6.896 Sublinear Time Algorithms	November 02, 2004
Lecture 15	
Lecturer: Eli Ben-Sasson	*Scribe: Swastik Kopparty*

1 Approximating Matrix Product in Sublinear Time

Today we consider the problem of multiplying matrices in sub-linear time. The results are from the work of Drineas, Kannan and Mahoney [1]. Given $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$, we wish to approximate $A \cdot B \in \mathbb{R}^{m\times p}$ in sublinear time. Think of $m, p \ll n$, so that the size of the output is sublinear in the size of the input.

For starters, we need to define what we mean by approximating a matrix. will use the *Frobenius norm* as our measure of proximity.

Definition 1 (Matrix Norms) *For a $m \times p$ matrix M, the* Frobenius Norm *of M is*

$$||M||_F \doteq \sqrt{\sum_{i\in[m],j\in[p]} M_{i,j}^2}.$$

The L_2 norm of M is

$$||M||_2 = \sup_{x\in\mathbb{R}^n} |Mx|.$$

In our context, we will say M approximates AB if $||M - AB||_F$ is small.

The Frobenius norm is related to the L_2 norm by the following inequalities:

$$||M||_2 \leq ||M||_F \leq \sqrt{n}||M||_2.$$

Thus an approximation in Frobenius Norm gives some approximation in L_2 norm. We point out that [1] give more results on approximating the L_2 norm, which we won't describe.

1.1 The Sublinear Approximation Algorithm

Key Observation The product AB, which is an $m \times p$ matrix, is a sum of n rank one $m \times p$ matrices. Thus, approximating AB ammounts to approximating this sum, which can be done in sublinear time by sampling a few of the summands. And now for the details.

Let $a^{(t)}$ be the t^{th} column of A and $b_{(t)}$ be the t^{th} row of B. Formally, the key observation is:

$$AB = \sum_{t=1}^{n} a^{(t)} b_{(t)}$$

Notice each summand is a rank one, $m \times p$ matrix. We need to estimate the sum of n such matrices, motivating the following algorithm, that has oracle access to matrices $A \in R^{m\times n}, B \in \mathbb{R}^{n\times p}$ and receives as explicit inputs the matrix dimensions m, n, p and a parameter c.

Matrix-Multiplication$^{A,B}[m,n,p,c]$

- Pick random $i_1, i_2, \dots i_c \in [n]$ independently and uniformly.
- Let $C = \left[a^{(i_1)} \dots a^{(i_c)}\right]$, $R = \begin{bmatrix} b_{(i_1)} \\ \vdots \\ b_{(i_c)} \end{bmatrix}$
- Output $\frac{n}{c}CR$

1.2 Analysis of Approximation Algorithm

Consider the matrix A that has one non-zero entry which is huge (say, $A_{1,1} = 10^{100}$). Then the Frobenius norm of $A \cdot A^T$ is also huge, but our approximation algorithm will be non-zero only with probability c/n ($= \Pr[1 \in \{i_1, \dots i_c\}]$). Thus, in our analysis we will impose a constraint on the Frobenius norm of the rank one matrices we are sampling, and arrive at the following Theorem.

Theorem 2 (Algorithm Analysis) *[1] Let $\beta \le 1$ be the maximal value such that*

$$\forall k \in [n], \beta \cdot ||a^{(k)} b_{(k)}||_F \le E||a^{(\ell)} b_{(\ell)}||_F$$

Where expectation is over uniformly random $\ell \in [n]$. Then,

1. *The expected output of our approximation algorithm is close to AB:*

$$E\left[||AB - \frac{n}{c}CR||_F^2\right] \le \frac{1}{\beta c}||A||_F^2 ||B||_F^2$$

2. *Furthermore, the approximation is usually close to the expectation:*

$$\forall \delta \in (0,1), \Pr\left[||AB - \frac{n}{c}CR||_F^2 \le \frac{\eta^2}{\beta c}||A||_F^2||B||_F^2\right] > 1 - \delta$$

where $\eta = 1 + \sqrt{\frac{8}{\beta}\log\frac{1}{\delta}}$.

1.3 Proof of Theorem 2

We will only prove the first part of the Theorem. We point out that the second part is proved using standard bounds on tails of the probability (in particular, Azuma's inequality). The first part follows by analyzing the expectation and variance of a single entry of our random matrix CR, and then summing up on all entries to bound the Frobenius norm of the random matrix. Now for the details.

Lemma 3 *Fix $i \in [m], j \in [p]$. Then,*

$$E\left[CR_{ij}\right] = \frac{c}{n}AB_{ij} \tag{1}$$

$$Var\left[CR_{ij}\right] \le \frac{c}{n}\sum_k (a^{(k)} b_{(k)})^2_{ij} \tag{2}$$

Proof Let $a_i^{(k)}$ be the i^{th} entry of (the column vector) $a^{(k)}$ and let $b^j_{(k)}$ be the j^{th} entry of (the row vector) $b_{(k)}$. Let $D \in \mathbb{R}^n$ be the vector with $D_k = a_i^{(k)} \cdot b^j_{(k)}$. Let X be CR_{ij}. Note $AB_{ij} = \sum_k D_k$. Also, $X = \sum_{t=1}^c X_t$, where X_t is a uniform sample from D. We first compute the expectation,

$$E\left[CR_{ij}\right] = E\left[X\right] = cE\left[X_t\right] = \frac{c}{n}AB_{ij}$$

Next, we compute the variance,

$$\mathrm{Var}(X) = c\mathrm{Var}(X_t) = c(E[X_t^2] - E[X_t]^2) = \left(\frac{c}{n}\sum_k D_k^2\right) - \frac{c^3}{n^2}(AB_{ij})^2 \le \frac{c}{n}\sum_k D_k^2$$

The last inequality follows because $\frac{c^3}{n^2}(AB_{ij})^2$ is non-negative. Our proof is complete by recalling $D_k = a_i^{(k)} \cdot b_{(k)}^j = (a^{(k)}b_{(k)})_{ij}$. ■

Proof [Of Theorem 2, Part 1:] Let us compute the expected square of the Frobenius distance between AB and our approximation.

$$E\left[||AB - \frac{n}{c}CR||_F^2\right] = \sum_{i,j} E\left[(AB_{ij} - \frac{n}{c}CR)_{ij}^2\right] = \sum_{i,j} E\left[\left(\frac{n}{c}E[CR_{ij}] - \frac{n}{c}CR_{ij}\right)^2\right] = \frac{n^2}{c^2}\sum_{i,j}\mathrm{Var}(CR_{ij})$$

The first equality above is linearity of expectation, the second uses Lemma 3, and the third follows by the definition of variance. By using Lemma 3 again (this time using its second part), we get

$$E\left[||AB - \frac{n}{c}CR||_F^2\right] \le \frac{n}{c}\sum_{i,j}\sum_k (a^{(k)}b_{(k)})_{ij}^2 = \frac{n}{c}\sum_{k\in[n]} ||a^{(k)}b_{(k)}||_F^2$$

The last equality above follows by changing order of summation. Recall our assumption

$$||a^{(k)}b_{(k)}||_F \le E||a^{(\ell)}b_{(\ell)}||_F/\beta = \frac{1}{\beta n}\sum_{k'} ||a^{(k')}b_{(k')}||_F$$

Using the assumption, we get

$$\frac{n}{c}\sum_{k\in[n]} ||a^{(k)}b_{(k)}||_F^2 \le \frac{n}{c}\sum_{k\in[n]} ||a^{(k)}b_{(k)}||_F \cdot \frac{1}{\beta n}\sum_{k'\in[n]} ||a^{(k')}b_{(k')}||_F = \frac{1}{\beta c}\left(\sum_{k\in[n]} ||a^{(k)}b_{(k)}||_F\right)^2$$

The first inequality uses the assumption and the fact that the Frobenius norm is non-negative. Notice $||a^{(k)}b_{(k)}||_F = |a^{(k)}| \cdot |b_{(k)}|$ (this is proved by expanding both sides of the equality). Using the Cauchy Schwartz inequality (second inequality below) we conclude our proof,

$$E\left[||AB - \frac{n}{c}CR||_F^2\right] \le \frac{1}{\beta c}\left(\sum_{k\in[n]} |a^{(k)}| \cdot |b_{(k)}|\right)^2 \le \frac{1}{c\beta}\left(\sum |a^{(k)}|^2\right)\left(\sum |b_{(k)}|^2\right) = \frac{1}{c\beta}||A||_F^2 \cdot ||B||_F^2.$$

■

References

[1] P. Drineas, R. Kannan, M. W. Mahoney. Fast Monte Carlo Algorithms for Matrices I: Approximating Matrix Multiplication. *YALEU/DCS/TR-1269* (submitted for publication), 2004.