# Complexity of Matrix Multiplication (contd...)

Wednesday, 17 May 2023 12:36

## Lecture 3

**Recall :-**

$M_{\langle k,m,n\rangle} : \mathbb{C}^{k\times m} \times \mathbb{C}^{m\times n} \longrightarrow \mathbb{C}^{k\times n}$ is the matrix mult. map
$$(A,B) \longmapsto A\times B$$

$M_{\langle k,m,n\rangle}$ is a bilinear map, so can be thought of as a tensor in
$$(\mathbb{C}^{k\times m})^* \otimes (\mathbb{C}^{m\times n})^* \otimes (\mathbb{C}^{k\times n}).$$

$$\omega = \inf \{ \tau \in \mathbb{R} \mid \underset{\text{Rank}}{R}(M_{\langle n,n,n\rangle}) = O(n^\tau) \}$$

We want to study $\omega$

Machinery that gives you an u.b. on $\omega$, given u.b. on $R(M_{\langle k,m,n\rangle})$ for fixed $k,m,n$

Recall Strassen showed $R(M_{\langle 2,2,2\rangle}) \le 7$.

**Defn** [Permutation of tensors] Let $t \in A \otimes B \otimes C$, where
$t = \sum_{i=1}^{r} t_j$, where $t_j = a_j^{(1)} \otimes a_j^{(2)} \otimes a_j^{(3)}$ Let $\sigma \in S_3$

$\sigma(t) = \sum_{i=1}^{r} \sigma(t_j)$, where $\sigma(t_j) = a_j^{(\sigma(1))} \otimes a_j^{(\sigma(2))} \otimes a_j^{(\sigma(3))}$

**Lemma** $R(\sigma(t)) = R(t)$. *Proof* obvious

**Defn** [Restriction of tensors]. Let $t \in A \otimes B \otimes C$ be such that
$t = \sum_{i=1}^{r} a_i \otimes b_i \otimes c_i$. Let $t' \in A' \otimes B' \otimes C'$ be such that

→ $t' = \sum_{i=1}^{r} f_1(a_i) \otimes f_2(b_i) \otimes f_3(c_i)$, where $f_1 : A \to A'$, $f_2 : B \to B'$,
$f_3 : C \to C'$ are homomorphisms.

We say $t'$ is a restriction of $t$ and we denote $t' \le t$.

$f_3 : C \to C$ are isomorphisms.

We say $t'$ is a restriction of $t$, and we denote $t' \lesssim t$.

**Lemma** If $t' \lesssim t$, $R(t') \le R(t)$. We have equality when $f_1, f_2, f_3$ are isomorphisms.

*Proof* by defn $\boxtimes$

**Lemma.** Take $\sigma \in S_3$. $R(M_{\langle k,m,n\rangle}) = R(M_{\sigma\langle k,m,n\rangle})$

*Proof*

$$M_{\langle k,m,n\rangle} : \mathbb{C}^{k\times m} \times \mathbb{C}^{m\times n} \longrightarrow \mathbb{C}^{k\times n}$$

$$(\mathbb{C}^{k\times m})^* \otimes (\mathbb{C}^{m\times n})^* \otimes \mathbb{C}^{k\times n}$$

Perm

$$m\times n \xrightarrow{\text{transpose}} n\times m \qquad k\times m \xrightarrow{\text{transpose}} m\times k$$

$\boxtimes$

**Defn** [Direct sum of tensors] $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$, $t' \in \mathbb{F}^{k'} \otimes \mathbb{F}^{m'} \otimes \mathbb{F}^{n'}$.

$t = \sum_{i=1}^{r} t_i$, where $t_i = \underset{\mathbb{F}^k}{a_i} \otimes \underset{\mathbb{F}^m}{b_i} \otimes \underset{\mathbb{F}^n}{c_i}$

$a_i$ is a $k$-dimensional co-ordinate vector.

$\to$ $t' = \sum_{i=1}^{r'} t'_i$ " $t'_i = \underset{\mathbb{F}^{k'}}{a'_i} \otimes \ldots$

$a'_i$ are $k'$-dim co-ordinate vectors

$\underline{a_i} = (a_i, \underbrace{-\text{ - - - }}_{k'}^{0's}\,)$, $\underset{\mathbb{F}^{m+m'}}{\underline{b_i}}$ . . . $\underset{\mathbb{F}^{n+n'}}{\underline{c_i}}$ . . . .

$\Downarrow$

$\underline{a_i} \in \mathbb{F}^{k+k'}$

$\underline{a_i}' = (\underbrace{\text{. - - -}}_{k}^{0's}, a'_i)$, $\underline{b_i}' = \ldots$, $\underline{c_i}'$

$r$ $r'$

$$t \oplus t' = \sum_{i=1}^{r} \underline{a_i} \otimes \underline{b_i} \otimes \underline{c_i} + \sum_{i=1}^{r'} \underline{a_i'} \otimes \underline{b_i'} \otimes \underline{c_i'}$$

**Lemma** $R(t \oplus t') \leq R(t) + R(t')$

*Proof* by defn $\boxtimes$

**Conjecture** This inequality is tight. (Strassen's additivity conjecture)

**Defn** [tensor product of tensors] $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$, $t' \in \mathbb{F}^{k'} \otimes \mathbb{F}^{m'} \otimes \mathbb{F}^{n'}$

$t = \sum_{i=1}^{r} t_i$ $\quad t_i = \underline{a_i} \otimes \underline{b_i} \otimes \underline{c_i}$ $\quad \rightarrow a_i = (a^{(1)}, \ldots, a^{(k)})$, $a_i \in \mathbb{F}^k$

$t' = \sum_{i=1}^{r'} t_i'$ $\quad t_i' = \underline{a_i'} \otimes \underline{b_i'} \otimes \underline{c_i'}$ $\quad \rightarrow a_i' = (\underline{a}^{(1)}, \ldots, \underline{a}^{(k')})$, $a_i' \in \mathbb{F}^{k'}$

$$(a_i \otimes a_i')_{i,j} = a^{(i)} \times \underline{a}^{(j)}, \qquad a_i \otimes a_i' \in \mathbb{F}^{kk'}$$

extend linearly to obtain $t \otimes t' \in \mathbb{F}^{kk'} \otimes \mathbb{F}^{mm'} \otimes \mathbb{F}^{nn'}$

**Lemma** $R(t \otimes t') \leq R(t) R(t')$

*Proof* by defn $\boxtimes$

**Observe** $M_{\langle k,m,n\rangle} \otimes M_{\langle k',m',n'\rangle} = M_{\langle kk', mm', nn'\rangle}$

**Thm** $R(M_{\langle k,m,n\rangle}) \leq r$. Then $\omega \leq 3 \log_{kmn} r$

**Proof**

$$R(M_{\langle k,m,n\rangle} \otimes M_{\langle m,n,k\rangle} \otimes M_{\langle n,k,m\rangle}) \leq r^3$$

$$\Rightarrow R(M_{\langle kmn, kmn, kmn\rangle}) \leq r^3 = (kmn)^{3 \log_{kmn} r}$$

$$\omega \leq 3 \log_{kmn} r \qquad \boxtimes$$

Strassen $R(M_{\langle 2,2,2\rangle}) \leq 7 \implies \omega \leq 3\log_8 7 \simeq 2.8074$

Pan $R(M_{\langle 70,70,70\rangle}) \leq 143640 \implies \omega \leq 2.796$

There is an intuitive way.

Timeline of matrix multiplication exponent

| Year | Bound on omega | Authors |
|---|---|---|
| 1969 | 2.8074 | Strassen[1] |
| 1978 | 2.796 | Pan[11] |
| 1979 | 2.780 | Bini, Capovani [it], Romani[12] |
| 1981 | 2.522 | Schönhage[13] |
| 1981 | 2.517 | Romani[14] |
| 1981 | 2.496 | Coppersmith, Winograd[15] |
| 1986 | 2.479 | Strassen[16] |
| 1990 | 2.3755 | Coppersmith, Winograd[17] |
| 2010 | 2.3737 | Stothers[18] |
| 2013 | 2.3729 | Williams[19][20] |
| 2014 | 2.3728639 | Le Gall[21] |
| 2020 | 2.3728596 | Alman, Williams[3] |
| 2022 | 2.37188 | Duan, Wu, Zhou[2] |

## Border rank

Consider $M_{\langle 2,2,3\rangle}$

$$2\times 2 \quad \times \quad 2\times 3 \longrightarrow 2\times 3$$

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \times \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \end{bmatrix}$$

$M_{\text{reduced} \langle 2,2,2\rangle}$

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

Idea Do $M_{\langle 2,2,3\rangle}$ using two $M_{\text{reduced}\langle 2,2,2\rangle}$

Idea: Do $M_{\langle 2,2,3\rangle}$ using two $M_{\text{reduced}\,\langle 2,2,2\rangle}$

$M_{\text{reduced}\,\langle 2,2,2\rangle}$ in 6 mults. trivially. Can you do in 5 mults?

shows $\mathbf{R}(M^{red}_{\langle 2\rangle}) \leq 6$. Bini et al. attempted to find a rank five expression for $M^{red}_{\langle 2\rangle}$. They searched for such an expression by computer. Their method was to minimize the norm of $M^{red}_{\langle 2\rangle}$ minus a rank five tensor that varied (see §4.6 for a description of the method), and their computer kept on producing rank five tensors with the norm of the difference getting smaller and smaller, but with larger and larger coefficients. Bini (personal communication) told me about how he lost sleep trying to understand what was wrong with his computer code. This went on for some time, when finally he realized *there was nothing wrong*

$$M_{\text{reduced}\,\langle 2,2,2\rangle} = \lim_{t\to 0} \frac{1}{t}\Big[ \underbrace{(x_{12}+t x_{11})\otimes(y_{12}+t y_{22})\otimes z_{21}}_{1}$$

B

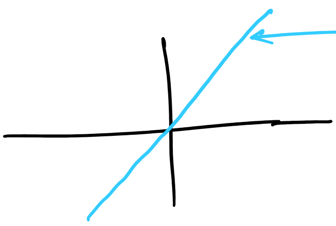
$$+ \underbrace{(x_{21}+t x_{11})\otimes y_{11}\otimes(z_{11}+t z_{12})}_{2} - \underbrace{\big(x_{12}\otimes y_{12}\otimes(z_{11}+z_{21}+t z_{22})\big)}_{3}$$

$$- \underbrace{x_{21}\otimes\big((y_{11}+y_{12})+t y_{21}\big)\otimes z_{11}}_{4} + \underbrace{(x_{12}+x_{21})\otimes(y_{12}+t y_{21})\otimes(z_{11}+t z_{22})}_{5}\Big]$$

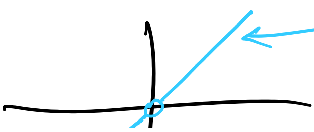
$M_{\text{reduced}\,\langle 2,2,2\rangle}$ can be arbitrarily well approximated by a rank 5 tensor

(*) $\underline{R}\left(M_{\text{reduced}\,\langle 2,2,2\rangle}\right) \leq 5$. We'll use u.b. on border ranks to get an upper bound on $\omega$.

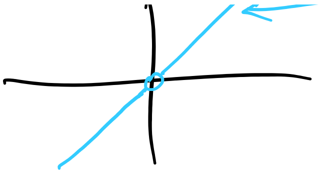
(*) $\underline{R}\left(M_{\langle 2,2,3\rangle}\right) \leq 10$.



$\{(x,x) \mid x \in \mathbb{R}\} = Z(y-x)$



Punctured line $Z(y-x) \setminus \{0,0\}$



Defn [Zariski Closure] Zariski closure of the punctured line is Zero set of the polynomial $y-x$.

If a set is Zariski closed, there is a finite set of polynomials whose common zeros vanish exactly on the set.

Defn [Segre Variety] (Parameterizes Rank 1 tensors)

$$Seg: \mathbb{P}(A_1) \times \cdots \mathbb{P}(A_n) \longrightarrow \mathbb{P}(A_1 \otimes \cdots \otimes A_n)$$

$$(a_1, \dots, a_n) \longmapsto a_1 \otimes \cdots \otimes a_n$$

$a_i$ have $\dim(A_i)$ homogeneous coordinates

$\sigma_1 = \mathrm{Im}(Seg)$ ← Segre Variety.

Defn [Secant line] Let $V$ be a projective variety. A line $L$ is called a secant line to $V$ if $L$ meets $V$ in two or more points.

Defn Let $X \subseteq \mathbb{P}(V)$ be a projective variety. Define

$$S_r(X)^0 = \{ (x_1, \dots, x_r, z) \in X^{\times r} \times \mathbb{P}(V) \mid z \in \mathrm{span}(x_1, \dots, x_r) \}$$

$$\subseteq Seg(X^{\times r} \times \mathbb{P}(V)) \subseteq \mathbb{P}(V^{\otimes r+1}) \leftarrow$$

Let $S_r(X) = \overline{S_r(X)^0}^{\text{Zariski closure}}$.

$S_r(X)$ is called the $r^{th}$ abstract Secant variety of $X$.

$\pi^0 : S_r(X)^0 \longrightarrow \mathbb{P}(V)$ & likewise $\pi : S_r(X) \longrightarrow \mathbb{P}(V)$

$(x_1, \dots, x_r, z) \longmapsto z$

– Image of $\pi^0$ is denoted $\sigma_r^0(X)$

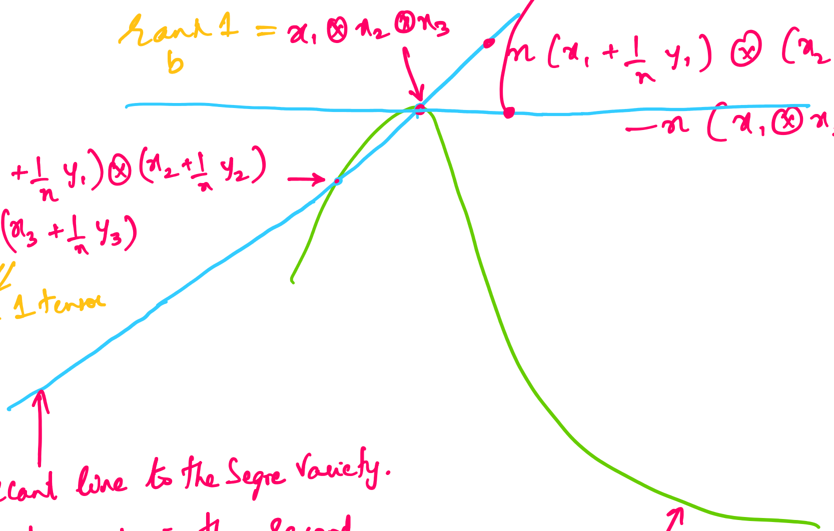
- Image of $\pi^0$ is denoted $\sigma_r^0(X)$
- $\sigma_r(X) = \mathrm{Im}(\pi)$ (Called $r$th Secant variety of $X$ ($\sigma_1(X) = X$))

When $X = \mathrm{Seg}(\mathbb{P}(A_1) \times \cdots \times \mathbb{P}(A_n))$

$\sigma_r(X) \to$ set of tensors of <u>border</u> rank at most $r$

(rank jumping tensor)

<u>Diagram</u>




(Ex) $M_{\text{reduced } \langle 2,2,2 \rangle}$ is a rank jumping tensor

Idea:- Interpolate between rank and border rank

Let $\varepsilon$ be an indeterminate

<u>Defn</u> [$\underset{\mathbb{Z}_{\geq 0}}{h}$-rank, border rank] $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$

① $R_h(t) = \min \Big\{ r \;\Big|\; \exists\, u_i \in \mathbb{F}[\varepsilon]^k, v_i \in \mathbb{F}[\varepsilon]^m, w_i \in \mathbb{F}[\varepsilon]^n : \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i = \varepsilon^h t + O(\varepsilon^{h+1}) \Big\}$

$\Updownarrow$

$t = \lim \frac{1}{\;} \big( \sum^{r} u_i \otimes v_i \otimes w_i \big)$

$$\Downarrow$$

$$t = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^h} \left( \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \right)$$

(2) $\underline{R}(t) = \min_{h \in \mathbb{N}} R_h(t)$

**Lemma** (1) $R_0(t) = R(t)$

(2) $R_0(t) \geq R_1(t) \geq \dots = \underline{R}(t)$

**Proof** By defn $\boxtimes$

**Lemma** (1) $\pi \in S_3$ : $R_h(\pi(t)) = R_h(t)$

(2) $R_{\max\{h,h'\}}(t \oplus t') \leq R_h(t) + R_{h'}(t')$

(3) $R_{h+h'}(t \otimes t') \leq R_h(t) \, R_{h'}(t')$

**Proof** Same as earlier lemmas .. use language of $R_h$ $\boxtimes$

**Lemma** [turn approx. computations into real ones] There is a constant $C_h \leq \binom{h+2}{2}$ s.t $\forall t$

$$R(t) \leq C_h R_h(t)$$

**Proof** Counting $\boxtimes$

**Thm** $\underline{R}(M_{\langle k,m,n \rangle}) \leq r \Rightarrow \omega \leq 3 \log_{kmn} r$

**Proof** Use $R_h$ ... and same logic as previous thm that gave you u.b. on $\omega$ using an u.b. on $R(M_{\langle * \rangle})$ ... make $h$ small .. $\boxtimes$

(*) $\underline{R}(M_{\text{reduced}\,\langle 2,2,2 \rangle}) \leq 5 \Rightarrow \underline{R}(M_{\langle 2,2,3 \rangle}) \leq 10$

$$\Rightarrow \omega \leq 3 \log_{12} 10 \simeq 2.78$$

# Schonhage's $\tau$ theorem

- Strassen & Pan turned u.b. of $R(M_{\langle * \rangle})$ into u.b. on $\omega$
- Bini et al. turned u.b. on $\underline{R}(M_{\langle * \rangle})$ into u.b. on $\omega$
- Schonhage turns u.b. on $\underline{R}$ (not a mat mult. tensor but it is a direct sum of matrix mult. tensor) into u.b. on $\omega$

**Lemma**
1. $R(M_{\langle k,1,n \rangle} \oplus M_{\langle 1,m,1 \rangle}) = kn+m$
2. $\underline{R}(M_{\langle k,1,n \rangle}) = kn$ & $\underline{R}(M_{\langle 1,m,1 \rangle}) = m$
3. $\underline{R}(M_{\langle k,1,n \rangle} \oplus M_{\langle 1,N,1 \rangle}) \leq kn+1$ when $N = (n-1)(k-1)$

**Defn** $I_{\langle r \rangle} \in \mathbb{F}^r \otimes \mathbb{F}^r \otimes \mathbb{F}^r$ is s.t

$$(I_{\langle r \rangle})_{i,i,i} = 1 \quad \forall i \in [r] \text{ \& } 0\text{'s elsewhere.}$$

$R(I_{\langle r \rangle}) = r$

**Lemma** $R(t) \leq r \iff t \lesssim I_{\langle r \rangle}$

**Proof** skipped $\boxtimes$

**Thm** [$\tau$-theorem] If $\underline{R}\left(\bigoplus_{i=1}^{p} M_{\langle k_i, m_i, n_i \rangle}\right) \leq r$, and $r > p$, then

$\omega \leq 3\tau$, where $\tau$ satisfies

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

**Proof** skipped $\boxtimes$

- $\underline{R}(M_{\langle 4,1,3 \rangle} \oplus M_{\langle 1,6,1 \rangle}) \leq 13$

Using the theorem $\omega \leq 2.55$

## Coppersmith-Winograd

**Defn** [easy C.W. tensor]

$$T_{q,cw} = \sum_{j=1}^{q} a_0 \otimes b_j \otimes c_j + a_j \otimes b_0 \otimes c_j + a_j \otimes b_j \otimes c_0$$

$$\mathbb{C}^{q+1} \otimes \mathbb{C}^{q+1} \otimes \mathbb{C}^{q+1}$$

dense sets of integers with no 3 term arithmetic progressions

expression has $3q$ terms. Turns out

$$\underline{R}(T_{q,cw}) = q+2.$$

**Thm**
$$\omega \leq \frac{\log\left(\frac{4}{27}\underline{R}(T_{q,cw})^3\right)}{\log q}.$$

Set $q = 8, \Rightarrow \omega \leq 2.41$

**Defn** [big CW tensor]

$$T^+_{q,cw} = T_{q,cw} + a_0 \otimes b_0 \otimes c_{q+1} + a_0 \otimes b_{q+1} \otimes c_0 + a_{q+1} \otimes b_0 \otimes c_0$$

$$\mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2}$$

**Thm** $\underline{R}(T^+_{q,cw}) = q+2$, and

$$\omega \leq \frac{\log\left(\frac{4}{27}\underline{R}\left(T^{+\,\otimes k}_{q,cw}\right)^{3/k}\right)}{\log q}$$

C.W. analyzed $T^+_{q,cw} \otimes T^+_{q,cw}$

Latest work analyzes $\left(T^+_{q,cw}\right)^{\otimes 32} \Longrightarrow \omega < 2.3728639$

Latest work analyses $\left(T^{+}_{q,CW}\right)$ $\Longrightarrow$ $\omega < 2.3728639$ $\uparrow$

People have shown that taking higher & higher powers of $T^{+}_{q,CW}$ won't improve beyond 2.3

Conjecture [Asymptotic Rank Conjecture]

$$\lim_{n\to\infty} R\left(T_{2,CW}^{\otimes n}\right)^{1/n} = 3$$

(Theorem $\Longrightarrow$ $\omega = 2$)

Conjecture [No 3 disjoint equivoluminous subsets Conjecture] Let $H$ be an abelian group.

Let $m_1, \ldots, m_n \in H$. This satisfies N3DES property if

$\forall\, S, T, U \subseteq [n]$, $S, T, U$ disjoint

$$\sum_{i\in S} m_i \neq \sum_{i\in U} m_i \neq \sum_{i\in T} m_i \,,$$

The actual conjecture is that there is an $H$ of size

$$|H| \leq 2^{o(n)}$$

[Theorem Conjecture $\Rightarrow$ asymptotic rank conjecture]

e.g. $H = \mathbb{Z}_2^n$ $\quad m_i = e_i$ satisfies N3DES property, but $|H| = 2^n$

$\star$ truth of the above conjecture $\Rightarrow$ one of the sunflower conjectures is false..