CS 2429 - Foundations of Communication Complexity

Lecturer: Toniann Pitassi

1 Proof Complexity Lower Bounds

Last class we proved communication complexity lower bounds for lifted search problems, from lower bounds for UDISJ. We saw how this implies monotone depth lower bounds.

Today we will show that from communication complexity lower bounds for lifted search problems, we can also obtain proof complexity lower bounds.

Recall that , $Tseitin$ on a graph with n vertices has n constraints, and we will select a gadget g with $c = 2$. Thus, our monotone function will have $4n$ inputs and its monotone depth will be at least the communication complexity of $S(Tseitin(G))og^n$.

Last lecture, we proved that for appropriate graphs G, the communication complexity the lifted Tseitin CNF search problem, $S(Tseitin(G))og^n$, is $\Omega(n/\log n)$.

1.1 Resolution and Cutting Planes

We will prove lower bounds for Resolution proofs as well as Cutting Planes Proofs. Both are refutation systems for showing that a CNF formula is unsatisfiable.

Resolution has one rule, the resolution rule: from the clause $(A \vee x)$ and $(B \vee \neg x)$ we can derive $(A \vee B)$ where A, B are disjunctions of literals. Given a CNF formula $f = C_1 \wedge C_2 \wedge \ldots \wedge C_m$ a resolution refutation of f is a sequence of clauses where each clause in the sequence is either a clause of f or follows from two previous clauses by the resolution rule, and where the final clause is the empty clause. Resolution is known to be sound and complete. The *size* of a resolution proof is the total number of clauses occuring in the proof; the *depth* is the maximum height of the underlying proof dag. A resolution refutation is *tree-like* if the underlying proof dag is a tree (input clauses from f can be repeated any number of times).

Like Resolution, Cutting Planes is a refutation system. However lines in a CP proof are linear inequalities $\sum_i a_i x_i \geq b$, where a_i are rational numbers. There are three CP rules:

(1) (Addition) Two previously derived inequalities can be added together;

(2) (Multiplication) Multiply a previously derived inequality by an integer;

(3) (Division by 2 with rounding) From the inequality $\sum_i a_i x_i \geq b$, we can derive $\sum_i a_i/2 x_i \geq \lceil b/2 \rceil$, as long as each a_i is divisible by 2.

Let f be a CNF formula. Then it can easily be converted into an equivalent set of inequalities. First, for each variable x_i we have the inequalities $x_i \geq 0$ and $x_i \leq 1$. Then we convert each clause into an inequality in the obvious way. For example, $(x_1 \vee \neg x_2)$ becomes $x_1 + (1 - x_2) \geq 1$. It is easy to see that f is satisfiable (over boolean assignments) if and only if the corresponding set of inequalities is satisfiable (over boolean assignments). A CP refutation of f is a sequence of inequalities where each inequality is either one from f, or derived from previous inequalities by one of the CP rules, and where the final inequality is $0 \geq 1$. The size is the total number of inequalities in the proof. (Note: actual size would measure the total length of all coefficients in the proof;

however it turns out that the coefficients without loss of generality have size at most exponential in n, and thus their bit length is $O(n)$ and thus we can just use the number of inequalities as a fairly good measure of the size.) Again we can define the depth of a CP proof, and the notion of a tree-like CP proof.

1.2 Reducing Depth and Tree-Size Lower bounds to Communication Complexity

Theorem 1. *For both Resolution and CPs, if f has a refutation of depth h, then the CNF search problem for f has a (randomized) communication complexity protocol of cost $O(\log n)h$.*

The proof idea is as follows. We will design an efficient communication complexity protocol for solving the CNF search problem for f from a small-depth Resolution/CPs refutation of f. The players will work their way from the root of the proof dag to a leaf of the dag, evaluating lines along the way. The key property is that for any partition of the variables underlying f, every line in a Resolution or CP refutation can be evaluated by the players with an efficient protocol, and thus the overall complexity is the complexity of evaluating a line times the depth. For example, for Resolution, given a clause C, some of the variables underlying C belong to Alice and some to Bob. Alice evaluates the literals in C that she owns, and sends "1" if one of these literals evaluates to true, and otherwise she sends "0"; Bob does the same. If at least one of the players announced "1" then the clause is true under their assignment, and otherwise it is false. They follow a path from the root to a leaf with the property that all clauses along the path evaluate to false. Since the rules are sound, and the root clause is identically false, it follows that for each clause that they encounter (which inductively is set to false by their assignment), one of the two children clauses will be set to false by their assignment. When they arrive at a leaf they have found a clause of f that is falsified. The argument for CPs is similar but is a bit more complicated because of the high weights. (In the case where the weights have length at most $O(\log n)$, the players can just compute the weighted sum of their variables and send the sum to the other player.)

Note that this lower bound method works much more generally for any sound proof system whose lines can be evaluated efficiently by a small 2-player communication protocol. The proof also easily generalizes to the case of sound proof systems (such as Sherali-Adams, Lasserre) where the lines in the proof are efficiently evaluated by a small k-player communication protocol.

1.3 Lower Bounds on depth and tree-like size for Resolution and CPs

The idea is as follows. We start with an unsatisfiable family of CNF formulas, $F_n(z_1, \ldots, z_n)$, whose underlying search problem has high (randomized) communication complexity. We lift F_n to obtain a new CNF formula $F' = F_n o g^n$, where g is our 2-player versatile function defined in the last lecture. (Recall that g takes x, y, $|x| = |y| = 2$ and outputs 1 if and only if the mod 2 sum of the numbers is either 2 or 3, modulo 4.) The lifted CNF, F' will be on variables x, y, $|x| = |y| = 2n$, and will be defined in the obvious way. If F_n is a k-CNF, then F' will be a $2k$-CNF formula. By the way that we define F', it will turn out that the CNF search problem for F' will be the lifted CNF search problem associated with F_n, g.

Let F be the CNF search problem, with clauses $C_1, \ldots, C_m$ over variables $z_1, \ldots, z_n$. Using g, the lifted CNF, F', will be over variables $.x_1, \ldots, x_{2n}$, $y_1, \ldots, y_{2n}$. We will denote the i^{th} variable in the j^{th} block of x by x_i^j. So for example, x_3, the first variable in the second block, will be

denoted by x_1^2. By x^j we mean the j^{th} block of x. (And similarly for y.) By way of example, suppose that $C_1 = (z_1 \vee \neg z_2 \vee z_4)$. Then C_1 will convert to a bunch of new clauses. For each assignment α to x^1, x^2, x^4 and assignment β to y^1, y^2, y^4 such that $g(\overline{\alpha}^1, \overline{\beta}^1) = 0$, and $g(\overline{\alpha}^2, \overline{\beta}^2) = 1$ and $g(\overline{\alpha}^4, \overline{\beta}^4) = 0$, we have the clause

$$(x^{1,\alpha^1} \vee y^{1,\beta^1} \vee x^{2,\alpha^2} \vee y^{2,\beta^2} \vee x^{4,\alpha^4} \vee y^{4,\beta^4}).$$

For example, let $\alpha^1 = \beta^1 = 11$, $\alpha^2 = \beta^2 = 00$, $\alpha^4 = \beta^4 = 11$. Then $g(00, 00) = 0$, and $g(11, 11) = 1$ and $g(00, 00) = 0$. (Because 00 is binary for 0, and $0 + 0(mod 4) = 0$ and similarly, 11 is binary for 3, and $3 + 3(mod 4) = 2$.) we add the following clause:

$$(x_1^1 \vee x_2^2 \vee y_1^1 \vee y_2^1 \vee \neg x_1^2 \vee \neg x_2^2 \vee \neg y_1^2 \vee \neg y_2^2 \vee x_1^4 \vee x_2^4 \vee y_1^4 \vee y_2^4).$$

It is easy to check that the CNF search problem associated with F' is none other than the lifted search problem associated with F, g.

To prove our lower bound, we will start with F_n that has very high (randomized) communication complexity, namely the Tseitin formulas.

Thus, it follows from our lower bound for the lifted Tseitin search problem together with the above theorem, that any Resolution or CP refutation of the lifted Tseitin formula requires depth $\Omega(n/\log n)$.

It can also be proven using a slightly more complicated argument that any tree-like Resolution or CP refutation for the lifted Tseitin formula requires exponential size.