On Some SAT-Variants over Linear Formulas

Stefan Porschen

and

Tatjana Schmidt

Institut für Informatik

Universität zu Köln

Overview

1. Motivation
2. Introduction and Preliminaries
3. NAE-SAT and XSAT-complexity of Linear Formulas
4. NAE-SAT and XSAT-complexity of Monotone Linear Formulas
5. NAE-SAT and XSAT-complexity of Exact Linear Formulas
6. Open Problems

Motivation

- SAT was shown to be NP-complete when restricted to the class of linear CNF formulas.
- There is some intuition that leads to the conjecture the linear formulas form the hard kernel for CNF-SAT.
- Considering NAESAT and XSAT for linear formulas: Is there any difference to SAT?

Introduction and Preliminaries

- A positive (negative) **literal** is a (negated) variable $\in \{0, 1\}$: x, $\overline{x}$.

- The **complement** of a literal l is its negation $\overline{l}$.

- Each **formula** is considered as a clause set, and each **clause** is represented as a literal set.
  Example:

$$\{\{x_1, \overline{x_2}\}, \{\overline{x_1}, x_2, x_3\}, \{x_1, \overline{x_3}\}\}$$

Let $C = \{\{x_1, \overline{x_2}, \overline{x_3}\}, \{\overline{x_2}, x_4, x_5\}, \{x_3, \overline{x_5}\}\}$.

- **SAT** asks, whether $C \in$ CNF has a **model** $t : V(C) \to \{0, 1\}$ assigning **at least one** literal in each clause of C to 1.
  C: $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, $x_4 = 1$, $x_5 = 1$.

- **NAE-SAT** asks for a **nae-model** setting **at least one** literal in each clause to 1 and **at least one** literal to 0.
  C: $x_1 = 0$, $x_2 = 0$, $x_3 = 1$, $x_4 = 0$, $x_5 = 1$.

- **XSAT** asks for a **x-model** setting **exactly one** literal in each clause to 1 and all other literals to 0.
  C: $x_1 = 1$, $x_2 = 1$, $x_3 = 1$, $x_4 = 0$, $x_5 = 1$.

Introduction and Preliminaries

- A CNF formula C is called **linear** if

(1) C contains no pair of complementary unit clauses and

(2) for all $c_1, c_2 \in C : c_1 \neq c_2$ we have $|V(c_1) \cap V(c_2)| \leq 1$.

Example: $\{\{x_1, x_2\}, \{\overline{x_2}, x_3\}, \{x_4, x_5\}\}$.

- C is called **exact linear** if

(1) C contains no pair of complementary unit clauses and

(2) for all $c_1, c_2 \in C : c_1 \neq c_2$ we have $|V(c_1) \cap V(c_2)| = 1$.

Example: $\{\{x_1, x_2\}, \{\overline{x_2}, x_3\}, \{\overline{x_1}, x_3, x_4\}\}$.

Introduction and Preliminaries

- A **monotone** formula has no negated variables;
  (L)CNF$_+$ denotes the set of all (linear) monotone formulas.

- A formula is k-**uniform** if all its clauses have length exactly k, where $k > 0$ is an integer.

NAESAT **and** XSAT-**complexity of Linear Formulas**

Theorem 1 *Both* XSAT *and* NAESAT *remain* NP-*complete when restricted to the class* LCNF.

Proof:

- T.J.Schaefer: CNF-XSAT is NP-complete.
- First provide a polynomial time reduction from CNF-XSAT to LCNF-XSAT:

NAESAT and XSAT-complexity of Linear Formulas

- Let $C \in CNF$ and $x_i \in V(C)$ occuring in the clauses $c_{j_1}, \ldots, c_{j_r} \in C$.

- Introduce a new variable $y_{x_i}^{j_s} \notin V(C)$, for each such occurrence $1 \leq s \leq r$.

- Replace the occurrence of x_i in c_{j_s} with $y_{x_i}^{j_s}$ not affecting the polarity, for $1 \leq s \leq r$.

- Add the following clauses to the current formula, for each fixed $x_i \in V(C)$:

$$\{\overline{x_i}, y_{x_i}^{j_1}\}, \{\overline{y}_{x_i}^{j_1}, y_{x_i}^{j_2}\}, \{\overline{y}_{x_i}^{j_2}, y_{x_i}^{j_3}\}, \ldots, \{\overline{y}_{x_i}^{j_{r-1}}, y_{x_i}^{j_r}\}, \{\overline{y}_{x_i}^{j_r}, x_i\}$$

NAESAT and XSAT-complexity of Linear Formulas

- These clauses are equivalent to the implicational chain:

$$x_i \to y_{x_i}^{j_1} \to y_{x_i}^{j_2} \ldots \to y_{x_i}^{j_r} \to x_i$$

  All variables contained in the 2-clauses of an implicational chain are forced to be set equally. Thus: $C \in$ XSAT if and only if $C' \in$ XSAT.

On 2-clauses XSAT and NAE-SAT coincide. Therefore the additional 2-clauses are all satisfied w.r.t. NAE-SAT.

Definition 1 *Let C be a x-satisfiable formula. A variable $y \in V(C)$ is called a* **x-backbone variable** *of C, if y has the same value in each x-model of C.*

Example:

Consider the x-satisfiable formula C:

$$C = \{\{x_1, x_2, x_5\}, \{x_2 x_3\}, \{x_1, x_3, x_4\}\}$$

The only x-models of C obviously are:

$$x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1, x_5 = 0$$

and

$$x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 1$$

Thus x_1 is a x-backbone variable.

NAESAT **and** XSAT**-complexity of Linear Formulas**

Lemma 2 *For each fixed $k \geq 3$, we can efficiently construct a monotone k-uniform linear formula C of $O(k^3)$ many variables and $O(k^2)$ many clauses such that C contains at least k x-backbone variables that all must be assigned 0.*

Proof: Example:

For case $k = 4$:

$$C' = \begin{array}{cccc} x & y_1 & y_2 & y_3 \\ x & a_{11} & a_{12} & a_{13} \\ x & a_{21} & a_{22} & a_{23} \\ x & a_{31} & a_{32} & a_{33} \\ y_1 & a_{11} & a_{21} & a_{31} \\ y_1 & a_{12} & a_{22} & a_{32} \\ y_1 & a_{13} & a_{23} & a_{33} \\ y_2 & a_{11} & a_{23} & a_{32} \\ y_2 & a_{12} & a_{21} & a_{33} \end{array}, \quad c = y_2\, a_{13}\, a_{22}\, a_{31}$$

- $C' \in$ XSAT: A (canonical) x-model t for C' is provided by assigning 1 to all variables in the removed clause c, and 0 to all other variables.

- Each x-model of C' assignes 0 to x and y_1. ∎

Definition 3 *A formula is called* **minimally nae-unsatisfiable** *if it is nae-unsatisfiable, but removing an arbitrary clause from it yields a nae-satisfiable formula. We call a set $U \subseteq L(C)$ of literals in a nae-satisfiable formula C a* **nae-backbone set**, *if each nae-model of C sets the literals in U either all to 0 or all to 1.*

Example:
The following formula is minimally nae-unsatisfiable:

$$\{x_1, x_2, x_3\}$$
$$\{x_1, x_4, x_5\}$$
$$\{x_1, x_6, x_7\}$$
$$\{x_2, x_4, x_6\}$$
$$\{x_2, x_5, x_7\}$$
$$\{x_3, x_4, x_7\}$$
$$\{x_3, x_5, x_6\}$$

Lemma 4 *Given a formula C that is nae-unsatisfiable. Then C contains a minimally nae-unsatisfiable subformula C'. Moreover, any clause c of C' forms a nae-backbone set in $C' - \{c\}$.*

Proof:

- There must exist a nae-satisfiable subformula of C, because any single clause of it has this property.

- Let C be a minimally nae-unsatisfiable formula, and $C' := C - \{c\}$, for arbitrary $c \in C$.
  Then C' is nae-satisfiable and c is a nae-backbone set in C'.

- Observe that $V(c) \subseteq V(C')$.

- Each nae-model of C' sets all literals in c either to 0 or all to 1, otherwise C would be nae-satisfiable yielding a contradiction. ■

Theorem 2 *For each fixed* $k \geq 3$, XSAT, *resp.* NAESAT *remain* NP-*complete restricted to* LCNF($\geq k$).

Proof:

Basic idea:

- Perform the reduction as shown before from CNF($\geq k$)-XSAT (resp. CNF($\geq k$)-NAE-SAT).
- Next step is **padding the added 2-clauses by x-backbone variables** resp. nae-backbone sets such that they get k-clauses and the XSAT (resp. NAE-SAT) status of the corresponding formulas is preserved.

Theorem 3 *Both* XSAT *and* NAESAT *remain NP-complete when restricted to the class* LCNF$_+$ *of monotone linear formulas.*

Proof: Polynomial time reduction from CNF$_+$-XSAT to LCNF$_+$-XSAT:

- Let $C \in$ CNF$_+$. For each fixed variable $x_i \in V(C)$ having $r \geq 2$ occurrences in C, in the clauses $c_{j_1}, \ldots, c_{j_r}$ of C, introduce the new variables $y_{x_i}^{j_1}, \ldots, y_{x_i}^{j_r} \notin V(C)$.
- Replace the occurrence of x_i in c_{j_s} with $y_{x_i}^{j_s}$, for $1 \leq s \leq r$.

- Introduce an auxiliary variable z_{x_i} also different from all variables.

- Add the following clauses to the formula such that all new variables (i.e. variables not in $V(C)$) are pairwise distinct:

$$(*) \quad \{x_i, z_{x_i}\} \cup \bigcup_{1 \le s \le r} \{y_{x_i}^{j_s}, z_{x_i}\}$$

- Any x-model of C which assigns to variable x_i a fixed truth value, assigns the same value to all new variables $y_{x_i}^{j_s}, 1 \le s \le r$, replacing x_i in C':

$$x_i = 1 \;\Rightarrow\; z_{x_i} = 0 \Rightarrow y_{x_i}^{j_s} = 1, \quad 1 \le s \le r$$
$$x_i = 0 \;\Rightarrow\; z_{x_i} = 1 \Rightarrow y_{x_i}^{j_s} = 0, \quad 1 \le s \le r$$

- Conversely, any x-model of C' assigning a fixed truth value to one of the new variables $y_{x_i}^{j_s}$ replacing x_i must assign the same value to x_i and also to all other variables replacing x_i:

$$y_{x_i}^{j_s} = 1 \;\Rightarrow\; z_{x_i} = 0 \Rightarrow x_i = 1,\; y_{x_i}^{j_{s'}} = 1, \quad 1 \leq s' \neq s \leq r$$
$$y_{x_i}^{j_s} = 0 \;\Rightarrow\; z_{x_i} = 1 \Rightarrow x_i = 0,\; y_{x_i}^{j_{s'}} = 0, \quad 1 \leq s' \neq s \leq r$$

altogether demonstrating the XSAT-equivalence of $x_i \leftrightarrow y_{x_i}^{j_s}$, $1 \leq s \leq r$. The last observation directly implies that $C \in$ XSAT iff $C' \in$ XSAT.

- The resulting formula C', obtained from C is linear and positive monotone. ■

Theorem 4 *For each* $k \geq 3$, XSAT *remains NP-complete when restricted to* $\text{LCNF}_+(\geq k)$*.* NAESAT *remains NP-complete on* $\text{LCNF}_+(\geq 4)$*.*

Proof:

- Concerning XSAT: The padding formulas used earlier for the non-monotone case can also be used here.

- For NAE-SAT this argument does not hold.
  We need: linear k-uniform monotone formulas that are not nae-satisfiable.
  Solved: $k = 3$: the following monotone *3-block formula* is not nae-satisfiable:

$$B_3 \quad := \quad \{xyz, xuv, xwq, yuw, yvq, zuq, zvw\}$$

$k = 4$: sophisticated construction scheme. ■

Corollary 6 NAESAT *restricted to exact linear formulas is polynomial time solvable and moreover* #NAESAT *is polynomial time solvable.*

Proof:

- S.P. and Speckenmeyer 2008:
  For $C \in$ CNF, such that the variable sets of *each pair* of clauses have exactly one or all members in common, SAT and moreover #SAT can be solved in polynomial time.

- Given an arbitrary CNF formula C, then $C \in$ NAESAT holds true if and only if $C \cup C^\gamma \in$ SAT.

- Let $C \in$ XLCNF be arbitrarily chosen, then $C \cup C^\gamma$ is a formula such that *each pair* of clauses has exactly one or all members in common. ■

Open Problems

Following open problems remain for **future work**:

- Whether we can provide an algorithm solving XSAT restricted to LCNF faster than the best XSAT algorithms working for unrestricted CNF.

- Whether NAE-SAT on LCNF can be done faster than in 2^n steps on input instances over n variables.

- The complexity issues of XSAT on exact linear formulas

- The question whether NAESAT remains NP-complete for the length restricted monotone classes LCNF$_+(\geq k)$, for $k \geq 5$, where clauses are not allowed to contain variables twice

Thank you very much for
your attention!