CMPSCI 711: More Advanced Algorithms

Vectors 6: Matrix Approximation

Andrew McGregor

Last Compiled: April 26, 2018

Problem

- Input: $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{n \times d}$
- Goal: Approximate $A^T B$. Specifically, we want C such that

$$\mathbb{P}\left[\|C - A^T B\|_F > \epsilon \|A\|_F \|B\|_F\right] \leq \delta$$

where for any matrix M,

$$\|M\|_F := \sqrt{\sum_{i,j} M_{i,j}^2}\ .$$

Sampling Approach

- $A^T B = \sum_i a_i b_i^T$ where a_i and b_i are the ith rows of A and B
- Sample m row where row i is sampled with probability p_i. Let

$$C = \sum_{i\text{: row } i \text{ was sampled}} \frac{a_i b_i^T}{p_i}$$

- Correct in expectation:

$$\mathbb{E}[C] = \sum_i p_i \frac{a_i b_i^T}{p_i} = A^T B$$

- Can show variance is low if $p_i \propto \|a_i\|_2 \|b_i\|_2$, i.e., for constant c,

$$\mathbb{E}\left[\|C - A^T B\|_F^2\right] \leq c\|A\|_F^2 \|B\|_F^2 / m$$

- And so for $m \geq c/(\epsilon^2 \delta)$ by Markov,

$$\mathbb{P}\left[\|C - A^T B\|_F > \epsilon \|A\|_F \|B\|_F\right] \leq \frac{\mathbb{E}\left[\|C - A^T B\|_F^2\right]}{\epsilon^2 \|A\|_F^2 \|B\|_F^2} \leq \delta$$

Reducing Dependency on Failure Probability

- Set $\delta = 1/3$ in above procedure. Generate $t = O(\log 1/\eta)$ estimates

$$C_1, \ldots, C_t$$

- With probability $\geq 1 - \eta$, more than $t/2$ estimates are good, i.e.,

$$\|C - A^T B\|_F \leq \epsilon \|A\|_F \|B\|_F$$

- Return any C_i with $|S_i| \geq t/2$ where

$$S_i = \{j : \|C_i - C_j\|_F \leq 2\epsilon \|A\|_F \|B\|_F\}$$

- Then if C_i is good and C_j is also good then

$$\|C_i - C_j\|_F \leq \|C_i - A^T B\|_F + \|C_j - A^T B\|_F \leq 2\epsilon \|A\|_F \|B\|_F$$

  and hence $|S_i| \geq t/2$.
- If C_i very bad, i.e., $\|C - A^T B\|_F > 4\epsilon \|A\|_F \|B\|_F$, and C_j is good,

$$\|C_i - C_j\|_F \geq \|C_i - A^T B\|_F - \|C_j - A^T B\|_F > 2\epsilon \|A\|_F \|B\|_F$$

  and hence $|S_i| < t/2$.

JL Approach

Definition (Moment Property)

Let D be a distribution over matrices $\Pi \in \mathbb{R}^{m \times n}$. We say D satisfies the (ϵ, δ, p)-JL moment property if for any unit x,

$$\left\| \|\Pi x\|_2^2 - 1 \right\|_p \leq \epsilon \delta^{1/p}$$

where for a random variable Z, $\|Z\|_p$ is defined as $\left(\mathbb{E}\left[|Z|^p\right]\right)^{1/p}$.

For example,

- CountSketch and F_2 estimation sketch satisfy $(\epsilon, \delta, 2)$-JL property with $m \approx \epsilon^{-2}\delta^{-1}$.
- If entries of Π are Gaussian then sketch satisfies $(\epsilon, \delta, \log 1/\delta)$-JL property with $m \approx \epsilon^{-2} \log \delta^{-1}$.

Theorem

Suppose D has (ϵ, δ, p)-JLMP for $p \geq 2$ then

$$\mathbb{P}\left[\|A^T B - (\Pi A)^T \Pi B\|_F > 3\epsilon \|A\|_F \|B\|_F\right] < \delta \ .$$

Proof of Theorem

- Let $M = A^T B - (\Pi A)^T \Pi B$. Then, by the Markov inequality,

$$\mathbb{P}\left[\|M\|_F > 3\epsilon\|A\|_F\|B\|_F\right] < \frac{\mathbb{E}\left[\|M\|_F^p\right]}{(3\epsilon\|A\|_F\|B\|_F)^p}$$

- Let x_i and y_i be columns of A and B,

$$M_{i,j}^2 = \left(\langle \Pi x_i, \Pi y_j\rangle - \langle x_i, y_j\rangle\right)^2 = T_{i,j}^2\|x_i\|_2^2\|y_i\|_2^2$$

  where $T_{i,j} = \langle \Pi \frac{x_i}{\|x_i\|_2}, \Pi \frac{y_j}{\|y_j\|_2}\rangle - \langle \frac{x_i}{\|x_i\|_2}, \frac{y_j}{\|y_j\|_2}\rangle$.
- Lemma: $\|T_{i,j}^2\|_{p/2} = \|T_{i,j}\|_p^2 \leq (3\epsilon\delta^{1/p})^2$
- $\mathbb{E}\left[\|M\|_F^p\right] = \|\sum_{i,j} M_{i,j}^2\|_{p/2}^{p/2}$ and

$$\begin{aligned}
\left\|\sum_{i,j} M_{i,j}^2\right\|_{p/2} = \left\|\sum_{i,j} T_{i,j}^2\|x_i\|_2^2\|y_j\|_2^2\right\|_{p/2} &\leq \sum_{i,j}\|x_i\|_2^2\|y_j\|_2^2\left\|T_{i,j}^2\right\|_{p/2} \\
&\leq (3\epsilon\delta^{1/p})^2\sum_{i,j}\|x_i\|_2^2\|y_j\|_2^2 \\
&= (3\epsilon\delta^{1/p})^2\|A\|_F^2\|B\|_F^2
\end{aligned}$$

- And so $\mathbb{E}\left[\|M\|_F^p\right] \leq ((3\epsilon\delta^{1/p})^2\|A\|_F^2\|B\|_F^2)^{p/2} = \delta(3\epsilon\|A\|_F\|B\|_F)^p$.

Proof of Lemma

Suffices to show that for any unit x, y, $\|\langle \Pi x, \Pi y\rangle - \langle x, y\rangle\|_p \leq 3\epsilon\delta^{1/p}$.

- Can rewrite $\langle \Pi x, \Pi y\rangle - \langle x, y\rangle$ as

$$\frac{\|\Pi x\|_2^2 + \|\Pi y\|_2^2 - \|\Pi(x-y)\|_2^2 - \|x\|_2^2 - \|y\|_2^2 + \|x-y\|_2^2}{2}$$

- By triangle inequality, $\|\langle \Pi x, \Pi y\rangle - \langle x, y\rangle\|_p$ is at most

$$\begin{aligned} & \frac{\|\|\Pi x\|_2^2 - \|x\|_2^2\|_p + \|\|\Pi y\|_2^2 - \|y\|_2^2\|_p + \|\|\Pi(x-y)\|_2^2 - \|x-y\|_2^2\|_p}{2} \\ \leq\ & \frac{\epsilon\delta^{1/p} + \epsilon\delta^{1/p} + \|x-y\|_2^2\epsilon\delta^{1/p}}{2} \\ \leq\ & 3\epsilon\delta^{1/p} \end{aligned}$$