on Bounds for Matrix Multiplication Complexity

Kimia Kazemian, Guanyu Li

December 2020

1 Introduction

Matrix multiplication is an important and fundamental topic in algorithm analysis. Strassen [S] made the startling discovery that one can multiply two $n \times n$ matrices in only $O(n^{2.81})$ field operations, compared with $2n^3$ for the standard algorithm. This immediately raises the question of the exponent of matrix multiplication, denoted by ω. It is conjectured that $\omega = 2$.

In the following, for a group G, matrix multiplication can be reduced to multiplication of elements of the group algebra of G using the structure theorem for $\mathbb{C}[G]$. This latter multiplication is performed via a Fourier transform, which reduces it to several smaller matrix multiplications, whose sizes are the character degrees of G. To elaborate, we will show that if it has 3 subsets of sizes m, n, p that satisfy a certain property (the triple product property aka TPP), then we can derive a bound for ω. We will show that such bound is non-trivial under certain conditions.

Our next step naturally is to construct such groups, and inspect the corresponding bounds for ω. We will construct a trivial example for illustration, and then we go on to construct those groups via uniquely solvable puzzles (USP). As part of it we will state the nice result that a set U is a USP if and only if certain subsets of U satisfy the triple product property within $\mathrm{Sym}(U)$, and therefore conclude with showing that there exists strong USP of a certain size. Most of our understandings happen here in Section 2 and Section 3, which includes a certain concrete result for $\omega < 2.48$.

After, we touch on the elegant wreath product construction, which is used to prove the corresponding result on the simultaneous triple product property (STPP). we conclude by showing that the STTP implies the existence of TTP in a lager set, hence TTP gives us as strong results as STPP. Also given two subsets satisfying simultaneous double product property we construct a triple of sets that satisfy the TPP, for that we need to show the existence of triangle-free sets, this will use a nice result in additive combinatorics, involving finding large subsets of $[1, N]$ that are 3-term arithmetic progression free.

2 Main Estimates

Definition. The asymptotic exponent of matrix multiplication is the smallest ω such that one may multiply two $n \times n$ matrices in $O(n^{\omega+\epsilon})$ complexity for $\epsilon > 0$ arbitrarily small.

It remains a mystery what ω is. Our notes focus on a method finding fast matrices multiplication, giving a partial answer to it. We start from the following definition, where later the sizes of matrices are captured by the group given.

Definition. A group G realizes $\langle n_1, n_2, n_3 \rangle$ if there are subsets $S_1, S_2, S_3 \subseteq G$ where $|S_i| = n_i$, 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.

Note: double product property is the special case of above for two subsets. Observe the following lemma tells us that the definition is actually well-defined.

Lemma. *If G realizes $\langle n_1, n_2, n_3 \rangle$, then it does so for every permutation of n_1, n_2, n_3.*

Definition. The group algebra $C[G]$ is the set of formal sums $\Sigma_{g \in G} g.a_g$ where $a_g \in \mathbb{C}$.

Theorem 2.1 ([CU]). *Let F be any field. If G realizes $\langle n, m, p \rangle$, then the number of field operations required to multiply $n \times m$ with $m \times p$ matrices over F is at most the number of operations required to multiply two elements of $F[G]$.*

We omit the proofs till the appendix. This is where the representation theory is involved. But one should notice that the number of operations are actually bounded by properties of the given group G, because the group algebra $F[G]$ is entirely determined by the group G. In particular, the following theorems assert that it suffices to know all the character degrees of G to bound ω.

Theorem 2.2. *For every group G, $\mathbb{C}[G] \simeq \mathbb{C}^{d_1^2} \times ... \times \mathbb{C}^{d_k^2}$, where The integers d_i are the character degrees of G, i.e. the dimensions of the irreducible representations of G, and the operations in the resulting algebra are component-wise matrix addition and multiplication. As a result, counting dimensions we see that $\Sigma_{1 \le i \le k} {d_i}^2 = |G|$.*

Theorem 2.3 ([CU]). *Suppose G realizes $\langle n, m, p \rangle$ and the character degrees of G are $\{d_i\}$. Then*

$$(nmp)^{\omega/3} \le \sum_i d_i^{\omega}.$$

Corollary 2.3.1. *Suppose G realizes $\langle n, m, p \rangle$ and has largest character degree d. Then*

$$(nmp)^{\omega/3} \le d^{\omega-2}|G|.$$

The above theorem yields a nontrivial bound on ω (by ruling out the possibility of $\omega = 3$) if and only if

$$nmp > \sum_i (d_i)^3$$

The natural question is whether such a group exists, we will construct one as an example in the next section, which will prove a nontrivial bound on ω. But actually we conjecture that $\omega = 2$, hence it seems like that one should construct good G realizing specific m, n, p, yielding good bound of ω.

However, in the next section, we would impose some other conjecture, with an approximation of existence of certain construction, which leads to that $\omega = 2$.

3 Key Notions and Key Results

3.1 Trivial Case [CKSU]

Let $H = (\mathbb{Z}/n\mathbb{Z})^3$, and let $G = H^2 \rtimes \mathbb{Z}/2\mathbb{Z}$, where $\mathbb{Z}/2\mathbb{Z}$ acts on H^2 by switching the two factors of H. Let z denote the generator of $\mathbb{Z}/2\mathbb{Z}$. We write elements of G in the form $(a, b)z_i$, with $a, b \in H$ and $i \in \{0, 1\}$. Note that $z(a, b)z = (b, a)$.

Let H_1, H_2, H_3 be the three factors of $\mathbb{Z}/n\mathbb{Z}$ in the product $H = (\mathbb{Z}/n\mathbb{Z})^3$, viewed as subgroups of H. For notational convenience, let $H_4 = H_1$. Define subsets $S_1, S_2, S_3 \le G$ by

$$S_i = \{(a, b)z_j \mid a \in H_i - \{0\}, b \in H_{i+1}, j \in \{0, 1\}\}.$$

A slight technical but not so important fact is

Lemma. *S_1, S_2, and S_3 satisfy the triple product property.*

With this, we can start our analysis. It is easy to see that the character degrees of G are all at most 2, because H^2 is an abelian subgroup of index 2. Then since the sum of the squares of the character degrees is $|G|$, the sum of their cubes is at most $2|G|$, which equals $4n^6$.

On the other hand, $|S_i| = 2n(n-1)$, so $|S_1||S_2||S_3| = 8n^3(n-1)^3$. For $n \ge 5$, this product is larger than $4n^6$. By Corollary 2.3.1, $(2n(n-1))^{\omega} \le 2^{\omega-2}2n^6$. The best bound on ω is achieved by setting $n = 17$, in which case we obtain $\omega < 2.9088$.

3.2 Improved case

However, the result from last section wasn't good enough. The reason is that the group we constructed could be better. In order to construct the group fitting in, we need to "solve" some combinatorial question, which is

Definition. A uniquely solvable puzzle (USP) of width k is a subset U of $\{1,2,3\}^k$ satisfying the following property. Let $\mathrm{Sym}(U)$ denote the ways of permuting the elements of U, so that it is isomorphic to $S_{|U|}$. The condition is that for all permutation $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}(U)$, either we have $\pi_1 = \pi_2 = \pi_3$ or there exists a $u \in U$ and $i \in \{1, 2, \cdots, k\}$ such that for at least two $j \in \{1,2,3\}$, $(\pi_j(u))_i = j$ is true.

It's not an easy thing to understand the definition above. One the one hand, we have two indices to keep track of. On the other hand, to distinguish different elements in U with the same symbol is also tedious. Therefore we tend to use an alternative way of defining it.

Definition. [M] A USP candidate of width k is a subset U of $\{1,2,3\}^k$, which we imagine as a grid with k columns and $|U|$ un-ordered rows each of whose entries is from the set $\{1,2,3\}$. For $u \in U, i \in \{1, 2, \cdots, k\}$, we refer to the element in the i-th column of the row u by writing u_i. Let U be a USP candidate and let (π_1, π_2, π_3) be three permutations of the rows of U. Let u_i be an element of the grid U. We say (π_1, π_2, π_3) satisfy the USP rule for this position if for at most one $j \in \{1,2,3\}$ we have $(\pi_j(u))_i = j$ is true.

Example 1.

$$\begin{matrix} 3 & 3 & 3 & 3 & 3 & 3 \\ 1 & 3 & 3 & 2 & 3 & 3 \\ 3 & 1 & 3 & 3 & 2 & 3 \\ 1 & 1 & 3 & 2 & 2 & 3 \\ 3 & 3 & 1 & 3 & 3 & 2 \\ 1 & 3 & 1 & 2 & 3 & 2 \\ 3 & 1 & 1 & 3 & 2 & 2 \\ 1 & 1 & 1 & 2 & 2 & 2 \end{matrix}$$

Above is an example of a USP candidate. The negation of the USP rule, that at least two of $(\pi_j(u))_i = j$ are true, which is the same language in the original definition, refers to the case where two of the permuted rows do overlap. So we can imagine each row of each grid as a jigsaw piece which must be fit with the others in a non-overlapping fashion. We refer to these overlaps as *collisions*. Therefore the definition for a USP states that if the USP rule is satisfied for all positions, we must have $\pi_1 = \pi_2 = \pi_3$. That is, all of the rows are permuted in the same way.

$$\begin{bmatrix} 3 & 3 & 3 & 3 & 3 & 3 \\ & 3 & 3 & & 3 & 3 \\ 3 & & 3 & 3 & & 3 \\ & & 3 & & & 3 \\ 3 & 3 & & 3 & 3 & \\ & 3 & & & 3 & \\ 3 & & & 3 & & \\ & & & & & \end{bmatrix} \quad \begin{bmatrix} & & & & & \\ 1 & & & & & \\ & 1 & & & & \\ 1 & 1 & & & & \\ & & 1 & & & \\ 1 & & 1 & & & \\ & 1 & 1 & & & \\ 1 & 1 & 1 & & & \end{bmatrix} \quad \begin{bmatrix} & & & & & \\ & & & 2 & & \\ & & & & 2 & \\ & & & 2 & 2 & \\ & & & & & 2 \\ & & & 2 & & 2 \\ & & & & 2 & 2 \\ & & & 2 & 2 & 2 \end{bmatrix}$$

Next we want to have an easy way of interpreting USP rule. The first thing to note is that the constraints it imposes on π_j for $j \in \{1,2,3\}$ mean that for π_j , only the j's in the grid are important. Imagine breaking the USP into three separate grids with all occurrences of j in each one, just as shown above. Each π_j moves around the rows of one of the grids independently of the others, and then the three grids are superimposed. The selection of u, i is the same as selecting a row and column respectively of the resulting table. Satisfying $(\pi_j(u))_i = j$ is equivalent to the grid having a j in the selected position. The USP rule's constraint that at most one of $(\pi_j(u))_i = j$ is true means there is no overlap between the three permuted rows.

So we fix a column i, and a row u, then $\pi_j(u)$ is the row after the permutation. Because we have the discussion above, it is OK to say this is a USP by testing all possible i's and choices of π_j's.

Example 2. Here below is a non-example:

$$\begin{matrix} 3 & 2 & 3 & 2 \\ 1 & 1 & 3 & 2 \\ 1 & 2 & 1 & 3 \\ 3 & 1 & 1 & 3 \\ 1 & 3 & 2 & 1 \end{matrix}$$

If we split it up into three pieces as before

$$\begin{bmatrix} 3 & & 3 & \\ & & 3 & \\ & & & 3 \\ 3 & & & 3 \\ & 3 & & \end{bmatrix} \qquad \begin{bmatrix} & 2 & & 2 \\ & & & 2 \\ & 2 & & \\ & & & \\ & & 2 & \end{bmatrix} \qquad \begin{bmatrix} & & & \\ 1 & 1 & & \\ 1 & & 1 & \\ & 1 & 1 & \\ 1 & & & 1 \end{bmatrix}$$

and notice that for $\pi_1 = \mathrm{id}, \pi_2 = (2,3,5), \pi_3 = (2,5,3)$ we get

$$\begin{matrix} 3 & 2 & 3 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 3 & 1 & 2 \\ 3 & 1 & 1 & 3 \\ 1 & 2 & 3 & 1 \end{matrix}$$

which is different from our first puzzle, where (a, b, c) means the permutation of rows $a \mapsto b, b \mapsto c, c \mapsto a$.

We also need this similar definition:

Definition. A strong USP is a USP in which for all permutations $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}(U)$, either $\pi_1 = \pi_2 = \pi_3$ or else there exist $u \in U$ and $i \in [k]$ such that exactly two of $(\pi_1(u))_i = 1$, $(\pi_2(u))_i = 2$, and $(\pi_3(u))_i = 3$ hold.

Remark. The authors have some thoughts on the definition of USPs. It is true that the conditions $(\pi_j(u))_i = j$ are linear (hence algebraic) equations. Also, the above discussions mean that the ambient space is the affine space

$$\mathbb{A}^{|U|\times k}_{\mathbb{Z}/3\mathbb{Z}}.$$

(One should view this as a scheme in the sense of algebraic geometry to use results from number theory). This means USP's form a complement of sub-varieties cut out by linear equations. This might give us a way of counting points. Then the existence is an immediate implication of counting.

Unfortunately, Bobby said no one has considered this. Also, The equations are too simple, hence we also don't know if number theory would help.

After a long discussion of USP / strong USP, we are able to bring the contribution to the estimate of them. This is another notion capturing the behavior of USP / strong USP in the large number cases.

Definition. The strong USP capacity is defined to be the largest constant C such that there exist strong USPs of size $(C - o(1))^k$ and width k for infinitely many values of k. The USP capacity is defined analogously.

It is sort of easy to prove that

Lemma. *The USP capacity is at most* $3/2^{2/3}$.

If one goes to deeper construction, it is also possible that one can show the following theorem using arithmetic progression.

Theorem 3.1. *The USP capacity equals* $3/2^{2/3}$. [1]

However, it is more important to know the strong USP capacity. The following sub-section will give an example. One conjectures that the strong USP capacity is also $3/2^{2/3}$, which implies $\omega = 2$!

3.3 The Realization of A Strong USP Capacity [CKSU]

Given a strong USP U of width k, let H be the abelian group of all functions from $U \times [k]$ to the cyclic group $\mathbb{Z}/m\mathbb{Z}$ (H is a group under point-wise addition). The symmetric group $\mathrm{Sym}(U)$ acts on H via

$$\pi(h)(u, i) = h(\pi^{-1}(u), i)$$

for $\pi \in \mathrm{Sym}(U)$, $h \in H$, $u \in U$, and $i \in [k]$.

[1] BIG THANKS TO BOBBY FOR HIS EXPLANATIONS IN THE EXTRA OFFICE HOURS!

Let G be the semidirect product $H \rtimes \mathrm{Sym}(U)$, and define subsets S_1, S_2, and S_3 of G by letting S_i consist of all products $h\pi$ with $\pi \in \mathrm{Sym}(U)$ and $h \in H$ satisfying

$$h(u,j) \neq 0 \ \text{ iff } \ u_j = i$$

for all $u \in U$ and $j \in [k]$.

We bring up a fact here:

Proposition 3.2. *Provided U is a strong USP, then S_1, S_2, and S_3 satisfy the triple product property.*

Immediately we have

Corollary 3.2.1. *If U is a strong USP of width k with $m \geq 3$ is an integer, then*

$$\omega \leq \frac{3\log m}{\log(m-1)} - \frac{3\log|U|!}{|U|k\log(m-1)}.$$

In particular, if C is the strong USP capacity, then

$$\omega \leq \frac{3\log m - \log C}{\log(m-1)}.$$

This is how strong USP capacity can be involved! Then the larger we can find such a C, the smaller ω we can get, which gives an explanation of the comments from last sub-section (where we take $m = 3$). Therefore, we bring up an actual construction of strong USP here to realize a result about ω.

Suppose $U \subseteq \{1,2,3\}^k$ is a subset with only two symbols occurring in each coordinate. Let H_1 be the subgroup of $\mathrm{Sym}(U)$ that preserves the coordinates in which only 1 and 2 occur, H_2 the subgroup preserving the coordinates in which only 2 and 3 occur, and H_3 the subgroup preserving the coordinates in which only 1 and 3 occur.

Lemma. *The set U is a USP if and only if H_1, H_2, H_3 satisfy the triple product property within* $\mathrm{Sym}(U)$

So one can show that

Theorem 3.3. *For each $k \geq 1$, there exists a strong USP of size $2^{k-1}(2^k+1)$ and width $3k$.*

Corollary 3.3.1. *If U is a USP of width k such that only two symbols occur in each coordinate, then $|U| \leq (2^{2/3}+o(1))^k$.*

So in our case, $C = 2^{2/3}$ and we get $\omega < 2.48$.

4 Further Discussion

From last section, we saw that the construction of the group G gives us the estimate that we want. However, morally the worse (more non-abelian) the group is, the smaller ω is. Hence we can generalize the notions mentioned above to loose the techniques we have to use in the constructions. Basically it is taking multiple of three sets / USP's and make them simultaneous.

4.1 The Simultaneous Double Product Property [CKSU]

The first generalization is as follows:

Definition. We say that n pairs of subsets A_i, B_i, C_i for $1 \leq i \leq n$ of a group H satisfy the simultaneous triple product property if

1. for all i the three subsets A_i, B_i satisfies the triple product property and
2. for all i, j, k, $a_i(a'_j)b_j(b'_k)^{-1}c_k(c'_i)^{-1} = 1$ implies $i = j = k$ where $a_i \in A_i, a'_j \in A_j, b_j \in B_j, b'_k \in B_k, c_k \in C_k$ and $c'_i \in C_i$.

note: the simultaneous double product property is a special case of STPP restricted to two subsets. A convenient reformulation is that if one looks at the sets $A_i^{-1}B_j = a^{-1}b : a \in A_i, b \in B_j$ those with $i = j$ are disjoint from those with $i \neq j$.

Define $\Delta_n = \{(a, b, c) \in \mathbb{Z} : a + b + c = n - 1, a, b, c \geq 0$, we call a subset S of Δ_n *triangle free* if for all $u, v, w \in S$ satisfying $u_1 = w_1, v_2 = u_2, w_3 = v_3$ it follows that $u = v = w$.

Given n pairs of subsets A_i, B_i in H for $i \in \{0, ..., n-1\}$ we define triples of subsets in H^3 indexed by $v = (v_1, v_2, v_3) \in \Delta_n$ as follows:

$$\hat{A}_v = A_{v_1} \times \{1\} \times B_{v_3} \qquad \hat{B}_v = B_{v_1} \times A_{v_2} \times \{1\} \qquad \hat{C}_v = \{1\} \times B_{v_2} \times A_{v_3}$$

these triples satisfy the following property, if $a_u(a'_v)^{-1}b_v(b'_w)^{-1}c_w(c'_u)^{-1} = 1$ then it follows from the simultaneous double product property that $u_1 = w_1, v_2 = u_2, w_3 = v_3$. Thus the triples $\hat{A}_v, \hat{B}_v, \hat{C}_v$ with v in a triangle-free subset of Δ_n satisfy the triple product property. So now the question is whether there exists a triangle-free subset of Δ_n of size $|\Delta_n|^{1-o(1)}$. Let T be a subset of $[|n/2|]$ of size $n^{1-o(1)}$ that contains no three-term arithmetic progression. by Theorem A.1 in appendix, such set exists. we can prove that the subset $(a, b, c) \in \Delta_n : b - a \in T$ is triangle-free and has size $|\Delta_n|^{1-o(1)}$. With this, we have

Theorem 4.1. *If H is a finite group with character degrees $\{d_k\}$, and n pairs of subsets $A_i, B_i \subseteq H$ satisfy the simultaneous double product property, then*

$$\sum_{i=1}^{n} (|A_i||B_i|)^{\omega/2} \leq \left(\sum_k d_k^\omega \right)^{3/2}.$$

Using this theorem, the construction above recovers the trivial bound $\omega \leq 3$ as $k \to \infty$. However we also have

Proposition 4.2. *For each $m \geq 2$, there is a construction in $\mathbb{Z}/n\mathbb{Z}^{2l}$ satisfying the simultaneous double product property with $\alpha = \log_2(m-1) + o(1)$ and $\beta = \log_2(m) + o(1)$ as $l \to \infty$.*

By taking $m = 3$, we get the same as Corollary 3.3.1

4.2 The wreath product construction [CKSU]

Here we mention that if any n triple of subsets satisfy STTP, we can construct subsets of the wreath product which satisfy the TTP, this will show that any bound that can be derived from Theorem 4.3, can be derived using Theorem 2.1.

Theorem 4.3. *If a group H simultaneously realizes $\langle a_1, b_1, c_1 \rangle, ..., \langle a_n, b_n, c_n \rangle$ and has character degree d_k then*

$$\sum_{i=1}^{n} (a_i b_i c_i)^{\omega/3} \leq \sum_{i=1}^{n} d_k^\omega$$

Theorem 4.4. *If n triple of subsets $A_i, B_i, C_i \subset H$ satisfy STTP the the following subsets H_1, H_2, H_3 of $G = \mathrm{Sym_n} \rtimes H^n$ satisfy the TPP:*

$$H_1 = \{h\pi \in \mathrm{Sym}_n, h_i \in A_i \textit{ for each } i\} \quad H_2 = \{h\pi \in \mathrm{Sym}_n, h_i \in B_i \textit{ for each } i\} \quad H_3 = \{h\pi \in \mathrm{Sym}_n, h_i \in C_i \textit{ for each } i\}$$

A Appendix

Proof of Theorem 2.1. Let G realize $\langle n, m, p \rangle$ through subsets S, T, U. Suppose A is an nm matrix, and B is an mp matrix. We will index the rows and columns of A with the sets S and T , respectively, those of B with T and U , and those of AB with S and U. Consider the product

$$\Big(\sum_{s \in S, t \in T} A_{st} s^{-1} t\Big)\Big(\sum_{t' \in T, u \in U} B_{t'u} t'^{-1} u\Big)$$

in the group algebra. We have

$$(s^{-1}t)(t'^{-1}u) = s'^{-1}u'$$

iff $s = s', t = t', u = u'$, so the coefficient of $s^{-1}u$ in the product is

$$\sum_{t \in T} A_{st} B_{tu} = (AB)_{su}$$

so the matrix product is determined from the group algebra product by looking at the coefficients of $s^{-1}u$ with $s \in S, u \in U$, and the assertions in the theorem statement follow. □

Theorem A.1. *Behrend's Theorem. There exists a set $A \in [1, N]$ with $|A| \gg Nexp(-c\sqrt{logN})$ containing no non-trivial three term arithmetic progressions, where c is an absolute positive constant.*

References

[CKSU] H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans, *Group-theoretic algorithms for matrix multiplication*, 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05).

[CU] H. Cohn and C. Umans, *A group-theoretic approach to fast matrix multiplication*, 44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.

[M] Palmer Mebane, *Uniquely solvable puzzles and fast matrix multiplication*, HMC Senior Theses **37** (2012).

[S] V. Strassen, *Relative bilinear complexity and matrix multiplication*, J. Reine Angew. Math. **375/376** (1987), 406–443. MR882307