

Lecture 5

Lecturer: Jakob Nordström

Scribe: Lukáš Poláček

1 Quick Recap of Lower Bounds on Length in Terms of Width

Recall that in previous lectures we proved two bounds by Ben-Sasson and Wigderson [BW01] on length in terms of width. For an (unsatisfiable) CNF formula F over n variables, we have the bound

$$L_{\mathcal{R}}(F \vdash \perp) \geq \exp \left(\Omega \left(\frac{(W(F \vdash \perp) - W(F))^2}{n} \right) \right) \quad (1.1)$$

in general resolution, whereas for tree-like resolution we proved the cleaner (and stronger) bound

$$L_{\mathcal{T}}(F \vdash \perp) \geq 2^{W(F \vdash \perp) - W(F)} . \quad (1.2)$$

A natural question is: can we improve (1.1) to an expression similar to (1.2)? The short answer is: no, we cannot. The purpose of today's lecture is to show that the bound (1.1) is in some sense essentially optimal for general resolution. As a byproduct, we will also see an example which shows that general resolution is exponentially stronger than tree-like resolution with respect to length.

2 Formulas with Wide and Short Refutations

Suppose F is a k -CNF formula for some bounded $k = O(1)$. Then what (1.1) says is that if $W(F \vdash \perp) = \omega(\sqrt{n \log n})$, it must hold that $L_{\mathcal{R}}(F \vdash \perp)$ is superpolynomial. Rephrasing our question above, what we are asking is whether a weaker lower bound on width can also provide guarantees for superpolynomial length lower bounds. We will prove that if one weakens the bound by only a $\sqrt{\log n}$ factor, there are no longer any guarantees. Namely, we will find k -CNF formulas that require width on the order of \sqrt{n} but nevertheless have refutations of polynomial length.

The following result was proved first by Bonet and Galesi in a conference paper [BG99] shortly after the result of Ben-Sasson and Wigderson was announced in [BW99]. As usual, we will cite the journal version [BG01] in what follows.

Theorem 2.1 ([BG01]). *There are 3-CNF formulas F_n over n variables with $\text{poly}(n)$ clauses such that $L_{\mathcal{R}}(F_n \vdash \perp) = \text{poly}(n)$ but $W(F_n \vdash \perp) = \Theta(\sqrt{n})$.*

As discussed above, this implies that (1.1) is essentially optimal. By (1.2), a refutation in tree-like resolution for F_n has to have length at least $2^{\Omega(\sqrt{n})}$. Thus, we have the following corollary.

Corollary 2.2. *General resolution is exponentially stronger than tree-like resolution with respect to length.*

There are stronger separations of tree-like and general resolution—the best one is in [BIW04] as far as the lecturer is aware—but the nice thing with Corollary 2.2 is that we get it “for free” from Theorem 2.1.

A word of caution: Above, the parameter n was the number of variables in the formula. This will change in a few seconds, and from then on n will instead be the natural parameter for

a family of formulas $\{F_n\}_{n=1}^\infty$ with $\text{poly}(n)$ variables. This might be a bit confusing, but these are standard conventions in the literature. Thus, rather than trying to shield course participants from the harsh realities of life by changing these conventions, we instead add this caveat to help deal with these realities head on.

Nevertheless, as an extra service to the reader we will also try to denote the number of variables by N below for increased clarity.

In the rest of the lecture, we will prove Theorem 2.1. We will use formulas which encode *ordering principles*. Suppose we have a finite set $S_n = \{e_1, \dots, e_n\}$ of partially (or totally) ordered elements, then S_n must have a minimal element. Note that this statement is not true for infinite sets, e.g. $\{1/n \mid n \in \mathbb{N}^+\}$.

Since we want unsatisfiable formulas, we will use CNF formulas saying that S_n is ordered but that despite of this there is no minimal element in the set. Below, we should interpret the variable x_{ij} to mean that $e_i < e_j$. We will use the following 4 types of clauses, where indices i, j, k range from 1 to n .

$$A(i, j, k) = \bar{x}_{ij} \vee \bar{x}_{jk} \vee x_{ik} \quad \text{for all } i \neq j \neq k \neq i \quad (\text{transitivity}) \quad (2.1a)$$

$$B(i, j) = \bar{x}_{ij} \vee \bar{x}_{ji} \quad \text{for all } i \neq j \quad (\text{anti-symmetry}) \quad (2.1b)$$

$$C_n(j) = \bigvee_{1 \leq i \leq n, i \neq j} x_{ij} \quad \text{for all } j \quad (\text{non-minimality}) \quad (2.1c)$$

$$D(i, j) = x_{ij} \vee x_{ji} \quad \text{for all } i \neq j \quad (\text{totality}) \quad (2.1d)$$

Clauses of type (2.1a), (2.1b) and (2.1c) form the *partial ordering principle formulas* and we denote the formula for S_n by POP_n . By adding clauses of type (2.1d), we get *linear ordering principle formulas* and we denote these formulas by LOP_n . We remark that these formulas go under a number of different names in the literature, but we will stick to POP_n and LOP_n in this course. It is easy to check that both types of formulas have $\Theta(n^2)$ variables and $\Theta(n^3)$ clauses.

3 An Upper Bound on Refutation Length of Ordering Principles

The ordering principle formulas were conjectured to be hard to refute in resolution by Krishnamurthy [Kri85] (note that this is close in time to the first exponential lower bounds for resolution by Haken [Hak85]), but were instead proven to be easy by Stålmarck [Stå96] a decade later. We will follow an adaptation of Stålmarck's resolution refutation by Bonet and Galesi [BG01].

Note that $POP_n \subseteq LOP_n$, so a refutation for POP_n is also a refutation for LOP_n and hence it holds that $L(LOP_n \vdash \perp) \leq L(POP_n \vdash \perp)$. Our goal is first to prove an upper bound on length for POP_n (which will automatically hold for LOP_n as well) and then in the next section a lower bound on width for LOP_n (which will also automatically hold for POP_n).

Theorem 3.1 ([Stå96]). *There exist resolution refutations of POP_n in length $\text{poly}(n)$.*

Proof. For $n = 1$, we have $POP_1 = (\bar{x}_{12} \vee \bar{x}_{21}) \wedge x_{12} \wedge x_{21}$, and this formula can clearly be refuted in a (small) constant number of steps by a resolution refutation π_2 . For bigger n our strategy is to derive POP_{n-1} from POP_n in polynomial length. If we have such resolution derivations $\pi_n : POP_n \vdash POP_{n-1}$, we can then string all these derivations $\pi_n, \pi_{n-1}, \dots, \pi_3, \pi_2$ together to get a refutation of POP_n .

Note that clauses of type (2.1a) and (2.1b) from POP_{n-1} are all present in POP_n , so we only need to show a way to derive clauses of type (2.1c). Namely, we will derive clauses $C_{n-1}(1), \dots, C_{n-1}(n-1)$ from $A(i, j, k), B(i, j)$ and $C_n(j)$ (for all needed i, j, k). The intuition behind this derivation is that we can extract a smaller set S_{n-1} from S_n which is also ordered and does not contain a minimal element by showing that since e_n is not minimal, some element in $\{e_1, \dots, e_{n-1}\}$ must be.

More formally, we claim that any clause $C_{n-1}(i)$ can be derived in polynomial length, and state this formally as Lemma 3.2 below. Assuming this lemma, which we will prove shortly, we can apply it for all $i = 1, 2, \dots, n-1$ to obtain all clauses of POP_{n-1} from those of POP_n , and Theorem 3.1 follows by induction. \square

Lemma 3.2. *For any $j \leq n-1$, the clause $C_{n-1}(j)$ is derivable in polynomial length from $C_n(1), \dots, C_n(n)$, $A(1, n, j), \dots, A(n-1, n, j)$, and $B(j, n)$.*

Proof. For any $i \neq j$, we can do the inference step

$$\frac{C_n(j) \quad A(i, n, j)}{C_{n-1}(j) \vee \bar{x}_{in}} \quad (3.1)$$

by resolving over x_{nj} . We can interpret this step as follows. Suppose e_j is not minimal in S_n and none of the elements $\{e_1, \dots, e_{j-1}, e_{j+1}, \dots, e_{n-1}\}$ is smaller than e_j (this corresponds to $C_{n-1}(j)$ being false), then e_n has to be smaller than e_j . This implies that $e_i \not\prec e_n$, since if on the contrary $e_i < e_n$, then we would also have $e_i < e_j$ by transitivity contrary to assumption. Thus, x_{in} must be false in this case.

To derive $C_{n-1}(j) \vee \bar{x}_{jn}$, we instead resolve

$$\frac{C_n(j) \quad B(j, n)}{C_{n-1}(j) \vee \bar{x}_{jn}} \quad (3.2)$$

over x_{nj} . In this way, we can derive $C_{n-j} \vee \bar{x}_{in}$ for all i .

Using these clauses one by one and resolving over the variables $x_{1n}, x_{2n}, \dots, x_{n-1,n}$ in that order, we get the resolution derivation

$$\frac{\frac{C_n(n) \quad C_{n-1}(j) \vee \bar{x}_{1,n}}{C_{n-1}(j) \vee \bigvee_{\ell=2}^{n-1} x_{\ell n}} \quad C_{n-1}(j) \vee \bar{x}_{2n}}{C_{n-1}(j) \vee \bigvee_{\ell=3}^{n-1} x_{\ell n}} \quad C_{n-1}(j) \vee \bar{x}_{3n}}{C_{n-1}(j) \vee \bigvee_{\ell=4}^{n-1} x_{\ell n}} \quad (3.3)$$

$$\vdots$$

$$\frac{C_{n-1}(j) \vee x_{n-1,n} \quad C_{n-1}(j) \vee \bar{x}_{n-1,n}}{C_{n-1}(j)}$$

which also clearly has polynomial length, and which derives $C_{n-1}(j)$. This concludes the proof of Lemma 3.2. \square

Note that all the clauses appearing in the resolution refutation have width $O(n)$, which is $O(\sqrt{N})$ if we let N denote the number of variables in POP_n . But what exactly is the length of the resolution refutation, and what is the clause space required to carry out the refutation in this length? These questions are left as an exercise for the reader.

4 3-CNF Versions of Wide CNF Formulas

So now we are done with the first half of our program for today, and if we can also prove a lower bound on width for LOP_n we will be done. Except...

Except that if we look closer at the definitions of POP_n and LOP_n , these formulas have clauses of width n , but Theorem 2.1 is stated for 3-CNF formulas. Intuitively, if the width of the clauses themselves is already n , then we would have to work very hard indeed to prove a lower bound on $W(LOP_n \vdash \perp)$ that could give anything interesting in (1.1).

So what can we do? One approach would be to proceed as for the pigeonhole principle formulas, where we defined “sparser versions” with constant-width clauses and showed that

these formulas were also hard. That is in fact something one can do for ordering principles as well,¹ although one has to be a bit more careful, but that is not what we are going to do today.

Instead, we will use a standard transformation of formulas of arbitrary width into equivalent 3-CNF formulas, and prove lower bounds for these formulas. To describe this transformation, let $F = C_1 \wedge \dots \wedge C_m$ be a CNF formula. For each C_j with $W(C_j) \leq 3$ we let $\tilde{C}_j = C_j$, and for each C_j with $W(C_j) > 3$, say $C_j = a_1 \vee a_2 \vee \dots \vee a_w$, we let \tilde{C}_j be the set of clauses

$$\tilde{C}_j = \begin{cases} \bar{y}_{0j} \\ y_{0j} \vee a_1 \vee \bar{y}_{1j} \\ y_{1j} \vee a_2 \vee \bar{y}_{2j} \\ \vdots \\ y_{w-1,j} \vee a_w \vee \bar{y}_{wj} \\ y_{wj} \end{cases} \quad (4.1)$$

where y_{ij} are new variables that are unique to \tilde{C}_j and do not appear anywhere else. With this notation, we define $\tilde{F} = \tilde{C}_1 \wedge \dots \wedge \tilde{C}_m$ to be the *3-CNF version* or *extended version* of F . We leave the verification of the following fact to the reader.

Proposition 4.1. *\tilde{F} is unsatisfiable if and only if F is unsatisfiable.*

As was pointed out in class, since we want to get the number of variables in our formulas right in order to prove that (1.1) is tight, we should check what happens to the ordering principle formulas.

Proposition 4.2. *The 3-CNF formulas \widetilde{POP}_n and \widetilde{LOP}_n have $\Theta(n^2)$ variables and $\Theta(n^3)$ clauses.*

Again we leave the verification to the reader, but it is a good idea to check that this really holds and understand why. (Is it always true, for instance, that the number of variables and clauses in F will be asymptotically the same as in \tilde{F} ?)

The refutation length for \tilde{F} will never be much larger than that of F and the refutation width will not increase at all, as we can see in the following proposition.

Proposition 4.3. *If F is a CNF formula with m clauses over n variables, then $L_{\mathcal{R}}(\tilde{F} \vdash \perp) \leq L_{\mathcal{R}}(F \vdash \perp) + O(nm)$ and $W(\tilde{F} \vdash \perp) \leq W(F \vdash \perp)$.*

Proof. We start by deriving C_i from \tilde{C}_i for all i (or, if we want to be nitpick, for all axiom clauses C_i that we will need to use in a short or narrow refutation of F). This can be done for each C_i in at most n resolution steps over auxiliary variables, giving $O(mn)$ steps in total. Now we have the original clauses of F , so we can just copy any resolution refutation of F to finish the refutation of \tilde{F} . \square

Thus, \widetilde{POP}_n is a 3-CNF formula that can be refuted in polynomial length and width $O(\sqrt{N})$ where N is the number of variables in \widetilde{POP}_n .

5 A Lower Bound on Refutation Width of Ordering Principles

Good, so after this short detour we are now back on track again, and we can conclude today's lecture by proving that the width of any resolution refutation of \widetilde{LOP}_n is $\Omega(n)$. Note that this is all we need to complete the proofs of Theorem 2.1 and Corollary 2.2.

To prove the lower bound on width, we will focus on a particular type of truth value assignments, namely assignments that satisfy every clause in LOP_n except $C_n(j)$ for some j . Such assignments define a total (a.k.a. linear) order on S_n with unique minimal element j .

¹See Lemma 8.17 in [SBI04] for the details.

Definition 5.1. A *critical assignment* $\alpha : \text{Vars}(LOP_n) \rightarrow \{0, 1\}$ for LOP_n is an assignment defining a total order, where $\alpha(x_{ij}) = 1$ if and only if $e_i < e_j$. The assignment α is j -critical if e_j is the unique minimal element in the ordering defined by α .

For a j -critical assignment α , $\alpha(LOP_n \setminus \{C_n(j)\}) = 1$ and $\alpha(C_n(j)) = 0$, since e_j is a minimal element. Let us extend this notion to the formulas \widetilde{LOP}_n and assignments β to $\text{Vars}(\widetilde{LOP}_n)$.

Definition 5.2. A j -critical assignment $\beta : \text{Vars}(\widetilde{LOP}_n) \rightarrow \{0, 1\}$ for \widetilde{LOP}_n is such that β restricted to variables $\{x_{ij}\}$ is j -critical for LOP_n , $\beta(\widetilde{LOP}_n \setminus \widetilde{C}_n(j)) = 1$ and $\beta(\widetilde{C}_n(j)) = 0$.

Proposition 5.3. Any j -critical assignment α for LOP_n can be extended to j -critical assignment β for \widetilde{LOP}_n .

Proof. Let $C_n(i)$ be a clause in LOP_n where $i \neq j$ and let $x_{ik} \in C_n(j)$ be one of the variables with $\alpha(x_{ik}) = 1$. Such a variable must exist, since $\alpha(C_n(i)) = 1$. We set $\beta(y_{lj}) = 0$ if $l < k$ and $\beta(y_{lj}) = 1$ otherwise. Notice that $\beta(\widetilde{C}_n(j))$ has to be false, since $\alpha(C_n(j)) = 0$ and so it is impossible to satisfy all the clauses in (4.1) regardless of how the auxiliary variables are set. \square

We next define the set of variables V_j that intuitively speaking are all the variables that can give us any information about the element e_j in S_n . Formally, we let $V_j = \text{Vars}(\widetilde{C}_n(j)) \cup \{x_{ji} \mid i \neq j\}$. We have the following simple but crucial proposition.

Proposition 5.4. If D is a clause with $W(D) = w$, then at most $2w$ sets V_j have variables appearing in D .

Proof. A variable x_{ij} appears only in V_j and V_i , and an auxiliary variable y_{ij} appears only in the set V_j . Thus, any variable in D can cover at most two sets. \square

We are now ready to prove our width lower bound. We will follow the same proof structure that we used successfully in a couple of previous lectures. Namely, we define a measure μ on the clauses which is small for axioms and big for the empty clause \perp , and then use this μ to measure the “progress” made in a resolution refutation. We prove that somewhere in any refutation there must be a clause that has made “medium-progress,” and that any such clause must contain many literals. This shows that any resolution refutation must be wide.

For $I \subseteq \{1, \dots, n\}$, let $\widetilde{C}_I = \bigwedge_{i \in I} \widetilde{C}_n(i)$. For a clause D , let $I_D \subseteq \{1, \dots, n\}$ be a set of minimal size such that all critical assignments to \widetilde{LOP}_n that satisfy \widetilde{C}_{I_D} also must satisfy D . Then the μ -measure of this clause is defined as $\mu(D) = |I_D|$ (and note that although I_D is not necessarily uniquely defined, the *size* of any minimal such set is uniquely defined).

Lemma 5.5. The following holds for the measure μ as defined above:

1. $\mu(D) \leq n$ for any clause D .
2. $\mu(D) \leq 1$ if D is an axiom.
3. $\mu(\perp) = n$.
4. $\mu(D \vee D') \leq \mu(D \vee x) + \mu(D \vee \bar{x})$, for clauses $D \vee x$ and $D \vee \bar{x}$.

Proof. For item 1, let $I'_D = \{1, \dots, n\}$. We have that $\widetilde{C}_{I'_D}$ is unsatisfiable for all critical assignments, so all critical assignments α for which $\alpha(\widetilde{C}_{I'_D}) = 1$ satisfy D as well for the vacuous reason that there are no such assignments. Hence $\mu(D) \leq n$.

For item 2, if D is an axiom in $\widetilde{C}_n(j)$, we can take $I_D = \{j\}$. Otherwise $I_D = \emptyset$ is fine, since we are only considering assignments defining total orders and such assignments clearly satisfy anti-symmetry and transitivity.

To prove the third item, note that there are assignments satisfying $\widetilde{C}_{I'}$ for any I' with at most $n - 1$ elements but that no assignment can satisfy the empty clause, so $\mu(\perp) > n - 1$. The inequality $\mu(\perp) \leq n$ then implies $\mu(\perp) = n$.

To see that the claim in item 4 holds, consider $I' = I_{D \vee x} \cup I_{D' \vee \bar{x}}$. Any assignment satisfying $\widetilde{C}_{I'}$ has to satisfy both $D \vee x$ and $D' \vee \bar{x}$ by assumption, which means it also has to satisfy $D \vee D'$. Hence, $|I'| \leq |I_{D \vee x}| + |I_{D' \vee \bar{x}}|$ is an upper bound on $\mu(D \vee D')$ and we conclude that $\mu(D \vee D') \leq \mu(D \vee x) + \mu(D \vee \bar{x})$. \square

Using the previous lemma we can finally prove the last missing piece in the proof of the main theorem in this lecture.

Lemma 5.6. *Every refutation π of \widetilde{LOP}_n contains a clause D such that $W(D) \geq n/6$.*

Proof. We will consider a clause D with “medium” measure $\mu(D)$ and assume $W(D) < n/6$ to get a contradiction. We will find two elements e_k and e_ℓ not mentioned in D , such that $k \in I_D$ and $\ell \notin I_D$. Their existence will follow by a simple counting argument from the fact that D is small. Then we will consider any k -critical assignment β falsifying D (such an assignment must always exist by assumption). Under this assignment, e_k is a minimal element, so $\beta(\widetilde{C}_{I_D}) = 0$. We will construct a new assignment β' where e_ℓ will be the new minimal element but otherwise it will be the same assignment as β . This will make \widetilde{C}_{I_D} satisfied, since the minimal element e_ℓ is outside I_D . However, D will remain falsified, since it does not contain any information about e_k or e_ℓ . Hence the new assignment β' will satisfy \widetilde{C}_{I_D} but not D , which contradicts the definition of I_D .

We will now formalize the previous informal argument. Fix I_D and the corresponding clause $D \in \pi$ such that $n/3 \leq |I_D| \leq 2n/3$. By part 4 of Lemma 5.5, such a clause D must exist, since otherwise it would be impossible to reach the progress level $\mu(\perp) = n$. Now we are going to assume $W(D) < n/6$ in order to derive a contradiction.

By Proposition 5.4 and the assumption that $2W(D) < |I_D|$, we conclude by counting that there must exist a $k \in I_D$ such that $V_k \cap \text{Vars}(D) = \emptyset$. Next, consider $\overline{I_D} = \{1, \dots, n\} \setminus I_D$. Since $|\overline{I_D}| \geq n/3$, by the same argument we get that there is an $\ell \in \overline{I_D}$ such that $V_\ell \cap \text{Vars}(D) = \emptyset$. These are the two elements $e_k \in I_D$ and $e_\ell \notin I_D$ not mentioned in D that we described in the informal proof overview above.

Let β be a k -critical assignment such that $\beta(D) = 0$. Such an assignment must exist since I_D was chosen of minimal size and $k \in I_D$, which means that there exists a critical assignment that satisfies $\widetilde{C}_{I_D \setminus \{k\}}$ but falsifies D . Note that $\beta(\widetilde{C}_{I_D}) = 0$ since β is k -critical.

We will now change the order given by β a little, namely by making e_ℓ the minimal element but leaving the internal order of all other elements unchanged. This will define a new ℓ -critical assignment β' . Since ℓ is not in I_D , for every $j \in I_D$ there exists an element which is smaller than e_j , so $\beta'(\widetilde{C}_{I_D}) = 1$.

We first make e_ℓ the new minimal element by setting $\beta'(x_{l_j}) = 1$ and $\beta'(x_{j_l}) = 0$ for all j . For all $j \notin \{l, k\}$, the only change from β is that x_{l_j} is potentially flipped to true. However, since $\widetilde{C}_n(j)$ was already satisfied we can keep the values assigned to the auxiliary variables in these clauses. It remains to examine $\widetilde{C}_n(k)$. We have $\beta'(x_{lk}) = 1$, so we can set the auxiliary variables in $\widetilde{C}_n(k)$ so that it is true under β' .

The assignment β' only differs from β in variables $x_{\ell j}, x_{j \ell}$ and auxiliary variables in $\widetilde{C}_n(k)$. All of these are not in D , so D remains falsified. Hence β' is a critical assignment satisfying \widetilde{C}_{I_D} but falsifying D , which is a contradiction. This concludes the proof of the lemma. \square

Now the proofs of Theorem 2.1 and Corollary 2.2 are complete, so this ends the lecture.

References

- [BG99] Maria Luisa Bonet and Nicola Galesi. A study of proof search algorithms for resolution and polynomial calculus. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS '99)*, pages 422–431, October 1999.
- [BG01] Maria Luisa Bonet and Nicola Galesi. Optimality of size-width tradeoffs for resolution. *Computational Complexity*, 10(4):261–276, December 2001. Preliminary version appeared in *FOCS '99*.
- [BIW04] Eli Ben-Sasson, Russell Impagliazzo, and Avi Wigderson. Near optimal separation of treelike and general resolution. *Combinatorica*, 24(4):585–603, September 2004.
- [BW99] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC '99)*, pages 517–526, May 1999.
- [BW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *Journal of the ACM*, 48(2):149–169, March 2001. Preliminary version appeared in *STOC '99*.
- [Hak85] Armin Haken. The intractability of resolution. *Theoretical Computer Science*, 39(2-3):297–308, August 1985.
- [Kri85] Balakrishnan Krishnamurthy. Short proofs for tricky formulas. *Acta Informatica*, 22(3):253–275, August 1985.
- [SBI04] Nathan Segerlind, Samuel R. Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k -DNF resolution. *SIAM Journal on Computing*, 33(5):1171–1200, 2004. Preliminary version appeared in *FOCS '02*.
- [Stå96] Gunnar Stålmarmark. Short resolution proofs for a sequence of tricky formulas. *Acta Informatica*, 33(3):277–280, May 1996.