On the Difference of Horn Theories*

Thomas Eiter
Institut und Ludwig Wittgenstein Labor für Informationssysteme,
Technische Universität Wien,
Favoritenstraße 9–11, A-1040 Wien, Austria.
(eiter@kr.tuwien.ac.at)

Toshihide Ibaraki
Department of Applied Mathematics and Physics,
Graduate School of Informatics,
Kyoto University, Kyoto 606, Japan.
(ibaraki@kuamp.kyoto-u.ac.jp)

Kazuhisa Makino
Department of Systems and Human Science,
Graduate School of Engineering Science,
Osaka University, Toyonaka, Osaka 560, Japan
(makino@sys.es.osaka-u.ac.jp)

April 27, 2000

*An extended abstract of this paper appears in *Proceedings of the 16th International Symposium on Theoretical Aspects of Computer Science (STACS '99)*, Germany, March 4-6, 1999 [8]. The authors gratefully acknowledge the partial support of the Scientific Grant in Aid by the Ministry of Education, Sports, Science and Culture of Japan. Part of this research was conducted while the first author visited Kyoto University in 1995 and 1998, by the support of the Scientific Grant in Aid by the Ministry of Education, Sports, Science and Culture of Japan (Grant 06044112).

Running head: On the Difference of Horn Theories

Mailing address:

Kazuhisa MAKINO
Department of Systems and Human Science,
Graduate School of Engineering Science,
Osaka University, Toyonaka, Osaka 560, Japan.
E-mail: makino@sys.es.osaka-u.ac.jp
Phone: +81-6-6850-6388
Fax: +81-6-6850-6341

Abstract

In this paper, we consider computing the difference between two Horn theories. This problem may arise, for example, if we take care of a theory change in a knowledge base. In general, the difference of Horn theories is not Horn. Therefore, we consider Horn approximations of the difference in terms of Horn cores (i.e., weakest Horn theories included in the difference) and the Horn envelope (i.e., the strongest Horn theory containing the difference), which have been proposed and analyzed extensively in the literature. We study the problem under the familiar representation of Horn theories by Horn CNFs, as well as under the recently proposed model-based representation in terms of the characteristic models. For all problems and representations, polynomial time algorithms or proofs of intractability for the propositional case are provided; thus, our work gives a complete picture of the tractability/intractability frontier in the propositional Horn theories.

1 Introduction

When we use logical formulas for representing a certain problem, it is in general necessary that different formulas are combined in way to obtain a more comprehensive and accurate description of the problem. Among the basic operations for combining logical formulas are the Boolean operations, i.e., conjunction $\wedge$, disjunction $\vee$, and complement $\neg$. In this paper, we consider another basic operation, the difference $\setminus$, which can represent the complement $\neg$ (because $\neg\Sigma$ can be represented by the difference of $\{0,1\}^n$ and Σ, i.e., $\{0,1\}^n \setminus \Sigma$, for a theory $\Sigma \subseteq \{0,1\}^n$), and can be represented by the complement $\neg$ and conjunction $\wedge$ (because $\Sigma_1 \setminus \Sigma_2 = \Sigma_1 \wedge \overline{\Sigma}_2$ for theories $\Sigma_1, \Sigma_2 \subseteq \{0,1\}^n$, where $\wedge$ is regarded as the set operation $\cap$). In principle, all these operations are easily accomplished, since the language of propositional logic provides the boolean connectives $\wedge$, $\vee$, and $\neg$ (or $\setminus$), which can be readily applied to combine formulas as desired. However, there are two obstacles to this immediate approach in practice.

One obstacle is that the problem description is frequently required to be in a special format. A predominant such format is conjunctive normal form (CNF), i.e., a logical theory Σ is represented by a set of clauses. Another is the model-based format, where a theory is represented by characteristic models. Combining two theories Σ_1 and Σ_2 in some format may then require that the result is also cast to the same format. For example, if Σ_1 and Σ_2 are represented by CNFs φ_1 and φ_2, respectively, then we can easily obtain the formula $\varphi_1 \setminus \varphi_2 \equiv \varphi_1 \wedge \overline{\varphi}_2$ of the difference $\Sigma_1 \setminus \Sigma_2$. Namely, it is represented by the conjunction of the CNF φ_1 and the disjunctive normal form (DNF) φ_3, where φ_3 is the formula obtained from $\overline{\varphi}_2$ by De Morgan's law. However, $\varphi_1 \wedge \overline{\varphi}_2$ is not a CNF, and moreover we do not know how to efficiently compute the CNF representation from it.

Another obstacle is computational complexity. As is well-known, even in propositional logic, the satisfiability problem of a CNF is NP-complete. In trading expressiveness for tractability, theories have been restricted to fragments of the full language. A most important such fragment is the class of Horn theories. Fields such as Logic Programming and Deductive Databases are based on their appealing semantical and computational properties, cf. [11]. In fact, deduction of a clause from a propositional Horn theory is possible in linear time [5], while this is co-NP-complete in general.

In this paper, we study the problem of computing the Boolean difference between two Horn theories Σ_1 and Σ_2, i.e., $\Sigma = \Sigma_1 \setminus \Sigma_2$. In general, the resulting theory Σ is not Horn. Therefore, we consider approximating Σ by a Horn theory, in order to maintain the desired closure property. Different such approximations are possible, but Horn cores and the Horn envelope, which have been introduced in [21], appear to be most natural.

A Horn theory Π is a *Horn core* (or a *Horn greatest lower bound*) of a theory Σ, if Π is included in Σ, i.e., Π logically implies Σ, and there is no weaker Π of this property. A theory R is the *Horn envelope* (or the *Horn least upper bound*) of Σ, if it is the strongest Horn theory R which includes Σ, i.e. Σ logically implies R. While there is more than one Horn core in general, the Horn envelope is always unique. Semantical and computational issues on Horn cores and the Horn envelope have been studied extensively, and a number of results have been obtained, cf. [21, 16, 2, 17, 18, 3, 4, 10, 1].

The main results of the present paper can be summarized as follows.

- We present characterizations of the Horn cores and the Horn envelope of a Horn difference $\Sigma_1 \setminus \Sigma_2$, which will form a basis of the algorithms discussed in this paper.
- We either present a polynomial time algorithm or prove intractability (unless P=NP) for each of the problems mentioned above. (For a concise overview, see Table 1 in Section 6.)
- Besides the familiar representation in terms of Horn CNFs, we also consider the model-based representation of Horn theories through their sets of characteristic models [14, 15]. This alternative has also been studied repeatedly, since it offers advantages to formula-based representation in certain cases; see [19, 18, 15, 6] for more details.
- Both formula-based and model-based representations allow polynomial time algorithms for many problems. In some cases, however, formula-based representation is polynomial while model-based representation is intractable, and vice versa. Thus, like with many other problems [15, 19, 6], formula-based and model-based representations complement each other with respect to their tractability/intractability profile.

Our results on the difference of Horn theories have applications in different contexts. On one hand, we extend the results on propositional knowledge base (KB) approximation in [16, 2, 10, 1], by providing a polynomial time algorithm for computing some Horn core of a KB which is expressible as the difference of two Horn theories. This should be compared with the general result [2] that computing a Horn core of a KB described by an arbitrary CNF formula φ is not possible in polynomial time, unless P = NP, and most likely impossible with a constant number of queries to an NP oracle. On the other hand, we establish that computing the Horn envelope of the difference $\varphi_1 \setminus \varphi_2$ is intractable. This complements the results in [17, 18, 16, 3, 4] that computing a Horn CNF of the Horn envelope is difficult from a CNF or the extensional description of an arbitrary theory (i.e., its set of models). (Section 6 contains a more detailed discussion.)

Our results find an application in the area of theory change, which has grown into an important research area within AI during the last decade. Here we consider the case in which the change does not need to be retracted. Taking the difference between Horn theories is meaningful in the following scenario, for example. Assume that we have a Horn theory Σ, which is a description of all the possible worlds of a state of affairs; i.e., the real "world" amounts to one of the models in Σ, say w, but is unknown to us. Suppose that we now obtain the information that a formula φ is false in w. Since w is a categorical description of the world, either ψ or $\neg\psi$ is true in it (but not both), for any formula ψ. Hence, in this case, we conclude that $\neg\varphi$ is true in w, and thus all models v that satisfy φ can be discarded from Σ. This amounts to updating Σ to $\Sigma_{new} = \Sigma \setminus \Sigma'$, where Σ' is the set of models satisfying φ; if φ is Horn, this gives an instance of our problem. If Σ_{new} is not Horn, its Horn envelope and the collection of all Horn cores provide sound and complete approximations of Σ_{new} in terms of Horn theories. Notice that for inference of Horn formulas, the Horn envelope of Σ_{new} is sound and complete [19, 3]. Our algorithms can be applied in this scenario for computing Σ_{new} or its Horn approximations. Observe that computing a Horn core and the Horn envelope in the context of theory change was studied in [10], in which the theory $\varphi \circ \psi$, given by two Horn CNFs φ and ψ, is studied for some revision operators "$\circ$" (see [13]). However, this is only weakly related to ours, since the operation of taking the difference of φ and ψ was not considered, and moreover, the formula ψ is restricted to be a single Horn clause.

The remainder of this paper is structured as follows. The next section recalls elementary concepts and provides basic notations. In Section 3, we consider the problem of deciding whether the difference of Horn formulas is Horn, and computing a representation of it if this is the case. After that, we study in Sections 4 and 5 computing Horn cores and the Horn envelope of the difference, respectively. The final Section 6 gives a discussion of the results and related work, as well as outlines further research issues.

2 Preliminaries

We assume a supply of propositional variables $x_1, x_2, \ldots, x_n$, where each x_i evaluates to either 1 (true) or 0 (false). Negated variables are denoted by $\overline{x}_i$. The x_i and $\overline{x}_i$ are called literals. A clause is a disjunction $c = \ell_1 \vee \cdots \vee \ell_k$ of literals, while a term is a conjunction $t = \ell_1 \wedge \cdots \wedge \ell_k$ of literals. By $P(c)$ and $N(c)$ (resp., $P(t)$ and $N(t)$) we denote the sets of variables occurring positively and negatively in c (resp., t); $\bot$ (resp., $\top$) denotes the empty clause (resp., empty term) representing falsity (resp., truth). A conjunction of clauses $\varphi = \bigwedge_i c_i$ (resp., a disjunction of terms $\varphi = \bigvee_i t_i$) is called a *conjunctive normal form*

(CNF) (resp., *disjunctive normal form* (DNF)).

A *model* is a vector $v \in \{0,1\}^n$, whose i-th component is denoted by v_i, and a *theory* is any set $\Sigma \subseteq \{0,1\}^n$ of models. The models $(0,0,\ldots,0)$ and $(1,1,\ldots,1)$ are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively. We use $v \leq w$ for the usual bitwise ordering of models, i.e., $v_i \leq w_i$ for all $i = 1,\ldots,n$, where $0 \leq 1$. A model $v \in \Sigma$ is *minimal* in Σ, if there is no $w \in \Sigma$ such that $w < v$.

For any formula φ, let $T(\varphi) = \{v \in \{0,1\}^n \mid \varphi(v) = 1\}$ be the set of its models. A formula φ represents a theory Σ if $T(\varphi) = \Sigma$. We often do not distinguish a formula from the theory it represents, if no confusion arises. For formulas φ and ψ, we write $\varphi \leq \psi$, if $T(\varphi) \subseteq T(\psi)$ holds. A nontautological clause c (resp., noncontradictory term t) is an *implicate* (resp., *implicant*) of a theory Σ if $c(v) = 1$ for all $v \in \Sigma$, i.e., $\{0,1\}^n \supset T(c) \supseteq \Sigma$ (resp., $t(v) = 0$ for all $v \notin \Sigma$, i.e., $\emptyset \subset T(t) \subseteq \Sigma$); it is *prime*, if no proper subclause (resp., subterm) is an implicate (resp., implicant) of Σ.

A theory Σ is *Horn* if $\Sigma = Cl_\wedge(\Sigma)$ holds, where $Cl_\wedge(S)$ is the closure of $S \subseteq \{0,1\}^n$ under bitwise AND (i.e., intersection) of models v and w, denoted by $v \bigwedge w$. Observe that any Horn theory Σ has the least (unique smallest) model, which is given by $\bigwedge_{v\in\Sigma} v$. For a Horn theory Σ, a model $v \in \Sigma$ is called *characteristic* [14], if $v \notin Cl_\wedge(\Sigma \setminus \{v\})$ holds. The set of all characteristic models of Σ, the *characteristic set of* Σ, is denoted by $C^*(\Sigma)$. Note that every Horn theory Σ has the unique characteristic set $C^*(\Sigma)$. For example, the theory $\Sigma = \{(0101), (1001), (1000), (0001), (0000)\}$ is Horn, and has $C^*(\Sigma) = \{(0101), (1001), (1000)\}$.

A clause c is *Horn* (resp., *negative, positive*) if $|P(c)| \leq 1$ (resp., $|P(c)| = 0$, $|N(c)| = 0$), and a CNF is *Horn* (resp., *negative, positive*) if it contains only Horn (resp., negative, positive) clauses. It is well-known that a theory Σ is Horn if and only if it is represented by some Horn CNF, and that all prime implicates of a Horn theory are Horn. A theory is *negative* (resp., *positive*), if it is represented by some negative (resp., positive) CNF. Observe that every negative theory is Horn.

A Horn theory Σ_c is a *Horn core* of a theory Σ, if $\Sigma_c \subseteq \Sigma$ holds and no Horn theory Σ' exists such that $\Sigma_c \subset \Sigma' \subseteq \Sigma$. Observe that, in general, Σ has more than one Horn core; e.g., $\Sigma = \{(110), (101)\}$ has two Horn cores $\Sigma_1 = \{(110)\}$ and $\Sigma_2 = \{(101)\}$, respectively. The *Horn envelope* of Σ is the Horn theory $\Sigma_e \supseteq \Sigma$ such that no Horn theory Σ' with $\Sigma_e \supset \Sigma' \supseteq \Sigma$ exists. For the above Σ, its Horn envelope is given by $\Sigma_e = \{(110), (101), (100)\}$. As easily seen, the Horn envelope is always unique (in fact, $\Sigma_e = Cl_\wedge(\Sigma)$). Let φ be a formula representing a theory Σ, and let φ_c and φ_e be Horn formulas representing a Horn core and the Horn envelope of Σ, respectively. Then φ_c and φ_e are also called a *Horn core* and the *Horn envelope* of φ, respectively.

3 Horn Property of the Difference

In this section, we consider the problem of deciding whether the difference $\Sigma_1 \setminus \Sigma_2$ of two Horn theories Σ_1 and Σ_2 is Horn. In Subsection 3.1, we study the problem under Horn CNF representation of Horn theories, and in Subsection 3.2, the model-based representation. As we will see, the problem is tractable in both cases.

3.1 Formula-based representation

Let $\varphi_1 = \bigwedge_{j=1}^{m_1} c_{1,j}$ and $\varphi_2 = \bigwedge_{j=1}^{m_2} c_{2,j}$ be Horn CNFs. Then

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad \bigvee_{j=1}^{m_2} (\varphi_1 \wedge \neg c_{2,j}) \quad = \quad \bigvee_{j=1}^{m_2} (\varphi_1 \wedge t_{2,j}), \tag{3.1}$$

where $t_{2,j}$ is the term equivalent to $\neg c_{2,j}$; e.g., if $c_{2,j} = (\overline{x}_1 \vee \overline{x}_2 \vee x_3)$, then $t_{2,j} = x_1 x_2 \overline{x}_3$. Note that each $\psi_j = \varphi_1 \wedge t_{2,j}$ is a Horn CNF. Hence $\varphi_1 \setminus \varphi_2$ can be represented by the disjunction of m_2 Horn theories ψ_j. It is known [7] that, in general, the problem of checking the Horn property of the disjunction of Horn theories is co-NP-complete. However, we shall show below that the Horn property of the difference of Horn theories can be checked in polynomial time.

If $\varphi_2 \equiv \bot$, then

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad \varphi_1 \tag{3.2}$$

clearly holds, and hence the difference is Horn. Otherwise, we compute the least model v in φ_2, and separately consider the two cases: (1) $\varphi_1(v) = 0$ and (2) $\varphi_1(v) = 1$.

Case (1). Since φ_1 is Horn, there is a Horn clause $c^* \in \varphi_1$ such that $c^* \geq \varphi_1$ and $c^*(v) = 0$. For a model u, let $HC(u)$ denote the set of all maximal Horn clauses c such that $c(u) = 0$. E.g., if $u = (111000)$, then $HC(u) = \{(\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3 \vee x_4), (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3 \vee x_5), (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3 \vee x_6)\}$. Now, some c in $HC(v)$ satisfies $c \geq c^* (\geq \varphi_1)$. We claim that if $v \neq \mathbf{1}$, then

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad \varphi_1 \setminus (\varphi_2 \wedge x_i) \tag{3.3}$$

holds for the x_i which appears in c (i.e., $P(c) = \{x_i\}$). To prove this, note that (3.3) is equivalent to

$$\varphi_1 \wedge \varphi_2 \quad \equiv \quad \varphi_1 \wedge \varphi_2 \wedge x_i, \tag{3.4}$$

which we shall prove. Clearly $\varphi_1 \wedge \varphi_2 \geq \varphi_1 \wedge \varphi_2 \wedge x_i$ holds. For the opposite inequality, since v is the least model of φ_2 and $c \geq \varphi_1$ holds, all models w with $(\varphi_1 \wedge \varphi_2)(w) = 1$ satisfy $w_i = 1$. This means $\varphi_1 \wedge \varphi_2 \leq \varphi_1 \wedge \varphi_2 \wedge x_i$.

On the other hand, if $v = \mathbf{1}$, then $T(\varphi_2) = \{\mathbf{1}\}$ and

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad \varphi_1 \tag{3.5}$$

immediately follows.

Case (2). Clearly $(\varphi_1 \setminus \varphi_2)(v) = 0$ holds. Thus, if $\varphi_1 \setminus \varphi_2$ represents a Horn theory, there exists a Horn clause c^* such that $c^* \geq \varphi_1 \setminus \varphi_2$ and $c^*(v) = 0$. By the definition of $HC(v)$, some c in $HC(v)$ then satisfies $c \geq c^* (\geq \varphi_1 \setminus \varphi_2)$. We claim that if $v \neq \mathbf{1}$, then

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad (\varphi_1 \wedge c) \setminus (\varphi_2 \wedge x_i) \tag{3.6}$$

holds, where the x_i appears in c (i.e., $P(c) = \{x_i\}$). First note that $\varphi_1 \setminus \varphi_2 \equiv (\varphi_1 \wedge c) \setminus \varphi_2$ holds by $c \geq \varphi_1 \setminus \varphi_2$. Thus, (3.6) is equivalent to

$$\varphi_1 \wedge c \wedge \varphi_2 \quad \equiv \quad \varphi_1 \wedge c \wedge \varphi_2 \wedge x_i. \tag{3.7}$$

Then this can be proved similarly to (3.4).

On the other hand, if $v = \mathbf{1}$, then

$$\varphi_1 \setminus \varphi_2 \quad \equiv \quad \varphi_1 \wedge c \tag{3.8}$$

can be proved in a manner similar to (3.5).

Now we can apply the above modifications repeatedly as long as possible. In this process, if $\varphi_2 = \bot$ or $v = \mathbf{1}$ holds, then we can conclude that $\varphi_1 \setminus \varphi_2$ is Horn by (3.2), (3.5) or (3.8). If there is no $c \in HC(v)$ such that $c \geq \varphi_1 \setminus \varphi_2$, then we can conclude that $\varphi_1 \setminus \varphi_2$ is not Horn. In the remaining cases, we apply (3.3) or (3.6) (i.e., φ_1 is modified to $\varphi_1 \wedge c$ in case of (3.6), and φ_2 is modified to $\varphi_2 \wedge x_i$ in both cases).

Let us show that this procedure halts in finitely many steps. Let $v^{(k)}$ be the least model of φ_2 in the k-th iteration. Then we can see that

$$v^{(k)} \quad < \quad v^{(k+1)} \tag{3.9}$$

holds for $k = 1, 2, \ldots$. Indeed, in either of Cases (1) and (2), φ_2 is updated to $\varphi \wedge x_i$ in each iteration, which implies $v^{(k)} \leq v^{(k+1)}$. Moreover, since the x_i that appears in $c \in HC(v^{(k)})$ satisfies $v_i^{(k)} = 0$ by $c(v^{(k)}) = 0$, and $v_i^{(k+1)} = 1$ by $(\varphi_2 \wedge x_i)(v^{(k+1)}) = 1$, we have $v^{(k)} < v^{(k+1)}$. This implies that the number of iterations (i.e., different $v^{(k)}$) is at most $n + 1$.

Formally, the algorithm can be written as follows.

Algorithm CHECK-HORN

Input: Horn CNFs φ_1 and φ_2.

Output: If $\varphi_1 \setminus \varphi_2$ is a Horn theory Σ, then output a Horn CNF for Σ; otherwise, "No".

Step 0. $\phi_1 := \varphi_1$; $\phi_2 := \varphi_2$;

Step 1. **if** $\phi_2 \equiv \bot$ **then** output ϕ_1 and halt
else begin compute the least model v of ϕ_2
if $\phi_1(v) = 0$ **then goto** Step 2
else (i.e., $\phi_1(v) = 1$) **goto** Step 3
end;

Step 2. **if** $v = \mathbf{1}$ **then** output ϕ_1 and halt
else begin
find a Horn clause c in $HC(v)$ such that $c \geq \phi_1$;
$\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$;
goto Step 1
end;

Step 3. Find a Horn clause c in $HC(v)$ such that $c \geq \phi_1 \setminus \phi_2$;
if no such c exists **then** output "No" and halt
else begin $\phi_1 := \phi_1 \wedge c$;
if $v = \mathbf{1}$ **then** output ϕ_1 and halt
else $\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$;
goto Step 1
end. □

Example 3.1 Let us apply CHECK-HORN to $\varphi_1 = (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge (\overline{x}_2 \vee x_4)$ and $\varphi_2 = \overline{x}_4 \wedge (\overline{x}_1 \vee \overline{x}_3) \wedge (\overline{x}_2 \vee \overline{x}_3)$. As we can check, the theory represented by $\varphi_1 \setminus \varphi_2$ is $\Sigma = \{(0111), (1011), (1010), (0011), (0001)\}$. Let $\phi_1 := \varphi_1$ and $\phi_2 := \varphi_2$. Since $\phi_2 \not\equiv \bot$ in Step 1, the least model $v = (0000)$ of ϕ_2 is computed. Since $\phi_1(v) = 1$, we branch to Step 3. Then $HC(v) = \{x_i \mid i = 1, 2, \ldots, 4\}$, and we can check that $\phi_1 \setminus \phi_2 \not\leq x_i$ holds for all $i = 1, 2, \ldots, 4$. Thus, the algorithm outputs "No" in Step 3. Observe that this is correct, since $(1010), (0001) \in \Sigma$ but $(0000) = (1010) \wedge (0001) \notin \Sigma$, which means that Σ is not Horn.

On the other hand, let φ_1 be as above and $\varphi_2' = \overline{x}_4 \wedge (\overline{x}_2 \vee \overline{x}_3)$. In this case, $\varphi_1 \setminus \varphi_2'$ represents $\Sigma' = \{(0111), (1011), (0011), (0001)\}$. Let $\phi_1 := \varphi_1$ and $\phi_2 := \varphi_2'$. Then, in Step 1 the least model is $v = (0000)$, and we again branch to Step 3, where $HC(v)$ is the same as the previous one. Now it holds that $\varphi_1 \setminus \varphi_2' \leq x_4$. Hence, we update ϕ_1 to $\phi_1 = \phi_1 \wedge x_4$ and ϕ_2 to $\phi_2 \wedge x_4$. After returning to Step 1, we find that $\phi_2 \equiv \bot$ now holds.

Hence, the formula

$$\begin{aligned}\phi_1 &= (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge (\overline{x}_2 \vee x_4) \wedge x_4 \\ &= (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge x_4\end{aligned}$$

is output. Observe that Σ' is represented by this ϕ_1, and thus the output is correct. □

Theorem 3.1 *Let φ_1 and φ_2 be Horn CNFs. Then, algorithm* CHECK-HORN *checks whether $\varphi_1 \setminus \varphi_2$ is Horn in $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$ time.*

Proof. The correctness of algorithm CHECK-HORN follows from the discussion before its description. We thus consider its time complexity.

Clearly, Step 0 can be done in $O(|\varphi_1| + |\varphi_2|)$ time. In Step 1, since ϕ_2 is Horn, we can check if $\phi_2 \not\equiv \bot$ (i.e., the satisfiability of ϕ_2) and compute the least model v of ϕ_2 in $O(|\phi_2|)$ time [5]. Then we can compute $\phi_1(v)$ in $O(|\phi_1|)$ time. Note that, in each iteration, ϕ_1 and ϕ_2 might be replaced by $\phi_1 \wedge c$ and $\phi_2 \wedge x_i$, respectively. Since the number of iterations is at most $n+1$, we have $O(|\phi_1|) = O(|\varphi_1| + n^2)$ and $O(|\phi_2|) = O(|\varphi_2| + n)$. Thus, Step 1 requires totally $O(n(|\phi_1| + |\phi_2|)) = O(n^3 + n(|\varphi_1| + |\varphi_2|))$ time.

In Step 2, since ϕ_1 is Horn, we can compute a Horn clause c of the desired property in $O(|\phi_1|) = O(n^2 + |\varphi_1|)$ time. Updating ϕ_2 to $\phi_2 \wedge x_i$ takes $O(1)$ time. Since we have at most n iterations, Step 2 can be executed in $O(n^3 + n|\varphi_1|)$ time.

Let us finally consider Step 3. Note that the difference $\phi_1 \setminus \phi_2$ can be represented by the disjunctions of m_2 Horn CNFs $\psi_1, \psi_2, \ldots, \psi_{m_2}$, where m_2 is the number of clauses in ϕ_2 (see (3.1)). It is easy to see that $c \geq \phi_1 \setminus \phi_2$ if and only if $c \geq \psi_k$ for all k. For each k, we can compute the set S_k of all (candidate) Horn clauses $c \in HC(v)$ such that $c \geq \psi_k$ in $O(|\psi_k|)$ time (the clauses are stored by their positive literals; e.g., we simply keep $S_k = \{x_3, x_4, x_5\}$ instead of $S_k = \{(\overline{x}_1 \vee \overline{x}_2 \vee x_3), (\overline{x}_1 \vee \overline{x}_2 \vee x_4), (\overline{x}_1 \vee \overline{x}_2 \vee x_5)\}$). Thus by taking the intersection of all S_k, we can find a $c \in HC(v)$ of the desired property in

$$O(\textstyle\sum_{k=1}^{m_2} |\psi_k| + nm_2) = O(|\phi_1||\phi_2| + nm_2) = O(n^3 + n^2|\varphi_2| + n|\varphi_1| + |\varphi_1||\varphi_2|) \qquad (3.10)$$

time. Furthermore, updating ϕ_1 and ϕ_2 can be done in $O(n^2 + |\varphi_1| + |\varphi_2|)$ time. Thus Step 3 can be done in $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$ time.

In total, algorithm CHECK-HORN requires $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$ time. □

Remark. Since the formulas ϕ_1 and ϕ_2 monotonically increase (by adding conjuncts), it is possible to use appropriate data structures which support incremental computation. This will lead to savings on the running time of the algorithm.

3.2 Model-based representation

In this section, we consider Horn theories represented by their characteristic sets. As in the case of formula representation, we find that deciding whether the difference $\varphi_1 \setminus \varphi_2$ of two Horn formulas is Horn is solvable in polynomial time.

Let Σ_1 and Σ_2 be Horn theories. Let S be the set of models defined by

$$S \quad = \quad C^*(\Sigma_1) \cup \{v \bigwedge w \mid v, w \in C^*(\Sigma_1)\}. \tag{3.11}$$

That is, S augments $C^*(\Sigma_1)$ by one step of the closure operator $Cl_\wedge(\cdot)$. We split S into S_1 and S_2 as follows:

$$\begin{aligned} S_1 \ \cap \ \Sigma_2 \ &= \ \emptyset \\ S_2 \ \subseteq \ \Sigma_2. & \end{aligned} \tag{3.12}$$

It is easy to see that

$$S_1 \quad \subseteq \quad \Sigma_1 \setminus \Sigma_2 \tag{3.13}$$

$$Cl_\wedge(S_2) \quad \subseteq \quad \Sigma_2. \tag{3.14}$$

We furthermore have the following lemma, where a similar result is obtained in [7].

Lemma 3.2 *For a Horn theory Σ, let $S = C^*(\Sigma) \cup \{v \bigwedge w \mid v, w \in C^*(\Sigma)\}$, and let S_1 and S_2 be sets of models such that $S_1 \cup S_2 = S$. Then*

$$Cl_\wedge(S_1) \cup Cl_\wedge(S_2) = \Sigma. \tag{3.15}$$

Proof. Since $S_1, S_2 \subseteq \Sigma$ and Σ is Horn, $Cl_\wedge(S_1) \cup Cl_\wedge(S_2) \subseteq \Sigma$ holds.

For the converse inclusion, assume that there exists a model $v \in \Sigma \setminus (Cl_\wedge(S_1) \cup Cl_\wedge(S_2))$. Then v can be represented by $v = \bigwedge_{w \in Q_1} w \wedge \bigwedge_{u \in Q_2} u$ for some $Q_1 \subseteq S_1 \cap C^*(\Sigma)$ and $Q_2 \subseteq S_2 \cap C^*(\Sigma)$. If $Q_1 = \emptyset$ (resp., $Q_2 = \emptyset$), then $v \in Cl_\wedge(S_2)$ (resp., $v \in Cl_\wedge(S_1)$) holds, which is a contradiction. Thus $Q_1, Q_2 \neq \emptyset$. For a model $w \in Q_1$, let $\Sigma_w = \{w \wedge u \mid u \in Q_2\}$. By the definition of S, $\Sigma_w \subseteq S$ holds for all w. Assume that some $w^* \in Q_1$ satisfies $\Sigma_{w^*} \subseteq S_1$. Then, since $Q_1, \Sigma_{w^*} \subseteq S_1$, and since v can be represented by $v = \bigwedge_{w \in Q_1} w \wedge \bigwedge_{u \in \Sigma_{w^*}} u$, we have $v \in Cl_\wedge(S_1)$, a contradiction. Thus, for every $w \in Q_1$, there is a model $u^{(w)} \in \Sigma_w$ such that $u^{(w)} \in S_2$. We can see that $v = \bigwedge_{w \in Q_1} w \wedge \bigwedge_{u \in Q_2} u = \bigwedge_{w \in Q_1} u^{(w)} \wedge \bigwedge_{u \in Q_2} u$ holds. Since this says $v \in Cl_\wedge(S_2)$, we again have a contradiction. □

By this lemma,

$$Cl_\wedge(S_1) \ \cup \ Cl_\wedge(S_2) \quad = \quad \Sigma_1 \tag{3.16}$$

holds for the S_1 and S_2 of (3.12).

The next lemma leads then to a polynomial time algorithm for our problem.

Lemma 3.3 *Let Σ_1 and Σ_2 be Horn theories, and let S_1 be the set of models defined by (3.12). Then $Cl_\wedge(S_1) \cap \Sigma_2 = \emptyset$ holds if and only if $\Sigma_1 \setminus \Sigma_2$ is a Horn theory. Furthermore, if $\Sigma_1 \setminus \Sigma_2$ is a Horn theory, then $Cl_\wedge(S_1) = \Sigma_1 \setminus \Sigma_2$ (i.e., $C^*(S_1) = C^*(\Sigma_1 \setminus \Sigma_2)$) holds.*

Proof. If $\Sigma_1 \setminus \Sigma_2$ is Horn, then $S_1 \subseteq \Sigma_1 \setminus \Sigma_2$ implies $Cl_\wedge(S_1) \subseteq \Sigma_1 \setminus \Sigma_2$. Hence $Cl_\wedge(S_1) \cap \Sigma_2 = \emptyset$ holds.

On the other hand, if $Cl_\wedge(S_1) \cap \Sigma_2 = \emptyset$, then $Cl_\wedge(S_1) \cap Cl_\wedge(S_2) = \emptyset$ holds, since $Cl_\wedge(S_2) \subseteq \Sigma_2$. This, combined with (3.16), implies that $(Cl_\wedge(S_1), Cl_\wedge(S_2))$ forms a partition of Σ_1. Moreover, it follows from (3.14) and the assumption $Cl_\wedge(S_1) \cap \Sigma_2 = \emptyset$ that

$$Cl_\wedge(S_1) \quad = \quad \Sigma_1 \setminus \Sigma_2 \tag{3.17}$$

holds. Hence $\Sigma_1 \setminus \Sigma_2$ is Horn. The second part of the lemma also follows from (3.17). □

Theorem 3.4 *Let Σ_1 and Σ_2 be Horn theories. Given $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$, we can check whether $\Sigma_1 \setminus \Sigma_2$ is Horn in $O(n|C^*(\Sigma_1)|^2|C^*(\Sigma_2)|)$ time. Furthermore, if $\Sigma_1 \setminus \Sigma_2$ is Horn, $C^*(\Sigma_1 \setminus \Sigma_2)$ can be computed in $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2)|))$ time.*

Proof. For the first statement, by Lemma 3.3, we only check if $Cl_\wedge(S_1) \cap \Sigma_2 = \emptyset$, where S_1 is given by (3.12). Let us first consider the time complexity to construct S_1. By the definition (3.11) of S, S can be constructed from $C^*(\Sigma_1)$ in $O(n|C^*(\Sigma_1)|^2)$ time. For each $v \in S$, we can check if $v \in S_1$ by checking whether

$$v \quad = \quad \bigwedge_{w \in C^*(\Sigma_2): w \geq v} w \tag{3.18}$$

holds. This can be done in $O(n|C^*(\Sigma_2)|)$ time. Since $|S| \leq |C^*(\Sigma_1)|^2$, therefore, S_1 can be constructed in $O(n|C^*(\Sigma_1)|^2|C^*(\Sigma_2)|)$ time.

It is known [6] that the condition $Cl_\wedge(Q_1) \cap Cl_\wedge(Q_2) = \emptyset$ can be checked from Q_1 and Q_2 in $O(n(|Q_1|+|Q_2|))$ time. Based on this, the condition $Cl_\wedge(S_1) \cap \Sigma_2$ $(= Cl_\wedge(S_1) \cap Cl_\wedge(\Sigma_2)) = \emptyset$ can be checked from S_1 and $C^*(\Sigma_2)$ in $O(n(|S_1| + |C^*(\Sigma_2)|)) = O(n(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2)|))$ time.

Therefore, in total, the algorithm requires $O(n|C^*(\Sigma_1)|^2|C^*(\Sigma_2)|)$ time.

For the second statement, by Lemma 3.3, $C^*(S_1) = C^*(\Sigma_1 \setminus \Sigma_2)$ holds. As already shown, S_1 can be constructed in $O(n|C^*(\Sigma_1)|^2|C^*(\Sigma_2)|)$ time. Furthermore, $C^*(S_1)$ can be computed from S_1 in $O(n|S_1|^2) = O(n|C^*(\Sigma_1)|^4)$ time, since $|S_1| \leq |\Sigma_1|^2$ holds. Thus, $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2)|))$ time is required in total. □

We close this section with a remark on an important special case of $\Sigma_1 = \{0,1\}^n$, i.e., Σ_1 contains all models. In [9], the class of double Horn functions has been introduced, which

are those functions f such that both f and its complement $\overline{f}$ are Horn.[1] In this case, we obtain the following relation.

Proposition 3.5 *Let $\Sigma_1 = \{0,1\}^n$, and let Σ_2 be an arbitrary Horn theory. Then the difference $\Sigma_1 \setminus \Sigma_2$ is Horn if and only if Σ_2 is double Horn.* □

By this proposition and the results in [9], if Σ_1 is fixed to $\{0,1\}^n$ (and thus it may be omitted from the input), then the problem of checking whether $\Sigma_1 \setminus \Sigma_2$ is Horn can be solved in linear time under formula-based representation, and, as a consequence of results in [9], in $O(n^2|C^*(\Sigma_2)|)$ time under model-based representation.

4 Horn Cores

In this section, we investigate computing Horn cores of the difference of Horn theories. Since in general, multiple Horn cores can exist, several variants of this task are of interest. One is computing some Horn core of the difference, while the other is the enumeration of all Horn cores.

A further interesting problem is whether the difference has a unique Horn core. As it is easy to see that the difference has a unique Horn core if and only if it is Horn, the algorithms from the previous section can be used to solve this problem. As previously, we first consider the problems under Horn CNFs, and then turn to model-based representation.

4.1 Formula-based representation

We first consider the problem of computing one Horn core of the difference $\varphi_1 \setminus \varphi_2$, where φ_1 and φ_2 are Horn CNFs. The algorithm is a modification of algorithm CHECK-HORN, which checks whether the difference $\varphi_1 \setminus \varphi_2$ is Horn.

In Step 3 of algorithm CHECK-HORN, if there is no Horn clause c in $HC(v)$ such that $c \geq \phi_1 \setminus \phi_2$, we conclude that the difference is not Horn and halt. To compute a Horn core, however, we update ϕ_1 and ϕ_2 as $\phi_1 := \phi_1 \wedge c$ and $\phi_2 := \phi_2 \wedge x_i$, respectively, for an *appropriate* $c \in HC(v)$. This is because no Horn core contains the model v.

Let us consider how to obtain the above appropriate $c \in HC(v)$. For a formula ψ and a model w, let $\min_{\geq w}(\psi)$ denote the set of minimal models u such that $\psi(u) = 1$ and $u \geq w$. We claim that a clause $c \in HC(v)$ such that $c(u) = 1$ for some $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ is

[1] Actually, in [9], DNF Horn functions have been considered, which are functions $f = \bigvee_i t_i$ such that each t_i is a Horn term, i.e., contains at most one negative literal. Clearly, the double Horn functions based on DNF and CNF representation coincide.

appropriate. The proof for this claim will be given (in different form) after presenting the algorithm and an example.

For models v, u with $v \leq u$, let $HC(v; u)$ denote the set of Horn clauses $c \in HC(v)$ whose positive literal x_j satisfies $u_j = 1$. For example, if $v = (111000)$ and $u = (111110)$, then $HC(v; u) = \{(\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3 \vee x_4), (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3 \vee x_5)\}$. To make the discussion clear and speed up the algorithm in practice, based on a model $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$, we update ϕ_1 and ϕ_2 respectively by

$$\phi_1 := \phi_1 \wedge \bigwedge_{c \in HC(v;u)} c, \tag{4.19}$$

$$\phi_2 := \phi_2 \wedge \bigwedge_{u_j = 1 \wedge v_j = 0} x_j. \tag{4.20}$$

Thus, we have the following algorithm.

Algorithm HORN-CORE1

Input: Horn CNFs φ_1 and φ_2.

Output: A Horn core of $\varphi_1 \setminus \varphi_2$.

Step 0. $\phi_1 := \varphi_1$; $\phi_2 := \varphi_2$;

Step 1. **if** $\phi_2 \equiv \bot$ **then** output ϕ_1 and halt
else begin compute the least model v of ϕ_2
if $\phi_1(v) = 0$ **then goto** Step 2
else (i.e., $\phi_1(v) = 1$) **goto** Step 3
end;

Step 2. **if** $v =$ **1 then** output ϕ_1 and halt
else begin
find a Horn clause c in $HC(v)$ such that $c \geq \phi_1$;
$\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$;
goto Step 1
end;

Step 3. **if** there exists a Horn clause c in $HC(v)$ such that $c \geq \phi_1 \setminus \phi_2$
then begin $\phi_1 := \phi_1 \wedge c$;
if $v =$ **1 then** output ϕ_1 and halt
else $\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$
end
else begin find a model $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$;

$\phi_1 := \phi_1 \wedge \bigwedge_{c \in HC(v;u)} c$;

$\phi_2 := \phi_2 \wedge \bigwedge_{u_j = 1 \wedge v_j = 0} x_j$

end

goto Step 1. □

Example 4.1 Let us reconsider the Horn CNFs φ_1 and φ_2 in Example 3.1. In Step 3 of CHECK-HORN, $\phi_1 \setminus \phi_2 \leq x_i$ was found false for all $i = 1, 2, \ldots, 4$. We thus choose in Step 3 of HORN-CORE1 a model u in $\min_{\geq v}(\phi_1 \setminus \phi_2)$. Since $v = (0000)$, $\min_{\geq v}(\phi_1 \setminus \phi_2)$ is the set of the minimal models in the theory Σ represented by $\phi_1 \setminus \phi_2$ (see Example 3.1). The model $u = (0001)$ is such a minimal model. Then, $HC(v; u) = \{x_4\}$, and we update

$$\begin{aligned} \phi_1 &:= (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge (\overline{x}_2 \vee x_4) \wedge x_4, \\ \phi_2 &:= \overline{x}_4 \wedge (\overline{x}_1 \vee \overline{x}_3) \wedge (\overline{x}_2 \vee \overline{x}_3) \wedge x_4. \end{aligned}$$

Returning to Step 1 of HORN-CORE1, we now find that $\phi_2 \equiv \bot$ is true; hence, the current ϕ_1 is output.

Observe that the output of CHECK-HORN for φ_1 and φ_2' in Example 3.1 is the same as the output ϕ_1. The formula $\varphi_1 \setminus \varphi_2$ and the output formula ϕ_1 represent the theories $\Sigma = \{(0111), (1011), (1010), (0011), (0001)\}$ and $\Sigma' = \{(0111), (1011), (0011), (0001)\}$, respectively. It is easily checked that the output is indeed a Horn core of $\varphi_1 \setminus \varphi_2$. □

Let $\phi_1^{(k)}$ and $\phi_2^{(k)}$ respectively denote the ϕ_1 and ϕ_2 in Step 1 of the k-th iteration, and let $v^{(k)}$ denote the least model in $\phi_2^{(k)}$. Then we claim that

$$v^{(k)} \quad < \quad v^{(k+1)} \tag{4.21}$$

holds for $k = 1, 2, \ldots$, which implies that the number of iterations is at most $n + 1$.

In the case that either Step 2 or the if-condition of Step 3 was reached during the k-th iteration, the proof is similar to (3.9). Thus, assume that the else-condition of Step 3 was reached during the k-th iteration. Let $u^{(k)} \in \min_{\geq v^{(k)}}(\phi_1^{(k)} \setminus \phi_2^{(k)})$ be chosen in algorithm HORN-CORE1. Then we show that

$$v^{(k)} \quad < \quad u^{(k)} \quad < \quad v^{(k+1)}, \tag{4.22}$$

which completes the proof of (4.21). It follows from the definition of $u^{(k)}$ that $u^{(k)} \geq v^{(k)}$ holds. Moreover, $\phi_2^{(k)}(v^{(k)}) = 1$ and $\phi_2^{(k)}(u^{(k)}) = 0$ imply $u^{(k)} \neq v^{(k)}$. We thus have $u^{(k)} > v^{(k)}$. As for the second inequality of (4.22), since $v^{(k)}$ is the least model in $\phi_2^{(k)}$, and $\phi_2^{(k)}$ is updated as (4.20), $w \geq u^{(k)}$ holds for all models w such that $\phi_2^{(k+1)}(w) = 1$ (In particular

for $w = v^{(k+1)}$). Moreover, since $\phi_2^{(k+1)} \leq \phi_2^{(k)}$ and $\phi_2^{(k)}(u^{(k)}) = 0$, $\phi_2^{(k+1)}(u^{(k)}) = 0$ holds, which implies $u^{(k)} \neq v^{(k+1)}$. Hence $u^{(k)} < v^{(k+1)}$ holds.

We next present some properties of $\phi_1^{(k)}$ without their proofs, since they can be easily found.

Lemma 4.1 *Let $\phi_1^{(k)}$ and $\phi_2^{(k)}$ be defined as above. Then $\phi_1^{(k)} \leq \varphi_1$ and $\phi_2^{(k)} \leq \varphi_2$ hold.*

Lemma 4.2 *Let $v^{(k)}$, $\phi_1^{(k)}$ and $\phi_2^{(k)}$ be defined as above.*

(*i*) *If Step 2 was reached during the k-th iteration, then $\phi_1^{(k+1)}(w) = \phi_1^{(k)}(w)$ holds for all models w.*

(*ii*) *If the if-condition of Step 3 was reached during the k-th iteration, then $\phi_1^{(k+1)}(w) = \phi_1^{(k)}(w)$ holds for all models w that satisfy either $w \not\geq v^{(k)}$ or $w \geq v^{(k+1)}$.*

(*iii*) *If the else-condition of Step 3 was reached during the k-th iteration, then $\phi_1^{(k+1)}(w) = \phi_1^{(k)}(w)$ holds for all models w that satisfy either $w \not\geq v^{(k)}$ or $w \geq u^{(k)}$, and $\phi_1^{(k+1)}(w) = 0$ holds for all models w that satisfy $v^{(k)} \leq w < u^{(k)}$, where $u^{(k)} \in \min_{\geq v^{(k)}}(\phi_1^{(k)} \setminus \phi_2^{(k)})$ is chosen in algorithm* HORN-CORE1.

By Lemma 4.2 (iii), if the else-condition of Step 3 was reached during the k-th iteration, then $\phi_1^{(k+1)}(u^{(k)}) = \phi_1^{(k)}(u^{(k)})$ holds. It follows from (4.21) and (4.22) that $u^{(k)} < v^{(l)}$ holds for all $l \geq k+1$. Since $\phi_1^{(k)}(u^{(k)}) = 1$, Lemma 4.2 implies that the final output in the h-th iteration, $\varphi_1^{(h)}$, also satisfies $\phi_1^{(h)}(u^{(k)}) = 1$.

Let us next consider a model $u' \in \min_{\geq v^{(k)}}(\phi_1^{(k)} \setminus \phi_2^{(k)})$ with $u' \neq u^{(k)}$. Since $u^{(k)}, u' \geq v^{(k)}$ and $u' \not\geq u^{(k)}$, there is an index j such that $v_j^{(k)} = u'_j = 0$ and $u_j^{(k)} = 1$. This means

$$\phi_1^{(k+1)}(u') \quad = \quad 0. \tag{4.23}$$

Moreover, since $u^{(k)} < v^{(k+1)}$ (by (4.22)) and $u' \not\geq u^{(k)}$, we have

$$u' \quad \not\geq \quad v^{(k+1)}. \tag{4.24}$$

By (4.23) and (4.24), Lemma 4.2 implies that the output $\phi_1^{(h)}$ satisfies $\phi_1^{(h)}(u') = 0$ for all models $u' \in \min_{\geq v^{(k)}}(\phi_1^{(k)} \setminus \phi_2^{(k)})$ with $u' \neq u^{(k)}$.

Therefore, the choice of u in the else-statement of Step 3 is in fact a choice of a Horn core $\psi\,(= \phi_1^{(h)})$ of $\varphi_1 \setminus \varphi_2$ among all possible cores by imposing $\psi(u) = 1$.

Theorem 4.3 *Let φ_1 and φ_2 be Horn CNFs. Then algorithm* HORN-CORE1 *outputs a Horn core of $\varphi_1 \setminus \varphi_2$ in $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$ time.*

Proof. Let us first prove that the output Horn CNF ϕ_1 $(= \phi_1^{(h)})$ satisfies $\phi_1^{(h)} \leq \varphi_1 \setminus \varphi_2$. To prove this, it is sufficient to show that $\phi_1^{(k)} \setminus \phi_2^{(k)} \leq \varphi_1 \setminus \varphi_2$ holds for all k. By the discussion before Algorithm CHECK-HORN, if either Step 2 or the if-statement of Step 3 was reached during the k-th iteration, then $\phi_1^{(k+1)} \setminus \phi_2^{(k+1)} = \phi_1^{(k)} \setminus \phi_2^{(k)}$ holds.

On the other hand, if the else-statement of Step 3 was reached during the k-th iteration, we have $\phi_1^{(k+1)} \setminus \phi_2^{(k+1)} \leq \phi_1^{(k)} \setminus \phi_2^{(k)}$. Indeed, as $\phi_2^{(k+1)} = \phi_2^{(k)} \wedge \bigwedge_{u_j^{(k)}=1 \wedge v_j^{(k)}=0} x_j$ and $v^{(k)}$ is the least model of $\phi_2^{(k)}$, $\phi_2^{(k+1)}(w) = \phi_2^{(k)}(w)$ holds for all w that satisfy either $w \not\geq v^{(k)}$ or $w \geq u^{(k)}$. This combined with Lemma 4.2(iii) implies our claim.

Let us next show $\phi_1^{(h)}$ is a Horn core of $\varphi_1 \setminus \varphi_2$. Assume the contrary, i.e., there is a Horn core ψ such that $\psi > \phi_1^{(h)}$. Let w be a minimal model such that $\psi(w) = 1$ and $\phi_1^{(h)}(w) = 0$. Then $\phi_1^{(1)}(w) = \varphi_1(w) = 1$ holds. By this and Lemma 4.1, there exists a k such that $\phi_1^{(k)}(w) = 1$ and $\phi_1^{(k+1)}(w) = 0$ holds. Moreover, we have $\phi_2^{(k)}(w) = 0$. This follows from $\varphi_2(w) = 0$ (by $\psi(w) = 1$) and Lemma 4.1. Thus $(\phi_1^{(k)} \setminus \phi_2^{(k)})(w) = 1$ holds. This implies that $\phi_1^{(k)}$ was modified in the else-statement of Step 3, in which $w > v^{(k)}$ and $w \not\geq u^{(k)}$ hold by Lemma 4.2 (iii). These facts, together with $u^{(k)} > v^{(k)}$, imply

$$u^{(k)} > (w \wedge u^{(k)}) \geq v^{(k)}.$$

Since $\psi(u^{(k)}) = \psi(w) = 1$, $\psi(w \wedge u^{(k)}) = 1$ must hold. However, $\phi_1^{(k)}(w \wedge u^{(k)}) = 0$ holds by Lemma 4.2 (iii), and hence $\phi_1^{(h)}(w \wedge u^{(k)}) = 0$ follows from Lemma 4.2 and (4.21). This contradicts $\psi > \phi_1^{(h)}$. Thus, algorithm HORN-CORE1 correctly outputs a Horn core of $\varphi_1 \setminus \varphi_2$.

For the time complexity, we only consider how to compute a model $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$, because the rest is similar to algorithm CHECK-HORN. Note that $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ holds if and only if $u \in \min((\phi_1 \setminus \phi_2) \wedge \bigwedge_{v_j=1} x_j)$. Let the CNF of ϕ_2 be $\phi_2 = \bigwedge_{j=1}^{m} c_j$. Then, similarly to (3.1),

$$(\phi_1 \setminus \phi_2) \wedge \bigwedge_{v_j=1} x_j \quad = \quad \bigvee_{j=1}^{m} (\phi_1 \wedge t_j \wedge \bigwedge_{v_j=1} x_j), \tag{4.25}$$

where t_j is the term equivalent to $\overline{c}_j$. Note that each $\phi_1 \wedge t_j \wedge \bigwedge_{v_j=1} x_j$ is a Horn CNF and (4.25) is a disjunction of m Horn CNFs. Since any Horn theory has the least model, (4.25) has at most m minimal models. Since each disjunct in (4.25) has size $O(|\phi_1| + n)$, all candidates for u (the least models in $\phi_1 \wedge t_j \wedge \bigwedge_{v_j=1} x_j$, $j = 1, 2, \ldots, m$) can be computed in $O(m(|\phi_1| + n))$ time [5]. Since finding a minimal model u among all such candidates can be done in $O(nm)$ time (i.e., linear time) by counting zero-bits, we can compute a model $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ in $O(m(|\phi_1| + n) + nm) = O(m(|\phi_1| + n))$ time. Similarly to algorithm

CHECK-HORN, we have $|\phi_1| \leq |\varphi_1| + n^2$ and $m \leq m_2 + n$, where m_2 is the number of clauses in φ_2. Thus $O(n^3 + n^2|\varphi_2| + n|\varphi_1| + |\varphi_1||\varphi_2|)$ time is required in this part.

Since the rest is similar to algorithm CHECK-HORN, and the number of iterations is at most $n+1$, algorithm HORN-CORE1 requires $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$ time in total. □

Let us next consider the problem of computing all Horn cores of $\varphi_1 \setminus \varphi_2$. We observe that in general, an exponential number of Horn cores might exist. This is shown by the following example.

Example 4.2 Let $\varphi_1 = \top$ and $\varphi_2 = \bigwedge_{i=1}^{n}[(\overline{x}_i \vee y_i) \wedge (x_i \vee \overline{y}_i)]$, i.e., φ_2 states that x_i is equivalent to y_i for all $i = 1, 2, \ldots, n$. Then $\varphi_1 \setminus \varphi_2$ is described by the formula

$$\bigvee_{i=1}^{n} (x_i \overline{y}_i \vee \overline{x}_i y_i), \tag{4.26}$$

which can be rewritten to the CNF

$$\varphi = \bigwedge_{S \subseteq \{1,2,\ldots,n\}} \Big(\bigvee_{i \in S} (x_i \vee y_i) \vee \bigvee_{j \in \overline{S}} (\overline{x}_j \vee \overline{y}_j) \Big), \tag{4.27}$$

where $\overline{S} = \{1, 2, \ldots, n\} \setminus S$. Consider then the Horn CNF

$$\psi = (\overline{x}_1 \vee \overline{y}_1 \vee \cdots \vee \overline{x}_n \vee \overline{y}_n) \wedge \bigwedge_{i=0}^{n-1} (\overline{x}_1 \vee \overline{y}_1 \vee \cdots \vee \overline{x}_i \vee \overline{y}_i \vee x_{i+1}), \tag{4.28}$$

where $(\overline{x}_1 \vee \overline{y}_1 \vee \cdots \vee \overline{x}_i \vee \overline{y}_i \vee x_{i+1}) = x_1$ if $i = 0$. Observe that every clause in φ contains some clause in ψ; hence, $\psi \leq \varphi$ holds. Moreover, ψ is a Horn core of φ. Indeed, suppose some Horn ψ' exists such that $\psi < \psi' \leq \varphi$. Then there exists a model v such that $\psi'(v) = \varphi(v) = 1$ but $\psi(v) = 0$. The latter implies that there is some $i \leq n-1$ such that v has value 1 on all components corresponding to x_j, y_j for $j = 1, 2, \ldots, i$ and value 0 for the component of x_{i+1}. Let w be the model which has value 1 on the components x_j, y_j, $j = 1, 2, \ldots, i$ and x_{i+1}, and has value 0 on all other components. Then, we have $\psi(w) = \psi'(w) = 1$, and since ψ' is Horn, $\psi'(v \wedge w) = 1$ follows. However, $\varphi(v \wedge w) = 0$ holds, since x_j and y_j of $v \wedge w$ are 1 for $j = 1, 2, \ldots, i$, and 0 for $j = i+1, i+2, \ldots, n$. This contradicts $\psi' \leq \varphi$. Hence, the ψ' as hypothesized does not exist, which means that ψ is a Horn core of φ.

Since x_{i+1} and y_{i+1} are symmetric in φ, it follows that we may replace in ψ arbitrarily the (unique) positive occurrence of x_i by y_i, and we obtain another Horn core of φ. Consequently φ has at least 2^n Horn cores. □

The previous example shows that computing all Horn cores is not always possible in polynomial time. However, this still leaves hope for a *polynomial total time* algorithm (called *output-polynomial* in [12]), i.e., which runs in polynomial time in the combined size of the input and the output. We will now show that such an algorithm does exist, and that even the enumeration of all Horn cores with polynomial delay [12] is possible.

Recall that in HORN-CORE1, the choice of $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ in the else-statement of Step 3 results in a Horn core ψ of $\varphi_1 \setminus \varphi_2$ such that $\psi(u) = 1$ and $\psi(u') = 0$ for all other models $u' \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ (see the discussion below Lemma 4.2). We show below that every Horn core ψ satisfies $\psi(u) = 1$ for some u. This means that every Horn core can be constructed by algorithm HORN-CORE1, if it properly chooses a model u in the else-statement of Step 3.

Let ψ be a Horn core of $\varphi_1 \setminus \varphi_2$. We execute algorithm HORN-CORE1 according to this ψ. That is, if the else-condition of Step 3 holds in some iteration, we find some $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ such that $\psi(u) = 1$, and then update ϕ_1 and ϕ_2 according to such a u. We now show that such a u can always be found.

Assume that such a u cannot be found in some iteration k; in what follows, we suppress superscripts $^{(k)}$ for convenience. Then $\phi_2 \leq \varphi_2$ holds by Lemma 4.1, and $\phi_1(w) = \varphi_1(w)$ holds for all $w \geq v$ by Lemma 4.2 and (4.21). These facts imply that

$$(\varphi_1 \setminus \varphi_2)(w) \quad \leq \quad (\phi_1 \setminus \phi_2)(w) \tag{4.29}$$

holds for all $w \geq v^{(k)}$. Since ψ is Horn, all models w such that $\psi(w) = 1$ and $w \geq v$ must satisfy $w \geq u$ for some $u \in \min_{\geq v}(\varphi_1 \setminus \varphi_2)$, and hence (4.29) implies

$$w \quad \geq \quad u \tag{4.30}$$

for some $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$. For such a u, we have $\psi \leq \bigwedge_{c \in HC(v;u)} c$, because any model w with $v \leq w < u$ satisfies $(\varphi_1 \setminus \varphi_2)(w) \leq (\phi_1 \setminus \phi_2)(w) = 0$, and hence $\psi(w) = 0$. Based on this u, we can update ϕ_1 and ϕ_2. Then continuing HORN-CORE1, we obtain a Horn core ψ' such that $\psi' \geq \psi$. Since ψ was a Horn core, this implies that $\psi' = \psi$ and hence $\psi(u) = 1$. It follows that a u as hypothesized can be found.

Thus, all Horn cores can be generated as follows.

Algorithm ALL-HORN-CORES(φ_1, φ_2)

Input: Horn CNFs φ_1 and φ_2.

Output: All Horn cores of $\varphi_1 \setminus \varphi_2$.

Step 0. $\phi_1 := \varphi_1$; $\phi_2 := \varphi_2$;

Step 1. **if** $\phi_2 \equiv \bot$ **then** output ϕ_1 and halt
else begin compute the least model v of ϕ_2
if $\phi_1(v) = 0$ **then goto** Step 2
else (i.e., $\phi_1(v) = 1$) **goto** Step 3
end;

Step 2. **if** $v =$ **1 then** output ϕ_1 and halt
else begin
find a Horn clause c in $HC(v)$ such that $c \geq \phi_1$;
$\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$;
goto Step 1
end;

Step 3. **if** there exists a Horn clause c in $HC(v)$ such that $c \geq \phi_1 \setminus \phi_2$
then begin $\phi_1 := \phi_1 \wedge c$;
if $v =$ **1 then** output ϕ_1 and exit;
else begin $\phi_2 := \phi_2 \wedge x_i$, where $P(c) = \{x_i\}$; **goto** Step 1 **end**
end
else for each a model $u \in \min_{\geq v}(\phi_1 \setminus \phi_2)$ **do begin**
$\phi_1 := \phi_1 \wedge \bigwedge_{c \in HC(v;u)} c$;
$\phi_2 := \phi_2 \wedge \bigwedge_{u_j = 1 \wedge v_j = 0} x_j$;
call ALL-HORN-CORES(ϕ_1,ϕ_2)
end{for}. □

Example 4.3 Consider again the Horn CNFs φ_1 and φ_2 from Examples 3.1 and 4.1. In Example 4.1, HORN-CORE1 found one Horn core $\psi_1 = (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge (\overline{x}_2 \vee x_4) \wedge x_4$ of $\varphi_1 \setminus \varphi_2$, based on the choice of $u = (0001)$ from $\min_{\geq v}(\phi_1 \setminus \phi_2)$. This ψ_1 is output by ALL-HORN-CORES. However, in Step 3, there is another model $u = (1010)$ in $\min_{\geq v}(\phi_1 \setminus \phi_2)$. Then, based on this $v = (0000)$, we have $H(v;u) = \{x_1, x_3\}$, and ϕ_1 and ϕ_2 are updated to

$$\begin{aligned} \phi_1 &= (\overline{x}_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_3) \wedge (\overline{x}_2 \vee x_3) \wedge (\overline{x}_2 \vee x_4) \wedge x_1 \wedge x_3 \;=\; x_1 \wedge \overline{x}_2 \wedge x_3, \\ \phi_2 &= \overline{x}_4 \wedge (\overline{x}_1 \vee \overline{x}_3) \wedge (\overline{x}_2 \vee \overline{x}_3) \wedge x_1 \wedge x_3 \;=\; \bot, \end{aligned}$$

respectively. Thus, in the recursive call of ALL-HORN-CORES, $\psi_2 = x_1 \wedge \overline{x}_2 \wedge x_3$ is also output before exist from the entire procedure. Observe that ψ_2 represents the theory Π_2 = {(1010), (1011)}, which is indeed a Horn core of Σ = {(0111), (1011), (1010), (0011), (0001)}. □

Theorem 4.4 *Algorithm* ALL-HORN-CORES *correctly generates all Horn cores of* $\varphi_1 \setminus \varphi_2$ *with polynomial delay, where the delay is bounded by the time of computing one Horn core, i.e.,* $O(n^4 + n^3|\varphi_2| + n^2|\varphi_1| + n|\varphi_1||\varphi_2|)$.

Proof. The correctness of algorithm ALL-HORN-CORES follows from the discussion before the algorithm, and its time complexity can be analyzed similarly to the proof of Theorem 4.3. □

4.2 Model-based representation

In Subsection 3.2, we have seen that S_1 of (3.12) characterizes the difference $\Sigma_1 \setminus \Sigma_2$ of two Horn theories Σ_1 and Σ_2. In this subsection, we show that any maximal set $Q \subseteq S_1$ such that $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$ gives a Horn core of $\Sigma_1 \setminus \Sigma_2$, and conversely any Horn core of $\Sigma_1 \setminus \Sigma_2$ can be generated from such a maximal set $Q \subseteq S_1$.

Note that, by the definition of S_1, any maximal set $Q \subseteq S$ with $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$ is contained in S_1. Thus we prove the above statements by using a set S instead of S_1.

Lemma 4.5 *Let* Σ_1 *and* Σ_2 *be Horn theories, and let* Q *be a maximal set* $Q \subseteq S$ *such that* $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$, *where* S *is given by* (3.11). *Then* $Cl_\wedge(Q)$ *is a Horn core of* $\Sigma_1 \setminus \Sigma_2$.

Proof. Since $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$, there exist a Horn core Π of $\Sigma_1 \setminus \Sigma_2$ such that $\Pi \supseteq Cl_\wedge(Q)$. Assume that there exists a model $v \in \Pi \setminus Cl_\wedge(Q)$. Since $v \in \Sigma_1$, v can be represented by

$$v \;=\; \bigwedge_{u \in Q_1} u \wedge \bigwedge_{w \in Q_2} w \tag{4.31}$$

for some $Q_1 \subseteq Q \cap C^*(\Sigma_1)$ and $Q_2 \subseteq (S \setminus Q) \cap C^*(\Sigma_1)$.

If $Q_2 = \emptyset$, then $v \in Cl_\wedge(Q)$, which is a contradiction. If $Q_1 = \emptyset$, then the maximality of Q implies that, for each $w \in Q_2$, there exists $Q_w \subseteq Q$ such that

$$\bigwedge_{u \in Q_w} u \;\wedge\; w \;\in\; \Sigma_2.$$

Let $y^{(w)} = \bigwedge_{u \in Q_w} u$ and $z^{(w)} = \bigwedge_{u \in Q_w} u \wedge w$. Then, since Σ_2 is Horn and $Q_1 = \emptyset$, we have

$$\bigwedge_{w \in Q_2} z^{(w)} \;=\; \bigwedge_{w \in Q_2} y^{(w)} \wedge \bigwedge_{w \in Q_2} w \;=\; \bigwedge_{w \in Q_2} y^{(w)} \wedge v \;\in\; \Sigma_2. \tag{4.32}$$

Note that $y^{(w)} \in Cl_\wedge(Q)$ holds for all $w \in Q_2$. Thus $\bigwedge_{w \in Q_2} y^{(w)}$ is also contained in $Cl_\wedge(Q) \subseteq \Pi$, and since Π is Horn and $v = \bigwedge_{w \in Q_2} w \in \Pi$, it follows $\bigwedge_{w \in Q_2} z^{(w)} \in \Pi$. But then $\Pi \subseteq \Sigma_1 \setminus \Sigma_2$ contradicts (4.32), showing that $Q_1 = \emptyset$ is impossible.

Finally, let us consider the remaining case (i.e., $Q_1, Q_2 \neq \emptyset$). For a model $w \in Q_2$, let $\Sigma_w = \{w \wedge u \mid u \in Q_1\}$. Since $Q_1, Q_2 \subseteq C^*(\Sigma_1)$, $\Sigma_w \subseteq S$ holds for all w. Assume that, for each $w \in Q_2$, there is a model $u^{(w)} \in \Sigma_w$ such that $u^{(w)} \in Q$. Then $v = \bigwedge_{w \in Q_2} u^{(w)} \wedge \bigwedge_{w \in Q_1} w$ holds. Since $u^{(w)} \in Q$ and $Q_1 \subseteq Q$, we have $v \in Cl_\wedge(Q)$, a contradiction. Thus, there is a model $w^* \in Q_2$ such that $\Sigma_{w^*} \subseteq S \setminus Q$. The maximality of Q implies that, for each $w \in Q_2 \cup \Sigma_{w^*}$, there exists $Q_w \subseteq Q$ such that

$$\bigwedge_{u \in Q_w} u \ \wedge \ w \ \in \ \Sigma_2.$$

Let $y^{(w)} = \bigwedge_{u \in Q_w} u$ and $z^{(w)} = \bigwedge_{u \in Q_w} u \wedge w$. Then similarly to the above case of $Q_1 = \emptyset$, we can derive a contradiction. □

Let us next show the converse direction.

Lemma 4.6 *Let Σ_1 and Σ_2 be Horn theories, and let Π be a Horn core of $\Sigma_1 \setminus \Sigma_2$. Then $\Pi = Cl_\wedge(Q)$ holds for some maximal set $Q \subseteq S$ such that $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$, where S is given by* (3.11).

Proof. Assume that Π is a Horn core of $\Sigma_1 \setminus \Sigma_2$ which cannot be represented by any maximal set $Q \subseteq S$ such that $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$. Then there is a model $v \in C^*(\Pi) \setminus S$. Since $v \in \Sigma_1$, v can be represented by $v = \bigwedge_{w \in Q_1} w \wedge \bigwedge_{w \in Q_2} w$ for some $Q_1 \subseteq \Pi \cap C^*(\Sigma_1)$ and $Q_2 \subseteq C^*(\Sigma_1) \setminus \Pi$. Since $v \in C^*(\Pi) \setminus S$, $Q_2 \neq \emptyset$ holds. Then by an argument similar to the case of $Q_2 \neq \emptyset$ in the proof of Lemma 4.5, we can derive a contradiction by showing $v \in \Sigma_2$. □

From Lemmas 4.5 and 4.6, we have the next characterization.

Theorem 4.7 *Let Σ_1 and Σ_2 be Horn theories, and let S be given by* (3.11). *Then Π is a Horn core of $\Sigma_1 \setminus \Sigma_2$ if and only if $\Pi = Cl_\wedge(Q)$ holds for some maximal set $Q \subseteq S$ such that $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$.* □

Based on Theorem 4.7, we obtain the following polynomial time algorithm for computing a Horn core of $\Sigma_1 \setminus \Sigma_2$.

Algorithm HORN-CORE2

Input: Characteristic sets $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$ of Horn theories Σ_1 and Σ_2.

Output: The characteristic set $C^*(\Pi)$ of a Horn core Π of $\Sigma_1 \setminus \Sigma_2$.

Step 0. $Q := \emptyset$;
compute the set of models S given by (3.11);

Step 1. **for** each $w \in S$ **do begin**

if $Cl_\wedge(Q \cup \{w\}) \cap \Sigma_2 = \emptyset$ **then** $Q := Q \cup \{w\}$

end{for};

Step 2. output $C^*(Q)$ and halt. □

Example 4.4 Let us apply HORN-CORE2 to the Horn CNFs φ_1 and φ_2 from Example 3.1. They represent the theories $\Sigma_1 = \{(0111), (1011), (1010), (0011), (0001), (0010), (0000)\}$ and $\Sigma_2 = \{(1100), (0010), (0100), (1000), (0000)\}$, respectively. Hence,

$$\begin{aligned} C^*(\Sigma_1) &= \{(0111), (1011), (1010), (0001)\}, \\ C^*(\Sigma_2) &= \{(1100), (0010), (0100), (1000)\}. \end{aligned}$$

Then, in Step 0, we compute

$$\begin{aligned} S &= C^*(\Sigma_1) \cup \{(0000), (0011), (0010)\} \\ &= \{(0111), (1011), (1010), (0001), (0000), (0011), (0010)\}. \end{aligned}$$

By considering the models $w \in S$ in the above order, we obtain in Step 1

$$Q = \{(0111), (1011), (0001), (0011)\}.$$

Thus, $C^*(Q) = \{(0111), (1011), (0001)\}$ is output in Step 2. Observe that $C^*(Q)$ is the characteristic set of the Horn core $\Pi_1 = \{(0111), (1011), (0011), (0001)\}$ of $\varphi_1 \setminus \varphi_2$, which was already obtained in Example 4.1. □

Theorem 4.8 *Let Σ_1 and Σ_2 be Horn theories. Then, algorithm* HORN-CORE2 *outputs the characteristic set of a Horn core of $\Sigma_1 \setminus \Sigma_2$ in $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |\Sigma_2|))$ time.*

Proof. The correctness of algorithm HORN-CORE2 follows from Theorem 4.7. For its time complexity, Step 0 clearly can be done in $O(n|C^*(\Sigma_1)|^2)$ time. In Step 1, we can check if $Cl_\wedge(Q \cup \{w\}) \cap \Sigma_2 = \emptyset$ for each w in $O(n(|Q| + |C^*(\Sigma_2|)) = O(n(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2|))$ time [6]. By $|S| \leq |C^*(\Sigma_1)|^2$, Step 1 requires $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2|))$ time. Step 2 is executable in $O(n|Q|^2) = O(n|C^*(\Sigma_1)|^4)$ time.

Therefore, algorithm HORN-CORE2 requires $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2|))$ time. □

Also deciding whether a given theory Π is a Horn core of the difference $\Sigma_1 \setminus \Sigma_2$ is possible in polynomial time. For that, modify the algorithm HORN-CORE2 as follows. In Step 0, set Q to $\Pi \cap (S \setminus \Sigma_2)$. In Step 1, check whether $\Pi = Cl_\wedge(Q)$ holds and no $w \in S \setminus Q$ exists

such that $Cl_\wedge(Q \cup \{w\}) \cap \Sigma_2 = \emptyset$. If these are true, then output "yes" in Step 2, else output "no". The correctness of this procedure, which runs obviously in polynomial time, follows from Theorem 4.7.

By Theorem 4.7, we have a good characterization of a Horn core of the difference. Theorem 4.8 says that this characterization leads to a polynomial time algorithm for computing some arbitrary Horn core. However, as we will show next, However, there is no polynomial total time algorithm for generating all Horn cores, unless P = NP.

For a (not necessarily Horn) theory Σ, let $\mathcal{HC}(\Sigma)$ denote the set of all Horn cores of Σ. In order to show the intractability of computing $\mathcal{HC}(\Sigma_1 \setminus \Sigma_2)$ from $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$, let us consider the following problem.

Problem ADDITIONAL-HCORE

Input: Characteristic sets $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$ of Horn theories Σ_1 and Σ_2, and a family of the characteristic sets $C^*(\Pi)$ of Horn theories Π from a subset $\mathcal{S} \subseteq \mathcal{HC}(\Sigma_1 \setminus \Sigma_2)$.

Question: Does there exist some $\Pi \in \mathcal{HC}(\Sigma_1 \setminus \Sigma_2) \setminus \mathcal{S}$?

Let $V = \{1, 2, \ldots, n\}$, and let $\Phi = \bigwedge_{c \in C} c$ be a CNF on $x_1, \ldots, x_n$. It is well-known [20] that the satisfiability problem (SAT) (i.e., deciding whether $\Phi(v) = 1$ for some $v \in \{0,1\}^n$) is NP-complete. Without loss of NP-hardness, we may assume that no literal appears in all $c \in C$. Let $V' = \{1', 2', \ldots, n'\}$, and define

$$\Phi_1 \quad = \quad \bigvee_{j \in V} x_j x_{j'}. \tag{4.33}$$

Furthermore, let Φ_2 be the positive CNF obtained from Φ by replacing every negative literal $\overline{x}_j$ by $x_{j'}$, i.e.,

$$\Phi_2 \quad = \quad \bigwedge_{c \in C} \Big(\bigvee_{x_j \in P(c)} x_j \vee \bigvee_{\overline{x}_j \in N(C)} x_{j'} \Big), \tag{4.34}$$

where $P(c)$ and $N(c)$ are the sets of positive and negative literals in c, respectively. Define

$$\Phi' \quad = \quad \Phi_1 \vee \Phi_2. \tag{4.35}$$

It is easy to see that Φ' represents a positive theory. By the assumption that no literal appears in all $c \in C$, all terms $x_j x_{j'}$ in Φ_1 are prime implicants of Φ'.

We then claim the following lemma.

Lemma 4.9 *Let Φ and Φ' be defined as above. Then Φ is satisfiable if and only if Φ' has a prime implicant t which does not appear in Φ_1, and Φ is not satisfiable if and only if $\Phi' \equiv \Phi_1$.*

Proof. Let us prove the first half. Let v be a model such that $\Phi(v) = 1$. Then construct a term t_v by

$$t_v \quad = \quad \bigwedge_{v_j=1} x_j \wedge \bigwedge_{v_j=0} x_{j'}.$$

Clearly, this t_v is positive, and $t_v \leq \Phi_2$ holds. Since t_v contains exactly one of x_j and $x_{j'}$ for every j, we have $t_v \not\leq \Phi_1$, which implies the only-if part.

On the other hand, suppose that Φ' has a prime implicant t which does not appear in Φ_1. Since t does not contain both x_j and $x_{j'}$ for any j, we can define a model $v \in \{0,1\}^V$ by

$$v_j \quad = \quad \begin{cases} 1, & \text{if } x_j \in P(t); \\ 0, & \text{otherwise.} \end{cases}$$

Since Φ_1, Φ_2 and t represent positive theories, $t \not\leq \Phi_1$ implies $t \leq \Phi_2$. Hence we have $\Phi(v) = 1$. This proves the if-part.

The second half of the lemma is a restatement of the first half. □

Let $W_1 = V \cup V'$, $W_2 = \{n+1, n+2, \ldots, 2n+1\}$, and $W = W_1 \cup W_2$. Let x^U denote the model with $x^U_j = 1$ if $j \in U$ and 0 otherwise. Define sets of models $M_1, M_2 \subseteq \{0,1\}^W$ by

$$M_1 \quad = \quad \{x^U \mid U \subseteq W_1, |U \cap W_1| \geq 2n-1\} \tag{4.36}$$

$$M_2 \quad = \quad M_{2,1} \cup M_{2,2}, \tag{4.37}$$

where

$$M_{2,1} \quad = \quad \{x^U \mid W \setminus U = \{n+j, j\} \text{ or } \{n+j, j'\},\ j \in V\},$$

$$M_{2,2} \quad = \quad \{x^U \mid W \setminus U = \{2n+1\} \cup c,\ c \in C_2\},$$

and C_2 corresponds to the set of clauses in Φ_2 of (4.34). (Since $c \in C_2$ is positive, we regard a clause $c \in C_2$ as the corresponding index set; e.g., we write $c = \{1, 2', 4\}$ instead of $c = (x_1 \vee x_{2'} \vee x_4)$.) Let

$$\Sigma_1 = Cl_\wedge(M_1) \text{ and } \Sigma_2 = Cl_\wedge(M_2).$$

As we will see in Lemma 4.12, these Σ_1 and Σ_2, together with a suitable $\mathcal{S}$, give rise to a desired instance of problem ADDITIONAL-HCORE. Then, since

$$\Sigma_1 \quad = \quad \{x^U \mid U \cap W_2 = \emptyset\}, \tag{4.38}$$

$\Sigma_1 \setminus \Sigma_2 = \Sigma_1 \setminus \Sigma_2'$ holds, where

$$\Sigma_2' \quad = \quad \{x^U \in \Sigma_2 \mid U \cap W_2 = \emptyset\}. \tag{4.39}$$

This allows us to restrict our attention to W_1. That is, we will consider $\Sigma_1[W_1] \setminus \Sigma_2'[W_1] \subseteq \{0,1\}^{W_1}$ in place of $\Sigma_1 \setminus \Sigma_2 \subseteq \{0,1\}^W$, where $\Sigma[U]$ denotes the projection of Σ on U for a theory Σ and an index set U.

The following lemma shows that $\Sigma_1[W_1] \setminus \Sigma_2'[W_1]$ can be represented by the formula Φ'.

Lemma 4.10 *Let Σ_2' and Φ' be defined by* (4.39) *and* (4.35)*, respectively. Then the theory $\Sigma_2'[W_1]$ can be represented by $\neg\Phi'$.*

Proof. Note that

$$\neg\Phi' \quad = \quad \bigwedge_{j \in V} (\overline{x}_j \vee \overline{x}_{j'}) \wedge \bigvee_{c \in C_2} (\bigwedge_{l \in c} \overline{x}_l). \tag{4.40}$$

Let $v \in \Sigma_2'$. Then $v_j = 0$ for all $j \in W_2 = \{n+1, n+2, \ldots, 2n+1\}$. In order to have $v_{n+k} = 0$ for $k = 1, 2, \ldots, n$, either $v_k = 0$ or $v_{k'} = 0$ (or both) must hold by the definition of Σ_2. This means that v satisfies the CNF

$$\neg\Phi_1 \quad = \quad \bigwedge_{j \in V} (\overline{x}_j \vee \overline{x}_{j'}). \tag{4.41}$$

Moreover, in order to make $v_{2n+1} = 0$, v must be generated by referring to some model in $M_{2,2}$. Hence, it follows that v also satisfies the next CNF,

$$\neg\Phi_2 \quad = \quad \bigvee_{c \in C_2} (\bigwedge_{l \in c} \overline{x}_l). \tag{4.42}$$

Thus (4.41) and (4.42) imply that all models v in Σ_2' satisfy $\neg\Phi'$. Since $\neg\Phi'$ depends only on W_1, all models u in $\Sigma_2'[W_1]$ satisfies $\neg\Phi'$.

Conversely, we can easily see that all models $u \in \{0,1\}^{W_1}$ such that $\neg\Phi'(u) = 1$ are contained in $\Sigma_2'[W_1]$. (Note that Σ_2' is a negative theory, and hence $\Sigma_2'[W_1]$ is also a negative theory). □

It follows from Lemma 4.10 and Equation (4.38) that $\Sigma_1[W_1] \setminus \Sigma_2'[W_1]$ can be represented by Φ'. Recall that Φ' represents a positive theory. The Horn cores of a positive theory can be characterized as follows.

Lemma 4.11 *Let Σ be a positive theory. Then Π is a Horn core of Σ if and only if $\Pi = \{w \mid w \geq v\}$ holds for some minimal model v of Σ.*

Proof. For a minimal model v in Σ, consider a theory $\Pi = \{w \mid w \geq v\}$. Clearly, Π is Horn. Since Σ is positive, $\Pi \subseteq \Sigma$ holds. If Π is not Horn core of Σ, then there is a model $u \in \Sigma \setminus \Pi$ such that $Cl_\wedge(\Pi \cup \{u\}) \subseteq \Sigma$. Since $u \notin \Pi$, $v \bigwedge u < v$ holds. Therefore,

$v \bigwedge u \in Cl_\wedge(\Pi \cup \{u\}) \subseteq \Sigma$ holds, a contradiction to the minimality of v. Thus Π is a Horn core of Σ.

Conversely, assuming that Π is a Horn core of Σ, we consider the least model u of Π. Let v be a minimal model in Σ such that $v \leq u$. Then it is easy to see that $\Pi \subseteq \{w \mid w \geq v\} \subseteq \Sigma$. By the maximality of Π, we have $\Pi = \{w \mid w \geq v\}$. □

Now we obtain the next lemma.

Lemma 4.12 *Problem* ADDITIONAL-HCORE *is* NP*-complete.*

Proof. We first show the membership in NP. Let $\Pi \in \mathcal{HC}(\Sigma_1 \setminus \Sigma_2) \setminus \mathcal{S}$. Then $\Pi \setminus \Pi_i \neq \emptyset$ holds for all $\Pi_i \in \mathcal{S}$. Choose an arbitrary model $v^{(i)} \in \Pi \setminus \Pi_i$ for each $\Pi_i \in \mathcal{S}$, and let Q be the set of all the chosen $v^{(i)}$. Then (1) $|Q| \leq |\mathcal{S}|$, and (2) $Cl_\wedge(Q) \setminus \Pi_i \neq \emptyset$ holds for all $\Pi_i \in \mathcal{S}$. Moreover, since $Q \subseteq \Pi$, Q satisfies (3) $Cl_\wedge(Q) \subseteq \Sigma_1$ and (4) $Cl_\wedge(Q) \cap \Sigma_2 = \emptyset$ (i.e., $Cl_\wedge(Q) \subseteq \Sigma_1 \setminus \Sigma_2$). Conversely, if we have a set $Q \subseteq \Sigma_1$ satisfying (1)–(4), we can conclude that $\mathcal{HC}(\Sigma_1 \setminus \Sigma_2) \setminus \mathcal{S} \neq \emptyset$. Now (1) can be trivially checked, (2)–(3) can be checked in polynomial time by applying a technique similar to (3.18), and polynomiality of the check (4) follows from the result in [6] that, given theories Σ and Σ', deciding whether $Cl_\wedge(\Sigma) \cap Cl_\wedge(\Sigma') = \emptyset$ is possible in polynomial time. Thus any Q satisfying (1)–(4) is a polynomial certificate of the problem, i.e., a polynomial size proof for a Yes-instance which can be checked in polynomial time. Consequently, ADDITIONAL-HCORE is in NP.

We next show the NP-hardness. Let Σ_1 and Σ_2 be constructed as above from an arbitrary CNF Φ. Lemmas 4.10 and 4.11 imply that $\Sigma_1[W_1] \setminus \Sigma_2'[W_1]$ has a Horn core represented by $x_j x_{j'}$, for each $j \in V$. From (4.38), it thus follows that $\Sigma_1 \setminus \Sigma_2$ has a Horn core Π_j represented by

$$t_j \quad = \quad (\bigwedge_{n+k \in W_2} \overline{x}_k) \wedge x_j \wedge x_{j'},$$

for every $j \in V$. Let $\mathcal{S} = \{\Pi_j \mid j \in V\}$ be the collection of all these Horn cores.

Consider now the instance $C^*(\Sigma_1)$, $C^*(\Sigma_2)$, $\{C^*(\Pi_j) \mid \Pi_j \in \mathcal{S}\}$ of ADDITIONAL-HCORE. It follows from Lemmas 4.9, 4.10, and 4.11 that $\mathcal{HC}(\Sigma_1 \setminus \Sigma_2) \setminus \mathcal{S} \neq \emptyset$ if and only if Φ is satisfiable.

To complete the proof of NP-hardness, it remains to show that $C^*(\Sigma_1)$, $C^*(\Sigma_2)$ and $\{C^*(\Pi_j) \mid \Pi_j \in \mathcal{S}\}$ are computable in polynomial time from the formula Φ. Clearly, $C^*(\Sigma_1) = M_1$ and $C^*(\Sigma_2) = C^*(M_2)$ hold, where M_1 and M_2 are defined in (4.36) and (4.37); hence, both $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$ are constructible in polynomial time from Φ. As easily seen, the characteristic set of $\Pi_j \in \mathcal{S}$ takes the form of

$$C^*(\Pi_j) \quad = \quad \{x^U \mid U \subseteq W_1, U \supseteq \{j, j'\}, |U| \geq |W_1| - 1\},$$

and hence, $|C^*(\Pi_j)| = 2n - 2 + 1 = 2n - 1$. Thus $C^*(\Pi_j)$ is constructible in polynomial time.

Consequently, it follows that SAT is reducible to ADDITIONAL-HCORE in polynomial time, completing the proof. □

Based on this result, we can show in the next theorem that the enumeration of all Horn cores of a difference $\Sigma_1 \setminus \Sigma_2$ is a hard problem under model-based representation.

Theorem 4.13 *There is no algorithm which, given the characteristic sets $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$ of Horn theories Σ_1 and Σ_2, computes the characteristic sets $C^*(\Pi)$ of all Horn cores $\Pi \in \mathcal{HC}(\Sigma_1 \setminus \Sigma_2)$ in polynomial total time (i.e., polynomial in the combined size of the input and the output), unless* P = NP.

Proof. Suppose such an algorithm A exists. We construct an algorithm B from it, which solves problem ADDITIONAL-HCORE in polynomial time. By Lemma 4.12, this implies P = NP.

The algorithm B simulates algorithm A for at most $p(n|C^*(\Sigma_1)| + n|C^*(\Sigma_2)|, n|\mathcal{S}|)$ many steps, where $p(I, O)$ is a polynomial bound on the runtime of A and I (resp., O) is the size of the input (resp., output). (As usual, any set S of n-ary vectors occupies $n|S|$ bits.) Without loss of generality, suppose that p is a monotone function. If A stops within that time, B checks whether the computed output is identical to $\mathcal{S}$; if yes, then it outputs "no"; otherwise, it outputs "yes". It is not hard to see that B indeed correctly solves problem ADDITIONAL-HCORE; observe that if A does not halt within $p(n|C^*(\Sigma_1)| + n|C^*(\Sigma_2)|, n|\mathcal{S}|)$ many steps and $\mathcal{S}$ coincides with the set of computed Horn cores so far, then some additional Horn core exists. By the hypothesis, B runs in polynomial time in the size of $C^*(\Sigma_1), C^*(\Sigma_2)$ and $\mathcal{S}$. The result follows. □

5 Horn Envelope

In this section, we investigate the Horn envelope of the difference of two Horn theories. Recall that in contrast to Horn cores, the Horn envelope of a theory is always unique.

We again start with Horn theories represented by Horn CNFs, and then consider the model-based case.

5.1 Formula-based representation

Prior to resolving the complexity of computing the Horn envelope, we consider the associated problem of its recognition. Intuitively, if it is difficult to recognize the Horn envelope, then so is computing it. As we will see, this is the case for the difference of Horn theories.

We now show that recognizing the Horn envelope of $\varphi_1 \setminus \varphi_2$ is co-NP-complete. To this end, we provide a reduction from the complement of SAT to this problem.

Let $V = \{1, 2, \ldots, n\}$, and let $\Phi = \bigwedge_{c \in C} c$ be a CNF on $x_1, \ldots, x_n$. Furthermore, let $V' = \{1', 2', \ldots, n'\}$ and define

$$\varphi_1 = \bigwedge_{j \in V} (\overline{x}_j \vee \overline{x}_{j'}) \tag{5.43}$$

$$\varphi_2 = \bigwedge_{c \in C} \Big(\bigvee_{x_j \in P(c)} \overline{x}_{j'} \vee \bigvee_{\overline{x}_j \in N(C)} \overline{x}_j \Big). \tag{5.44}$$

Note that φ_1 and φ_2 represent negative (and hence Horn) theories.

Lemma 5.1 *Let Φ, φ_1 and φ_2 be defined as above. Then Φ is not satisfiable if and only if φ_1 is the Horn envelope of $\varphi_1 \setminus \varphi_2$.*

Proof. Let u be a model in $\{0,1\}^V$ such that $\Phi(u) = 1$. Then we construct the model $v \in \{0,1\}^{V \cup V'}$ by $v_j = u_j$ for $j \in V$ and $v_{j'} = 1 - u_j$ for $j' \in V'$. Since either $v_j = 0$ or $v_{j'} = 0$ holds for every $j \in V$, $\varphi_1(v) = 1$ holds. Moreover, by the definition of φ_2, $\varphi_2(v) = 1$ holds. Thus $(\varphi_1 \setminus \varphi_2)(v) = 0$. Let $c_v = \bigvee_{l \in V \cup V' : v_l = 1} \overline{x}_l$. We claim that

$$\varphi_1 \;\; > \;\; \varphi_1 \wedge c_v \;\; \geq \;\; \varphi_1 \setminus \varphi_2,$$

which completes the if-part, since $\varphi_1 \wedge c_v$ is Horn.

The first inequality follows from $\varphi_1 \geq \varphi_1 \wedge c_v$, $\varphi_1(v) = 1$ and $(\varphi_1 \wedge c_v)(v) = 0$. For the second inequality, it suffices to show that any model $w \neq v$ such that $c_v(w) = 0$ satisfies $(\varphi_1 \setminus \varphi_2)(w) = 0$. Clearly, $w_l = 1$ holds for every $l \in V \cup V'$ such that $v_l = 1$. Since exactly one of v_j and $v_{j'}$ has the value 1 for each $j \in V$, $w \neq v$ implies that there exists an index $h \in V$ such that $w_h = w_{h'} = 1$. This means $\varphi_1(w) = 0$, and hence $(\varphi_1 \setminus \varphi_2)(w) = 0$. Consequently, the second inequality holds.

To prove the only-if part, assume that Φ is not satisfiable. Then take a model $v \in \{0,1\}^{V \cup V'}$ with $\varphi_1(v) = \varphi_2(v) = 1$ (and hence $(\varphi_1 \setminus \varphi_2)(v) = 0$). Such a v has an index $h \in V$ such that $v_h = v_{h'} = 0$. Let us consider the models $w \geq v$ such that exactly one of w_j or $w_{j'}$ has the value 0 for every $j \in V$. Clearly, $\varphi_1(w) = 1$, and moreover it follows from the above discussion that $(\varphi_1 \setminus \varphi_2)(w) = 1$ holds. Since the model v is obtained by taking the intersection of all such w, this means that v is in the Horn envelope of $\varphi_1 \setminus \varphi_2$. Thus the Horn envelope of $\varphi_1 \setminus \varphi_2$ contains all models v with $\varphi_1(v) = \varphi_2(v) = 1$, and hence φ_1 is the Horn envelope. □

Theorem 5.2 *Given Horn CNFs φ_1, φ_2, and φ, deciding whether φ represents the Horn envelope of $\varphi_1 \setminus \varphi_2$ is* co-NP*-complete. This is true even if $\varphi = \varphi_1$.* □

Proof. We first show the membership in co-NP. The CNF φ does not represent the Horn envelope of $\varphi_1 \setminus \varphi_2$, if either (i) $\varphi_1 \setminus \varphi_2 \not\leq \varphi$ holds or (ii) there exists a Horn clause c such that $\varphi \not\leq c$ and $\varphi_1 \setminus \varphi_2 \leq \varphi \wedge c$. It is easily seen that $\varphi_1 \setminus \varphi_2 \not\leq \varphi$ holds if and only if $\varphi_1 \wedge \neg c \not\leq \varphi_2$ holds for all clauses c in φ. The latter condition amounts to testing the implication of Horn formulas, which is well-known to be polynomial. Thus, the condition in (i) can be checked in polynomial time. For (ii), a suitable clause c can be guessed and then verified in polynomial time, similarly as in (i). (Observe that (ii) need only be checked if (i) fails.) It follows that deciding whether φ represents the Horn envelope of $\varphi_1 \setminus \varphi_2$ is in co-NP.

The hardness part follows from Lemma 5.1. □

As an immediate consequence, we obtain the following result.

Corollary 5.3 *Let φ_1 and φ_2 be Horn CNFs. Then computing the Horn envelope is* co-NP*-hard.* □

By this result, an algorithm for computing the Horn envelope in polynomial time is unlikely to exist. Such an algorithm is in fact impossible, since the problem requires provably exponential time. This follows from the fact that the size of the Horn envelope of $\varphi_1 \setminus \varphi_2$ can be exponential in the sizes of φ_1 and φ_2. This is shown by the following example.

Example 5.1 Let Horn CNFs φ_1 and φ_2 be defined as follows.

$$\varphi_1 \;=\; \bigwedge_{i=1}^{n} (\overline{x}_i \vee \overline{z}_i), \qquad \varphi_2 \;=\; \bigwedge_{i=1}^{n} (\overline{x}_i \vee y_i).$$

Then, the complement of φ_2 amounts to the formula $\bigvee_{i=1}^{n} x_i \overline{y}_i = \bigwedge_{\lambda_i \in \{x_i, \overline{y}_i\}, i=1,2,\ldots,n} (\lambda_1 \vee \cdots \vee \lambda_n)$, and hence the difference of φ_1 and φ_2 is described by

$$\varphi_1 \setminus \varphi_2 \;=\; \bigwedge_{i=1}^{n} (\overline{x}_i \vee \overline{z}_i) \;\wedge \bigwedge_{\lambda_i \in \{x_i, \overline{y}_i\},\, i=1,2,\ldots,n} (\lambda_1 \vee \cdots \vee \lambda_n).$$

The prime implicates of $\varphi_1 \setminus \varphi_2$ which are definite (or pure) Horn (i.e., containing exactly one positive literal) are all clauses

$$x_i \vee \lambda_{i_1} \vee \cdots \vee \lambda_{i_{n-1}}, \tag{5.45}$$

such that $\{i_1, \ldots, i_{n-1}\} = \{1, \ldots, n\} \setminus \{i\}$ for some i and $\lambda_{i_j} \in \{\overline{y}_{i_j}, \overline{z}_{i_j}\}$ for $j = 1, 2, \ldots, n-1$. Observe that resolution is not applicable to any pair of such clauses. Consequently, any irredundant prime DNF for $\varphi_1 \setminus \varphi_2$ must contain all clauses (5.45). There are $n2^{n-1}$ such clauses, and hence any Horn CNF for the Horn envelope of $\varphi_1 \setminus \varphi_2$ is exponential in the sizes of φ_1 and φ_2. □

Since the size of the Horn envelope can be exponential, we have to resort to polynomial total time computability as an appropriate concept of tractability. However, from the previous results, we also obtain for this notion a negative result.

Theorem 5.4 *There is no polynomial total time algorithm, which, given two Horn CNFs φ_1 and φ_2, computes an (irredundant and prime) Horn CNF for the Horn envelope of $\varphi_1 \setminus \varphi_2$, unless* P = NP.

Proof. The proof is similar to the proof of Theorem 4.13. Assuming that such an algorithm A exists, we can construct an algorithm B that which decides in polynomial time whether φ_1 is the Horn envelope of $\varphi_1 \setminus \varphi_2$ or not. By Theorem 5.2, this implies P = NP. □

5.2 Model-based representation

In Subsection 3.2, we showed that $Cl_\wedge(S_1) = \Sigma_1 \setminus \Sigma_2$ holds if $\Sigma_1 \setminus \Sigma_2$ is Horn, where S_1 is given by (3.12). It is not difficult to see that, in general, this gives the Horn envelope of $\Sigma_1 \setminus \Sigma_2$.

Lemma 5.5 *Let Σ_1 and Σ_2 be Horn theories, and let S_1 be the set of models defined by* (3.12). *Then $Cl_\wedge(S_1)$ is the Horn envelope of $\Sigma_1 \setminus \Sigma_2$.*

Proof. By (3.13), the Horn envelope Π of $\Sigma_1 \setminus \Sigma_2$ must contain S_1. Since Π is Horn, $Cl_\wedge(S_1) \subseteq \Pi$ holds. On the other hand, it follows from (3.14) and (3.16) that $Cl_\wedge(S_1) \supseteq \Sigma_1 \setminus \Sigma_2$. Thus, $\Pi = Cl_\wedge(S_1)$ holds, since Π is the Horn envelope. □

Theorem 5.6 *Let Σ_1 and Σ_2 be Horn theories, and let Π be the Horn envelope of $\Sigma_1 \setminus \Sigma_2$. Then, given $C^*(\Sigma_1)$ and $C^*(\Sigma_2)$, $C^*(\Pi)$ is computable in $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2)|))$ time.*

Proof. By Lemma 5.5, we have to compute $C^*(S_1)$. As already shown in the proof of Theorem 3.4, $C^*(S_1)$ can be constructed in $O(n|C^*(\Sigma_1)|^2(|C^*(\Sigma_1)|^2 + |C^*(\Sigma_2)|))$ time. □

6 Discussion and Conclusion

In this paper, we have considered the issue of computing a Horn representation of the difference of two Horn theories Σ_1 and Σ_2. For this purpose, we have presented characterizations of the differences $\Sigma_1 \setminus \Sigma_2$ which are Horn, and algorithms for computing its Horn representation. Moreover, for the case in which $\Sigma_1 \setminus \Sigma_2$ is not Horn, we have characterized and presented algorithms for computing Horn cores and the Horn envelope of $\Sigma_1 \setminus \Sigma_2$. Our

results on the complexity of these issues are summarized in Table 1, which gives a complete picture of the tractability/intractability frontier of these problems. We have considered both the standard representation of Horn theories by means of Horn CNFs, and by the set of its characteristic models [14, 17, 18, 19].

Table 1: Complexity of problems on the Horn difference $\Sigma_1 \setminus \Sigma_2$ (Σ_1, Σ_2 and Π are Horn)

	representation	
Problem	*formula-based (Horn CNF)*	*model-based (characteristic set)*
Is $\Sigma_1 \setminus \Sigma_2$ Horn?	P	P
Compute $\Sigma_1 \setminus \Sigma_2$, if it is Horn	P	P
Is Π a Horn core of $\Sigma_1 \setminus \Sigma_2$?	P	P
Compute one Horn core of $\Sigma_1 \setminus \Sigma_2$	P	P
Compute all Horn cores of $\Sigma_1 \setminus \Sigma_2$	polynomial delay	not polynomial total time unless P = NP
Is Π the Horn envelope of $\Sigma_1 \setminus \Sigma_2$?	co-NP-complete	P
Compute the Horn envelope of $\Sigma_1 \setminus \Sigma_2$	not polynomial total time unless P = NP	P

Our results on formula-based representation complement the results known from [21, 2, 16, 3, 4]. Observe that these works have focused on theories which are represented by CNFs or restricted classes of CNFs, i.e., the problem input is a (possibly restricted) CNF. In contrast, we deal with non-CNF formulas as a result of the difference operation $\setminus$.

The tractability of computing some Horn core of a Horn difference is a positive result, since this problem is intractable for a general theory. For the case, in which a theory is represented by a CNF formula, [2] contains a polynomial time algorithm which consults an NP oracle, and shows that the problem is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-hard. Most recently, in [1] another algorithm for computing a Horn core from a given CNF is represented, which is based on the classical Davis-Putnam procedure for the satisfiability problem. Observe, however, that

these algorithms are not efficiently applicable to our problem, since they require a CNF formula for input; by rewriting the difference of Horn CNFs φ_1 and φ_2 to a CNF, the size of the formula might exponentially increase. Our positive result that all Horn cores of a Horn difference $\varphi_1 \setminus \varphi_2$ can be enumerated with polynomial delay parallels the analogous result that the Horn cores of a given theory Σ can be enumerated with polynomial delay [17].

Concerning the Horn envelope, our results complement the intractability results presented in [21, 16, 17, 18, 3]. It was known [21] that the Horn envelope of a CNF φ can be exponential in the size of φ; as already shown, the same holds for the difference of very simple Horn CNFs φ_1 and φ_2. Furthermore, the result that computing the Horn envelope of $\varphi_1 \setminus \varphi_2$ is difficult even if a possible exponential size is taken into account, parallels a similar result for computing an irredundant prime CNF for the Horn envelope of a given theory Σ [17]. A result in [18] suggests that this problem is in fact difficult, and no polynomial total time algorithm is known. On the other hand, it remains open [18] whether such an algorithm would imply P = NP. Thus, finding a polynomial time reduction of computing the Horn envelope of a Horn difference $\varphi_1 \setminus \varphi_2$ to the problem setting in [17, 18] appears to be an intriguing nontrivial and perhaps impossible problem, even though a converse reduction is feasible. (For algorithms computing the Horn envelope of classes of CNFs, see [16, 3, 4] and references therein.)

The results of the present and companion papers [6, 7] establish operations towards a Boolean "calculus" on Horn theories, in which Horn theories are combined using the operations of conjunction ($\wedge$), disjunction ($\vee$), and difference ($\setminus$). These operations are useful, for example, in the context of changing theories that are possible world representations of a state of affairs. The paper [6] considered conjunction $\wedge$ for model-based representations of Horn theories. Note that the conjunction of Horn theories is also Horn, and the conjunction of Horn CNFs is also a Horn CNF. Therefore, in this case, we do not need to consider CNF representations and Horn approximation (such as core and envelope). As shown in [6], there exists no polynomial total time algorithm for computing model-based representation for the conjunction of Horn theories, but deciding consistency and deductive question for the conjunction of Horn theories are both polynomial. The paper [7] investigated the problems on the disjunction of Horn theories $\Sigma_1, \Sigma_2, \ldots, \Sigma_l$ similar as in this paper. In particular, we showed that deciding whether the disjunction represents a Horn theory is co-NP-complete in general, but is polynomial if l is bounded by some constant.

Several issues remain for further work. One issue is giving a precise account to computing Horn cores and the Horn envelope of the complement $\overline{\Sigma} = \{0,1\}^n \setminus \Sigma$ of a Horn theory. Our results imply polynomial time algorithms in some cases, but a complete picture remains to be drawn. Another issue is a more accurate account of the complexity of the polynomial

time solvable problems in Table 1. Under formula-based representation, all these problems are complete for P under logspace reductions; this is an easy consequence of the fact that deciding the satisfiability of a Horn CNF is complete for P under logspace reductions, cf. [20]. Therefore, parallelization of these problems is most likely not possible. It remains to see which results can be obtained for the model-based representation.

Another interesting issue is a generalization of the results from the propositional to the first order case. Here, in general only formula-based representation is meaningful, and the problems in Table 1 are nonrecursive in general. It remains to single out cases and restrictions under which these problems are effectively solvable. Finally, a study of the effect of Horn renamings (cf. [1]) could be worthwhile.

References

[1] Y. Boufkhad. Algorithms for Propositional KB Approximation. In *Proceedings National Conference on AI (AAAI '98)*, pages 280–285, Madison, Wisconsin, July 26–30 1998. AAAI Press.

[2] M. Cadoli. Semantical and Computational Aspects of Horn Approximations. In *Proceedings International Join Conference on Artificial Intelligence (IJCAI '93)*, pages 39–44, Chambery, France, August 1993. Full version, by M. Cadoli and F. Scarcello, (1997) available from the author.

[3] A. del Val. An Analysis of Approximate Knowledge Compilation. In *Proceedings IJCAI '95*, pages 830–836, 1995.

[4] A. del Val. Approximate Knowledge Compilation: The First-Order Case. In *Proceedings AAAI-96*, pages 498–503, 1996.

[5] W. Dowling and J. H. Gallier. Linear-time Algorithms for Testing the Satisfiability of Propositional Horn Theories. *Journal of Logic Programming*, 3:267–284, 1984.

[6] T. Eiter, T. Ibaraki, and K. Makino. Computing Intersections of Horn Theories for Reasoning with Models. *Artificial Intelligence*, 110: 57–101, 1999.

[7] T. Eiter, T. Ibaraki, and K. Makino. Disjunctions of Horn Theories and their Cores. In K.-Y. Chwa and O. H. Ibarra, editors, *Proceedings Ninth Annual Symposium on Algorithms and Computation (ISAAC '98)*, LNCS 1533, pages 49–58, Korea, December 1998. Springer.

[8] T. Eiter, T. Ibaraki, and K. Makino. On the Difference of Horn Theories. In *Proceedings 16th International Symposium on Theoretical Aspects of Computer Science (STACS '99)*, LNCS 1563, pages 467–477, Germany, March, 1999, Springer.

[9] T. Eiter, T. Ibaraki, and K. Makino. Double Horn Functions. *Information and Computation*, 144:155–190, 1998.

[10] G. Gogic, C. Papadimitriou, and M. Sideri. Incremental Recompilation of Knowledge. *Journal of Artificial Intelligence Research*, 8:23–37, 1998.

[11] W. Hodges. Logical Features of Horn Clauses. In D. Gabbay, C. Hogger, and J. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume I : Logical Foundations, pages 449–503. Clarendon Press, Oxford; New York, 1994.

[12] D. S. Johnson, M. Yannakakis, and C. H. Papadimitriou. On Generating All Maximal Independent Sets. *Information Processing Letters*, 27:119–123, 1988.

[13] H. Katsuno and A. O. Mendelzon. Propositional Knowledge Base Revision and Minimal Change. *Artificial Intelligence*, 52:253–294, 1991.

[14] H. Kautz, M. Kearns, and B. Selman. Reasoning With Characteristic Models. In *Proceedings AAAI-93*, pages 34–39, 1993.

[15] H. Kautz, M. Kearns, and B. Selman. Horn Approximations of Empirical Data. *Artificial Intelligence*, 74:129–245, 1995.

[16] H. Kautz and B. Selman. Knowledge Compilation and Theory Approximation. *Journal of the ACM*, 43(2):193–224, 1996.

[17] 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. Springer.

[18] R. Khardon. Translating between Horn Representations and their Characteristic Models. *Journal of Artificial Intelligence Research*, 3:349–372, 1995.

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

[20] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[21] B. Selman and H. Kautz. Knowledge Compilation Using Horn Approximations. In *Proceedings AAAI-91*, pages 904–909, Anaheim, August 1991.