In this lecture, we will show Coppersmith-Winograd's algorithm for matrix multiplication that uses the "small" CW tensor. We will achieve $\omega < 2.404$.

Recall for any integer $q \geq 2$, the "small" CW tensor is defined as

$$cw_q = \left(\sum_{i=1}^{q} x_0 y_i z_i\right) + \left(\sum_{i=1}^{q} x_i y_0 z_i\right) + \left(\sum_{i=1}^{q} x_i y_i z_0\right),$$

which lies in $\mathcal{K}^{(q+1)\times(q+1)\times(q+1)}$. The volume of a matrix multiplication tensor $\langle K, M, N\rangle$ is defined as KMN. As we can see, cw_q is a sum of three matrix multiplication tensors of volume q. Coppersmith and Winograd showed that $\underline{R}(cw_q) \leq q + 2$. It's easy to see that $\underline{R}(cw_q) \geq q + 1$

We will use the following special case of Schönhage's τ-theorem.

Theorem 1. *If $\underline{R}(\bigoplus_{i=1}^{p}\langle k_i, m_i, n_i\rangle) \leq r$ for some $r > p$, and $k_i m_i n_i = v$ for every i, then $\omega \leq \frac{3\log(r/p)}{\log v}$.*

The strategy for the Coppersmith-Winograd algorithm is to first take a big tensor power of cw_q, then attempt to apply Schönhage's τ-theorem. However, since cw_q consists of sum of some tensors instead of direct sum, we have to carefully *zero out* some variables to transform the tensor power of cw_q to a direct sum of some matrix multiplication tensors.

The following is the main theorem we will prove in this lecture.

Theorem 2 (Coppersmith-Winograd)**.** *If $R(cw_q^{\otimes n}) \leq c^{n+o(n)}$ for some constant c and some integer q, then $\omega \leq \frac{\log\left(\frac{4c^3}{27}\right)}{\log q}$.*

We get the following two corollaries from the theorem.

Corollary 1. *If $\underline{R}(cw_q) = q + 1$, then $\omega = 2$.*

Proof. If $\underline{R}(cw_q) = q + 1$, then by definition of border rank, there exists h such that $R_h(cw_q) = q + 1$. Therefore, for all $n \geq 1$, $R_{nh}(cw_q^{\otimes n}) \leq (q + 1)^n$, which further implies $R(cw_q^{\otimes n}) \leq (q + 1)^n \text{poly}(n) = (q + 1)^{n+o(n)}$. Thus, we can apply Theorem 2 by using $q = 2$ and $c = 3$, which gives that

$$\omega \leq \frac{\log\left(\frac{4\cdot 3^3}{27}\right)}{\log 2} = 2.$$

□

Unfortunately, it is now known that $\underline{R}(cw_q) = q + 2$, so the above corollary is not useful. However, an easy extension of the corollary is: if for some k, $\underline{R}(cw_q^{\otimes k}) = (q + 1)^k$, then $\omega = 2$. Recently it was proven that $\underline{R}(cw_q^{\otimes 2}) = (q + 2)^2$, and if $q > 4$ then $\underline{R}(cw_q^{\otimes 3}) = (q + 2)^3$. However, it is not ruled out that for some larger power k the border rank goes below $(q + 2)^k$, or that maybe in the limit it goes to $(q + 1)^k$, which would also show that $\omega = 2$.

Our next corollary of Theorem 2 is:

Corollary 2. $\omega \leq 2.404$.

Proof. We know that $\underline{R}(cw_q) \leq q + 2$. Similar to the previous corollary, we can show that $R(cw_q^{\otimes n}) \leq (q + 2)^{n+o(n)}$. We can apply Theorem 2 by using $q = 8$ and $c = 10$, which gives that

$$\omega \leq \frac{\log\left(\frac{4\cdot 10^3}{27}\right)}{\log 8} = 2.404.$$

□

1 Preliminaries

In this section, we define some notations and introduce some theorems necessary for the proof of the main theorem.

1.1 Zeroing out, a special type of restriction

Let $t \in \mathcal{K}^{K\times M\times N}$ and $t' \in \mathcal{K}^{K'\times M'\times N'}$. Recall that $t \le t'$ (t is a restriction of t') if there exist homomorphisms $A : \mathcal{K}^K \to \mathcal{K}^{K'}, B : \mathcal{K}^M \to \mathcal{K}^{M'}$ and $C : \mathcal{K}^N \to \mathcal{K}^{N'}$ so that $t = (A \otimes B \otimes C)t'$. We showed that if $t \le t'$, then $R(t) \le R(t')$.

Zeroing out is a special type of restriction. Suppose t is a tensor in the form $\sum_{i,j,k} t_{i,j,k} x_i y_j z_k$. We pick a set of variables $x_{i_1}, \ldots, x_{i_p}, y_{j_1}, \ldots, y_{j_q}, z_{k_1} \ldots, z_{k_r}$, and set them to 0. We can define homomorphism $A : \mathcal{K}^K \to \mathcal{K}^K$ such that $A(x_1, \ldots, x_k) = (x_1, \ldots, x_{i_1-1}, 0, x_{i_1+1}, \ldots, x_{i_2-1}, 0, x_{i_2+1}, \ldots)$, which just set $x_{i_1}, \ldots, x_{i_p}$ to zero. We can similarly define homomorphisms B and C. Therefore, the zeroing out of t equals $(A \otimes B \otimes C)t$, so it is a restriction of t. Thus, if t' is a zeroing out of t, then $t' \le t$ and thus $R(t') \le R(t)$.

Note that zeroing out is limited, in the sense that it might not be possible to get all subsets of the terms of a tensor via zeroing out. For instance, let $t = x_0y_1z_1 + x_1y_0z_0 + x_0y_0z_0$. We cannot get the tensor $x_0y_1z_1 + x_1y_0z_0$ by zeroing out some variables, because in order to eliminate the $x_0y_0z_0$ term, we must set one of x_0, y_0, z_0 to zero. However, no matter which one we set to zero, another term becomes zero as well.

1.2 Variable Blocks

The tensor cw_q uses variables $x_0, \ldots, x_q, y_0, \ldots, y_q, z_0, \ldots, z_q$. We split these variables into groups, two "blocks" per variable set X, Y or Z.

$$\begin{aligned} X_0 &= \{x_0\}, \quad X_1 = \{x_1, \ldots, x_q\}, \\ Y_0 &= \{y_0\}, \quad Y_1 = \{y_1, \ldots, y_q\}, \\ Z_0 &= \{z_0\}, \quad Z_1 = \{z_1, \ldots, z_q\}. \end{aligned}$$

Using this block notation, let us now define for every choice $a, b, c \in \{0, 1\}$, the subtensor of $t = cw_q$ using only variables in X_a, Y_b, Z_c:

$$T_{a,b,c} := \sum_{x_i \in X_a, y_j \in Y_b, z_k \in Z_k} t_{ijk} x_i y_j z_k.$$

Then

$$cw_q = T_{011} + T_{101} + T_{110},$$

and $T_{011} = \sum_{i=1}^q x_0 y_i z_i$, $T_{101} = \sum_{i=1}^q x_i y_0 z_i$, $T_{110} = \sum_{i=1}^q x_i y_i z_0$. Notice that each of the subtensors $T_{011}, T_{101}, T_{110}$ are matrix products of volume q. $T_{0,1,1}$ is $\langle 1, 1, q\rangle$, $T_{1,0,1}$ is $\langle q, 1, 1\rangle$, and $T_{1,1,0}$ is $\langle 1, q, 1\rangle$.

As a simple observation, we see that for each $T_{i,j,k}$ is in cw_q iff $i + j + k = 2$.

In the proof of the main theorem, we will use tensor powers of cw_q. For instance,

$$cw_q^{\otimes 2} = (T_{011} + T_{101} + T_{110})^{\otimes 2} = \sum_{a,a',b,b',c,c' \in \{0,1\}} T_{a,b,c} \otimes T_{a',b',c'}.$$

Since each of $T_{011}, T_{101}, T_{110}$ is a matrix product of volume q, each of $T_{a,b,c} \otimes T_{a',b',c'}$ is a matrix product of volume q^2.

Take the term $T_{0,1,1} \otimes T_{1,0,1} \equiv \langle 1, 1, q\rangle \otimes \langle q, 1, 1\rangle$ as an example, we can write it as $\sum_{i=1}^q \sum_{j=1}^q x_{0,j} y_{i,0} z_{i,j}$, and it is a matrix multiplication tensor $\langle q, 1, q\rangle$. We will denote $T_{0,1,1} \otimes T_{1,0,1}$ by $T_{01,10,11}$.

More generally, for $I \in \{0, 1\}^k$, we can use X_I to denote the set of x variables $x_{p_1,\ldots,p_k}$ of $cw_q^{\otimes k}$ such that if $I_i = 0$ then $p_i = 0$, and if $I_i = 1$ then $p_i \in \{1, \ldots, q\}$. We can similarly define Y_J and Z_K. Then, for for $I, J, K \in \{0, 1\}^k$, we can define $T_{I,J,K}$ the subtensor of $cw_q^{\otimes k}$ whose variables lie in X_I, Y_J, Z_K.

It is not hard to see that

$$cw_q^{\otimes k} = \sum_{\substack{I,J,K \in \{0,1\}^k \\ I_i + J_i + K_i = 2 \ \forall i}} T_{I,J,K}.$$

1.3 Avoiding Arithmetic Progression

We will use the following theorem (by Salem, Spencer and also Behrend) that gives a large subset of $[M]$ that does not contain any nontrivial arithmetic progressions.

Theorem 3 ([2, 3]). *For any $\epsilon > 0$, there exists m_0 such that for all integer $M > m_0$, there exists $S \subseteq [M]$, $|S| \geq M^{1-\epsilon}$, such that if $a, b, c \in S$ and $a + c \equiv 2b \pmod M$, then $a = b = c$.*

(The real theorem doesn't have $\pmod M$, but has the equality be just over the integers, but it's easy to see that the modular version follows from the integer version.)

2 Proof of the Main Theorem

Let $\epsilon > 0$ be any constant. Let N be a big enough integer that depends on ϵ. Consider the $(3N)$-th tensor power of cw_q,

$$cw_q^{\otimes 3N} = \sum_{\substack{I,J,K \in \{0,1\}^{3N} \\ I_i + J_i + K_i = 2 \ \forall i}} T_{I,J,K}.$$

We aim to zero out some variables so the remaining tensor is a direct sum of some matrix multiplication tensors (and they will be of the same volume). We will always zero out all variables in the same block at once, so when we say we set a block to 0, we set all variables variables in the block to 0. The number of triples in the sum of $cw_q^{\otimes 3N}$ is 3^{3N}, since for each i, there are 3 choices for X_i, Y_i, Z_i. Also, note that each $T_{I,J,K}$ is a matrix multiplication tensor with volume q^{3N}, since each of the three terms in cw_q is a matrix multiplication tensor with volume q.

Let us call two triples $(I, J, K), (I', J', K')$ with $I, J, K, I', J', K' \in \{0,1\}^{3N}$ *independent* if $I \neq I', J \neq J', K \neq K'$. A direct sum of $T_{I,J,K}$ coming from $cw_q^{\otimes 3N}$ comes of a choice of triples (I, J, K) that are all pairwise independent.

We would like to show that there is a way to zero out variable blocks so that we obtain a large number of independent triples.

We first zero out the blocks X_I, Y_J, Z_K such that I, J or K does not have exactly N 0's and $2N$ 1's. We will not lose much here at all, but it will be easier to analyze the tensor. The remaining tensor becomes

$$t = \sum_{\substack{I,J,K \in \{0,1\}^{3N} \\ I_i + J_i + K_i = 2 \ \forall i \\ I,J,K \text{ have } N \text{ 0's}}} T_{I,J,K}.$$

Since t is a zeroing out of $cw_q^{\otimes 3N}$, we have that $R(t) \leq R(cw_q^{\otimes 3N}) \leq c^{3N+o(N)}$.

Now, let us define a **combinatorial problem** that we would like to solve. Let $L = \{(I, J, K) \mid I, J, K \in \{0,1\}^{3N}, I_i + J_i + K_i = 2 \ \forall i \in [3N], I, J, K \text{ have } N \text{ zeros}\}$. Consider all choices of $S^X, S^Y, S^Z \subseteq \{0,1\}^{3N}$ and for every such choice define $L(S^X, S^Y, S^Z) := L \cap \{(I, J, K) \mid I \in S^X, J \in S^Y, K \in S^Z\}$. What is the largest size of $L(S^X, S^Y, S^Z)$ such that the triples in it are independent?

Claim 1. *If the largest number of independent triples of L in the solution of the combinatorial problem above is $\geq \phi^{N-o(N)}$ for some ϕ, then*

$$\omega \leq \frac{\log \frac{(q+2)^3}{\phi}}{\log(q)}.$$

Proof. From the assumption of the claim we get that using just zeroing outs we can get from cw_q^{3N} down to a direct sum of at least $p := \phi^{N-o(N)}$ matrix products of volume $v := q^{3N}$, and the border rank is at most $r := (q+2)^{3N}$. From Schönhage we get: $\omega \leq \frac{3\log(r/p)}{\log v} \leq 3\log((q+2)^{3N}/\phi^{N-o(N)})/\log(q^{3N})$, which gives

$$\omega \leq 3N\log((q+2)^3/\phi)/(3N\log(q)) + 3\log_{q^3}(\phi)(o(N)/N),$$

and as N goes to ∞, the second summand goes away, and we get.

$$\omega \leq \frac{\log \frac{(q+2)^3}{\phi}}{\log(q)}.$$

□

Let us now focus on solving the combinatorial problem. We will show that the solution is $\phi \geq 27/4$ which will give $\omega = 2.404$.

We start with $L = \{(I,J,K) \mid I,J,K \in \{0,1\}^{3N}, I_i + J_i + K_i = 2\ \forall i \in [3N], I,J,K \text{ have } N \text{ zeros}\}$. We will have two phases of zeroing out Is, Js and Ks and all the triples containing them.

Phase 1: Use hash functions to zero out and partition the triples into buckets, so that two triples in different buckets are independent.

Phase 2: Use a greedy procedure to remove dependencies within individual buckets.

We begin with Phase 1.

Phase 1. Let M be a prime that depends on N. We will pick M later to be any prime in the interval $[3\binom{2N}{2}, 4\binom{2N}{2}]$.

Let $S \subseteq [M]$, $|S| \geq M^{1-\epsilon}$ be from Theorem 3 such that if $a + c \equiv 2b \pmod M$ for some $a,b,c \in S$, then $a = b = c$.

Pick $w_0, w_1, \ldots, w_{3N}$ be independent uniformly random variables from $[M]$. We define three hash functions as follows.

$$h_X(I) = \sum_{i=1}^{3N} w_i \cdot I_i \bmod M,$$
$$h_Y(J) = \left(\sum_{i=1}^{3N} w_i \cdot J_i\right) + w_0 \bmod M,$$
$$h_Z(K) = \frac{1}{2}\left(\left(\sum_{i=1}^{3N} w_i \cdot (2-k_i)\right) + w_0\right) \bmod M.$$

Since we will pick M to be a large prime, the $\frac{1}{2}$ factor is well-defined.

Notice that if $I \neq I'$, then $h_X(I)$ and $h_X(I')$ are independent. Similarly, if $J \neq J'$, then $h_Y(J)$ and $h_Y(J')$ are independent, and if $K \neq K'$, then $h_Z(K)$ and $h_Z(K')$ are independent. Also, because of w_0, each $h_X(I)$ and $h_Y(J)$ are indepedent.

The remaining triples are

$$L' := \{(I,J,K) \mid I,J,K \in \{0,1\}^{3N}, I_i+J_i+K_i = 2\ \forall i \in [3N], I,J,K \text{ have } N \text{ zeros}, h_X(I), h_Y(J), h_Z(K) \in S\}.$$

Let $(I,J,K) \in L'$. Then

$$\begin{aligned} h_X(I) + h_Y(J) &\equiv \sum_{i=1}^{p} w_i \cdot (I_i + J_i) + w_0 \pmod M \\ &\equiv 2h_Z(K) \pmod M. \end{aligned}$$

Since $h_X(I), h_Y(J), h_Z(K) \in S$, and $h_X(I) + h_Y(J) \equiv 2h_Z(K) \pmod{M}$, we must have $h_X(I) = h_Y(J) = h_Z(K)$ by the property of S.

If we define $B_s = \{(I, J, K) \in L' | h_X(I) = h_Y(J) = h_Z(K) = s\}$ for every $s \in S$, we see that the B_s partition L', and also two triples in different buckets B_s are independent. Thus, it suffices to get rid of the dependence inside each B_s (this would be phase 2).

We have finished phase 1. Now we want to analyze how many triples we have left.

The number of triples in L is $\binom{3N}{N,N,N}$ since there must be N indices i where $I_i = 0, J_i = 1, K_i = 1$, N indices j where $I_j = 1, J_j = 0, K_j = 1$, and N indices k where $I_k = 1, J_k = 1, K_k = 0$. For each of the $\binom{3N}{N,N,N}$ triples (I, J, K),

$$\Pr[(I, J, K) \in B_s] = \Pr[h_X(I) \in B_s \wedge h_Y(J) \in B_s \wedge h_Z(K) \in B_s] = \Pr[h_X(I) \in B_s \wedge h_Y(J) \in B_s],$$

where the last equality is because K is completely determined from I and J.

Note that $h_X(I)$ and $h_Y(J)$ are independent from our previous discussion. Thus, $\Pr[X_I Y_J Z_K \in B_s] = \frac{1}{M^2}$.

Thus the total number of triples remaining in L' is in expectation

$$\binom{3N}{N,N,N}|S|/M^2 \geq \binom{3N}{N,N,N} \cdot \frac{1}{M^{1+\epsilon}}.$$

Phase 2. Now we begin Phase 2, where we take L' and its partitioning into the B_s and remove all dependencies within the B_s.

For each B_s, we build a list L_s of independent triples in B_s greedily. Initially, $L_s = \emptyset$. For each $(I, J, K) \in B_s$, we try to add (I, J, K) to L_s. If no triples in L_s share any variable with (I, J, K), then we successfully add (I, J, K) to L_s. Otherwise, without loss of generality assume L_s contains (I, J', K') (a triple sharing block I with (I, J, K); more generally, the shared block could be J or K, but these cases are symmetric).

In this case, we kill off all triples containing J as their Y-block. Here we keep (I, J', K') but may kill off many triples containing J. We will show we do not kill off too many triples overall.

Let's see how many triples get killed this way. Suppose ℓ triples get killed by killing off the triples containing J. Then there are $\binom{\ell}{2}$ **pairs** of triples that share J, and the pair (I, J, K) and (I, J', K') share I.

Thus, the remaining set B_s loses at least $\binom{\ell}{2} + 1 \geq \ell$ **pairs of triples** that share a block. Therefore, the total number of triples we kill is at most the number of pairs of triples sharing a block in B_s.

Now we analyze the *expected* number of pairs of triples sharing a block in B_s.

As we showed earlier, the expected number of triples in B_s is $\frac{\binom{3N}{N,N,N}}{M^2}$.

Now we analyze the expected number of unordered pairs $((I, J, K), (I, J', K'))$ where both are in B_s and $J \neq J', K \neq K'$. The expected number of pairs of triples sharing a block in B_s can be bounded by three times of the previous value, by symmetry of I, J, K.

First, consider the number of pairs $((I, J, K), (I, J', K'))$ such that (I, J, K) and (I, J', K') are both in the original list L before phase 1. The number of choices for I is $\binom{3N}{N}$ since I contains N 0's and $2N$ 1's. After we fix I, then each valid J that is together with I in a triple contains N 0's on the indices i where $I_i = 1$, and it must have 1s for all indices i where $I_i = 0$. Thus, there are $\binom{2N}{N}$ choices for J.

Since J' must be different from J, the number of choices for J' is $\binom{2N}{N} - 1$. After we fix I, J, J', the other two sets K and K' will be uniquely determined. Thus, the number of unordered pairs $((I, J, K), (I, J', K'))$ of triples in L is $\frac{1}{2}\binom{3N}{N}\binom{2N}{N}\left(\binom{2N}{N} - 1\right)$.

For each such pair $((I, J, K), (I, J', K'))$, we bound the probability that both of them fall into B_s.

$$\Pr[(I, J, K) \in B_s \wedge (I, J', K') \in B_s] = \Pr[h_X(I) = s \wedge h_Y(J) = s \wedge h_Y(J') = s].$$

Recall that $h_Y(J)$ is independent from $h_Y(J')$ since $J \neq J'$ and there will be a term w_i in $h_Y(J)$ but not in $h_Y(J')$. $(h_Y(J), h_Y(J'))$ is independent from $h_X(I)$ because of the w_0 term. Therefore,

$$\Pr[(I, J, K) \in B_s \wedge (I, J', K') \in B_s] = \frac{1}{M^3}.$$

Therefore, the expected number of pairs sharing a block in B_s is at most

$$P_s = 3 \cdot \frac{1}{2}\binom{3N}{N}\binom{2N}{N}\left(\binom{2N}{N} - 1\right) \cdot \frac{1}{M^3}.$$

By previous discussions,

$$\begin{aligned}\mathbb{E}[|L_s|] &\geq \binom{3N}{N,N,N} \cdot \frac{1}{M^2} - \frac{3}{2}\binom{3N}{N}\binom{2N}{N}\left(\binom{2N}{N} - 1\right) \cdot \frac{1}{M^3} \\ &\geq \frac{1}{M^2}\binom{3N}{N,N,N}\left(1 - \frac{3}{2M}\binom{2N}{N}\right).\end{aligned}$$

Now we can set M to be an arbitrary prime in the range $\left[3\binom{2N}{N}, 4\binom{2N}{N}\right]$, so that

$$\mathbb{E}[|L_s|] \geq \frac{1}{2M^2}\binom{3N}{N,N,N}.$$

We can sum over all $s \in S$ to get that the expected number of independent triples remaining is at least:

$$\begin{aligned}\mathbb{E}\left[\sum_{s\in S}|L_s|\right] &\geq |S| \cdot \frac{1}{2M^2}\binom{3N}{N,N,N} \\ &\geq \frac{1}{2M^{1+\epsilon}}\binom{3N}{N,N,N} \\ &\geq \frac{1}{2M^{\epsilon}} \cdot \frac{1}{4\binom{2N}{N}}\binom{3N}{N,N,N} \\ &= \frac{1}{8M^{\epsilon}}\binom{3N}{N}.\end{aligned}$$

We will then use Stirling's approximation which says $n! = \Theta(\sqrt{2\pi n}\,(n/e)^n)$ to bound the number of independent triples further:

Thus,

$$\mathbb{E}\left[\sum_{s\in S}|L_s|\right] \geq \Theta\left(\frac{1}{8M^{\epsilon}} \frac{\sqrt{6\pi N}(3N/e)^{3N}}{\sqrt{2\pi N}(N/e)^N\sqrt{4\pi N}(2N/e)^{2N}}\right) \geq c' \frac{1}{M^{\epsilon}\sqrt{N}}\left(\frac{27}{4}\right)^N,$$

for some constant c'.

There exist a choice of random variables so that $\sum_{s\in S}|L_s| \geq \mathbb{E}\left[\sum_{s\in S}|L_s|\right]$. We pick this setting of our variables $w_0, \ldots, w_{3N}$.

In this case, we obtain $\phi = \frac{27}{4}$ and $\phi^{N-o(N)}$ as our solution to the combinatorial problem.

References

[1] Don Coppersmith and Shmuel Winograd. *Matrix multiplication via arithmetic progressions.* Proceedings of the nineteenth annual ACM symposium on Theory of computing. 1987.

[2] Raphaël Salem, and Donald C. Spencer. *On sets of integers which contain no three terms in arithmetical progression.* Proceedings of the National Academy of Sciences of the United States of America 28.12 (1942): 561.

[3] Felix A. Behrend. *On sets of integers which contain no three terms in arithmetical progression.* Proceedings of the National Academy of Sciences of the United States of America 32.12 (1946): 331.