Communication Complexity and Applications

Lecturer: Yuhao Li

1 Log-rank Conjecture

The log-rank conjecture is one of the most fundamental open problems in the communication complexity. It states that the deterministic communication complexity of a two-party Boolean matrix is polynomially related to the logarithm of the rank of the input matrix over reals.

The motivation that we try to use logarithm of the rank to fully capture the deterministic communication complexity (up to some polynomial factors) starts from the following theorem by Mehlhorn and Schmidt:

Theorem 1 ([11]). *Let $F : X \times Y \to \{0, 1\}$. Then*

$$\mathrm{P}^{\mathrm{CC}}(F) \geq \log \operatorname{rank}(M_F).$$

Throughout this lecture notes, we will use $\mathrm{P}^{\mathrm{CC}}(\cdot)$ to denote the deterministic communication complexity of the input Boolean matrix.

The log-rank conjecture, proposed by Lovász and Saks [9], states that the lower bound is tight up to polynomial factors.

Conjecture 2 (Log-rank Conjecture [9]). *There exists a universal constant C such that for every $F : X \times Y \to \{0, 1\}$,*

$$\mathrm{P}^{\mathrm{CC}}(F) \leq O((\log \operatorname{rank}(M_F))^C).$$

Despite it being fundamental, important, and fascinating, little progress has been made in the past three decades. Note that there is also a simple upper bound $\mathrm{P}^{\mathrm{CC}}(F) \leq \operatorname{rank}(M_F)$, but it is exponentially worse than the conjectured bound. Both of the upper bound and the lower bound have been non-trivially improved. We state the state-of-the-art as follows.

The best-known upper bound is from Lovett [10].

Theorem 3 ([10]). *For any Boolean function F,*

$$\mathrm{P}^{\mathrm{CC}}(F) \leq O(\sqrt{\operatorname{rank}(M_F)} \cdot \log \operatorname{rank}(M_F)).$$

The best-known lower bound is from Göös, Pitassi, and Watson [4], which implies that if the log-rank conjecture is true, then the universal constant C should be at least 2.

Theorem 4 ([4]). *There exists a sequence of functions f_n, whose log-rank goes to infinity, such that*

$$\mathrm{P}^{\mathrm{CC}}(F_n) = \tilde{\Omega}(\log^2 \operatorname{rank}(M_{F_n})).$$

There is another line studying the log-rank conjecture restricted to a special families of functions. In particular, there has been extensive attention on the lifting functions where the inner function is XOR (i.e., $F(x, y) = f(x \oplus y)$ for a boolean function f). Maybe coincidentally, the best-known upper bounds for these XOR functions are the same as the general upper bounds, with different techniques by Tsang, Wong, Xie, and Zhang [14].

Theorem 5 ([14])**.** *For any n-variable boolean function f, we let $f^{\oplus}(x, y) := f(x \oplus y)$. Then*

$$\mathrm{P}^{\mathrm{CC}}(f^{\oplus}) \le O(\sqrt{\operatorname{rank}(M_{f^{\oplus}})} \cdot \log \operatorname{rank}(M_{f^{\oplus}})).$$

In the rest of this notes, we will prove these two upper bounds for the XOR functions and general functions. Note that although Theorem 5 is a corollary of Theorem 3, the techniques are different and might be of independent interests. For the motivation to study log-rank conjecture for the lifting functions, we refer interested readers to a recent survey by Knop, Lovett, McGuire, and Yuan [6].

2 Proof of Theorem 5

Before we go into the details of the proof, let's recall the Fourier analysis and see how it connects with our log-rank conjecture.

2.1 Fourier Analysis

For any real function $f : \{0, 1\}^n \to \mathbb{R}$, the Fourier coefficients are defined as $\hat{f}(s) = 2^{-n} \sum_x f(x)\chi_s(x)$, where $\chi_s(x) := (-1)^{s \cdot x}$. Then the function f can be written as $f = \sum_s \hat{f}_s \chi_s$. The Fourier sparsity of f, denoted as $\|\hat{f}\|_0$, is defined as the number of nonzero Fourier coefficients. The ℓ_1-norm of $\hat{f}$, denoted as $\|\hat{f}\|_1$, is defined as $\sum_s |\hat{f}_s|$. We include the following simple fact:

Fact 6. *For any boolean function $f : \{0, 1\}^n \to \{0, 1\}$, we have $\|\hat{f}\|_1 \le \sqrt{\|\hat{f}\|_0}$.*

Proof. By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&(\|\hat{f}\|_1)^2 \\
=&\left(\sum_s \hat{f}_s \cdot \operatorname{sgn}(\hat{f}_s)\right)^2 \\
\le&\left(\sum_s \hat{f}_s^2\right) \cdot \left(\sum_s (\operatorname{sgn}(\hat{f}_s))^2\right) \\
=&\|\hat{f}\|_0.
\end{aligned}$$

□

A very helpful observation here is that, the rank of the communication matrix $\operatorname{rank}(M_{f^{\oplus}})$ is exactly equal to the Fourier sparsity $\|\hat{f}\|_0$.

Observation 7 ([1])**.** *For any function $f : \{0,1\}^n \to \mathbb{R}$, we have*

$$\operatorname{rank}(M_{f\oplus}) = \|\hat{f}\|_0.$$

We have related the RHS of the log-rank conjecture to Fourier sparsity. Now we will convert the LHS $\mathrm{P}^{\mathrm{CC}}(f^{\oplus})$ to the parity decision tree.

2.2 Parity Decision Tree

Parity decision tree is an extension of the standard decision tree model, where each internal node can query an arbitrary parity. For a boolean function $f : \{0,1\}^n \to \{0,1\}$, we denote the minimal depth of a parity decision tree by $\mathrm{PDT}(f)$. A simple observation is that $\mathrm{P}^{\mathrm{CC}}(f^{\oplus}) \leq 2 \cdot \mathrm{PDT}(f)$, since Alice and Bob can finish the communication by simulating the parity decision tree: At each internal node of the parity decision tree, when they need to compute a parity of $x \oplus y$ (where Alice knows x and Bob knows y), they locally compute the parity of x and y and exchange these two bits.

Observation 8. *For any boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\mathrm{P}^{\mathrm{CC}}(f^{\oplus}) \leq 2 \cdot \mathrm{PDT}(f).$$

2.3 Main Protocol

The main communication protocol is to construct a parity decision tree via a notion so called *degree kill rank*[1].

Definition 1 (Degree Kill Rank [14])**.** *Let $f \in \mathbb{F}_2[x_1, \cdots, x_n]$ be a boolean function and suppose its $\mathbb{F}_2$-degree $\deg_2(f) = d$. Then the degree kill rank of f, denoted by $\mathrm{DKR}(f)$, is defined as the minimum integer r such that f can be expressed as*

$$f = \ell_1 f_1 + \cdots + \ell_r f_r + f_0,$$

where the $\mathbb{F}_2$-degree of each ℓ_i is 1 and the $\mathbb{F}_2$-degree of each f_i is at most $d-1$.

The construction of parity decision based on the degree kill rank is as follows: Since each the $\mathbb{F}_2$-degree of each ℓ_i is 1, each ℓ_i is a parity. Then we query all $\ell_i(x)$ and get the answers a_i. Then we face a new function $f' = \sum_{i=1}^{r} a_i f_i + f_0$. Notice that the $\mathbb{F}_2$-degree of f' is at most $d-1$ and it can also be proven that both $\|\hat{f}\|_0$ and $\|\hat{f}\|_1$ are non-increasing. We recursively find the degree kill rank until f is a $\mathbb{F}_2$-degree 0 function.

Recall that $\mathbb{F}_2$-degree is upper bounded by $\log \|\hat{f}\|_0$ for any boolean function f. Notice that if we can prove the $\mathrm{DKR}(f)$ is upper bounded by $O(\log^C \|\hat{f}\|_0)$ for some universal constant C, then we know that $\mathrm{P}^{\mathrm{CC}}(f^{\oplus}) \leq O(\log^{C+1} \operatorname{rank}(M_{f\oplus}))$, which implies the validity of log-rank conjecture for XOR functions. Unfortunately we haven't known it yet. Tsang, Wong, Xie, and Zhang proved the following lemma to get Theorem 5.

Lemma 9 ([14])**.** *For any boolean function $f : \{0,1\}^n \to \{0,1\}$, we have $\mathrm{DKR}(f) \leq O(\|\hat{f}\|_1)$.*

[1] The name of degree kill rank is from the lecturer. In the papers [2, 14] the authors called it polynomial rank. The notion of rank of polynomials is from mathematics, but using "rank" would be confusing in the context of log-rank conjecture.

2.4 The Notion of Rank

Before proving the Lemma 9, I would like to explain a bit the notion of rank (of polynomials) in mathematics, and its connection and difference from our degree kill rank.

We quote the definition of rank (of polynomials) from a paper by Green and Tao [5].

Definition 2 (Rank, Definition 1.5 in [5])**.** *Let $d \geq 0$, and let $P : \mathbb{F}^n \to \mathbb{F}$ be a function. We define the degree d rank* $\mathrm{rank}_d(P)$ *of P to be the least interger $k \geq 0$ for which there exist polynomials $Q_1, \cdots, Q_k \in \mathcal{P}_d(\mathbb{F}^n)$ and a function $B : \mathbb{F}^k \to \mathbb{F}$ such that we have the representation $P = B(Q_1, \cdots, Q_k)$. If no such k exists, we declare* $\mathrm{rank}_d(P)$ *to be infinite (since $\mathbb{F}^n$ is finite-dimensional, this only occurs when $d = 0$ and P is non-constant.)*

From the view point of theoretical computer science, especially the query complexity, the notion of rank can be understood as the minimal *non-adaptive* queries of degree-d queries to know the function P. However, in our context, the communication protocol/parity decision tree should be *adaptive*. Note that in the main protocol, the queries of ℓ_i are also *non-adaptive*, but the adaptivity is from the construction of degree kill rank after knowing the $\mathbb{F}_2$-degree $d-1$ function f'.

2.5 Proof of Lemma 9

To finish the proof of Theorem 5, it remains for us to prove the Lemma 9 given that $\|\hat{f}\|_1 \leq \sqrt{\|\hat{f}\|_0} = \sqrt{\mathrm{rank}(M_{f\oplus})}$. We introduce a complexity measure called parity kill number, denoted by $C^{\oplus}_{\min}(f)$, where the name is from [13]. The parity kill number is defined as

$$C^{\oplus}_{\min}(f) = \min\{\text{co-dim}(H) | H \text{ is an affine subspace on which } f \text{ is a constant}\}.$$

We include a simple observation:

Observation 10 (Corollary 20 in [14])**.** *For any $f : \{0,1\}^n \to \{0,1\}$, we have* $\mathrm{DKR}(f) \leq C^{\oplus}_{\min}(f)$.

So it suffices for us to upper bound the parity kill number by $O(\|\hat{f}\|_1)$.

Proof. We consider a greedy folding process. We sort the parities to be $\gamma^1, \cdots, \gamma^m$ such that $|\hat{f}(\gamma^1)| \geq \cdots \geq |\hat{f}(\gamma^m)|$. We fold $\beta = \gamma^1 + \gamma^2$ and let $b = 0$ if $\mathrm{sgn}(\hat{f}(\gamma^1)) = \mathrm{sgn}(\hat{f}(\gamma^2))$; let $b = 1$ otherwise. Then we use $\sum_{i \in \beta} x_i = b$ as a linear restriction.

The crucial claims of this greedy folding process are as follows.

- After at most $O(\|\hat{f}\|_1)$ steps, the largest absolute value of Fourier coefficient will be at least 1/2, i.e., $|\hat{f}(\gamma^1)| \geq 1/2$;
- For each step, $\|\hat{f}\|_1$ will decrease $2 \cdot |\hat{f}(\gamma^1)|$.
- When $\|\hat{f}\|_1 \leq 1$, it will be a constant function after one more step.

So we have $C^{\oplus}_{\min} \leq O(\|\hat{f}\|_1)$.

For a more formal and complete proof, we refer interested readers to Lemma 30 in he original paper [14].

□

3 Proof of Theorem 3

Now we are going to prove the general upper bounds Theorem 3. (Maybe) Totally different from the previous proof, the techniques here are mainly based on analyzing discrepancy and large monochromatic rectangles.

We first recall the theorem by Nisan and Wigderson [12], which shows the equivalence between the log-rank conjecture and large monochromatic rectangles.

Theorem 11 ([12]). *Assume that for any function $f : X \times Y \to \{-1, 1\}$ of* $\mathrm{rank}(M_f) = r$ *there exists a monochromatic rectangle of size $|R| \geq 2^{-c(r)}|X \times Y|$. Then we have*

$$\mathrm{P}^{\mathrm{CC}}(f) \leq O(\log^2 r + \sum_{i=1}^{\log r} c(r/2^i)).$$

In particular, by master theorem, if $c(r) \leq p(r)$ for some polynomial p, then we have $\mathrm{P}^{\mathrm{CC}}(f) \leq p(r)$; if $c(r) \leq O(\log^C r)$ for some universal constant C, then we have $\mathrm{P}^{\mathrm{CC}}(f) \leq O(\log^{C+1} r)$.

We will include the Lovett's proof of $c(r) \leq O(\sqrt{r} \log r)$ by discrepancy, which leads to Theorem 3. Before that, let us introduce a lemma by Gavinsky and Lovett [3], which relaxes the condition in Theorem 11, showing that it actually suffices for us to find a large rectangle that is "closed" to be monochromatic. Let $\mathbb{E}[f] := \frac{1}{|X \times Y|} \sum_{(x,y)} f(x, y)$. For a subrectangle R, we let $\mathbb{E}[f] := \frac{1}{|R|} \sum_{(x,y)\in R} f(x, y)$

Lemma 12 ([3]). *Let $f : X \times Y \to \{-1, 1\}$ be a boolean function with* $\mathrm{rank}(M_f) = r$ *and* $\mathbb{E}[f] \geq 1 - 1/2r$*, then there exists a monochromatic rectangle R with $|R| \geq |X \times Y|/8$.*

3.1 Discrepancy Method

Let $f : X \times Y \to \{-1, 1\}$ and μ be a distribution over $X \times Y$. The discrepancy of f w.r.t. μ is

$$\mathrm{disc}_\mu := \max_R \left| \sum_{(x,y)\in R} \mu(x, y) f(x, y) \right|.$$

The discrepancy of f is its discrepancy for the worse case distribution,

$$\mathrm{disc} := \min_\mu \max_R \left| \sum_{(x,y)\in R} \mu(x, y) f(x, y) \right|.$$

From [7, 8], we know that $\mathrm{disc}(f) \geq 1/8\sqrt{\mathrm{rank}(M_f)}$. And this lower bound is sharp when the function f is the inner product function.

3.2 Put everything together

Given Theorem 11, Lemma 12, and $\mathrm{disc}(f) \geq 1/8\sqrt{\mathrm{rank}(M_f)}$, it suffices to prove the following lemma:

Lemma 13 (Main Lemma, [10]). *For every $f : X \times Y \to \{-1, 1\}$, there exists a rectangle R such that*

- $|R| \geq 2^{-O(\log r/\delta)}|X \times Y|$;
- $\mathbb{E}_R[f] \geq 1 - 1/2r$,

where $r = \operatorname{rank}(M_f)$ *and* $\delta = \operatorname{disc}(f)$.

The high-level idea to prove Lemma 13 is to iteratively purify the matrix. More precisely, we can relax Lemma 13 to following lemma.

Lemma 14 ([10]). *Let* $f : X \times Y \to \{-1, 1\}$ *and* $\mathbb{E}[f] = 1 - \beta \geq 0$ *and* $\operatorname{disc}(f) = 3\delta$. *Then there exists a rectangle* R *such that*

- $|R| \geq 2^{-O(2/\delta)}|X \times Y|$;
- $\mathbb{E}_R[f] \geq 1 - \beta/2$,

Notice that we can iteratively use Lemma 14 $O(\log r)$ times, then Lemma 13 follows. To prove Lemma 14, we need the last technical lemma, in whose proof we will use the (definition of) discrepancy.

Lemma 15 ([10]). *Let* $f : X \times Y \to \{-1, 1\}$ *with* $\mathbb{E}[f] = \alpha \geq 0$ *and* $\operatorname{disc}(f) = 3\delta$. *Then there exists a rectangle* R *such that*

$$\mathbb{E}_R[f] \geq \alpha + \delta(1 - \alpha^2)\frac{|X \times Y|}{|R|}.$$

Intuitively, this tells us that we can always find a relatively large subrectangle that is slightly more biased.

Proof of Lemma 15. We will design a specific distribution μ such that we can take advantage of the discrepancy. We define μ as

$$\mu(x, y) = \begin{cases} 1/(1+\alpha)|X \times Y| & f(x, y) = 1; \\ 1/(1-\alpha)|X \times Y| & f(x, y) = -1. \end{cases} \tag{1}$$

Then we have $\sum_{(x,y)} \mu(x, y)f(x, y) = 0$. By the definition of discrepancy, we know that there exists a subrectangle R_1 such that

$$\left| \sum_{(x,y)\in R_1} \mu(x, y)f(x, y) \right| \geq 3\delta.$$

We partition the whole matrix according R_1 to $\{R_1, R_2, R_3, R_4\}$, then we know that there exists a $R \in \{R_1, \cdots, R_4\}$ such that $\sum_{(x,y)\in R} \mu(x, y)f(x, y) \geq \delta$.

By some calculation, we have for this subrectangle R,

$$\mathbb{E}_R[f] \geq \alpha + \delta(1 - \alpha^2)\frac{|X \times Y|}{|R|}.$$

□

The proof of Lemma 14 based on Lemma 15 is a bit technical and doesn't use any property of $\operatorname{rank}(M_f)$ or $\operatorname{disc}(f)$. So we refer the interested readers to the original paper [10].

References

[1] A. Bernasconi and B. Codenotti, *Spectral analysis of boolean functions as a graph eigenvalue problem*, IEEE transactions on computers, 48 (1999), pp. 345–351. 7

[2] L. E. Dickson, *Linear groups: With an exposition of the Galois field theory*, vol. 6, BG Teubner, 1901. 1

[3] D. Gavinsky and S. Lovett, *En route to the log-rank conjecture: New reductions and equivalent formulations*, in International Colloquium on Automata, Languages, and Programming, Springer, 2014, pp. 514–524. 3, 12

[4] M. Göös, T. Pitassi, and T. Watson, *Deterministic communication vs. partition number*, SIAM Journal on Computing, 47 (2018), pp. 2435–2450. 1, 4

[5] B. Green and T. Tao, *The distribution of polynomials over finite fields, with applications to the gowers norms*, Contributions to Discrete Mathematics, 4 (2009). 2.4, 2

[6] A. Knop, S. Lovett, S. McGuire, and W. Yuan, *Guest column: Models of computation between decision trees and communication*, ACM SIGACT News, 52 (2021), pp. 46–70. 1

[7] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, *Complexity measures of sign matrices*, Combinatorica, 27 (2007), pp. 439–463. 3.1

[8] N. Linial and A. Shraibman, *Learning complexity vs communication complexity*, Combinatorics, Probability and Computing, 18 (2009), pp. 227–245. 3.1

[9] L. Lovász and M. Saks, *Lattices, mobius functions and communications complexity*, in [Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science, IEEE Computer Society, 1988, pp. 81–90. 1, 2

[10] S. Lovett, *Communication is bounded by root of rank*, Journal of the ACM (JACM), 63 (2016), pp. 1–9. 1, 3, 13, 14, 15, 3.2

[11] K. Mehlhorn and E. M. Schmidt, *Las vegas is better than determinism in vlsi and distributed computing*, in Proceedings of the fourteenth annual ACM symposium on Theory of computing, 1982, pp. 330–337. 1

[12] N. Nisan and A. Wigderson, *On rank vs. communication complexity*, Combinatorica, 15 (1995), pp. 557–565. 3, 11

[13] R. ODonnell, J. Wright, Y. Zhao, X. Sun, and L.-Y. Tan, *A composition theorem for parity kill number*, in 2014 IEEE 29th Conference on Computational Complexity (CCC), IEEE, 2014, pp. 144–154. 2.5

[14] H. Y. Tsang, C. H. Wong, N. Xie, and S. Zhang, *Fourier sparsity, spectral norm, and the log-rank conjecture*, in 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, IEEE, 2013, pp. 658–667. 1, 5, 1, 9, 1, 10, 2.5