Tensor decompositions and Theoretical computer science

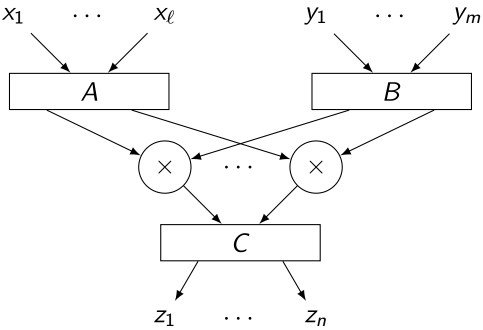
Lecture 2: Basic techniques for the complexity of matrix multiplication

Vladimir Lysikov

13.09.2022

Recall: bilinear circuits

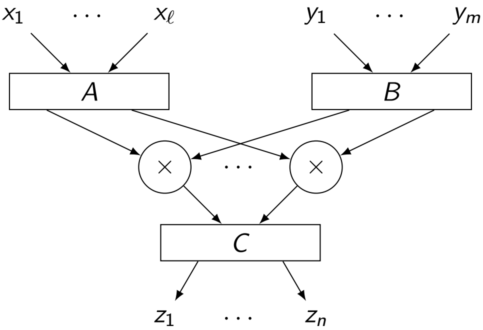
$T\colon X \times Y \to Z$ is a bilinear map. Let $\ell = \dim X$, $m = \dim Y$, $n = \dim Z$.



$$T(x, y) = \sum_{a=1}^{r} f_a(x) g_a(y) w_k$$

Recall: bilinear circuits

$T \colon X \times Y \to Z$ is a bilinear map. Let $\ell = \dim X$, $m = \dim Y$, $n = \dim Z$.



$$T = \sum_{a=1}^{r} f_a \otimes g_a \otimes w_a$$

Recall: Rank and circuit complexity

$R(T)$ — minimal number of proper multiplications in a bilinear circuit

$L(T)$ — number of algebraic operations required to compute $T(x, y)$

$$R(T) \leq L_{mlin}(T) \leq \mathrm{const} \cdot L(T)$$

From rank to circuit complexity

$T\colon X \times Y \to Z$ is a bilinear map, $\ell = \dim X$, $m = \dim Y$, $n = \dim Z$, $r = R(T)$.

$$T(x, y) = \sum_{a=1}^{r} f_a(x) g_a(y) w_k$$

- To compute one linear form $f_a(x)$, we need at most ℓ constant multiplications and $\ell - 1$ additions
- To compute r linear forms $f_1(x), \dots, f_r(x)$, we need $\approx 2\ell r$ linear operations
- Same for g — $2mr$ linear operations
- Then we multiply $f_a(x)$ by $g_a(y)$ — r proper multiplications
- To compute the output as linear combinations of these products, we need $2nr$ linear operations

$$\text{Total: } r + 2r(\ell + m + n)$$

Complexity of matrix multiplication

- Consider $n \times n$ matrix multiplication
- Let $R(n)$ be the rank of $n \times n$ matrix multiplication tensor
- $L(n)$ —- the circuit complexity of matrix multiplication

$$L(n) \leq R(n) + 2R(n)(n^2 + n^2 + n^2) = R(n) + 6n^2 R(n)$$

- Additions and constant multiplications dominate the cost

Complexity of matrix multiplication

- Consider $n \times n$ matrix multiplication
- Let $R(n)$ be the rank of $n \times n$ matrix multiplication tensor
- $L(n)$ —- the circuit complexity of matrix multiplication

$$L(n) \leq R(n) + 2R(n)(n^2 + n^2 + n^2) = R(n) + 6n^2 R(n)$$

- Additions and constant multiplications dominate the cost
- Bilinear algorithms also work work block matrix multiplication

$$\sum_{j=1}^{n} x_{ij} y_{jk} = \sum_{a=1}^{R_n} w_{aik} \left(\sum_{i,j=1}^{n} f_{aij} x_{ij} \right) \left(\sum_{j,k=1}^{n} g_{ajk} y_{jk} \right)$$

- Consider $nm \times nm$ matrix multiplication as multiplication of $n \times n$ block matrices with $m \times m$ blocks

$$L(nm) \leq R(n) \cdot L(m) + 6n^2 R(n) \cdot m^2$$

Complexity of matrix multiplication: Recurrence

$$L(nm) \leq R(n) \cdot L(m) + 6n^2 R(n) \cdot m^2$$

- If $R(n) > n^2$, then $L(m) = O(m^{\log_n R(n)})$
- If $R(n) = n^2$, then $L(m) = O(m^2 \log m)$

Definition

The *exponent of matrix multiplication* ω is defined as

$$\omega = \inf\{p \mid L(m) = O(m^p)\}$$

We have almost proven

Lemma

$$\omega \leq \log_n R(n)$$

Conciseness

- Let $T\colon U \times V \to W$ be a bilinear map of rank r

$$T = \sum_{a=1}^{r} f_a \otimes g_a \otimes w_a$$

$$T(x, y) = \sum_{a=1}^{r} f_a(x) g_a(y) w_a$$

- If $r < \dim W$, then

$$\operatorname{im} T \subset \operatorname{span}(w_1, \dots, w_r) \subsetneq W$$

- Thus, if $\operatorname{im} T = W$, then $r \geq \dim W$
- Note that $\operatorname{im} T$ stays the same if we consider it as a linear map $T\colon U \otimes V \to W$

Definition

A tensor $T \in U \otimes V \otimes W$ is *concise* in the third factor if the corresponding map $T\colon U^* \otimes V^* \to W$ is surjective

Exponent of matrix multiplication

Definition

The *exponent of matrix multiplication* ω is defined as

$$\omega = \inf\{p \mid L(m) = O(m^p)\}$$

Lemma

$$\omega \leq \log_n R(n)$$

Theorem

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

- If $R(n) \leq Cn^p$, then $\omega \leq \log_n R(n) \leq p + O(\frac{1}{\log n})$
- If $\omega \leq p$, then $R(n) \leq \text{const}\, L(n) = O(n^p)$

Matrix multiplication tensor

- Consider multiplication of an $\ell \times m$ matrix by an $m \times n$ matrix

$$xy = \sum_{i=1}^{\ell} \sum_{k=1}^{n} e_{ik} \sum_{j=1}^{m} x_{ij} y_{jk}$$

- As a tensor, this matrix multiplication is

$$\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} e_{ij}^* \otimes e_{jk}^* \otimes e_{ik} \in (\mathbf{k}^{\ell \times m})^* \otimes (\mathbf{k}^{m \times n})^* \otimes \mathbf{k}^{\ell \times n}$$

- It is more convenient to work with an equivalent tensor obtained by applying some isomorphisms to the factors

$$\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} e_{ij} \otimes e_{jk} \otimes e_{ki} \in \mathbf{k}^{\ell \times m} \otimes \mathbf{k}^{m \times n} \otimes \mathbf{k}^{n \times \ell}$$

- This tensor is often denoted by $\langle \ell, m, n \rangle$

Strassen algorithm

$$
\begin{aligned}
\langle 2,2,2\rangle &= (e_{11}+e_{22}) \otimes (e_{11}+e_{22}) \otimes (e_{11}+e_{22}) \\
&+ (e_{21}+e_{22}) \otimes e_{11} \otimes (e_{12}-e_{22}) \\
&+ (e_{12}-e_{22}) \otimes (e_{21}+e_{22}) \otimes e_{11} \\
&+ e_{11} \otimes (e_{12}-e_{22}) \otimes (e_{21}+e_{22}) \\
&+ (e_{12}+e_{11}) \otimes e_{22} \otimes (e_{21}-e_{11}) \\
&+ (e_{21}-e_{11}) \otimes (e_{12}+e_{11}) \otimes e_{22} \\
&+ e_{22} \otimes (e_{21}-e_{11}) \otimes (e_{12}+e_{11})
\end{aligned}
$$

$$\omega \leq \log_2 7 < 2.81$$

Strassen algorithm

- Take a 2×2 matrix D with such that $D^3 = \mathbf{1}$, for example, $D = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$
- Its characteristic polynomial is $\lambda^2 + \lambda + 1$, so $D^{-1} + D + \mathbf{1} = D^2 + D + 1 = 0$
- Choose a vector x which is not an eigenvector of D, for example, $x = [1\ 0]^T$
- Take a vector y such that $y^T x = 0$ and normalize it so that $y^T D x = 1$
- We have $y^T D^{-1} x = -y^T D x - y^T x = -1$
- The matrix $M = xy^T$ is traceless because $\operatorname{tr} M = y^T x = 0$
- $MDM = xy^T D x y^T = xy^T = M$
- $MD^{-1}M = xy^T D^{-1} x y^T = -xy^T = -M$

Strassen algorithm

Decompose

$$X = x_1 D \quad + x_2 M + x_3 D^{-1} M D + x_4 D M D^{-1}$$
$$Y = y_1 D^{-1} + y_2 M + y_3 D^{-1} M D + y_4 D M D^{-1}$$

and multiply according to the multiplication table

	D^{-1}	M	$D^{-1}MD$	DMD^{-1}
D	$\mathbf{1}$	DM	MD	$D^{-1}MD^{-1}$
M	MD^{-1}	0	$MD^{-1}MD = -MD$	$MDMD^{-1} = MD^{-1}$
$D^{-1}MD$	$D^{-1}M$	$D^{-1}MDM = D^{-1}M$	0	$D^{-1}MD^{-1}MD^{-1} = -D^{-1}MD^{-1}$
DMD^{-1}	DMD	$DMD^{-1}M = -DM$	$DMDMD = DMD$	0

Kronecker products

- To express the trick with block matrix multiplication in tensor language, we need to define Kronecker products

Kronecker products

- To express the trick with block matrix multiplication in tensor language, we need to define Kronecker products

- Abstract definition of the Kronecker product:
- Let $T \in U \otimes V \otimes W$ and $S \in U' \otimes V' \otimes W'$ be two tensors.
- Form the tensor product $T \otimes S \in (U \otimes V \otimes W) \otimes (U' \otimes V' \otimes W')$
- Then apply the natural isomorphism

$$(U \otimes V \otimes W) \otimes (U' \otimes V' \otimes W') \cong (U \otimes U') \otimes (V \otimes V') \otimes (W \otimes W')$$

- To get a (trilinear) tensor

$$T \boxtimes S \in (U \otimes U') \otimes (V \otimes V') \otimes (W \otimes W')$$

Kronecker products

- To express the trick with block matrix multiplication in tensor language, we need to define Kronecker products

- Kronecker product in coordinates:
- Let

$$T = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} t_{ijk} e_i \otimes e_j \otimes e_k$$

$$S = \sum_{i=1}^{\ell'} \sum_{j=1}^{m'} \sum_{k=1}^{n'} s_{ijk} e_i \otimes e_j \otimes e_k$$

- Then

$$T \boxtimes S = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{i'=1}^{\ell'} \sum_{j'=1}^{m'} \sum_{k'=1}^{n'} (t_{ijk} \cdot s_{i'j'k'}) e_{ii'} \otimes e_{jj'} \otimes e_{kk'}$$

Kronecker product: Computational interpretation

- Let $T\colon U \times V \to W$ and $S\colon U' \times V' \to W'$ be bilinear maps

$$T(u, v) = \sum_{ijk} t_{ijk} u_i v_j e_k$$

$$S(x, y) = \sum_{ijk} s_{ijk} x_i y_j e_k$$

- Consider elements of $U \otimes U'$
- They have the form $\sum_i e_i \otimes x_i$, so they consists of several blocks $x_i \in U'$
- Same for $V \otimes V'$ and $W \otimes W'$
- For the Kronecker product

$$[T \boxtimes S](\sum_i e_i \otimes x_i, \sum_j e_j \otimes y_j) = \sum_{ijk} t_{ijk}(e_k \otimes S(x_i, y_j))$$

- It operates blockwise like T, and the blocks are combined using S

Kronecker product: properties

Lemma

$$\langle \ell_1, m_1, n_1 \rangle \boxtimes \langle \ell_2, m_2, n_2 \rangle = \langle \ell_1 \ell_2, m_1 m_2, n_1 n_2 \rangle$$

Lemma

$$R(T \boxtimes S) \leq R(T) \cdot R(S)$$

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

$$S = \sum_{b=1}^{r'} x_b \otimes y_b \otimes z_b$$

$$\Rightarrow T \boxtimes S = \sum_{a=1}^{r} \sum_{b=1}^{r'} (u_a \otimes x_b) \otimes (v_a \otimes y_b) \otimes (w_a \otimes z_b)$$

Matrix multiplication: rectangular blocks

Lemma

$$R(\langle \ell, m, n\rangle) = R(\langle m, n, \ell\rangle)$$

$$\langle \ell, m, n\rangle = \sum_{a=1}^{r} u_a \otimes v_a \otimes w_a$$

$$\Rightarrow \langle m, n, \ell\rangle = \sum_{a=1}^{r} v_a \otimes w_a \otimes u_a$$

Theorem

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

$$R(\langle \ell mn, \ell mn, \ell mn\rangle) \leq R(\langle \ell, m, n\rangle) R(\langle m, n, \ell\rangle) R(\langle n, \ell, m\rangle) = r^3$$

$$\omega \leq \log_{\ell mn} r^3 = 3 \log_{\ell mn} r$$

Direct sum of tensors

- Let $T\colon U \times V \to W$ and $S\colon U' \times V' \to W'$ be two bilinear maps
- Their direct sum is defined as $T \oplus S\colon (U \oplus U') \times (V \oplus V') \to (W \oplus W')$

$$[T \oplus S]((u,x),(v,y)) = (T(u,v), S(x,y))$$

- Abstractly

$$(U \oplus U') \otimes (V \oplus V') \otimes (W \otimes W') = (U \otimes V \otimes W) \oplus (U' \otimes V' \otimes W') \oplus \dots$$

- We inject tensors $T \in U \otimes V \otimes W$ and $S \in U' \otimes V' \otimes W'$ into the tensor product of sums and add them
- Rank is subadditive with respect to direct sums

$$R(T \oplus S) < R(T) + R(S)$$

- [Shitov 17] shows that the inequality can be strict

Matrix multiplication: several blocks

- Suppose we have a decomposition which allows us to multiply many pairs of matrices simultaneously

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

- Can we construct from this another decomposition for multiplication of large matrices?
- Computationally, we need to dissect a large matrix into appropriate blocks and multiply them using an existing decomposition
- Focus on the case of uniform blocks for now
- Denote $p \odot T = T \oplus T \oplus \dots \oplus T$ (p times)

Theorem

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

Another perspective on tensor rank

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

- Define a map $A\colon \mathbf{k}^r \to U$ by $Ae_a = u_a$
- Similarly $Be_a = v_a$ and $Ce_a = w_a$

$$T = \sum_{a=1}^{r} u_a \otimes v_a \otimes w_a = \sum_{a=1}^{r} (Ae_a) \otimes (Be_a) \otimes (Ce_a) = (A \otimes B \otimes C) \sum_{a=1}^{r} e_a \otimes e_a \otimes e_a$$

- Denote $E_r = \sum_{a=1}^{r} e_a \otimes e_a \otimes e_a \in \mathbf{k}^{r \times r \times r}$

Another perspective on tensor rank

Definition

A tensor $T \in U \otimes V \otimes W$ is a *restriction* of a tensor $S \in U' \otimes V' \otimes W'$ (notation $T \leq S$) if

$$T = (A \otimes B \otimes C)S$$

for some linear maps $A\colon U' \to U$, $B\colon V' \to V$, $C\colon W' \to W$

Tensors T and S are *equivalent* if $T \leq S$ and $S \leq T$

Theorem

$$R(T) \leq r \Leftrightarrow T \leq E_r$$

Properties

$$T \leq T \oplus S$$

$$S \neq 0 \Rightarrow T \leq T \boxtimes S$$

$$T_1 \leq T_2 \Rightarrow T_1 \boxtimes S \leq T_2 \boxtimes S$$

$$T_1 \leq T_2 \Rightarrow T_1 \oplus S \leq T_2 \oplus S$$

$$E_r \boxtimes T \cong r \odot T$$

$$E_r \boxtimes E_{r'} \cong E_{rr'}, E_r \oplus E_{r'} \cong E_{r+r'}$$

Simple asymptotic sum inequality for rank

Theorem

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

- Denote $R(m) = R(\langle m, m, m \rangle)$

$$\langle nm, nm, nm \rangle = \langle n, n, n \rangle \boxtimes \langle m, m, m \rangle \leq \langle n, n, n \rangle \boxtimes E_{R(m)} \cong R(m) \odot \langle n, n, n \rangle \leq$$

$$\leq (\lceil \frac{R(m)}{p} \rceil p) \odot \langle n, n, n \rangle \cong \lceil \frac{R(m)}{p} \rceil \odot (p \odot \langle n, n, n \rangle) \leq \lceil \frac{R(m)}{p} \rceil \odot E_r$$

$$R(nm) \leq \lceil \frac{R(m)}{p} \rceil r \leq R(m) \frac{r}{p} + r$$

$$R(m) = O(m^{\log_n \frac{r}{p}})$$

$$\omega \leq \log_n \frac{r}{p}$$

Markus Bläser. *Fast Matrix Multiplication*,
Theory of Computing Library Graduate Surveys 5

https://theoryofcomputing.org/articles/gs005/