1 Preliminaries

We recall several definitions and lemmas from previous lectures.

**Definition 1.1.** $t$ 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 1.1.** 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 1.2.** If $t \leq t'$ then $t \otimes t'' \leq t' \otimes t''$. The same is true for $\otimes$ replaced by $\oplus$.

**Definition 1.2.** Tensor $t$ is isomorphic to tensor $t'$, denoted $t \sim t'$, if $t \leq t'$ and $t' \leq t$. Note that in this case $A$, $B$, and $C$ are isomorphisms.

**Lemma 1.3.** The isomorphism classes of tensors form a ring.

**Definition 1.3.** If $a$ is an integer, we let $a \triangleleft t$ 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 1.4.** For any $s \geq 1$, $R(\langle K^{s+1}, M^{s+1}, N^{s+1} \rangle) \leq \langle K^s, M^s, N^s \rangle \otimes \langle K, M, N \rangle$.

**Lemma 1.5.** If $R(\langle K, M, N \rangle) \leq r$, then $\omega \leq 3 \log r / \log(KMN)$.

**Lemma 1.6.** $(\langle K, M, N \rangle) \otimes (\langle K', M', N' \rangle) = \langle KK', MM', NN' \rangle$.

2 Schönhage’s $\tau$ theorem

**Theorem 2.1** (Schönhage’s $\tau$ theorem). Suppose $r > p$ and the border rank $R(\oplus_{i=1}^p \langle k_i, m_i, n_i \rangle) \leq r$. Then $\omega \leq 3\tau$ where $\tau$ is the solution to $\sum_{i=1}^p (k_i \cdot m_i \cdot n_i)^r = r$.

Schönhage’s $\tau$ 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 $\tau$, which bounds $\omega$.

**Proof.** We begin with a lemma.

**Lemma 2.1.** Suppose the rank (not to be confused with border rank) $R(a \circ \langle K, M, N \rangle) \leq b$. Then for all integers $s \geq 1$, $R(a \circ \langle K^s, M^s, N^s \rangle) \leq [b/a]^s \cdot a$.

**Proof.** We proceed by induction on $s$.

**Base case:** $s = 1$. In this case, $R(a \circ \langle K^s, M^s, N^s \rangle) = R(a \circ \langle K, M, N \rangle) \leq b \leq [b/a]^s \cdot a$.

**Inductive hypothesis:** Suppose that $R(a \circ \langle K^s, M^s, N^s \rangle) \leq [b/a]^s \cdot a$. By lemma 1.1, this is equivalent to supposing that $a \circ \langle K^s, M^s, N^s \rangle \leq \langle [b/a]^s \cdot a \rangle$. 
Inductive step: Our goal is to show that \( R(a \circ \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle) \leq [b/a]^{s+1} \cdot a \). By Lemma 1.4, we have

\[
a \circ \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle \cong (a \circ \langle K^s, M^s, N^s \rangle) \circ \langle K, M, N \rangle
\]

\[
\leq ([b/a]^s \cdot a) \circ \langle K, M, N \rangle \quad \text{by the inductive hypothesis and Lemma 1.2}
\]

\[
\cong ([b/a]^s \cdot a) \circ \langle K, M, N \rangle
\]

\[
\cong ([b/a]^s) \circ (a \cdot \langle K, M, N \rangle)
\]

\[
\leq ([b/a]^s) \circ \langle b \rangle
\]

\[
\cong ([b/a]^s \cdot b).
\]

Thus, \( R(a \circ \langle K^{s+1}, M^{s+1}, N^{s+1} \rangle) \leq [b/a]^s \cdot b \leq [b/a]^{s+1} \cdot a \).

Now we prove a corollary of Lemma 2.1, which we will use to prove Schönhage’s τ theorem.

**Corollary 2.1.** If \( R(a \circ \langle K, M, N \rangle) \leq b \), then \( \omega \leq 3 \log [b/a] / \log (KMN) \).

**Proof.** By Lemma 2.1, for all \( s \), \( R((K^s, M^s, N^s)) \leq [b/a]^s \cdot a \). Thus, by Lemma 1.5, for all \( s \),

\[
\omega \leq \frac{3 \log ([b/a]^s \cdot a)}{\log (K^sM^sN^s)}
\]

\[
= \frac{3s \log [b/a]}{s \log (KMN)} + \frac{3 \log a}{s \log (KMN)}
\]

\[
= \frac{3 \log [b/a]}{\log (KMN)} + O(1/s).
\]

Since \( 1/s \to 0 \) as \( s \to \infty \) and \( \omega \) is an infimum, we have that \( \omega \leq \frac{3 \log [b/a]}{\log (KMN)} \). □

Now we will use Corollary 2.1 to prove Schönhage’s τ theorem. Note that we will need to overcome the fact that Corollary 2.1 is about rank while Schönhage’s τ theorem is about border rank. We will use a similar trick to last lecture.

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

\[
R(t^\otimes s) \leq r^s \cdot \text{poly}(h \cdot s).
\] (1)

Now, by Lemmas 1.6 and 1.3, and the distributive property of rings, we have

\[
t^\otimes s \equiv (\bigoplus_{i=1}^p \langle k_i, m_i, n_i \rangle)^\otimes s
\]

\[
\cong \bigoplus_{s_1, s_2, \ldots, s_p : \sum_{i=1}^p s_i = s} \frac{s!}{s_1!s_2! \ldots s_p!} \cdot \langle k_1^{s_1} \cdot m_1^{s_1} \cdot n_1^{s_1}, \ldots, k_p^{s_p} \cdot m_p^{s_p} \cdot n_p^{s_p} \rangle.
\]

Now we will pick one of the summands from the above expression and apply Corollary 2.1 on it. Let \( \tau \) with \( 2/3 < \tau < 1 \) be the solution to

\[
\sum_{s_1, s_2, \ldots, s_p : \sum_{i=1}^p s_i = s} \frac{s!}{s_1!s_2! \ldots s_p!} \cdot \langle k_1^{s_1} \cdot m_1^{s_1} \cdot n_1^{s_1}, \ldots, k_p^{s_p} \cdot m_p^{s_p} \cdot n_p^{s_p} \rangle^\tau = r^s \cdot \text{poly}(h \cdot s).
\]

Such a \( \tau \) exists but we will not prove it.

Let \( \bar{s}_1, \bar{s}_2, \ldots, \bar{s}_p \) be such that \( \sum_{i=1}^p \bar{s}_i = s \) and the inner summand \( \frac{s!}{s_1!s_2! \ldots s_p!} \cdot \langle k_1^{s_1} \cdot m_1^{s_1} \cdot n_1^{s_1}, \ldots, k_p^{s_p} \cdot m_p^{s_p} \cdot n_p^{s_p} \rangle^\tau \) 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 $\binom{s+p-1}{p-1} \cdot \left( \frac{n!}{s_1!s_2!\ldots s_p!} \right) \cdot \left( \prod_i k_i^r \cdot \prod_i m_i^r \cdot \prod_i n_i^r \right)^\tau$.

Let $K = \prod_{i=1}^{p} k_i^r$, let $M = \prod_{i=1}^{p} m_i^r$, and let $N = \prod_{i=1}^{p} n_i^r$. Then, the entire summation above is at most $\binom{s+p-1}{p-1} \cdot \left( \frac{n!}{s_1!s_2!\ldots s_p!} \right) \cdot (KMN)^\tau$. That is,

$$r^s \cdot \text{poly}(h \cdot s) < \left( \frac{s + p - 1}{p - 1} \right) \cdot \left( \frac{n!}{s_1!s_2!\ldots s_p!} \right) \cdot (KMN)^\tau. \quad (2)$$

Let $a = \left( \frac{n!}{s_1!s_2!\ldots s_p!} \right)$ and $b = \binom{s+p-1}{p-1} \cdot (KMN)^\tau$. Then, by Equations 1 and 2, we have $R(a \odot (K, M, N)) \leq b$.

Then, by Corollary 2.1,

$$\omega \leq \frac{3\log[b/a]}{\log(KMN)} \leq \frac{3\log(KMN)^\tau}{\log(KMN)} + \frac{3\log\binom{s+p-1}{p-1} + 1}{\log(KMN)} = 3\tau + \frac{3\log\binom{s+p-1}{p-1} + 1}{\log(KMN)}. \quad (3)$$

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

By Equation 2, we have

$$\log(KMN)^\tau \geq \log\left( \frac{r^s}{\text{poly}(s) \left( \frac{n!}{s_1!s_2!\ldots s_p!} \right)} \right) \geq \log\left( \frac{(r/p)^s}{\text{poly}(s)} \right) = s\log(r/p) - O(\log s).$$

Thus, we have

$$\frac{3\log\binom{s+p-1}{p-1} + 1}{\log(KMN)} \leq \frac{3\log(\text{poly}(s))}{\log(KMN)} \leq \frac{\log s}{s\log(r/p) - O(\log s)} \leq \frac{\log s}{s} \text{ since } r > p.$$ 

Since $\frac{\log s}{s} \to 0$ as $s \to \infty$, we have that $\frac{3\log\binom{s+p-1}{p-1} + 1}{\log(KMN)} \to 0$ as $s \to \infty$, so by Equation 3, $\omega \leq 3\tau$. \hfill \square

The Schönhage’s $\tau$ theorem was used to bound $\omega$ below 2.5, and subsequently also used in the Coppersmith-Winograd approach, which achieves nearly the best known bound on $\omega$.

### 3 Introduction to Coppersmith-Winograd

Coppersmith-Winograd use the following special case of Schönhage’s $\tau$ theorem.

**Theorem 3.1** (Special case of Schönhage’s $\tau$ theorem). If $R(p \odot (k_i, m_i, n_i)) \leq r$ and for all $i$, $k_i \cdot m_i \cdot n_i = V$, then $\omega \leq \frac{3\log(r/p)}{\log V}$. 

3
3.1 Trilinear notation

Recall that a bilinear problem is to compute \( z_k = \sum_{i,j} t_{i,j,k} \cdot x_i \cdot y_j \). Similarly, trilinear notation is \( \sum_{i,j,k} t_{i,j,k} \cdot x_i \cdot y_j \cdot z_k \), where the goal is to find for every \( z_k \) the coefficient \( \sum_{i,j} t_{i,j,k} \cdot x_i \cdot y_j \) in front of \( z_k \).

For example, the bilinear problem for matrix multiplication is \( z_{ij} = \sum_k x_{ik} y_{kj} \), which in trilinear notation is \( \sum_{i,j,k} x_{ik} y_{kj} z_{ij} \). In research on matrix multiplication, it is written slightly differently as \( \sum_{ijk} x_{ik} y_{kj} z_{ji} \) (the difference is that we take the transpose of \( z \)). The reason for this is that the new version is “super symmetric” i.e. the order of the variables is \( ik, kj, ji \), which forms a cycle. This way, it’s easier to see that permutations of the matrix multiplication tensor preserve the rank and border rank.

3.2 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 \( cw_q \in K^{(q+1)\times(q+1)\times(q+1)} \) 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:

\[
\begin{array}{cccc}
  y_q & z_q & 0 & 0 \\
  \ldots & \ldots & 0 & z_0 \\
  y_1 & z_1 & z_0 & 0 \\
  y_0 & z_1 & \ldots & z_q \\
  x_0 & x_1 & \ldots & x_q \\
\end{array}
\]

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 \( CW_q \in K^{(q+2)\times(q+2)\times(q+2)} \) 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:

\[
\begin{array}{cccc}
  y_{q+1} & z_0 & 0 & 0 \\
  y_q & z_q & 0 & 0 \\
  \ldots & \ldots & 0 & z_0 \\
  y_1 & z_1 & z_0 & 0 \\
  y_0 & z_{q+1} & z_1 & \ldots \\
  x_0 & x_1 & \ldots & x_{q+1} \\
\end{array}
\]

There’s also a “rotated” version of \( CW_q \), where the diagonal of \( z_0 \)’s is rotated, as follows:
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 $\mathbb{K}^{(q+2)\times(q+2)\times(q+2)}$. However, $cw_q \in \mathbb{K}^{(q+1)\times(q+1)\times(q+1)}$ and it is not known if this is tight. In particular, if one could show that $R(cw_q) = q+2$, then one might be able to 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 $\omega$ 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 $\tau$ 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 $\tau$ 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 $\tau$ theorem to get a bound on $\omega$.

We will not see how to get a bound on $\omega$ 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 $\tau$ theorem. As a result, we use the general version of Schönhage’s $\tau$ theorem rather than just the special case, which makes the proof more involved. To bound $\omega$ 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.