Resolution proof lower bounds for random k-SAT

T-79.7001
Postgraduate course in Theoretical Computer Science

Siert Wieringa

22.10.2007
Subject

- Proving the theoretical complexity of random k-SAT formulas for resolution

- *Simplified and Improved Resolution Lower Bounds* by Paul Beame and Toniann Pitassi
Table of contents

Proof idea
Definitions
Lemma’s
Results
Conclusion
Proof idea

• First choose a restriction that removes all large clauses

• Argue that the restricted formula is random enough to require any proof it to contain long clauses

• Contradiction!
Sparsity (1)

Definition (n′ – sparsity)
A formula $F$ is $n′$ – sparse if every set of $s \leq n'$ variables contains at most $s$ clauses of $F$. 

Consider the following unsatisfiable set of four clauses:
- {$1, 2$}
- {$1, -2$}
- {$-1, 3$}
- {$-1, -3$}
This formula is $2$ – sparse as for every possible set of two variables from this formula there are at most two clauses that contain all variables in that set.
Sparsity (1)

Definition ($n' - sparsity$)
A formula $\mathcal{F}$ is $n' - sparse$ if every set of $s \leq n'$ variables contains at most $s$ clauses of $\mathcal{F}$.

Excuse me?
Consider the following unsatisfiable set of four clauses:

- $\{ 1, 2 \}$
- $\{ 1, -2 \}$
- $\{ -1, 3 \}$
- $\{ -1, -3 \}$

This formula is $2 - sparse$ as for every possible set of two variables from this formula there are at most two clauses that contain all variables in that set.
Sparsity (2)

Definition $((n', n'', y) - sparsity)$
A formula $\mathcal{F}$ is $(n', n'', y) - sparse$ if every set of $s$ variables, $n' < s \leq n''$, contains at most $ys$ clauses.
Boundary set

Definition (Boundary set)
The boundary set of a set $S$ is the set of variables that appear in only one clause of $S$. 
Lemma (5.4.11)

*If a CNF formula $\mathcal{F}$ is $n'$ – sparse then every subset of up to $n'$ of its clauses is satisfiable.*
Satisfiable subsets

Lemma (5.4.11)
If a CNF formula $\mathcal{F}$ is $n'$ – sparse then every subset of up to $n'$ of its clauses is satisfiable.

Proof.
Every subset $S$ of the $n'$ – sparse formula $\mathcal{F}$ with $|S| \leq n'$ contains at least $|S|$ distinct variables and it is therefore satisfiable.
## Size of boundary set

**Lemma (5.4.12)**

Let $F$ be a CNF formula with clause size at most $k$ and suppose $F$ is:

$$\left(n', \frac{k + \epsilon}{2}, n'', \frac{k + \epsilon}{2}, \frac{2}{k + \epsilon}\right) - \text{sparse}. $$

Then every set $S$ of size $l$ clauses of $F$, with $n' < l \leq n''$ has a boundary size of at least $\epsilon l$. 
Size of boundary set

Proof.
Suppose $S$ has boundary of size less then $\epsilon / l$. There are at most $k/l$ variable occurrences in $S$. So, the maximum number of different variables occurring in $S$ must be less than:

Since each boundary variable occurs once and every one of the remaining variables occurs at least twice. This contradicts with the assumption that $F$ is $(n' k + \epsilon^2, n'' k + \epsilon^2, 2 k + \epsilon)$ sparse.
Proof.
Suppose $S$ has boundary of size less than $\epsilon l$. There are at most $kl$ variable occurrences in $S$. So, the maximum number of different variables occurring in $S$ must be less than:

$$\epsilon l + \frac{kl - \epsilon l}{2} \leq \frac{kl}{2} + \frac{\epsilon l}{2} \leq l\left(\frac{k + \epsilon}{2}\right) \leq n'' \frac{k + \epsilon}{2}$$

Since each boundary variable occurs once
Size of boundary set

Proof.
Suppose $S$ has boundary of size less than $\epsilon l$. There are at most $kl$ variable occurrences in $S$. So, the maximum number of different variables occurring in $S$ must be less than:

$$\epsilon l + \frac{kl - \epsilon l}{2} \leq \frac{kl}{2} + \frac{\epsilon l}{2} \leq l\frac{k + \epsilon}{2} \leq k'' \frac{k + \epsilon}{2}$$

Since each boundary variable occurs once and every one of the remaining variables occurs at least twice.
Size of boundary set

Proof.
Suppose $S$ has boundary of size less then $\epsilon l$. There are at most $kl$ variable occurrences in $S$. So, the maximum number of different variables occurring in $S$ must be less than:

$$\epsilon l + \frac{kl - \epsilon l}{2} \leq \frac{kl}{2} + \frac{\epsilon l}{2} \leq l \frac{k + \epsilon}{2} \leq n'' \frac{k + \epsilon}{2}$$

Since each boundary variable occurs once and every one of the remaining variables occurs at least twice. This contradicts with the assumption that $\mathcal{F}$ is:

$$(n' \frac{k + \epsilon}{2}, n'' \frac{k + \epsilon}{2}, \frac{2}{k + \epsilon}) - \text{sparse}.$$
Size of boundary set

Excusé-moi?

Why does the maximum number of different variables occurring in $S$ must be less than $l \frac{k + \epsilon}{2}$ contradict with the assumption that $\mathcal{F}$ is:

$$(n' \frac{k + \epsilon}{2}, n'' \frac{k + \epsilon}{2}, \frac{2}{k + \epsilon}) – \text{sparse} \ ???$$

Note

Analysing this proof with the right hand side of the expression $l \frac{k + \epsilon}{2} \leq n'' \frac{k + \epsilon}{2}$ leads to an incomplete result, I therefore continue with the left hand side expression.
Size of boundary set

Excusé-moi?
Why does the maximum number of different variables occurring in $S$ must be less than $l \frac{k + \epsilon}{2}$ contradict with the assumption that $F$ is:

$$(n' \frac{k + \epsilon}{2}, n'' \frac{k + \epsilon}{2}, \frac{2}{k + \epsilon}) - \text{sparse}$$

$$z = \frac{k + \epsilon}{2}$$
Size of boundary set

Excusé-moi?

*Why does the maximum number of different variables occurring in $S$ must be less than $lz$ contradict with the assumption that:*

$$z = \frac{k + \epsilon}{2} \text{ and } F \text{ is } (n'z, n''z, \frac{1}{z}) - \text{sparse} ??$$
Size of boundary set

Excusé-moi?

*Why does the maximum number of different variables occurring in S must be less than $lz$ contradict with the assumption that:*

$$z = \frac{k + \epsilon}{2} \text{ and } \mathcal{F} \text{ is } (n'z, n''z, \frac{1}{z}) - \text{sparse}???$$

Definition $((n', n'', y) - \text{sparsity})$

A formula $\mathcal{F}$ is $(n', n'', y) - \text{sparse}$ if every set of $s$ variables, $n' < s \leq n''$, contains at most $ys$ clauses.
Size of boundary set

Excusé-moi?
Why does the maximum number of different variables occurring in S must be less than l\(z\) contradict with the assumption that:

\[ z = \frac{k + \epsilon}{2} \text{ and } F \text{ is } (n'z, n''z, \frac{1}{Z}) - \text{sparse} \]

Definition \(((n'z, n''z, \frac{1}{Z}) - \text{sparseity})\)
A formula \(F\) is \((n'z, n''z, \frac{1}{Z}) - \text{sparse}\) if every set of \(s\) variables, \(n'z < s \leq n''z\), contains at most \(\frac{s}{Z}\) clauses.
Size of boundary set

Excusé-moi?
S should contain less then \(lz\) variables. This means that it must contain less then \(\frac{lz}{z} = l\) clauses. However, \(S\) is of size \(l\) which is a contradiction.

Definition ((\(n'z, n''z, \frac{1}{z}\) – sparsity))
A formula \(F\) is \((n'z, n''z, \frac{1}{z}) – sparse\) if every set of \(s\) variables, \(n'z < s \leq n''z\), contains at most \(\frac{s}{z}\) clauses.
Lemma (5.4.13)

Let $n' \leq n$ and $\mathcal{F}$ be an unsatisfiable $k$–CNF formula with $n$ variables. If $\mathcal{F}$ is $n'$–sparse and:

$$\left( n' \frac{k + \epsilon}{4}, n' \frac{k + \epsilon}{2}, \frac{2}{k + \epsilon} \right)$$

then every resolution refutation of $\mathcal{F}$ must include a clause of length at least $\frac{en'}{2}$.
Complex clause lemma

Definition (Clause complexity)
The complexity of a clause $C$ is the smallest number of clauses whose conjunction implies $C$.

Start of proof.

- *The complexity of the empty clause must be* $> n'$.
Complex clause lemma

Definition (Clause complexity)
The complexity of a clause $C$ is the smallest number of clauses whose conjunction implies $C$.

Start of proof.

- The complexity of the empty clause must be $> n'$.
- Since the complexity of the resolvent is at most the sum of the complexities of the clauses from which it is derived there must exist a clause $C$ in the proof whose complexity is bigger then $\frac{n'}{2}$ and at most $n'$. 
Complex clause lemma

Continued proof.

• Let $S$ be a set of clauses witnessing the complexity of $C$ with $\frac{n'}{2} < |S| \leq n'$.
Complex clause lemma

Continued proof.

• Let $S$ be a set of clauses witnessing the complexity of $C$ with $\frac{n'}{2} < |S| \leq n'$.
• The boundary set $b(S)$ is at least of size $\epsilon |S| > \epsilon \frac{n'}{2}$.
Complex clause lemma

Continued proof.

- Let $S$ be a set of clauses witnessing the complexity of $C$ with $\frac{n'}{2} < |S| \leq n'$.
- The boundary set $b(S)$ is at least of size $\epsilon|S| > \epsilon \frac{n'}{2}$.
- $S$ implies $C$, and $S - \{C'\}$ does not imply $C$. 
Complex clause lemma

Continued proof.

• Let $S$ be a set of clauses witnessing the complexity of $C$ with $\frac{n'}{2} < |S| \leq n'$.
• The boundary set $b(S)$ is at least of size $\epsilon |S| > \epsilon \frac{n'}{2}$.
• $S$ implies $C$, and $S - \{C'\}$ does not imply $C$.
• $C$ must contain all variables in $b(S)$ and is therefore of length $> \epsilon \frac{n'}{2}$
Restriction effect

Lemma (5.4.14)

Let $P$ be a resolution refutation of $F$. The large clauses of $P$ are those clauses mentioning more then $\alpha n$ distinct variables. With probability greater then $1 - 2^{(1 - \frac{\alpha t}{4})|P|}$, a random restriction of size $t$ sets all large clauses to 1.
Restriction effect

Start of proof.

- Let $C$ be a large clause of $P$
Restriction effect

Start of proof.

- Let $C$ be a large clause of $P$
- Expected number of variables assigned a value by random restriction of size $t$ is $\alpha n \frac{t}{n} = \alpha t$
Restriction effect

Start of proof.

- **Let C be a large clause of P**
- **Expected number of variables assigned a value by random restriction of size t is** $\alpha n \frac{t}{n} = \alpha t$
- $Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}}$
Restriction effect

Start of proof.

- Let $C$ be a large clause of $P$
- Expected number of variables assigned a value by random restriction of size $t$ is $\alpha n \frac{t}{n} = \alpha t$
- $\Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}}$

Anteeski?

The probability that the number of variables in a clause is less then or equal to a quarter of the expected number. This includes the case where $|C \cap D| = \emptyset$. 
Restriction effect

Continued proof.

- $\Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}}$
Restriction effect

Continued proof.

- \( \Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}} \)
- Given that \( |C \cap D| = s \) the probability that \( C_p \) is not satisfied is \( 2^{-s} \)
Continued proof.

- \( \Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}} \)
- Given that \( |C \cap D| = s \) the probability that \( C \subseteq_p \) is not satisfied is \( 2^{-s} \)
- The probability that \( |C \cap D| > \frac{\alpha t}{4} \) and \( C \) is not satisfied is at most \( 2^{-\frac{\alpha t}{4}} \)
Restriction effect

Continued proof.

- \( \Pr[|C \cap D| \leq \frac{\alpha t}{4}] \leq 2^{-\frac{\alpha t}{2}} \)
- Given that \(|C \cap D| = s\) the probability that \(C_p\) is not satisfied is \(2^{-s}\)
- The probability that \(|C \cap D| > \frac{\alpha t}{4}\) and \(C\) is not satisfied is at most \(2^{-\frac{\alpha t}{4}}\)
- The probability that \(C\) is not satisfied is at most:
  \[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{1 - \frac{\alpha t}{4}} \]
Restriction effect

Entschuldigen Sie bitte!

\[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{(1-\frac{\alpha t}{4})} \]
Restriction effect

Entschuldigen Sie bitte!

\[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{\left(1-\frac{\alpha t}{4}\right)} \]

\[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{-\frac{\alpha t}{4}} + 2^{-\frac{\alpha t}{4}} \]
Restriction effect

Entschuldigen Sie bitte!

\[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{(1-\frac{\alpha t}{4})} \]

\[ 2^{-\frac{\alpha t}{2}} + 2^{-\frac{\alpha t}{4}} < 2^{-\frac{\alpha t}{4}} + 2^{-\frac{\alpha t}{4}} \]

\[ 2^{-\frac{\alpha t}{4}} + 2^{-\frac{\alpha t}{4}} = 2^{1} \times 2^{-\frac{\alpha t}{4}} = 2^{(1-\frac{\alpha t}{4})} \]
Restriction effect

Lemma (5.4.14)

Let $P$ be a resolution refutation of $F$. The large clauses of $P$ are those clauses mentioning more then $\alpha n$ distinct variables. With probability greater then $1 - 2^{(1 - \frac{\alpha t}{4})|P|}$, a random restriction of size $t$ sets all large clauses to 1.
Restriction effect

Lemma (5.4.14)

Let $P$ be a resolution refutation of $\mathcal{F}$. The large clauses of $P$ are those clauses mentioning more than $\alpha n$ distinct variables. With probability greater than $1 - 2^{(1 - \frac{\alpha t}{4}) |P|}$, a random restriction of size $t$ sets all large clauses to 1.

Proof.

- The probability that a clause $C$ in $P$ is not satisfied is at most $2^{(1 - \frac{\alpha t}{4})}$
- The probability that a clause is satisfied is therefore at least $1 - 2^{(1 - \frac{\alpha t}{4})}$
- The probability that all clauses are satisfied is therefore at least $1 - 2^{(1 - \frac{\alpha t}{4}) |P|}$
Probability of sparsity

Lemma (5.4.15)

Let $x, y, z$ be such that $x \leq 1, \frac{1}{k-1} < y \leq 1, 2^{\frac{1}{k}} \leq z$, and let $\rho$ be any restriction of size $t$ variables with

$$t \leq \min\left\{ \frac{xn}{2}, \frac{x^{(1-\frac{1+1/y}{k})}n^{1-2/k}}{z} \right\}.$$  

If $F$ is chosen as a random $k$ – CNF formula with at most

$$\frac{y}{e^{1+1/y}2^{k+1/y}}x^{1/y-(k-1)}n$$

clauses then:

$$\Pr[F[p \text{ is both } xn \text{ – and } (\frac{xn}{2}, xn, y) \text{ – sparse}] \geq 1 - 2^{-t} - 2z^{-k} - \frac{1}{n}$$
What is the general idea?

- Basically, with large probability after applying this type of refutation $\rho$ the random $k$–CNF formula still has a certain sparsity.
What is the general idea?

- Basically, with large probability after applying this type of refutation $\rho$ the random $k$ – CNF formula still has a certain sparsity.
- By the complex clause lemma each resolution refutation for a formula with that sparsity must contain a long clause.
What is the general idea?

- Basically, with large probability after applying this type of refutation $\rho$ the random $k - CNF$ formula still has a certain sparsity.
- By the complex clause lemma each resolution refutation for a formula with that sparsity must contain a long clause.
- The refutation $\rho$ removed all long clauses from the formula.
What is the general idea?

- Basically, with large probability after applying this type of refutation $\rho$ the random $k$ – CNF formula still has a certain sparsity.
- By the complex clause lemma each resolution refutation for a formula with that sparsity must contain a long clause.
- The refutation $\rho$ removed all long clauses from the formula.
- Contradiction!
The result

- Exponential size proofs are required for random $k$–CNF formulas with $m \leq n^{(k-1)/4}$. 
Conclusion

- Proving that refutations for random $k-\text{CNF}$ formulas are of exponential size is far from trivial.
- We have seen some definitions and lemma’s that helped us get the general idea behind the proof.
- And as an analogue to Petri’s conclusion:
Conclusion

- Proving that refutations for random $k - CNF$ formulas are of exponential size is far from trivial.
- We have seen some definitions and lemma’s that helped us get the general idea behind the proof.
- And as an analogue to Petri’s conclusion:

  Bravery and stupidity are closely related.