COMS 6998: Algebraic Techniques in TCS (Fall'21) December 2, 2021

Lecture 11: The Laser Method

Instructor: *Josh Alman* Scribe notes by: *Elena Gribelyuk, Yunhao Wang*

Disclaimer: This draft may be incomplete or have errors. Consult the course webpage for the most up-to-date version.

1 Definitions and Background

First, we begin by providing the background information and definitions we will need to understand the laser method. Let T be a tensor over sets of variables X, Y, and Z, and partition the set of variables $X = X_0 \cup X_1 \cup \cdots \cup X_q$, $Y = Y_0 \cup Y_1 \cup \cdots \cup Y_q$, and $Z = Z_0 \cup Z_1 \cup \cdots \cup Z_q$. Also, for $i, j, k \in \{0, 1, \cdots, q\}$, write $T_{ijk} = T|_{X_i, Y_j, Z_k}$. Then, we see that

$$T^{\otimes N} = (\sum_{i,j,k} T_{ijk})^{\otimes N} = \sum_{(T_1, \cdots, T_N) \in \{T_{ijk}\}^N} T_1 \otimes T_2 \otimes \cdots \otimes T_N.$$

In other words, when T is partitioned into these subtensors, the Nth power of T is also partitioned into products of subtensors. Now, we will introduce a key idea which will motivate the laser method.

Definition 1. *Fix a probability distribution $\alpha_{ijk} \geq 0$ for each T_{ijk}, so that $\sum_{i,j,k} \alpha_{ijk} = 1$.*

In particular, we say that $(T_1, \cdots, T_N)$ *conforms* to distribution α if $\forall i, j, k$, we have that

$$|\{l \text{ such that } T_l = T_{ijk}\}| = \alpha_{ijk} \cdot N.$$

We make two assumptions. (1) We assume that our tensor T is symmetric, i.e., $T_{ijk} = T_{jki}$ up to rotations. This assumption is without loss of generality since we can make the tensor symmetric by adding all rotations in the same way as the last lecture. (2) We assume that α is also symmetric ($\alpha_{ijk} = \alpha_{jki}$). We can assume this because for all known methods the bound on ω is minimized when α is symmetric.

Our goal is to find $T^{\otimes N} \geq_{zo} T'$, where T' is a direct sum of c subtensors that conform to the probability distribution α.

Question: how big can we hope for c to be? Notice that $T_{i_1 j_1 k_1} \otimes \cdots \otimes T_{i_N j_N k_N}$ uses X-variables from $X_{i_1} \times \cdots \times X_{i_N}$. We say that $X_{i_1} \times \cdots \times X_{i_N}$ conforms to α if $\forall i \in \{0, 1, \cdots, q\}$,

$$|\{l \text{ such that } i_l = i\}| = \alpha_i \cdot N, \text{ where } \alpha_i := \sum_{j,k} \alpha_{ijk}.$$

If $T_{i_1 j_1 k_1} \otimes \cdots \otimes T_{i_N j_N k_N}$ conforms to α, then $X_{i_1} \times \cdots \times X_{i_N}$ must also conform to α. So c is bounded by the number of $X_{i_1} \times \cdots \times X_{i_N}$ that conforms to α, which is

$$\binom{N}{\alpha_0 N, \alpha_1 N, \cdots, \alpha_q N} \geq c.$$

As a shorthand we will denote this multinomial coefficient as $\binom{N}{\alpha_i N}$.

Now, we're ready to state the laser method.

Theorem 2 (Laser method)**.** *If T satisfies the following two conditions, then $T^{\otimes N} \geq_{zo} T'$, where T' is the direct sum of $c = \binom{N}{\alpha_i N}^{1-o(1)}$ subtensors that conform to the probability distribution α.*

1. *The marginals of α uniquely determine α, i.e., given α_i for all $i \in [q]$, there is a unique choice of α_{ijk} for all $i, j, k \in [q]$.*

2. *There is a number P such that if $T_{ijk} \neq 0$, then $i + j + k = P$.*

Proof. **Definitions.** We first define a few notations. For any $I = (i_1, \cdots, i_N)$, we define $X_I := X_{i_1} \times \cdots \times X_{i_N}$. We define $S_\alpha \subset \{0, 1, \cdots, q\}^N$ to be the set of all $(i_1, \cdots, i_N)$'s which conforms to α. Note that $S_\alpha = \binom{N}{\alpha_i N}$. For any $I, J, K \in S_\alpha$, we define $T_{IJK} := \bigotimes_{l=1}^N T_{i_l j_l k_l}$.

Step 1. For the first step, we will zero out all variables in X_I, Y_I, Z_I, where I does not conform to α. As a result, we get a tensor

$$T' = \sum_{I,J,K \in S_\alpha} T_{IJK}. \tag{1}$$

Note that by condition 1, every subtensor T_{IJK} in T' conforms to our distribution α. So, by zeroing-out the X_I, Y_I, Z_I which don't conform to the marginals of α, we've actually zero-ed out every subtensor which doesn't conform to α. Thus, the total number of subtensors T_{IJK} in T' is

$$\binom{N}{\alpha_{ijk} N} \tag{2}$$

For each fixed X_I, the number of subtensors in T' that use X_I is

$$\frac{\binom{N}{\alpha_{ijk} N}}{\binom{N}{\alpha_i N}} =: R. \tag{3}$$

More definitions. Let M be a prime number in the range $[300R, 600R]$. Recall that by condition 2, there exists a number P such that if $T_{ijk} \neq 0$ then $i+j+k = P$. Define hash functions $h_x, h_j, h_z : S_\alpha \to \mathbb{Z}_M$ as follows: Pick a random $w \in \mathbb{Z}_M^N$ and a random $w_0 \in \mathbb{Z}_M$. For any $I, J, K \in S_\alpha \subset \{0, 1, \cdots, q\}^N$, let

$$\begin{aligned} h_x(I) &= 2\langle I, w \rangle \pmod{M}, \\ h_y(J) &= 2w_0 + 2\langle J, w \rangle \pmod{M}, \\ h_z(K) &= w_0 + \langle P - K, w \rangle \pmod{M}, \end{aligned}$$

where $P - K := (P - k_1, P - k_2, \cdots, P - k_N)$ for $K = (k_1, k_2, \cdots, k_N)$.

Let's observe some key properties of these hash functions:

1. If T_{IJK} conforms to α and $T_{IJK} \neq 0$, then $h_x(I) + h_y(J) = 2h_z(K)$. (Proof: $2\langle I, w \rangle + 2\langle J, w \rangle + 2w_0 = 2\langle I + J, w \rangle + 2w_0 = 2\langle P - K, w \rangle + 2w_0 = 2h_z(K)$).

2. If T_{IJK} conforms to α, then $h_x(I), h_y(J), h_z(K)$ are uniformly random numbers mod M, even when conditioned on one of the other two.

The last ingredient is the following lemma:

Lemma 3. *There is a subset $A \subseteq \mathbb{Z}_M$ of size $|A| \geq M^{1-o(1)}$, such that if $a, b, c \in A$ and $a + b = 2c$, then $a = b = c$.*

This lemma basically states that there is a subset $A \subseteq \mathbb{Z}_M$ which doesn't contain any 3-term arithmetic progressions. We will prove the lemma later.

Step 2. Given the set A of the lemma, we zero out all X_I for which $h_x(I) \notin A$, similarly we zero out all Y_J and Z_K for which $h_y(J) \notin A$ and $h_z(K) \notin A$.

Consider any subtensor T_{IJK} in T' of Eq. (1). What is the probability that we did not zero it out, i.e., the probability that $h_x(I) \in A$, $h_y(J) \in A$, and $h_z(K) \in A$?

- $h_x(I) \in A$ with probability $\frac{|A|}{M} = \frac{1}{M^{o(1)}}$, since $h_x(I)$ is an uniformly random integer mod M.
- Notice that since there is no 3-term arithmetic progression in A, but we have $h_x(I), h_y(J), h_z(K)$ satisfy the arithmetic progression $h_x(I) + h_y(J) = 2h_z(K)$, so according to Lemma 3, we need to have $h_x(I) = h_y(J) = h_z(K)$.

  $h_y(J) = h_x(I)$ with probability $\frac{1}{M}$, since $h_y(J)$ is an uniformly random integer mod M even when h_x is fixed.
- Finally, notice that whenever we fix two of the three hash functions, the last one is also fixed, so $h_z(K) = h_x(I)$ with probability 1.

So in total, the probability that we don't zero out T_{IJK} is

$$\frac{1}{M^{o(1)}} \cdot \frac{1}{M} \cdot 1 = \frac{1}{M^{1+o(1)}}. \tag{4}$$

Consider any T_{IJK} and $T_{I'J'K'}$ in T' that share variables. What is the probability that we did not zero out either of them?

Notice that T_{IJK} and $T_{I'J'K'}$ share variables means that either $I = I'$, or $J = J'$, or $K = K'$. They can share at most one variable, because if they share two variables, then by condition 2 (if $T_{ijk} \neq 0$ then $i + j + k = P$) they must share all of the three variables, in which case we have $T_{IJK} = T_{I'J'K'}$. Then, without loss of generality, we can assume that $I = I', J \neq J', K \neq K'$. By a similar argument as before, neither of T_{IJK} and $T_{I'J'K'}$ is zeroed out when $h_x(I) \in A$, $h_y(J) = h_x(I)$, and $h_y(J') = h_x(I') = h_x(I)$.

- $h_x(I) \in A$ with probability $\frac{1}{M^{o(1)}}$.
- $h_y(J) = h_x(I)$ with probability $\frac{1}{M}$.
- $h_y(J') = h_x(I)$ with probability $\frac{1}{M}$.

Therefore, in total the probability is

$$\frac{1}{M^{o(1)}} \cdot \frac{1}{M} \cdot \frac{1}{M} = \frac{1}{M^{2+o(1)}}. \tag{5}$$

Step 3. Finally, we repeatedly pick I such that two or more subtensors use X_I, and we zero out X_I. We zero out such Y_J and Z_K in the same way.

If we remove $b \geq 2$ subtensors when zeroing out X_I, we effectively removed $b \cdot (b-1) \geq b$ pairs that share variables. So as long as the number of subtensors is much larger than the number of pairs that share variables, after this step we are still left with a large number of subtensors. Using Eq. (4) and (2), the expected number of subtensors after Step 2 is $\frac{1}{M^{1+o(1)}} \cdot \binom{N}{\alpha_{ijk}N}$. Using Eq. (3), each subtensor T_{IJK} in T' shares X_I with R other subtensors, and similarly it shares Y_J or Z_K with R other subtensors, so there are in total $\binom{N}{\alpha_{ijk}N} \cdot 3R$ number of pairs that share variables in T'. Then using Eq. (5), the expected number of pairs that share variables after Step 2 is $\frac{1}{M^{2+o(1)}} \cdot \binom{N}{\alpha_{ijk}N} \cdot 3R$.

Thus, the expected number of remaining subtensors after Step 3 is

$$\begin{aligned} &\geq \frac{1}{M^{1+o(1)}} \cdot \binom{N}{\alpha_{ijk}N} - \frac{1}{M^{2+o(1)}} \cdot \binom{N}{\alpha_{ijk}N} \cdot 3R \\ &\geq \frac{1}{M^{1+o(1)}} \cdot \binom{N}{\alpha_{ijk}N} \cdot (1 - \frac{3R}{M}) \\ &\geq \frac{0.99}{M^{1+o(1)}} \cdot \binom{N}{\alpha_{ijk}N} \geq \binom{N}{\alpha_i N}^{1-o(1)} =: c. \end{aligned}$$

where the second step follows from $M \geq 300R$, and the third step follows from $M \leq 600R$ and $R = \frac{\binom{N}{\alpha_{ijk}N}}{\binom{N}{\alpha_i N}}$.

This finishes the proof of the laser method. $\square$

At last, we prove Lemma 3. The set $A \subseteq \mathbb{Z}_M$ in which no three numbers form an arithmetic progression is called a Salem–Spencer set [SS42]. It was later improved by Behrend [Beh46].

Proof of Lemma 3. Let n, d be two integers that depend on M and will be fixed later. Consider the set $\{1, 2, \cdots, n\}^d$ which has n^d points. Consider the spheres ${x_1}^2 + {x_2}^2 + \cdots + {x_d}^2 = t$ for $t \in \{1, 2, 3, \cdots, dn^2\}$, and in total there are dn^2 spheres. There must exist at least one sphere which contains $\geq \frac{n^d}{dn^2}$ points from $\{1, 2, \cdots, n\}^d$. We fix such a sphere. Notice that since all these $\geq \frac{n^d}{dn^2}$ points are on a sphere, there is no arithmetic progression.

Then we map d-dimension points to integers:

$$(x_1, \cdots, x_d) \to x_1 + (2n+1)x_2 + \cdots + (2n+1)^{d-1}x_d.$$

We let A be the set of $\geq \frac{n^d}{dn^2}$ integers obtained by this mapping. A good property of this mapping is that since $x_i \in \{1, \cdots, n\}$, the sum of two x_i's is at most $2n$, and since we use $2n+1$ as the base, there is no carry-over. Therefore, since there is no arithmetic progression in the original points, there is also no arithmetic progression in the set A.

It remains to fix the values of d and n. Pick $d = \sqrt{\log M}$ and $n = \frac{M^{1/d}-1}{2}$, so that the maximum possible integer in A is $(2n+1)^d = M$. The size of A is bounded by

$$|A| \geq \frac{n^d}{dn^2} = \frac{M}{2^d} \cdot \frac{1}{dn^2} = M^{1-o(1)}. \qquad \square$$

2 Omega Bound with Copersmith-Winograd Tensor

'Simple' Coppersmith-Winograd tensor. The 'simple' Coppersmith-Winograd tensor is defined as

$$T = \sum_{i=1}^{q} (x_0 y_i z_i + x_i y_0 z_i + x_i y_i z_0).$$

We have $\underline{R}(T) \le q + 2$ (proof see Handout 3).

Partition the set X as $X_0 = \{x_0\}$, $X_1 = \{x_1, \cdots, x_q\}$, and similarly partition $Y = Y_0 \cup Y_1$ and $Z = Z_0 \cup Z_1$. The three non-zero tensors are $T_{011} = \langle 1, 1, q\rangle$, $T_{101} = \langle q, 1, 1\rangle$, and $T_{110} = \langle 1, q, 1\rangle$. The Kronecker product of N such tensors always has volume q^N.

We set the probability α as $\alpha_{110} = \alpha_{101} = \alpha_{011} = \frac{1}{3}$, hence $\alpha_0 = \frac{1}{3}$ and $\alpha_1 = \frac{2}{3}$.

It's easy to check that both T and α are symmetric, and the two conditions of Theorem 2 are satisfied. Applying the laser method (Theorem 2), we can zero out $T^{\otimes N}$ into c disjoint tensors of volume q^N, and

$$c = \binom{N}{\frac{1}{3}N, \frac{2}{3}N}^{1-o(1)} = \binom{N}{\frac{1}{3}N}^{1-o(1)} = \left(\frac{3}{2^{2/3}}\right)^{N-o(N)},$$

where the last step follows from the bionomial bound that $\binom{N}{pN} = \left(\frac{1}{p^p(1-p)^{1-p}}\right)^{N-o(N)}$.

Then applying the asymptotic sum inequality, we have

$$\omega \le 3 \cdot \frac{\log\left(\frac{(q+2)^N}{c}\right)}{\log\left(q^N\right)} \approx 3 \cdot \frac{\log\left(\frac{q+2}{3/2^{2/3}}\right)}{\log\left(q\right)} \overset{q=8}{\Longrightarrow} \omega \le 2.404.$$

Coppersmith-Winograd tensor. Next consider the Coppersmith-Winograd tensor

$$T = \sum_{i=1}^{q} (x_0 y_i z_i + x_i y_0 z_i + x_i y_i z_0) + x_0 y_0 z_{q+1} + x_0 y_{q+1} z_0 + x_{q+1} y_0 z_0.$$

We have $\underline{R}(T) \le q + 2$ (proof see Handout 3).

Partition the set X into $X_0 = \{x_0\}$, $X_1 = \{x_1, \cdots, x_q\}$, and $X_2 = \{x_{q+1}\}$. Similarly we partition $Y = Y_0 \cup Y_1 \cup Y_2$, and $Z = Z_0 \cup Z_1 \cup Z_2$. The non-zero tensors are $T_{011} = \langle 1, 1, q\rangle$, $T_{101} = \langle q, 1, 1\rangle$, $T_{110} = \langle 1, q, 1\rangle$, and $T_{002} = T_{020} = T_{200} = \langle 1, 1, 1\rangle$. (Note that all non-zero T_{ijk} have $i + j + k = 2$.)

We set $\alpha_{110} = \alpha_{101} = \alpha_{011} = a$ and $\alpha_{002} = \alpha_{020} = \alpha_{200} = \frac{1}{3} - a$, so the marginal probabilities are $\alpha_0 = \frac{2}{3} - a$, $\alpha_1 = 2a$ and $\alpha_2 = \frac{1}{3} - a$.

Applying the laser method (Theorem 2) and the asymptotic sum inequality in the same way as before, and optimize over the parameters q and a, we have that the best bound reached when $q = 6$ and $a \approx 0.3$, and $\omega \le 2.387$.

Square of Coppersmith-Winograd tensor. Let T be the Coppersmith-Winograd tensor, and let $T' = T^{\otimes 2}$. For more details see Section 3 of Handout 3. By applying the laser method to T', we have $\omega \le 2.376$ [CW87].

And by applying the laser method to the 32-th power of T, we have $\omega \le 2.37287$ [LG14].

Current best ω. By applying a whole new idea, people are able to reach $\omega \leq 2.37286$ [AW21], which is the current best upper bound on ω.

References

[AW21] Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 522–539. SIAM, 2021.

[Beh46] 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):331, 1946.

[CW87] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 1–6, 1987.

[LG14] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation*, pages 296–303, 2014.

[SS42] 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):561, 1942.