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

Lecture on 03/31/2022

Lecturer: Xiaorui Sun *Scribe: Sai Goutham Reddy Inturi*

1 Matrix Multiplication

To understand how different algorithms work, we put matrix multiplication through a tensor.
Let us consider a 3-tensor : $x_i \; y_j \; z_k$
The matrix multiplication tensor would be :

$$t = \sum t_{ijk} x_i y_j z_k$$

$t_{ij',jk',ki'} = 1$, if i=i',j=j',k=k'

else $t_{ij',jk',ki'} = 0$

Rank of Tensor: Rank of a tensor R(t) for tensor t is defined as the minimum number of rank 1 tensors whose sum is equal to t.
Traid (rank 1 tensor) : $(\sum u_i x_i)(\sum v_j y_j)(\sum w_k z_k)$
We have the freedom to select any coefficients

Observation: If $R(\langle K,M,N \rangle) \leq r$, then $\omega \leq \frac{3logr}{log(NMK)}$
To prove the observation we have to assume the matrix is symmetric
$R(\langle K,M,N \rangle) = R(\langle K,N,M \rangle) = R(\langle N,K,M \rangle)$
Our goal is to find an upper bound on $R(\langle n,n,n \rangle)$

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

Observation: $R(t \otimes t') \leq R(t).R(t')$
if we have T = N x M x K, then $\langle T,T,T \rangle = \langle K,N,M \rangle \otimes \langle N,M,K \rangle \otimes \langle M,K,N \rangle$

Strassen proved that $R(\langle 2,2,2 \rangle) \leq 7$

2 DIRECT SUM

Direct Sum: Given $t_{ijk} \in \mathbb{F}^{K \times M \times N}$ and $t'_{i'j'k'} \in \mathbb{F}^{K' \times M' \times N'}$,we have
$t \oplus t' = t_{ijk}$ if $i \leq K, j \leq M, k \leq N$
$t \oplus t' = t'_{i-K,j-M,k-N}$ if $i > K, j > M, k > N$
$t \oplus t' = 0$ otherwise
The dimension of $t \oplus t' = (K+K') \times (M+M') \times (N+N')$

Observation: $R(t \oplus t') \leq R(t).R(t')$
As per Strassen's observation if $R(\langle 2,2,2 \rangle) = 7$ then $\omega \leq 2.81$
It was also shown that $R(\langle 2,2,3 \rangle) = 11$ and $14 \leq R(\langle 2,3,3 \rangle) \leq 15$ which are both not better than $\langle 2,2,2 \rangle$ tensor.
We have $19 \leq R(\langle 3,3,3 \rangle) \leq 23$ and if $R(\langle 3,3,3 \rangle) \leq 21$ then $\omega \leq 2.79$
Pan showed that if $R(\langle 70,70,70 \rangle) \leq 143640$ then $\omega < 2.8$

3 APPROXIMATE TENSOR

Lets say we have an infinite set of matrices $M_1, M_2, ...$ where $j \to \infty$ and $M_j \to M$
Suppose $r(M_j) \leq r$, then $r(M) \leq r$
Look at any $(r+1)x(r+1)$ submatrix P_j of M_f
Here the determinant of $P_j = 0$ hence determinant of $P = 0$

Suppose we have a tensor t with rank 3 such as $\{x_0,x_1\},\{y_0,y_1\},\{z_0,z_1\}$ then,

$$t = x_0y_0z_0 + x_1y_0z_1 + x_0y_1z_1$$

A tensor with parameter (ϵ)

$$t(\epsilon) = (x_0 + \epsilon x_1) . (y_0 + \epsilon y_1) . 1/\epsilon . z_1 + x_0 y_0(z_0 - z_1/\epsilon)$$
$$t(\epsilon) = x_0y_0(1/\epsilon)z_1 + x_0y_1z_1 + x_1y_0z_1 + \epsilon x_1y_1z_1 + x_0y_0z_0 - 1/\epsilon x_0y_0z_1$$
$$t(\epsilon) = x_0y_1z_1 + x_1y_0z_1 + \epsilon x_1y_1z_1 + x_0y_0z_0$$

Here, the rank of $t(\epsilon)$ $R(t(\epsilon)) = 2$ when $\epsilon \to 0$ and $t(\epsilon) \to t$

In approximation of a tensor we assume the coefficients of a tensor $\mathbb{F}$. We extend this with ϵ. So, the approximation would be $\mathbb{F}[\epsilon]$.

4 BORDER RANK

Border Rank: Given a tensor t and an integer h, the border rank of the tensor $R_h(t)$ be the smallest integer l such that:

$$t(\epsilon) = \sum_{\lambda=1}^{l} (\sum U_{\lambda i}\ X_i)(\sum V_{\lambda j}\ y_j)(\sum W_{\lambda k}\ y_k)$$

$$t(\epsilon) = \epsilon^h\ t + O\ (\epsilon^{h+1})$$

Where $U_{\lambda i}$, $V_{\lambda j}$ and $W_{\lambda k}$ are of $\sum_{i=1}^{n} a_i \epsilon^{\ i}$ for $a_i \in \mathbb{F}$.

The border rank of the tensor is defined as $R(t) = \min_{h \geq 0} R_h(t)$. This will hold for the previous example:

R(t)=3
if h=1
$R_1(t) \leq 2 \Rightarrow R(t) \leq 2$

Theorem: Given a tensor t where $R_h(t) \leq r$, then $R(t) \leq \binom{h+2}{2} r$.

This will not hold for the previous example but will hold for other tensor where t is huge.

So, let us use border rank instead of rank to try and prove the observation if $R(\langle K,M,N \rangle) \leq r$, then $\omega \leq \frac{3logr}{log(KMN)}$

Goal is bound the border rank of $R(\langle\ 2,2,3\ \rangle)$. From the above we have already shown that the rank of $R(\langle\ 2,2,3\ \rangle) = 11$. Hence we can say that the border rank of $R(\langle\ 2,2,3\ \rangle) \leq 10$. Which signifies that $\omega \leq 2.78$.

Let us consider the below matrix multiplication:

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

If we dot not consider z_{22} and only consider the other three entries of z to be a tensor then its rank would be R(t)=6. Hence the border rank of the tensor would be R(t) $\leq$ 5.

P1 = $(x_{12} + \epsilon x_{22})y_{21}$

P2 = $x_{11}(y_{11} + \epsilon y_{12})$

P3 = $x_{12}(y_{12} + y_{21} + \epsilon y_{22})$

P4 = $(x_{11} + x_{12} + \epsilon x_{21})y_{11}$

P5 = $(x_{12} + \epsilon x_{21})(y_{11} + \epsilon y_{22})$

$\epsilon P1 + \epsilon P2 = \epsilon . z_{11} + O(\epsilon^2)$

P2 - P4 + P5 = $\epsilon . z_{12} + O(\epsilon^2)$

P1 - P3 + P5 = $\epsilon . z_{21} + O(\epsilon^2)$

The tensor $\langle$ 2,2,3 $\rangle$ is equivalent to two copies of t. This proves that the rank of the tensor R($\langle$ 2,2,3 $\rangle$) is upper bounded by 2 and the R(t) $\leq$ 10 which gives $\omega \leq$ 2.78.

5 NEXT CLASS

In the next class, we will discuss about Coppersmith-Winograd algorithm.

————————————THE END————————————