1 Tools

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

Definition 1.1 (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. Thus the rank of $\langle r\rangle$ is at most r. It is also known that $R(\langle r\rangle) = r$.

Definition 1.2 (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' ($t \le t'$) 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 $t \in \mathcal{K}^{K\times M\times N}$, $t \le \langle r\rangle$ if and only if $R(t) \le r$.*

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

Suppose $R(t) \le r$, 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 $1 \le i \le r$. 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 $t \le \langle r\rangle$. $\square$

If $t \le t'$ and $t' \le t$, 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 1.3. We call $t \cong' t'$ if 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.

2 Preliminaries

We recall several definitions and lemmas from previous lectures.

Definition 2.1. t is a *restriction* of tensor t', denoted $t \leq t'$, if there exist homomorphisms A, B, and C, such that $t = (A \otimes B \otimes C)t'$.

Lemma 2.1. *A tensor t is the restriction of the $r \times r \times r$ diagonal tensor (i.e. $t \leq \langle r \rangle$) if and only if $R(t) \leq r$.*

Lemma 2.2. *If $t \leq t'$ then $t \otimes t'' \leq t' \otimes t''$. The same is true for $\otimes$ replaced by $\oplus$.*

Definition 2.2. Tensor t is *isomorphic* to tensor t', denoted $t \cong t'$, if $t \leq t'$ and $t' \leq t$. Note that in this case A, B, and C are isomorphisms.

Lemma 2.3. *The isomorphism classes of tensors form a ring.*

Definition 2.3. If a is an integer, we let $a \odot t$ denote the direct sum of t with itself a times. Also, we let $t^{\otimes a}$ denote the Kronecker product of t with itself a times.

Lemma 2.4. *For any s, $\langle K^{s+1}, M^{s+1}, N^{s+1} \rangle \cong \langle K^s, M^s, N^s \rangle \otimes \langle K, M, N \rangle$*

Lemma 2.5. *If $R(\langle K, M, N \rangle) \leq r$, then $\omega \leq 3 \log r / \log(KMN)$.*

Lemma 2.6. $(\langle K, M, N \rangle) \otimes (\langle K', M', N' \rangle) = \langle KK', MM', NN' \rangle$.

3 Schönhage's τ theorem

Theorem 3.1 (Schönhage's τ theorem)**.** *Suppose $r > p$ and the border rank $\underline{R}(\oplus_{i=1}^{p} \langle k_i, m_i, n_i \rangle) \leq r$. Then $\omega \leq 3\tau$ where τ is the solution to $\sum_{i=1}^{p} (k_i \cdot m_i \cdot n_i)^\tau = r$.*

Schönhage's τ theorem suggests a new approach to matrix multiplication: identify the direct sum of matrix multiplication tensors, show that its border rank is at most r, and then solve for τ, which bounds ω.

Proof. We begin with a lemma.

Lemma 3.1. *Suppose the rank (not to be confused with border rank) $R(a \odot \langle K, M, N \rangle) \leq b$. Then for all integers $s \geq 1$, $R(a \odot \langle K^s, M^s, N^s \rangle) \leq \lceil b/a \rceil^s \cdot a$.*

Proof. We proceed by induction on s.

Base case: $s = 1$. In this case, $R(a \odot \langle K^s, M^s, N^s \rangle) = R(a \odot \langle K, M, N \rangle) \le b \le \lceil b/a \rceil^s \cdot a$.

Inductive hypothesis: Suppose that $R(a \odot \langle K^s, M^s, N^s \rangle) \le \lceil b/a \rceil^s \cdot a$. By lemma 2.1, this is equivalent to supposing that $a \odot \langle K^s, M^s, N^s \rangle \le \langle \lceil b/a \rceil^s \cdot a \rangle$.

Inductive step: Our goal is to show that $R(a \odot \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle) \le \lceil b/a \rceil^{s+1} \cdot a$. By Lemma 2.4, we have

$$\begin{aligned}
a \odot \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle &\cong (a \odot \langle K^s, M^s, N^s \rangle) \otimes \langle K, M, N \rangle \\
&\le \langle \lceil b/a \rceil^s \cdot a \rangle \otimes \langle K, M, N \rangle \quad \text{by the inductive hypothesis and Lemma 2.2} \\
&\cong (\lceil b/a \rceil^s \cdot a) \odot \langle K, M, N \rangle \\
&\cong (\lceil b/a \rceil^s) \odot (a \odot \langle K, M, N \rangle) \\
&\le \langle \lceil b/a \rceil^s \rangle \otimes \langle b \rangle \\
&\cong \langle \lceil b/a \rceil^s \cdot b \rangle.
\end{aligned}$$

Thus, $R(a \odot \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle) \le \lceil b/a \rceil^s \cdot b \le \lceil b/a \rceil^{s+1} \cdot a$. □

Now we prove a corollary of Lemma 3.1, which we will use to prove Schönhage's τ theorem.

Corollary 3.1. *If $R(a \odot \langle K, M, N \rangle) \le b$, then $\omega \le 3 \log \lceil b/a \rceil / \log(KMN)$.*

Proof. By Lemma 3.1, for all s, $R(\langle K^s, M^s, N^s \rangle) \le \lceil b/a \rceil^s \cdot a$. Thus, by Lemma 2.5, for all s,

$$\begin{aligned}
\omega &\le \frac{3 \log(\lceil b/a \rceil^s \cdot a)}{\log(K^s M^s N^s)} \\
&= \frac{3s \log \lceil b/a \rceil}{s \log(KMN)} + \frac{3 \log a}{s \log(KMN)} \\
&= \frac{3 \log \lceil b/a \rceil}{\log(KMN)} + O(1/s).
\end{aligned}$$

Since $1/s \to 0$ as $s \to \infty$ and ω is an infimum, we have that $\omega \le \frac{3 \log \lceil b/a \rceil}{\log(KMN)}$. □

Now we will use Corollary 3.1 to prove Schönhage's τ theorem. Note that we will need to overcome the fact that Corollary 3.1 is about rank while Schönhage's τ theorem is about border rank. We will use a similar trick to last lecture.

Let $t = \oplus_{i=1}^p \langle k_i, m_i, n_i \rangle$ and let h be an integer such that $R_h(t) \le r$. Let s be a large integer. Then we have $R_{h \cdot s}(t^{\otimes s}) \le r^s$. Last lecture we saw how to turn a border rank expression into a rank expression; we have

$$R(t^{\otimes s}) \le r^s \cdot \mathrm{poly}(h \cdot s). \tag{1}$$

Now, by Lemmas 2.6 and 2.3, and the distributive property of rings, we have

$$\begin{aligned}
t^{\otimes s} &\cong (\oplus_{i=1}^p \langle k_i, m_i, n_i \rangle)^{\otimes s} \\
&\cong \oplus_{s_1, s_2, \ldots, s_p : \sum_{i=1}^p s_i = s} \Big(\frac{s!}{s_1! s_2! \ldots s_p!} \Big) \odot \Big\langle \prod_{i=1}^p k_i^{s_i}, \prod_{i=1}^p m_i^{s_i}, \prod_{i=1}^p n_i^{s_i} \Big\rangle.
\end{aligned}$$

Now we will pick one of the summands from the above expression and apply Corollary 3.1 on it. Let τ with $2/3 < \tau < 1$ be the solution to

$$\sum_{i=1}^p (k_i \cdot m_i \cdot n_i)^\tau = r.$$

Such a τ exists but we will not prove it. Taking both sides of the equation to the s power we get:

$$\sum_{s_1,s_2,\ldots,s_p:\sum_{i=1}^p s_i=s} \Big(\frac{s!}{s_1!s_2!\ldots s_p!}\Big)\cdot\Big(\prod_i k_i^{s_i}\cdot\prod_i m_i^{s_i}\cdot\prod_i n_i^{s_i}\Big)^\tau = r^s.$$

Let $\overline{s}_1,\overline{s}_2,\ldots,\overline{s}_p$ be such than $\sum_{i=1}^p \overline{s}_i = s$ and the inner summand $\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)\cdot\big(\prod_i k_i^{\overline{s}_i}\cdot\prod_i m_i^{\overline{s}_i}\cdot\prod_i n_i^{\overline{s}_i}\big)^\tau$ is maximized.

The number of choices $s_1,s_2,\ldots,s_p$ such that $\sum_{i=1}^p s_i = s$ is $\binom{s+p-1}{p-1}$. Thus, the entire summation above is at most $\binom{s+p-1}{p-1}\cdot\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)\cdot\big(\prod_i k_i^{\overline{s}_i}\cdot\prod_i m_i^{\overline{s}_i}\cdot\prod_i n_i^{\overline{s}_i}\big)^\tau$.

Let $K=\prod_{i=1}^p k_i^{\overline{s}_i}$, let $M=\prod_{i=1}^p m_i^{\overline{s}_i}$, and let $N=\prod_{i=1}^p n_i^{\overline{s}_i}$. Then, the entire summation above is at most $\binom{s+p-1}{p-1}\cdot\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)\cdot(KMN)^\tau$. That is,

$$r^s \le \binom{s+p-1}{p-1}\cdot\Big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\Big)\cdot(KMN)^\tau. \tag{2}$$

Further:

$$r^s\cdot\mathrm{poly}(h\cdot s) \le \binom{s+p-1}{p-1}\cdot\Big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\Big)\cdot(KMN)^\tau\mathrm{poly}(h\cdot s). \tag{3}$$

Let $a=\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)$ and let $b=\binom{s+p-1}{p-1}\cdot\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)\cdot(KMN)^\tau\mathrm{poly}(h\cdot s)$. Then, by Equations 1 and 3, we have $R(a\odot\langle K,M,N\rangle)\le b$.

Notice that $\lceil b/a\rceil = \binom{s+p-1}{p-1}\cdot(KMN)^\tau\mathrm{poly}(h\cdot s)$.

Then, by Corollary 3.1,

$$\begin{aligned}\omega &\le \frac{3\log\lceil b/a\rceil}{\log(KMN)}\\ &\le \frac{3\log(KMN)^\tau}{\log(KMN)}+\frac{3\log\binom{s+p-1}{p-1}\mathrm{poly}(h\cdot s)}{\log(KMN)}\\ &= 3\tau+\frac{3\log\binom{s+p-1}{p-1}\mathrm{poly}(h\cdot s)}{\log(KMN)}.\end{aligned} \tag{4}$$

We claim that as $s\to\infty$, $\frac{3\log\binom{s+p-1}{p-1}\mathrm{poly}(h\cdot s)}{\log(KMN)}\to 0$.

By Equation 3, we have

$$\begin{aligned}\log(KMN)^\tau &\ge \log\Big(\frac{r^s}{\mathrm{poly}(s)\big(\frac{\overline{s}!}{\overline{s}_1!\overline{s}_2!\ldots\overline{s}_p!}\big)}\Big)\\ &\ge \log\Big(\frac{(r/p)^s}{\mathrm{poly}(s)}\Big)\\ &= s\log(r/p)-O(\log s).\end{aligned}$$

Thus, we have

$$\begin{aligned}\frac{3\log\binom{s+p-1}{p-1}\mathrm{poly}(h\cdot s)}{\log(KMN)} &\le \frac{3\log(\mathrm{poly}(s))}{\log(KMN)}\\ &\le O\Big(\frac{\log s}{s\log(r/p)-O(\log s)}\Big)\\ &\le O(\frac{\log s}{s}) \qquad \text{since } r>p.\end{aligned}$$

Since $\frac{\log s}{s}\to 0$ as $s\to\infty$, we have that $\frac{3\log\binom{s+p-1}{p-1}\mathrm{poly}(h\cdot s)}{\log(KMN)}\to 0$ as $s\to\infty$, so by Equation 4, $\omega\le 3\tau$. □

The Schönhage's τ theorem was used to bound ω below 2.5, and subsequently also used in the Coppersmith-Winograd approach, which achieves nearly the best known bound on ω.

4 Introduction to Coppersmith-Winograd

Coppersmith-Winograd use the following special case of Schönhage's τ theorem.

Theorem 4.1 (Special case of Schönhage's τ theorem). *If $\underline{R}(\oplus_{i=1}^{p}\langle k_i, m_i, n_i\rangle) \leq r$ and for all i, $k_i \cdot m_i \cdot n_i = V$, then $\omega \leq \frac{3\log(r/p)}{\log V}$.*

4.1 The Coppersmith-Winograd tensors

Coppersmith-Winograd use several families of tensors. We present them in trilinear notation.

Easy tensors One type of Coppersmith-Winograd tensor is known as the "easy tensor" or "small tensor". For any integer $q \geq 1$ we define the easy tensor $cw_q \in \mathcal{K}^{(q+1)\times(q+1)\times(q+1)}$ as

$$cw_q = \sum_{i=1}^{q} x_0 y_i z_i + x_i y_0 z_i + x_i y_i z_0.$$

Note that the portion of the tensor $\sum_{i=1}^{q} x_0 y_i z_i$ is $\langle 1, 1, q\rangle$, $\sum_{i=1}^{q} x_i y_0 z_i$ is $\langle q, 1, 1\rangle$, and $\sum_{i=1}^{q} x_i y_i z_0$ is $\langle 1, q, 1\rangle$. That is, the easy tensor is the sum of three matrix products, but it's not a direct sum since the terms are not independent of each other e.g. $x_0 y_i z_i$ and $x_i y_0 z_i$ share the variable z_i.

The following is a representation of the easy tensor:

y_q	z_q	0	0	z_0
$\ldots$	$\ldots$	0	z_0	0
y_1	z_1	z_0	0	0
y_0	0	z_1	$\ldots$	z_q
	x_0	x_1	$\ldots$	x_q

Complicated tensors The second type of Coppersmith-Winograd tensor is known as the "complicated tensor" or "big tensor". For any integer $q \geq 1$ we define the complicated tensor $CW_q \in \mathcal{K}^{(q+2)\times(q+2)\times(q+2)}$ as

$$CW_q = cw_q + x_0 y_0 z_{q+1} + x_0 y_{q+1} z_0 + x_{q+1} y_0 z_0.$$

The following is a representation of the complicated tensor:

y_{q+1}	z_0	0	0	0	0
y_q	z_q	0	0	z_0	0
$\ldots$	$\ldots$	0	z_0	0	0
y_1	z_1	z_0	0	0	0
y_0	z_{q+1}	z_1	$\ldots$	z_q	z_0
	x_0	x_1	$\ldots$	x_q	x_{q+1}

There's also a "rotated" version of CW_q, where the diagonal of z_0's is rotated, as follows:

y_{q+1}	z_0	0	0	0	0
y_q	z_q	z_0	0	0	0
$\dots$	$\dots$	0	z_0	0	0
y_1	z_1	0	0	z_0	0
y_0	z_{q+1}	z_1	$\dots$	z_q	z_0
	x_0	x_1	$\dots$	x_q	x_{q+1}

Using the Coppersmith-Winograd tensors Coppersmith-Winograd showed that the border rank of the easy, complicated, and rotated complicated tensors are all $q+2$. For the complicated and rotated complicated tensors, this is tight since they are both in $\mathcal{K}^{(q+2)\times(q+2)\times(q+2)}$. However, $cw_q \in \mathcal{K}^{(q+1)\times(q+1)\times(q+1)}$ and it is not known if this is tight. In particular, if one could show that $\underline{R}(cw_q) = q+1$, then one would show that $\omega = 2$.

One reason for defining the rotated complicated tensor is that it is easier to show that its border rank is at most $q+2$. In particular, the rotated complicated tensor is similar to the tensor for multiplying polynomials, which we saw in a previous lecture, as well as similar to the tensor for addition mod q. In particular, the rotated complicated tensor has a subset of the entries of the tensor for addition mod q. More specifically, the rotated complicated tensor is a *degeneration* of the tensor for addition mod q. Degeneration is like restriction except it preserves border rank instead of rank. One can show via FFT that the addition mod q tensor has rank q, and using degeneration it follows that the rotated complicated tensor has border rank $q+2$.

Next lecture we will see how to get a bound on ω using the easy tensor. An outline is as follows. We take a sum of cw_q's, take this to a large tensor power, and use distributive property in a similar way to our proof of Schönhage's τ theorem. This yields a huge sum of matrix products (but not direct sum). Now, we want to make this huge sum into a direct sum since Schönhage's τ theorem is about direct sums. To do this, we set some variables to 0 in the huge sum. This does not change the border rank. If we choose these variables very carefully, we end up with a huge direct sum of matrix multiplication tensors. This allows us to apply the special case of Schönhage's τ theorem to get a bound on ω.

We will not see how to get a bound on ω using the complicated tensor. The reason it is more difficult than the easy tensor is because it is no longer true that for all i $k_i \cdot m_i \cdot n_i = V$ for some V, which is a precondition for the special case of Schönhage's τ theorem. As a result, we use the general version of Schönhage's τ theorem rather than just the special case, which makes the proof more involved. To bound ω in this case, Coppersmith-Winograd define the "value function" of a tensor, which captures how big of a matrix multiplication tensor you can handle in a big power of your tensor.