Reasoning with Models

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Reasoning with Models[*]

Roni Khardon[†]   **Dan Roth**[‡]

Aiken Computation Laboratory,
Harvard University,
Cambridge, MA 02138.
{roni,danr}@das.harvard.edu

January 1994
(Revised June 30, 1994)

Abstract

We develop a model-based approach to reasoning, in which the knowledge base is represented as a set of models (satisfying assignments) rather then a logical formula, and the set of queries is restricted. We show that for every propositional knowledge base (KB) there exists a set of *characteristic models* with the property that a query is true in KB if and only if it is satisfied by the models in this set. We fully characterize a set of theories for which the model-based representation is compact and provides efficient reasoning. These include cases where the formula-based representation does not support efficient reasoning. In addition, we consider the model-based approach to *abductive reasoning* and show that for any propositional KB, reasoning with its model-based representation yields an abductive explanation in time that is polynomial in its size. Some of our technical results make use of the *Monotone Theory*, a new characterization of Boolean functions introduced in [Bsh93].

The notion of *restricted queries* is inherent to our approach. This is a wide class of queries for which reasoning is very efficient and exact, even when the model-based representation KB provides only an approximate representation of the "world".

Moreover, we show that the theory developed here generalizes the model-based approach to reasoning with Horn theories [KKS93], and captures even the notion of reasoning with Horn-approximations [SK91]. Our result characterizes the Horn theories for which the approach suggested in [KKS93] is useful and the phenomena observed there, regarding the relative sizes of the formula-based representation and model-based representation of KB is explained and put in a wider context.

1 Introduction

A widely accepted framework for reasoning in intelligent systems is the knowledge-based system approach [McC58]. The idea is to store the knowledge in some *representation language* with a well defined meaning assigned to those sentences. The sentences are stored in a Knowledge Base (KB) which is combined with a reasoning mechanism, used to determine what can be inferred from the sentences in the KB. Reasoning is abstracted as a deduction task of determining whether a sentence α, assumed to capture the situation at hand, is implied from KB (denoted KB $\models \alpha$). This can be understood as the question: "is α consistent with the current state of knowledge ?"

Solving the question KB $\models \alpha$, even in the propositional case, is co-NP-Hard and under the current complexity theoretic beliefs requires exponential time. Many other forms of reasoning which have been developed, at least partly, to avoid these computational difficulties were also shown to be hard to compute [Sel90, Rot93].

A significant amount of recent work on reasoning is influenced by convincing arguments of Levesque [Lev86] who argued that common-sense reasoning is a distinct mode of reasoning and that we should give a computational theory that accounts for both its speed and flexibility. Most of the work in this direction still views reasoning as a kind of theorem proving process and is based on the belief that a careful study of the sources of the computational difficulties may lead to a formalism expressive enough to capture common-sense knowledge, while still allowing for efficient reasoning. Thus, the work aims at identifying classes of limited expressiveness, with which one can perform theorem proving efficiently [BL84, Lev92, Rot93, Sel90]. None of these works, however, meets the strong tractability requirements required for common-sense reasoning (e.g. see [Sha93]), even though the limited expressiveness of classes discussed there has been argued to be implausible [DP91].

Levesque argues [Lev86, Lev92] that reasoning with a more direct representation is easier and better suits common-sense reasoning. He suggests to represent the knowledge base KB in a vivid form, one that bears a strong and direct relationship to the real world. This might be just a model of KB [EBBK89, Pap91] on which one can evaluate the truth value of the query α. It is not clear, however, how one might derive a vivid form of the knowledge base, and as expected, selecting a model that is, informally, the most likely model of the real world based on any reasonable criterion is a computationally hard task [Pap91, SK90]. Most importantly, in order to achieve an efficient solution to the reasoning problem this approach modifies the problem: reasoning with a vivid representation no longer solves the problem KB $\models \alpha$, but rather a different problem, whose exact relation to the original inference problem depends on the method selected to simplify the knowledge base.

A Model-Based Approach to Reasoning

In this work we embark on the development of a model-based approach to common sense reasoning. It is not hard to motivate a model-based approach to reasoning from a cognitive point of view and indeed, most of the proponents of this approach to reasoning have been cognitive psychologists [JL83, JLB91, Kos83]. In the AI community this approach can be seen as an example of Levesque's notion of "vivid" reasoning and has already been studied in [KKS93].

The problem KB $\models \alpha$ can be approached using the following model-based strategy:

Test Set: A set S of possible assignments.

Test: If there is an element $x \in S$ which satisfies KB, but does not satisfy α, deduce that KB $\not\models \alpha$; Otherwise, KB $\models \alpha$.

Clearly, this approach solves the inference problem if S is the set of *all* models (satisfying assignments) of KB, but this set might be too large. A model-based approach becomes useful if one can show that it is possible to use a fairly small set of models as the Test Set, and still perform reasonably good inference, under some criterion.

We define a set of models, the *characteristic models* of the knowledge base, and show that performing the model-based test with it suffices to deduce that KB $\models \alpha$, for a restricted set of queries. We prove that for a fairly wide class of representations, this set is sufficiently small, and thus the model-based approach is feasible. The notion of *restricted queries* is inherent to our approach; since we are interested in formalizing common-sense reasoning, we take the view that a reasoner need not answer efficiently *all* possible queries. For a wide class of queries we show that exact reasoning can be done efficiently, even when the reasoner keeps in KB an "approximate" representation of the "world".

We show that the theory developed here generalizes the model-based approach to reasoning with Horn theories, suggested in [KKS93], and captures even the notion of reasoning with approximate theories [SK91]. In particular, our results characterize the Horn theories for which the approach suggested in [KKS93] is useful and explain the phenomena observed there, regarding the relative sizes of the logical formula representation and model-based representation of KB. We also give other examples of expressive families of propositional theories, for which our approach is useful.

In addition, we consider the problem of performing *abduction* using a model-based approach and show that for any propositional knowledge base, using a model-based representation yields an abductive explanation in time that is polynomial in the size of the model-based representation. Some of our technical results make use of a new characterization of Boolean functions, called the *Monotone Theory*, introduced recently by Bshouty [Bsh93].

Most of the work on reasoning assumes that the knowledge base is given in some form, and the question of how this knowledge might be acquired is not considered. While in this paper we also take this point of view, we are interested in studying the entire process of *learning* a knowledge base representation and reasoning with it. In particular, Bshouty [Bsh93] gives an algorithm that learns the model-based representation we consider here when given access to a *Membership Oracle* and an *Equivalence Oracle*. In [KR94a] we discuss the issue of "learning to reason" and illustrate the importance of the model-based approach for this problem.

Summary of Results

We now briefly describe the main applications of the model-based approach developed in this paper. We consider two types of queries with which reasoning is efficient. Queries are called *relevant* if they belong to the propositional language that represents the "world". Queries are called *common* if they belong to some set $\mathcal{L}_C$ of *common* propositional languages[1] (see Definition 6.2). These include for example Horn queries and $\log n$ CNF queries. Our results can be grouped into 3 categories that can be informally described as follows:
(1) A function has a small set of characteristic models if either (a) it has a small DNF representation and a small CNF representation, or (b) it is in $\mathcal{L}_C$ and it has a small DNF representation.

For these functions, model-based deduction is correct and efficient for relevant and for common queries.
(2) The set $\Gamma^{\mathcal{G}}$, of characteristic models with respect to a propositional language $\mathcal{G}$, describes the least upper bound of f with respect to $\mathcal{G}$.

---

[1]We use interchangeably the terms propositional expression and Boolean function. Similarly, a family of Boolean functions is used interchangeably with a propositional language.

Model-based deduction, using $\Gamma^{\mathcal{G}}$, is correct and efficient for queries in $\mathcal{G}$.
(3) For the functions defined in (1), efficient and correct model-based abduction can be performed.

Clearly, our algorithms do not solve NP-complete problems. Most hardness results for reasoning assume that KB is given as a CNF formula. The fact that we can perform reasoning efficiently relies on the fact that we change the knowledge representation into a more accessible form (another knowledge representation which enables reasoning, yet for some reason is considered less interesting, is DNF). Another important point is that our algorithms do not need to know the CNF nor the DNF representation of KB. We only require that a polynomial size representation exists.

The rest of this paper is organized as follows: In Section 2 we introduce some notations and review the monotone theory. In Section 3 we consider the deduction problem. We introduce the set of characteristic models, and analyze the correctness and efficiency of model-based deduction with this set. In Section 4 we show that in the case of Horn theories our theory reduces to the work in [KKS93]. In Section 5 we discuss the size of model-based representations. In Section 6 we discuss applications of the our theory to particular propositional languages. In Section 7 we consider the abduction problem, and in Section 8 we conclude with some reference to future work.

2   Monotone Theory

In this section we introduce the notations, definitions and results of the Monotone Theory of Boolean functions, introduced by Bshouty [Bsh93].

We consider a Boolean function $f : \{0,1\}^n \to \{0,1\}$. The elements in the set $\{x_1, \ldots, x_n\}$ are called variables. Assignments in $\{0,1\}^n$ are denoted by x, y, z, and $weight(x)$ denotes the number of 1 bits in the assignment x. A literal is either a variable x_i (called a positive literal) or its negation $\overline{x_i}$ (a negative literal). A clause is a disjunction of literals, and a CNF formula is a conjunction of clauses. For example $(x_1 \vee \overline{x_2}) \wedge (x_3 \vee \overline{x_1} \vee x_4)$ is a CNF formula with two clauses. A term is a conjunction of literals, and a DNF formula is a disjunction of terms. For example $(x_1 \wedge \overline{x_2}) \vee (x_3 \wedge \overline{x_1} \wedge x_4)$ is a DNF formula with two terms. A CNF formula is Horn if every clause in it has at most one positive literal. We note that every Boolean function has many possible representations, and in particular both a CNF representation and a DNF representation. The size of CNF and DNF representation are, respectively, the number of clauses and the number of terms in the representation.

An assignment $x \in \{0,1\}^n$ satisfies f if $f(x) = 1$. Such an assignment x is also called a model of f. By "f implies g", denoted $f \models g$, we mean that every model of f is also a model of g. Throughout the paper, when no confusion can arise, we identify a Boolean function f with the set of its models, namely $f^{-1}(1)$. Observe that the connective "implies" ($\models$) used between Boolean functions is equivalent to the connective "subset or equal" ($\subseteq$) used for subsets of $\{0,1\}^n$. That is, $f \models g$ if and only if $f \subseteq g$.

Definition 2.1 (Order) *We denote by $\leq$ the usual partial order on the lattice $\{0,1\}^n$, the one induced by the order $0 < 1$. That is, for $x, y \in \{0,1\}^n$, $x \leq y$ if and only if $\forall i, x_i \leq y_i$. For an assignment $b \in \{0,1\}^n$ we define $x \leq_b y$ if and only if $x \oplus b \leq y \oplus b$ (where $\oplus$ is the bitwise addition modulo 2).*

Intuitively, if $b_i = 0$ then the order relation on the ith bit is the normal order; if $b_i = 1$, the order relation is reversed and we have that $1 <_{b_i} 0$.

The *monotone extension of $z \in \{0,1\}^n$* with respect to b is:

$$\mathcal{M}_b(z) = \{x \mid x \geq_b z\}.$$

The *monotone extension of f* with respect to b is:

$$\mathcal{M}_b(f) = \{x \mid x \geq_b z, \; for \; some \; z \in f\}.$$

The set of *minimal assignments of f* with respect to b is:

$$\min_b(f) = \{z \mid z \in f, \;\; such \; that \; \forall y \in f, z \not\succ_b y\}.$$

The following claim lists some properties of $\mathcal{M}_b$, all are immediate from the definitions:

Claim 2.1 *Let $f, g : \{0,1\}^n \to \{0,1\}$ be Boolean functions. The operator $\mathcal{M}_b$ satisfies the following properties:*
(1) If $f \subseteq g$ then $\mathcal{M}_b(f) \subseteq \mathcal{M}_b(g)$.
(2) $\mathcal{M}_b(f \wedge g) \subseteq \mathcal{M}_b(f) \wedge \mathcal{M}_b(g)$.
(3) $\mathcal{M}_b(f \vee g) = \mathcal{M}_b(f) \vee \mathcal{M}_b(g)$.
(4) $f \subseteq \mathcal{M}_b(f)$.

Claim 2.2 *Let $z \in f$. Then, for every $b \in \{0,1\}^n$, there exists $u \in \min_b(f)$ such that $\mathcal{M}_b(z) \subseteq \mathcal{M}_b(u)$.*

Proof: If $z \notin \min_b(f)$ then $\exists y \in f$ such that $y \leq_b z$. Let u be a minimal element in f with this property. Then, $u \in \min_b(f)$ and clearly $\{x \mid x \geq_b z\} \subseteq \{x \mid x \geq_b u\}$, as needed. ∎

Using Claims 2.2 and 2.1 we get a characterization of the monotone extension of f:

Claim 2.3 *The monotone extension of f with respect to b is:*

$$\mathcal{M}_b(f) = \bigvee_{z \in f} \mathcal{M}_b(z) = \bigvee_{z \in \min_b(f)} \mathcal{M}_b(z).$$

Clearly, for every assignment $b \in \{0,1\}^n$, $f \subseteq \mathcal{M}_b(f)$. Moreover, if $b \notin f$, then $b \notin \mathcal{M}_b(f)$ (since b is the smallest assignment with respect to the order $\leq_b$). Therefore:

$$f = \bigwedge_{b \in \{0,1\}^n} \mathcal{M}_b(f) = \bigwedge_{b \notin f} \mathcal{M}_b(f).$$

The question is if we can find a small set of negative examples b, and use it to represent f as above.

Definition 2.2 (Basis) *A set B is a* basis *for f if $f = \bigwedge_{b \in B} \mathcal{M}_b(f)$. B is a basis for a class of functions $\mathcal{F}$ if it is a basis for all the functions in $\mathcal{F}$.*

Using this definition, the representation

$$f = \bigwedge_{b \in B} \mathcal{M}_b(f) = \bigwedge_{b \in B} \bigvee_{z \in \min_b(f)} \mathcal{M}_b(z) \tag{1}$$

yields the following necessary and sufficient condition describing when $x \in \{0,1\}^n$ is positive for f:

Corollary 2.4 *Let B be a basis for f, $x \in \{0,1\}^n$. Then, $x \in f$ (i.e., $f(x) = 1$) if and only if for every basis element $b \in B$ there exists $z \in \min_b(f)$ such that $x \geq_b z$.*

The following claim bounds the size of the basis of a function f:

Claim 2.5 *Let $f = C_1 \wedge C_2 \wedge \cdots \wedge C_k$ be a CNF representation for f and let B be a set of assignments in $\{0,1\}^n$. If every clause C_i is falsified by some $b \in B$ then B is a basis for f. In particular, f has a basis of size $\leq k$.*

Proof: Let $B = \{b^1, b^2, \ldots, b^k\}$ be a collection of assignments such that b^i falsifies C_i. We show that $f = \bigwedge_{b \in B} \mathcal{M}_b(f)$. First observe that using Claim 2.1 part (4) we get $f \subseteq \bigwedge_{b \in B} \mathcal{M}_b(f)$. In order to show $f \supseteq \bigwedge_{b \in B} \mathcal{M}_b(f)$ we show that for all $y \notin f$ there exists $b \in B$ such that $y \notin \mathcal{M}_b(f)$, and therefore $y \notin \bigwedge_{b \in B} \mathcal{M}_b(f)$.

Consider $y \in \{0,1\}^n$ such that $y \notin f$ and assume, w.l.o.g, that $C_1(y) = 0$. Let $b = b^1$ be the corresponding element in B, and assume, by way of contradiction that $\mathcal{M}_b(f)(y) = 1$. Then , by Corollary 2.4 there exists $z \in \min_b(\mathcal{M}_b(f)) = \min_b(f)$ such that $z \leq_b y$. We therefore have that $b \leq_b z \leq_b y$. Let x_i be a variable the appears in the clause C_1. Since $C_1(y) = C_1(b) = 0$, we must have $y_i = z_i = b_i$. Since this holds for all variables that appear in C_1, it implies that $C_1(z) = 0$ and contradicts the assumption that $z \in f$. ∎

The set of *floor* assignments of an assignment x, with respect to the order relation b, denoted $\lfloor x \rfloor_b$, is the set of all elements $z <_b x$ such that there does not exist y for which $z <_b y <_b x$ (i.e., z is strictly smaller than x relative to b and is different from x in exactly one bit).

The set of *local minimal assignments of* f with respect to b is:

$$\min_b^*(f) = \{x \mid x \in f, \text{ and } \forall y \in \lfloor x \rfloor_b, \ y \notin f\}.$$

The following claims bound the size of $\min_b(f)$:

Claim 2.6 *Let $f = D_1 \vee D_2 \vee \cdots \vee D_k$ be a DNF representation for f. Then for every $b \in \{0,1\}^n$, $|\min_b^*(f)| \leq k$.*

Proof: Let D be one of the terms in the representation, and let p be the number of literals in D. That is $D = \bigwedge_{j=1}^p x_{i_j}^{c_{i_j}}$ (here $x^1 = x$ and $x^0 = \overline{x}$). Clearly, the set $\min_b(D) = \min_b^*(D)$ contains a single element, z, defined by $z_i = c_i$ if x_i appears in D and $z_i = b_i$ if x_i does not appear in D. Further, for any two functions g_1, g_2, $\min_b^*(g_1 \cup g_2) \subseteq \min_b^*(g_1) \cup \min_b^*(g_2)$ and therefore $|\min_b^*(f)| \leq |\cup_{i=1}^k \min_b^*(D_i)| \leq k$. ∎

Corollary 2.7 *Let $f = D_1 \vee D_2 \vee \cdots \vee D_k$ be a DNF representation for f. Then for every $b \in \{0,1\}^n$, $|\min_b(f)| \leq k$.*

Proof: This follows from Claim 2.6, observing that by definition $\min_b(f) \subseteq \min_b^*(f)$. ∎

Example: Let f have the CNF representation:

$$f = (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_4) \wedge (\overline{x_1} \vee \overline{x_2} \vee x_3 \vee \overline{x_4})$$

The function f has 12 (out of the 16 possible) satisfying assignments. The non-satisfying assignments of f are[2]: $\{0000, 0001, 0010, 1101\}$. Using Claim 2.5 we get that the set $B = \{0000, 1101\}$ is a basis for f.

---

[2]An element of $\{0,1\}^n$ denotes an assignment to the variables $x_1, \ldots, x_n$ (i.e., 0011 means $x_1 = x_2 = 0$, and $x_3 = x_4 = 1$).

The sets of minimal assignments with respect to this basis are: $\min_{0000}(f) = \{1000, 0100, 0011\}$ and $\min_{1101}(f) = \{1100, 1111, 1001, 0101\}$. These can be easily found by drawing the corresponding lattices and checking which of the satisfying assignments of f are minimal. It is also easy to check that f can be represented as in equation (1) using the minimal elements identified.

3   Deduction with Models

We consider the deduction problem $KB \models \alpha$. KB is the knowledge base, which is taken to be a propositional expression (i.e., some Boolean function), and α is also a propositional expression. The assertion $KB \models \alpha$ means that if some model $x \in \{0,1\}^n$ satisfies KB, then it must also satisfy α.

Let $\Gamma \subseteq KB \subseteq \{0,1\}^n$ be a set of models. To decide whether $KB \models \alpha$ we consider the straight forward model-based approach to deduction: for all the models $z \in \Gamma$ check whether $\alpha(z) = 1$. If for some z, $\alpha(z) = 0$, say "no"; otherwise say "yes".

By definition, if Γ=KB this approach yields correct deduction, but representing KB by explicitly holding *all* the possible models of KB is not plausible. A model-based approach becomes feasible if one can make correct inferences when working with a small subset of models.

In this section we define a special collection Γ of *characteristic models* of KB and show that performing the model-based test on Γ yields correct deduction. We fully characterize Γ in terms of the Boolean function KB and the query α. Thus we can explore the trade-off between the logical and a model-based approach in terms of size of representation as well as efficiency of reasoning.

3.1   Exact Deduction

Definition 3.1 *Let $\mathcal{F}$ be a class of functions, and let B be a basis for $\mathcal{F}$. For a knowledge base $KB \in \mathcal{F}$ we define the set $\Gamma = \Gamma_{KB}^{B}$ of* characteristic models *to be the set of of all minimal assignments of KB with respect to the basis B. Formally,*

$$\Gamma_{KB}^{B} = \cup_{b \in B} \{z \in min_b(KB)\}.$$

Theorem 3.1 *Let $KB, \alpha \in \mathcal{F}$ and let B be a basis for $\mathcal{F}$. Then $KB \models \alpha$ if and only if for every $u \in \Gamma_{KB}^{B}$, $\alpha(u) = 1$.*

Proof: Clearly, $\Gamma = \Gamma_{KB}^{B} \subseteq$ KB and therefore, if there exists $z \in \Gamma$ such that $\alpha(z) = 0$ then $KB \not\models \alpha$. For the other direction assume that for all $u \in \Gamma$, $\alpha(u) = 1$. We will show that if $y \in$KB, then $\alpha(y) = 1$. From Corollary 2.4, since B is a basis for α and for all $u \in \Gamma$, $\alpha(u) = 1$, we have that

$$\forall u \in \Gamma, \ \forall b \in B, \ \exists v_{u,b} \in min_b(\alpha) \text{ such that } u \geq_b v_{u,b}. \tag{2}$$

Consider now a model $y \in$KB. Again, Corollary 2.4 implies that

$$\forall b \in B, \ \exists z \in min_b(KB) \text{ such that } y \geq_b z. \tag{3}$$

By the assumption, since $min_b(KB) \subseteq \Gamma$, all the elements z identified in Equation 3 satisfy α and therefore, as in Equation 2 we have that

$$\forall z \in min_b(KB), \ \exists v_{z,b} \in min_b(\alpha) \text{ such that } z \geq_b v_{z,b}. \tag{4}$$

Substituting Equation 4 into Equation 3 gives the required condition on $y \in$KB:

$$\forall b \in B, \ \exists v_{(z),b} \in min_b(\alpha) \text{ such that } y \geq_b v_{(z),b}$$

which implies, by Corollary 2.4, that $\alpha(y) = 1$. ∎

The above theorem requires that KB and α can be described by the same basis B. This requirement is somewhat relaxed in the following theorem.

Theorem 3.2 *Let KB be a propositional theory with basis B and let α be a query with basis B'. Then $\Gamma_{KB}^{B \cup B'}$ is a model-based representation for the inference problem $KB \models \alpha$. That is, $KB \models \alpha$ if and only if for every $u \in \Gamma_{KB}^{B \cup B'}$, $\alpha(u) = 1$.*

Proof: It is clear, from Eq. 1 and Claim 2.1 part (4) that $B \cup B'$ is a basis both for KB and α. Therefore, Theorem 3.1 implies the result. ∎

Example: (continued) The set Γ relative to $B = \{0000, 1101\}$ is:
$\Gamma = \{1000, 0100, 0011, 1100, 1111, 1001, 0101\}$. Note that it includes only 7 out of the 12 satisfying assignments of f. Since model-based deduction does not make mistakes on queries that are implied by f we concentrate in our examples on queries that are not implied by f.

To exemplify Theorem 3.1 consider the query $\alpha_1 = \overline{x_2} \wedge \overline{x_3} \rightarrow x_4$. This is equivalent to $x_2 \vee x_3 \vee x_4$ which is falsified by 0000 so B is a basis for α. Reasoning with Γ will find the counterexample 1000 and will therefore conclude $f \not\models \alpha_1$.

The query $\alpha_2 = x_1 \wedge x_3 \rightarrow x_2$ is equivalent to $\overline{x_1} \vee x_2 \vee \overline{x_3}$ which is not falsified by the basis B. Therefore B is not a basis for α_2 and model-based deduction might be wrong. Indeed reasoning with Γ will not find a counterexample and will conclude $f \models \alpha_2$ (it is wrong since the assignments $1010, 1011$ satisfy f but not α_2).

Next, to exemplify Theorem 3.2 consider adding a basis element for α_2. This could be either 1010, or 1011. Choosing 1010, the set of additional minimal elements in Γ is $\{1010\}$, and reasoning with Γ would be correct on α_2.

3.2 Exact Deduction with Approximate Theories

We have shown in the discussion above how to perform deduction with the set of characteristic models Γ_{KB}^{B}, were B is a basis for the knowledge base KB. In this section we consider the natural generalization to the case in which the set of characteristic models of KB is constructed with respect to a basis B that is *not* a basis for the knowledge base KB.

We show that even in this case we can perform exact deduction. As we show, reasoning with characteristic models of KB with respect to a basis B is equivalent to reasoning with the *least upper bound* (LUB) [SK91] of KB in the class of functions with basis B. The significance of this, as proved in Theorem 3.4, is that for queries with basis B this yields exact deduction.

A theory of knowledge compilation using Horn approximation was developed by Selman and Kautz in a series of papers [SK91, KS91, KS92]. Their goal is to speed up inference by replacing the original theory by two Horn approximations of it, one that implies the original theory (a lower bound) and one the is implied by it (an upper bound). While reasoning with the approximations instead of the original theory does not guarantee exact reasoning, it sometimes provides a quick answer to the inference problems. Of course, computing the approximations is a hard computational problem, and this is why it is suggested as a *compilation process*. The computational problems of computing Horn approximations and reasoning with them are studied also in [Cad93, GS92, Rot93].

To facilitate the presentation we first define the notion of an approximation of KB. We then show that representing KB with a set of characteristic models with respect to a basis B yields a function which is the LUB of KB. We then proceed to show the implication to reasoning, and in particular exact deduction with common queries. In particular, since Horn theories have small basis (see Claim 4.1) we can construct Horn LUB and reason with it, generalizing the concept defined and discussed in [SK91, KS91, KS92].

Definition 3.2 (Least Upper-bound) *Let $\mathcal{F}, \mathcal{G}$ be families of propositional languages. Given $f \in \mathcal{F}$ we say that $f_{lub} \in \mathcal{G}$ is a $\mathcal{G}$-least upper bound of f if and only if $f \subseteq f_{lub}$ and there is no $f' \in \mathcal{G}$ such that $f \subset f' \subset f_{lub}$.*

These bounds are called $\mathcal{G}$-*approximations* of the original theory f. The next theorem characterizes the $\mathcal{G}$-LUB of a function and shows that it is unique.

Theorem 3.3 *Let f be any propositional theory and $\mathcal{G}$ a class of all propositional theories with basis B. Then*

$$f_{lub} = \bigwedge_{b \in B} \mathcal{M}_b(f).$$

Proof: Define $g = \bigwedge_{b \in B} \mathcal{M}_b(f)$. We need to prove that (1) $f \subseteq g$, (2) $g \in \mathcal{G}$ and (3) there is no $f' \in \mathcal{G}$ such that $f \subset f' \subset g$. (1) is immediate from Claim 2.1 part (4). To prove (2) we need to show that B is a basis for g. Indeed,

$$
\begin{aligned}
\bigwedge_{b \in B} \mathcal{M}_b(g) &= \bigwedge_{b \in B} \mathcal{M}_b(\bigwedge_{b \in B} \mathcal{M}_b(f)) \subseteq (\bigwedge_{b \in B} \mathcal{M}_b(f)) \bigwedge (\bigwedge_{b_i, b_j \in B, b_i \neq b_j} \mathcal{M}_{b_i} \mathcal{M}_{b_j}(f)) \\
&= g \bigwedge (\bigwedge_{b_i, b_j \in B, b_i \neq b_j} \mathcal{M}_{b_i} \mathcal{M}_{b_j}(f)) \\
&\subseteq g.
\end{aligned}
$$

Using Claim 2.1 part (4) again we get $g \subseteq \bigwedge \mathcal{M}_b(g)$ and therefore $\bigwedge_{b \in B} \mathcal{M}_b(g) = g$, that is $g \in \mathcal{G}$. Finally, to prove (3) assume that there exists $f' \in \mathcal{G}$ such that $f \subseteq f'$. Then,

$$g = \bigwedge_{b \in B} \mathcal{M}_b(f) \subseteq \bigwedge_{b \in B} \mathcal{M}_b(f') = f',$$

where the last equality results from the fact that $f' \in \mathcal{G}$. Therefore, $g = f_{lub}$. ∎

The following theorem can be seen as a generalization of Theorem 3.1, in which we do not require that the basis B is the basis of KB.

Theorem 3.4 *Let $KB \in \mathcal{F}$, $\alpha \in \mathcal{G}$ and let B be a basis for $\mathcal{G}$. Then $KB \models \alpha$ if and only if for every $u \in \Gamma^B_{KB}$, $\alpha(u) = 1$.*

Proof: We have shown in Theorem 3.3 that

$$KB_{lub} = \bigwedge_{b \in B} \mathcal{M}_b(KB) = \bigwedge_{b \in B} \bigvee_{z \in \min_b(KB)} \mathcal{M}_b(z).$$

By Theorem 3.1, since $\alpha(u) = 1$ for every $u \in \Gamma^B_{KB}$, we have that $KB_{lub} \models \alpha$ and therefore KB $\models \alpha$. On the other hand, since $\Gamma^B_{KB} \subseteq$ KB, if for some $u \in \Gamma^B_{KB}$, $\alpha(u) = 0$, KB $\not\models \alpha$. ∎

A result similar to the corollary that follows, for the case in which $\mathcal{G}$ is the class of Horn theories, is discussed in [KS91, Cad93].

Corollary 3.5 *Model-based Reasoning with the least upper bound (with respect to the language $\mathcal{G}$) of a theory KB is correct for all queries in $\mathcal{G}$.*

Example: (continued) The Horn basis for our example is: $B_H = \{1111, 1110, 1101, 1011, 0111\}$ (see Claim 4.1). The minimal elements with respect to 1101 were given before. Each of the models $1111, 0111, 1011, 1110$ satisfies f and therefore for each of these, $\min_b(f) = b$ and together we get that $\Gamma_f^{B_H} = \{1111, 0111, 1011, 1100, 1001, 0101, 1110\}$.

For the query $\alpha_3 = x_1 \wedge x_3 \rightarrow (x_2 \vee x_4)$, which is not Horn, reasoning with $\Gamma_f^{B_H}$ will be wrong. For the Horn query $\alpha_2 = x_1 \wedge x_3 \rightarrow x_2$, reasoning with $\Gamma_f^{B_H}$ will find the counterexample 1011 and therefore be correct.

4 Horn Theories

We consider in detail the case of Horn formulas and show that in this case our notion of *characteristic models* coincides with the notion introduced in [KKS93]. We further discuss the issue of answering all CNF queries. In [KKS93] the deduction theorem was extended to answer any such query. This extension relies on a special property of Horn formulas and does not hold in the general case. We give an example that explains this phenomenon. We start by showing that Horn formulas have a small basis.

Claim 4.1 *The set $B_H = \{u \in \{0,1\}^n \mid weight(u) \geq n-1\}$ is a basis for any Horn CNF function.*

Proof: Let KB be any Horn function. Denote by $b^{(0)}$ the basis element with weight n and by $b^{(i)}$ the basis element with the ith bit set to zero and all the others set to one. By Claim 2.5 it is enough to show that if C is a clause in the CNF representation of KB then it is falsified by one of the basis elements in B. Indeed, if C is a clause in which all the literals are negative, then it is falsified by $b^{(0)}$. If x_k is the only variable that appears un-negated in C then C is falsified by $b^{(k)}$. ∎

4.1 Characteristic Models

In order to relate to the results from [KKS93] we need a few definitions presented there.

For $u, v \in \{0,1\}^n$, we define the *intersection* of u and v to be the assignment $z \in \{0,1\}^n$ such that $z_i = 1$ if and only if $u_i = 1$ and $v_i = 1$ (i.e., the bitwise logical-and of u and v.).

The *closure* of $S \subseteq \{0,1\}^n$, denoted $closure(S)$, is defined as the smallest set containing S that is closed under intersection.

Let KB be a Horn theory. The set of the *Horn characteristic models* of KB, denoted here $char_H(\text{KB})$ is defined as the set of models of KB that are not the intersection of other models of KB. Formally,

$$char_H(KB) = \{u \in KB \mid u \notin closure(KB \setminus \{u\})\}. \tag{5}$$

The following claim should be attributed to Mckinsey [McK43] (note that the definitions there are different; an adaption of this proof to the propositional terminology can be found in [KR94c]). A different proof of this property appears in [DP92].

Claim 4.2 ([McK43]) *A theory is Horn if and only if its set of models is closed under intersection.*

Based on this characterization of Horn theories, it is clear that if KB is a Horn theory and $M \subseteq$ KB any subset of models, then $closure(M) \subseteq closure(\text{KB}) = \text{KB}$. In [KKS93] it is shown that if we take $M = char_H(\text{KB})$, then we get

$$closure(char_H(KB)) = closure(KB) = KB.$$

In particular, Equation 5 implies that $char_H(\text{KB})$ is the smallest subset of KB with that property. Based on this it is then shown that model-based deduction using $char_H(\text{KB})$ yields correct deduction. In the following we show that with respect to the basis B_H from Claim 4.1, and for any Horn theory KB, $char_H(\text{KB}) = \Gamma_{KB}^{B_H}$. Therefore $char_H(\text{KB})$ is an instance of the theory developed in Section 3, and we can reason with it according to Theorem 3.1.

Theorem 4.3 *Let KB be a Horn theory and* $B_H = \{u \in \{0,1\}^n \mid weight(u) \geq n - 1\}$. *Then,* $char_H(KB) = \Gamma_{KB}^{B_H}$.

Proof: Denote $\Gamma = \Gamma_{KB}^{B_H}$. In order to show that $char_H(\text{KB}) \subseteq \Gamma$, it is sufficient to prove that KB $= closure(\Gamma)$. This is true since $char_H(\text{KB})$ is the smallest subset of KB with that property.

Consider $x \in KB$, Corollary 2.4 implies that for all $b^{(i)} \in B$, there exists $u^{(i)} \in \min_{b^{(i)}}(f)$ such that $x \geq_{b^{(i)}} u^{(i)}$. We claim that

$$x = \wedge_{\{k : x_k = 0\}} u^{(k)} \in closure(\Gamma).$$

To see that, consider first the zero bits of x. Let $x_j = 0$, this implies that $u^{(j)}$ is in the intersection and that it satisfies $x \geq_{b^{(j)}} u^{(j)}$. Since $x_j = 0$ and $b_j^{(j)} = 0$ the fact $x_j \geq_{b_j^{(j)}} u_j^{(j)}$ implies $u_j^{(j)} = 0$, and the intersection on this bit is also 0.

Consider now the case $x_j = 1$. Since all the $u^{(k)}$ in the intersection are such that $x_k = 0$, the order relation on the jth bit is always the reversed order, $\leq_1$. That is, all the $u^{(k)}$ in the intersection satisfy $1 = x_j \geq_1 u_j^{(k)}$. This implies that for all the $u^{(k)}$ in the intersection $u_j^{(k)} = 1$ and the intersection on this bit is also 1. This completes the proof of $char_H(\text{KB}) \subseteq \Gamma$.

To prove $\Gamma \subseteq char_H(\text{KB})$, we show that if $x \in \Gamma$, x cannot be represented as $x = y \wedge z$ where $y, z \in KB$ and $x \neq y, z$. Since $char_H(\text{KB})$ is the collection of *all* those elements in KB (from Equation 5), we get the result.

Consider $x \in \min_{b^{(k)}}(\text{KB}) \subseteq \Gamma$, and suppose by way of contradiction that $\exists y, z \in KB$ such that $x = y \wedge z$ and $x \neq y, z$. Fix the order relation $b^{(k)}$ and consider the indices of x. First consider an index $i \neq k$. Since $b_i^{(k)} = 1$ the order relation of the ith index is the reversed one. Now, if $y_i = z_i$ then $x_i = y_i = z_i$, and if $y_i \neq z_i$ then $x_i = 0$. Therefore, in both cases we get that for all $i \neq k$ $x_i \geq_{b_i^{(k)}} y_i$ and $x_i \geq_{b_i^{(k)}} z_i$. For the case $k=0$, the indices $i \neq k$ include all the bits. This implies $x \geq_{b^{(k)}} y$ and $x \geq_{b^{(k)}} z$ and since $x \in \min_{b^{(k)}}(\text{KB})$, this contradicts the assumption that $x \neq y, z$, and therefore proves the claim.

Otherwise, when $k \neq 0$ we consider also the order relation of the kth index, which is the usual order. Again, if $y_k = z_k$ then $x_k = y_k = z_k$ and if $y_i \neq z_i$ then $x_i = 0$. This implies that $x_k \geq_{b_k^{(k)}} y_k$ or $x_k \geq_{b_k^{(k)}} z_k$.

Together with the case $i \neq k$ we get that $x \geq_{b^{(k)}} z$ or $x \geq_{b^{(k)}} y$ (depends on whether $z_k = 0$ or $y_k = 0$). But since $x \in \min_{b^{(k)}}(\text{KB})$, this contradicts the assumption that $x \neq y, z$, and completes the proof. ∎

4.2 General Queries

In [KKS93] it is shown that in case of Horn theories one can answer any CNF query without re-computing the characteristic models. While we have shown that our model-based representation coincides with that of [KKS93] it turns out that the ability to answer any query relies on a special characteristic of Horn theories, and does not generalize to other propositional theories. We next

give a counterexample that exemplifies this. The deduction scheme in [KKS93] when α is a general CNF expression , utilizes the following theorem[3].

Theorem 4.4 *Let KB be a Horn theory and α any disjunction. If $KB \models \alpha$ then there exists a Horn disjunction β such that $KB \models \beta$ and $\beta \models \alpha$.*

Together with the following observations:

1. Every disjunction α can be represented as $\alpha = \beta_1 \vee \ldots \vee \beta_k$, where the β_i are Horn disjunctions.

2. $KB \models \alpha_1 \alpha_2$ if and only if $KB \models \alpha_1$ and $KB \models \alpha_2$.

Observation (2) implies that it is enough to consider queries that are disjunctions. Given α, the deduction scheme in [KKS93] decomposes it into the Horn disjunction β_is and tests deduction against the β_is. By Theorem 4.4, at least one of the β_is is implied by KB. While the observations above are true even for non-Horn theories, Theorem 4.4 depends on KB being Horn, as the following example shows.
Example [Theorem 4.4 does not hold for non-Horn languages]
Let:
$\quad$ KB $= (x_1 \vee x_2 \vee \overline{x_3} \vee \overline{x_4}) \wedge (x_3 \vee x_5 \vee \overline{x_6})$.
$\quad$ $\alpha = x_1 \vee x_2 \vee \overline{x_4} \vee x_5 \vee \overline{x_6}$
The knowledge base KB is not a Horn theory, and it is easy to check that KB $\models \alpha$. However, there is no disjunction β such that KB$\models \beta \models \alpha$.

5 The Size of Model Based Representations

The complexity of model-based reasoning is directly related to the number of models in the representation. It is therefore important to compare this size with the size of other representations of the same function. In the previous section we have shown that our model-based representation is the same as that in [KKS93] when the theory is Horn. In [KKS93] examples are given for large Horn theories with a small set of characteristic models and vice versa, but it was not yet understood when and why it happens. Our results imply that the set of characteristic models of a Horn theory is small if the size of a DNF description for the same theory is small. The other direction is however not true (i.e., there are Horn theories with a small set of characteristic models but an exponential size DNF). We start with a bound on the size of the model-based representation.

Lemma 5.1 *Let B be a basis for the knowledge base KB, and denote by $|DNF(KB)|$ the size of its DNF representation. Then, the size of the model-based representation of a knowledge base KB is*

$$|\Gamma_{KB}^B| \leq \sum_{b \in B} |min_b(KB)| \leq |B| \cdot |DNF(KB)|.$$

Proof: The lemma follows from Corollary 2.7. $\blacksquare$

As the following claim shows, this bound is actually achieved for some functions. For the next claim we would need the following terminology. A term t is an *implicant* of a function f, if $t \models f$. A term t is a *prime implicant* of a function f, if t is an implicant of f and the conjunction of any proper subset of the literals in t is not a prime implicant.

---

[3]This theorem follows from McKinsey's proof of Claim 4.2. In [KKS93] it is derived, in a different way, using a completeness theorem for resolution given in [SCL69] (see also [KR94c] for a discussion).

Claim 5.2 *For any b-monotone function f, $|min_b(f)| = |DNF(f)|$.*

Proof: We first consider monotone functions (i.e. 0^n-monotone). It is well known that for a monotone function there is a unique DNF representation in which each term is a prime implicant. Let f be a monotone function and consider this representation for f. Similar to Claim 2.6 we can map every term in the representation to its corresponding minimal element. Moreover, since the terms are monotone and the order relation is 0^n, each of these minimal elements is indeed a minimal element of f (otherwise one of the terms in the representation is not a prime implicant). So there is a one to one correspondence between prime implicants and minimal assignments of f with respect to $b = 0^n$, and $|min_{0^n}(f)| = |DNF(f)|$. The same arguments hold for any b-monotone function with respect to the order relation b (one can simply rename the variables) and therefore $|min_b(f)| = |DNF(f)|$. ∎

Claim 5.2 explains the two examples in [KKS93]. Both examples are 1^n-monotone Horn functions, one has a small DNF and the other has an exponentially large DNF. We note that exponential size model-based representations are not restricted to happen in b-monotone functions. One can easily construct such functions by using, for example, a conjunction of several functions each b-monotone with a different b (of course the DNF size has to be exponential here too). The following claim shows that DNF size is not always needed. There are theories for which the DNF size is exponential but the size of the model-based representation, and therefore also model-based reasoning is polynomial.

Claim 5.3 *There exist Horn formulas with an exponential size DNF and a set $\Gamma_f^{B_H}$ of linear size.*

Proof: For each n we exhibit a formula f with the required property. The function

$$f = (\overline{x_1} \vee \overline{x_2} \ldots \vee \overline{x_{\sqrt{n}-1}} \vee x_{\sqrt{n}}) \wedge \ldots \wedge (\overline{x_{n-\sqrt{n}+1}} \vee \overline{x_{n-\sqrt{n}+2}} \vee \ldots \vee \overline{x_{n-1}} \vee x_n)$$

is clearly in Horn form.

The size of its DNF representation is $\sqrt{n}^{\sqrt{n}}$. This is easy to observe by renaming the negative literals as its negation. This yields a monotone formula in which each term we get, by multiplying one variable from each clause, is a prime implicant.

The set Γ is of size $< 2n$. Recall that $b^{(i)} \in B_H$ denotes the basis element in which the variable x_i is assigned 0, and $b^{(0)} = 1^n$. First observe that for $i \neq k\sqrt{n}$, $b^{(i)}$ is a satisfying assignment of f and therefore has only one minimal element (that is, itself). For $i = k\sqrt{n}$, $b^{(i)}$ is not a satisfying assignment of f. There is however only one clause, C^i, not satisfied by $b^{(i)}$, the clause which includes the variable x_i. Now, since each variable appears only once in f, each of the variables in C^i we flip yields a satisfying assignment which is minimal. This contributes $\sqrt{n}$ minimal assignments. (Flipping variables not in C^i does not contribute minimal assignments with respect to $b^{(i)}$.) One last note is that each of these $b^{(i)}$ would have 1^n as one of the minimal assignments, so we need to count it only once, and count $(\sqrt{n} - 1)$ for each of the $b^{(i)}$'s.

Altogether there are $(n - \sqrt{n}) + \sqrt{n}(\sqrt{n} - 1) + 1 = 2(n - \sqrt{n}) + 1$ minimal elements. ∎

Considering model-based representations, Claim 5.2 implies that for every basis there is a function which has an exponential number of characteristic models. Nevertheless, one might hope that there is a basis for which least upper bounds will always have small representations in some (maybe other) form that admits fast reasoning. Kautz and Selman [KS92] show that for Horn representations this is not the case. In particular they show that unless NP $\subseteq$ non-uniform P there exists

a function whose Horn LUB does not have a short representation that allows for efficient reasoning. This can be generalized[4], using essentially the same proof, to hold for every fixed basis and in particular k-quasi-Horn, $\log n$ CNF, and monotone functions. We therefore have the following theorem:

Theorem 5.4 *Unless $NP \subseteq$ non-uniform P, for every fixed basis B there exists a function whose LUB does not have a short representation which admits efficient reasoning.*

6 Applications

In Section 3 we developed the general theory for model based deduction. In this section we discuss applications of this theory to specific propositional languages.

Our basic result (Theorem 3.1) assumed that the knowledge base and the query share the same basis. We give such queries a special status.

Definition 6.1 *Let B be a basis for KB. A query α is* relevant *to KB if B is a basis for α.*

Theorem 3.2 suggests a way in which one can overcome the difficulty in the case where the basis B of KB is not a basis for the query α. This can be done by: (1) adding the basis B' of the query to the knowledge base basis, and (2) computing additional characteristic models based on the new basis.

Claim 2.5 suggests a simple way for computing the basis for a given query, as required in (1). The problem of computing additional characteristic models, however, is in general a hard problem that we do not address here. Neither do we consider computing additional models in an on-line process performed for each query. At this point we assume that the knowledge base is given in the form of its set of characteristic models. We note however that Bshouty [Bsh93] gives an algorithm that learns the model-based representation we consider here when given access to a *Membership Oracle* and an *Equivalence Oracle*. In [KR94a] we discuss the issue of "learning to reason" and illustrate the importance of the model-based approach for this problem.

We say that queries are *common* if they are taken from some common propositional language[5] as defined below.

Definition 6.2 *A language $\mathcal{F}$ is* common *if there is a small (polynomial size) fixed basis for all $f \in \mathcal{F}$. The set $\mathcal{L}_C$ is the set of* common *languages.*

Important examples of common languages are: (1) Horn-CNF formulas, (2) reversed Horn-CNF formulas (CNF with clauses containing at most one *negative* literal), (3) k-quasi-Horn formulas (a generalization of Horn theories in which there are at most k positive literals in each clause), (4) k-quasi-reversed-Horn formulas and (5) $\log n$CNF formulas (CNF in which the clauses contain at most $O(\log n)$ literals). Any formula that can be represented as a CNF with clauses from any combination of the above is also in $\mathcal{L}_C$. The fixed bases for these languages are discussed in the following subsection.

---

[4]This issue has been brought to our attention by Henry Kautz and Bart Selman.

[5]Note that a fixed basis uniquely characterizes a family of Boolean functions which can be represented using it. There are of course other ways to characterize classes of functions which do not correspond to any basis (e.g. some subset of DNF).

6.1 Languages with a Small Basis

In Claim 4.1 we have shown that Horn formulas have a short basis. A similar construction yields a basis for reversed Horn, k-quasi-Horn formulas, and k-quasi-reversed-Horn formulas.

Claim 6.1 *There is a polynomial size basis for: reversed Horn formulas, k-quasi-Horn formulas, and k-quasi-reversed-Horn formulas.*

Proof: The analysis is very similar to the one in Claim 4.1. By flipping the polarity of all bits in B_H we can get a basis for reversed Horn. Similarly, using the set $B_{H_k} = \{u \in \{0,1\}^n \mid \text{weight}(u) \geq n - k\}$ we get a basis for k-quasi-Horn, and flipping the polarity of all bits in B_{H_k} we get a basis for k-quasi-reversed-Horn formulas. ∎

We next consider the expressive class of $\log n$ CNF formulas, in which there are up to $O(\log n)$ variables in a clause, and show that it has a polynomial size basis.

An (n,k)-*universal set* is a set of assignments $\{d_1, \ldots d_t\} \subseteq \{0,1\}^n$ such that every subset of k variables assumes all of its 2^k possible assignments in the d_i's. It is known [ABN$^+$92] that for $k = \log n$ one can construct (n,k)-universal sets of polynomial size.

Claim 6.2 ([Bsh93]) *Let B be an (n,k)-universal set. Then B is a basis for any k-CNF KB.*

Proof: By Claim 2.5 it is enough to show that if C is a clause in the k-CNF representation of KB then it is falsified by one of the basis elements in B. Let $C = l_{i_1} \vee \ldots l_{i_k}$ be a clause in the CNF representation of KB, where $l_{i_k} \in \{x_{i_k}, \overline{x_{i_k}}\}$. Let $a \in \{0,1\}^n$ be an assignment. Then the value $C(a)$ is determined only by $a_{i_1}, \ldots, a_{i_k}$ and since B is an (n,k)-universal set, there must be an element $b \in B$ for which $C(b) = 0$. ∎

6.2 Main Applications

In the case of common or relevant queries, reasoning involves the evaluation of a propositional formula on a polynomial number of assignments. This is a very simple and easily parallelizable procedure. Moreover, Theorem 3.4 shows that in order to reason with common queries, we need not use the basis of KB at all, and it is enough to represent KB by the set of characteristic models with respect to the basis of the query language. Lemma 5.1 together with Theorems 3.1,3.2,3.3 and 3.4 imply the following general applications of our theory:

Theorem 6.3 *Any function $f : \{0,1\}^n \rightarrow \{0,1\}$ that has a polynomial size representation in both DNF and CNF form can be described with a polynomial size set of characteristic models.*

Theorem 6.4 *Any function $f : \{0,1\}^n \rightarrow \{0,1\}$ in any language $\mathcal{F} \in \mathcal{L}_C$ which has a polynomial size DNF representation can be described with a polynomial size set of characteristic models.*

Theorem 6.5 *Let KB be a knowledge base (on n variables) that can be described with a polynomial size set Γ of characteristic models. Then, for any relevant or common query, model-based deduction using Γ, is both correct and efficient.*

Theorem 6.6 *Let KB be a knowledge base (on n variables) that can be described with a polynomial size DNF. Then there exists a fixed, polynomial size set of models Γ, such that for any common query, a model-based deduction using Γ, is both correct and efficient.*

The results in this paper are concerned mainly with propositional languages. While many AI formalizations use first order logic as its main tool, some applications do not need the full power of first order logic. It is quite easy to observe that any such formalization which is function free and has a finite number of constants can be mapped into a finite propositional language. Furthermore, a function free universally quantified sentence in Horn form (or any other language with a fixed basis) remains in Horn form[6] in the new propositional domain. These observations imply that our results hold for these restricted first order logic formalization, where the polynomial bounds are relative to the number of variables in the propositional domain.

7   Abduction with Models

We consider in this section the question of performing abduction using a model-based representation. In [KKS93] it is shown that for a Horn theory KB, Abduction can be done in polynomial time using characteristic models, although using formula based representation the problem is NP-Hard [Sel90]. In this section we show that if we add a few base assignments to our basis, the algorithm presented there works in the general case too.

Abduction is the task of finding a minimal explanation to some observation. Formally [RDK87], the reasoner is given a knowledge base KB (the *background theory*), a set of propositional letters[7] A (the *assumption set*), and a query letter q. An *explanation* of q is a minimal subset $E \subseteq A$ such that

1. $KB \wedge (\wedge_{x \in E} x) \models q$ and

2. $KB \wedge (\wedge_{x \in E} x) \neq \emptyset$.

Thus, abduction involves tests for entailment and consistency, but also a search for an explanation that passes both tests. We now show how one can use the algorithm from [KKS93] for any propositional theory KB.

Theorem 7.1 *Let KB be a background propositional theory with a basis B, let A be an assumption set and q be a query. Let $B_H = \{x \in \{0,1\}^n | weight(x) \geq n - 1\}$. Then, using the set of characteristic models $\Gamma = \Gamma_{KB}^{B \cup B_H}$ one can find an abductive explanation of q in time polynomial in $|\Gamma|$ and $|A|$.*

Proof: We use the algorithm *Explain* suggested in [KKS93] for the case of a Horn knowledge base. For a Horn theory KB the algorithm uses the set $char_H(KB) = \Gamma_{KB}^{B_H}$ defined in Section 4. We show that adding the Horn basis B_H and the additional characteristic models to a general model-based representation is sufficient for it to work in the general case.

The abduction algorithm *Explain* starts by enumerating all the characteristic models. When it finds a model in which the query holds, (i.e., $q = 1$) it sets E to be the conjunction of all the variables in A that are set to 1 in that model. (This is the strongest set of assumptions that are valid in this model.)

---

[6]We note that the CNF formula size grows exponentially with the number of quantifiers. This size however does not effect our results as we are interested in the size of the basis and the size of the DNF formula.

[7]The task of abduction is normally defined with arbitrary literals for explanations. For Horn theories explanations turn out to be composed of positive literals (this can be concluded from Corollary 4 in [RDK87]). Here we restrict ourselves to explanations composed of positive literals (by allowing only positive literals in the assumption set) when using general theories. One may therefore talk about "positive explanations" instead of explanations. We nevertheless continue with the term explanation.

The algorithm then performs the entailment test ((1) in the definition above) to check whether E is a valid explanation. This test is equivalent to testing the deduction KB $\models (q \vee (\vee_{x \in E} \bar{x}))$, that is a deductive inference with a Horn clause as the query. According to Theorem 3.2 this can be done efficiently with $\Gamma_{KB}^{B \cup B_H}$.

If the test succeeds, the assumption set is minimized in a greedy fashion by eliminating variables from E and using the entailment test again. It is clear that if the algorithm outputs a *minimal* assumption set E (in the sense that no subset of E is a valid explanation, not necessarily of smallest cardinality) then it is correct. Minimality is guaranteed by the greedy algorithm, the requirement (1) by the deductive test, and the requirement (2) by the existence of the original model that produced the explanation.

It remains to show that if an explanation exists, the algorithm will find one. To prove this, it is sufficient to show that in such a case there exists a model $x \in \Gamma$ in which both the bit q and a superset of E are set to 1.

The existence of x is a direct consequence of including the base assignment $b = 1^n$ in the basis. This is true as relative to b we have $1 <_b 0$ for each bit. Therefore if there exists a model y which satisfies some explanation E, either it is a minimal assignment relative to b, or $\exists x \leq_b y$ and x is in Γ. In the first case $x = y$ is the required assignment, in the second case we observe that $y_i = 1$ implies $x_i = 1$ which is what we need. ∎

It is quite easy to see that the above theorem can be generalized in several ways. First, we can allow the assumptions set A to have up to k negative literals for some constant k and use the basis for k-quasi-Horn instead of B_H. Second, we can allow the query q to have more than just one literal. In particular it is easy to verify that the same proof works if q is a conjunction of positive literals, or if q is any Horn disjunction.

8 Conclusions and Further Work

This paper develops a formal theory of model-based reasoning. We show that a simple model-based approach can support exact deduction and abduction even when an exponentially small portion of the model space is tested. Our approach builds on (1) the characterization of a set of characteristic models of the knowledge base that together capture all the information needed to reason with (2) a restricted set of queries. We prove that for a fairly large class of propositional theories, including theories that do not allow efficient formula-based reasoning, the model-based representation is compact and provides efficient reasoning.

The restricted set of queries, which we call *relevant queries* and *common queries*, can come from a wide class of propositional languages, (and include, for example, quasi-Horn theories and $\log n$ CNF), or from the same propositional language that represents the "world". We argue that this is a reasonable approach to take in the effort to give a computational theory that accounts for both the speed and flexibility of common-sense reasoning.

The usefulness of the approach developed here is exemplified by the fact that it explains, generalizes and unifies many previous investigations, and in particular the fundamental works on reasoning with Horn models [KKS93] and Horn approximations [SK91, KS91, KS92].

We are currently studying several extensions to this theory. First order logic formalizations of AI problems may have a very large (or even infinite) number of attributes. This may rule out the basic approach since even storing the models may constitute a problem. This consideration, as well as other reasons, led us to consider reasoning with partial assignments. In [KR94b] we report some results in this line.

We are also studying the relations of the model-based approach to planning. Since the original formalizations of planning were in the form of deduction queries, one can reduce a planning problem to several deduction queries. The question here is whether this reduction can be done in a way that the queries can be answered efficiently using a model-based approach.

This work is part of a more general framework which views *learning* as an integral part of the reasoning process. We believe that some of the difficulties in constructing an adequate computational theory to reasoning result from the fact that these two tasks are viewed as separate. In [KR94a] we discuss the issue of "learning to reason" and illustrate the importance of the model-based approach for this problem.

Acknowledgments

References

[ABN+92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

[BL84] R. Brachman and H. Levesque. The tractability of subsumption in framebased description languages. In *Proceedings of the National Conference on Artificial Intelligence*, pages 34–37, 1984.

[Bsh93] N. H. Bshouty. Exact learning via the monotone theory. In *Proceedings of the IEEE Symp. on Foundation of Computer Science*, pages 302–311, Palo Alto, CA., 1993.

[Cad93] M. Cadoli. Semantical and computational aspects of Horn approximations. In *Proceedings of the International Joint Conference of Artificial Intelligence*, pages 39–44, August 1993.

[DP91] J. Doyle and R. Patil. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. *Artificial Intelligence*, 48:261–297, 1991.

[DP92] R. Dechter and J. Pearl. Structure identification in relational data. *Artificial Intelligence*, 58:237–270, 1992.

[EBBK89] D. Etherington, A. Borgida, R. Brachman, and H. Kautz. Vivid knowledge and tractable reasoning: Preliminary report. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 1989.

[GS92] R. Greiner and D. Schuurmans. Learning useful Horn approximations. In *Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning*, pages 383–392, 1992.

[JL83] P. N. Johnson-Laird. *Mental Models*. Harvard Press, 1983.

[JLB91] P. N. Johnson-Laird and R. M. J. Byrne. *Deduction*. Lawrence Erlbaum Associates, 1991.

[KKS93]   H. Kautz, M. Kearns, and B. Selman. Reasoning with characteristic models. In *Proceedings of the National Conference on Artificial Intelligence*, pages 34–39, 1993.

[Kos83]   S. M. Kosslyn. *Image and Mind.* Harvard Press, 1983.

[KR94a]   R. Khardon and D. Roth. Learning to reason. In *Proceedings of the National Conference on Artificial Intelligence*, pages xx–yy, 1994. Full version: Technical Report TR-2-94, Aiken Computation Lab., Harvard University, January 1994.

[KR94b]   R. Khardon and D. Roth. Learning to reason with a restricted world view. 1994. in preparation.

[KR94c]   R. Khardon and D. Roth. A note on Horn theories. 1994. Manuscript.

[KS91]   H. Kautz and B. Selman. A general framework for knowledge compilation. In *Proceedings of the International Workshop on Processing Declarative Knowledge, Kaiserlautern, Germany*, July 1991.

[KS92]   H. Kautz and B. Selman. Forming concepts for fast inference. In *Proceedings of the National Conference on Artificial Intelligence*, pages 786–793, 1992.

[Lev86]   H. Levesque. Making believers out of computers. *Artificial Intelligence*, 30:81–108, 1986.

[Lev92]   H. Levesque. Is reasoning too hard ? In *Proceeding of the 3rd NEC research Symposium.* 1992.

[McC58]   J. McCarthy. Programs with common sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, volume 1, pages 77–84. National Physical Laboratory, 1958. Reprinted in Minsky's (ed.) *Semantic Information Processing*, MIT Press(1968), 403–409. Also in R. Brachman and H. Levesque, Readings in Knowledge Representation, 1985.

[McK43]   J. C. C McKinsey. The decision problem for some classes of sentences without quantifier. *Journal of Symbolic Logic*, 8(3):61–76, 1943.

[Pap91]   C. H. Papadimitriou. On selecting a satisfying truth assignment. In *Proc. 32nd Ann. IEEE Symp. on Foundations of Computer Science*, pages 163–169, 1991.

[RDK87]   R. Reiter and J. De Kleer. Foundations of assumption-based truth maintenance systems. In *Proceedings of the National Conference on Artificial Intelligence*, pages 183–188, 1987.

[Rot93]   D. Roth. On the hardness of approximate reasoning. In *Proceedings of the International Joint Conference of Artificial Intelligence*, pages 613–618, August 1993.

[SCL69]   J.R. Slage, C.L. Chang, and R.C.T. Lee. Completeness theorems for semantic resolution in consequence finding. In *Proceedings of the International Joint Conference of Artificial Intelligence*, pages 281–285, August 1969.

[Sel90]   B. Selman. *Tractable Default Reasoning.* PhD thesis, Department of Computer Science, University of Toronto, 1990.

[Sha93]   L. Shastri. A computational model of tractable reasoning - taking inspiration from cognition. In *Proceedings of the International Joint Conference of Artificial Intelligence*, pages 202–207, August 1993.

[SK90]   B. Selman and H. Kautz. Model-preference default theories. *Artificial Intelligence*, 45:287–322, 1990.

[SK91]   B. Selman and H. Kautz. Knowledge compilation using Horn approximations. In *Proceedings of the National Conference on Artificial Intelligence*, pages 904–909, 1991.