B &= \|\boldsymbol{Z}_t\|_2 \leq \|\boldsymbol{A}\|_F^2 \\
\nu^2 &= \Big\|\sum_{t=1}^k \mathbb{E}[\boldsymbol{Z}_t^2]\Big\|_2 \leq \sum_{t=1}^k \left\|\mathbb{E}[\boldsymbol{Z}_t^2]\right\|_2 \leq k\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2 \\
\delta &= \epsilon\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2
\end{aligned}$$

Hence

$$
\begin{aligned}
\mathbb{P}\left(\left\|\sum_{t=1}^{k} \boldsymbol{Z}_t\right\|_2 \geq k\delta\right) &\leq 4ne^{-\frac{k^2\delta^2}{v^2 + \frac{1}{3}Bk\delta}} \\
&\leq 4ne^{-\frac{k^2\left(\epsilon\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F\right)^2}{k\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2 + \frac{1}{3}\|\boldsymbol{A}\|_F^2 k\epsilon\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2}} \\
&\leq 4ne^{-\frac{k^2\left(\epsilon\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F\right)^2}{k\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2 + \frac{k}{3}\|\boldsymbol{A}\|_F^2 \, 3\frac{\|\boldsymbol{A}\|_2}{\|\boldsymbol{A}\|_F}\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F}} \\
&= 4ne^{-\frac{k^2\left(\epsilon\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F\right)^2}{k\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2 + k\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_F^2}} \\
&= 4ne^{-\frac{k\epsilon^2}{2}}
\end{aligned}
$$

which gives

$$
\mathbb{P}\left[\|\boldsymbol{C}\boldsymbol{C}^\top - \boldsymbol{A}\boldsymbol{A}^\top\|_2 \geq \epsilon\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_F\right] \leq 4ne^{-\frac{k\epsilon^2}{2}}
$$

□