Randomized Algorithm Note *Algorithm for Mordern Dataset*

Matrix Multiplication Approximation

Zijian Zhen *Dec-12-2020*

Algorithm 1: MatrixMultiplicationApproximation

Data: An $m \times n$ matrix A, and $n \times p$ matrix B, a positive integer c, and probabilites $\{p_i\}_{i=1}^n$
Result: Matrix C and R such that $CR \approx AB$
for $t = 1$ *to* c **do**
  Pick $i_t\{1, \ldots, n\}$ with probability $\Pr[i_t = k] = p_k$, in i.i.d. trails, with replacement;
  Set $C^{(t)} = A^{(i_t)}/\sqrt{cp_{i_t}}$ and $R_{(t)} = B_{(i_t)}$;
end

Lemma 1 *Given matrices A and B, by applying the algorithm to construct matrices C and R. Then,*

$$E[(CR))_{ij}] = (AB)_{ij}$$

and

$$\mathrm{Var}\left[(CR)_{ij}\right] = \frac{1}{c}\sum_{k=1}^{n}\frac{A_{ik}^2 B_{kj}^2}{p_k} - \frac{1}{c}(AB)_{ij}^2$$

Proof: Fix i, j. For $t = 1, ..., c$, define $X_t = \left(\frac{A^{(i_t)}B_{(i_t)}}{cp_{i_t}}\right)_{ij} = \frac{A_{ii_t}B_{i_tj}}{cp_{i_t}}$. Thus,

$$\mathbf{E}\left[X_t\right] = \sum_{k=1}^{n} p_k \frac{A_{ik}B_{kj}}{cp_k} = \frac{1}{c}(AB)_{ij} \quad \text{and} \quad \mathbf{E}\left[X_t^2\right] = \sum_{k=1}^{n}\frac{A_{ik}^2 B_{kj}^2}{c^2 p_k}$$

Since by construct $(CR)_{ij} = \sum_{t=1}^{c} X_t$, we have $\mathbf{E}\left[(CR)_{ij}\right] = \sum_{t=1}^{c}\mathbf{E}\left[X_t\right] = (AB)_{ij}$. Since $(CR)_{ij}$ is the sum of c independent random variables, $\mathrm{Var}\left[(CR)_{ij}\right] = \sum_{t=1}^{c}\mathrm{Var}\left[X_t\right]$. Since $\mathrm{Var}\left[X_t\right] = \mathbf{E}\left[X_t^2\right] - \mathbf{E}\left[X_t\right]^2$, we have:

$$\mathrm{Var}\left[X_t\right] = \sum_{k=1}^{n}\frac{A_{ik}^2 B_{kj}^2}{c^2 p_k} - \frac{1}{c^2}(AB)_{ij}^2$$

Thus,

$$\mathrm{Var}\left[(CR)_{ij}\right] = \frac{1}{c}\sum_{k=1}^{n}\frac{A_{ik}^2 B_{kj}^2}{p_k} - \frac{1}{c}(AB)_{ij}^2$$

Lemma 2 *Given matrices A and B, by applying the algorithm to construct matrices C and R. Then,*

$$\mathbf{E}\left[\|AB - CR\|_F^2\right] = \sum_{k=1}^{n}\frac{\left\|A^{(k)}\right\|_2^2\left\|B_{(k)}\right\|_2^2}{cp_k} - \frac{1}{c}\|AB\|_F^2$$

And if,

$$p_k = \frac{\left\|A^{(k)}\right\|_2\left\|B_{(k)}\right\|_2}{\sum_{k'=1}^{n}\left\|A^{(k')}\right\|_2\left\|B_{(k')}\right\|_2}$$

then,

$$\mathbf{E}\left[\|AB - CR\|_F^2\right] = \frac{1}{c}\left(\sum_{k=1}^{n}\left\|A^{(k)}\right\|_2\left\|B_{(k)}\right\|_2\right)^2 - \frac{1}{c}\|AB\|_F^2$$

Proof: Firstly, according to the property of Variance,

$$\mathbf{E}\left[\|AB - CR\|_F^2\right] = \sum_{i=1}^{m}\sum_{j=1}^{p}\mathbf{E}\left[(AB - CR)_{ij}^2\right] = \sum_{i=1}^{m}\sum_{j=1}^{p}\mathbf{Var}\left[(CR)_{ij}\right]$$

From **Lemma 1** we have:

$$\begin{aligned}\mathbf{E}\left[\|AB - CR\|_F^2\right] &= \frac{1}{c}\sum_{k=1}^{n}\frac{1}{p_k}\left(\sum_i A_{ik}^2\right)\left(\sum_j B_{kj}^2\right) - \frac{1}{c}\|AB\|_F^2 \\ &= \frac{1}{c}\sum_{k=1}^{n}\frac{1}{p_k}\left\|A^{(k)}\right\|_2^2\left\|B_{(k)}\right\|_2^2 - \frac{1}{c}\|AB\|_F^2\end{aligned}$$

If the value $p_k = \frac{\|A^{(k)}\|_2\|B_{(k)}\|_2}{\sum_{k'=1}^{n}\left\|A^{(k')}\right\|_2\left\|B_{(k')}\right\|_2}$ is applied in the expression (the proof for the magic value p_k will be introduced in **Lemma 3**), then,

$$\mathbf{E}\left[\|AB - CR\|_F^2\right] = \frac{1}{c}\left(\sum_{k=1}^{n}\left\|A^{(k)}\right\|_2\left\|B_{(k)}\right\|_2\right)^2 - \frac{1}{c}\|AB\|_F^2$$

The expression shows that with the given value p_k, the algorithm minimize the expected value of the Frobenius norm of the error between AB and CR.

Lemma 3 *Sampling probabilities* $p_k = \frac{\|A^{(k)}\|_2\|B_{(k)}\|_2}{\sum_{k'=1}^{n}\left\|A^{(k')}\right\|_2\left\|B_{(k')}\right\|_2}$ *minimize* $\mathbf{E}\left[\|AB - CR\|_F^2\right]$

Proof: Define the function:

$$f\left(p_1, \ldots p_n\right) = \sum_{k=1}^{n}\frac{1}{p_k}\left\|A^{(k)}\right\|_2^2\left\|B_{(k)}\right\|_2^2$$

Function $f\left(p_1, \ldots p_n\right)$ dominate the value of $\mathbf{E}\left[\|AB - CR\|_F^2\right]$ on the p_k's for $k = 1, 2, \ldots, n$. To minimize the function, notice that $\sum_{k=1}^{n} p_k = 1$, we will apply the Lagrange multiplier λ and define the function:

$$\mathcal{L}\left(p_1, \ldots p_n\right) = f\left(p_1, \ldots p_n\right) - \lambda\left(\sum_{k=1}^{n} p_k - 1\right)$$

Let the partial derivative on p_k equals to 0, we can solve the optimal p_k:

$$\frac{\partial\mathcal{L}}{\partial p_i} = \frac{-1}{p_i^2}\left\|A^{(i)}\right\|_2^2\left\|B_{(i)}\right\|_2^2 + \lambda = 0$$

Notice again that $\sum_{k=1}^{n} p_k = 1$, we can solve λ by applying this fact:

$$p_i = \frac{\left\|A^{(i)}\right\|_2\left\|B_{(i)}\right\|_2}{\sqrt{\lambda}} = \frac{\left\|A^{(i)}\right\|_2\left\|B_{(i)}\right\|_2}{\sum_{i'=1}^{n}\left\|A^{(i')}\right\|_2\left\|B_{(i')}\right\|_2}$$