Today we will continue discussing matrix multiplication algorithms in the arithmetic circuit model. As the previous lecture, these notes are based on lecture notes by Markus Bläser [1]. The figures in these notes are from Bläser's notes as well. In the lecture video you'll see slightly different figures.

1 Recap

Recall some useful definitions and theorems from last lecture:

- In a *bilinear program*, we are given a length n vector x and a length m vector y and we want to compute a vector z such that $z_k = \sum_{ij} t_{ijk} x_i y_j$. The coefficients t_{ijk} form an order 3 tensor t.
- The *rank* $R(t)$ of a tensor t is the minimum number of rank 1 tensors that sum to t. That is, $t = \sum_{\ell=1}^{r} a_\ell \otimes b_\ell \otimes c_\ell$, where a_ℓ, b_ℓ, and c_ℓ are all vectors and the (i,j,k) entry of $a_\ell \otimes b_\ell \otimes c_\ell$ is $a_\ell[i] \cdot b_\ell[j] \cdot c_\ell[k]$. This is a natural extension of the outer product of vectors.
- In the *arithmetic circuit model*, also known as the straight-line program (SLP) model, the goal is to find the shortest SLP for a problem. The *Ostrowski measure* $C^{*,/}$ of an SLP is the combined number of multiplication and division operations in the SLP, i.e. all other operations are free. The *Ostrowski cost* of a bilinear problem is the Ostrowski measure of the minimum cost SLP for this bilinear problem.
- Strassen's vermeidung von divisionen theorem, combined with a lemma proved last lecture implies that for infinite fields, for any tensor t defined by a bilinear problem, $C^{*,/}(t) \leq R(t) \leq 2C^{*,/}(t)$.
- We denote by $\langle K, M, N\rangle$ the matrix multiplication tensor for multiplying a $k \times m$ matrix and an $m \times n$ matrix. $(\langle K, M, N\rangle)_{i'j,j'k,k'i} = \delta_{ii'}\delta_{jj'}\delta_{kk'}$ where $\delta_{kk'}$ is 1 whenever $k = k'$ (and analogously for i, i' and j, j').

Today we'll see some properties of the matrix multiplication tensor.

2 Omega and rank

In this section we will show how ω and rank are related. We will give three definitions of ω and show that they are all the same. We begin with the usual definition of ω.

Definition 2.1. $\omega = \inf\{p \mid C(\langle n, n, n\rangle) \leq O(n^p)\}$ where C(t) is minimum length SLP for $\langle n, n, n\rangle$ (i.e. $n \times n$ matrix multiplication).

Now, we introduce two alternate definitions of ω.

Definition 2.2. $\overline{\omega} = \inf\{p \mid R(\langle n, n, n\rangle) \leq O(n^p)\}$.

Definition 2.3. $\overline{\omega}' = \inf\{p \mid C^{*,/}(\langle n, n, n\rangle) \leq O(n^p)\}$.

Note that Strassen's Vermeidung von Divisionen theorem, together with $C^{*,/}(\langle n, n, n\rangle) \leq R(\langle n, n, n\rangle) \leq 2C^{*,/}(\langle n, n, n\rangle)$, implies that $\overline{\omega} = \overline{\omega}'$. We will show that $\omega = \overline{\omega}$.

Theorem 2.1. *For infinite fields,* $\omega = \overline{\omega}$.

Proof. By definition, $C^{*,/}(\langle n,n,n\rangle) \le C(\langle n,n,n\rangle)$, so we have that $\overline{\omega} \le \omega$. It remains to show that $\overline{\omega} \ge \omega$.

By definition, $\overline{\omega} < p$ means that for every ϵ, there exists m_0 such that for all $m > m_0$, $R(\langle m,m,m\rangle) \le m^{p+\epsilon}$. Fix a constant $\epsilon > 0$ and pick a constant m such that $R(\langle m,m,m\rangle) \le m^{p+\epsilon}$. Let ℓ be the number of summations plus scalar multiplications in the rank expression for $\langle m,m,m\rangle$.

Now, consider multiplying $m^i \times m^i$ matrices. Block each $m^i \times m^i$ into m^2 blocks where each block has dimension $m^{i-1} \times m^{i-1}$. Let $A(i)$ be the minimum size of an SLP for $\langle m^i, m^i, m^i\rangle$. That is, $A(i) = C(\langle m^i, m^i, m^i\rangle)$. Let $r = m^{p+\epsilon}$. Notice $r > m^2$.

Using our blocking, we can devise a recursive expression for $A(i)$:

$$A(i) \le rA(i-1) + \ell m^{2(i-1)}.$$

Expanding this recurrence, we get

$$\begin{aligned}
A(i) &\le r^{i-1}A(0) + \sum_{j=0}^{i-1} r^j \ell m^{2(i-j-1)} \\
&\le r^{i-1}A(0) + \ell m^{2(i-1)} \sum_{j=0}^{i-1} (r/m^2)^j \\
&\le r^{i-1}A(0) + \ell m^{2(i-1)} \frac{(r/m^2)^i - 1}{r/m^2 - 1} \\
&\le r^{i-1}A(0) + \ell \frac{m^{2(i-1)}(r^i - m^{2i})}{m^{2i-2}(r-m^2)} \\
&\le r^i \Big(\frac{A(0)}{r} + \frac{\ell}{r-m^2} \Big) \\
&\le O(r^i)
\end{aligned}$$

where the last inequality is since $r = m^{p+\epsilon} > m^2$.

Thus, we have

$$\begin{aligned}
C(< n,n,n >) &\le C(\langle m^{\lceil \log_m n \rceil}, m^{\lceil \log_m n \rceil}, m^{\lceil \log_m n \rceil}\rangle) \\
&\le A(\lceil \log_m n \rceil) \\
&\le O(r^{\lceil \log_m n \rceil}) \qquad \text{by the above string of inequalities} \\
&\le O(n^{\log_m r}) \\
&\le O(n^{p+\epsilon}).
\end{aligned}$$

Thus, $\omega \le p + \epsilon$ for all ϵ, and since ω is an infimum, $\omega \le \overline{\omega}$. □

From now on, we will only consider ω in terms of the rank definition rather than counting the number of operations in an SLP.

3 Permutations of tensors

We will consider two types of permutations of tensors that preserve the rank.

3.1 Permuting the indices

Consider a tensor t that lies in the field $\mathcal{K}^{K \times M \times N}$. That is, each entry of t is denoted t_{ijk} where $i \in [1,\dots,K]$, $j \in [1,\dots,M]$, and $k \in [1,\dots,N]$.

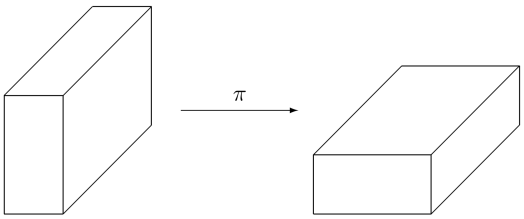

Figure 1: Permuting the indices (figure from [1])

Consider $\pi \in S_3$, a permutation over 3 elements that acts on the indices i, j, and k. For example, if $\pi = (1\ 2\ 3)$, then $(\pi t)_{jki} = t_{ijk}$. See Figure 1.

Suppose t has rank decomposition $t = \sum_{\ell=1}^{r} a_{\ell_1} \otimes a_{\ell_2} \otimes a_{\ell_3}$. Then, $\pi t = \sum_{\ell=1}^{r} a_{\ell_{\pi^{-1}(1)}} \otimes a_{\ell_{\pi^{-1}(2)}} \otimes a_{\ell_{\pi^{-1}(3)}}$.

One can check that the rank decomposition of πt is well-defined in the sense that if you write two different rank decompositions for t and then write πt for both, you will get the same thing (we won't prove this).

The rank expression for πt implies the following claim.

Claim 1. *For all t and all $\pi \in S_3$, $R(t) = R(\pi t)$.*

Now we will derive a consequence of Claim 1 for the matrix multiplication tensor. Let $t = \langle K, M, N \rangle$. Recall that $(\langle K, M, N \rangle)_{i'j,j'k,k'i} = \delta_{ii'}\delta_{jj'}\delta_{kk'}$ where $\delta_{kk'}$ is 1 whenever $k = k'$ (and analogously for i, i' and j, j').

Let $\pi = (1\ 2\ 3)$. Then $(\pi t)_{k'i,i'j,j'k} = \delta_{ii'}\delta_{jj'}\delta_{kk'}$. Thus, (πt) is the matrix multiplication tensor $\langle N, K, M \rangle$. Then, by Claim 1, $R(\langle K, M, N \rangle) = R(\langle N, K, M \rangle)$.

By the same argument, $(\pi^2 t) = \langle M, N, K \rangle$. Thus, by Claim 1 we have $R(\langle K, M, N \rangle) = R(\langle N, K, M \rangle) = R(\langle M, N, K \rangle))$.

3.2 Permuting the slices

Again, consider a tensor t that lies in the field $\mathcal{K}^{K \times M \times N}$. Let $\sigma \in S_K$ be a permutation of K elements. Then, the permuted tensor $t' \in \mathcal{K}^{K \times M \times N}$ is defined as $t'_{ijk} = t_{\sigma^{-1}(i)jk}$. Analogously $\sigma \in S_M$ could permute the j indices and $\sigma \in S_N$ could permute the k indices. See Figure 2.

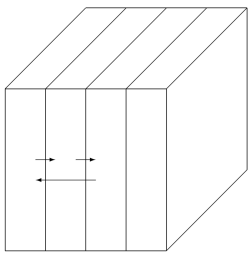

Figure 2: Permuting the slices (figure from [1])

More generally, we can define homomorphisms for each of the three indices. Let

$$A : \mathcal{K}^K \to \mathcal{K}^{K'}, \qquad B : \mathcal{K}^M \to \mathcal{K}^{M'}, \qquad C : \mathcal{K}^N \to \mathcal{K}^{N'}.$$

Now, we can define the tensor $(A \otimes B \otimes C)t$. Suppose t has rank decomposition $t = \sum_{\ell=1}^{r} a_\ell \otimes b_\ell \otimes c_\ell$. Then we define $(A \otimes B \otimes C)t = \sum_{\ell=1}^{r} A(a_\ell) \otimes B(b_\ell) \otimes C(c_\ell) \in \mathcal{K}^{K' \times M' \times N'}$.

Again, one can check that the rank decomposition of $(A \otimes B \otimes C)t$ is well-defined in the sense that if you write two different rank decompositions for t and then write $(A \otimes B \otimes C)t$ for both, you will get the same thing (we won't prove this).

The rank expression for $(A \otimes B \otimes C)t$ implies the following claim.

Claim 2. *For all t and all homomorphisms A, B, C, $R((A \otimes B \otimes C)t) \leq R(t)$, with equality if A, B, C are isomorphisms.*

Now we will derive a consequence of Claims 1 and 2 for the matrix multiplication tensor.

Let $t = \langle K, M, N \rangle$. First we will permute the indices and then we will permute the slices. Let $\pi = (1\ 2)(3)$. Let $t' = \pi t$. Then $t'_{j'k,i'j,k'i} = \delta_{ii'}\delta_{jj'}\delta_{kk'}$. By Claim 1, $R(\langle K, M, N \rangle) = R(t')$.

Now we permute the slices by swapping j' with k, i' with j, and k' with i to get a new tensor t'' where $t''_{kj',ji',ik'} = \delta_{ii'}\delta_{jj'}\delta_{kk'}$. Thus, t'' is the matrix multiplication tensor $\langle N, M, K \rangle$.

Since these permutations are isomorphisms, by Claim 2 $R(t'') = R(t')$, so $R(\langle N, M, K \rangle) = R(\langle K, M, N \rangle)$.

We can apply the same argument to show that the rank of the matrix multiplication tensor for *all* permutations of K, M, and N are equal.

4 Tensor sums and products

4.1 Direct sum

Definition 4.1. Given two tensors $t \in \mathcal{K}^{K \times M \times N}$ and $t' \in \mathcal{K}^{K' \times M' \times N'}$, the *direct sum* $t \oplus t' \in \mathcal{K}^{(K+K') \times (M+M') \times (N+N')}$ is defined as

$$(t \oplus t')_{i,j,k} = \begin{cases} t_{i,j,k} & i \leq K, j \leq M, k \leq N \\ t'_{i-K,j-M,k-N} & i > K, j > M, k > N \\ 0 & o.w. \end{cases}$$

See Figure 3.

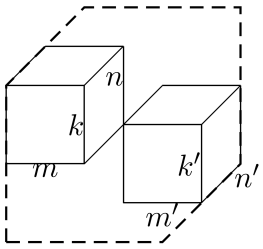

Figure 3: Direct sum of two tensors (figure from [1])

Claim 3. *For all t, t', $R(t \oplus t') \leq R(t) + R(t')$.*

Proof. Let $t = \sum_{\ell=1}^{r} a_\ell \otimes b_\ell \otimes c_\ell$ and $t' = \sum_{\ell=1}^{s} a'_\ell \otimes b'_\ell \otimes c'_\ell$.

For all $\ell \in [r]$, let $\hat{a}_\ell$ be a vector of length $K + K'$ which is a_ℓ in the first K indices and 0 in the last K' indices. Similarly, for all $\ell \in [s]$, let $\hat{a}'_\ell$ be a vector of length $K + K'$ which is 0 in the first K indices and a'_ℓ in the last K' indices. Define $\hat{b}_\ell$, $\hat{b}'_\ell$, $\hat{c}_\ell$, and $\hat{c}'_\ell$ analogously.

By definition, we have $t \oplus t' = \sum_{\ell=1}^{r} \hat{a}_\ell \otimes \hat{b}_\ell \otimes \hat{c}_\ell + \sum_{\ell'=1}^{s} \hat{a}'_{\ell'} \otimes \hat{b}'_{\ell'} \otimes \hat{c}'_{\ell'}$. This completes the proof. □

Strassen made the "wild" conjecture that this claim holds with equality, but nobody has yet found a counterexample.

Conjecture 1 (Strassen's additivity conjecture)**.** *For every t, t' $R(t \oplus t') = R(t) + R(t')$.*

4.2 Kronecker product

Definition 4.2. Let $t \in \mathcal{K}^{K \times M \times N}$ and $t' \in \mathcal{K}^{K' \times M' \times N'}$. The Kronecker product $(t \otimes t') \in \mathcal{K}^{(KK') \times (MM') \times (NN')}$ is defined as

$$(t \otimes t')_{ii',jj',kk'} = t_{ijk} \cdot t'_{i'j'k'}.$$

See Figure 4.

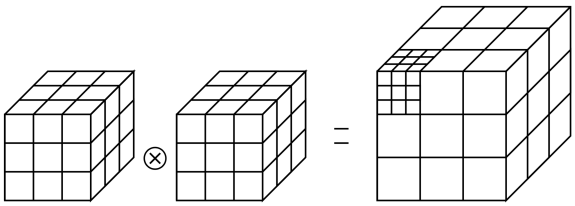

Figure 4: Kronecker product of two tensors (figure from [1])

Claim 4. *For all t, t', $R(t \otimes t') \leq R(t) \cdot R(t')$.*

Proof. Let $t = \sum_{\ell=1}^{r} a_\ell \otimes b_\ell \otimes c_\ell$ and $t' = \sum_{\ell=1}^{s} a'_{\ell'} \otimes b'_{\ell'} \otimes c'_{\ell'}$.

For all $\ell \in [r]$ and $\ell' \in [s]$, define $(\hat{a}_{\ell\ell'})_{ii'} = a_{\ell_i} \cdot a'_{\ell'_{i'}}$. Define $\hat{b}_{\ell\ell'}$ and $\hat{c}_{\ell\ell'}$ analogously.

By definition, we have $t \otimes t' = \sum_{\ell'=1}^{s} \sum_{\ell=1}^{r} \hat{a}_{\ell\ell'} \otimes \hat{b}_{\ell\ell'} \otimes \hat{c}_{\ell\ell'}$. This completes the proof. □

Now, we will apply the Kronecker product to the matrix multiplication tensor. For $i, i' \in [K]$, $j, j' \in [M]$, $k, k' \in [N]$, $p, p' \in [K']$, $q, q' \in [M']$, and $s, s' \in [N']$, we have

$$\begin{aligned}(\langle K, M, N \rangle) \otimes (\langle K', M', N' \rangle)_{ij'pq',jk'qs',ki'sp'} &= \delta_{ii'}\delta_{jj'}\delta_{kk'}\delta_{pp'}\delta_{qq'}\delta_{ss'} \\ &= \delta_{(ip),(i'p')}\delta_{(jq),(j'q')}\delta_{(ks),(k's')} \\ &= \langle KK', MM', NN' \rangle.\end{aligned}$$

Claim 5. *If $R(\langle K, M, N \rangle) \leq r$, then $\omega \leq 3 \log r / \log(KMN)$.*

Proof. Let $T = KMN$. Note that $\langle T, T, T \rangle = \langle K, M, N \rangle \otimes \langle M, N, K \rangle \otimes \langle N, K, M \rangle$. Thus, by Claim 4, $R(\langle T, T, T \rangle) \leq R(\langle K, M, N \rangle) \cdot R(\langle M, N, K \rangle) \cdot R(\langle N, K, M \rangle)$, which is at most r^3 since we proved (in Section 3) that $R(\langle K, M, N \rangle) = R(\langle M, N, K \rangle) = R(\langle N, K, M \rangle)$.

Then, by definition, $\omega \leq \log(r^3) / \log T = 3 \log r / \log(KMN)$. □

Claim 5 shows that to get a bound on ω it suffices to get a bound on the rank of a rectangular matrix multiplication tensor of constant dimension. Next we will summarize what is currently known about the rank of rectangular matrix multiplication tensors of constant dimension.

4.2.1 Known bounds on the rank of rectangular matrix multiplication tensors of constant dimension

- $R(\langle 2, 2, 2 \rangle) = 7$. This was shown by Strassen and implies that $\omega \leq 2.81$. Note that the equals sign means that both upper and lower bounds are known.
- $R(\langle 2, 2, 3 \rangle) = 11$. This does not imply a better bound on ω than Strassen.
- $R(\langle 2, 3, 3 \rangle) \in \{14, 15\}$.

- $19 \leq R(\langle 3,3,3\rangle) \leq 23$. Also, it is known that if $R(\langle 3,3,3\rangle) \leq 21$ then $\omega \leq 2.79$. Even though the current best bound on ω is much better than 2.79, this would be a nice result because it would imply a *simple* way to get a better bound on ω than Strassen. People have tried hard to show using software that $R(\langle 3,3,3\rangle) < 23$, but they have not succeeded since the search space is so big.
- $R(\langle 70,70,70\rangle) \leq 143,640$. This was shown by Pan in 1980 and implies that $\omega < 2.8$.

5 Preview of next lecture

Here's a motivating example for what we'll discuss next lecture. Suppose we have a sequence of matrices A_j so that as $j \to \infty$, $A_j \to A$. Suppose that for every j, $R(A_j) \leq r$. Then, one can show that $R(A) \leq r$.

Although this is true for matrices, it is not true for tensors, which seems strange. We will exploit this fact in the next lecture. The following is a counterexample showing that the above property is not true for tensors.

Let $a = (a_0, a_1)$, $b = (b_0, b_1)$, $c = (c_0, c_1)$ and suppose

$$c_0 = a_0 b_0$$

$$c_1 = a_1 b_0 + a_0 b_1.$$

Let t be the tensor for the bilinear program c. $R(t) = 3$.

Now consider a sequence of tensors $t(\epsilon)$ so that as $\epsilon \to 0$, $t(\epsilon) \to t$. Specifically, let

$$t(\epsilon) = (1, \epsilon) \otimes (1, \epsilon) \otimes (0, 1/\epsilon) + (1, 0) \otimes (1, 0) \otimes (1, -1/\epsilon).$$

By definition, for all ϵ, $R(t(\epsilon)) = 2$.

Now we will show that as $\epsilon \to 0$, $t(\epsilon) \to t$. Let $p_0 = (a_0 + \epsilon a_1) \cdot (b_0 + \epsilon b_1)$ and let $p_1 = a_0 \cdot b_0$. Then, $c_0 = p_1$ and $c_1 = (p_0 - p_1)/\epsilon = a_0 b_1 + a_1 b_0 + \epsilon a_1 b_1$. As $\epsilon \to 0$, the term $\epsilon a_1 b_1 \to 0$. Thus, as $\epsilon \to 0$, we have $c_0 = a_0 b_0$ and $c_1 = a_1 b_0 + a_0 b_1$, which is identical to the definition of t. This concludes the counterexample.

Next lecture, we will see a sequence of tensors that approach the matrix multiplication tensor as $\epsilon \to 0$. Then, we will show that bounding the rank of this sequence still allows us to get a bound on ω. Furthermore, this sequence of tensors will have a lower rank than the tensor that the sequence approaches. This allows us to get a better bound on ω than we would get just by looking at the rank of the individual matrix multiplication tensor that the sequence approaches.

References

[1] Bläser, Markus, *Complexity of Bilinear Problems.*, Lecture Notes (2009).