COMS 6998: Algebraic Techniques in TCS (Fall'21)	Nov 23, 2021

Lecture 10: Reductions, Asymptotic Sum Ineq, Laser Method

Instructor: *Josh Alman* Scribe notes by: *Ziyao Zhang, Yun Lu*

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

1 Reductions Between Tensors

We introduce three types of reductions between tensors. Let A, B be tensors. These three reductions will satisfy that if A "$\geq$" B, then $R(A) \geq R(B)$ or $\underline{R}(A) \geq \underline{R}(B)$.

1.1 Zeroing Out

We say B can be *zeroing out* from A, denoted by $A \geq_{zo} B$, if we can get from A to B by setting some variables in A to 0.

Example.

$$A = x_0 y_0 z_0 + x_0 y_0 z_1,$$
$$B = x_0 y_0 z_0.$$

We can get from A to B by setting z_1 to 0.

Rank. If $A \geq_{zo} B$, then $R(A) \geq R(B)$. This is because a rank-one tensor is still a rank-one tensor after setting some variables to zero. So zeroing out could only decrease the rank.

1.2 Restriction

Suppose A is over variables X, Y, Z and B is over variables X', Y', Z'. We say B is a *restriction* of A, denoted by $A \geq B$, if there are linear maps

$$M_X : \mathbb{F}^X \to \mathbb{F}^{X'},$$
$$M_Y : \mathbb{F}^Y \to \mathbb{F}^{Y'},$$
$$M_Z : \mathbb{F}^Z \to \mathbb{F}^{Z'}$$

such that if

$$A = \sum_{x \in X, y \in Y, z \in Z} A[x, y, z] \cdot xyz$$

where $A[x, y, z]$ denotes the coefficient, then

$$B = \sum_{x \in X, y \in Y, z \in Z} A[x, y, z] \cdot \, M_X(x) M_Y(y) M_Z(z).$$

Example.

$$\begin{aligned} A &= x_1 y_1 z_1, \\ B &= x_1 y_1 z_1 + x_1 y_1 z_2, \\ M_X &= M_Y = I, \\ M_Z(z_1) &= z_1 + z_2. \end{aligned}$$

Here I denote the identity map. We can see that $A \geq M_X(x_1) M_Y(y_1) M_Z(z_1) = x_1 y_1 (z_1 + z_2) = B$.

Rank. If $A \geq B$, then $R(A) \geq R(B)$. This is because a rank-one tensor is the product of three linear combinations of X, Y, Z respectively, and after combining with the linear maps M_X, M_Y, M_Z, it is still a product of three linear combinations.

Proposition 1. *$R(T) \leq r$ if and only if T is a restriction of $\langle r \rangle$, where $\langle r \rangle$ denotes the size-r diagonal tensor:*

$$\langle r \rangle = \sum_{i=1}^{r} x_i y_i z_i.$$

Proof. "if": If $T \leq \langle r \rangle$, from the previous claim we have $R(T) \leq R(\langle r \rangle) = r$.

"only if": This follows by the definition. Recall that any rank-one tensor can be represented as $(\sum_i \alpha_i x_i)(\sum_i \beta y_i)(\sum_i \gamma_i z_i)$, so it is always a restriction of a single monomial $x_1 y_1 z_1$. □

1.3 Monomial Degeneration

Suppose A is over variables X, Y, Z. We say B is a *monomial degeneration* of A, denoted by $A \trianglerighteq_{md} B$, if there is a map that sends variables in A to a single variable ϵ:

$$m : X \cup Y \cup Z \to \{\epsilon^h \mid h \in \mathbb{Z}\}$$

such that for $\forall x \in X, y \in Y, z \in Z$,

1) If $B[x, y, z] \neq 0$, then $B[x, y, z] = A[x, y, z]$ and $m(x)m(y)m(z) = 1$.
2) If $B[x, y, z] = 0$, then $m(x)m(y)m(z) = \epsilon^h$ for some $h > 0$.

Example.

$$\begin{aligned} A &= x_0 y_1 z_1 + x_1 y_0 z_1 + x_1 y_1 z_0 + x_0 y_0 z_0, \\ B &= x_0 y_1 z_1 + x_1 y_0 z_1 + x_1 y_1 z_0, \\ m(x_0) &= m(y_0) = \epsilon, \\ m(x_1) &= m(y_1) = 1, \\ m(z_0) &= 1, m(z_1) = 1/\epsilon. \end{aligned}$$

Then $m(x_0)m(y_0)m(z_0) = \epsilon^2$, and other terms get 1.

Rank. If $A \trianglerighteq_{md} B$, then $\underline{R}(A) \geq \underline{R}(B)$. This follows from the definition of border rank.

Monomial degeneration is useful for proving border rank upper bounds. In the example above, $R(A) = 2$, $R(B) = 3$, but $\underline{R}(B) \leq 2$.

1.4 Extra: Degeneration

This type of reduction is a combination of restriction and monomial degeneration. We use $A \trianglerighteq B$ to denote that B is a *degeneration* of A. Similar to restriction, there are linear maps M_X, M_Y, M_Z. The difference is that

$$M_X : \mathbb{F}^X \to \mathbb{F}[\epsilon, 1/\epsilon]^{X'},$$

where $\mathbb{F}[\epsilon, 1/\epsilon]$ denotes polynomials over ϵ and $1/\epsilon$. An example of the linear map is $M_X(x_1) = \epsilon x_1 + \frac{1}{\epsilon} x_2$.

Rank. If $A \trianglerighteq B$, then $\underline{R}(A) \geq \underline{R}(B)$.

2 Asymptotic Sum Inequality

See Section 2 (Disjoint Sum Identity) of Handout 2. We have

$$\underline{R}(\langle 4, 1, 4 \rangle \oplus \langle 1, 9, 1 \rangle) \leq 17, \tag{1}$$

where $\oplus$ denotes the disjoint sum, which is the sum of two tensors that do not share any variable. This is a surprising result because $\underline{R}(\langle 4, 1, 4 \rangle) = 16$ and $\underline{R}(\langle 1, 9, 1 \rangle) = 9$, but the border rank of their direct sum is much smaller than $16 + 9$!

To make use of Eq. (1) which bounds the border rank of a direct sum, we use the following theorem by Schönhage [Sch81].

Theorem 2 (Asymptotic Sum Inequality). *If $\underline{R}(\bigoplus_{i=1}^p \langle k_i, m_i, n_i \rangle) \leq r$, then $\sum_{i=1}^p (k_i \cdot m_i \cdot n_i)^{\omega/3} \leq r$.*

As a sanity check, when $p = 1$ we have $(kmn)^{\omega/3} \leq \underline{R}(\langle k, m, n \rangle)$, which is proved last time.

Using this theorem and Eq. (1), we have

$$16^{\omega/3} + 9^{\omega/3} \leq 17 \implies \omega \leq 2.55.$$

To prove this theorem - a more general case, we first prove a lemma, which considers the case where all k_i's are equal, all m_i's are equal and all n_i's are equal.

Lemma 3. *If $R(f \odot \langle k, m, n \rangle) \leq g$, then $\omega \leq 3 \cdot \frac{\log \lceil g/f \rceil}{\log (kmn)}$.*

Here $f \odot \langle k, m, n \rangle$ denotes the disjoint sum of f many copies of tensor $\langle k, m, n \rangle$, i.e. $\bigoplus_{i=1}^f \langle k, m, n \rangle$. Notice that $f \odot \langle k, m, n \rangle = \langle f \rangle \otimes \langle k, m, n \rangle$. (Recall that $\langle f \rangle := \sum_{i=1}^f x_i y_i z_i$.)

Be patient. To prove this lemma, we need to prove a claim.

Claim 4. *If $R(f \odot \langle k, m, n \rangle) \leq g$, then $R(f \odot \langle k^s, m^s, n^s \rangle) \leq \lceil g/f \rceil^s \cdot f$ for any positive integer s.*

Proof. We prove by induction on s. When $s = 1$, it is exactly the condition. In the induction step,

$$\begin{aligned} f \odot \left\langle k^{s+1}, m^{s+1}, n^{s+1} \right\rangle &= \langle f \rangle \otimes \left\langle k^{s+1}, m^{s+1}, n^{s+1} \right\rangle \\ &= \langle f \rangle \otimes \langle k, m, n \rangle \otimes \langle k^s, m^s, n^s \rangle \\ &\leq \langle g \rangle \otimes \langle k^s, m^s, n^s \rangle \\ &= g \odot \langle k^s, m^s, n^s \rangle . \end{aligned}$$

The second last step is due to the fact that $\langle f \rangle \otimes \langle k, m, n \rangle$ has rank at most g, so it is a restriction of $\langle g \rangle$. Therefore,

$$\begin{aligned} R(f \odot \left\langle k^{s+1}, m^{s+1}, n^{s+1} \right\rangle) &\leq R(g \odot \langle k^s, m^s, n^s \rangle) \\ &\leq R(\lceil g/f \rceil \cdot f \odot \langle k^s, m^s, n^s \rangle) \\ &\leq \lceil g/f \rceil \cdot \lceil g/f \rceil^s \cdot f \\ &= \lceil g/f \rceil^{s+1} \cdot f. \end{aligned}$$

□

Now we can prove the lemma.

Proof of Lemma 3. If $R(f \odot \langle k, m, n \rangle) \leq g$, using Claim 4 we have

$$\begin{aligned} R(\langle k^s, m^s, n^s \rangle) &\leq R(f \odot \langle k^s, m^s, n^s \rangle) \\ &\leq \lceil \frac{g}{f} \rceil^s \cdot f. \end{aligned}$$

Then we have

$$\begin{aligned} & \omega \leq 3 \cdot \frac{\log(\lceil \frac{g}{f} \rceil^s \cdot f)}{\log((kmn)^s)} = 3 \cdot \frac{s \cdot \log \lceil \frac{g}{f} \rceil + \log f}{s \cdot \log(kmn)} \\ \longrightarrow \quad & \omega \leq 3 \cdot \frac{\log \lceil g/f \rceil}{\log(kmn)} \qquad \text{As } s \text{ tends to be very large, } \log f \text{ becomes insignificant} \end{aligned}$$

□

Then we can prove the theorem.

Proof of Theorem 2. Let tensor $T = \bigoplus_{i=1}^p \langle k_i, m_i, n_i \rangle$. Taking its s-th Kronecker power,

$$T^{\otimes s} = \bigoplus_{a_1 + \cdots + a_p = s} \binom{s}{a_1, \cdots, a_p} \odot \left\langle \prod_{i=1}^p k_i^{a_i}, \prod_{i=1}^p m_i^{a_i}, \prod_{i=1}^p n_i^{a_i} \right\rangle .$$

In particular, for any choice of $a_1, \cdots, a_p$, we can zero out everything else to get just

$$\binom{s}{a_1, \cdots, a_p} \odot \left\langle \prod_{i=1}^p k_i^{a_i}, \prod_{i=1}^p m_i^{a_i}, \prod_{i=1}^p n_i^{a_i} \right\rangle$$

That means, by our assumption, $\underline{R}(\binom{s}{a_1, \cdots, a_p} \odot \langle \prod_{i=1}^p k_i^{a_i}, \prod_{i=1}^p m_i^{a_i}, \prod_{i=1}^p n_i^{a_i} \rangle) \leq \underline{R}(T^{\otimes s}) \leq r^s$. And similar to what we proved in the last lecture, we have

$$R\left(\binom{s}{a_1,\cdots,a_p}\odot\left\langle\prod_{i=1}^{p}k_i^{a_i},\prod_{i=1}^{p}m_i^{a_i},\prod_{i=1}^{p}n_i^{a_i}\right\rangle\right)=R(T^{\otimes s})\leq r^s\cdot p(s),$$

where $p(s)$ is some polynomial of s. Then we can apply Lemma 3 to get a specific bound on ω:

$$\binom{s}{a_1,\cdots,a_p}\cdot\left(\prod_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{a_i}\right)^{\frac{\omega}{3}}\leq r^s\cdot p(s) \tag{2}$$

Now we sum it over all the possible choices of a_i's, so to show the average of these values are high. Then by binomial theorem,

$$\sum_{a_1+\cdots+a_p=s}\binom{s}{a_1,\cdots,a_p}\cdot\prod_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{a_i\cdot\frac{\omega}{3}}=\left(\sum_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{\frac{\omega}{3}}\right)^s.$$

Pick $a_1,\cdots,a_p$ that maximize the term, then we have

$$\binom{s}{a_1,\cdots,a_p}\cdot\prod_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{a_i\cdot\frac{\omega}{3}}\geq\frac{(\sum_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{\frac{\omega}{3}})^s}{\binom{p+s-1}{p-1}}. \tag{3}$$

Note that $\binom{p+s-1}{p-1}$ is the number of choices of $a_1,\cdots,a_p$ that sums to s, and it is a polynomial of s with degree p. Combining Eq. (2) and (3) and take $s\to\infty$, we get $\sum_{i=1}^{p}(k_i\cdot m_i\cdot n_i)^{\omega/3}\leq r$. □

3 Strassen's Tensor and Laser Method

After proving $\omega\leq 2.55$, people conjectured that $\omega=2.5$. This was disproved by Strassen, who showed that $\omega\leq 2.48$ [Str87].

Strassen's tensor. Strassen's tensor is defined as

$$Str=\sum_{i=1}^{q}(x_iy_0z_i+x_0y_iz_i).$$

Note that $Str=\langle q,1,1\rangle+\langle 1,q,1\rangle$ (not direct sum). Strassen proved that

$$\underline{R}(Str)\leq q+1,$$

which is much smaller than $2q$. See Handout 2 for proof.

We cannot directly apply the asymptotic sum inequality since Strassen's tensor has sum instead of direct sum, i.e., the two tensors share variables. To deal with this, Strassen developed a technique called the laser method.

Laser Method. Strassen observed that the tensor Str has an outer structure and an inner structure. Let $X=\{x_0,\cdots,x_q\}$, $Y=\{y_0,\cdots,y_q\}$, $Z=\{z_1,\cdots,z_q\}$, so that Str is a tensor over X,Y,Z. Define $X_0=\{x_0\}$, $X_1=\{x_1,\cdots,x_q\}$. Similarly define Y_0 and Y_1. Note that $X_0\cup X_1=X$ and $Y_0\cup Y_1=Y$.

Define $Z_1 = Z$. The inner structure are defined as the set of tensors $Str_{ijk} = Str|_{X_i,Y_j,Z_k}$, e.g.,

$$Str_{011} = Str|_{X_0,Y_1,Z_1} = \sum_{i=1}^{q} x_0 y_i z_i = \langle 1, 1, q\rangle.$$

It's easy to see the only two non-zero tensors are $Str_{011} = \langle 1, 1, q\rangle$ and $Str_{101} = \langle q, 1, 1\rangle$, and they both have volume q. (The volume is defined as $\mathrm{Vol}(\langle k, m, n\rangle) := kmn$.)

The outer structure is defined as a tensor T over $\{0, 1\}$, $\{0, 1\}$, and $\{1\}$ such that $T[i, j, k] = 1$ if $Str_{ijk} \neq 0$ and $T[i, j, k] = 0$ otherwise. It's easy to see that

$$T = x_0 y_1 z_1 + x_1 y_0 z_1 = \langle 1, 2, 1\rangle.$$

We use " $\otimes$ " to denote the operation that combines the outer structure with the inner structure: $Str = \langle 1, 2, 1\rangle$"$\otimes$"$\{\langle 1, 1, q\rangle, \langle q, 1, 1\rangle\}$. Strassen makes the tensor Str symmetric. After permutations on x, y, z, we get the following two tensors which also have border rank $\leq q + 1$:

$$Str' = \langle 1, 1, 2\rangle\text{“}\otimes\text{”}\{\langle q, 1, 1\rangle, \langle 1, q, 1\rangle\},$$
$$Str'' = \langle 2, 1, 1\rangle\text{“}\otimes\text{”}\{\langle 1, q, 1\rangle, \langle 1, 1, q\rangle\}.$$

We leave it to the readers to check that the " $\otimes$ " operation satisfies that for tensors T, T' and sets S, S' $(T$"$\otimes$"$S) \otimes (T'$"$\otimes$"$S') = (T \otimes T')$"$\otimes$"$(S \otimes S')$, where $S \otimes S' := \{a \otimes a' \mid a \in S, a' \in S'\}$. Thus, after taking the Kronecker product of the three Strassen vectors, we get

$$Str \otimes Str' \otimes Str'' = \langle 2, 2, 2\rangle\text{“}\otimes\text{”}\mathrm{Vol}(q^3),$$

where with an abuse of notation we use $\mathrm{Vol}(q^3)$ to denote a set of tensors with volume q^3. Since $\underline{R}(Str \otimes Str' \otimes Str'') \leq (q + 1)^3$, taking the s-th Kronecker product of this tensor, we get

$$\underline{R}(\langle 2^s, 2^s, 2^s\rangle\text{“}\otimes\text{”}\mathrm{Vol}(q^{3s})) \leq (q + 1)^{3s}. \tag{4}$$

We then show that the tensor $\langle 2, 2, 2\rangle$"$\otimes$"$\mathrm{Vol}(q^3)$ can be reduced to a direct sum using monomial degeneration. We first prove the following proposition.

Proposition 5. $\langle n, n, n\rangle \trianglerighteq_{md} \langle \frac{3}{4}n^2\rangle$.

Proof. What we want to do is to take the $\langle n, n, n\rangle$ tensor and multiply each variable by a power of ϵ. For simplicity assume that n is odd and we write $n = 2m + 1$. The proof for even n is similar. Then $\langle n, n, n\rangle = \sum_{i=-m}^{m} \sum_{j=-m}^{m} \sum_{k=-m}^{m} x_{ik} y_{kj} z_{ij}$. We multiply x_{ik} by ϵ^{i^2+2ik}, y_{kj} by ϵ^{k^2+2kj} and z_{ij} by ϵ^{k^2+2ij}. Then the tensor $\langle n, n, n\rangle$ becomes

$$\sum_{i,j,k=-m}^{m} \epsilon^{(i+j+k)^2} x_{ik} y_{kj} z_{ij} \trianglerighteq_{md} \sum_{\substack{i,j,k=-m \\ s.t.\ i+j+k=0}}^{m} x_{ik} y_{kj} z_{ij}$$
$$\geq \langle \frac{3}{4} n^2\rangle,$$

where the last step follows by observing that after fixing any i and j, k is uniquely determined by

$i + j + k = 0$, so the tensors $x_{ik}y_{kj}z_{ij}$ do not share variables under the constraint $i + j + k = 0$. $\square$

This monomial degeneration implies that $\langle 2^s, 2^s, 2^s\rangle \trianglerighteq_{md} \langle \frac{3}{4}2^{2s}\rangle$. In fact, applying the same power of ϵ of the proposition to the inner tensors of $\langle 2^s, 2^s, 2^s\rangle$ "$\otimes$" $\mathrm{Vol}(q^{3s})$, we can reduce it to a direct sum of $\frac{3}{4}(2^{2s})$ number of volume-q^{3s} tensors. (For more details see e.g. Section 8 of [Blä13].) Thus using the asymptotic sum inequality (Theorem 2) we have

$$\underline{R}(\langle 2^s, 2^s, 2^s\rangle \text{“}\otimes\text{”} \mathrm{Vol}(q^{3s})) \geq \frac{3}{4}(2^{2s}) \cdot (q^{3s})^{\omega/3}. \tag{5}$$

Combining Eq. (4) and (5), we have

$$\frac{3}{4}(2^{2s})(q^{3s})^{\omega/3} \leq (q+1)^{3s}.$$

Then we take $q = 5$ and $s \to \infty$, and we get $\omega \leq 2.48$.

References

[Blä13] Markus Bläser. Fast matrix multiplication. *Theory of Computing*, pages 1–60, 2013.

[Sch81] Arnold Schönhage. Partial and total matrix multiplication. *SIAM Journal on Computing*, 10(3):434–455, 1981.

[Str87] Volker Strassen. Relative bilinear complexity and matrix multiplication. 1987.