

Generalizations of Matched CNF Formulas

Stefan Szeider* (szeider@cs.toronto.edu)

*Department of Computer Science, University of Toronto, M5S 3G4 Toronto,
Ontario, Canada*

Abstract. A CNF formula is called *matched* if its associated bipartite graph (whose vertices are clauses and variables) has a matching that covers all clauses. Matched CNF formulas are satisfiable and can be recognized efficiently by matching algorithms. We generalize this concept and cover clauses by collections of bicliques (complete bipartite graphs). It turns out that such generalization indeed gives rise to larger classes of satisfiable CNF formulas which we term *biclique satisfiable*. We show, however, that the recognition of biclique satisfiable CNF formulas is NP-complete, and remains NP-hard if the size of bicliques is bounded.

A satisfiable CNF formula is called *var-satisfiable* if it remains satisfiable under arbitrary replacement of literals by their complements. Var-satisfiable CNF formulas can be viewed as the best possible generalization of matched CNF formulas as every matched CNF formula and every biclique satisfiable CNF formula is var-satisfiable. We show that recognition of var-satisfiable CNF formulas is Π_2^P -complete, answering a question posed by Kleine Büning and Zhao.

Keywords: SAT problem, bipartite graph, matched formula, deficiency, biclique cover, NP-completeness, Π_2^P -completeness, polynomial hierarchy

1. Introduction

The *satisfiability problem* (SAT) asks whether a given propositional formula F in conjunctive normal form (CNF) has a satisfying truth assignment. SAT is the first problem that has been shown NP-complete. Since this problem is important in many areas (e.g., Artificial Intelligence and Operations Research) considerable efforts have been made to identify classes of satisfiable CNF formulas which can be recognized in polynomial time; see [6, 9] for surveys.

Our point of departure is a certain class of satisfiable CNF formulas which can be recognized in polynomial time by matching algorithms. To define this class, we associate to a CNF formula F a bipartite graph $I(F)$, the “incidence graph” of F , where one class of vertices consists of the variables occurring in F , and the other class consists of the clauses of F ; a variable v and a clause C are joined by an edge if v occurs in C (positively or negatively). For example, Figure 1 exhibits the incidence

* The author gratefully acknowledges the support of the Austrian Science Fund (FWF) projects J2111 and J2295.

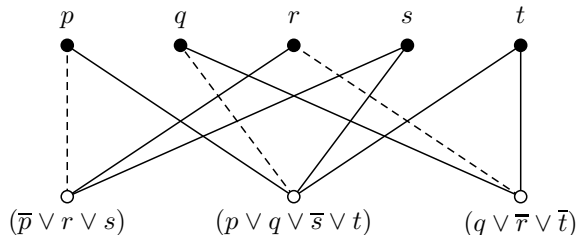


Figure 1. Incidence graph of a matched CNF formula.

graph of the formula

$$F = (\bar{p} \vee r \vee s) \wedge (p \vee q \vee \bar{s} \vee t) \wedge (q \vee \bar{r} \vee \bar{t}) \quad (1)$$

(the meaning of dashed lines will be explained later). Note that polarities of literals in F are not taken into account in forming $I(F)$.

Tovey [18] and independently Aharoni and Linial [1] make the following simple but nevertheless fruitful observation.

If $I(F)$ has a matching M (i.e., a set of independent edges) such that every clause of F is incident with some edge in M , then F is satisfiable.

For, we can obtain a satisfying truth assignment by setting a variable TRUE if it occurs positively in the clause to which it is matched, and FALSE otherwise (the value of variables which are not matched to any clause is arbitrary).

Following Franco and Van Gelder [6] we call a CNF formula *matched* if its incidence graph $I(F)$ has a matching that covers all clauses. For example, the graph in Figure 1 has such matching (indicated by dashed lines); thus the formula in (1) is satisfiable. Since a matching of maximum size can be found in polynomial time (see, e.g., [13]), we can decide efficiently whether a given CNF formula is matched.

Matched CNF formulas can be characterized in terms of a graph theoretic notion as observed by Franco and Van Gelder [6]: The *deficiency* of a CNF formula F on m clauses and n variables is defined by $\delta(F) = m - n$; the *maximum deficiency* $\delta^*(F)$ is the number of clauses which remain exposed by a maximum matching of $I(F)$. Thus F is matched if and only if $\delta^*(F) = 0$. It follows by Hall's Theorem that $\delta^*(F)$ equals the maximum $\delta(F')$ taken over all CNF formulas F' which arise by removing any number of clauses (possibly none or all) from F (see [13], Theorem 1.3.1).

Fleischer et al. [5] show that satisfiability of CNF formulas whose maximum deficiency is bounded by a fixed integer can be decided in polynomial time. The algorithm of [5] is based on the fact that a CNF

formula F is satisfiable if and only if there is a truth assignment t to at most $\delta^*(F)$ variables such that the CNF formula F_t is matched, where F_t results from F by the usual reduction w.r.t. the assignment t , i.e., by removing clauses satisfied by t and literals which are set to FALSE by t . Recently, the quoted result of Fleischner et al. could be improved: Szeider [17] presents an algorithm which decides satisfiability of formulas with n variables and maximum deficiency k in time $\mathcal{O}(2^k \cdot n^3)$; hence SAT for instances of bounded maximum deficiency is “fixed-parameter tractable” in the sense of Downey and Fellows [3].

Kullmann [12, 11] develops a framework for the systematic simplification of CNF formulas by means of certain classes of partial assignments (“autarkies”). The concept of matched CNF formulas provides the basis for an important class of such assignments.

It would be desirable to find suitable generalizations of matched CNF formulas to extend the quoted results.

In Section 3 we generalize matched CNF formulas, covering incidence graphs by disjoint bicliques (complete bipartite graphs). This approach yields a class of satisfiable CNF formulas (“biclique satisfiable CNF formulas”) which properly contains all matched CNF formulas (a matching is nothing but a collection of disjoint bicliques on two vertices each). We show, however, that recognition of biclique satisfiable CNF formulas is NP-complete (and remains NP-hard if the size of bicliques is bounded from above). We present a reduction from graph 3-colorability in Section 5.

The question rises whether “covering by bicliques” is a too specific concept, since a more general approach might yield a larger class of satisfiable CNF formulas which could possibly be recognized efficiently. To tackle this question, we consider in Section 4 the class $\text{Var-SAT} \cap \text{CNF}$ as studied by Kleine Büning and Zhao [10]—a CNF formula belongs to $\text{Var-SAT} \cap \text{CNF}$ if and only if it is satisfiable and remains satisfiable under arbitrary replacement of literals by their complements. We also term formulas in $\text{Var-SAT} \cap \text{CNF}$ to be *var-satisfiable* (in earlier versions of this paper, not being aware of [10], we called such formulas “graph satisfiable”).

Note that a CNF formula F is var-satisfiable if and only if every CNF formula F' with $I(F') = I(F)$ is satisfiable; thus satisfiability of such F is determined solely by its incidence graph. Clearly, if a CNF formula is biclique satisfiable (or matched), then it is also graph satisfiable, but the converse does not hold.

Kleine Büning and Zhao [10] show that var-satisfiable 2-CNF formulas can be recognized in linear time, and they ask whether there are polynomial time algorithms for var-satisfiable CNF formulas in general. We answer this question negatively and pinpoint the exact complexity

of $\text{Var-SAT} \cap \text{CNF}$ by showing its Π_2^{P} -completeness ($\Pi_2^{\text{P}} = \text{co-NP}^{\text{NP}}$ is a complexity class in the second level of the *Polynomial Hierarchy*, see [14, 16]). Thus var-satisfiable CNF formulas are apparently harder to recognize than biclique satisfiable ones. We obtain this completeness result in two steps: first we show Π_2^{P} -completeness of an enhanced graph coloring problem (given a graph G and a set S of vertices of G ; can every 2-coloring of the vertices in S be extended to a proper 3-coloring of all vertices of G ?). Second we reduce this enhanced graph coloring problem to the recognition problem for $\text{Var-SAT} \cap \text{CNF}$.

Recognition of $\text{Var-SAT} \cap \text{CNF}$ and the enhanced graph coloring problem contribute to the (relatively short) list of natural problems which are known to be Π_2^{P} -complete [15]. The following diagram visualizes our main results (the inclusions are proper).

$$\begin{array}{ccccc} \text{matched} & \subseteq & \text{biclique satisfiable} & \subseteq & \text{var-satisfiable} \\ (\text{polynomial time}) & & (\text{NP-complete}) & & (\Pi_2^{\text{P}}\text{-complete}) \end{array}$$

Both completeness results hold even for 3-CNF formulas.

The remainder of this paper is organized as follows. In the next section we present our basic terminology on formulas, assignments and graphs. In Sections 3 and 4 we introduce the concepts of biclique satisfiable and var-satisfiable CNF formulas, respectively, and state our main results. The more technical NP-completeness and Π_2^{P} -completeness proofs are presented in Sections 5 and 6, respectively.

2. Notation and basic definitions

We assume an infinite supply of (propositional) *variables*. A *literal* ℓ is a variable x or a negated variable \bar{x} ; if $\ell = \bar{x}$ then we put $\bar{\ell} = x$. A *clause* C is a finite disjunction of literals $\ell_1 \vee \dots \vee \ell_k$; k is the *width* of C . A *CNF formula* F is a finite conjunction of clauses $C_1 \wedge \dots \wedge C_n$; if all clauses of F have width $\leq k$ then F is a *k-CNF formula*. We allow the empty clause \perp and the empty CNF formula \top ; we also allow multiple occurrences of literals in a clause and multiple occurrences of clauses in a CNF formula.

A *truth assignment* is a map $t : X \rightarrow \{\text{TRUE}, \text{FALSE}\}$ defined on a finite set X of variables; t is *total* for a CNF formula F if all variables of F belong to X . For a variable $x \in X$ we put $t(\bar{x}) := \text{TRUE}$ if $t(x) = \text{FALSE}$, and $t(\bar{x}) := \text{FALSE}$ if $t(x) = \text{TRUE}$. We say that t *satisfies* a clause C if $t(\ell) = \text{TRUE}$ for some literal of C , and we say that t *satisfies* a CNF formula F if t satisfies all clauses of F . Thus no truth assignment satisfies \perp , although every truth assignment satisfies \top .

A CNF formula F is called *satisfiable* if it is satisfied by some truth assignment; otherwise F is *unsatisfiable*. F is *not-all-equal satisfiable* (*nae-satisfiable*, for short) if it is satisfied by some truth assignment t such that every clause of F contains a literal ℓ with $t(\ell) = \text{FALSE}$; in this case we say that t *nae-satisfies* F . Occasionally we will also consider so-called *quantified boolean formulas* (*QBFs*, for short), propositional formulas with a prefix of quantifiers. For example, if F is a CNF formula and X, Y is a bipartition of the set variables of F , then the quantified boolean formula $\forall X \exists Y \cdot F$ is called *satisfiable* if every possible way of setting the variables in X to TRUE or FALSE can be extended to a satisfying truth assignment of F . For detailed definitions of QBFs we refer the reader to [9, 14].

All graphs considered are finite and simple. We denote the vertex set and the edge set of a graph G by $V(G)$ and $E(G)$, respectively. The *degree* $d_G(v)$ of a vertex $v \in V(G)$ is the number of incident edges. If G is bipartite, then we assume that sets $V_1(G), V_2(G) \subseteq V(G)$ give a fixed bipartition. Further graph theoretic terminology can be found in [2].

The *incidence graph* of a CNF formula F is the bipartite graph $I(F)$ defined as follows: $V_1(I(F))$ consists of the variables of F and $V_2(I(F))$ consists of the clauses of F ; a variable x and a clause C are adjacent if and only if x occurs (positively or negatively) in C .

3. Covering by bicliques

A single edge can be considered as a biclique (i.e., a complete bipartite graph) on two vertices, and so we ask whether we can use a collection of bicliques instead of a matching to certify satisfiability. The following observation gives a reason for such approach.

LEMMA 1. *Let F be a CNF formula on n variables and m clauses. If $I(F)$ is a biclique and $m < 2^n$, then F is satisfiable.*

Proof. Every clause of F has exactly n literals; hence every clause of F discards exactly one of the 2^n truth assignments. \square

For example, the CNF formula $(p \vee q) \wedge (p \vee \bar{q}) \wedge (\bar{p} \vee q)$ is satisfiable by means of Lemma 1 (however, it is not matched).

The next idea is that we can certify satisfiability by using several bicliques to cover all clauses in $I(F)$, as long as every variable belongs to at most one biclique.

LEMMA 2. A CNF formula F is satisfiable if there is a collection $\mathcal{C} = \{X_1, \dots, X_k\}$ of sets of vertices of $I(F)$ such that

- (i) every clause of F belongs to some X_i , $1 \leq i \leq k$,
- (ii) every variable of F belongs to at most one X_i , $1 \leq i \leq k$,
- (iii) X_i induces in $I(F)$ a biclique on n_i variables and $m_i < 2^{n_i}$ clauses, $1 \leq i \leq k$.

Proof. By means of Lemma 1, we find for every $i = 1, \dots, k$ some truth assignment t_i to the variables in X_i such that all clauses in X_i are satisfied. Since every variable of F belongs to at most one X_i , we can combine t_1, \dots, t_k consistently to a satisfying truth assignment of F . \square

We call a CNF formula *biclique satisfiable* if it fulfills the conditions stated in Lemma 2. See Figure 3 for an example, where bicliques are indicated by different types of dashed lines.

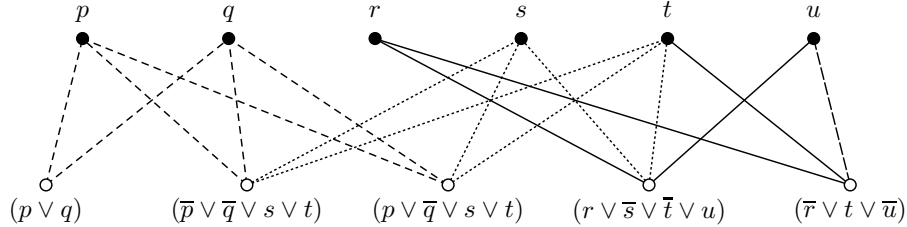


Figure 2. Graph of a biclique satisfiable CNF formula.

In Lemma 2 we only require that different elements of \mathcal{C} do not share variables; however nothing is lost if we require disjointness:

LEMMA 3. A CNF formula F is biclique satisfiable if and only if there is a collection \mathcal{C} which satisfies the conditions of Lemma 2 such that all elements of \mathcal{C} are mutually disjoint.

Proof. If $\mathcal{C} = \{X_1, \dots, X_k\}$ satisfies the conditions of Lemma 2 (w.r.t. some CNF formula F), then putting $X'_i := X_i \setminus \bigcup_{1 \leq j < i} X_j$ yields a collection $\mathcal{C}' \subseteq \{X'_1, \dots, X'_k\}$ which satisfies the conditions of Lemma 2 as well; however, elements of \mathcal{C}' are mutually disjoint. \square

The question rises whether biclique satisfiable CNF formulas can be recognized efficiently. Unfortunately, our results of Section 5 show the following.

THEOREM 1. Recognition of biclique satisfiable 3-CNF formulas is NP-complete.

Are there any restrictions of the general concept of biclique satisfiability which admit efficient recognition? A natural candidate for such approach is to restrict the size of bicliques used in the cover by some positive integer $k \geq 1$. More formally, we say that a CNF formula F is *k-biclique satisfiable* if there is a collection \mathcal{C} which satisfies the conditions of Lemma 2, and each X_i contains at most k variables. The following is obvious.

LEMMA 4. *A CNF formula is matched if and only if it is 1-biclique satisfiable. Hence 1-biclique satisfiable CNF formulas can be recognized in polynomial time.*

Our results of Section 5 yield that recognition of k -biclique satisfiable CNF formulas becomes intractable as soon as we go beyond matched CNF formulas:

THEOREM 2. *For any $k \geq 2$, recognition of k -biclique satisfiable 3-CNF formulas is NP-complete.*

4. Var-satisfiable CNF formulas

We call a CNF formula F *var-satisfiable* if every CNF formula F' with $I(F') = I(F)$ is satisfiable. Since a CNF formula is var-satisfiable if and only if it is satisfiable and remains satisfiable under arbitrary replacement of literals by their complements, the class of var-satisfiable CNF formulas is nothing but the class $\text{Var-SAT} \cap \text{CNF}$ considered by Kleine Büning and Zhao [10].

Clearly matched and biclique satisfiable CNF formulas are var-satisfiable, since polarity of literals is not taken into account in forming incidence graphs. However, there are var-satisfiable CNF formulas which are not biclique satisfiable; take, for example,

$$F = (p \vee q \vee r) \wedge (\bar{p} \vee q \vee r) \wedge (\bar{p} \vee q \vee \bar{r}) \wedge (p \vee \bar{q}) \wedge (\bar{p} \vee \bar{q}).$$

(a similar argument as used in the proof of Lemma 1 shows that F is var-satisfiable.)

The following result is due to Kleine Büning and Zhao [10].

THEOREM 3. *Var-satisfiable 2-CNF formulas can be recognized in linear time.*

Kleine Büning and Zhao [10] ask whether a polynomial time algorithm for recognition of var-satisfiable CNF formulas with clause widths ≥ 3 exists. We settle this question as follows (for a proof, see Section 6).

THEOREM 4. *Recognition of var-satisfiable 3-CNF formulas is Π_2^P -complete.*

We recall that the complexity class Π_2^P belongs to the second level of the Polynomial Hierarchy [16]; Π_2^P contains the complements of all properties which can be checked in nondeterministic polynomial time by an NP-oracle (or, equivalently, by a co-NP-oracle).

It is easy to see that $\text{Var-SAT} \cap \text{CNF}$ belongs to Π_2^P : if a given CNF formula F is not var-satisfiable, then we can guess an unsatisfiable CNF formula F' with $I(F) = I(F')$; unsatisfiability of F' can be checked by a co-NP-oracle (for an alternate proof using quantified boolean formulas, see Section 6).

For showing that $\text{Var-SAT} \cap \text{CNF}$ is Π_2^P -complete, we will use a reduction from an enhanced 3-colorability problem (see Section 6).

5. NP-completeness of biclique satisfiability

We will obtain the quoted NP- and Π_2^P -completeness results in a graph theoretic framework. Let G be a bipartite graph. A *biclique cover* \mathcal{C} of G is a set $\mathcal{C} = \{X_1, \dots, X_r\}$ of mutually disjoint subsets of $V(G)$ such that each X_i induces in G a biclique and

$$1 \leq |V_2(G) \cap X_i| < 2^{|V_1(G) \cap X_i|}.$$

If, additionally, $|V_2(G) \cap X_i| \leq q$ for $i = 1, \dots, r$, then we call \mathcal{C} a q -biclique cover. A vertex v is *covered* by \mathcal{C} if v is contained in $\bigcup_{i=1}^r X_i$; otherwise v is *exposed*. A biclique cover is *total* if it covers all vertices in $V_2(G)$ (vertices in $V_1(G)$ may remain exposed). In view of Lemma 3 it follows that a CNF formula F is biclique satisfiable (q -biclique satisfiable) if and only if its incidence graph $I(F)$ has a total biclique cover (q -biclique cover, respectively).

Let \mathcal{C} be a biclique cover of G and G' a subgraph of G . Then \mathcal{C} induces a biclique cover

$$\mathcal{C} \upharpoonright G' := \{ X \in \mathcal{C} \mid X \subseteq V(G') \}$$

of G' . Observe that it is possible that \mathcal{C} is a total biclique cover of G , but $\mathcal{C} \upharpoonright G'$ is not a total biclique cover of G' .

Let k be a positive integer. A k -coloring of a graph G is a map $\chi : V(G) \rightarrow \{1, \dots, k\}$. A k -coloring is *proper* if $\chi(v) \neq \chi(w)$ holds for every edge $vw \in E(G)$. Furthermore, G is called k -colorable if G has some proper k -coloring.

It is well-known that the problem of whether a given graph is 3-colorable is NP-complete, see e.g., [7]. Evidently the problem remains NP-

hard for graphs with minimum degree 2. We will reduce this problem to the problem of whether a given bipartite graph has a total biclique cover. We note that the latter problem has some similarity to “graph packing” problems studied elsewhere (see Hell [8] for a recent survey), but is apparently not entailed by known results from this area.

For the reduction, we will use the following two types of gadgets.

A *color gadget* is a bipartite graph H_c with $V_1(H_c) = \{x_1, x_2\}$ and $V_2(H_c) = \{y_1, y_2, y_3, w\}$ as depicted in Figure 3. We call the vertices y_i *outputs* of H_c , and we will also use the notation $\text{out}_i(H_c) := y_i$, $i = 1, 2, 3$.

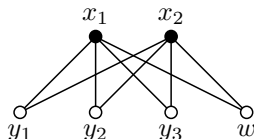


Figure 3. Color gadget H_c .

Evidently, a color gadget H_c has no total biclique cover. However, for every $i \in \{1, 2, 3\}$ there is a (unique) biclique cover \mathcal{C}_i of H_c such that y_i is the only vertex which remains exposed (e.g., $\mathcal{C}_1 = \{\{x_1, x_2, y_2, y_3, w\}\}$).

Let $q \geq 2$ be an integer. A *fan-out gadget* is a bipartite graph H_f with $V_1(H_f) = \{a_i, b_i, z_i \mid i = 1, \dots, q-1\} \cup \{z_q\}$ and $V_2(H_f) = \{c_i, d_i, e_i \mid i = 1, \dots, q-1\}$ as depicted in Figure 4(a) for $q = 2$ and in Figure 4(b) for $q > 2$. We call the vertex $\text{in}(H_f) := a_1$ the *input* of H_f , and the vertices in $\text{out}(H_f) := \{z_1, \dots, z_q\}$ the *outputs* of H_f .

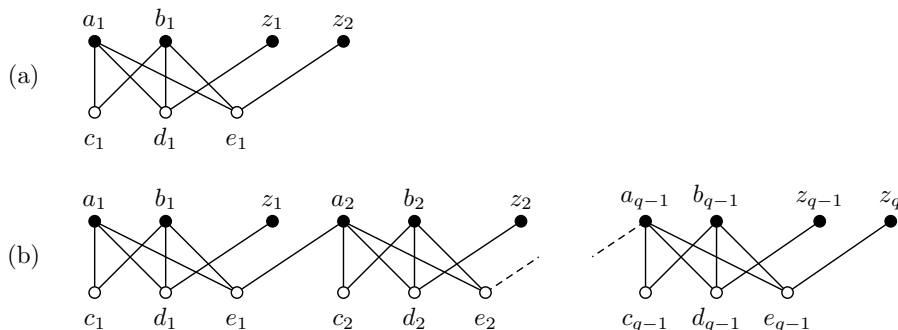


Figure 4. Fan-out gadgets H_f .

LEMMA 5. *A fan-out gadget H_f has a total biclique cover which exposes all outputs. There is exactly one total biclique cover of H_f which exposes the input; such cover, however, covers all outputs.*

Proof. Consider a fan-out gadget H_f with q outputs. Evidently

$$\{ \{a_i, b_i, c_i, d_i, e_i\} \mid i = 1, \dots, q-1 \}$$

is a total biclique cover of H_f which exposes all outputs. For the second part of the lemma, we construct a total biclique cover \mathcal{C} inductively as follows. Since c_1 must be covered without using the input a_1 , we have no choice but using $\{b_1, c_1\}$ to cover c_1 . For $q = 2$ we must use $\{z_1, d_1\}$ and $\{z_2, e_1\}$ to cover d_1 and e_1 , respectively. For $q > 2$, we must use $\{e_1, a_2\}$ to cover e_1 , and as above, $\{b_2, c_2\}$ to cover c_2 , and so forth. The biclique cover \mathcal{C} defined by this process is total and covers all outputs of H_f . \square

Consider any finite graph G with minimum degree 2. We are going to construct a bipartite graph H_G from G . For each vertex v of G we will construct a graph $H(v)$; then we will put all these graphs together to form H_G .

Let v be any vertex of G . We take a color gadget $H_c(v)$, and fan-out gadgets $H_f(v, j)$, $j = 1, 2, 3$, with $d_G(v)$ outputs each. For $j = 1, 2, 3$ we connect $H_c(v)$ with $H_f(v, j)$ by adding an edge between $\text{out}_j(H_c(v))$ and $\text{in}(H_f(v, j))$. We denote the obtained graph by $H(v)$.

Let

$$Z := \bigcup_{1 \leq j \leq 3, v \in V(G)} \text{out}(H_f(v, j)).$$

We construct a set \mathcal{P} of pairs $\{z, z'\} \subseteq Z$ such that

- (i) for every $\{z, z'\} \in \mathcal{P}$ there is some $j \in \{1, 2, 3\}$ and an edge $vw \in E(G)$ such that $z \in \text{out}(H_f(v, j))$ and $z' \in \text{out}(H_f(w, j))$,
- (ii) every $z \in Z$ belongs to exactly one pair in \mathcal{P} .

Such \mathcal{P} exists, since by construction each $H_f(j, v)$ has $d_G(v)$ outputs; moreover, \mathcal{P} can be constructed efficiently by a greedy approach. Finally, we obtain a bipartite graph H_G from the disjoint union of all $H(v)$, $v \in V(G)$, by adding for each pair $\{z, z'\} \in \mathcal{P}$ a new vertex $p_{z, z'} \in V_2(H_G)$ and the edges $zp_{z, z'}$ and $z'p_{z, z'}$.

Note that $d_{H_G}(v) \leq 3$ for all $v \in V_2(H_G)$; thus any CNF formula whose incidence graph is H_G is a 3-CNF formula.

LEMMA 6. *For every graph G with minimum degree 2 the following statements are equivalent.*

- (i) G is 3-colorable;
- (ii) H_G has a total 2-biclique cover;
- (iii) H_G has a total biclique cover.

Proof. (i) \Rightarrow (ii). Let χ be a proper 3-coloring of G . We define for each $v \in V(G)$ a biclique cover $\mathcal{C}(v)$ of $H(v)$ as follows. Consider $v \in V(G)$ and put $j := \chi(v)$. For the color gadget $H_c(v)$, we take the unique biclique cover \mathcal{C}_j which only exposes $\text{out}_j(H_c(v))$. To cover $\text{out}_j(H_c(v))$, we have no choice but using the set $X := \{\text{out}_j(H_c(v)), \text{in}(H_f(v, j))\}$. For the fan-out gadgets $H_f(v, i)$, $i = 1, 2, 3$, we define total biclique covers $\mathcal{C}^{(i)}$ according to Lemma 5 as follows. For $i \neq j$ we choose $\mathcal{C}^{(i)}$ such that all outputs of $H_f(v, i)$ are exposed; $\mathcal{C}^{(j)}$, however, must expose $\text{in}(H_f(v, j))$, since $\text{in}(H_f(v, j))$ is already contained in X . Therefore we must take for $\mathcal{C}^{(j)}$ the unique biclique cover which covers all outputs of $H_f(v, j)$; cf. Lemma 5. Thus we have defined a total biclique cover

$$\mathcal{C}(v) := \mathcal{C}_j \cup \{X\} \cup \mathcal{C}^{(1)} \cup \mathcal{C}^{(2)} \cup \mathcal{C}^{(3)} \quad (2)$$

of $H(v)$. By construction, $\mathcal{C}(v)$ exposes all outputs of $H_f(v, i)$ for $i \neq j = \chi(v)$. Consequently, $\bigcup_{v \in V(G)} \mathcal{C}(v)$ is a biclique cover of H_G which covers all vertices in

$$\bigcup_{v \in V(G)} V_2(H(v)) = V_2(H_G) \setminus \{p_{z, z'} \mid \{z, z'\} \in \mathcal{P}\}.$$

Now consider a vertex $p_{z, z'}$ for some pair $\{z, z'\} \in \mathcal{P}$. By definition of H_G , there is an edge $vw \in E(G)$ and some $i \in \{1, 2, 3\}$ such that $z \in \text{out}(H_f(v, i))$ and $z' \in \text{out}(H_f(w, i))$. Since χ is a proper coloring, $\chi(v) \neq \chi(w)$ follows of necessity. Hence $i \neq \chi(v)$ or $i \neq \chi(w)$; w.l.o.g., we assume $i \neq \chi(v)$. It follows that $\mathcal{C}(v)$ exposes the outputs of $H_f(v, i)$, and so we can use $X_{z, z'} := \{z, p_{z, z'}\}$ to cover $p_{z, z'}$. Whence

$$\mathcal{C} := \{X_{z, z'} \mid \{z, z'\} \in \mathcal{P}\} \cup \bigcup_{v \in V(G)} \mathcal{C}(v)$$

is a total biclique cover of H_G . By construction, $|X \cap V_1(H_G)| \leq 2$ for all $X \in \mathcal{C}$.

(ii) \Rightarrow (iii). Trivial.

(iii) \Rightarrow (i). Let \mathcal{C} be a total biclique cover of H_G . For every $v \in V(G)$, the biclique cover $\mathcal{C} \upharpoonright H_c(v)$ exposes at least one output of $H_c(v)$. Hence we can define 3-coloring of G by choosing $\chi(v)$ such that $\text{out}_{\chi(v)}(H_c(v))$

is exposed by $\mathcal{C} \upharpoonright H_c(v)$ for all $v \in V(G)$. We show that χ is a proper coloring. Suppose to the contrary that $\chi(v) = \chi(w) = j$ for some edge $vw \in E(G)$. Since $\text{out}_j(H_c(v))$ is exposed by $\mathcal{C} \upharpoonright H_c(v)$, it follows that $\{\text{out}_j(H_c(v)), \text{in}(H_f(v, j))\} \in \mathcal{C}$; consequently, by Lemma 5, all outputs of $H_f(v, j)$ are covered by $\mathcal{C} \upharpoonright H_f(v, j)$. The same argument shows that all outputs of $H_f(w, j)$ are covered by $\mathcal{C} \upharpoonright H_f(w, j)$. By construction of H_G , there is a vertex $p_{z, z'} \in V_2(H_G)$ of degree 2 which is adjacent with some output z of $H_f(v, j)$ and some output z' of $H_f(w, j)$. Since z and z' are covered by $\mathcal{C} \upharpoonright H_f(v, j)$ and $\mathcal{C} \upharpoonright H_f(w, j)$, respectively, it follows that $p_{z, z'}$ is exposed by \mathcal{C} , a contradiction, as \mathcal{C} is assumed to be a total biclique cover. \square

Clearly the problem of whether a bipartite graph has a total biclique cover is in NP. Moreover, the construction of H_G can be carried out in polynomial time w.r.t. the size of G . Hence Lemma 6 implies the following.

THEOREM 5. *Recognition of bipartite graphs having a total biclique cover is NP-complete. This problem remains NP-hard if we consider only 2-biclique covers of bipartite graphs H with $d_H(v) \leq 3$ for all $v \in V_2(H)$.*

Thus Theorems 1 and 2 are established.

6. Π_2^P -completeness of $\text{Var-SAT} \cap \text{CNF}$

This final section is devoted to a proof of Theorem 4.

Let G be a graph and $S \subseteq V(G)$. We say that G is *S -extending 3-colorable* if every coloring $\chi_S : S \rightarrow \{1, 2\}$ can be extended to a proper 3-coloring of G . (If S contains a pair of adjacent vertices, then G is evidently not S -extending 3-colorable; hence it suffices to consider S independent.)

THEOREM 6. *The problem of whether a graph G with $S \subseteq V(G)$ is S -extending 3-colorable is Π_2^P -complete.*

Proof. The problem belongs to Π_2^P , since if G is not S -extending 3-colorable, then we can guess a 2-coloring χ_S of S and check by a co-NP-oracle that χ_S cannot be extended to a proper 3-coloring of G .

For Π_2^P -hardness we use a reduction from the following problem which is shown Π_2^P -complete in [4].

Given a quantified boolean formula $\Phi = \forall Y \exists X \cdot F$ (i.e., X, Y are finite disjoint sets of variables such that every variable of F belongs

to $X \cup Y$) where F is a CNF formula with exactly three literals in every clause; is Φ nae-satisfiable? (i.e., every truth assignment to Y can be extended to a truth assignment to $X \cup Y$ which nae-satisfies F .)

Consider $\Phi = \forall Y \exists X \cdot F$ as above. We obtain from F a graph G by a standard transformation used elsewhere to reduce nae-satisfiability from 3-colorability (see, e.g., Theorem 9.8 of [14]). For each variable z of F we take two vertices v_z and $v_{\bar{z}}$, and for each clause $C = (l_1 \vee l_2 \vee l_3)$ of F we take three vertices $v_{C,1}, v_{C,2}, v_{C,3}$; finally, we take one more vertex a . We add edges such that $v_z, v_{\bar{z}}, a$ forms a triangle for every variable z , and $v_{C,1}, v_{C,2}, v_{C,3}$ forms a triangle for every clause C . Finally, for every clause $C = (l_1 \vee l_2 \vee l_3)$ and $i = 1, 2, 3$, we join $v_{C,i}$ and v_{l_i} by an edge (see Figure 5 for an illustration).

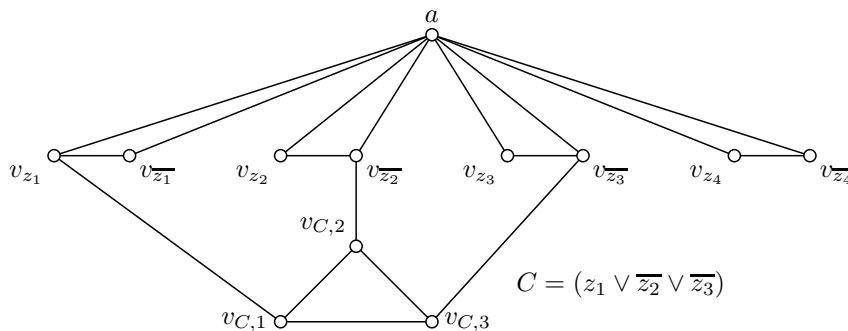


Figure 5. Reduction from nae-satisfiability.

F is nae-satisfiable if and only if G is 3-colorable (see [14, Theorem 9.8]). In fact, every nae-satisfying total assignment of F corresponds to a proper 3-coloring of G such that vertex v_l has color 1 if and only if l is set TRUE by the corresponding assignment (fixing the color of vertex a to 3.)

Putting $S := \{ v_y \mid y \in Y \}$, we conclude that Φ is nae-satisfiable if and only if every coloring $\chi_S : S \rightarrow \{1, 2\}$ can be extended to a proper 3-coloring of G . \square

At this juncture we give an outline for a polynomial-time transformation which reduces recognition of $\text{Var-SAT} \cap \text{CNF}$ to satisfiability of quantified boolean formulas with $\forall \exists$ prefix; since the latter problem is well-known to be Π_2^P -complete (see, e.g., [14]), this transformation yields $\text{Var-SAT} \cap \text{CNF} \in \Pi_2^P$ (cf. the discussion in Section 4). Consider a CNF formula $F = C_1 \wedge \dots \wedge C_m$ and assume, w.l.o.g., that all literals occurring in F are positive (for, deciding var-satisfiability, polarity of literals is inessential). Let X denote the set of variables of F . Consider

some (positive) occurrence of a variable x in a clause C_i of F . We obtain from F a CNF formula F' by replacing x in C_i by a new variable y , and by addition of the clauses $D_1 = (\bar{x} \vee y \vee z)$, $D_2 = (x \vee \bar{y} \vee z)$, $D_3 = (x \vee y \vee \bar{z})$, $D_4 = (\bar{x} \vee \bar{y} \vee \bar{z})$; z is a new variable as well. For any truth assignment t which satisfies D_1, \dots, D_4 , $t(z) = \text{TRUE}$ implies $t(x) \neq t(y)$, and $t(z) = \text{FALSE}$ implies $t(x) = t(y)$. Hence the quantified boolean formula $\forall z \exists (X \cup \{y\}) \cdot F'$ is satisfiable if and only if F is satisfiable and remains satisfiable under replacing x in C_i by \bar{x} . Applying this construction for all variable occurrences of F we end up with a quantified boolean formula $\Phi_F = \forall Z \exists (X \cup Y) \cdot F''$; Φ_F is satisfiable if and only if F is var-satisfiable. Since the construction of Φ_F can be carried out in polynomial time, we have in fact reduced recognition of $\text{Var-SAT} \cap \text{CNF}$ to satisfiability of QBFs with $\forall \exists$ prefix.

For showing Π_2^P -completeness of $\text{Var-SAT} \cap \text{CNF}$ we use a similar construction as in Section 5. Let G be a graph with minimum degree 2 and $S \subseteq V(G)$. We construct a graph $H_{G,S}$ as above, except for vertices in S we use different gadgets instead of color gadgets (thus $H_{G,\emptyset}$ is nothing but H_G). A *precolor gadget* is a bipartite graph H_p with $V_1(H_p) = \{x_1, x_2\}$ and $V_2(H_p) = \{y_1, y_2, w_1, w_2\}$ as depicted in Figure 6. Similarly as for color gadgets, we call $\text{out}_i(H_p) = y_i$, $i = 1, 2$, the *outputs* of H_p .

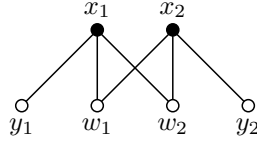


Figure 6. Precolor gadget H_p .

Let G be a graph, $S \subseteq V(G)$, and $H_{G,S}$ the bipartite graph as defined above. Consider $v \in S$ and the precolor gadget $H_p(v)$. For $i \in \{1, 2\}$ we write $f_{v,i}(H_{G,S})$ for the result of removing all vertices of $H_p(v)$ from $H_{G,S}$ except the vertex $y_i = \text{out}_i(H_p(v))$; i.e., $f_{v,i}(H_{G,S}) := H_{G,S} - (V(H_p(v)) \setminus \{y_i\})$. If F is a CNF formula with $I(F) = H_{G,S}$, then we can define $F' = f_{v,i}(F)$ as the CNF formula which emerges from F by the corresponding modification yielding $I(F') = f_{i,v}(H_{G,S})$.

Moreover, we write $H_{G,S}^*$ for the result of applying such modification for all $v \in S$; we call $H_{G,S}^*$ a *gadget reduction* of $H_{G,S}$. Since for each $v \in S$ we have two possibilities for forming $f_{i,v}(H_{G,S})$, it follows that $H_{G,S}$ has $2^{|S|}$ different gadget reductions.

LEMMA 7. *Let F be a CNF formula with $I(F) = H_{G,S}$, and let $v \in S$. If both $f_{v,1}(F)$ and $f_{v,2}(F)$ are satisfiable, then F is satisfiable. Conse-*

quently, if for every gadget reduction $H_{G,S}^*$ of $H_{G,S}$ all CNF formulas F^* with $I(F^*) = H_{G,S}^*$ are satisfiable, then F is satisfiable.

Proof. Using the notation $x^1 := x$, $x^0 := \bar{x}$, we can write the clauses of F which correspond to $H_p(v)$ as

$$(\ell_1 \vee x_1^{\alpha_1}) \wedge (x_1^{\alpha_2} \vee x_2^{\beta_3}) \wedge (x_1^{\alpha_3} \vee x_2^{\beta_2}) \wedge (x_2^{\beta_1} \vee \ell_2)$$

where $\alpha_i, \beta_i \in \{0, 1\}$. Consider the CNF formulas

$$\begin{aligned} F_1 &:= (x_1^{\alpha_2} \vee x_2^{\beta_3}) \wedge (x_1^{\alpha_3} \vee x_2^{\beta_2}) \wedge (x_2^{\beta_1}), \\ F_2 &:= (x_1^{\alpha_1}) \wedge (x_1^{\alpha_2} \vee x_2^{\beta_3}) \wedge (x_1^{\alpha_3} \vee x_2^{\beta_2}). \end{aligned}$$

F_1 is unsatisfiable if and only if $\alpha_2 \neq \alpha_3$ and $\beta_1 \neq \beta_2 = \beta_3$. Symmetrically, F_2 is unsatisfiable if and only if $\beta_2 \neq \beta_3$ and $\alpha_1 \neq \alpha_2 = \alpha_3$. Since these two cases exclude each other, it follows that we can choose $i \in \{1, 2\}$ such that F_i is satisfiable. By assumption, $f_{i,v}(F)$ is satisfiable. Since F_i and $f_{i,v}(F)$ do not share any variable, we can combine satisfying truth assignments of F_i and $f_{i,v}(F)$, respectively, to a satisfying truth assignment of F . Hence the first part of the lemma is shown true. The second part follows by repeated application of the first part. \square

LEMMA 8. *Let G be a graph with minimum degree 2, $S \subseteq V(G)$ an independent set of vertices, and let $H_{G,S}$ be the corresponding bipartite graph as defined above. Then G is S -extending 3-colorable if and only if every CNF formula F with $I(F) = H_{G,S}$ is satisfiable.*

Proof. (\Rightarrow). Assume that G is S -extending 3-colorable, and choose some gadget reduction $H_{G,S}^*$ of $H_{G,S}$ arbitrarily. We define a coloring $\chi_S : S \rightarrow \{1, 2\}$ by setting $\chi_S(v) = i$ if and only if $f_{v,i}$ was used by obtaining $H_{G,S}^*$ from $H_{G,S}$. Since G is S -extending 3-colorable, there is a proper 3-coloring χ of G which extends χ_S . We are going to define a total biclique cover of $H_{G,S}^*$.

Using the same notation $\mathcal{C}^{(i)}$ as in the first part of the proof of Lemma 6 and letting $X := \{\text{out}_{\chi(v)}(H_p(v)), \text{in}(H_f(v, \chi(v)))\}$, we define for vertices $v \in S$ a cover

$$\mathcal{C}'(v) := \{X\} \cup \mathcal{C}^{(1)} \cup \mathcal{C}^{(2)}.$$

For the remaining vertices $v \in V(G) \setminus S$ we define $\mathcal{C}'(v) = \mathcal{C}(v)$ as in (2) of the proof of Lemma 6; the same argument as used there shows that

$$\mathcal{C}' := \{X_{z,z'} \mid \{z, z'\} \in \mathcal{P}\} \cup \bigcup_{v \in V(G)} \mathcal{C}'(v)$$

is a total biclique cover of $H_{G,S}^*$. Therefore, every CNF formula F^* with $I(F^*) = H_{G,S}^*$ is satisfiable. Since the gadget reduction $H_{G,S}^*$ was chosen arbitrarily, it follows now by Lemma 7 that every CNF formula F with $I(F) = H_{G,S}$ is satisfiable.

(\Leftarrow). Assume that every CNF formula F with $I(F) = H_{G,S}$ is satisfiable, and choose an arbitrary coloring $\chi_S : S \rightarrow \{1, 2\}$. We show that χ_S can be extended to a proper 3-coloring of G . We define a CNF formula F with $I(F) = H_{G,S}$ as follows (naming of variables is chosen w.r.t. Figures 3, 4, and 6).

- For every precolor gadget $H_p(v)$, $v \in S$, we take clauses

$$(\overline{x_1} \vee a^{(1)}), (x_1 \vee x_2), (x_1 \vee \overline{x_2}), (x_2 \vee a^{(2)}) \quad \text{if } \chi_S(v) = 1 \quad (3)$$

$$(x_1 \vee a^{(1)}), (\overline{x_1} \vee x_2), (x_1 \vee x_2), (\overline{x_2} \vee a^{(2)}) \quad \text{if } \chi_S(v) = 2 \quad (4)$$

where $a^{(i)} = \text{in}(H_f(v, i))$, $i = 1, 2$.

- For every color gadget $H_c(v)$ of H_G we take clauses

$$(\overline{x_1} \vee \overline{x_2} \vee a^{(1)}), (\overline{x_1} \vee x_2 \vee a^{(2)}), (x_1 \vee \overline{x_2} \vee a^{(3)}), (x_1 \vee x_2), \quad (5)$$

where $a^{(i)} = \text{in}(H_f(v, i))$, $i = 1, 2, 3$.

- For every fan-out gadget $H_f(v, i)$ we take clauses of the form

$$\begin{aligned} &(\overline{a_1} \vee b_1), (\overline{a_1} \vee \overline{b_1} \vee z_1), (\overline{a_1} \vee \overline{b_1} \vee a_2), \\ &\quad \dots \\ &(\overline{a_{q-1}} \vee b_{q-1}), (\overline{a_{q-1}} \vee \overline{b_{q-1}} \vee z_{q-1}), (\overline{a_{q-1}} \vee \overline{b_{q-1}} \vee z_q) \end{aligned} \quad (6)$$

where $a_i = \text{in}(H_f(v, i))$ and $\{z_1, \dots, z_q\} = \text{out}(H_f(v, i))$.

- For every vertex $p_{z,z'} \in V(H_G)$ with $z \in \text{out}(H_f(v, j))$ and $z' \in \text{out}(H_f(w, j))$, we take the clause

$$(\overline{z} \vee \overline{z'}). \quad (7)$$

Since F is assumed to be var-satisfiable, there is a truth assignment t which satisfies F . Next we make some simple observations.

- Consider clauses defined in (3). It follows that $t(x_1) = \text{TRUE}$, otherwise one of the clauses $(x_1 \vee x_2)$, $(x_1 \vee \overline{x_2})$ remains unsatisfied. Consequently $t(a^{(1)}) = \text{TRUE}$. A symmetric argument shows $t(a^{(2)}) = \text{TRUE}$ for clauses defined in (4).
- Consider clauses defined in (5). Clearly $t(x_1) = \text{TRUE}$ or $t(x_2) = \text{TRUE}$ in order to satisfy $(x_1 \vee x_2)$; consequently $t(a^{(i)}) = \text{TRUE}$ for some $i \in \{1, 2, 3\}$.

- Consider clauses defined in (6). If $t(a_1) = \text{TRUE}$, then $t(b_1) = \text{FALSE}$ to satisfy $(\overline{a_1} \vee b_1)$; consequently $t(z_1) = \text{TRUE}$. A repeated application of this argument shows that $t(a_1) = \text{TRUE}$ implies $t(z_i) = \text{TRUE}$ for $i = 1, \dots, q$.

We extend χ_S to a 3-coloring of G choosing $\chi(v)$ for $v \in V(G) \setminus S$ as specified in the following table (x_1, x_2 are the vertices in $V_1(H_c(v))$).

$t(x_1)$	$t(x_2)$	$\chi(v)$
TRUE	TRUE	1
TRUE	FALSE	2
FALSE	TRUE	3

In view of the above observations, it follows that $t(z) = \text{TRUE}$ for all $z \in \text{out}(H_f(v, \chi(v)))$, $v \in V(G)$. It remains to show that χ is a proper 3-coloring. Suppose to the contrary that there is a pair of adjacent vertices $v, w \in V(G)$ with $\chi(v) = \chi(w) = j$. By construction of $H_{G,S}$, there is a vertex $p_{z,z'} \in V_2(H_{G,S})$ of degree 2 being incident with $z \in \text{out}(H_f(v, j))$ and $z' \in \text{out}(H_f(w, j))$; thus F contains the clause $(\overline{z} \vee \overline{z'})$ because of (7). However, $t(z) = t(z') = \text{TRUE}$, since $z \in \text{out}(H_f(v, \chi(v)))$ and $z' \in \text{out}(H_f(w, \chi(w)))$. Hence $(\overline{z} \vee \overline{z'})$ is not satisfied, a contradiction. Therefore we conclude that the coloring χ is a proper. Since the coloring χ_S of S was chosen arbitrarily, G is indeed S -extending 3-colorable. \square

The combination of Lemma 8 and Theorem 6 yields Π_2^P -hardness of var-satisfiability. Since the degree of vertices in $V_2(H_{G,S})$ is at most 3, Theorem 4 is shown true.

Acknowledgment

The idea to generalize matchings by collections of bicliques emerged through discussions with Oliver Kullmann at his research visit in Vienna, August 2001.

References

1. Aharoni, R. and N. Linial: 1986, ‘Minimal Non-Two-Colorable Hypergraphs and Minimal Unsatisfiable Formulas’. *J. Combin. Theory Ser. A* **43**, 196–204.
2. Diestel, R.: 2000, *Graph Theory*, Vol. 173 of *Graduate Texts in Mathematics*. New York: Springer Verlag, 2nd edition.
3. Downey, R. G. and M. R. Fellows: 1999, *Parameterized Complexity*. Springer Verlag.

4. Eiter, T. and G. Gottlob: 2000, 'Complexity results for some eigenvector problems'. *International Journal of Computer Mathematics* **76**(1), 59–74.
5. Fleischner, H., O. Kullmann, and S. Szeider: 2002, 'Polynomial-Time Recognition of Minimal Unsatisfiable Formulas with Fixed Clause-Variable Difference'. *Theoretical Computer Science* **289**(1), 503–516.
6. Franco, J. and A. Van Gelder: 2003, 'A Perspective on Certain Polynomial Time Solvable Classes of Satisfiability'. *Discrete Applied Mathematics* **125**, 177–214.
7. Garey, M. R. and D. R. Johnson: 1979, *Computers and Intractability*. W. H. Freeman and Company, New York.
8. Hell, P.: 2000, 'Graph Packings'. In: I. Rusu (ed.): *Proceedings for the 6th International Conference on Graph Theory*, Vol. 5 of *Electronic Notes in Discrete Mathematics*.
9. Kleine Büning, H. and T. Lettman: 1999, *Propositional logic: deduction and algorithms*. Cambridge University Press, Cambridge.
10. Kleine Büning, H. and X. Zhao: 2001, 'Satisfiable Formulas Closed Under Replacement'. In: H. Kautz and B. Selman (eds.): *Proceedings for the Workshop on Theory and Applications of Satisfiability*, Vol. 9 of *Electronic Notes in Discrete Mathematics*.
11. Kullmann, O.: 2000, 'Investigations on autark assignments'. *Discrete Applied Mathematics* **107**(1-3), 99–137.
12. Kullmann, O.: 2003, 'Lean clause-sets: Generalizations of minimally unsatisfiable clause-sets'. *Discrete Applied Mathematics* **130**(2), 209–249.
13. Lovász, L. and M. D. Plummer: 1986, *Matching Theory*, Vol. 29 of *Annals of Discrete Mathematics*. Amsterdam: North-Holland Publishing Co.
14. Papadimitriou, C. H.: 1994, *Computational Complexity*. Addison-Wesley.
15. Schaefer, M.: 2001, 'Completeness in the polynomial-time hierarchy'. Technical Report TR01-009, School of CTI, DePaul University, Chicago, IL.
16. Stockmeyer, L. J.: 1976, 'The polynomial-time hierarchy'. *Theoretical Computer Science* **3**(1), 1–22.
17. Szeider, S.: 2003, 'Minimal unsatisfiable formulas with bounded clause-variable difference are fixed-parameter tractable'. In: T. Warnow and B. Zhu (eds.): *Proceedings of the 9th International Computing and Combinatorics Conference (COCOON'03)*, Vol. 2697 of *Lecture Notes in Computer Science*. pp. 548–558.
18. Tovey, C. A.: 1984, 'A simplified NP-complete satisfiability problem'. *Discrete Applied Mathematics* **8**(1), 85–89.