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

Why is this? Fix any j. Consider all $r+1 \times r+1$ minors of A_j; they have to vanish because the rank is r. The functions representing these minors are continuous, and since these functions vanish for all j, then in the limit they need to vanish as well. But at the limit these are the $r+1 \times r+1$ minors of A, and so A must have rank at most r as well.

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

Let us consider the problem of computing two bilinear forms on inputs $a = (a_0, a_1)$, $b = (b_0, b_1)$, $c = (c_0, c_1)$, where the two forms are

$$g_0 = a_0 b_0$$

$$g_1 = a_1 b_0 + a_0 b_1.$$

In trilinear notation we have

$$a_0 b_0 c_0 + a_1 b_0 c_1 + a_0 b_1 c_1.$$

Let t be the tensor for the set of bilinear forms c. Its rank is $R(t) = 3$. The upper bound is trivial, and the lower bound is not hard to prove by something called the substitution method.

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's consider the bilinear forms that the tensor $t(\epsilon)$ defines. It says, first compute the products $p_0 = (a_0 + \epsilon a_1) \cdot (b_0 + \epsilon b_1)$ and $p_1 = a_0 \cdot b_0$.

Then, the forms to be computed are $f_0 = p_1$ and $f_1 = (p_0 - p_1)/\epsilon = a_0 b_1 + a_1 b_0 + \epsilon a_1 b_1$.

(In trilinear notation we have that the tensor $t(\epsilon)$ corresponds to $a_0 b_0 c_0 + a_0 b_1 c_1 + a_1 b_0 c_1 + \epsilon a_1 b_1 c_1$.)

As $\epsilon \to 0$, the term $\epsilon a_1 b_1 \to 0$. Thus, as $\epsilon \to 0$, we have $f_0 = a_0 b_0$ and $f_1 = a_1 b_0 + a_0 b_1$, which is identical to the definition of the original tensor t.

(In trilinear notation we have that as ϵ goes to 0, $\epsilon a_1 b_1 c_1$ goes to 0 and the tensor goes to $a_0 b_0 c_0 + a_0 b_1 c_1 + a_1 b_0 c_1$ which is t. We will later say that $t(\varepsilon) = t + O(\varepsilon)$.)

This concludes the counterexample.

Thus we have a sequence of tensors whose ranks are smaller than the rank of the limit of this sequence. Today, we will exploit this fact to get better bounds on ω. As before, the notes are adapted from notes by Bläser [1].

1 Border Rank

Let $\mathcal{K}$ be a field. The polynomial ring $\mathcal{K}[\varepsilon]$ consists of polynomials of the form $\sum_{i=0}^{m} a_i \varepsilon^i$ for some integer $m \ge 0$ and $a_0, \ldots, a_m \in \mathcal{K}$.

Let $t \in \mathcal{K}^{K \times M \times N}$ be a tensor. Let $h \ge 0$ be a nonnegative integer. We define $R_h(t)$ to be the minimum integer r such that there exists $u_\ell \in \mathcal{K}[\varepsilon]^K, v_\ell \in \mathcal{K}[\varepsilon]^M, w_\ell \in \mathcal{K}[\varepsilon]^N$ for $1 \le \ell \le r$ so that

$$\sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell = \varepsilon^h \cdot t + O(\varepsilon^{h+1}). \tag{1}$$

Above, $O(\varepsilon^{h+1})$ contains all terms of the tensor that have a power of ε that is at least $h+1$.

Informally, we can think that $\frac{1}{\varepsilon^h} \cdot \sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$ approaches t when ϵ approaches 0. If any of u_ℓ, v_ℓ, w_ℓ has an entry that has degree greater than h, we can eliminate all terms with degree greater than h so that Equation (1) still holds. Thus, for the rank expression of R_h, we can assume the entries of u_ℓ, v_ℓ, w_ℓ are degree $\leq h$ polynomials in ε.

Using $R_h(t)$, we can define border rank as follows.

Definition 1.1 (border rank)**.** The border rank of a tensor t, $\underline{R}(t)$, is defined as $\min_{h\geq 0} R_h(t)$.

As an observation, $R(t) = R_0(t) \geq R_1(t) \geq \cdots \geq \underline{R}(t)$.

In the last lecture, we showed many properties for the rank function $R(\cdot)$. It turns out many of those properties still hold for $R_h(\cdot)$. In the following lemma, we list those properties that will be used later.

Lemma 1.1. *The followings are true for any tensors* $t \in \mathcal{K}^{K\times M\times N}, t' \in \mathcal{K}^{K'\times M'\times N'}$.

(1) For any $h \geq 0$ *and any* $\pi \in S_3$, $R_h(t) = R_h(\pi t)$.

(2) For any $h, h' \geq 0$, $R_{\max(h,h')}(t \oplus t') \leq R_h(t) + R_{h'}(t')$.

(3) For any $h, h' \geq 0$, $R_{h+h'}(t \otimes t') \leq R_h(t) \cdot R_{h'}(t')$.

Proof. (1) We first show $R_h(\pi t) \leq R_h(t)$. Suppose $R_h(t) = r$, and suppose

$$\sum_{\ell=1}^{r} u_{\ell,1} \otimes u_{\ell,2} \otimes u_{\ell,3} = \varepsilon^h t + O(\varepsilon^{h+1}).$$

Then it is not hard to verify that

$$\sum_{\ell=1}^{r} u_{\ell,\pi^{-1}(1)} \otimes u_{\ell,\pi^{-1}(2)} \otimes u_{\ell,\pi^{-1}(3)} = \varepsilon^h(\pi t) + O(\varepsilon^{h+1}),$$

which implies $R_h(\pi t) \leq r$.

We can show $R_h(\pi t) \geq R_h(t)$ analogously.

(2) Without loss of generality, assume $h \geq h'$. Let $R_h(t) = r$ and $R_{h'}(t') = s$. Suppose

$$\sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell = \varepsilon^h t + O(\varepsilon^{h+1}), \tag{2}$$

and

$$\sum_{\ell=1}^{s} u'_\ell \otimes v'_\ell \otimes w'_\ell = \varepsilon^{h'} t + O(\varepsilon^{h'+1}). \tag{3}$$

We can multiply both sides of Equation (3) by $\varepsilon^{h-h'}$ to get

$$\sum_{\ell=1}^{s} (u'_\ell \varepsilon^{h-h'}) \otimes v'_\ell \otimes w'_\ell = \varepsilon^h t + O(\varepsilon^{h+1}).$$

We can define $\widehat{u}_\ell$ to be u_ℓ but padded with K' zeros at the end. We can similarly define $\widehat{v}_\ell$ and $\widehat{w}_\ell$. For $\widehat{u}'_\ell, \widehat{v}'_\ell, \widehat{w}'_\ell$, we can define them to be $u'_\ell, v'_\ell, w'_\ell$ padded with K, M and N zeros at the beginning, respectively. Then we have

$$\sum_{\ell=1}^{r} \widehat{u}_\ell \otimes \widehat{v}_\ell \otimes \widehat{w}_\ell + \sum_{\ell=1}^{s} (\widehat{u}'_\ell \varepsilon^{h'-h}) \otimes \widehat{v}'_\ell \otimes \widehat{w}'_\ell = \varepsilon^h (t \oplus t') + O(\varepsilon^{h+1}),$$

which implies $R_h(t \oplus t') \leq r + s = R_h(t) + R_{h'}(t')$.

(3) Let $R_h(t) = r$ and $R_{h'}(t') = s$. Suppose

$$\sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell = \varepsilon^h t + O(\varepsilon^{h+1}) \quad \text{and} \quad \sum_{\ell=1}^{s} u'_\ell \otimes v'_\ell \otimes w'_\ell = \varepsilon^{h'} t + O(\varepsilon^{h'+1}).$$

Then

$$\sum_{\ell=1}^{r} \sum_{\ell'=1}^{s} (u_\ell \otimes u'_{\ell'}) \otimes (v_\ell \otimes v'_{\ell'}) \otimes (w_\ell \otimes w'_{\ell'}) = \varepsilon^{h+h'} (t \otimes t') + O(\varepsilon^{h+h'+1}).$$

□

The border rank is nicely related to the rank of a tensor with the following lemma.

Lemma 1.2. *For any tensor $t \in \mathcal{K}^{K \times M \times N}$, If $R_h(t) \le r$, then $R(t) \le c_h \cdot r$, where $c_h \le \binom{h+2}{2}$.*

When the field $\mathcal{K}$ is infinite, we can actually improve the bound on c_h to $2h + 1$. However, in the later proofs, what we really need is the fact that c_h is bounded by a polynomial in h. Thus, we won't prove the tighter upper bound on c_h.

Proof of Lemma 1.2. Let $\sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell = \varepsilon^h \cdot t + O(\varepsilon^{h+1})$. Note that the entries of u_ℓ, v_ℓ, w_ℓ are polynomials of degree bounded by h. Thus, we can rewrite u_ℓ as $\sum_{i=0}^{h} u_{\ell,i} \cdot \varepsilon^i$, where $u_{\ell,i} \in \mathcal{K}^K$. Similarly, $v_\ell = \sum_{j=0}^{h} v_{\ell,j} \cdot \varepsilon^j$ and $w_\ell = \sum_{k=0}^{h} w_{\ell,k} \cdot \varepsilon^k$.

Therefore, we can express t as

$$\sum_{i+j+k=h} \sum_{\ell=1}^{r} u_{\ell,i} \otimes v_{\ell,j} \otimes w_{\ell,k},$$

which has at most $\binom{h+2}{2} \cdot r$ terms. Thus, $R(t) \le \binom{h+2}{2} \cdot r$.

□

In the last lecture, we showed that if the rank of a matrix multiplication tensor $\langle K, M, N \rangle$ is r, then $\omega \le \frac{3 \log r}{\log(KMN)}$. A similar result also holds for border rank. Since border rank can be smaller than rank, using border rank can give better bounds on ω.

Theorem 1.1. *If $\underline{R}(\langle K, M, N \rangle) \le r$, then $\omega \le \frac{3 \log r}{\log(KMN)}$*

Proof. Let $T = KMN$. By the results from last lecture, we know that

$$\langle T, T, T \rangle = \langle K, M, N \rangle \otimes \langle M, N, K \rangle \otimes \langle N, K, M \rangle.$$

Let h be such that $\underline{R}(\langle K, M, N \rangle) = R_h(\langle K, M, N \rangle)$. By Lemma 1.1 (1), $R_h(\langle K, M, N \rangle) = R_h(\langle M, N, K \rangle) = R_h(\langle N, K, M \rangle) \le r$. Then by Lemma 1.1 (3), $R_{3h}(\langle T, T, T \rangle) \le r^3$.

For any integer $s \ge 1$, $\langle T^s, T^s, T^s \rangle$ is $\langle T, T, T \rangle$ tensor product with itself s times. Thus, $R_{3hs}(\langle T^s, T^s, T^s \rangle) \le r^{3s}$. Thus, by Lemma 1.2, $R(\langle T^s, T^s, T^s \rangle) \le c_{3hs} \cdot r^{3s}$, where c_{3hs} is $\binom{3hs+2}{2}$.

A result from last lecture says that if $R(\langle T^s, T^s, T^s \rangle) \le c_{3hs} \cdot r^{3s}$, then

$$\omega \le \frac{3 \log(c_{3hs} r^{3s})}{\log(T^{3s})},$$

which simplifies to

$$\omega \le \frac{3 \log r}{\log(KMN)} + \frac{3 \log(c_{3hs})}{3s \log(NMK)} = \frac{3 \log r}{\log(KMN)} + O\left(\frac{\log s}{s}\right).$$

Thus, the bound approaches $\frac{3 \log r}{\log(KMN)}$ when s approaches ∞. Since ω is defined to be an infimum, it means that $\omega \le \frac{3 \log r}{\log(KMN)}$. □

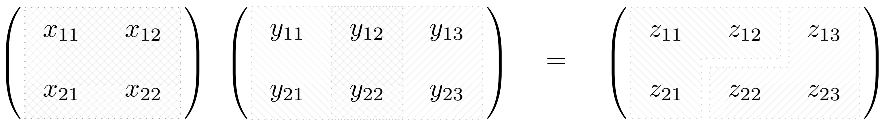

Figure 1: An illustration of how to decompose $\langle 2,2,3\rangle$ to two copies of t. The directions of the lines indicate which copy of t the entries are involved.

2 Bini et al.'s Example

Now we have established the relationship between border rank and ω, we can bound the border rank of some small matrix multiplication tensors to get a better bound on ω. The example in this section is due to Bini et al. [2].

First, we consider the following partial matrix multiplication.

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

where we want to compute a 2 by 2 matrix multiplication but we don't need the z_{22} entry. It can be shown that the tensor t for this partial matrix multiplication has rank 6. However, we will show that $R_1(t) \le 5$.

Consider the following five products

$$\begin{aligned} P_1 &= (x_{12} + \varepsilon x_{22}) y_{21} \\ P_2 &= x_{11}(y_{11} + \varepsilon y_{12}) \\ P_3 &= x_{12}(y_{12} + y_{21} + \varepsilon y_{22}) \\ P_4 &= (x_{11} + x_{12} + \varepsilon x_{21}) y_{11} \\ P_5 &= (x_{12} + \varepsilon x_{21})(y_{11} + \varepsilon y_{22}) \end{aligned}$$

Then it is easy to verify that

$$\begin{aligned} \varepsilon P_1 + \varepsilon P_2 &= \varepsilon z_{11} + O(\varepsilon^2) \\ P_2 - P_4 + P_5 &= \varepsilon z_{12} + O(\varepsilon^2) \\ P_1 - P_3 + P_5 &= \varepsilon z_{21} + O(\varepsilon^2) \end{aligned}$$

Thus, $\underline{R}(t) \le R_1(t) \le 5$.

We can actually use this construction to get an upper bound for the border rank of $\langle 2,2,3\rangle$. As illustrated in Figure 1, $\langle 2,2,3\rangle$ can be decomposed to two copies of t, and thus $\underline{R}(\langle 2,2,3\rangle) \le 2\underline{R}(t) \le 10$. Thus, we can apply Theorem 1.1 to get $\omega \le \frac{3\log 10}{\log 12} \le 2.78$.

3 Schönhage's τ-theorem

Bini et al.'s bound on ω was improved by Schönhage only two years later. Schönhage [3] proved the following general theorem, called the Schönhage τ Theorem, that considers the border rank of direct sums of several matrix multiplication tensors. The theorem is also known as the Asymptotic Sum Inequality.

Theorem 3.1 (Schönhage's τ-Theorem)**.** *Suppose for some integers $r > p$ and integers k_i, m_i, n_i for $1 \le i \le p$, $\underline{R}\left(\bigoplus_{i=1}^{p} \langle k_i, m_i, n_i\rangle\right) \le r$, then $\omega \le 3\tau$ where τ is the solution to $\sum_{i=1}^{p} (k_i m_i n_i)^\tau = r$.*

We delay the proof of Schönhage's τ-theorem until next lecture. In this lecture, we will discuss the implication of the theorem and several tools needed for the proof.

3.1 Better Bound on ω

We will consider the direct sum in the form $\langle k,1,n\rangle \oplus \langle 1,m,1\rangle$. It is known that $R(\langle k,1,n\rangle \oplus \langle 1,m,1\rangle) = kn+m$; also, it is known that $\underline{R}(\langle k,1,n\rangle) = kn$ and $\underline{R}(\langle 1,m,1\rangle) = m$. Thus, it is essential to consider the border rank of the direct sum of multiple matrix multiplication tensors in order to get this improvement. Schönhage showed that $\underline{R}(\langle k,1,n\rangle \oplus \langle 1,m,1\rangle) \leq kn+1$ if $m=(k-1)(n-1)$. In this lecture, we will prove the special case $\underline{R}(\langle 3,1,3\rangle \oplus \langle 1,4,1\rangle) \leq 10$. It is a good exercise to generalize this proof.

When $k=3, n=3, m=4$, we want to compute the $a_i b_j$ for all $i,j \in [3]$ and $\sum_{\ell=1}^{4} u_\ell v_\ell$ together. Consider the following ten products.

$$\begin{aligned}
P_1 &= (a_1+\varepsilon u_1)(b_1+\varepsilon v_1)\\
P_2 &= (a_1+\varepsilon u_2)(b_2+\varepsilon v_2)\\
P_3 &= (a_2+\varepsilon u_3)(b_1+\varepsilon v_3)\\
P_4 &= (a_2+\varepsilon u_4)(b_2+\varepsilon v_4)\\
P_5 &= (a_3-\varepsilon u_1-u_3)b_1\\
P_6 &= (a_3-\varepsilon u_2-u_4)b_2\\
P_7 &= a_1(b_3-\varepsilon v_1-\varepsilon v_2)\\
P_8 &= a_2(b_3-\varepsilon v_3-\varepsilon v_4)\\
P_9 &= a_3 b_3\\
P_{10} &= (a_1+a_2+a_3)(b_1+b_2+b_3)
\end{aligned}$$

Clearly, P_1 through P_9 compute $a_i b_j$ for all $i,j \in [3]$. Also, since

$$\varepsilon^2 \sum_{\ell=1}^{4} u_\ell v_\ell = P_1 + \cdots P_9 - P_{10},$$

these ten products are sufficient. Thus, $\underline{R}(\langle 3,1,3\rangle \oplus \langle 1,4,1\rangle) \leq 10$.

We can apply Schönhage's τ-theorem using the condition $\underline{R}(\langle k,1,n\rangle \oplus \langle 1,m,1\rangle) \leq kn+1$ for $k=4, n=3$ and $m=6$. We get that $\omega \leq 3\tau$ where τ is the solution to $12^\tau + 6^\tau = 13$. This implies $\omega \leq 2.57$, a much better bound!

3.2 Tools

In this section, we introduce several tools necessary to prove Schönhage's τ-theorem.

Definition 3.1 (identity tensor)**.** $\langle r\rangle$ is the identity tensor in $\mathcal{K}^{r\times r\times r}$, where

$$\langle r\rangle_{i,j,k} = \begin{cases} 1 & : i=j=k \\ 0 & : \text{otherwise} \end{cases}$$

It is easy to see that $\langle r\rangle = \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i$, where e_i is the vector whose i-th coordinate is 1 and all other coordinates are 0. It is known that $R(\langle r\rangle) = r$.

Definition 3.2 (restriction)**.** Let $t \in \mathcal{K}^{K\times M\times N}$ and $t' \in \mathcal{K}^{K'\times M'\times N'}$. We say t is a restriction of t' ($t \leq t'$) if there exist homomorphisms

$$\begin{aligned}
A &: \mathcal{K}^{K'} \to \mathcal{K}^{K}\\
B &: \mathcal{K}^{M'} \to \mathcal{K}^{M}\\
C &: \mathcal{K}^{N'} \to \mathcal{K}^{N},
\end{aligned}$$

such that $t = (A\otimes B\otimes C)t'$.

The following claim relates all tensors with identity tensors.

Claim 1. *For any tensor* $t \in \mathcal{K}^{K\times M\times N}$, $t \leq \langle r \rangle$ *if and only if* $R(t) \leq r$.

Proof. From last lecture, we know that $t \leq t'$ implies $R(t) \leq R(t')$. Thus, if $t \leq \langle r \rangle$, then $R(t) \leq R(\langle r \rangle) = r$. Thus, it suffices to prove the other direction.

Suppose $R(t) \leq r$, then there exist u_ℓ, v_ℓ, w_ℓ such that $t = \sum_{i=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$. We define a homomorphism $A : \mathcal{K}^r \to \mathcal{K}^K$ such that $A(e_i) = u_i$ for any $1 \leq i \leq r$. Note that since A is a homomorphism, the values on the basis $\{e_i\}_i$ determine A. We similarly define $B : \mathcal{K}^r \to \mathcal{K}^M$ such that $B(e_i) = v_i$ and $C : \mathcal{K}^r \to \mathcal{K}^N$ such that $B(e_i) = w_i$.

Then we have

$$\begin{aligned}(A \otimes B \otimes C)\langle r \rangle &= (A \otimes B \otimes C) \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \\ &= \sum_{i=1}^{r} A(e_i) \otimes B(e_i) \otimes C(e_i) \\ &= \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i = t,\end{aligned}$$

so $t \leq \langle r \rangle$. □

If $t \leq t'$ and $t' \leq t$, we say that t is isomorphic to t' ($t \cong t'$). This notion of isomorphism is not too nice, since there are cases when t is essentially t' but padded with some zeros. Thus, we define the following notion of isomorphism.

Definition 3.3. We call $t \cong' t'$ is there exist all zero tensors n, n' such that $t \oplus n \cong t' \oplus n'$.

Proposition 1. *The isomorphism classes of tensors form a ring. In other words, the followings are true.*

1. $t \oplus \langle 0 \rangle \cong' t$.
2. $t \otimes \langle 1 \rangle \cong' t$.
3. $t \oplus (t' \oplus t'') \cong' (t \oplus t') \oplus t''$.
4. $t \otimes (t' \otimes t'') \cong' (t \otimes t') \otimes t''$.
5. $t \oplus t' \cong' t' \oplus t$.
6. $t \otimes t' \cong' t' \otimes t$.
7. $t \otimes (t' \oplus t'') \cong' (t \otimes t') \oplus (t \otimes t'')$.

All facts in this proposition should be easy to check, so we omit its proof.

References

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

[2] Dario Bini, Milvio Capovani, Grazia Lotti, and Francesco Romani. $O(n^{2.7799})$ *complexity for matrix multiplication, Inform.*' Proc. Lett 8 (1979): 234-235.

[3] Schönhage, Arnold. *Partial and total matrix multiplication.* SIAM Journal on Computing 10.3 (1981): 434-455.