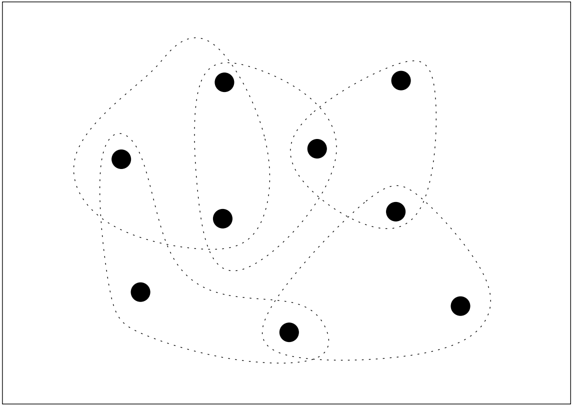
Linear Formulas Are Large and Complex

Dominik Scheder, ETH Zürich

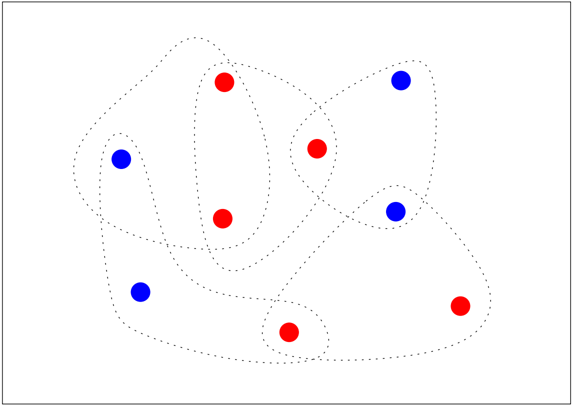
STACS 2010, Nancy, March 4

Some Extremal Combinatorics

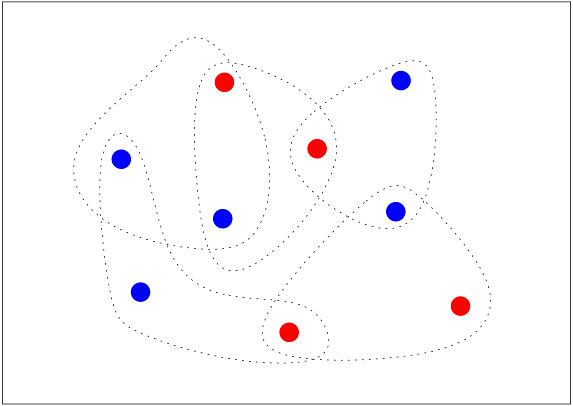
Some Extremal Combinatorics



Some Extremal Combinatorics



Some Extremal Combinatorics



Colorability of Hypergraphs

Theorem (Erdős)

A k-uniform hypergraph with less than 2^{k-1} edges is 2-colorable.

Colorability of Hypergraphs

Theorem (Erdős)

A k-uniform hypergraph with less than 2^{k-1} edges is 2-colorable.

Proof: Take a random coloring.

Colorability of Hypergraphs

Theorem (Erdős)

A k-uniform hypergraph with less than 2^{k-1} edges is 2-colorable.

Proof: Take a random coloring.

Theorem (Erdős)

There exists a k-uniform hypergraph with $k^2 2^k$ edges that is not 2-colorable.

Theorem (Erdős)

A k-uniform hypergraph with less than 2^{k-1} edges is 2-colorable.

Proof: Take a random coloring.

Theorem (Erdős)

There exists a k-uniform hypergraph with $k^2 2^k$ edges that is not 2-colorable.

Proof: Take k^2 vertices and randomly select $k^2 2^k$ hyperedges from the $\binom{k^2}{k}$ possible ones.

CNF Formulas

CNF Formulas

$$(x_2) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_3 \vee x_4)$$

CNF Formulas

$$(x_2) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_3 \vee x_4)$$

assignment: $x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1$

CNF Formulas

$$(x_2) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_3 \vee x_4)$$

assignment: $x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1$

CNF Formulas

$$(x_2) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3) \wedge (\bar{x}_3 \vee x_4)$$

assignment: $x_1 \mapsto 1, x_2 \mapsto 1, x_3 \mapsto 0, x_4 \mapsto 1$

k-CNF formulas: $(x_1 \vee x_2 \vee \bar{x}_4) \wedge (\bar{x}_1 \vee x_3 \vee \bar{x}_4) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3)$

Extremal Results on CNF Formulas

Theorem (Folklore)

A k-CNF formula with less than 2^k clauses is satisfiable.

Extremal Results on CNF Formulas

Theorem (Folklore)

A k-CNF formula with less than 2^k clauses is satisfiable.
There exists an unsatisfiable k-CNF formula with 2^k clauses.

Extremal Results on CNF Formulas

Theorem (Folklore)

A k-CNF formula with less than 2^k clauses is satisfiable.
There exists an unsatisfiable k-CNF formula with 2^k clauses.

$$
\begin{aligned}
&(x \vee y \vee z) \quad \wedge \\
&(x \vee y \vee \bar{z}) \quad \wedge \\
&(x \vee \bar{y} \vee z) \quad \wedge \\
&(x \vee \bar{y} \vee \bar{z}) \quad \wedge \\
&(\bar{x} \vee y \vee z) \quad \wedge \\
&(\bar{x} \vee y \vee \bar{z}) \quad \wedge \\
&(\bar{x} \vee \bar{y} \vee z) \quad \wedge \\
&(\bar{x} \vee \bar{y} \vee \bar{z})
\end{aligned}
$$

Linear CNF Formulas

Linear CNF Formulas

Definition

F is linear if any two clauses share at most one variable.

Linear CNF Formulas

Definition

F is linear if any two clauses share at most one variable.

Forbidden: $(x \vee y \vee z) \wedge (x \vee \bar{z} \vee u)$

Linear CNF Formulas

Definition

F is linear if any two clauses share at most one variable.

Forbidden: $(x \vee y \vee z) \wedge (x \vee \bar{z} \vee u)$
Allowed: $(x \vee u \vee v) \wedge (\bar{x} \vee y \vee w)$

Existence of Unsatisfiable Linear k-CNF Formulas

Question (Porschen, Randerath, Speckenmeyer)

Are there unsatisfiable linear k-CNF formulas?

Existence of Unsatisfiable Linear k-CNF Formulas

Question (Porschen, Randerath, Speckenmeyer)

Are there unsatisfiable linear k-CNF formulas?

- $k = 0$: $\Box$ (also called 0, or f, or $\bot$)

Existence of Unsatisfiable Linear k-CNF Formulas

Question (Porschen, Randerath, Speckenmeyer)

Are there unsatisfiable linear k-CNF formulas?

- $k = 0$: $\Box$ (also called 0, or f, or $\bot$)
- $k = 1$: $(x) \wedge (\bar{x})$

Existence of Unsatisfiable Linear k-CNF Formulas

Question (Porschen, Randerath, Speckenmeyer)

Are there unsatisfiable linear k-CNF formulas?

- $k = 0$: $\square$ (also called 0, or f, or $\perp$)
- $k = 1$: $(x) \wedge (\bar{x})$
- $k = 2$: $(\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$

Existence of Unsatisfiable Linear k-CNF Formulas

Question (Porschen, Randerath, Speckenmeyer)

Are there unsatisfiable linear k-CNF formulas?

- $k = 0$: $\Box$ (also called 0, or f, or $\bot$)
- $k = 1$: $(x) \wedge (\bar{x})$
- $k = 2$: $(\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$
- $k = 3$: 30 clauses, maybe less...

Existence and Upper Bounds

Theorem

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Existence and Upper Bounds

Theorem

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Proof:

$$(x \vee y \vee z) \wedge (x \vee u \vee v) \wedge (y \vee u \vee w) \wedge (y \vee z \vee v)$$

Theorem

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Proof:

$$(\bar{x} \vee y \vee \bar{z}) \wedge (x \vee \bar{u} \vee \bar{v}) \wedge (\bar{y} \vee \bar{u} \vee w) \wedge (y \vee z \vee v)$$

Theorem

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Proof:

$$(x \vee \bar{y} \vee z) \wedge (\bar{x} \vee \bar{u} \vee v) \wedge (y \vee u \vee w) \wedge (\bar{y} \vee \bar{z} \vee \bar{v})$$

Theorem

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Proof:

$$(\bar{x} \vee y \vee z) \wedge (x \vee \bar{u} \vee \bar{v}) \wedge (\bar{y} \vee \bar{u} \vee w) \wedge (y \vee \bar{z} \vee v)$$

Lower Bounds

Theorem

Any linear k-CNF formulas with less than $c4^k/k^2$ clauses is satisfiable.

Lower Bounds

Theorem

Any linear k-CNF formulas with less than $c4^k/k^2$ clauses is satisfiable.

Proof: Lovász Local Lemma $+$ Tweaking

Probabilistic vs. Explicit Constructions

Question

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.

Question

There are unsatisfiable linear k-CNF formulas with $O(k^2 4^k)$ clauses.
Yeah, but show me one. . .

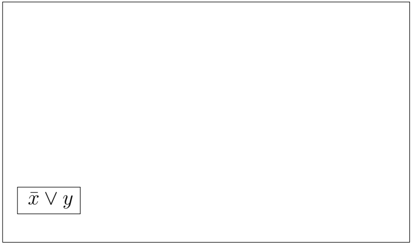
Resolution Trees

What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$

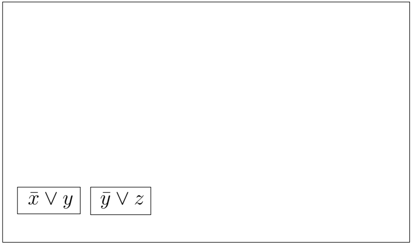
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



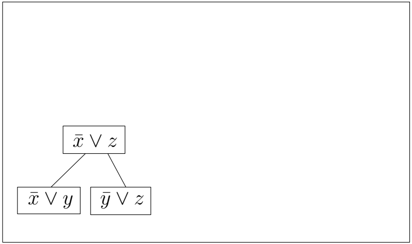
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



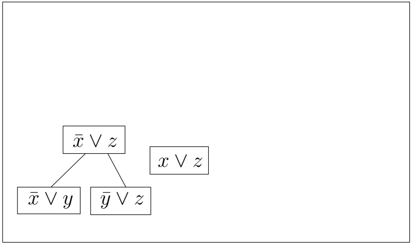
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



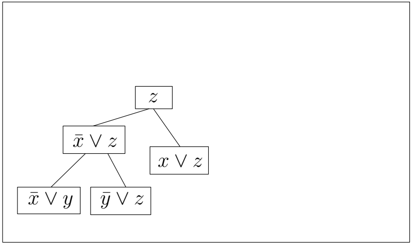
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$

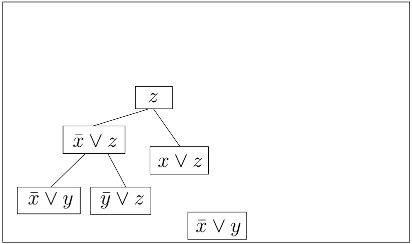


What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$

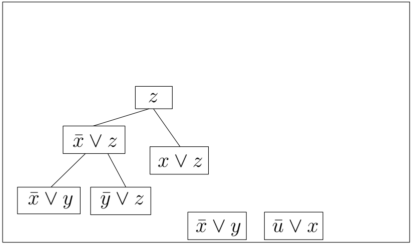


$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



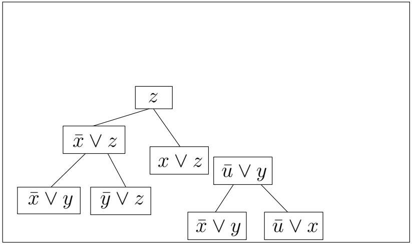
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



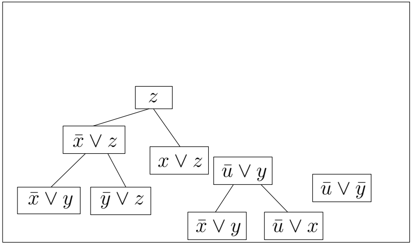
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



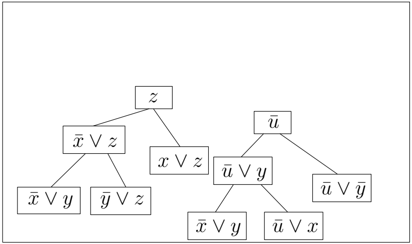
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



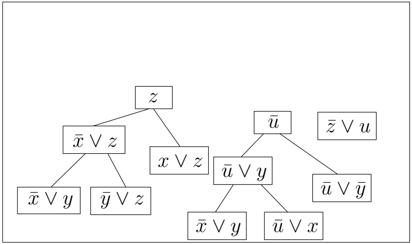
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



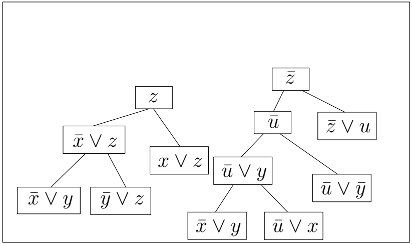
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



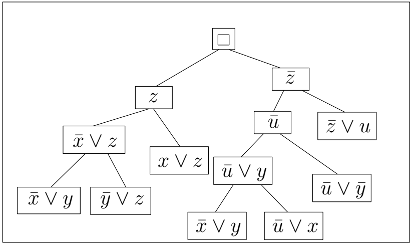
What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



What Are Resolution Trees

$$F = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



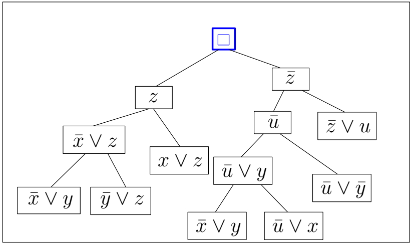
Quite Large Resolution Trees

Theorem

Suppose F is an unsatisfiable linear k-CNF formula. Any resolution tree for F has at least $2^{2^{k/2-1}}$ nodes.

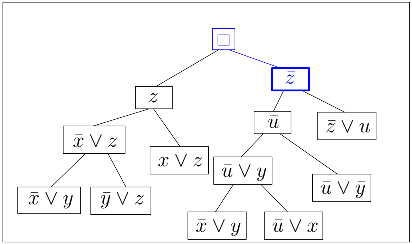
Taking a Random Walk

$$F_0 = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



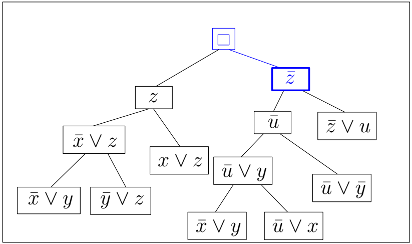
Taking a Random Walk

$$F_0 = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



Taking a Random Walk

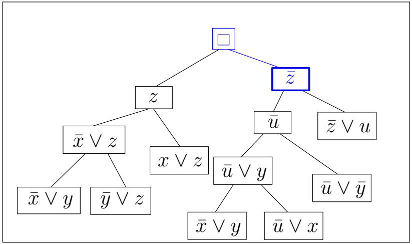
$$F_0 = (\bar{x} \vee y) \wedge (\bar{y} \vee z) \wedge (\bar{z} \vee u) \wedge (\bar{u} \vee x) \wedge (x \vee z) \wedge (\bar{y} \vee \bar{u})$$



$z \mapsto 1$

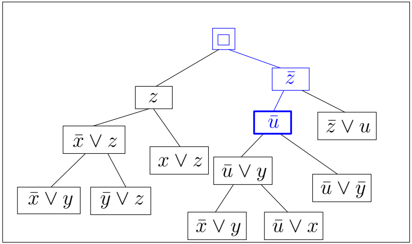
Taking a Random Walk

$$F_1 = (\bar{x} \vee y) \wedge \qquad (\quad u) \wedge (\bar{u} \vee x) \wedge \qquad (\bar{y} \vee \bar{u})$$



$z \mapsto 1$

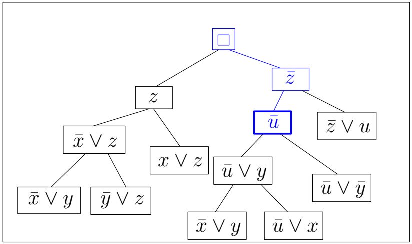
$$F_1 = (\bar{x} \vee y) \wedge \qquad (\quad u) \wedge (\bar{u} \vee x) \wedge \qquad (\bar{y} \vee \bar{u})$$



$z \mapsto 1$

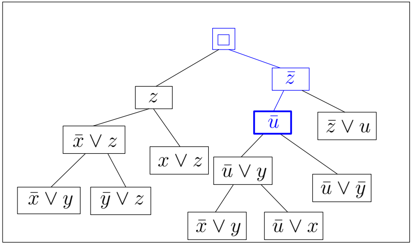
Taking a Random Walk

$$F_1 = (\bar{x} \vee y) \wedge \qquad (\quad u) \wedge (\bar{u} \vee x) \wedge \qquad (\bar{y} \vee \bar{u})$$



$z \mapsto 1$, $u \mapsto 1$

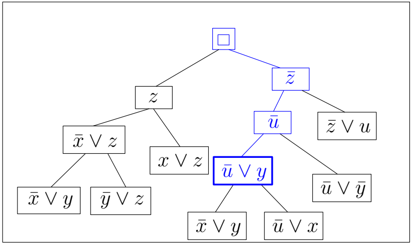
$$F_2 = (\bar{x} \vee y) \wedge \qquad (\quad x) \wedge \qquad (\bar{y} \quad)$$



$z \mapsto 1$, $u \mapsto 1$

Taking a Random Walk

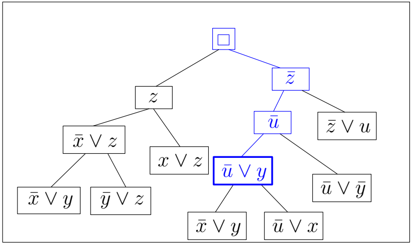
$F_2 = (\bar{x} \vee y) \wedge \qquad (\quad x) \wedge \qquad (\bar{y} \quad)$



$z \mapsto 1$, $u \mapsto 1$

Taking a Random Walk

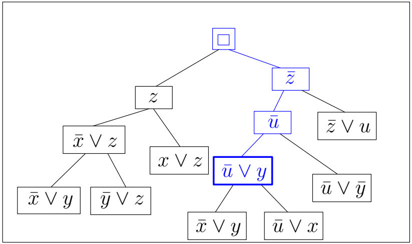
$$F_2 = (\bar{x} \vee y) \wedge \quad (\quad x) \wedge \quad (\bar{y} \quad)$$



$$z \mapsto 1 \, , u \mapsto 1 \, , y \mapsto 0$$

Taking a Random Walk

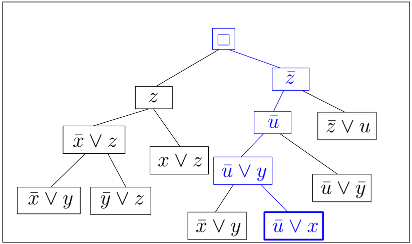
$F_3 = (\bar{x} \quad) \wedge \qquad\qquad (\quad x)$



$z \mapsto 1 \, , u \mapsto 1 \, , y \mapsto 0$

Taking a Random Walk

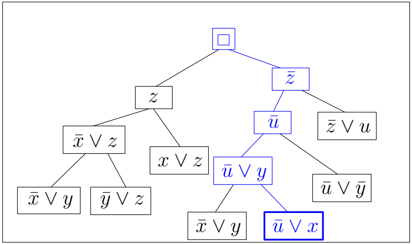
$$F_3 = (\bar{x} \quad) \wedge \qquad\qquad\qquad (\quad x)$$



$$z \mapsto 1 \,, u \mapsto 1 \,, y \mapsto 0$$

Taking a Random Walk

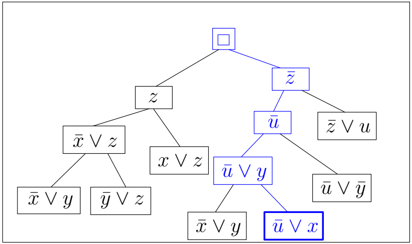
$$F_3 = (\bar{x} \quad) \wedge \qquad (\quad x)$$



$$z \mapsto 1\,, u \mapsto 1\,, y \mapsto 0\,, x \mapsto 0$$

Taking a Random Walk

$F_4 = \qquad\qquad\qquad (\quad)$



$z \mapsto 1\ ,\ u \mapsto 1\ ,\ y \mapsto 0\ ,\ x \mapsto 0$

Taking a Random Walk

- A function $\{L, R\}^* \rightarrow \{\text{nodes of the resolution tree}\}$

Taking a Random Walk

- A function $\{L, R\}^* \rightarrow \{$nodes of the resolution tree$\}$
- Produces a sequence $F_0, F_1, F_2, \ldots$

Taking a Random Walk

- A function $\{L, R\}^* \to \{\text{nodes of the resolution tree}\}$
- Produces a sequence $F_0, F_1, F_2, \ldots$
- Walk reaches a leaf $\Rightarrow F_i$ contains the empty clause

Taking a Random Walk

- A function $\{L, R\}^* \to \{\text{nodes of the resolution tree}\}$
- Produces a sequence $F_0, F_1, F_2, \ldots$
- Walk reaches a leaf $\Rightarrow$ F_i contains the empty clause

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\#\text{ inner nodes} \quad \geq \quad 2^\ell \;\; \Pr[\text{walk doesn't reach a leaf}]$$

Taking a Random Walk

- A function $\{L, R\}^* \to \{\text{nodes of the resolution tree}\}$
- Produces a sequence $F_0, F_1, F_2, \ldots$
- Walk reaches a leaf $\Rightarrow$ F_i contains the empty clause

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned}\#\text{ inner nodes} &\geq 2^\ell \ \Pr[\text{walk doesn't reach a leaf}] \\ &\geq 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}]\end{aligned}$$

Bounding Pr[F_ℓ contains empty clause]

$$w(F) := \sum_{\text{clauses } C \in F : |C| \leq k-2} 2^{-|C|}$$

Bounding Pr[F_ℓ contains empty clause]

$$w(F) := \sum_{\text{clauses } C \in F : |C| \leq k-2} 2^{-|C|}$$

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned} \#\text{ inner nodes} &\geq 2^\ell \ \Pr[\text{walk doesn't reach a leaf}] \\ &\geq 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}] \end{aligned}$$

Bounding Pr[F_ℓ contains empty clause]

$$w(F) := \sum_{\text{clauses } C \in F : |C| \leq k-2} 2^{-|C|}$$

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned}
\#\text{ inner nodes} &\geq 2^\ell \ \Pr[\text{walk doesn't reach a leaf}] \\
&\geq 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}] \\
&\geq 2^\ell \Pr[w(F_\ell) < 1]
\end{aligned}$$

Bounding $\Pr[F_\ell$ contains empty clause $]$

$$w(F) := \sum_{\text{clauses } C \in F : |C| \le k-2} 2^{-|C|}$$

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned}
\#\text{ inner nodes} &\ge 2^\ell \Pr[\text{walk doesn't reach a leaf}] \\
&\ge 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}] \\
&\ge 2^\ell \Pr[w(F_\ell) < 1] \\
&\ge 2^\ell (1 - \mathbb{E}[w(F_\ell)])
\end{aligned}$$

Bounding Pr[F_ℓ contains empty clause]

$$w(F) := \sum_{\text{clauses } C \in F : |C| \leq k-2} 2^{-|C|}$$

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned}
\#\text{ inner nodes} &\geq 2^\ell \text{ Pr[walk doesn't reach a leaf]} \\
&\geq 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}] \\
&\geq 2^\ell \Pr[w(F_\ell) < 1] \\
&\geq 2^\ell (1 - \mathbb{E}[w(F_\ell)])
\end{aligned}$$

a little bit more work

Bounding Pr[F_ℓ contains empty clause]

$$w(F) := \sum_{\text{clauses } C \in F : |C| \leq k-2} 2^{-|C|}$$

Observation

For a random walk $F_0, F_1, \ldots, F_\ell$ of ℓ steps:

$$\begin{aligned}
\#\text{ inner nodes} &\geq 2^\ell \; \Pr[\text{walk doesn't reach a leaf}] \\
&\geq 2^\ell \Pr[F_\ell \text{ doesn't contain empty clause}] \\
&\geq 2^\ell \Pr[w(F_\ell) < 1] \\
&\geq 2^\ell (1 - \mathbb{E}[w(F_\ell)])
\end{aligned}$$

a little bit more work

$$\geq 2^\ell (1 - \ell^2 2^{-k})$$

Size of Resolution Trees

$$\#\text{ inner nodes} \geq 2^{\ell}(1 - \ell^2 2^{-k})$$

Size of Resolution Trees

$$\#\text{ inner nodes} \geq 2^{\ell}(1 - \ell^2 2^{-k})$$

Plug in $\ell := \sqrt{2^{k-1}}$

$$\#\text{ inner nodes} \geq 2^{\ell}(1 - \ell^2 2^{-k})$$

Plug in $\ell := \sqrt{2^{k-1}}$

Theorem

Suppose F is an unsatisfiable linear k-CNF formula. Any resolution tree for F has at least $2^{2^{k/2-1}}$ nodes.

Conclusion

Conclusion

- unsatisfiable linear k-CNF formulas exist

Conclusion

- unsatisfiable linear k-CNF formulas exist
- large, but not too large

Conclusion

- unsatisfiable linear k-CNF formulas exist
- large, but not too large
- hard to construct explicitly

Conclusion

- unsatisfiable linear k-CNF formulas exist
- large, but not too large
- hard to construct explicitly

They might help to understand the “Hay in a Haystack Paradox”