CS 594: Representations in Algorithm Design **Spring 2022**

Lecture 22: March 31, 2022

Lecturer: Xiaorui Sun *Scribe: Rishi Advani*

1 Last time and today

Today, we will be continuing to study matrix multiplication algorithms with time complexity less than $O(n^3)$.

We have already seen Strassen's algorithm, which has time complexity $O(n^{2.81})$. In order to do significantly better, we have to introduce tensors.

This is the matrix multiplication tensor:

$$t_{ij',jk',ki'} = \begin{cases} 1 & \text{if } i = i', j = j', k = k' \\ 0 & \text{otherwise} \end{cases}$$

Next, we define the rank of a tensor.

Definition 1 (Rank)**.** The *rank* of a tensor t is the minimum number $R(t)$ such that there exists a set of $R(t)$ rank-1 tensors that sum to t.

Observation 2. *If $R(\langle K, M, N\rangle) \leq r$, then $\omega \leq \frac{3 \log r}{\log NMK}$, where ω is the optimal matrix multiplication exponent.*

Our goal is to find an upper bound on $R(\langle n, n, n\rangle)$. It is difficult to bound the rank of this large tensor, so people often try to bound the sums and products of smaller tensors to get new bounds for ω.

Definition 3 (Kronecker product)**.** Given $t \in \mathbb{F}^{K \times M \times N}$ and $t' \in \mathbb{F}^{K' \times M' \times N'}$, we have $(t \otimes t')_{\substack{KK',MM',NN' \\ ii',jj',kk'}} = t_{ijk} \cdot t'_{i',j',k'}$.

We also have the following bound on the rank of a Kronecker product.

Observation 4.

$$R(t \otimes t') \leq R(t) \cdot R(t')$$

At the end of the last lecture, we mentioned that if we have $T = N \times M \times K$, then $\langle T, T, T\rangle = \langle K, N, M\rangle \otimes \langle N, M, K\rangle \otimes \langle M, K, N\rangle$. Once we can get a bound on the rank for the RHS, then we can get one for the LHS.

Remark. Strassen essentially showed that $R(\langle 2,2,2\rangle) \leq 7$ and then used our observation to get an upper bound for ω.

Relevant Readings:

- Bini, Dario Andrea, Milvio Capovani, Francesco Romani, and Grazia Lotti. "O (N2. 7799) COMPLEXITY FOR N BY N APPROXIMATE MATRIX MULTIPLICATION." (1979): 234-235.

2 Direct sum

The next operation we will be using is the direct sum.

Definition 5 (Direct sum)**.** Given $t_{i,j,k} \in \mathbb{F}^{K\times M\times N}$ and $t'_{i',j',k'} \in \mathbb{F}^{K'\times M'\times N'}$, then the direct sum is defined as follows.

$$t \oplus t' = \begin{cases} t_{ijk} & \text{if} i \leq K, j \leq M, k \leq N \\ t'_{i-K,j-M,k-N} & \text{if} i > K, j > M, k > N \\ 0 & \text{otherwise} \end{cases}$$

We also have the following bound on the rank of a direct sum.

Observation 6.

$$R(t \oplus t') \leq R(t) + R(t')$$

3 History

Many people used this idea of analyzing the rank of smaller tensors to get better bounds on ω.

Strassen proved that $R(\langle 2,2,2\rangle) = 7$. It was also shown that $R(\langle 2,2,3\rangle) = 11$, but this did not lead to better bounds on ω than the previous case. It has also been shown that $19 \leq R(\langle 3,3,3\rangle) \leq 23$. If $R(\langle 3,3,3\rangle) \leq 21$, then we have $\omega \leq 2.79$. Pan (1980) found that $R(\langle 70,70,70\rangle) \leq 143640$, which implies $\omega < 2.8$.

This was the status of the problem in the 1980s. To make more progress, new ideas other than directly bounding the rank of small tensors were needed.

4 Approximate tensors

The idea is to somehow "approximate" tensors.

To get an intuition, suppose we have an infinite sequence of matrices $M_1, M_2, \ldots$.

Claim 7. Suppose as j goes to infinity, M_j converges to some matrix M. If $r(M_j) \leq r$ for all j, then we can say $r(M) \leq r$.

Proof sketch. Look at any $(r+1) \times (r+1)$ submatrix P_j of M_j. Fix the submatrix that we are using for all M_j. The determinant of P_j is 0, and we can show that it is a continuous function, so then P has determinant 0. Then, M has rank at most r. □

A similar idea makes it so that bounding the rank of an approximate matrix multiplication tensor gives us a bound on the exact matrix multiplication tensor. The same idea cannot be directly translated, though, since it is difficult to define the determinant for tensors.

Suppose we have a tensor t of rank 3 that is with respect to $\{x_0, x_1\}, \{y_0, y_1\}, \{z_0, z_1\}$.

$$t = x_0y_0z_0 + x_1y_0z_1 + x_0y_1z_1$$

Define a tensor with a parameter ϵ.

$$\begin{aligned} t(\epsilon) &= (x_0 + \epsilon x_1) \cdot (y_0 + \epsilon y_1) \cdot 1/\epsilon \cdot z_1 + x_0 \times y_0(z_0 - z_1/\epsilon) \\ &= x_0y_0 1/\epsilon z_1 + x_0y_1z_1 + x_1y_0z_1 + \epsilon x_1y_1z_1 + x_0y_0z_0 - 1/\epsilon x_0y_0z_1 \\ &= x_0y_1z_1 + x_1y_0z_1 + \epsilon x_1y_1z_1 + x_0y_0z_0 \end{aligned}$$

The rank of $t(\epsilon)$ is exactly 2. As ϵ goes to 0, $t(\epsilon)$ goes to t. $t(\epsilon)$ is an approximation of t as ϵ shrinks, but its rank is smaller. Thus, for a fixed tensor t, we can find an approximation with smaller rank.

Now we are going to try making use of an approximation with smaller rank to speed up matrix multiplication. Before we had coefficients from a field $\mathbb{F}$. For the approximation, we extend this field with ϵ, so we have the field $\mathbb{F}[\epsilon]$.

Definition 8 (Border rank)**.** Given a tensor t and an integer h, let $R_h(t)$ be the smallest integer ℓ such that the following equation holds for some $t'(\epsilon)$.

$$\epsilon^{h-1}t + \epsilon^h t'(\epsilon) = \sum_{\lambda=1}^{\ell} \left(\sum U_{\lambda_i} X_i\right) \left(\sum V_{\lambda_j} y_j\right) \left(\sum W_{\lambda_k} y_k\right),$$

where U_{λ_i}, V_{λ_j}, and W_{λ_k} are of the form $\sum_{i=0}^{n} a_i \epsilon^i$ for $a_i \in \mathbb{F}$. Then, the *border rank* $\underline{R}(t)$ is defined as $\min_{h \geq 0} R_h(t)$.

Once we have the border rank of a tensor, we can use it to argue about the rank of the original tensor.

Theorem 9. *If you have a tensor t with $R_h(t) \leq r$, then $R(t) \leq \binom{h+2}{2} r$.*

Theorem 9 isn't useful for our old example, but it is helpful for some examples where t is very large.

Based on this, we can prove that if $\underline{R}(\langle K, M, N\rangle) \leq r$, then $\omega \leq \frac{3 \log r}{\log(KMN)}$. So, we can use the border rank instead of rank for the purposes of bounding ω.

Our goal is to bound the border rank of $\langle 2, 2, 3\rangle$. we know that its rank is 11 already, but we can show that its border rank is at most 10, which gives us $\omega \leq 2.78$.

Proof. Suppose we have the following.

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix}$$

If we only care about the top left 3 entries of z, and consider those to be a tensor t, then its rank is ≤ 6. We want to show that the border rank is ≤ 5.

$$\begin{aligned} P1 &= (x_{12} + \epsilon x_{22}) y_{21} \\ P2 &= x_{11}(y_{11} + \epsilon \cdot y_{12}) \\ P3 &= x_{12}(y_{12} + y_{21} + \epsilon y_{22}) \\ P4 &= (x_{11} + x_{12} + \epsilon x_{21}) y_{11} \\ P5 &= (x_{12} + \epsilon x_{21})(y_{11} + \epsilon y_{22}) \\ \epsilon P1 + \epsilon P2 &= \epsilon \cdot z_{11} + O(\epsilon^2) \\ P2 - P4 + P5 &= \epsilon \cdot z_{12} + O(\epsilon^2) \\ P1 - P3 + P5 &= \epsilon \cdot z_{21} + O(\epsilon^2) \end{aligned}$$

The tensor $\langle 2, 2, 3\rangle$ is equivalent to two copies of t. This gives us that $\underline{R}(\langle 2, 2, 3\rangle)$ is upper bounded by $2 \cdot \underline{R}(t) \leq 10$. □

This proof gives us $\omega \leq 2.78$! It was proposed by Bini et al. in 1979.

5 Next time

Next week, we will begin studying the Coppersmith-Winograd algorithm.