# Hard examples for DPLL algorithms

Dmitry Itsykson

Steklov Institute of Mathematics at St. Petersburg

Theory Days
Ocober 8, 2011

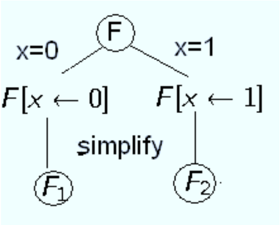
# Outline

1. DPLL
2. Lower bounds on unsatisfiable instances
3. Lower bounds on satisfiable instances
4. Goldreich's one-way candidate
5. DPLL algorithms with cut heuristic

# DPLL algorithm

$(x \vee y \vee \neg z) \wedge (\neg x \vee \neg y) \wedge (\neg y \vee z)$

$(x \vee y \vee \neg z) \wedge (\neg x \vee \neg y) \wedge (\neg y \vee z)$, $x := 0, y := 1, z := 1$




- Heuristic **A** chooses the variable $x$
- Heuristic **B** chooses the branch to be examinated first
- Simplification rules

Examples of heuristics:

- **A** chooses
  - the most frequent variable
  - a variable from the shortest clause
- **B** chooses the most frequent sign

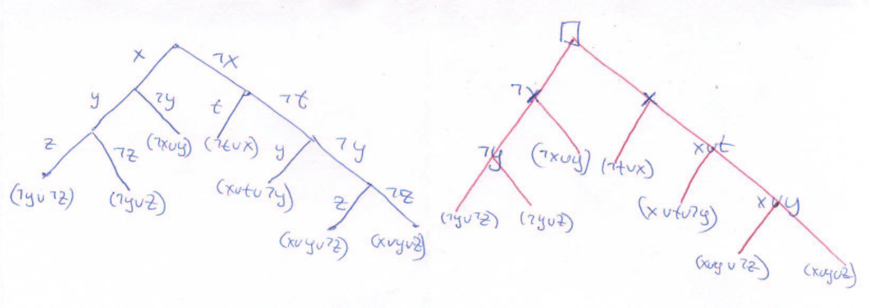
Simplification rules:

- Unit clause elimination
- Pure literal rule

# DPLL on unsatisfiable formulas

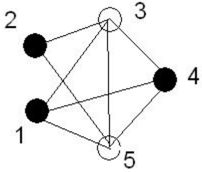
- Resolution rule: $\frac{(A \vee x) \qquad (B \vee \neg x)}{A \vee B}$.

$$(x \vee y \vee z) \wedge (x \vee y \vee \neg z) \wedge (x \vee t \vee \neg y) \wedge (\neg t \vee x) \wedge (\neg x \vee y) \wedge (\neg y \vee z) \wedge (\neg y \vee \neg z)$$



# Hard examples for resolutions

- Tseitin formulas [1968].



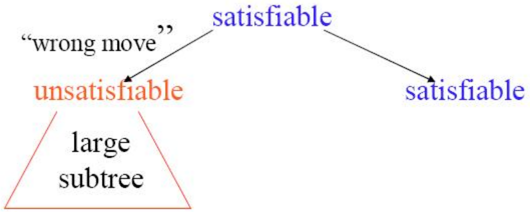

1. $e_{13} \oplus e_{14} \oplus e_{15} = 1$
2. $e_{23} \oplus e_{25} = 1$
3. $e_{13} \oplus e_{23} \oplus e_{34} \oplus e_{35} = 0$
4. $e_{14} \oplus e_{34} \oplus e_{45} = 1$
5. $e_{15} \oplus e_{25} \oplus e_{35} \oplus e_{45} = 0$

    - Constant degree
    - $\forall A \subseteq V$ if $\frac{|V|}{3} \leq |A| \leq \frac{2|V|}{3}$, then $E[A, V \setminus A] \geq \alpha |V|$.

- Pigeonhole Principle
    - $n+1$ pigeons and $n$ holes
    - $p_{ij}$: $i$-th pigeon is in $j$-th hole
    - $\forall i \in [1 \ldots n+1]$: $(p_{i1} \vee p_{i2} \vee \cdots \vee p_{in})$
    - $\forall k \in [1 \ldots n] \forall i, j \in [1..n+1]$: $(\neg p_{ik} \vee \neg p_{jk})$

# Lower bounds on satisfiable formulas

- If **P** = **NP** then no superpolynomial lower bounds for DPLL algorithms since heuristic **B** may choose corect value.
- Satisfiable formulas are much easier for solvers
- [Nikolenko, 2002], [Achilioptas,Beame, Molloy, 2003-2004] exponential lower bound for specific DPLL algoritms
- [Alekhnovich, Hirsch, Itsykson, 2005] Exponential lower bound for myopic and drunken algorithms.
- Inverting of functions corresponds to satisfiable formulas



# Myopic algorithms

- Myopic heuristics **A**, **B**:
  - Read formula with erased negations
  - Read $K = n^{1-\varepsilon}$ clauses
  - Query the number of positive and negative occurrences of variable

$$\begin{array}{ccc} (x_1 \vee x_3 \vee x_5) & & (x_1 \vee x_3 \vee x_5) \\ \color{red}{(x_2 \vee x_3)} & & (x_2 \vee \color{red}{\neg} x_3) \\ (x_2 \vee x_4 \vee x_5) & \Rightarrow & (x_2 \vee x_4 \vee x_5) \\ \color{red}{(x_1 \vee x_4 \vee x_6)} & & (x_1 \vee \color{red}{\neg} x_4 \vee x_6) \end{array}$$

- Lower bound:
  - $Ax = b$ over $\mathbb{F}_2$
  - $A$ is a randomly constructed 0/1 matrix
  - exactly 3 ones per row; full rank
  - $x \oplus y \oplus z = 1 \iff (\neg x \vee \neg y \vee \neg z) \wedge (\neg x \vee y \vee z) \wedge (x \vee \neg y \vee z) \wedge (x \vee y \vee \neg z)$
  - $x \oplus y \oplus z = 0 \iff (x \vee y \vee z) \wedge (\neg x \vee \neg y \vee z) \wedge (\neg x \vee y \vee \neg z) \wedge (x \vee \neg y \vee \neg z)$

## Drunken algorithms

Drunken heuristics:

- **A**: any!
- **B**: random 50:50

Lower bound:

- $F$ is a hard unsatisfiable formula
- $F' = F +$ one satisfying assignmet
- Wrong substitution during first several step w.h.p.
- Fall to hard unsatisfiable formula

Combination:

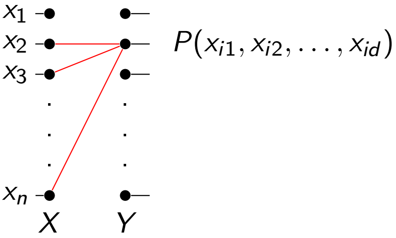
- **A**: any!
- **B**: myopic

Cheating:

- **A** chooses variable that satisfy first clause.

# Goldreich's one-way candidate

$f : \{0,1\}^n \to \{0,1\}^n$




- $G(X, Y, E)$ is a bipartite graph;
- $\forall y \in Y \quad deg(y) = d$
- $d$ is a constant.

Goldreich's conjecture:

- $P$ is a random predicate;
- $G$ is an expander;

then function $f$ is a one-way.

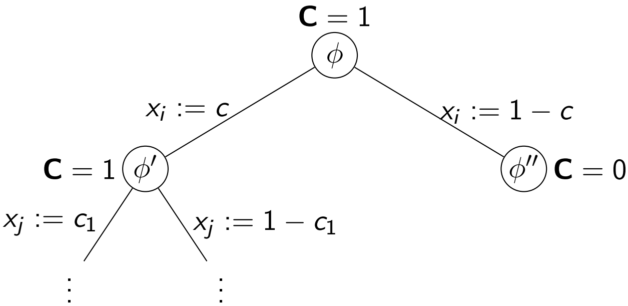
- $f$ is computed by constant depth circuit;
- [Applebaum, Ishai, Kushilevitz 2006] If one-way functions exist then there is a one-way function that can be computed by constant depth circuit.

# Exponential lower bounds on the complexity of invertion of Goldreich's function

$P(x_1, \dots x_d) = x_1 \oplus \dots \oplus x_{d-k} \oplus Q(x_{d-k+1}, \dots, x_d), k < d/4.$

| Paper | Graph | DPLL |
|---|---|---|
| Cook, Etesami, Miller, Trevisan, 2009 | Random | Myopic |
| Itsykson, 2010 | Random | Drunken |
| Itsykson, Sokolov, 2011 | Explicit (based on expander) | Drunken Myopic |

# DPLL with cut heuristic




- Heuristic **A** chooses a variable for splitting.
- Heuristic **B** chooses first value.
- Heuristic **C** cuts unpromising branches.

- Algorithm is correct on unsatisfiable formulas.
- Possible errors on satisfiable formulas
- Correctnes vs. effectiveness tradeoff

## Correctnes vs. effectiveness

[Itsykson, Sokolov 2011]
Theorem. There exists family of unsatisfiable formulas $\Phi_n$ such that $\forall$ deterministic myopic $\mathbf{A}, \mathbf{C}$ there exists polynomial-time samplable ensemble of distributions $D_n$ with $\operatorname{supp} D_n \subset SAT$ such that $\forall \mathbf{B}$ either

- $\Pr_{\varphi \leftarrow D_n}[DPLL_{\mathbf{A},\mathbf{B},\mathbf{C}}(\varphi) = 1] < 1/100$ or
- Running time of $DPLL_{\mathbf{A},\mathbf{B},\mathbf{C}}(\Phi_n)$ is $2^{\Omega(n)}$.

Theorem. There exists family of unsatisfiable formulas $\Phi_n$ and polynomial-time samplable ensemble of distributions $R_n$ with $\operatorname{supp} R_n \subset SAT$ such that $\forall$ deterministic myopic $\mathbf{A}, \mathbf{C}$ and $\forall \mathbf{B}$ if $\Pr_{\varphi \leftarrow D_n}[DPLL_{\mathbf{A},\mathbf{B},\mathbf{C}}(\varphi) = 1] = 1 - o(1)$, then running time of $DPLL_{\mathbf{A},\mathbf{B},\mathbf{C}}(\Phi_n)$ is $2^{\Omega(n)}$.