Counting Complexity of
Minimal Cardinality and Minimal Weight Abduction[⋆]

Miki Hermann[1] and Reinhard Pichler[2]

[1] LIX (CNRS, UMR 7161), École Polytechnique, 91128 Palaiseau, France.
`hermann@lix.polytechnique.fr`
[2] Institut für Informationssysteme, Technische Universität Wien, A-1040 Wien, Austria.
`pichler@dbai.tuwien.ac.at`

Abstract. Abduction is an important method of non-monotonic reasoning with many applications in artificial intelligence and related topics. In this paper, we concentrate on propositional abduction, where the background knowledge is given by a propositional formula. We have recently started to study the counting complexity of propositional abduction. However, several important cases have been left open, namely, the cases when we restrict ourselves to solutions with minimal cardinality or with minimal weight. These cases – possibly combined with priorities – are now settled in this paper. We thus arrive at a complete picture of the counting complexity of propositional abduction.

1 Introduction

Abduction is a method of non-monotonic reasoning which has taken a fundamental importance in artificial intelligence and related topics. It aims at giving explanations for observed symptoms and is, therefore, widely used in diagnosis – notably in the medical domain (see [17]). Other important applications of abduction can be found in planning, database updates, data-mining and many more areas (see e.g. [11, 12, 16]).

Logic-based abduction is formally described as follows. Given a logical theory T, a set M of manifestations, and a set H of hypotheses, find a solution $\mathcal{S}$, i.e., a set $S \subseteq H$ such that $T \cup S$ is consistent and logically entails M. In this paper, we consider *propositional abduction problems* (PAPs, for short), where the theory T is represented by a propositional formula over a Boolean algebra $\mathbb{B} = (\{0, 1\}; \vee, \wedge, \neg, \rightarrow, \equiv)$ and the sets H and M consist of variables from some set V. A *diagnosis problem* can be represented by a PAP $\mathcal{P} = \langle V, H, M, T \rangle$ as follows: The theory T is the system description. The hypotheses $H \subseteq V$ describe the possibly faulty system components. The manifestations $M \subseteq V$ are the observed symptoms, describing the malfunction of the system. The solutions $\mathcal{S}$ of $\mathcal{P}$ are the possible explanations of the malfunction.

Example 1. Consider the following football knowledge base.

$$T = \{\, weak\_defense \wedge weak\_attack \rightarrow match\_lost,$$
$$match\_lost \rightarrow manager\_sad \wedge press\_angry$$
$$star\_injured \rightarrow manager\_sad \wedge press\_sad \,\}$$

Moreover, let the set of observed manifestations and the set of hypotheses be

$$M = \{\ manager\_sad\ \}$$
$$H = \{\ star\_injured,\ weak\_defense,\ weak\_attack\ \}$$

This PAP has the following five abductive explanations (= "solutions").

$$\mathcal{S}_1 = \{\ star\_injured\ \}$$
$$\mathcal{S}_2 = \{\ weak\_defense,\ weak\_attack\ \}$$
$$\mathcal{S}_3 = \{\ weak\_attack,\ star\_injured\ \}$$
$$\mathcal{S}_4 = \{\ weak\_defense,\ star\_injured\ \}$$
$$\mathcal{S}_5 = \{\ weak\_defense,\ weak\_attack,\ star\_injured\ \}$$

Obviously, in the above example, not all solutions are equally intuitive. Indeed, for many applications, one is not interested in *all* solutions of a given PAP $\mathcal{P}$ but only in *all acceptable* solutions of $\mathcal{P}$. *Acceptable* in this context means *minimal* with respect to some preorder $\preceq$ on the powerset 2^H. Two natural preorders are subset-minimality and cardinality-minimality, where the preorder is $\subseteq$ and $\leq$, respectively. In Example 1, both $\mathcal{S}_1$ and $\mathcal{S}_2$ are subset-minimal but only $\mathcal{S}_1$ is cardinality-minimal. If we have a weight function on the hypotheses then we may define the acceptable solutions as the weight-minimal ones. This preorder (i.e., smaller or equal weight) is denoted as $\sqsubseteq$.

All three criteria $\subseteq$, $\leq$, and $\sqsubseteq$ can be further refined by a hierarchical organization of our hypotheses according to some *priorities* (cf. [5]). In this context, priorities may be used to represent a qualitative version of probability. The resulting preorder is denoted by $\subseteq_P$, $\leq_P$, and $\sqsubseteq_P$. For instance, suppose that for some reason we know that (for a specific team) *star_injured* is much less likely to occur than *weak_defense* and *weak_attack*. This judgment can be formalized by assigning lower priority to the former. Then $\mathcal{S}_2$ is the only minimal solution with respect to the preorders $\subseteq_P$ and $\leq_P$. Actually, in this simple example, $\mathcal{S}_2$ is also the only $\sqsubseteq_P$-minimal solution independently of the concrete weight function. Finally, if indeed all solutions are acceptable, then the corresponding preorder is the syntactic equality $=$.

The usually observed algorithmic problem in logic-based abduction is the existence problem, i.e. deciding whether at least one solution $\mathcal{S}$ exists for a given abduction problem $\mathcal{P}$. Another well-studied decision problem is the so-called relevance problem, i.e. Given a PAP $\mathcal{P}$ and a hypothesis $h \in H$, is h part of at least one acceptable solution? However, this approach is not always satisfactory. Especially in database applications, in diagnosis, and in data-mining there exist situations where we need to know *all* acceptable solutions of the abduction problem or at least an important part of them. Consequently, the enumeration problem (i.e., the computation of all acceptable solutions) has received much interest (see e.g. [3,4]). Another natural question is concerned with the total number of solutions to the considered problem. The latter problem refers to the *counting complexity* of abduction. Clearly, the counting complexity provides a lower bound for the complexity of the enumeration problem. Moreover, counting the number of abductive explanations can be useful for probabilistic abduction problems (see e.g. [18]). Indeed, in order to compute the probability of failure of a given component in a diagnosis problem (under the assumption that all preferred explanations are equiprobable), we need to count the number of preferred explanations as well as the number of preferred explanations that contain a given hypothesis.

#-Abduction	$=$	$\subseteq$	$\subseteq_P$	$\leq$	$\leq_P$	$\sqsubseteq, \sqsubseteq_P$
General case	#·coNP	#·coNP	#·Π_2P	#·Opt$_2$P[$\log n$]	#·Opt$_2$P	#·Opt$_2$P
Horn	#P	#P	#·coNP	#·OptP[$\log n$]	#·OptP	#·OptP
definite Horn	#P	#P	#P	#·OptP[$\log n$]	#·OptP	#·OptP
dual Horn	#P	#P	#P	#·OptP[$\log n$]	#·OptP	#·OptP
bijunctive	#P	#P	#·coNP	#·OptP[$\log n$]	#·OptP	#·OptP

Table 1. Counting complexity of propositional abduction

The study of counting complexity has been initiated by Valiant [19, 20] and is now a well-established part of the complexity theory, where the best known class is #P. Many counting variants of decision problems have been proved #P-complete. Higher counting complexity classes do exist, but they are not commonly known. A counting equivalent of the polynomial hierarchy was defined by Hemaspaandra and Vollmer [8], whereas generic complete problems for these counting hierarchy classes were presented in [1]. We enlarged in [10] the approach of Hemaspaandra and Vollmer to classes of optimization problem, obtaining this way a new hierarchy of classes #·Opt$_k$P[$\log n$] and #·Opt$_k$P for arbitrary $k \in \mathbb{N}$. These classes are sandwiched between the previously known counting classes #·Π_kP, i.e., for each $k \in \mathbb{N}$ we have

$$\text{\#·}\Pi_k\text{P} \subseteq \text{\#·Opt}_{k+1}\text{P}[\log n] \subseteq \text{\#·Opt}_{k+1}\text{P} \subseteq \text{\#·}\Pi_{k+1}\text{P}.$$

It was shown in [10] that these inclusions are proper unless the polynomial hierarchy collapses to the k-th level. The most important special case is $k = 1$, where we write #·OptP[$\log n$] and #·OptP as a short-hand for #·Opt$_1$P[$\log n$] and #·Opt$_1$P. On the first two levels, we thus have the inclusions #P $\subseteq$ #·OptP[$\log n$] $\subseteq$ #·OptP $\subseteq$ #·coNP $\subseteq$ #·Opt$_2$P[$\log n$] $\subseteq$ #·Opt$_2$P $\subseteq$ #·Π_2P. It will turn out that these new counting complexity classes are precisely the ones needed to pinpoint the exact counting complexity of the open cases in propositional abduction.

Results. We considered in [9] propositional abduction counting problems with the three preorders $=$, $\subseteq$, and $\subseteq_P$. Together with the general case where T can be an arbitrary propositional formula, we also considered the special cases where T is Horn, definite Horn, dual Horn, and bijunctive. These are the most frequently studied subcases of propositional formulas. Our results from [9] are summarized in the first three columns of Table 1. In this paper we continue the investigation on counting complexity of propositional abduction, focusing on the preorders $\leq$, $\sqsubseteq$, $\leq_P$, and $\sqsubseteq_P$. Note that these are practically highly relevant cases for the following reasons: If the failure of any component in a system is independent of the failure of the other components and all components have equal failure probability, then explanations with minimum cardinality are the ones with highest probability. If we have numeric values available for the repair cost or for the robustness of each component (e.g., based on data such as the empirically collected mean time to failure and component age), then weight-minimal abduction seeks for the cheapest repair respectively for the most likely explanation. If in addition different sets of components can be ranked according to some criterion that is not well

suited for numeric values (like, e.g., a qualitative rather than a quantitative robustness measure of components, the accessibility of components, or how critical the failure of a certain component would be), then this ranking can be expressed by priorities on the hypotheses, for both the cardinality and weight minimal case. Our results obtained in this work are summarized in the last three columns of Table 1. In total, we have thus achieved a complete picture of the counting complexity of propositional abduction.

2 Preliminaries

2.1 Propositional Abduction

A *propositional abduction problem* (PAP) $\mathcal{P}$ consists of a tuple $\langle V, H, M, T \rangle$, where V is a finite set of *variables*, $H \subseteq V$ is the set of *hypotheses*, $M \subseteq V$ is the set of *manifestations*, and T is a consistent *theory* in the form of a propositional formula. A set $\mathcal{S} \subseteq H$ is a *solution* (also called *explanation*) to $\mathcal{P}$ if $T \cup \mathcal{S}$ is consistent and $T \cup \mathcal{S} \models M$ holds. *Priorities* $P = \langle H_1, \ldots, H_K \rangle$ are a stratification of the hypotheses $H = H_1 \cup \cdots \cup H_K$ into a fixed number of disjoint sets. The *minimal cardinality with priorities* relation $A \leq_P B$ holds if $A = B$ or there exists an $i \in \{1, \ldots, K\}$ such that $A \cap H_j = B \cap H_j$ for all $j < i$ and $|A \cap H_i| < |B \cap H_i|$. The *minimal weight with priorities* relation $A \sqsubseteq_P B$ holds if $A = B$ or there exists an $i \in \{1, \ldots, K\}$ such that $A \cap H_j = B \cap H_j$ for all $j < i$ and $\sum_{a \in A \cap H_i} w(a) < \sum_{b \in B \cap H_i} w(b)$, where $w \colon H \to \mathbb{N}$ is the weight function on the hypotheses H.

   We study the following family of counting problems, which are parameterized by a preorder $\preceq$ on 2^H.

Problem: #-$\preceq$-ABDUCTION
Input: A propositional abduction problem $\mathcal{P} = \langle V, H, M, T \rangle$.
Output: Number of $\preceq$-minimal solutions (explanations) of $\mathcal{P}$.

We considered the abduction counting problems with the preorders of equality $=$, subset minimality $\subseteq$, and subset minimality with priorities $\subseteq_P$ in [9]. In this paper we consider the preorders of minimal cardinality $\leq$, minimal weight $\sqsubseteq$, as well as their versions with priorities $\leq_P$ and $\sqsubseteq_P$, respectively. It is clear that an upper bound for a minimal weight decision or counting abduction problem subsumes that for the corresponding abduction problem for minimal cardinality. Similarly, a lower bound for a minimal cardinality abduction problem subsumes that for minimal weight abduction. In both cases, setting the weight of each hypothesis $x \in H$ to $w(x) = 1$ corresponds to the cardinality version. Throughout this paper, we follow the formalism of Eiter and Gottlob [2], allowing only positive literals in the solutions.

   Together with the general case where T can be an arbitrary propositional formula, we consider the special cases where T is Horn, definite Horn, dual Horn, and bijunctive. A propositional clause C is said to be *Horn*, *definite Horn*, *dual Horn*, or *bijunctive* if it has at most one positive literal, exactly one positive literal, at most one negative literal, or at most two literals, respectively. A theory T is Horn, definite Horn, dual Horn, or bijunctive if it is a conjunction (or, equivalently, a set) of Horn, definite Horn, dual Horn, or bijunctive, clauses, respectively.

2.2 Counting Complexity

The study of *counting problems* was initiated by Valiant in [19, 20]. While decision problems ask if at least one solution of a given problem instance exists, counting problems ask for the number of different solutions. The most intensively studied counting complexity class is #P, which denotes the functions that count the number of accepting paths of a non-deterministic polynomial-time Turing machine. In other words, #P captures the counting problems corresponding to decision problems in NP. By allowing the non-deterministic polynomial-time Turing machine access to an oracle in NP, $\Sigma_2 P$, $\Sigma_3 P, \ldots$, we can define an infinite hierarchy of counting complexity classes.

Alternatively, a *counting problem* is presented using a *witness* function which for every input x returns a set of *witnesses* for x. A *witness* function is a function $w \colon \Sigma^* \to \mathcal{P}^{<\omega}(\Gamma^*)$, where Σ and Γ are two alphabets, and $\mathcal{P}^{<\omega}(\Gamma^*)$ is the collection of all finite subsets of Γ^*. Every such witness function gives rise to the following *counting problem*: given a string $x \in \Sigma^*$, find the cardinality $|w(x)|$ of the *witness* set $w(x)$. According to [8], if $\mathcal{C}$ is a complexity class of decision problems, we define #·$\mathcal{C}$ to be the class of all counting problems whose witness function w satisfies the following conditions.

1. There is a polynomial $p(n)$ such that for every $x \in \Sigma^*$ and every $y \in w(x)$ we have $|y| \le p(|x|)$;
2. The problem "given x and y, is $y \in w(x)$?" is in $\mathcal{C}$.

It is easy to verify that #P $=$ #·P. The counting hierarchy is ordered by linear inclusion [8]. In particular, we have that #P $\subseteq$ #·coNP $\subseteq$ #·Π_2P $\subseteq$ #·Π_3P, etc

In [10] we introduced new counting complexity classes for counting *optimal* solutions. We followed the aforementioned approach, where the complexity class $\mathcal{C}$ was chosen among OptP and OptP[$\log n$], or, more generally, Opt_kP and Opt_kP[$\log n$] for arbitrary $k \in \mathbb{N}$, respectively. These classes were previously defined by Krentel [14, 15]. A large collection of completeness results for these classes is given in [7]. As Krentel observed, the classes OptP[$\log n$] and OptP, which are closely related to $\text{FP}^{\text{NP}[\log n]}$ and FP^{NP}, contain problems computing optimal solutions with a logarithmic and polynomial number of calls to an NP-oracle, respectively.

The application of the counting operator to the aforementioned optimization classes allowed us to define in [10] the counting complexity classes #·OptP, #·OptP[$\log n$] and, more generally, #·Opt_kP, #·Opt_kP[$\log n$] for each $k \in \mathbb{N}$. To formally introduce these classes, we need some supplementary notions.

A *non-deterministic transducer* M is a non-deterministic polynomial-time bounded Turing machine, which writes a binary number on the output at the end of every accepting path. If M is equipped with an oracle from the complexity class $\mathcal{C}$, then it is called a *non-deterministic transducer with $\mathcal{C}$-oracle*. A Σ_kP*-transducer* M is a non-deterministic transducer with a Σ_{k-1}P oracle. We identify non-deterministic transducers without oracle and Σ_1P-transducers. For $x \in \Sigma^*$, we write $\text{opt}_M(x)$ to denote the *optimal* value, which can be either the *maximum* or the *minimum*, on any accepting path of the computation of M on x. If no accepting path exists then $\text{opt}_M(x)$ is undefined.

We say that a counting problem #·$A \colon \Sigma^* \to \mathbb{N}$ is in the class #·Opt_kP for some $k \in \mathbb{N}$, if there is a Σ_kP-transducer M, such that #·$A(x)$ is the number of accepting paths of the computation of M on x yielding the optimum value $\text{opt}_M(x)$. If

no accepting path exists then $\#{\cdot}A(x) = 0$. If the length of the binary number written by M is bounded by $O(\log |x|)$, then $\#{\cdot}A$ is in the class $\#{\cdot}\mathrm{Opt}_k\mathrm{P}[\log n]$. For $k = 1$, we write $\#{\cdot}\mathrm{OptP}[\log n]$ and $\#{\cdot}\mathrm{OptP}$ as a short-hand for $\#{\cdot}\mathrm{Opt}_1\mathrm{P}[\log n]$ and $\#{\cdot}\mathrm{Opt}_1\mathrm{P}$, respectively. It was shown in [10] that these new classes $\#{\cdot}\mathrm{Opt}_k\mathrm{P}[\log n]$ and $\#{\cdot}\mathrm{Opt}_k\mathrm{P}$ are robust, i.e., they do not collapse to already known counting complexity classes unless the polynomial hierarchy collapses as well. Finally, these new counting classes were shown to be sandwiched between the classes $\#{\cdot}\Pi_k\mathrm{P}$, i.e., we obtained the inclusions $\#\mathrm{P} \subseteq \#{\cdot}\mathrm{OptP}[\log n] \subseteq \#{\cdot}\mathrm{OptP} \subseteq \#{\cdot}\mathrm{coNP} \subseteq \#{\cdot}\mathrm{Opt}_2\mathrm{P}[\log n] \subseteq \#{\cdot}\mathrm{Opt}_2\mathrm{P} \subseteq \#{\cdot}\Pi_2\mathrm{P}$, etc.

The prototypical $\#{\cdot}\Pi_k\mathrm{P}$-complete problem for $k \in \mathbb{N}$ is $\#\Pi_k\mathrm{SAT}$ [1], defined as follows. Given a formula

$$\varphi(X) = \forall Y_1 \exists Y_2 \cdots Q_k Y_k \; \psi(X, Y_1, \ldots, Y_k)$$

where ψ is a Boolean formula and X, Y_1, ..., Y_k are sets of propositional variables, count the number of truth assignments to the variables in X that satisfy φ. We obtain the prototypical $\#{\cdot}\mathrm{Opt}_{k+1}\mathrm{P}[\log n]$-complete problem #MIN-CARD-Π_kSAT and the prototypical $\#{\cdot}\mathrm{Opt}_{k+1}\mathrm{P}$-complete problem #MIN-WEIGHT-Π_kSAT [10] by asking for the number of cardinality-minimal and weight-minimal models of $\varphi(X)$. In the latter case, there exists a weight function $w\colon X \to \mathbb{N}$ assigning positive values to each variable $x \in X$. As usual, the counting problems #MIN-CARD-Π_0SAT and #MIN-WEIGHT-Π_0SAT are just denoted by #MIN-CARD-SAT and #MIN-WEIGHT-SAT, being respectively $\#{\cdot}\mathrm{OptP}[\log n]$- and $\#{\cdot}\mathrm{OptP}$-complete.

3 General Case

Theorem 2. *#-$\leq$-ABDUCTION is $\#{\cdot}\mathrm{Opt}_2\mathrm{P}[\log n]$-complete and #-$\sqsubseteq$-ABDUCTION is $\#{\cdot}\mathrm{Opt}_2\mathrm{P}$-complete.*

Proof. In order to prove the membership, we show that these problems can be solved by an appropriate $\Sigma_2\mathrm{P}$-transducer M, i.e., M works in non-deterministic polynomial time with access to an NP-oracle and, in case of #-$\leq$-ABDUCTION, the output of M is logarithmically bounded. We give a high-level description of M: It takes an arbitrary PAP $\mathcal{P} = \langle V, H, M, T \rangle$ as input and non-deterministically enumerates all subsets $\mathcal{S} \subseteq H$, such that every computation path of M corresponds to exactly one $\mathcal{S} \subseteq H$. By two calls to an NP-oracle, M checks on every path whether $T \cup \mathcal{S}$ is consistent (i.e., satisfiable) and if $T \cup \mathcal{S} \models M$ holds. If both oracle calls answer "yes", then $\mathcal{S}$ is a solution of $\mathcal{P}$ and the computation path is accepting. The output written by M on each path is the cardinality of the corresponding set $\mathcal{S}$ (resp. the sum of the weights of the elements in $\mathcal{S}$) for the #-$\leq$-ABDUCTION problem (resp. for the #-$\sqsubseteq$-ABDUCTION problem). Finally, we define the optimal value of M to be the minimum. Obviously, the accepting paths of M outputting the optimal value correspond one-to-one to the cardinality-minimal (resp. weight-minimal) solutions of the PAP $\mathcal{P}$.

The hardness of #-$\leq$-ABDUCTION (resp. of #-$\sqsubseteq$-ABDUCTION) is shown by reduction from #MIN-CARD-Π_1SAT (resp. from #MIN-WEIGHT-Π_1SAT). Let an arbitrary instance of #MIN-CARD-Π_1SAT (resp. of #MIN-WEIGHT-Π_1SAT) be given

by the quantified Boolean formula $\varphi(X) = \forall Y\ \psi(X, Y)$ with $X = \{x_1, \ldots, x_k\}$ and $Y = \{y_1, \ldots, y_l\}$. In case of #MIN-WEIGHT-Π_1SAT, we additionally have a weight function w defined on the variables in X. Let $X' = \{x_1', \ldots, x_k'\}$, $X'' = \{x_1'', \ldots, x_k''\}$, $Q = \{q_1, \ldots, q_k\}$, $R = \{r_1, \ldots, r_k\}$, and t be fresh variables. Then we define the PAP $\mathcal{P} = \langle V, H, M, T \rangle$ as follows.

$$V = X \cup X' \cup X'' \cup Y \cup Q \cup R \cup \{t\}, \quad H = X \cup X' \cup X'', \quad M = Q \cup R \cup \{t\}$$
$$T = \{\psi(X, Y) \to t\} \cup \{\neg x_i \vee \neg x_i', x_i \to q_i, x_i' \to q_i \mid i = 1, \ldots, k\}$$
$$\cup \{\neg x_i' \vee \neg x_i'', x_i' \to r_i, x_i'' \to r_i \mid i = 1, \ldots, k\}.$$

In case of #-$\sqsubseteq$-ABDUCTION, we leave the weights of the variables in X unchanged. For the remaining hypotheses, we set $w(x_i) = w(x_i') = w(x_i'')$ for every $i \in \{1, \ldots, k\}$.

For each i, the clauses $\neg x_i \vee \neg x_i', x_i \to q_i, x_i' \to q_i$ in T ensure that every solution $\mathcal{S}$ of $\mathcal{P}$ contains exactly one of $\{x_i, x_i'\}$. Similarly, the clauses $\neg x_i' \vee \neg x_i'', x_i' \to r_i, x_i'' \to r_i$ ensure that every solution contains exactly one of $\{x_i', x_i''\}$. The sets of variables X' and X'' both represent the complement $X \smallsetminus A$, but X'' is there to get the cardinalities right, since without it, the cardinality $|A \cup (X \smallsetminus A)'|$ would be the same for all $\mathcal{S}$.

For a subset of variables $A \subseteq X$, let A' and A'' be defined as $A' = \{x' \mid x \in A\}$ and $A'' = \{x'' \mid x \in A\}$. Then, the effect of the conjunct $\psi(X, Y) \to t$ in T is that, for every subset $A \subseteq X$ the following equivalence holds: The assignment I on X with $I^{-1}(1) = A$ is a model of $\varphi(X)$ if and only if $A \cup (X \smallsetminus A)' \cup \{\psi(X, Y) \to t\} \models \{t\}$. Thus, for every $A \subseteq X$, we have the following equivalences. The assignment I on X with $I^{-1}(1) = A$ is a model of $\varphi(X)$ if and only if $A \cup (X \smallsetminus A)' \cup A''$ is a solution of $\mathcal{P}$. Moreover, the previous assignment I is cardinality-minimal (resp. weight-minimal) if and only if $A \cup (X \smallsetminus A)' \cup A''$ is a cardinality-minimal (resp. a weight-minimal) solution of $\mathcal{P}$. This accomplishes a parsimonious reduction to #-$\leq$-ABDUCTION (resp. #-$\sqsubseteq$-ABDUCTION). $\qquad\square$

#-$\leq_P$-ABDUCTION with no restriction on the number of priorities requires some preparatory work. For this purpose, we first consider the appropriate version of #SAT.

Problem: #MIN-LEX–Π_kSAT
Input: A quantified Boolean formula $\varphi(X) = \forall Y_1 \exists Y_2 \cdots QY_k\ \psi(X, Y_1, \ldots, Y_k)$ and a subset $X' = \{x_1, \ldots, x_\ell\} \subseteq X$, such that $Q = \forall$ (resp. $Q = \exists$) and $\psi(X, Y_1, \ldots, Y_k)$ is in DNF (resp. in CNF) if k is odd (resp. k is even).
Output: Number of satisfying assignments $I: X \to \{0, 1\}$ of the formula $\varphi(X)$, such that $(I(x_1), \ldots, I(x_\ell))$ is lexicographically minimal.

As usual, #MIN-LEX–Π_0SAT represents the aforementioned problem for unquantified formulas, therefore we denote it as #MIN-LEX-SAT.

Theorem 3. *#MIN-LEX–Π_kSAT is #·Opt_{k+1}P-complete. In particular, #MIN-LEX-SAT is #·OptP-complete.*

Proof. We only give the proof for #MIN-LEX-SAT, since the generalization to higher levels of the hierarchy is obvious.

In order to prove the membership, we show that #MIN-LEX-SAT can be solved by an appropriate NP-transducer M. We give a high-level description of M: It takes as input an arbitrary propositional formula φ with variables in X plus a subset $X' = \{x_1, \ldots, x_\ell\} \subseteq X$ of distinguished variables. M non-deterministically enumerates all possible truth assignments $I: X \rightarrow \{0, 1\}$, such that every computation path of M corresponds to exactly one assignment I. On each path, M checks in polynomial time if I is a model of φ. If this is the case, then the computation path is accepting. The output written by M on each path is the binary string $(I(x_1), \ldots, I(x_\ell))$. Finally, we define the optimal value of M to be the minimum. Obviously, the accepting paths of M outputting the optimal value correspond one-to-one to the satisfying assignments I of φ, such that $(I(x_1), \ldots, I(x_\ell))$ is lexicographically minimal.

For the hardness proof, let L be an arbitrary *minimum* problem in #·OptP. We show that there exists a parsimonious reduction from L to #MIN-LEX-SAT. Since L is in #·OptP, there exists an NP-transducer M for L. On input w, the transducer M produces an output of length $\leq p(|w|)$ on every branch for some polynomial p. Without loss of generality, we may assume that M actually produces an output of length exactly $= p(|w|)$. Now let w be an arbitrary instance of L and let $N = p(|w|)$ denote the length of the output on every computation path. Analogously to Cook's theorem (see [6]), there exists a propositional formula φ with variables X, such that there is a one-to-one correspondence between the satisfying truth assignment of φ and the successful computations of M on w. Moreover, X and φ can be extended in such a way that the output on each successful computation path is encoded by the variables $X' = \{x_1, \ldots, x_N\}$, i.e., for every successful computation path π, the truth values $(I(x_1), \ldots, I(x_N))$ of the corresponding model I of φ represent exactly the output on the path π. But then there is indeed a one-to-one correspondence between the computation paths of M on w, such that M outputs the minimum on these paths and the satisfying assignments of the (extended) formula φ, such that the truth values on $(x_1, \ldots, x_N)$ are lexicographically minimal. □

We also need the usual restriction of the previous problem to three literals per clause.

Problem: #MIN-LEX-3SAT
Input: A propositional formula φ in conjunctive normal form over the variables X with at most three literals per clause and a subset $X' = \{x_1, \ldots, x_\ell\} \subseteq X$.
Output: Number of satisfying assignments $I: X \rightarrow \{0, 1\}$ of the formula φ, such that $(I(x_1), \ldots, I(x_\ell))$ is lexicographically minimal.

Since there exists a parsimonious reduction from #SAT to #3SAT (see [13]), the same reduction implies the following consequence of Theorem 3.

Corollary 4. #MIN-LEX-3SAT *is* #·OptP-*complete.*

Theorem 5. #-$\leq_P$-ABDUCTION *without restriction on the number of priorities and* #-$\sqsubseteq_P$-ABDUCTION *with or without restriction on the number of priorities are* #·Opt$_2$P-*complete.* #-$\leq_P$-ABDUCTION *is* #·Opt$_2$P[$\log n$]-*complete if the number of priorities is bounded by a constant.*

Proof. For the membership proof, we slightly modify the $\Sigma_2\mathrm{P}$-transducer M from the membership proof of Theorem 2. Again, M non-deterministically enumerates all subsets $\mathcal{S} \subseteq H$, such that every computation path of M corresponds to exactly one $\mathcal{S} \subseteq H$. By two calls to an NP-oracle, M checks on every path whether $T \cup \mathcal{S}$ is consistent (i.e., satisfiable) and whether $T \cup \mathcal{S} \models M$ holds. If both oracle calls answer "yes", then $\mathcal{S}$ is a solution of $\mathcal{P}$ and the computation path is accepting. Only the output written by M on each path has to be modified with respect to the proof of Theorem 2: Suppose that the input PAP $\mathcal{P}$ has K priorities $H_1, \ldots, H_K$. Then M computes on every computation path the vector $(c_1, \ldots, c_K)$, where c_i is the cardinality (resp. the total weight) of $\mathcal{S} \cap H_i$ for every i. Without loss of generality we may assume for every i that, on all paths, the binary representation of the numbers c_i has identical length (by adding appropriately many leading zeros). Then M simply outputs this vector $(c_1, \ldots, c_K)$, considered as a single number in binary. Finally, we again define the optimal value of M as the minimum. Obviously, the accepting paths of M outputting the optimal value correspond one-to-one to the $\leq_P$-minimal (resp. $\sqsubseteq_P$-minimal) solutions of the PAP $\mathcal{P}$. If there are no restrictions on the number K of priorities or if we consider weight-minimality, then the output of M has polynomial length. Indeed, Since $K \leq |H|$ always holds, because in the extremal case each hypothesis has its own priority class, we need at most $|H|$ bits. The length of each c_i is bounded by $\log|H|$ bits, since $c_i \leq |H|$ holds. We need $O(K \log|H|)$ bits to represent the vector $(c_1, \ldots, c_K)$. If K is constant, this becomes $O(\log|H|)$.

For the hardness part, only the $\#\cdot\mathrm{Opt}_2\mathrm{P}$-hardness of $\#\text{-}\leq_P\text{-ABDUCTION}$ without restriction on the number of priorities has to be shown. The remaining cases follow from the corresponding hardness result without priorities in Theorem 2. We reduce the $\#\text{MIN-LEX–}\Pi_1\text{SAT}$ problem to $\#\text{-}\leq_P\text{-ABDUCTION}$. Let an arbitrary instance of $\#\text{MIN-LEX–}\Pi_1\text{SAT}$ be given by the quantified Boolean formula $\varphi(X) = \forall Y\ \psi(X, Y)$ with $X = \{x_1, \ldots, x_n\}$ and the subset $X' = \{x_1, \ldots, x_\ell\} \subseteq X$. Let t, $Q = \{q_1, \ldots, q_n\}$ $R = \{r_1, \ldots, r_\ell\}$, $Z = \{z_1, \ldots, z_n\}$, and $Z' = \{z_1', \ldots, z_\ell'\}$ be fresh variables. Then we define the PAP $\mathcal{P} = \langle V, H, M, T \rangle$ as follows:

$$V = X \cup Y \cup Z \cup Z' \cup Q \cup R \cup \{t\}$$
$$H = X \cup Z \cup Z' \text{ with}$$
$$\qquad H_1 = \{x_1\}, \ldots, H_\ell = \{x_\ell\}, \text{ and } H_{\ell+1} = (X \setminus X') \cup Z \cup Z'$$
$$M = Q \cup R \cup \{t\}$$
$$T = \{\psi(X, Y) \to t\} \cup \{\neg x_i \vee \neg z_i, x_i \to q_i, z_i \to q_i \mid 1 \leq i \leq n\}$$
$$\qquad \cup \{\neg z_i \vee \neg z_i', z_i \to r_i, z_i' \to r_i \mid 1 \leq i \leq \ell\}$$

The idea of the variables in Q, R, Z, and Z' is similar to the the variables Q, R, X', and X'' in the proof of Theorem 2. They ensure that every solution $\mathcal{S}$ of $\mathcal{P}$ contains exactly n variables out of the $2n$ variables in $H_{\ell+1}$. This can be seen as follows. By the clauses $\neg x_i \vee \neg z_i, x_i \to q_i, z_i \to q_i$ with $i \in \{1, \ldots, n\}$, every solution contains exactly one of $\{x_i, z_i\}$. Of course, the variables x_i with $i \in \{1, \ldots, \ell\}$ are not in $H_{\ell+1}$. However, the clauses $\neg z_i \vee \neg z_i', z_i \to r_i, z_i' \to r_i$ with $i \in \{1, \ldots, \ell\}$ ensure that every solution contains exactly one of $\{z_i, z_i'\}$. In other words, for every $i \in \{1, \ldots, \ell\}$ every solution contains either $\{x_i, z_i'\}$ or $\{z_i\}$.

There is a one-to-one correspondence between the models of $\varphi(X)$ which are lexicographically minimal on X' and the $\leq_P$-minimal solutions of $\mathcal{P}$. Indeed, let I be a model of $\varphi(X)$ which is lexicographically minimal on X'. Then I can be extended to exactly one $\leq_P$-minimal solution $\mathcal{S}$ of $\mathcal{P}$, namely $\mathcal{S} = I^{-1}(1) \cup \{z_i \mid 1 \leq i \leq n$ and $I(x_i) = 0\} \cup \{z_i' \mid 1 \leq i \leq \ell$ and $I(x_i) = 1\}$.

Conversely, let $\mathcal{S}$ be a $\leq_P$-minimal solution of $\mathcal{P}$. Then we obtain a lexicographically minimal model I of $\varphi(X)$ simply by restricting $\mathcal{S}$ to X, i.e. $I(x) = 1$ for all $x \in \mathcal{S} \cap X$ and $I(x) = 0$ otherwise. $\qquad\square$

4 Special Cases

We consider the special cases of propositional abduction problems, where the theory is presented by Horn, definite Horn, dual Horn, or bijunctive formulas. Recall the following counting problem introduced in [10].

Problem: #MIN-CARD-VERTEX-COVER (RESP. #MIN-WEIGHT-VERTEX-COVER)
Input: Graph $G = (V, E)$ (plus a weight function $w \colon V \to \mathbb{N}$ in case of #MIN-WEIGHT-VERTEX-COVER).
Output: Number of vertex covers of G with minimal cardinality (resp. with minimal weight), i.e., cardinality-minimal (resp. weight-minimal) subsets $C \subseteq V$ such that $(u, v) \in E$ implies $u \in C$ or $v \in C$.

In [10], it was shown that #MIN-CARD-VERTEX-COVER is $\#{\cdot}\text{OptP}[\log n]$-complete while #MIN-WEIGHT-VERTEX-COVER is $\#{\cdot}\text{OptP}$-complete.

Theorem 6. *#-$\leq$-ABDUCTION is $\#{\cdot}\text{OptP}[\log n]$-complete and #-$\sqsubseteq$-ABDUCTION is $\#{\cdot}\text{OptP}$-complete for Horn, definite Horn, dual Horn, or bijunctive theories.*

Proof. For the membership part, we construct a transducer M exactly as in the proof of Theorem 2. The only difference is that we can now check in *deterministic polynomial time* whether $T \cup \mathcal{S}$ is consistent (i.e., satisfiable) and whether $T \cup \mathcal{S} \models M$ holds. Hence, we end up with the desired NP-transducer (rather than a Σ_2P-transducer) since we no longer need an NP-oracle.

The hardness is shown by a reduction from #MIN-CARD-VERTEX-COVER (resp. #MIN-WEIGHT-VERTEX-COVER). Let an arbitrary instance of #MIN-CARD-VERTEX-COVER be given by the graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ and $E = \{e_1, \ldots, e_m\}$. By slight abuse of notation, we consider the elements in V and E also as propositional variables and set $X = \{v_1, \ldots, v_n\}$ and $R = \{e_1, \ldots, e_m\}$. In case of #MIN-WEIGHT-VERTEX-COVER, we additionally have a weight function w defined on the variables in X. Then we define the PAP $\mathcal{P} = \langle W, H, M, T \rangle$ as follows.

$$W = X \cup R, \quad H = X, \quad M = R$$
$$T = \{v_i \to e_j \mid v_i \in e_j, 1 \leq i \leq n, 1 \leq j \leq m\}$$

The resulting theory contains only clauses which are, at the same time, Horn, definite Horn, dual Horn, and bijunctive. Obviously, for every subset $X' \subseteq X = V$

the following equivalence holds: X' is a solution of $\mathcal{P}$ if and only if X' is a vertex cover of G. But then there exists also a one-to-one correspondence between the cardinality-minimal (resp. weight-minimal) solutions of $\mathcal{P}$ and the cardinality-minimal (resp. weight-minimal) vertex covers of G. □

Again, #-$\leq_P$-ABDUCTION with no restriction on the number of priorities requires some preparatory work. For this purpose, we first consider an appropriate variant of counting the vertex covers of a graph.

Problem: #MIN-LEX-VERTEX-COVER
Input: Graph $G = (V, E)$ and a subset $V' = \{v_1, \dots, v_\ell\} \subseteq V$.
Output: Number of vertex covers C of G, such that $(\chi(v_1), \dots, \chi(v_\ell))$ is lexicographically minimal, where χ is the characteristic function of the vertex cover C.

Theorem 7. *#MIN-LEX-VERTEX-COVER is #·OptP-complete.*

Proof. In order to prove the membership, we show that #MIN-LEX-VERTEX-COVER can be solved by the following NP-transducer M. It takes as input an arbitrary graph $G = (V, E)$ with distinguished vertices $V' = \{v_1, \dots, v_\ell\}$. M non-deterministically enumerates all subsets $C \subseteq V$, such that every computation path of M corresponds to exactly one such subset C. If C is a vertex cover of G, then the computation path is accepting. The output written by M on each path is the binary vector $(\chi_C(v_1), \dots, \chi_C(v_\ell))$. Obviously, the accepting paths of M outputting the minimal value correspond one-to-one to the vertex covers C of G, such that $(\chi_C(v_1), \dots, \chi_C(v_\ell))$ is lexicographically minimal.

The hardness proof is by a parsimonious reduction from #MIN-LEX-3SAT. In fact, this is the same reduction as in the standard NP-completeness proof of VERTEX COVER by reduction from 3SAT to VERTEX COVER, see e.g. [6]. Let $\varphi(x_1, \dots, x_k)$ be a propositional formula in CNF with three literals per clause. We construct the graph $G = (V, E)$ as follows. For each variable x_i we construct an edge $e_i = (x_i, x_i')$. For each clause $c_i = l_i^1 \vee l_i^2 \vee l_i^3$ we construct three edges $(l_i^1, l_i^2), (l_i^2, l_i^3), (l_i^3, l_i^1)$ forming a triangle t_i. Finally, we connect each positive literal z in the triangle t_i to its counterpart z in an edge $e_j = (z, z')$, as well as each negative literal $\neg z$ in the triangle t_i to its counterpart z'. The set of distinguished variables X' from #MIN-LEX-3SAT becomes the set of distinguished vertices V' in #MIN-LEX-VERTEX-COVER. □

Theorem 8. *#-$\leq_P$-ABDUCTION without restriction on the number of priorities and #-$\sqsubseteq_P$-ABDUCTION with or without restriction on the number of priorities are #·OptP-complete for Horn, definite Horn, dual Horn, or bijunctive theories. #-$\leq_P$-ABDUCTION for Horn, definite Horn, dual Horn, or bijunctive theories is #·OptP[$\log n$]-complete if the number of priorities is restricted by a constant.*

Proof. For the membership part, we construct a transducer M exactly as in the proof of Theorem 5. The only difference is that we get an NP-transducer (rather than a Σ_2P-transducer) since we no longer need an NP-oracle for checking whether $T \cup \mathcal{S}$ is consistent (i.e., satisfiable) and whether $T \cup \mathcal{S} \models M$ holds.

For the hardness part, only the #·OptP-hardness of #-$\leq_P$-ABDUCTION without restriction on the number of priorities has to be shown. The remaining cases follow

from the corresponding hardness result without priorities in Theorem 6. Let an arbitrary instance of #Min-Lex-Vertex-Cover be given by the graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ and $E = \{e_1, \ldots, e_m\}$ and let $V' = \{v_1, \ldots, v_\ell\}$ with $\ell \leq n$. As in the proof of Theorem 6, we consider the elements in V and E also as propositional variables and set $X = \{v_1, \ldots, v_n\}$ and $R = \{e_1, \ldots, e_m\}$. In addition, let $Q = \{q_{\ell+1}, \ldots, q_n\}$, and $Z = \{z_{\ell+1}, \ldots, z_n\}$ be fresh variables. Then we define the PAP $\mathcal{P} = \langle V, H, M, T \rangle$ as follows.

$$V = X \cup R \cup Q \cup Z, \quad M = R \cup Q$$
$$H = X \cup Z \text{ with } H_1 = \{v_1\}, \ldots, H_\ell = \{v_\ell\}, \text{ and } H_{\ell+1} = (X \smallsetminus V') \cup Z$$
$$T = \{v_i \rightarrow e_j \mid v_i \in e_j, 1 \leq i \leq n, 1 \leq j \leq m\} \cup$$
$$\{v_i \rightarrow q_i, z_i \rightarrow q_i \mid \ell + 1 \leq i \leq n\}$$

The resulting theory contains only clauses which are, at the same time, Horn, definite Horn, dual Horn, and bijunctive. The variables Q and Z realize the familiar idea that in every $\leq_P$-minimal solution $\mathcal{S}$ of $\mathcal{P}$, for every $i \in \{\ell + 1, \ldots, n\}$, exactly one of v_i and z_i is contained in $\mathcal{S}$. It can then be easily shown that there is a one-to-one correspondence between the lexicographically minimal vertex covers of G and the $\leq_P$-minimal solutions of $\mathcal{P}$. $\qquad\square$

5 Conclusion

In this paper, we have completed the analysis of the counting complexity of propositional abduction. Together with previous results presented in [9], we have thus achieved a full picture. Recall from [19] that counting problems may display a significantly different complexity behavior from the corresponding decision problems. Hence, the complexity of a class of problems is better understood when we analyse the counting complexity in addition to the decision complexity. By complementing the complexity results of Eiter and Gottlob [2] on decision problems related to propositional abduction with our counting complexity results in Table 1, we have thus arrived at a better understanding of the complexity of propositional abduction in various settings.

From a complexity theoretic point of view, there is another interesting aspect to the counting complexity results shown here. The class $\#\mathrm{P}$ has been studied intensively and many completeness results for this class can be found in the literature. In contrast, for the higher counting complexity classes $\#\cdot\Pi_k\mathrm{P}$, $\#\cdot\mathrm{Opt}_k\mathrm{P}[\log n]$, and $\#\cdot\mathrm{Opt}_k\mathrm{P}$ (with $k \geq 1$) very few problems had been shown to be complete. Our results on the counting complexity of propositional abduction thus also lead to a better understanding of these counting complexity classes.

For future work, we plan to extend the complexity analysis of many more families of decision problems in the artificial intelligence domain (like, e.g., closed-world reasoning in various settings) to counting problems. Moreover, we would also like to extend the abduction cases studied in this paper to yet another case, namely the case of affine theories, i.e.: the theory T is an affine system $AX = b$ over $\mathbb{Z}_2$. This case was in fact dealt with in [9] for $\#$-$\preceq$-abduction with $\preceq \in \{=, \subseteq, \subseteq_P\}$. There are obvious upper and lower bounds also for $\#$-$\preceq$-abduction with affine theories when the preorder $\preceq$ is in

$\{\leq, \sqsubseteq, \leq_P, \sqsubseteq_P\}$. However, proving tight complexity bounds also for these cases has to be left as an open problem for future work.

References

1. A. Durand, M. Hermann, and P. G. Kolaitis. Subtractive reductions and complete problems for counting complexity classes. *Theoretical Computer Science*, 340(3):496–513, 2005.
2. T. Eiter and G. Gottlob. The complexity of logic-based abduction. *Journal of the Association for Computing Machinery*, 42(1):3–42, 1995.
3. T. Eiter and K. Makino. On computing all abductive explanations. In *Proc. 18th AAAI, Edmonton (Alberta, Canada)*, pages 62–67. AAAI Press, 2002.
4. T. Eiter and K. Makino. Generating all abductive explanations for queries on propositional Horn theories. In M. Baaz and J. A. Makowsky (eds), *Proc. 12th CSL and 8th KGC, Vienna (Austria)*, LNCS 2803, pages 197–211. Springer, 2003.
5. R. Fagin, J. D. Ullman, and M. Y. Vardi. On the semantics of updates in databases. In *Proc. 2nd PODS, Atlanta (Georgia, USA)*, pages 352–265. ACM, 1983.
6. M. R. Garey and D. S. Johnson. *Computers and intractability: A guide to the theory of NP-completeness*. W.H. Freeman and Co, 1979.
7. W. I. Gasarch, M. W. Krentel, and K. J. Rappoport. OptP as the normal behavior of NP-complete problems. *Mathematical Systems Theory*, 28(6):487–514, 1995.
8. L. A. Hemaspaandra and H. Vollmer. The satanic notations: Counting classes beyond #P and other definitional adventures. *SIGACT News*, 26(1):2–13, 1995.
9. M. Hermann and R. Pichler. Counting complexity of propositional abduction. In M. M. Veloso (ed), *Proc. 20th IJCAI 2007, Hyderabad (India)*, pages 417–422. AAAI Press, 2007.
10. M. Hermann and R. Pichler. Complexity of counting the optimal solutions. In X. Hu and J. Wang (eds), *Proc. 14th COCOON, Dalian (China)*, LNCS 5092, pages 149–159. Springer, 2008.
11. A. Herzig, J. Lang, P. Marquis, and T. Polacsek. Updates, actions, and planning. In B. Nebel (ed), *Proc. 17th IJCAI 2001, Seattle (Washington, USA)*, pages 119–124. Morgan Kaufmann, 2001.
12. A. C. Kakas and P. Mancarella. Database updates through abduction. In D. McLeod, R. Sacks-Davis, and H.-J. Schek (eds), *Proc. 16th VLDB, Brisbane (Queensland, Australia)*, pages 650–661. Morgan Kaufmann, 1990.
13. D. C. Kozen. *The design and analysis of algorithms*, chapter 26: Counting problems and #P, pages 138–143. Springer, 1992.
14. M. W. Krentel. The complexity of optimization problems. *Journal of Computer and System Sciences*, 36(3):490–509, 1988.
15. M. W. Krentel. Generalizations of OptP to the polynomial hierarchy. *Theoretical Computer Science*, 97(2):183–198, 1992.
16. I. Papatheodorou, A. C. Kakas, and M. J. Sergot. Inference of gene relations from microarray data by abduction. In Ch. Baral, G. Greco, N. Leone, and G. Terracina (eds), *Proc. 8th LPNMR, Diamante (Italy)*, LNCS 3662, pages 389–393. Springer, 2005.
17. Y. Peng and J. A. Reggia. *Abductive inference models for diagnostic problem solving*. Springer, 1990.
18. D. Poole. Probabilistic Horn abduction and bayesian networks. *Artificial Intelligence*, 64(1):81–129, 1993.
19. L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979.
20. L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.