Electronic Colloquium on Computational Complexity, Report No. 29 (2007)

A Dichotomy Theorem within SCHAEFER for the Boolean Connectivity Problem

Kazuhisa Makino[1], Suguru Tamaki[2], and Masaki Yamamoto[3]

[1] Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 113-8656, Japan. **makino@mist.i.u-tokyo.ac.jp**

[2] Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan. **tamak@kuis.kyoto-u.ac.jp**

[3] Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan. **masaki.yamamoto@kuis.kyoto-u.ac.jp**

Abstract

Gopalan et al. studied in [14] connectivity properties of the solution-space of Boolean formulas, and investigated complexity issues on connectivity problems in Schaefer's framework [26]. A set S of logical relations is SCHAEFER if all relations in S are either bijunctive, Horn, dual Horn, or affine. They conjectured that the connectivity problem for SCHAEFER is in $\mathcal{P}$. We disprove their conjecture by showing that it is co$\mathcal{NP}$-complete for Horn and dual Horn relations. This, together with the results in [14], implies a dichotomy theory within SCHAEFER and a trichotomy theory for the connectivity problem. We also show that the connectivity problem for bijunctive relations can be solved in $O(\min\{n|\varphi|, T(n)\})$ time, where n denotes the number of variables, φ denotes the corresponding 2-CNF formula, and $T(n)$ denotes the time needed to compute the transitive closure of a directed graph of n vertices. Furthermore, we investigate a tractable aspect of Horn and dual Horn relations.

1 Introduction

The Boolean satisfiability problem (satisfiability problem for short) is one of the central problems in the computational complexity theory. Schaefer proposed in [26] a framework for expressing variants of the satisfiability problem, and showed a dichotomy theorem: the satisfiability problem for certain classes of Boolean formulas is in $\mathcal{P}$ while it is $\mathcal{NP}$-complete for all other classes. From this theorem, we have that 2-SAT and HORN-SAT are in $\mathcal{P}$, while k-SAT for $k \geq 3$, NAE-SAT (or NOT-ALL-EQUAL SAT), and XSAT (or EXACT SAT) are all $\mathcal{NP}$-complete. Since then, dichotomies or trichotomies have been established for several aspects of the satisfiability problem such as optimization [3, 5, 24], counting [4], inverse satisfiability [23], minimal satisfiability [16], unique satisfiability [15], 3-valued satisfiability [2] and propositional abduction [6].

Very recently, Gopalan et al. studied in [14] connectivity properties of the solution-space of Boolean formulas, and investigated complexity issues on connectivity problems in Schaefer's framework [26], where the connectivity properties of disjunctive normal forms (DNFs) were studied by Ekin et al. [11]. The connectivity problem (CONN) is to decide if the solutions of a given Boolean

ISSN 1433-8092

Relations	SAT(S)	CONN(S)	ST-CONN(S)
bijunctive	$\mathcal{P}$ [26]	$\mathcal{P}$ [$*$] (implied by [14])	$\mathcal{P}$ [14]
affine			
Horn		co$\mathcal{NP}$-complete [$*$]	
dual Horn			
tight and non-SCHAEFER	$\mathcal{NP}$-complete [26]	co$\mathcal{NP}$-complete [14]	
non-tight		$\mathcal{PSPACE}$-complete [14]	$\mathcal{PSPACE}$-complete [14]

Figure 1: Complexity of the connectivity and st-connectivity problems

formula φ on n variables induce a connected subgraph of the n-dimensional hypercube, while the st-connectivity problem (ST-CONN) is to decide if two specific solutions s and t of φ are connected. As mentioned in [14], connectivity properties of Boolean satisfiability merit study in their own right, since they shed light on the structure of the solution-space, and moreover, structural studies on the solution-space are important to analyze the satisfiability problem and algorithms for it. They [14] established a dichotomy for the st-connectivity problem: It is polynomially solvable if given Boolean relations are *tight*, while it is $\mathcal{PSPACE}$-complete in all other cases. This reveals that the tractable side is larger than the one for the satisfiability problem. Namely, the tight class properly contains SCHAEFER that consists of the classes of bijunctive, Horn, dual Horn, and affine relations. For the connectivity problem, they established a dichotomy with the same boundary: One side is in co$\mathcal{NP}$ and the other side is $\mathcal{PSPACE}$-complete. Furthermore, they showed that the connectivity problem for the class of non-SCHAEFER and tight is co$\mathcal{NP}$-complete. However, they did not give us a complete picture of the complexity status of the connectivity, and conjectured that the connectivity problem for SCHAEFER is in $\mathcal{P}$ [1].

In this paper, we disprove their conjecture by showing that it is co$\mathcal{NP}$-complete for Horn and dual Horn relations. This, together with the results in [14], implies a dichotomy theory within SCHAEFER and a trichotomy theory for the connectivity problem. We also show that the connectivity problem for bijunctive relations can be solved in $O(\min\{n|\varphi|, T(n)\})$ time, where n denotes the number of variables, φ denotes the corresponding 2-CNF formula, and $T(n)$ denotes the time needed to compute the transitive closure of a directed graph of n vertices [2]. It is known [27] that $T(n) = \tilde{O}(n^{\omega})$, where $\omega \leq 2.376$ and we write $f = \tilde{O}(g)$ if $f = O(g \log^k g)$ for some constant k. We also give the proof of the polynomially solvability for affine relations to have a self-contained paper within SCHAEFER.

Figure 1 summarizes complexity of the satisfiability, connectivity and st-connectivity problems, where our results are marked by $*$.

We also investigate a tractable aspect of the intractable side (i.e., Horn and dual Horn re-

[1] Actually, they mentioned (without proofs) that the connectivity problem can be solved in polynomial time for bijunctive and affine relations. Thus, what remains is to show the exact complexity of the connectivity problem for Horn and dual Horn relations.

[2] We again note that the polynomiality is only mentioned in [14].

lations). We consider the *semantic* (i.e., model-based) representation of Horn relations, instead of the traditional *syntactic* (i.e., formula-based) one. The model-based representation has been proposed as an alternative form of representing and accessing a logical knowledge base, e.g., [7, 8, 9, 17, 18, 19, 21, 22]. In contrast to the formula-based representation, if we have the model-based representation, that is, if we are given the *characteristic set* of Horn relations, the connectivity problem is solvable in polynomial time. This strengthens the result in [11] that the connectivity problem for DNF formulas can be solved in polynomial time, since model-based representation M is more compact than DNF representation. More precisely, for any DNF formula ψ, we have $|M| \leq n|\psi|$, where $|M| \ll |\psi|$ is expected in most cases.

The rest of the paper is organized as follows. In the next section, we review the basic Boolean concepts and fix notations. Section 3 shows a dichotomy theorem within SCHAEFER for the connectivity problem, and Section 4 considers the connectivity problem for model-based representation of Horn relations.

2 Preliminaries

We review the basic concepts of the classification of Boolean constraint satisfaction problems, which were introduced by Schaefer [26]. A *logical relation* R over k Boolean variables, which is called a *k-arity relation*, is a mapping from $\{0,1\}^k$ to $\{0,1\}$. We say that a k-arity relation R is *satisfied* by an assignment $t \in \{0,1\}^k$ if and only if $R(t) = 1$. Let S be a (finite) set of relations, and X be a set of Boolean variables. An *S-constraint* over X is defined as the form of $R(y_1, \cdots, y_k)$ for some k-arity relation $R \in S$ and some $\{y_1, \cdots, y_k\} \subseteq X$. We say that a collection φ of S-constraints over n variables is *satisfied* by an assignment $t \in \{0,1\}^n$, denoted by $\varphi(t) = 1$, if every S-constraint of φ is satisfied by t. We call such an assignment t a *satisfying assignment* or a *solution* for φ. In this framework, the satisfiability problem SAT(S) is to decide if there exists a solution for a given collection φ of S-constraints. In this framework, several problems have been investigated. In this paper, we consider the connectivity problem, denoted by CONN(S), which was introduced by [14].

Let H_n be the n-dimensional hypercube. Given a collection φ of S-constraints over n variables, we denote by $G(\varphi) = (V_\varphi, E_\varphi)$ the subgraph of H_n induced by the solutions of φ, that is, $V_\varphi = \{t \in \{0,1\}^n : \varphi(t) = 1\}$, and $(t, t') \in E_\varphi$ for $t, t' \in V$ if and only if the Hamming distance $d(t, t')$ between t and t' is one. The connectivity problem CONN(S) is to decide if $G(\varphi)$ is connected for a given collection φ of S-constraints. In this paper, we assume that readers are familiar with the standard notions and notations of graph theory such as path, cycle and connected component.

Let X be a set of Boolean variables. A *literal* is a variable $x \in X$ or its negation $\overline{x}$, which are respectively called *positive* and *negative*. A *clause* is a disjunction of literals, whose *length* is defined as the number of literals in it. A clause is called *unit* if its length is one. A formula is called *conjunctive normal form* (*CNF*) if it is a conjunction of clauses. A CNF formula is called *2-CNF* if each clause is of length at most two, and *Horn* (resp., *dual Horn*) if each clause has at most one positive (resp., negative) literal. Given a formula φ over X, a set of $\{y_1, \cdots, y_k\} \subseteq X$, and $a_1, \cdots, a_k \in \{0,1\}$, we denote by $\varphi|_{y_1=a_1,\cdots,y_k=a_k}$ the formula obtained from φ by assigning y_i to a_i for $i = 1, \cdots, k$.

In this paper, we are interested in the connectivity problem CONN(S) with respect to the following four types of relations.

Definition 1 Let R be a relation. We say that R is (1) *bijunctive* if it is expressible as a 2-CNF formula, (2) *Horn* if it is expressible as a Horn formula, (3) *dual Horn* if it is expressible as a dual Horn formula, and (4) *affine* if it is expressible as a system of linear equations over $\mathbb{GF}(2)$.

Definition 2 A set S of relations is SCHAEFER if at least one of the following holds: (1) Every relation in S is bijunctive; (2) Every relation in S is Horn; (3) Every relation in S is dual Horn; (4) Every relation in S is affine.

Since we are concerned with only these four types of relations, we simply deal with CNF formulas and systems of linear equations over $\mathbb{GF}(2)$, instead of relations representing them.

Given two assignments $t, t' \in \{0,1\}^n$, we define a *coordinate-wise partial order* $\leq$ as follows: $t \leq t'$ if $t_i \leq t'_i$ for all $1 \leq i \leq n$. Given a formula φ, we say that a satisfying assignment t is *locally minimal* for φ if t has no satisfying neighbor t' with $t' \leq t$, i.e., $\varphi(t') = 1$, $d(t', t) = 1$ and $t' \leq t$. Observe that t is locally minimal for φ if and only if, for each i with $t_i = 1$, there exists a clause C in φ that is falsified by flipping the value of t_i from t. In the latter case, we say that t satisfies the *locally minimal condition.* A path $P = t^{(0)}, t^{(1)}, \cdots, t^{(k)}$ in $G(\varphi)$ is called *monotone* (*decreasing*) if $t^{(i-1)} \geq t^{(i)}$ for all $i = 1, \cdots, k$.

For the connectivity of Horn formulas, the following characterization is known.

Lemma 1 (Gopalan et al. [14]) *Let φ be a Horn formula. Then, every component of $G(\varphi)$ contains a unique locally minimal assignment. Moreover, every satisfying assignment is connected to the locally minimal solution in the same component by a monotone path.*

We make use of the following lemma, which is easily derived from the above.

Lemma 2 *Let φ be a Horn formula without unit clauses over n variables (i.e., a Horn formula such that $\varphi(0^n) = 1$). Then, $G(\varphi)$ is connected if and only if there exists no locally minimal assignment other than 0^n.*

3 A Dichotomy Theorem within SCHAEFER

In this section, we provide a dichotomy theorem within SCHAEFER.

3.1 Tractable cases for CONN(S)

This subsection shows that CONN(S) is polynomially solvable, if S is either bijunctive or affine.

First, we briefly see the affine case. Given an affine formula φ, we note that $d(t, t') \geq 2$ for every pair of (distinct) satisfying assignments t and t' for φ. From this observation, $G(\varphi)$ is connected if and only if φ has at most one satisfying assignment. Thus, it suffices to check whether φ is satisfiable and whether φ is uniquely satisfiable, if so. Any affine formula φ with n variables and m clauses can be regarded as a linear system $Ax = \mathbf{1}$ over the finite field $\mathbb{GF}(2)$, where A is an $m \times n$ 0-1 matrix, and x and $\mathbf{1}$ are respectively the transposes of $(x_1, \cdots, x_n)$ and 1^n. We can easily see that φ is satisfiable if and only if $\text{rank}(A) = \text{rank}([A, \mathbf{1}])$, and φ is uniquely satisfiable if and only if $\text{rank}(A) = \text{rank}([A, \mathbf{1}]) = n$. Since we can obtain the rank of a matrix in polynomial time, we have the following result.

Theorem 1 *Let S be a set of affine relations. Then* $\textsc{Conn}(S)$ *is polynomially solvable.*

We next consider the bijunctive case, i.e., 2-CNF formulas. In what follows, we assume that a given 2-CNF formula is satisfiable, since otherwise, we can easily decide the connectivity.

We first note that we may assume that a given (not necessary 2-CNF) formula has no unit clause.

Proposition 1 *Let φ be a formula over $\{x_1, \cdots, x_n\}$. For an i with $1 \leq i \leq n$ and an $a \in \{0,1\}$, if $V_\varphi = V_{\varphi|_{x_i=a}} \times \{a\}$ (i.e., unit clause x_i is implied by φ if $a = 1$, and $\overline{x}_i$ is implied by φ if $a = 0$), then $G(\varphi)$ is connected if and only if $G(\varphi|_{x_i=a})$ is connected.*

From this proposition, we deal with formulas without unit clauses: If a given formula φ contains a unit clause, say x_i, then we regard $\varphi|_{x_i=1}$ as an input formula. This can be applied until the resulting formula contains no unit clause. Note that this is possible in linear time.

We next note that a given 2-CNF formula may be assumed to be Horn.

Proposition 2 *Let φ be a formula over n variables. For an assignment $a \in \{0,1\}^n$, let ψ be a formula obtained from φ by renaming a, i.e., $\psi(x) = \varphi(x \oplus a)$, where $\oplus$ denotes the component-wise exclusive-or. Then $G(\varphi)$ is connected if and only if $G(\psi)$ is connected.*

Proof. It follows from the fact that $d(t,t') = 1$ if and only if $d(t \oplus a, t' \oplus a) = 1$ for any assignments t and t'. $\square$

From this proposition, we deal with Horn 2-CNF formulas: If a given 2-CNF formula φ is not Horn, then we construct a Horn formula $\psi(x) = \varphi(x \oplus a)$ by computing a satisfying assignment a for φ. Since $\psi(0^n) = 1$, we can see that ψ is Horn, and it can be computed in linear time [1].

In what follows, we assume that a given 2-CNF formula φ is satisfiable and Horn without unit clauses, which is equivalent to the condition that φ is satisfied by 0^n. We now present a notion of *core set*, followed by our key lemma.

Definition 3 Let φ be 2-CNF formula over X. We say that a subset $Y = \{y_1, \cdots, y_k\}$ of X is a *core set* for φ, if $k \geq 2$ and φ contains clauses $y_1 \vee \overline{y}_2, y_2 \vee \overline{y}_3, \cdots, y_{k-1} \vee \overline{y}_k, y_k \vee \overline{y}_1$ that form a cycle called a *core cycle* for φ. Furthermore, we say that a core set Y is *satisfiable* if $\varphi|_{y=1:y \in Y}$ is satisfiable.

Lemma 3 *Let φ be a 2-CNF formula, which is satisfiable and Horn without unit clauses. Then, $G(\varphi)$ is connected if and only if there exists no satisfiable core set for φ.*

Proof. Since $\varphi(0^n) = 1$ by assumption, we recall Lemma 2, i.e., $G(\varphi)$ is connected if and only if φ has no locally minimal non-zero assignment. We first show the only-if part. Let $X = \{x_1, \ldots, x_n\}$ be a variable set, and let $Y = \{x_1, \cdots, x_k\}$ be a satisfiable core set for φ, We assume that φ is satisfied by $t \in \{0,1\}^n$ such that $t_1 = 1, \cdots, t_k = 1, t_{k+1} = 1, \cdots, t_l = 1, t_{l+1} = 0, \cdots, t_n = 0$, where $2 \leq k \leq l \leq n$. Let t' be an assignment such that $t' \leq t$ and $t'_j = 0$ for exactly one j with $1 \leq j \leq k$. Since φ contains a core cycle, this means that t' does not satisfy φ. Hence we have no monotone path from t to 0^n, which proves the only-if part.

We next show the if part. Assume that $G(\varphi)$ is not connected. Then φ has a locally minimal non-zero assignment t, say, $t_1 = 1, \cdots, t_l = 1, t_{l+1} = 0, \cdots, t_n = 0$. Since φ is not satisfied by $t - e^{(j)}$, $j = 1, \cdots, l$, where $e^{(j)}$ is the j-th unit assignment, φ contains a $x_j \vee \overline{x}_{j'}$ with $1 \leq j' \leq l$. This implies that there exists a core cycle in these clauses. Hence there is a satisfiable core set Y for φ such that $Y \subseteq \{x_1 \cdots, x_l\}$, which completes the if part. $\Box$

By this lemma, it is not difficult to see that the connectivity problem for bijunctive relations is solvable in polynomial time.

Let φ be a 2-CNF formula φ, where we assume that it is satisfiable and Horn without unit clauses. We construct a directed graph $G = (V, E)$ from φ in the standard way [1]; i.e., $V = \{x, \overline{x} : x \in X\}$ and $E = \{(\overline{x}, y), (\overline{y}, x) : x \vee y \in \varphi\}$. As shown in [1], G represents implications for φ. Namely if G has a path from x to y, then $x = 1$ always implies $y = 1$, i.e., $\overline{x} \vee y$. Note that, by the assumption on φ, each strongly connected component consists of either only positive literals or only negative literals. Thus by the symmetricity of G (i.e., $(x, y) \in E$ if and only if $(\overline{y}, \overline{x}) \in E$), we write strongly connected components in G by $G_i = (V_i, E_i)$ and $G_i' = (V_i', E_i')$ for $1 \leq i \leq k$, where $V_i \subseteq X$ and $V_i' = \{\overline{x} : x \in V_i\}$. We note that any core cycle is contained in a single connected component G_i, and any connected component G_i with $|V_i| \geq 2$ contains a core cycle. Thus it suffices to check if, for each component $G_i = (V_i, E_i)$ with $|V_i| \geq 2$, $\varphi|_{x=1:x\in V_i}$ is satisfiable. This simply can be done as follows.

Let $H = (V_H = \{V_i, V_i' : 1 \leq i \leq k\}, E_H)$ be a directed graph obtained from G by identifying each connected component to a single vertex.

Lemma 4 *Let φ be a 2-CNF formula φ, which is satisfiable and Horn without unit clauses, and let H be defined as above. Then $G(\varphi)$ is connected if and only if, for every V_i with $|V_i| \geq 2$, there exists a path in H from V_i to V_i'.*

Proof. Let us first show the if part. Since H has a path from V_i to V_i', $x = 1$, $x \in V_i$, always implies $\overline{x} = 1$. Thus we have no satisfiable core set for φ, which proves the if part by Lemma 3.

On the other hand, if $G(\varphi)$ is connected, we have no satisfiable core set for φ. Since any core set is contained in a connected component, say V_i, $\varphi|_{x=1:x\in V_i}$ is unsatisfiable. This means that H contains two directed paths P_1 from V_i to V_j and P_2 from V_i to V_j' for some j. By the symmetricity of H, P_2 implies that H has a path P_3 from V_j to V_i'. Therefore, by concatenating P_1 and P_3, we have a path from V_i to V_i'. $\Box$

It follows from Lemma 4 that the connectivity problem for bijunctive relations can be solved in $O(n|\varphi|)$ time by checking the existence of n paths in H, where we note that H can be computed in linear time. If we first compute the transitive closure H^* of H, it can be computed in $\tilde{O}(n^{\omega})$ time, where $\omega \leq 2.376$ and this is the current best bound for computing the transitive closure of a graph with n vertices.

Formally, our algorithm can be described in Figure 2.

Lemma 5 *Let φ be a 2-CNF formula of n variables. Then* `two-sat-conn`(φ) *correctly solves the connectivity problem for φ in $O(\min\{n|\varphi|, T(n)\})$ time, where $T(n)$ denotes the time needed to compute the transitive closure of a directed graph of n vertices.*

`two-sat-conn`(φ) /* φ: a 2-CNF formula over X */

If φ is not satisfiable, then we output `YES` and halt.

Update φ to a Horn 2-CNF formula without unit clauses by using a satisfying assignment a for φ.

Construct a directed graph $G = (V, E)$ from φ by $V = \{x, \overline{x} : x \in X\}$ and $E = \{(\overline{x}, y), (\overline{y}, x) : x \vee y \in \varphi\}$.

/* Let $G_i = (V_i, E_i)$ and $G'_i = (V'_i, E'_i)$, $1 \leq i \leq k$, be the strongly connected components of G, where $V_i \subseteq X$ and $V'_i = \{\overline{x} : x \in V_i\}$. */

Construct H from G by identifying each connected component to a single vertex.

for each V_i of $V_1, \cdots, V_k$

  if $|V_i| \geq 2$ and there is no path in H from V_i to V'_i, then output `NO` and halt.

end-for-each

Output `YES`.

`end-of-two-sat-conn`

Figure 2: An algorithm for bijunctive relations

Therefore, we have the following positive result.

Theorem 2 *Let S be a set of bijunctive relations. Then* CONN(S) *is polynomially solvable.*

3.2 Co$\mathcal{NP}$-hardness for Horn and dual Horn relations

In this subsection, we prove our main theorem.

Theorem 3 *Let S be a set of relations such that either of the following holds:*

1. *Every relation in S is Horn.*
2. *Every relation in S is dual Horn.*

Then CONN(S) *is co$\mathcal{NP}$-complete.*

We only consider Horn relations S and show that the complement of CONN(S) is NP-complete, since dual Horn relations are handled in a similar way.

First we characterize a necessary and sufficient condition for non-connectivity of Horn relations. In this section, we assume formulas contain no unit clause to make discussion easier. By Lemma 2, a Horn formula is not connected if and only if there exists a locally minimal non-zero satisfying assignment, which can be represented by the following Boolean formula.

$$\Phi(\varphi) \quad = \quad \varphi \wedge \bigwedge_{x_i \in X} \left(\overline{x}_i \vee \Big(\bigvee_{C \in \varphi : P(C) = \{x_i\}} \bigwedge_{y \in N(C)} y \Big) \right). \tag{1}$$

Here, for a clause C, $P(C)$ and $N(C)$ respectively denote sets of variables that occur positively and negatively in C. Note that if $\{C \in \varphi : P(C) = \{x_i\}\}$ is empty, then $(\bigvee_{C \in \varphi : P(C) = \{x_i\}} \bigwedge_{y \in N(C)} y)$ is interpreted as false.

Lemma 6 (Logical formulation of non-connectivity) *Let φ be a Horn formula without unit clauses. Then there exists a locally minimal non-zero assignment of φ if and only if $\Phi(\varphi)$ is satisfied by a non-zero assignment.*

Proof. For the if part, let t be a non-zero satisfying assignment of $\Phi(\varphi)$. Note that t also satisfies φ. To confirm that this t satisfies the locally minimal condition, pick an arbitrary variable x_i such that $t_i = 1$. Since t satisfies $\overline{x}_i \vee \bigvee_{C\in\varphi:P(C)=\{x_i\}} \left(\bigwedge_{y\in N(C)} y\right)$, φ contains a clause C such that t satisfies $\bigwedge_{y\in N(C)} y$, which implies that $t - e^{(i)}$ does not satisfy φ. Here $e^{(i)}$ denotes the i-th unit assignment. This completes the if part.

For the only-if part, let t be a locally minimal non-zero assignment of φ. For all $x_i \in X$, we show that t satisfies $\overline{x}_i \vee \bigvee_{C\in\varphi:P(C)=\{x_i\}} \left(\bigwedge_{y\in N(C)} y\right)$. It is obvious for x_i with $t_i = 0$. For x_i with $t_i = 1$, we have a clause C in φ such that $P(C) = \{x_i\}$ and $N(C) \subseteq \{x_j \in X : t_j = 1\}$. This proves the claim. □

By Lemmas 2 and 6, we have the following characterization.

Corollary 1 (Characterization of non-connectivity) *Let φ be a Horn formula without unit clauses. Then $G(\phi)$ is non-connected if and only if $\Phi(\varphi)$ is satisfied by a non-zero assignment.*

Now we are ready to prove the theorem.

The proof of Theorem 3. It follows from Proposition 1 and Corollary 1 that the complement of $\textsc{Conn}(S)$ for Horn relations S belongs to $\mathcal{NP}$. To show the $\mathcal{NP}$-hardness, we reduce to it 3-Uniform Hypergraph 2-Colorability, which is known to be $\mathcal{NP}$-complete (see SP4 in [13]). Let $\mathcal{H} = (V, \mathcal{E})$ be a 3-uniform hypergraph, where a hypergraph is called 3-*uniform* if $|E| = 3$ holds for all $E \in \mathcal{E}$. From $\mathcal{H}$, we construct a Horn formula $\varphi_{\mathcal{H}}$ over a variable set $X \cup X' \cup Y \cup \{q\}$, where $X = \{x_v : v \in V\}$, $X' = \{x'_v : v \in V\}$, and $Y = \{y_E : E \in \mathcal{E}\}$, as follows:

$$\varphi_{\mathcal{H}} \equiv \bigwedge_{E\in\mathcal{E}} \left(\bigvee_{v\in E} \overline{x}_v\right) \wedge \bigwedge_{E\in\mathcal{E}} \left(\bigwedge_{v\in E} (y_E \vee \overline{x}_v)\right) \tag{2}$$

$$\wedge \left(\bigvee_{E\in\mathcal{E}} \overline{y}_E \vee q\right) \wedge \bigwedge_{v\in V} \left((\overline{x}_v \vee x'_v \vee \overline{q})(x_v \vee \overline{x'_v} \vee \overline{q})\right). \tag{3}$$

Example 1 Let $\mathcal{H} = (V, \mathcal{E})$ be a 3-uniform hypergraph defined by $V = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{E_1 = \{1, 2, 3\}, E_2 = \{2, 3, 4\}\}$. Then $\varphi_{\mathcal{H}}$ in (3) is given by

$$\begin{aligned}\varphi_{\mathcal{H}} \;=\; & (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3)(\overline{x}_2 \vee \overline{x}_3 \vee \overline{x}_4) \\ & \wedge (y_{E_1} \vee \overline{x}_1)(y_{E_1} \vee \overline{x}_2)(y_{E_1} \vee \overline{x}_3)(y_{E_2} \vee \overline{x}_2)(y_{E_2} \vee \overline{x}_3)(y_{E_2} \vee \overline{x}_4) \\ & \wedge (\overline{y}_{E_1} \vee \overline{y}_{E_2} \vee q)(\overline{x}_1 \vee x'_1 \vee \overline{q})(x_1 \vee \overline{x'_1} \vee \overline{q})(\overline{x}_2 \vee x'_2 \vee \overline{q})(x_2 \vee \overline{x'_2} \vee \overline{q}) \\ & \wedge (\overline{x}_3 \vee x'_3 \vee \overline{q})(x_3 \vee \overline{x'_3} \vee \overline{q})(\overline{x}_4 \vee x'_4 \vee \overline{q})(x_4 \vee \overline{x'_4} \vee \overline{q}).\end{aligned}$$

Lemma 7 *Let $\mathcal{H}$ be a 3-uniform hypergraph and $\varphi_{\mathcal{H}}$ be a Horn formula constructed from $\mathcal{H}$ as above. Then $\mathcal{H}$ is 2-colorable if and only if $G(\varphi_{\mathcal{H}})$ is non-connected.*

Proof. Since $\varphi_{\mathcal{H}}$ constructed from a given $\mathcal{H}$ contains no unit clause by 3-uniformity of $\mathcal{H}$, we make use of the characterization of non-connectivity. The corresponding formula $\Phi(\varphi_{\mathcal{H}})$ in (1) can be written as follows:

$$
\begin{aligned}
\Phi(\varphi_{\mathcal{H}}) \quad = \quad & \varphi_{\mathcal{H}} && (4)\\
& \wedge \bigwedge_{E\in\mathcal{E}} \left(\overline{y}_E \vee \bigvee_{v\in E} x_v \right) && (5)\\
& \wedge \left(\overline{q} \vee \bigwedge_{E\in\mathcal{E}} y_E \right) && (6)\\
& \wedge \bigwedge_{v\in V} \left((\overline{x}_v \vee x'_v q)(\overline{x'_v} \vee x_v q) \right). && (7)
\end{aligned}
$$

Example 2 For a 3-uniform hypergraph $\mathcal{H}$ in Example 1, we have

$$
\begin{aligned}
\Phi(\varphi_{\mathcal{H}}) \quad = \quad & \varphi_{\mathcal{H}} \wedge (\overline{y}_{E_1} \vee x_1 \vee x_2 \vee x_3)(\overline{y}_{E_2} \vee x_2 \vee x_3 \vee x_4)\\
& \wedge (\overline{q} \vee y_{E_1} y_{E_2})(\overline{x}_1 \vee x'_1 q)(\overline{x'_1} \vee x_1 q)(\overline{x}_2 \vee x'_2 q)(\overline{x'_2} \vee x_2 q)\\
& \wedge (\overline{x}_3 \vee x'_3 q)(\overline{x'_3} \vee x_3 q)(\overline{x}_4 \vee x'_4 q)(\overline{x'_4} \vee x_4 q).
\end{aligned}
$$

To see the condition that $\Phi(\varphi_{\mathcal{H}})$ is satisfied by a non-zero assignment, we consider two cases depending on the value of q.

Fact 1 $\Phi(\varphi_{\mathcal{H}})|_{q=0}$ *is satisfied only if all the remaining variables are set to be 0.*

Proof. Setting $q = 0$ induces unit clauses $\overline{x}_v, \overline{x'_v}$ for every $v \in V$ by (7). This, together with (5) induces unit clause $\overline{y}_E$ for every $E \in \mathcal{E}$. These imply Fact 1. □

Fact 2 $\Phi(\varphi_{\mathcal{H}})|_{q=1}$ *is satisfiable if and only if* $\mathcal{H}$ *is* 2*-colorable.*

Proof. Setting $q = 1$ induces $\bigwedge_{E\in\mathcal{E}} y_E$ by (6), that simplifies $\Phi(\varphi_{\mathcal{H}})|_{q=1}$ as follows:

$$
\begin{aligned}
\Phi(\varphi_{\mathcal{H}})|_{q=1} \quad = \quad & \bigwedge_{E\in\mathcal{E}} \left(\left(\bigvee_{v\in E} \overline{x}_v \right) \wedge \left(\bigvee_{v\in E} x_v \right) \right) && (8)\\
& \wedge \bigwedge_{v\in V} \left((\overline{x}_v \vee x'_v)(x_v \vee \overline{x'_v}) \right). && (9)
\end{aligned}
$$

Here we note that (8) is obtained from (4) and (5), while (9) is obtained from (7). It is easy to see that (8) is satisfiable if and only if H is 2-colorable. Since (9) just forces $x_v = x'_v$ for every $v \in V$, we have Fact 2. □

These facts, combined with Corollary 1, prove Lemma 7. □

This completes the proof of Theorem 3. □

Remark 1 *The connectivity problem for Horn formulas is $\mathcal{NP}$-complete, even if we restrict Horn formulas to 3-CNF, i.e., each clause contains at most three literals. This can be shown by slightly modifying $\varphi_{\mathcal{H}}$ as follows. Let $\mathcal{E} = \{E_1, \cdots, E_k\}$, and let z_i, $i = 1, \ldots k-2$, be new variables. We replace clause $(\overline{y}_{E_1} \vee \overline{y}_{E_2} \vee \cdots \vee \overline{y}_k \vee q)$ in (3) by*

$$(\overline{y}_{E_1} \vee \overline{y}_{E_2} \vee z_1)(\overline{z}_1 \vee \overline{y}_{E_3} \vee z_2)(\overline{z}_2 \vee \overline{y}_{E_4} \vee z_3) \cdots (\overline{z}_{k-2} \vee \overline{y}_{E_k} \vee q).$$

The correctness can be proved in a similar manner.

This together with Theorem 2 gives us a tractability-intractability border for Horn connectivity problem. Namely, Horn connectivity problem is solvable in polynomial time if a given Horn formula φ is 2-CNF, while it is co$\mathcal{NP}$-complete if φ is k-CNF for $k \geq 3$.

4 Horn Relations Represented by Characteristic Set

In this section, we investigate a tractable aspect of Horn and dual Horn relations. Specifically, we show that if we are given the characteristic set of Horn relations, the connectivity problem is solvable in polynomial time.

We recall that Horn relations have a well-known semantical characterization. Let for assignments $t, t' \in \{0,1\}^n$ denote $t \bigwedge t'$ their component-wise AND, and let for a set of assignment $M \subseteq \{0,1\}^n$ denote $Cl_\wedge(M)$ the closure of M under $\bigwedge$. Then, for every $M \subseteq \{0,1\}^n$, it holds that M is the set of satisfying assignments for some Horn formula φ (i.e, $M = V_\varphi$) if and only if $M = Cl_\wedge(M)$ [25] (see e.g., [7, 21] for proofs). Namely, the set of satisfying assignments of a Horn formula is closed under the intersection $\bigwedge$, and any set of assignments which is closed under the intersection can be represented by a Horn formula. By this characterization, it is easy to see that any Horn formula has a unique minimal satisfying assignment. Here a satisfying assignment t for φ is called *minimal* (resp., *maximal*) if there exists no other satisfying assignment t' for φ such that $t' \leq t$ (resp., $t' \geq t$). By definition, minimal assignments for φ are locally minimal, but not vice versa.

As discussed by Kautz *et al.* [17], a Horn formula φ is semantically represented by its characteristic assignment, where $v \in V_\varphi$ is called *characteristic* (or *extreme* [7]), if $v \notin Cl_\wedge(V_\varphi \setminus \{v\})$. The set of all characteristic assignments of φ, the *characteristic set of* φ, is denoted by $char(\varphi)$. Note that $char(\varphi)$ is unique and that $char(\varphi)$ contains all maximal satisfying assignments for φ.

Lemma 8 *For a Horn formula φ, let t^* be the unique minimal satisfying assignment for φ. Then $G(\varphi)$ is connected if and only if, for each maximal satisfying assignment t for φ, $G(\varphi)$ contains a monotone path between t^* and t.*

Proof. Since the only-if part is easily derived from Lemma 1, we only show the if part. We assume that, for each maximal assignment t of φ, $G(\varphi)$ contains a monotone path from t to t^*. We show that there is no locally minimal assignment other than t^*. This, together with Lemma 1 implies that $G(\varphi)$ is connected.

Let v be an arbitrary satisfying assignment for φ which is not maximal. Let t be a maximal assignment such that $t \geq v$. Since $G(\varphi)$ contains a monotone path from t to t^*, there exists an edge (u, w) in the path such that $u \geq v$ and $v \gneq (v \wedge w)$. Note that $v \wedge w$ is a satisfying assignment, and $d(v, v \wedge w) = 1$. This means that v is not locally minimal. $\square$

By Lemma 8, the following simple algorithm checks the connectivity of Horn relations represented by the characteristic set.

Since $t \in V_\varphi$ if and only if $t = \bigwedge_{v \in char(\varphi): v \geq t} v$, for each assignment t in $char(\varphi)$ that includes maximal assignments for φ, the algorithm checks if $G(\varphi)$ contains a monotone path between t and t^*. Thus, from Lemma 8, algorithm `horn-sat-conn-from-charset`(φ) checks the connectivity of Horn functions.

`horn-sat-conn-from-charset`$(char(\varphi))$

/* $char(\varphi)$: the characteristic set of a Horn formula φ */

Let $t^* := \bigwedge_{t \in char(\varphi)} t$ and $M := char(\varphi)$

while $(M \neq \{t^*\})$

Let t be an arbitrary element in $M \setminus \{t^*\}$

if there exists an index j s.t. $t_j = 1$ and $t - e^{(j)} = \bigwedge_{v \in char(\varphi): v \geq t - e^{(j)}} v$

then $M := (M \setminus \{t\}) \cup \{t - e^{(j)}\}$

Otherwise, output `NO` and halt

end-while

Output `YES`

`end-of-horn-sat-conn-from-charset`

Figure 3: A naive algorithm for the Horn connectivity from the characteristic set

Theorem 4 *Given the characteristic set $char(\varphi)$ of a Horn formula φ, algorithm `horn-sat-conn-from-charset`(φ) checks its connectivity in $O(n^3|char\ (\varphi)|^2)$ time.*

Proof. Since the correctness of the algorithm follows from Lemma 8 and the discussion before the description, we only show its time complexity.

Clearly, we can initialize M and t^* in $O(n|char(\varphi)|)$ time. For each $t \in M \setminus \{t^*\}$, we can test if there exists an index j such that $t_j = 1$ and $t - e^{(j)} = \bigwedge_{w \in char(\varphi): w \geq t - e^{(j)}} w$ in $O(n^2|char(\varphi)|)$ time. Since we have at most $n|char(\varphi)|$ such t's, this requires $O(n^3|char(\varphi)|^2)$ time. Therefore, in total, the algorithm requires $O(n^3|char(\varphi)|^2)$ time. □

We now improve the complexity. For an assignment t, let $S_t = \{j \mid t_j = 0\}$. It is not difficult to see that algorithm `horn-sat-conn-from-charset2`$(char(\varphi))$ checks the connectivity of a Horn formula φ: In the for-loop, we check if there exists a monotone path from each $t \in char(\varphi)$ to t^*. In the while-loop, we maintain a variable set X such that the corresponding assignment t^X (i.e., $t^X_j = 0$ if $j \in X$, and 1 otherwise) is reachable from t by a monotone path. Observe that t^X is not locally minimal if and only if there is a set S in the current $\mathcal{S}$ such that $|S \setminus X| = 1$. Moreover, the while-loop requires $O(n|char(\varphi)|)$ time, if $\mathcal{S}$ is stored in the proper data structure. Thus we have the following theorem.

Theorem 5 *Given the characteristic set $char(\varphi)$ of a Horn formula φ, algorithm `horn-sat-conn-from-charset2`(φ) checks its connectivity in $O(n|char\ (\varphi)|^2)$ time.*

```
horn-sat-conn-from-charset2(char(φ))
    /* char(φ): the characteristic set of a Horn formula φ */
        Let t* := ⋀_{t∈char(φ)} t
        for each t of char(φ)
            Let 𝒮 := {S_v | v ∈ char(φ)} and X := S_t
            while (∃S ∈ 𝒮 with |S \ X| ≤ 1)
                Let S be an arbitrary element in 𝒮
                if |S \ X| = 0, then 𝒮 := 𝒮 \ {S}
                if |S \ X| = 1, then 𝒮 := 𝒮 \ {S} and X := X ∪ S
            end-while
            if X ≠ S_{t*}, then output NO and halt
        end-for-each
        Output YES
end-of-horn-sat-conn-from-charset2
```

Figure 4: A faster algorithm for the Horn connectivity from the characteristic set

Remark 2 *This strengthens the result in [11] that the connectivity problem for DNF formulas can be solved in polynomial time, since the characteristic set $char(\varphi)$ is more compact than DNF representation ψ. More precisely, for any DNF formula ψ, we have $|char(\varphi)| \leq n|\psi|$, where $|char(\varphi)| \ll |\psi|$ is expected in most cases.*

Remark 3 *For Horn relations, formula-based (i.e., CNFs) and model-based (characteristic sets) representations are* orthogonal *in the sense that the one side may be exponentially larger than the other one. Therefore, the results in this section do not conflict with Theorem 3 in the previous section. We further remark that the transformation between a Horn formula φ and the characteristic set $char(\varphi)$ is at least as difficult as the monotone dualization problem [19, 20], which is known to be solved in output quasi-polynomial time [10, 12].*

References

[1] B. Aspvall, M. F. Plass and R. E. Tarjan. A linear-time algorithm for testing the truth of certain quantified Boolean formulas. *Information Processing Letters* 8; pp. 121-123, 1979.

[2] A. Bulatov. A dichotomy theorem for constraints on a three-element set. in *Proceeding of 43rd IEEE Symposium on Foundations of Computer Science*, pp. 649–658, 2002.

[3] N. Creignou. A dichotomy theorem for maximum generalized satisfiability problems. *Journal of Computer and System Sciences*, 51: pp. 511-522,1995.

[4] N. Creignou and M. Hermann. Complexity of generalized satisfiability counting problems. *Information and Computation*, 125: pp. 1-12, 1996.

[5] N. Creignou, S. Khanna, and M. Sudan. *Complexity classification of Boolean constraint satisfaction problems.* SIAM Monographs on Discrete Mathematics and Applications, 2001.

[6] N. Creignou and B. Zanuttini. A complete classification of the complexity of propositional abduction. *SIAM Journal on Computing*, 36: pp. 207 - 229, 2006.

[7] R. Dechter and J. Pearl. Structure identification in relational data. *Artificial Intelligence*, 58: pp. 237–270, 1992.

[8] T. Eiter, T. Ibaraki and K. Makino. Computing intersections of Horn theories for reasoning with models. *Artificial Intelligence* 110: pp. 57-101, 1999.

[9] T. Eiter and K. Makino. *Generating all abductive explanations for queries on propositional Horn theories*, KBS Research Report INFSYS RR-1843-03-09, Institute of Information Systems, Vienna University of Technology, 2006.

[10] T. Eiter, K. Makino, and G. Gottlob. *Computational aspects of monotone dualization: A brief survey.* KBS Research Report INFSYS RR-1843-06-01, Institute of Information Systems, Vienna University of Technology, 2006.

[11] O. Ekin, P. L. Hammer and A. Kogan. On connected Boolean functions. *Discrete Applied Mathematics*, 96–97: pp. 337–362, 1999.

[12] M. L. Fredman and L. Khachiyan. On the complexity of dualization of monotone disjunctive normal forms. *Journal of Algorithms*, 21: pp. 618–628, 1996.

[13] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York, 1978.

[14] P. Gopalan, P. G. Kolaitis, E. N. Maneva, and C. H. Papadimitriou. The connectivity of Boolean satisfiability: Computational and structural dichotomies. in *Proceeding of 33rd International Colloquium on Automata, Languages and Programming* (*ICALP'06*), pp. 346-357, 2006.

[15] L. Juban. Dichotomy theorem for the generalize unique satisfiability problem. in *Proceedings of 12ht International Symposium of Fundamentals of Computation Theory*, LNCS 1684: pp. 327-337, 1999.

[16] L. Kirousis and P. Kolaitis. The complexity of minimal satisfiability problems. *Information and Computation*, 187: pp. 20-39, 2003.

[17] H. Kautz, M. Kearns, and B. Selman. Reasoning With characteristic models. In *Proceedings AAAI-93*, pp. 34–39, 1993.

[18] H. Kautz, M. Kearns, and B. Selman. Horn approximations of empirical data. *Artificial Intelligence*, 74: pp. 129–245, 1995.

[19] D. Kavvadias, C. Papadimitriou, and M. Sideri. On Horn envelopes and hypergraph transversals. In W. Ng, editor, *Proceedings 4th International Symposium on Algorithms and Computation (ISAAC-93)*, LNCS 762, pages 399–405, Hong Kong, December 1993.

[20] R. Khardon. Translating between Horn representations and their characteristic models. *Journal of AI Research* 3: pp. 349-372, 1995.

[21] R. Khardon and D. Roth. Reasoning with models. *Artificial Intelligence*, 87(1/2): pp. 187–213, 1996.

[22] R. Khardon and D. Roth. Defaults and relevance in model-based reasoning. *Artificial Intelligence*, 97: pp. 169–193, 1997.

[23] D. Kavvadias and M. Sideri. The inverse satisfiability problem. *SIAM Journal on Computing*, 28: pp. 152-163, 1998.

[24] S. Khanna, M. Sudan, L. Trevisan, and D. Williamson. The approximability of constraint satisfaction problems. *SIAM Journal on Computing*, 30: pp. 1863-1920, 2001.

[25] J. McKinsey. The decision problem for some classes of sentences without quantifiers. *Journal of Symbolic Logic*, 8: pp. 61–76, 1943.

[26] T. J. Schaefer. The complexity of satisfiability problems. in *Proceeding of 10th ACM Symposium of Theory of Computing*, pp. 216-226, 1978.

[27] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/