Tensor decompositions and Theoretical computer science

Lecture 3: Sums of matrix multiplication tensors / Border rank

Vladimir Lysikov

14.09.2022

Recall: complexity of matrix multiplication

- We defined the *exponent of matrix mult.* ω in terms of circuit complexity
- And then we proved that it can be equivalently defined in terms of tensor rank

Matrix multiplication tensor

$$\langle \ell, m, n \rangle = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} e_{ij} \otimes e_{jk} \otimes e_{ki}$$

Theorem

$$\omega = \inf\{p \mid R(\langle n, n, n \rangle) = O(n^p)\}$$

If $\omega < p$, then $n \times n$ matrix multiplication can be performed using $O(n^p)$ algebraic operations.

If $\omega = p$, then — using $n^{p+o(1)}$ operations

Recall: Basic tools for matrix multiplication

Theorem

$$\omega \leq \log_n R(\langle n, n, n \rangle)$$

$$R(\langle n, n, n \rangle) \leq r \Rightarrow n^\omega \leq r$$

Theorem

$$\omega \leq 3 \log_{\ell mn} R(\langle \ell, m, n \rangle)$$

$$R(\langle \ell, m, n \rangle) \leq r \Rightarrow (\ell mn)^{\omega/3} \leq r$$

Theorem (Pan 79)

$$R(p \odot \langle n, n, n \rangle) \leq r \Rightarrow pn^\omega \leq r$$

Pan used this statement to prove $\omega < 2.8$, but the construction is complicated

Asymptotic sum inequality: Uniform version

Theorem

$$R(p \odot \langle \ell, m, n\rangle) \leq r \Rightarrow p(\ell mn)^{\omega/3} \leq r$$

- Proven in the same way as for one rectangular map:
- $R(p \odot \langle \ell, m, n\rangle) = R(p \odot \langle m, n, \ell\rangle) = R(p \odot \langle n, \ell, m\rangle)$ by shifting tensor factors
- $R(p \odot \langle \ell, m, n\rangle) \leq r \Rightarrow R(p^3 \odot \langle \ell mn, \ell mn, \ell mn\rangle) \leq r^3$
- Applying the square version, we get $p^3(\ell mn)^{\omega} \leq r^3$

Asymptotic sum inequality

Theorem (≈ Schönhage 81)

$$R\left(\bigoplus_{k=1}^{p} \langle \ell_k, m_k, n_k \rangle\right) \leq r \Rightarrow \sum_{k=1}^{p} (\ell m n)^{\omega/3} \leq r$$

Asymptotic sum inequality

We will prove this for the case of 2 summands

Theorem

$$\text{If } R(\langle \ell_1, m_1, n_1\rangle \oplus \langle \ell_2, m_2, n_2\rangle) \leq r$$
$$\text{then } (\ell_1 m_1 n_1)^{\frac{\omega}{3}} + (\ell_2 m_2 n_2)^{\frac{\omega}{3}} \leq r$$

Asymptotic sum inequality: Binomial formula

- Properties of tensor operations

$$T \oplus S \cong S \oplus T$$

$$T \boxtimes S \cong S \boxtimes T$$

$$(T_1 \oplus T_2) \boxtimes S \cong (T_1 \boxtimes S) \oplus (T_2 \boxtimes S)$$

- Here $T \cong S$ if $T = (A \otimes B \otimes C)S$ with invertible A, B, C
- This is enough to prove the binomial formula

$$(T_1 \oplus T_2)^{\boxtimes N} \cong \bigoplus_{K=1}^{N} \binom{N}{K} \odot T_1^{\boxtimes K} \boxtimes T_2^{\boxtimes (N-K)}$$

Asymptotic sum inequality

Theorem

If $R(\langle \ell_1, m_1, n_1 \rangle \oplus \langle \ell_2, m_2, n_2 \rangle) \leq r$

then $(\ell_1 m_1 n_1)^{\frac{\omega}{3}} + (\ell_2 m_2 n_2)^{\frac{\omega}{3}} \leq r$

$$(\langle \ell_1, m_1, n_1 \rangle \oplus \langle \ell_2, m_2, n_2 \rangle)^{\boxtimes N} \cong \bigoplus_{K=1}^{N} \binom{N}{K} \odot \left\langle \ell_1^K \ell_2^{N-K}, m_1^K m_2^{N-K}, n_1^K n_2^{N-K} \right\rangle$$

- Take only one of the “monomials”

$$R\left(\binom{N}{K} \odot \left\langle \ell_1^K \ell_2^{N-K}, m_1^K m_2^{N-K}, n_1^K n_2^{N-K} \right\rangle\right) \leq r^N$$

Entropy enters

- We need some facts involving entropies
- First, the binomial coefficients can be estimated using entropies

$$\binom{N}{K} \leq \left(\frac{N}{K}\right)^K \left(\frac{N}{N-K}\right)^{N-K} = 2^{N \cdot H(\frac{K}{N}, 1-\frac{K}{N})}$$

- This estimate becomes better if $N \to \infty$

$$\lim_{N \to \infty} \frac{1}{N} \log \binom{N}{\lfloor pN \rfloor} = H(p, 1-p)$$

- Finally,

$$\max_p H(p, 1-p) + p \log a + (1-p) \log b = \log(a+b)$$

- Maximum is attained on $p = \frac{a}{a+b}$
- Note that $\log \binom{N}{K} a^K b^{N-K} \approx H(\frac{K}{N}, 1-\frac{K}{N}) + \frac{K}{N} \log a + (1-\frac{K}{N}) \log b$

Asymptotic sum inequality

$$R\left(\binom{N}{K} \odot \langle \ell_1^K \ell_2^{N-K}, m_1^K m_2^{N-K}, n_1^K n_2^{N-K} \rangle\right) \leq r^N$$

- Apply the uniform version

$$\binom{N}{K} \left[(\ell_1 m_1 n_1)^K (\ell_2 m_2 n_2)^{N-K}\right]^{\omega/3} \leq r^N$$

- Let $p = \frac{K}{N}$

$$\frac{1}{N} \log \binom{N}{K} + p\frac{\omega}{3} \log \ell_1 m_1 n_1 + (1-p)\frac{\omega}{3} \log \ell_2 m_2 n_2 \leq \log r$$

- Fix the probability and let $N \to \infty$

$$H(p, 1-p) + p\frac{\omega}{3} \log \ell_1 m_1 n_1 + (1-p)\frac{\omega}{3} \log \ell_2 m_2 n_2 \leq \log r$$

Asymptotic sum inequality: final steps

$$H(p, 1-p) + p\frac{\omega}{3}\log \ell_1 m_1 n_1 + (1-p)\frac{\omega}{3}\log \ell_2 m_2 n_2 \leq \log r$$

$$\log\left[(\ell_1 m_1 n_1)^{\omega/3} + (\ell_2 m_2 n_2)^{\omega/3}\right] \leq \log r$$

Theorem

If $R(\langle \ell_1, m_1, n_1\rangle \oplus \langle \ell_2, m_2, n_2\rangle) \leq r$

then $(\ell_1 m_1 n_1)^{\frac{\omega}{3}} + (\ell_2 m_2 n_2)^{\frac{\omega}{3}} \leq r$

Story of border rank

$$T = \sum_{a=1}^{r} u_a \otimes v_a \otimes w_a$$

- Tensor rank decomposition can be thought as an equation on u_a, v_a, w_a
- Numerical methods to solve: nonlinear least squares

$$\| T - \sum_{a=1}^{r} u_a \otimes v_a \otimes w_a \|^2 \to \min$$

- Works very well for $\langle 2, 2, 2 \rangle$
- Can be made to work for slightly larger tensors, but diverges often
- Why divergence?

Border rank

$$W = a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a$$

- It is possible to prove that $R(W) = 3$
- One approach: if $R(W) = 2$, then $W = x \otimes M + y \otimes N$ with M, N of rank 1

$$W = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[(a + \varepsilon b)^{\otimes 3} - a^{\otimes 3} \right]$$

- W is a limit of rank 2 tensors

Definition

Consider a tensor space $U \otimes V \otimes W$. Let

$$X_r = \{ T \in U \otimes V \otimes W \mid R(T) \leq r \}$$

be the set of all tensors of rank at most r. The *border rank* $\underline{R}(T)$ is the minimal r such that T lies in the closure $\overline{X_r}$

Border rank

Definition

$$X_r = \{T \in U \otimes V \otimes W \mid R(T) \leq r\}$$

The *border rank* $\underline{R}(T)$ is the minimal r such that T lies in the closure $\overline{X_r}$

- Over arbitrary field take Zariski closure
- Over $\mathbb{C}$ the closure in analytic topology is the same, because X_r is a constructible set

Secant varieties

- The set X_1 of rank 1 tensors is the affine cone over the Segre variety $Seg(\mathbb{P}U \times \mathbb{P}V \times \mathbb{P}W) \subset \mathbb{P}(U \otimes V \otimes W)$
- We have $\mathbb{P}\overline{X_r} = \sigma_r(Seg(\mathbb{P}U \times \mathbb{P}V \times \mathbb{P}W))$
- Border rank is captured by secant varieties of the Segre variety
- For example, $[W]$ lies on a tangent to the Segre variety at the point $[a^{\otimes 3}]$

Orbit closures

- Recall that $R(T) \leq r$ if and only if $T \leq E_r$
- Consider tensors in $\mathbf{k}^{r \times r \times r}$
- $\overline{X_r} = \overline{\{T \mid T = (A \otimes B \otimes C)E_r \text{ where } A, B, C \colon \mathbf{k}^r \to \mathbf{k}^r\}}$
- $\overline{X_r} = \overline{\{T \mid T = (A \otimes B \otimes C)E_r \text{ where } A, B, C \in \mathrm{GL}_r\}}$
- $\overline{X_r} = \overline{(\mathrm{GL}_r^{\times 3}) \cdot E_r}$

Definition

$T \in U \otimes V \otimes W$ is a *degeneration* of $S \in U' \otimes V' \otimes W'$ if there is a tensor $T' \in \overline{\mathrm{GL}(U') \times \mathrm{GL}(V') \times \mathrm{GL}(W') \cdot S}$ such that $T \cong T'$

Notation: $T \trianglelefteq S$

Theorem

$$\underline{R}(T) \leq r \Leftrightarrow T \trianglelefteq S$$

Algebraic characterization

- We need a more algebraic approach

Theorem

Let $T, S \in \mathbf{k}^{\ell \times m \times n}$. We have $T \trianglelefteq S$ if and only if

$$\varepsilon^p T + \varepsilon^{p+1} T_{p+1} + \cdots = [A(\varepsilon) \times B(\varepsilon) \times C(\varepsilon)]S$$

for some p and some matrices A, B, C with polynomial entries.

- The easy direction: divide by ε^p and let $\varepsilon \to 0$ to prove $T \in \overline{\{(A \otimes B \otimes C)S\}}$
- Example: $\varepsilon W + o(\varepsilon) = (a + \varepsilon b)^{\otimes 3} + (-a)^{\otimes 3} = \begin{bmatrix} 1 & \varepsilon \\ -1 & 0 \end{bmatrix} E_2$
- p is called the *approximation degree*

Algebraic degenerations: properties

- Write $T \trianglelefteq_p S$ if the approximation degree is p

Lemma

$$T_1 \trianglelefteq_{p_1} S_1, T_2 \trianglelefteq_{p_2} S_2 \Rightarrow T_1 \boxtimes T_2 \trianglelefteq_{p_1+p_2} S_1 \boxtimes S_2$$

$$F_p = \sum_{\substack{0 \le i,j,k \le p \\ i+j+k=p}} e_i \otimes e_j \otimes e_k$$

- F_p is equivalent to the tensor of truncated polynomial multiplication ($\mathbf{k}[\varepsilon]/\left\langle \varepsilon^{p+1} \right\rangle$)

Lemma

$$T \trianglelefteq_p S \Rightarrow T \le S \boxtimes F_p$$

From degenerations to restrictions

Lemma

$$T \trianglelefteq_p S \Rightarrow T \leq S \boxtimes F_p$$

- Suppose $\epsilon^p T + o(\varepsilon^p) = [A(\varepsilon) \otimes B(\varepsilon) \otimes C(\varepsilon)]S$
- Let $A(\varepsilon) = A_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \ldots$ and the same for B, C.
- The coefficients before ε^p in $A(\varepsilon) \otimes B(\varepsilon) \otimes C(\varepsilon)$ is

$$\sum_{i+j+k=p} A_i \otimes B_j \otimes C_k$$

- Therefore

$$T = \sum_{i+j+k=p} (A_i \otimes B_j \otimes C_k)S$$

From degenerations to restrictions

Lemma

$$T \trianglelefteq_p S \Rightarrow T \leq S \boxtimes F_p$$

- Define $\hat{A}\colon U' \otimes \mathbf{k}^{p+1} \to U$

$$\hat{A}(\sum_i x_i \otimes e_i) = \sum_i A_i x_i$$

- and analogously $\hat{B}, \hat{C}$
- We have

$$(\hat{A} \otimes \hat{B} \otimes \hat{C})(S \boxtimes F_p) = \sum_{i+j+k=p} (\hat{A} \otimes \hat{B} \otimes \hat{C})(S \boxtimes (e_i \otimes e_j \otimes e_k)) =$$

$$= \sum_{i+j+k=p} (A_i \otimes B_j \otimes C_k) S = T$$

- Thus $T \leq S \boxtimes F_p$

Border rank and matrix multiplication

Theorem (Bini 80)

$\omega \leq \log_n \underline{R}(\langle n, n, n \rangle)$

- Let $\langle n, n, n \rangle \trianglelefteq_p E_r$
- Take Kronecker powers

$$\langle n, n, n \rangle^{\boxtimes N} \trianglelefteq_{Np} E_r^{\boxtimes N} \cong E_{r^N}$$

- Get rid of degenerations

$$\left\langle n^N, n^N, n^N \right\rangle \leq E_{r^N} \boxtimes F_{Np}$$

- Trivial upper bound: $R(F_p) \leq (p+1)^2$

$$R(\left\langle n^N, n^N, n^N \right\rangle) \leq (Np+1)^2 r^N$$

- $\omega \leq \log_{n^N} r^N (Np+1)^2 = r + O(\frac{\log N}{N}) \to r$ as $N \to \infty$

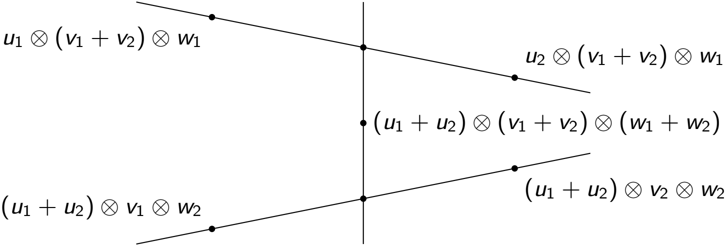
Bini-Capovani-Lotti-Romani algorithm

- [Bini *et al* 79] presents an approximate decomposition for $\langle 3,2,2\rangle$, $\underline{R} \leq 10$
- Later, [Bini 80] proves the statement connecting border rank to rank
- The actual decomposition is a bit complicated
- It is based (anachronistically) on the following [Landsberg & Ryder 15]
- Consider a W-like decomposition

$$\varepsilon(x \otimes v \otimes w + u \otimes y \otimes w + u \otimes v \otimes z) + o(\varepsilon^2) =$$

$$= (u + \varepsilon x) \otimes (v + \varepsilon y) \otimes (w + \varepsilon z) - u \otimes v \otimes w$$

- We can combine several decompositions of this form if several corresponding $u \otimes v \otimes w$ terms can be combined into one term



Border rank: additional results

- All results that we proved yesterday and at the start of the lecture only use algebraic properties of tensor operations and the relation $\leq$
- The degeneration relation shares all relevant properties

Theorem (Asymptotic sum inequality, Schönhage 81)

$$\underline{R}\Big(\bigoplus_{k=1}^{p} \langle \ell_k, m_k, n_k \rangle\Big) \leq r \Rightarrow \sum_{k=1}^{p} (\ell mn)^{\omega/3} \leq r$$

Schönhage's construction

- Consider matrix multiplication tensor $\langle n, 1, m \rangle$

$$\langle n, 1, m \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} e_i \otimes e_j \otimes e_{ij}$$

- We have $R(\langle n, 1, m \rangle) = nm$ by conciseness

Schönhage's construction

- Modify the decomposition

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(e_i + \varepsilon f_{ij}) \otimes (e_j + \varepsilon f_{ij}) \otimes (\varepsilon^2 e_{ij} + f) =$$

$$= (\sum_i e_i) \otimes (\sum_j e_j) \otimes f$$

$$+\varepsilon \sum_i e_i \otimes (\sum_j f_{ij}) \otimes f$$

$$+\varepsilon \sum_j (\sum_i f_{ij}) \otimes e_j \otimes f$$

$$+\varepsilon^2 \left[\langle n, 1, m \rangle + \sum_{ij} f_{ij} \otimes f_{ij} \otimes f \right] + \dots$$

Schönhage's construction

- Choose f_{ij} such that $\sum_i f_{ij} = 0$ and $\sum_j f_{ij} = 0$
- $(n-1)(m-1)$ linearly independent

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (e_i + \varepsilon f_{ij}) \otimes (e_j + \varepsilon f_{ij}) \otimes (\varepsilon^2 e_{ij} + f) - (\sum_i e_i) \otimes (\sum_j e_j) \otimes f =$$

$$= \varepsilon \sum_i e_i \otimes (\sum_j f_{ij}) \otimes f$$

$$+ \varepsilon \sum_j (\sum_i f_{ij}) \otimes e_j \otimes f$$

$$+ \varepsilon^2 \left[\langle n, 1, m \rangle + \sum_{ij} f_{ij} \otimes f_{ij} \otimes f \right] + \dots$$

- For every matrix M, the tensor $M \otimes v$ is equivalent to

$$\langle 1, p, 1 \rangle = \sum_{i=1}^{p} e_i \otimes e_i \otimes e \text{ where } p = \operatorname{rk} M$$

- If $\sum_i f_{ij} = 0$ and $\sum_j f_{ij} = 0$, then $\operatorname{rk} \sum_{ij} f_{ij} \otimes f_{ij} = (n-1)(m-1)$

Schönhage's upper bound

Theorem (Schönhage 81)

$$\underline{R}(\langle n, 1, m\rangle + \langle 1, (n-1)(m-1), 1\rangle) \leq nm + 1$$

- In fact this is an equality, by conciseness for border rank
- $\underline{R}(\langle 4, 1, 4\rangle + \langle 1, 9, 1\rangle) \leq 17$ gives $\omega < 2.55$

- [Coppersmith & Winograd 82] show how to improve the "scalar product" term $\langle 1, p, 1\rangle$ after taking Kronecker powers
- They get $\omega < 2.5$ from $\underline{R} < 10000$ for

$$\langle 81, 1, 81\rangle \oplus 2 \odot \langle 54, 4, 27\rangle \oplus \langle 36, 16, 9\rangle \oplus 2 \odot \langle 9, 46, 9\rangle \oplus 2 \odot \langle 6, 184, 3\rangle \oplus \langle 1, 3502, 1\rangle$$