Defence

Group-theoretical Method of Matrix Multiplication

Jiayue Qi

Supervisor: Xiao-Shan Gao

2017.05.24

Contents

Matrix multiplication exponent ω

Definition (matrix multiplication exponent ω)

The matrix multiplication exponent ω is the smallest real number ω for which $n \times n$ matrix multiplication can be performed in $O(n^{\omega+\varepsilon})$ operations for each $\varepsilon > 0$.

It is clear: $2 \leq \omega \leq 3$

A Major Conjecture: $\omega = 2$.

Strassen's algorithm

Let $A, B, C \in R^{2^n \times 2^n}$.

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{1}$$

Let

$$\begin{aligned} M_1 &:= (A_{11} + A_{22})(B_{11} + B_{22}) \\ M_2 &:= (A_{21} + A_{22})B_{11} \\ M_3 &:= A_{11}(B_{12} - B_{22}) \\ M_4 &:= A_{22}(B_{21} - B_{11}) \\ M_5 &:= (A_{11} + A_{12})B_{22} \\ M_6 &:= (A_{21} - A_{11})(B_{11} + B_{12}) \\ M_7 &:= (A_{12} - A_{22})(B_{21} + B_{22}) \end{aligned} \tag{2}$$

Strassen's algorithm

$C_{11}, C_{12}, C_{21}, C_{22}$ can be obtained from M_i by additions.
Then we only need 7 multiplication operations in each step!
We repeat this step n times till the sub-matrix becomes number.

Denote $f(n)$ as the total number of calculations for multiplying two $2^n \times 2^n$ matrices.

$$f(n+1) = 7f(n) + l \cdot 4^n,$$

where l is the number of additions in one step of the algorithm.
Thus,

$$f(n) = (7 + o(1))^n,$$

then for two $N = 2^n$ matrices, the asymptotic complexity of Strassen's algorithm is:

$$O([7 + o(1)]^n) = O(N^{\log_2 7 + o(1)}) \approx O(N^{2.8074}).$$

History of the complexity of matrix multiplication

- Volker Strassen, 1969, $\omega \leq 2.8074$.
- Don Coppersmith, Shmuel Winograd, 1990, tensor algorithm $\omega \leq 2.375477$. (CW1990)
- Andrew Stothers, 2010, improve CW90 algorithm, $\omega \leq 2.374$.
- Virginia Williams, 2011, $\omega \leq 2.3728642$.
- Francois Le Gall, 2014, simplify Williams' algorithm, $\omega \leq 2.3728639$.

History of the complexity of matrix multiplication

- Henry Cohn, Robert Kleinberg, Balazs Szegedy, Chris Umans, 2005, the Group-theoretical Method of Matrix Multiplication, two conjectures $\implies \omega = 2$, best bound: $\omega \leq 2.41$.
- Andris Ambainis, Yuval Filmus, Francois Le Gall, 2015, "the framework of analyzing higher and higher tensor powers of a certain identity of Coppersmith and Winograd cannot result in an algorithm within running time $O(n^{2.3725})$ thus cannot prove $\omega = 2$".
- Hence the main topic of this thesis is the group-theoretical method of matrix multiplication.

Contents

Group Method of Matrix Multiplication: Notions

$\mathbb{C}$: the field of complex numbers.

- The group algebra $\mathbb{C}[G]$ of a finite group G decomposes as the direct product $\mathbb{C}[G] \cong \mathbb{C}^{d_1 \times d_1} \times \ldots \times \mathbb{C}^{d_k \times d_k}$ of matrix algebras of orders $d_1, \ldots, d_k$. These orders are the character degrees of G.
- If we compute the dimensions of both sides, we have $|G| = \sum_i d_i^2$.
- If G has an abelian subgroup A, then all the character degrees of G are less than or equal to the index $[G : A]$.

Group Method of Matrix Multiplication: Notions

- If S is a subset of a group, let $Q(S)$ denote the right quotient set of S,i.e., $Q(S) = s_1 s_2^{-1} : s_1, s_2 \in S$.

Definition (*double product property*)

We say that subsets S_1, S_2 of a group H satisfy the *double product property if*
$q_1 q_2 = 1$ *implies* $q_1 = q_2 = 1$, *where* $q_i \in Q(S_i)$.

Definition

A group realizes $\langle n_1, n_2, n_3 \rangle$ if there are subsets $S_1, S_2, S_3 \subseteq G$ such that $|S_i| = n_i$, and for $q_i \in Q(S_i)$, if $q_1 q_2 q_3 = 1$ then $q_1 = q_2 = q_3 = 1$. We call this condition on S_1, S_2, S_3 the **triple product property**.

Group-theoretical Method of Matrix Multiplication

Suppose G realizes $\langle n, m, p\rangle$ and has character degrees $\{d_i\}$.

Theorem (CU03)

Suppose G realizes $\langle n, m, p\rangle$ and the character degrees of G are $\{d_i\}$. Then $(nmp)^{\omega/3} \leq \sum_i d_i^{\omega}$.

Theorem (CU03)

Suppose G realizes $\langle n, m, p\rangle$ and has largest character degree d. Then $(nmp)^{\omega/3} \leq d^{\omega-2}|G|$.

Beating the sum of the cubes

Since $\omega \leq 3$,by ruling out the possibility of $\omega = 3$, Thm1.8[CU03] yields a nontrivial bound on ω if and only if $nmp > \sum_i d_i^3$.

Triple product property of Sylow subgroups

Theorem (TPP)

Suppose group G has Sylow p-subgroup P, Sylow q-subgroup Q and Sylow r-subgroup R, p, q, r are pairwisely coprime. Then G realizes $\langle |P|, |Q|, |R| \rangle$ via P, Q, R.

Corollary (DPP)

Group G has Sylow p-subgroup P and Sylow q-subgroup Q, $|P|, |Q|$ coprime. Then $P, Q \subset G$ satisfy double product property.

The simultaneous double product property

Definition (CKSU05)

We say that n pairs of subsets A_i, B_i(for $1 \leq i \leq n$) of a group H satisfy the *simultaneous double product property if*

- *for all i, the pair A_i, B_i satisfies the double product property, and*
- *for all i,j,k, $a_i(a_j')^{-1}b_j(b_k')^{-1} = 1$ implies $i = k$, where $a_i \in A_i, a_j' \in A_j, b_j \in B_j, and b_k' \in B_k$.*

The simultaneous double product property

Theorem (CKSU05)

If n pairs of subsets $A_i, B_i \subseteq H$ (with $0 \leq i \leq n-1$) satisfy the simultaneous double product property, then the following subsets S_1, S_2, S_3 of $G = (H^3)^{\Delta_n} \rtimes Sym(\Delta_n)$ satisfy the triple product property:

$S_1 = \widehat{a}\pi : \pi \in Sym(\Delta_n), \widehat{a_v} \in \widehat{A_v}$ for all v

$S_2 = \widehat{b}\pi : \pi \in Sym(\Delta_n), \widehat{b_v} \in \widehat{B_v}$ for all v

$S_3 = \widehat{c}\pi : \pi \in Sym(\Delta_n), \widehat{c_v} \in \widehat{C_v}$ for all v

An example: a nontrivial bound for ω

Example

$H = Cyc_n^k \times Cyc_n$, $A_i = \{(x,i) : x \in Cyc_n^k\}$, $B_i = \{(0,i)\}$, then for $i \in Cyc_n$, A_i, B_i satisfy the The simultaneous double product property.

Let $G = (H^3)^{\Delta_n} \rtimes Sym(\Delta_n)$

$S_1 = \{\widehat{a}\pi : \pi \in Sym(\Delta_n), \widehat{a_v} \in \widehat{A_v}\text{for all } v\}$

$S_2 = \{\widehat{b}\pi : \pi \in Sym(\Delta_n), \widehat{b_v} \in \widehat{B_v}\text{for all } v\}$

$S_3 = \{\widehat{c}\pi : \pi \in Sym(\Delta_n), \widehat{c_v} \in \widehat{C_v}\text{for all } v\}$

where $\Delta_n = \{(a,b,c) \in \mathbb{Z}^3 : a+b+c = n-1 \text{ and} a,b,c \geq 0\}$ for n pairs subsets $A_i, B_i$of H, $0 \leq i \leq n-1$, we define subset triples in H^3, $v = (v_1, v_2, v_3) \in \Delta_n$ is the index set:

$\widehat{A_v} = A_{v_1} \times \{1\} \times B_{v_3}$

$\widehat{B_v} = B_{v_1} \times A_{v_2} \times \{1\}$

$\widehat{C_v} = \{1\} \times B_{v_2} \times A_{v_3}$

An example

Example

from CKSU05 theorem 4.3(as showed above)we know that $S_1, S_2, S_3 \subset G$ satisfy the triple product property. From CKSU05 thm1.8 and cor1.9, we have $(|S_1||S_2||S_3|)^{\omega/3} \leq \sum_i d_i^{\omega}$, denote as equation (1)

$|S_1| = (|\Delta_n|!)(n^k)^{|\Delta_n|} = |S_2| = |S_3|$,

$|\Delta_n| = \binom{n+1}{2} = \frac{1}{2}n(n+1)$.

$|G| = |\Delta_n|! \cdot (n^{k+1})^{3|\Delta_n|}$, substitute into (1), $d_G \leq |\Delta_n|!$

$\Longrightarrow$

$$\omega \leq 3 + \frac{6}{k \cdot n \cdot (n+1)} - \frac{2 \cdot \log_n((\frac{n \cdot (n+1)}{2})!)}{k \cdot n(n+1)},$$

By calculation we know when $n = 4$, $k = 3$ ω has a best bound $\omega \leq 2.63682$.

Contents

Small matrix multiplication—background

The famous result $O(n^{2.81})$ is based on an algorithm that can compute the product of two 2×2 matrices with only 7 multiplications.

- Winograd: cannot produce better results with 2×2 matrices.
- Hedtke and Murthy: the group-theoretic framework is not able to produce better bounds for 3×3 and 4×4 matrices.
- Sarah Hart, Ivo Hedtke, Matthias M$\ddot{u}$ller-Hannemann and Sandeep Murthy in 2013: the group-theoretic framework is not able to produce better bounds for 5×5 matrices.

We consider the case for 6×6 matrices multiplication and to see whether this particular TPP approach can give us a better bound.

Rank

Definition (BCS1997 chap 14, def14.7)

Let k be a field and U, V, W finite dimensional k-vector space. Let $\eta : U \times V \to W$ be a k-bilinear map. For $i \in \{1, ..., r\}$ let $f_i \in U^*$, $g_i \in V^*$(dual spaces of U and V resp. over k) and $w_i \in W$ such that $\eta(u, v) = \sum_{i=1}^{r} f_i(u)g_i(v)w_i$ for all $u \in U$, $v \in V$. Then $\{f_1, g_1, w_1; ...; f_r, g_r, w_r\}$ is called a *k-bilinear algorithm of length r for η*, or simply a *bilinear algorithm* when k is fixed. The minimal length of all bilinear algorithms for η is called the *rank* $R(\eta)$ of η. Let A be a k-algebra. The *rank* $R(A)$ of A is defined as the rank of its bilinear multiplication map.

6×6 small matrix multiplication

Problem Statement: Is there a group with order less than 90 that can realize $\langle 6, 6, 6 \rangle$ TPP property and have multiplication rank less than 161[DIStable]?

Since the search space is too large, my main thinking is to reduce the search space by lots of necessary conditions.

Theorem

If G is an abelian group realizing $\langle 6, 6, 6 \rangle$, then $R(G) \geq 216$.

So we only need to consider non-abelian groups from now on.

Necessary conditions for 6×6 small matrix multiplication

For a finite group G, let $T(G)$ be the number of irreducible complex characters of G and $b(G)$ the largest degree of an irreducible character of G.

Theorem (APlowerbounds, Theorem 6)

Let G be a group.
(1)If $b(G) = 1$, then $R(G) = |G|$.
(2)If $b(G) = 2$, then $R(G) = 2|G| - T(G)$.
(3)If $b(G) \geq 3$, then $R(G) \geq 2|G| + b(G) - T(G) - 1$.

Remark

We write $\bar{R}(G) := \sum_i R(d_i)$ for the best known upper bound and $\underline{R}(G)$ for the best known upper bound(can be the theorem above sometimes) for $R(G)$.

Necessary conditions for 6×6 small matrix multiplication

Theorem (HHMM555, lemma3.3)

If G is non-abelian, then $T(G) \leq \frac{5}{8}|G|$. *Equality implies that* $|G : Z(G)| = 4$.

we have:
$R(G) \geq 2|G| - T(G) \geq (11/8)|G|$
Since we want $R(G) < 161$, then we have:
$(11/8)|G| < 161$
$|G| \leq 117$.

Necessary conditions for 6×6 small matrix multiplication

Definition ($\langle 6,6,6\rangle$C1 candidate)

If a group G realizes $\langle 6,6,6\rangle$ and has $\underline{R}[G] < 161$, we call this group a $\langle 6,6,6\rangle$ *C1 candidate*.

Proposition

If group G is a $\langle 6,6,6\rangle$ C1 candidate, then $66 \le |G| \le 117$.

Necessary conditions for 6×6 small matrix multiplication

Definition (HHMM555, definition3.4)

Let G be a group with a TPP triple (S, T, U), and suppose H is a subgroup of index 2 in G. We define $S_0 = S \cap H, T_0 = T \cap H, U_0 = U \cap H,\ S_1 = S \setminus H,\ T_1 = T \setminus H$ and $U_1 = U \setminus H$.

Theorem (generalized)

If group G realizes $\langle n, n, n\rangle$. When n is odd, if G has a subgroup H of index 2, then H realizes $\langle \frac{n+1}{2}, \frac{n+1}{2}, \frac{n+1}{2}\rangle$; When n is even, if G has a subgroup H of index 2, then H realizes$\langle \frac{n}{2}, \frac{n}{2}, \frac{n}{2}\rangle$.

Lemma

Suppose G realizes $\langle 6, 6, 6\rangle$. If G has a subgroup H of index 2, then H realizes $\langle 3, 3, 3\rangle$.

Necessary conditions for 6×6 small matrix multiplication

Lemma

If G realizes $\langle 6, 6, 6 \rangle$ and $|G| < 90$, then G has no abelian subgroups of index 2.

6×6 small matrix multiplication—result

Remark

After all these necessary conditions and GAP calculations on the bound of $R(G)$ (rule out those groups G with $R(G) \geq 161$).

Among all the groups of order less than 90, *possible C1 candidates are listed as below by their GAP ID (56 groups in total):*
(68,3),(72,3),(72,15),(72,16),(72,19),(72,20),(72,21),(72,22), (72,23),(72,24),(72,25), (72,39),(72,40),(72,41),(72,42),(72,43), (72,44),(72,45),(72,46),(72,47),(75,2),(78,1), (78,2),(80,3), (80,15),(80,18),(80,28),(80,29),(80,30),(80,31),(80,32),(80,33), (80,34), (80,39),(80,40),(80,41),(80,42),(80,49),(80,50),(81,3), (81,4),(81,6),(81,7),(81,8), (81,9),(81,10),(81,12),(81,13), (81,14),(84,1),(84,2),(84,7),(84,8),(84,9),(84,10),(84,11).

Contents

Constructing TPP triples

Definition (IHupgrade2015, TPP capacity)

Denote the *TPP capacity* of group G as $\beta(G)$,
$\beta(G) := max\{npm, \text{where} G \text{ realize } \langle n, p, m \rangle\}$.

Theorem

A_4 realizes $\langle 3, 3, 2 \rangle$, $\beta(A_4) = 18$.

TPP triples $S : \{(1), (13)(24)\}$; $T : \{(1), (243), (234)\}$;
$U : \{(1), (124), (142)\}$.

constructing TPP triples

Denote $G := C_6 \times A_4$.

Proposition

G realizes $\langle 6, 6, 3\rangle$ via S_1, T_1, U_1:
$S_1 :=$
$\{(1,1),(1,(13)(24)),(\bar{3}^{(1)},1),(\bar{3}^{(1)},(13)(24)),(\bar{3}^{(2)},1),(\bar{3}^{(2)},(13)(24))\}$;
$T_1 :=$
$\{(1,1),(1,(243)),(1,(234)),(\bar{2}^{(1)},1),(\bar{2}^{(1)},(243)),(\bar{2}^{(1)},(234))\}$;
$U_1 := \{(1,1),(1,(124)),(1,(142))\}$.

Constructing TPP triples

Denote $H := C_3 \times A_4$.

Proposition

H realizes $\langle 6, 4, 3 \rangle$ via S, T, U:
$S :=$
$\{(1,1),(1,(13)(24)),(\bar{3}^{(1)},(13)(24)),(\bar{3}^{(2)},(13)(24)),(\bar{3}^{(1)},1),(\bar{3}^{(2)},1)\}$;
$T := \{(1,1),(1,(14)(23)),(1,(143)),(1,(134))\}$;
$U := \{(1,1),(1,(123)),(1,(132))\}$.

constructing TPP triples—some principles

First explain $S_2, T_2, U_2, X, Y, Z, S, T, U, D, S_3, T_3, U_3, Q$!

Theorem

If $S_2, T_2, U_2 \subset D$ satisfy TPP and $S \cap X \neq \phi$, then $Y \cap T = \phi$ and $Z \cap U = \phi$ must hold.

Theorem (generalized)

If $S_3, T_3, U_3 \subset Q$ satisfy TPP and $S \cap X \neq \phi$, then we have $Y \cap T = \phi$ and $Z \cap U = \phi$.

Constructing TPP triples—some principles

Proposition

If $S_2, T_2, U_2 \subset D$ satisfy TPP, then the subset triples (S, Y, U), (S, Y, Z), (S, T, Z), (X, T, U), (X, T, Z), (X, Y, U), (X, Y, Z) of B all satisfy TPP.

Theorem

If $S_2, T_2, U_2 \subset D$ satisfy TPP, and $S_2|_B$ contains some repeated elements, then B realizes $\langle a, b, c \rangle$, where $a = r + 1$ (r is the number of elements that has more than one occurrence), $b = |T_2|$, $c = |U_2|$.

Constructing TPP triples—some principles

Theorem (generalized)

If $S', T', U' \subset F$ satisfy TPP and $S_i|_B$ contains some repeated elements, then B realizes $\langle a, b, c\rangle$, where $a = max\{r+1, |S_i|\}$ (r is the number of elements that has more than one occurrence), $b = \max\{|T_i|\}$, $c = \max\{|U_i|\}$.(explain S_i, T_i, U_i, division of $S'|_B$, $T'|_B$,$U'|_B$)

Contents

Main results

- An example leading to a non-trivial bound: $\omega \leq 2.63682$
- TPP and DPP property of Sylow subgroups of a given group.
- 6×6 small matrix multiplication: Reduces to 56 candidates for groups of order < 90.
- Relations between the TPP of an abstract group B and the group $C_n \times B$.

Reference

- (CKSU05) Cohn H, Kleinberg R, Szegedy B, e al. Group-theoretic algorithms for matrix multiplication[C]. Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on. IEEE, 2005: 379-388.
- (CU03) Cohn H, Umans C. A group-theoretic approach to fast matrix multiplication[C]. Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on. IEEE, 2003: 438-449.
- (HHMM555) Hart S, Hedtke I, Mller-Hannemann M, et al. A fast search algorithm for$\langle m, m, m \rangle$ Triple Product Property triples and an application for 5 5 matrix multiplication[J]. Groups Complexity Cryptology, 2015, 7(1): 31-46.

Reference

- (APlowerbounds) Pospelov A. Group-theoretic lower bounds for the complexity of matrix multiplication[C]. International Conference on Theory and Applications of Models of Computation. Springer Berlin Heidelberg, 2011: 2-13.
- (strassen1969) Strassen V. Gaussian elimination is not optimal[J]. Numerische mathematik, 1969, 13(4): 354-356.
- (CW90) Coppersmith D, Winograd S. Matrix multiplication via arithmetic progressions[J]. Journal of symbolic computation, 1990, 9(3): 251-280.

Reference

- (AS2010) Davie A M, Stothers A J. Improved bound for complexity of matrix multiplication[J]. Proceedings of the Royal Society of Edinburgh: Section A Mathematics, 2013, 143(02): 351-369.
- (VW2012) Williams V V. Multiplying matrices faster than Coppersmith-Winograd[C]. Proceedings of the forty-fourth annual ACM symposium on Theory of computing. ACM, 2012: 887-898.
- (LeGall2014) Le Gall F. Powers of tensors and fast matrix multiplication[C]. Proceedings of the 39th international symposium on symbolic and algebraic computation. ACM, 2014: 296-303.

Reference

- (AFL2015) Ambainis A, Filmus Y, Le Gall F. Fast matrix multiplication: limitations of the coppersmith-winograd method[C]. Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing. ACM, 2015: 585-593.
- (DIStable) DrevetC, Islam M N, Schost. Optimization techniques for small matrix multiplication[J]. Theoretical Computer Science, 2011, 412(22): 2219-2236.
- (IHupgrade2015) Hedtke I. Upgrading Subgroup Triple-Product-Property Triples[J]. Journal of Experimental Algorithmics (JEA), 2015, 20: 1.1.

Reference

- (BCS1997) Brgisser P, Clausen M, Shokrollahi A. Algebraic Complexity Theory[M]. Springer Science&Business Media, 1996.

Thank You