Communication Complexity and Applications

Lecturer: Toniann Pitassi

1 Deterministic Communication and Combinatorial Rectangles

The success in proving good lower bounds on the communication complexity comes from the *combinatorial* view of protocols. The idea is to view protocols as a way to partition the space of all possible input pairs, $X \times Y$, into special sets called combinatorial rectangles.

We can view a deterministic protocol P as a binary tree, where each vertex in the tree is labelled by one of Alice or Bob (the player who owns this vertex). For each internal vertex v in the protocol tree, there is a function Π_v that maps each input $\alpha \in \{0,1\}^n$ to either 0 or 1. Vertex v has two outedges, one edge is labelled by 0 and the other labelled by 1. Each input (x, y) induces a unique path from the root of the protocol tree to a leaf.

For a node v of the protocol tree, we denote by R_v is the set of inputs (x, y) that reach node v. It is easy to see by induction that for every node v, the set R_v is a combinatorial rectangle. Furthermore, if L is the set of leaf vertices of the protocol P, then the set $\{R_l\}_{l \in L}$ gives a *partition* of $X \times Y$ into disjoint combinatorial rectangles. This discussion leads to the following fundamental element in the combinatorics of protocols.

Definition 1 (Rectangle)**.** *A rectangle in $X \times Y$ is a subset $R \subseteq X \times Y$ such that $R = A \times B$ for some $A \subseteq B$ and $B \subseteq Y$.*

The connection between rectangles and protocols is implicit in the following proposition.

Proposition 1. *For all $l \in L$, the set R_l is a rectangle.*

Proof. By induction on the depth of the protocol tree. □

Definition 2 (f-monochromatic)**.** *A subset $R \subseteq X \times Y$ is f-monochromatic if f is fixed*[1] *on R.*

The following two statement are immediate from the above definitions.

Fact 2. *Any protocol P for f induces a partition of $X \times Y$ into f-monochromatic rectangles. The number of (f-monochromatic) rectangles equals the number of leaves of P.*

Fact 3. *If any partition of $X \times Y$ into f-monochromatic rectangles requires at least t rectangles, then $D(f) \geq \log_2 t$.*

[1]There exists $z \in \{0, 1\}$ such that for all $(x, y) \in R$, $f(x, y) = z$.

1.1 Balancing Protocols

As mentioned earlier, we can view a deterministic protocol as a binary tree, and we measured the complexity of the communication protocol by the *height* of the binary tree. Another natural measure of complexity is the *size* of the tree, or the number of leaves. It is clear that if Π is a protocol tree of height d, then the size of Π is at most 2^d. It turns out that the converse is also true. The following lemma shows that deterministic protocols can always be balanced and therefore the minimal size of a protocol is always expponential in the minimum protocol depth.

Lemma 4. *If Π is a protocol for a two-party function F with l leaves, then there is a protocol for F of depth at most $O(\log l)$.*

Proof. The proof is a simple inductive application of the one-third/two-thirds lemma which states that in any binary tree with $l > 1$ leaves, there exists a vertex v such that the subtree rooted at v contains r leaves for $\frac{l}{3} \leq r < \frac{2l}{3}$. Given a protocol Π with l leaves, the players first determine (without communication) a vertex v such that the subtree rooted at v contains between a 1/3 and 2/3 fraction of leaves. Then the players communicate at one bit each in order to determine whether or not their input (x, y) is contained in R_v. If $(x, y) \in R_v$ then the proceed inductively on the subtree rooted at R_v and otherwise they delete the vertex v and its subtree and continue the simulation on the pruned tree. At each step the number of leaves is reduced by a factor of at least 2/3, so the total number if iterations is at most $\log_{3/2} l$ and each iteration requires 2 bits. □

1.2 Partitions versus Protocols

The partition number of a two-player function f is the minimum C such that M_f can be partitioned into C monochromatic subrectangles. We argued above that every protocol of cost c implies a partition of the underlying matrix into 2^c monochromatic subrectangles. Thus the log of the partition number is at most the deterministic communication complexity of f. What about the converse? Can we characterize the communication complexity of f by the partition number of f? A clever argument due to Yannakakis shows that this is true, albeit with a quadratic blowup.

Lemma 5. *Let P be a partition of M_f into $C = 2^c$ monochromatic rectangles. Then there is a deterministic protocol for f of cost $O(c^2)$.*

We will defer the proof of this result until later when we discuss nondeterministic protocols.

For a long while it was open whether or not this quadratic simulation was optimal. We will see in future lectures that it is in fact tight. The lower bound will be one of many applications of the method of lifting in communication complexity.

2 Deterministic Lower Bounds

2.1 The Fooling Set Argument

Consider the following $2^n \times 2^n$ matrix associated with equality function $EQ(x, y)$, $|x| = |y| = n$.

$$M_{EQ} := \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

Each “1” has to be in its own 1-monochromatic rectangle. Thus the number of monochromatic rectangles is greater than 2^n. This observation motivates the following definition of a “fooling set”.

Definition 3. *Let $f : X \times Y \to \{0,1\}$. A subset $S \subseteq X \times Y$ is a fooling set for f if there exists $z \in \{0,1\}$ such that*

(i) $\forall\ (x,y) \in S,\ f(x,y) = z$;

(ii) for any two distinct $(x_1,y_1), (x_2,y_2) \in S$, *either* $f(x_1,y_2) \neq z$ *or* $f(x_2,y_1) \neq z$.

Lemma 6. *If f has a fooling set S of size t, then $D(f) \geq \log_2 t$.*

2.2 The Rank Lower Bound Method

Given any boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ we can associate a $2^n \times 2^n$ matrix M_f, where $M_f(x,y) = f(x,y)$. In words, M_f specifies the values of the function f on any input $(x,y) \in X \times Y$. The rank lower bound method is an algebraic method to give lower bounds on $D(f)$ by computing the rank of M_f.

Definition 4. *For any function f, $\mathsf{rank}(f)$ is the linear rank of M_f over $\mathbb{R}$.*

The following lemma gives a lower bound on the deterministic communication complexity of f through the rank of M_f.

Lemma 7. *Let a function f . Then $\mathsf{P}^{\mathsf{cc}}(f) \geq \log_2 \mathsf{rank}(f)$.*

Proof. Let L_1 be the set of leaves of any protocol tree that gives output 1. For each $l \in L_1$, let M_l be a $2^n \times 2^n$ matrix which is 1 on all $(x,y) \in R_l$ and 0 otherwise. It is clear that

$$M_f = \sum_{l \in L_1} M_l.$$

Fact : The rank function is a sub-additive function, i.e., $\mathsf{rank}(A+B) \leq \mathsf{rank}(A) + \mathsf{rank}(B)$ for any matrix A, B. Therefore,

$$\mathsf{rank}(M_f) \leq \sum_{l \in L_1} \mathsf{rank}(M_l).$$

Notice that $\mathsf{rank}(M_l) = 1$ for any $l \in L_1$ since M_l can be expressed as an outer-product of two vectors[2]. Therefore $\mathsf{rank}(M_f) \leq |L_1| \leq |L|$, which implies that

$$\mathsf{P}^{\mathsf{cc}}(f) \geq \log_2 \mathsf{rank}(M_f).$$

□

[2] These vectors are the characteristic vectors for the rectangle that reaches l.

The above fact shows that communication complexity lower bounds can be proven from rank lower bounds. The Log Rank Conjecture (LRC) asserts that there is a polynomial relationsip between the log of $\mathsf{rank}(M_f)$ and the deterministic commmunication complexity of f. We will discuss the LRC in the next lecture.

3 Other Communication Models and Norms

3.1 Zero-error Communication and α-Discrepancy

Define zero-error model (see gavinsky and lovett paper) show $zero{-}cc \leq infocomplex \leq randomized \leq det$. So zero-error is the weakest of them all. And we know that if low rank implies low zero-error then logrank conjecture is true!