Fast Matrix Multiplication = Calculating Tensor Rank

Caltech Math 10 Presentation

Chinmay Nirkhe

November 15th, 2016

Statement of the Problem

Question

Given two matrices $A, B \in \mathbb{C}^{n \times n}$, what is the algorithmic complexity of calculating $C = AB$?

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

Assume that arithmetic operations (i.e. $+, \times$) of two values in $\mathbb{C}$ is constant time ($O(1)$).

Big-O Notation

Definition (Big-O)

For functions $f, g : \mathbb{N} \to \mathbb{N}$, we say $f \in O(g)$ (or $f = O(g)$) if $\exists\ c > 0$ s.t. $f(n) \leq cg(n)$ for all but finitely many n.

A naïve $O(n^3)$ algorithm

Calculating each C_{ij} can be done in $O(n)$ computations by

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

n^2 such values $C_{ij} \implies n^2 \cdot O(n) = O(n^3)$.

Definition of ω

Definition (Coefficient of Matrix Multiplication)

We define the *coefficient of matrix multiplication* ω as

$$\omega = \inf \left\{ \alpha \text{ s.t. } \forall \epsilon > 0, \exists \text{ alg. of complexity } O(n^{\alpha+\epsilon}) \right\}.$$

Definition of ω

Definition (Coefficient of Matrix Multiplication)

We define the *coefficient of matrix multiplication* ω as

$$\omega = \inf \left\{ \alpha \text{ s.t. } \forall \epsilon > 0, \exists \text{ alg. of complexity } O(n^{\alpha+\epsilon}) \right\}.$$

Remark (Trivial Bounds)

$2 \leq \omega \leq 3$.

We saw an $O(n^3)$ algorithm. C has n^2 entries so it takes at least $O(n^2)$ computations to just write C.

Goal revisited

The goal is to (ideally) show $\omega = 2$. The latest result by Coppersmith-Winnograd is $\omega \leq 2.3728639$.

How do we improve ω?

Start by considering $A, B \in \mathbb{C}^{2\times 2}$. Then,

$$\begin{cases} C_{11} &= A_{11}B_{11} + A_{12}B_{21} \\ C_{12} &= A_{11}B_{12} + A_{12}B_{22} \\ C_{21} &= A_{21}B_{11} + A_{22}B_{21} \\ C_{22} &= A_{21}B_{12} + A_{22}B_{22} \end{cases}$$

Clearly there are 8 multiplications terms here. We are going to reduce this to 7 multiplications.

7 multiplications

$$
\begin{aligned}
M_1 &= (A_{11} + A_{22})(B_{11} + B_{22}) \\
M_2 &= (A_{21} + A_{22})B_{11} \\
M_3 &= A_{11}(B_{12} - B_{22}) \\
M_4 &= A_{22}(B_{21} - B_{11}) \\
M_5 &= (A_{11} + A_{12})B_{22} \\
M_6 &= (A_{21} - A_{11})(B_{11} + B_{12}) \\
M_7 &= (A_{12} - A_{22})(B_{21} + B_{22})
\end{aligned}
$$

$$
\begin{cases}
C_{11} &= M_1 + M_4 - M_5 + M_7 \\
C_{12} &= M_3 + M_5 \\
C_{21} &= M_2 + M_4 \\
C_{22} &= M_1 - M_2 + M_3 + M_6
\end{cases}
$$

A bound on ω due to 7 multiplications and recursion

Consider matrices $A, B \in \mathbb{C}^{2^m \times 2^m}, m \in \mathbb{N}$. Write $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ for $A_{ij} \in \mathbb{C}^{2^{m-1} \times 2^{m-1}}$ and B similarly. Use the 7 multiplication trick to recursively multiply A and B.

A bound on ω due to 7 multiplications and recursion

Consider matrices $A, B \in \mathbb{C}^{2^m \times 2^m}, m \in \mathbb{N}$. Write $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ for $A_{ij} \in \mathbb{C}^{2^{m-1} \times 2^{m-1}}$ and B similarly. Use the 7 multiplication trick to recursively multiply A and B.

Let $f(m)$ be the complexity of multiplying $A, B \in \mathbb{C}^{2^m \times 2^m}$.

$$f(m) = 7f(m-1) + c_0(2^m \times 2^m)$$

A bound on ω due to 7 multiplications and recursion

Consider matrices $A, B \in \mathbb{C}^{2^m \times 2^m}, m \in \mathbb{N}$. Write $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ for $A_{ij} \in \mathbb{C}^{2^{m-1} \times 2^{m-1}}$ and B similarly. Use the 7 multiplication trick to recursively multiply A and B.

Let $f(m)$ be the complexity of multiplying $A, B \in \mathbb{C}^{2^m \times 2^m}$.

$$f(m) = 7f(m-1) + c_0(2^m \times 2^m)$$

$$f(m) = O(7^m)$$

A bound on ω due to 7 multiplications and recursion

Let $N = 2^m$, then $f(m) = O(N^{\log_2 7})$.

Theorem (Strassen's)

$\omega \leq \log_2 7 \approx 2.8074$

Can we multiply $\mathbb{C}^{2\times 2}$ with < 7 multiplications?

Can we multiply $\mathbb{C}^{2\times 2}$ with < 7 multiplications?

No.

Can we multiply $\mathbb{C}^{2\times 2}$ with < 7 multiplications?

No.

Can we find a simpler way to multiply a 3×3 matrix using k multiplications s.t. $\log_3 k < \log_2 7$?

Can we multiply $\mathbb{C}^{2\times 2}$ with < 7 multiplications?

No.

Can we find a simpler way to multiply a 3×3 matrix using k multiplications s.t. $\log_3 k < \log_2 7$?

No.

The Matrix Multiplication Tensor

Definition (The Matrix Multiplication Tensor)

For fixed n, the *matrix multiplication tensor* $T \in \mathbb{C}^{nm \times mp \times pn}$ defined by

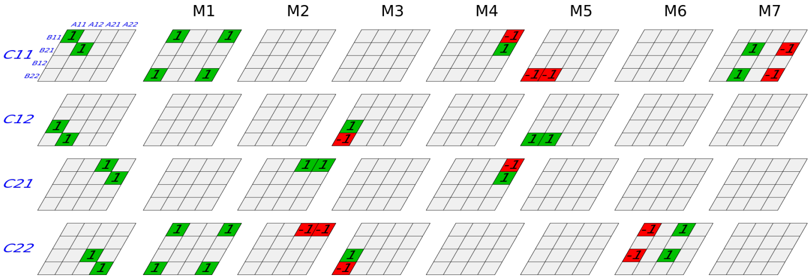
$$T_{ij',jk',ki'} = \mathbf{1}_{\{i=i',j=j',k=k'\}}$$

for $i, i', j, j', k, k' \in \{1, \ldots, n\}$.

Notation

The tensor is often notated as $\langle n, m, p \rangle$.

M.M. Tensor of $\langle 2, 2, 2\rangle$



Strassen's equations

$$
\begin{aligned}
M_1 &= (A_{11} + A_{22})(B_{11} + B_{22}) \\
M_2 &= (A_{21} + A_{22})B_{11} \\
M_3 &= A_{11}(B_{12} - B_{22}) \\
M_4 &= A_{22}(B_{21} - B_{11}) \\
M_5 &= (A_{11} + A_{12})B_{22} \\
M_6 &= (A_{21} - A_{11})(B_{11} + B_{12}) \\
M_7 &= (A_{12} - A_{22})(B_{21} + B_{22})
\end{aligned}
$$

Rank of a Tensor

Definition (Rank 1 Tensor)

A tensor $T \in \mathbb{C}^{nm \times mp \times np}$ is a rank 1 tensor if it can be expressed as $T = u \otimes v \otimes w$ where $u \in \mathbb{C}^{nm}, v \in \mathbb{C}^{mp}, w \in \mathbb{C}^{np}$. i.e.

$$T_{ijk} = u_i v_j w_k$$

Rank of a Tensor

Definition (Rank 1 Tensor)

A tensor $T \in \mathbb{C}^{nm \times mp \times np}$ is a rank 1 tensor if it can be expressed as $T = u \otimes v \otimes w$ where $u \in \mathbb{C}^{nm}, v \in \mathbb{C}^{mp}, w \in \mathbb{C}^{np}$. i.e.

$$T_{ijk} = u_i v_j w_k$$

Definition (Tensor Rank)

For an arbitrary tensor T, the rank of T (not. $R(T)$) is the minimum number of rank 1 tensors summing to T.

$R(\langle 2,2,2\rangle) \leq 7.$

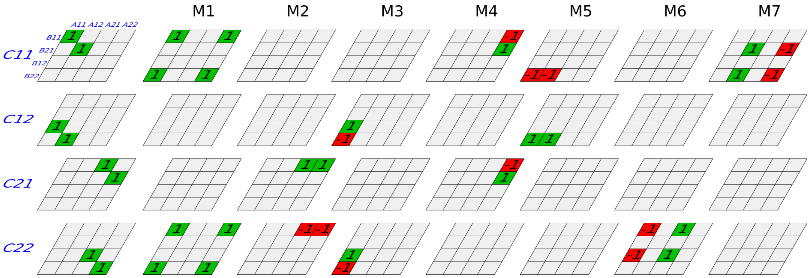
$$
\begin{aligned}
T_1 &= (1,0,0,1) \otimes (1,0,0,1) \otimes (1,0,0,1) \\
T_2 &= (0,0,1,1) \otimes (1,0,0,0) \otimes (0,0,1,-1) \\
T_3 &= (1,0,0,0) \otimes (0,1,0,-1) \otimes (0,1,0,1) \\
T_4 &= (0,0,0,1) \otimes (-1,0,1,0) \otimes (1,0,1,0) \\
T_5 &= (1,1,0,0) \otimes (0,0,0,1) \otimes (-1,1,0,0) \\
T_6 &= (-1,0,1,0) \otimes (1,1,0,0) \otimes (0,0,0,1) \\
T_7 &= (0,1,0,-1) \otimes (0,0,1,1) \otimes (1,0,0,0)
\end{aligned}
$$

Strassen's equations

$$
\begin{aligned}
M_1 &= (A_{11} + A_{22})(B_{11} + B_{22}) \\
M_2 &= (A_{21} + A_{22})B_{11} \\
M_3 &= A_{11}(B_{12} - B_{22}) \\
M_4 &= A_{22}(B_{21} - B_{11}) \\
M_5 &= (A_{11} + A_{12})B_{22} \\
M_6 &= (A_{21} - A_{11})(B_{11} + B_{12}) \\
M_7 &= (A_{12} - A_{22})(B_{21} + B_{22})
\end{aligned}
$$

$$
\begin{cases}
C_{11} & = M_1 + M_4 - M_5 + M_7 \\
C_{12} & = M_3 + M_5 \\
C_{21} & = M_2 + M_4 \\
C_{22} & = M_1 - M_2 + M_3 + M_6
\end{cases}
$$

M.M. Tensor of $\langle 2, 2, 2 \rangle$



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

Theorem

For any $n > 2$, if $r = R(\langle n, n, n \rangle)$ where $\langle n, n, n \rangle$ is the mat.mult. tensor for $n \times n$ matrices, then $\omega \leq \log_n r$

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

Theorem

For any $n > 2$, if $r = R(\langle n, n, n \rangle)$ where $\langle n, n, n \rangle$ is the mat.mult. tensor for $n \times n$ matrices, then $\omega \leq \log_n r$

Let $T = \sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$ be the rank decomposition of T.

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

Theorem

For any $n > 2$, if $r = R(\langle n, n, n \rangle)$ where $\langle n, n, n \rangle$ is the mat.mult. tensor for $n \times n$ matrices, then $\omega \leq \log_n r$

Let $T = \sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$ be the rank decomposition of T. Divide matrices $A, B \in \mathbb{C}^{n^m \times n^m}$ into $n \times n$ equal parts $A_{11}, \ldots, A_{nn}$ and $B_{11}, \ldots, B_{nn}$.

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

Theorem

For any $n > 2$, if $r = R(\langle n, n, n\rangle)$ where $\langle n, n, n\rangle$ is the mat.mult. tensor for $n \times n$ matrices, then $\omega \leq \log_n r$

Let $T = \sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$ be the rank decomposition of T. Divide matrices $A, B \in \mathbb{C}^{n^m \times n^m}$ into $n \times n$ equal parts $A_{11}, \ldots, A_{nn}$ and $B_{11}, \ldots, B_{nn}$. Compute (recursively)

$$M_\ell = \left(\sum_{i,j=1}^{n,n} u_\ell^{(ij)} A_{ij}\right)\left(\sum_{i,j=1}^{n,n} v_\ell^{(ij)} B_{ij}\right)$$

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

Theorem

For any $n > 2$, if $r = R(\langle n, n, n \rangle)$ where $\langle n, n, n \rangle$ is the mat.mult. tensor for $n \times n$ matrices, then $\omega \leq \log_n r$

Let $T = \sum_{\ell=1}^{r} u_\ell \otimes v_\ell \otimes w_\ell$ be the rank decomposition of T. Divide matrices $A, B \in \mathbb{C}^{n^m \times n^m}$ into $n \times n$ equal parts $A_{11}, \ldots, A_{nn}$ and $B_{11}, \ldots, B_{nn}$. Compute (recursively)

$$M_\ell = \left(\sum_{i,j=1}^{n,n} u_\ell^{(ij)} A_{ij} \right) \left(\sum_{i,j=1}^{n,n} v_\ell^{(ij)} B_{ij} \right)$$

$$C_{ij} = \sum_{\ell=1}^{r} w_\ell M_\ell$$

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

Calculation of M takes r multiplications of matrices of size $n^{m-1} \times n^{m-1}$ + $c_0 N^2$ additions.

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

Calculation of M takes r multiplications of matrices of size $n^{m-1} \times n^{m-1}$ + $c_0 N^2$ additions.

$$f(m) = rf(m-1) + c_0(n^m \times n^m)$$

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

Calculation of M takes r multiplications of matrices of size $n^{m-1} \times n^{m-1} + c_0 N^2$ additions.

$$f(m) = rf(m-1) + c_0(n^m \times n^m)$$

$$f(m) = O(r^m) = O(n^{(\log_n r)m}) \implies$$

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

Calculation of M takes r multiplications of matrices of size $n^{m-1} \times n^{m-1} + c_0 N^2$ additions.

$$f(m) = rf(m-1) + c_0(n^m \times n^m)$$

$$f(m) = O(r^m) = O(n^{(\log_n r)m}) \implies \boxed{\omega \leq \log_2^m}$$

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

By the construction of the tensor,

$$C_{ab} = \sum_{i,i',j,j'} T_{(i,j)(i',j')(a,b)} A_{ij} B_{i'j'}$$

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

By the construction of the tensor,

$$\begin{aligned} C_{ab} &= \sum_{i,i',j,j'} T_{(i,j)(i',j')(a,b)} A_{ij} B_{i'j'} \\ &= \sum_{\ell=1}^{r} \sum_{i,i',j,j'} T^{(\ell)}_{(i,j)(i',j')(a,b)} A_{ij} B_{i'j'} \end{aligned}$$

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

By the construction of the tensor,

$$\begin{aligned} C_{ab} &= \sum_{i,i',j,j'} T_{(i,j)(i',j')(a,b)} A_{ij} B_{i'j'} \\ &= \sum_{\ell=1}^{r} \sum_{i,i',j,j'} T^{(\ell)}_{(i,j)(i',j')(a,b)} A_{ij} B_{i'j'} \\ &= \sum_{\ell=1}^{r} w_\ell^{(ab)} \left(\sum_{i,j=1}^{n,n} u_\ell^{(ij)} A_{ij} \right) \left(\sum_{i,j=1}^{n,n} v_\ell^{(ij)} B_{ij} \right) \\ &= \sum_{\ell=1}^{r} w_\ell^{(ab)} M_\ell \end{aligned}$$

Does this help reduce ω bound?

Theorem

For $n > 2$, $\omega \leq \log_n R(\langle n, n, n \rangle)$.

The only other small n found for which this gives a stronger bound than Strassen's is $n = 70$, where $R = 143,640$ so $\omega \leq 2.80$.

Does this help reduce ω bound?

Theorem

For $n > 2$, $\omega \leq \log_n R(\langle n, n, n \rangle)$.

The only other small n found for which this gives a stronger bound than Strassen's is $n = 70$, where $R = 143,640$ so $\omega \leq 2.80$.

What about non cubic-tensors?

The Matrix Multiplication Tensor

Definition (The Matrix Multiplication Tensor)

For fixed n, the *matrix multiplication tensor* $T \in \mathbb{C}^{nm \times mp \times pn}$ defined by

$$T_{ij',jk',ki'} = \mathbf{1}_{\{i=i',j=j',k=k'\}}$$

for $i, i', j, j', k, k' \in \{1, \ldots, n\}$.

Notation

The tensor is often notated as $\langle n, m, p \rangle$.

Bounding Non-Cubic Tensors

Lemma (Permutation invariant)

$R(\langle n, m, p\rangle) = R(\langle \pi(n), \pi(m), \pi(p)\rangle)$ *for any* $\pi \in S_3$.

Bounding Non-Cubic Tensors

Lemma (Permutation invariant)

$R(\langle n, m, p\rangle) = R(\langle \pi(n), \pi(m), \pi(p)\rangle)$ for any $\pi \in S_3$.

Take the decomposition into rank 1 tensors and permute $(u, v, w) \mapsto (\pi(u), \pi(v), \pi(w))$ for each rank 1 tensor. □

Bounding Non-Cubic Tensors

Lemma (Tensor Product Upper bound)

$$\begin{aligned} R(\langle nn', mm', pp' \rangle) &= R(\langle n, m, p \rangle \otimes \langle n', m', p' \rangle) \\ &\leq R(\langle n, m, p \rangle) R(\langle n', m', p' \rangle). \end{aligned}$$

Bounding Non-Cubic Tensors

Lemma (Tensor Product Upper bound)

$$\begin{aligned} R(\langle nn', mm', pp' \rangle) &= R(\langle n, m, p \rangle \otimes \langle n', m', p' \rangle) \\ &\leq R(\langle n, m, p \rangle) R(\langle n', m', p' \rangle). \end{aligned}$$

Think of $\langle nn', mm', pp' \rangle$ as deconstruction into a tensor of tensors.

Bounding Non-Cubic Tensors

Lemma (Tensor Product Upper bound)

$$\begin{aligned} R(\langle nn', mm', pp' \rangle) &= R(\langle n, m, p \rangle \otimes \langle n', m', p' \rangle) \\ &\leq R(\langle n, m, p \rangle) R(\langle n', m', p' \rangle). \end{aligned}$$

Think of $\langle nn', mm', pp' \rangle$ as deconstruction into a tensor of tensors.

Let $\langle n, m, p \rangle = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$, $\langle n', m', p' \rangle = \sum_{j=1}^{r'} u'_j \otimes v'_j \otimes w'_j$. Then

$$\langle nn', mm', pp' \rangle = \sum_{i,j=1}^{r,r'} (u_i \otimes u'_j) \otimes (v_i \otimes v'_j) \otimes (w_i \otimes w'_j)$$

General Tensor Bound

Theorem (General Bound)

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

General Tensor Bound

Theorem (General Bound)

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

$$\begin{aligned} R(\langle nmp, nmp, nmp \rangle) &\leq R(\langle n, m, p \rangle) R(\langle m, p, n \rangle) R(\langle p, n, m \rangle) \\ &\leq R(\langle n, m, p \rangle)^3 \end{aligned}$$

General Tensor Bound

Theorem (General Bound)

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

$$\begin{aligned} R(\langle nmp, nmp, nmp \rangle) &\leq R(\langle n, m, p \rangle) R(\langle m, p, n \rangle) R(\langle p, n, m \rangle) \\ &\leq R(\langle n, m, p \rangle)^3 \end{aligned}$$

$$\omega \leq \log_{nmp} R(\langle n, m, p \rangle)^3 = 3 \log_{nmp} R(\langle n, m, p \rangle)$$

□

The Quest for Tensor Rank and other methods

Computational bashing! People have tried to find tensors but non-trivial in general and no serious improvements have been achieved.

Border Rank

Let $T^{(k)} \in (\mathbb{C}[\epsilon^k])^{nm \times mp \times np}$ be a tensor such that as $\epsilon \to 0$, $T^{(k)} = \epsilon^k T + O(\epsilon^{k+1})$. Then, define $\underline{R}(T) = \min_k R(T^{(k)})$.

Border Rank

Let $T^{(k)} \in (\mathbb{C}[\epsilon^k])^{nm \times mp \times np}$ be a tensor such that as $\epsilon \to 0$, $T^{(k)} = \epsilon^k T + O(\epsilon^{k+1})$. Then, define $\underline{R}(T) = \min_k R(T^{(k)})$.

Theorem

$$\omega \leq 3 \log_{nmp} \underline{R}(\langle n, m, p \rangle)$$

Theorem

$$\underline{R}(\langle 2, 2, 3 \rangle) = 10 \implies \omega \leq 2.79$$

Group Theoretic Approach

Recently, the biggest push in this subject and the way that Coppersmith and Winnograd achieved their lower bound was to find the matrix multiplication tensor embedded within the Cayley table of a group G and use properties of generators on the group to bound ω. But for another day...