Decidable Fragments of First-Order Language Under Stable Model Semantics and Circumscription

Heng Zhang[a] and **Mingsheng Ying**[a,b]

[a]State Key Laboratory of Intelligent Technology and Systems,
Department of Computer Science and Technology,
Tsinghua University, Beijing, China
[b]Center of Quantum Computation and Intelligent Systems,
Faculty of Engineering and Information Technology,
University of Technology, Sydney, NSW, Australia

Abstract

The stable model semantics was recently generalized by Ferraris, Lee and Lifschitz to the full first-order language with a syntax translation approach that is very similar to McCarthy's circumscription. In this paper, we investigate the decidability and undecidability of various fragments of first-order language under both semantics of stable models and circumscription. Some maximally decidable classes and undecidable classes are identified. The results obtained in the paper show that the boundaries between decidability and undecidability for these two semantics are very different in spite of the similarity of definition. Moreover, for all fragments considered in the paper, decidability under the semantics of circumscription coincides with that in classical first-order logic. This seems rather counterintuitive due to the second-order definition of circumscription and the high undecidability of first-order circumscription.

Introduction

The stable model semantics was defined by (Gelfond and Lifschitz 1988) to provide a declarative semantics for logic programming with negation as failure. Answer set programming was then proposed based on the stable model semantics, and it has emerged as a flourishing paradigm of declarative programming. Many answer set solvers have been designed, and various combinatorial problems from different areas have been solved by the techniques of answer set programming (Gelfond 2008). On the other hand, it was suggested by (Pearce 2008) that the stable model semantics can be applied in a new method of non-monotonic reasoning, namely stable reasoning.

The earlier versions of the stable model semantics apply only for logic programs without functions or propositional languages, and most of them are based on the grounding techniques. To increase the expressive power, the stable model semantics was recently generalized to the full first-order language (Ferraris, Lee, and Lifschitz 2007; 2010)[1]. Their approach is to find a suitable translation of first-order formulas into a second-order language. This is similar to the semantics of circumscription (McCarthy 1980; 1986) where minimal models are employed.

Inference and satisfiability testing are two of the most important computational tasks in each logical system. Under the stable model semantics, there is an effective procedure to reduce the inference problem to the satisfiability problem (see Fact 1 and the remark on it in the next section). If the formula to be inferred is just the negation of a negative formula, there is a similar reduction for the circumscriptive semantics (Lifschitz 1994). So, in this paper we will focus our attention on the satisfiability problem. It is unrealistic to consider this problem in the full first-order language. For instance, it was proved in (Schlipf 1987) that the problem of deciding whether an arbitrary first-order sentence has a countable infinite minimal model is Σ_2^1-complete over the integers, and even the existence of infinite minimal models is dependent upon the continuum hypothesis (Schlipf 1986). Moreover, results obtained in this paper indicate that the satisfiability problem for the stable model semantics may be more difficult than that for circumscription. This forces us to consider fragments of first-order language.

The purpose of this paper is to examine decidability of various fragments of first-order language under the stable model semantics and circumscription. As in the case of classical semantics, a fragment of first-order language is said to be decidable under the semantics of stable models (or circumscription) if the problem whether there is a stable model (or a minimal model respectively) for a sentence in it is decidable. For circumscription, there have been many attempts at finding fragments of first-order language in which the circumscription for every formula that is a second-order formula is equivalent to a first-order formula (Lifschitz 1985; Rabinov 1989; Kolaitis and Papadimitriou 1990). Then semi-decidability of these fragments under circumscription immediately follows from semi-decidability of first-order logic. Our aim is quite different: we will give direct characterizations of some decidable fragments and some undecidable fragments. This can help us to clarify the boundary of decidability and undecidability under circumscriptive semantics. The same problem is also considered for the stable model semantics. The fragments considered in the present paper are standard prefix-vocabulary classes (or simply standard classes) which have be thoroughly studied in the classical first-order logic (Börger, Grädel, and Gurevich 1997).

[1]A different definition was given by (Lin and Zhou 2007).

The main contributions of this paper are as follows:

- We obtain six standard classes which are maximally decidable under the semantics of circumscription. We show that the Rabin class is maximally decidable under the semantics of stable models. It is also proved that the existential class with functions is semi-decidable and the existential class without functions is decidable under the stable model semantics.

- Three standard classes are found to be maximally decidable for the circumscriptive semantics but undecidable under the stable model semantics.

- Some standard classes undecidable under the semantics of circumscription or the semantics of stable models are identified.

The paper is organized as follows. In the next section we review the semantics of circumscription and stable models and fix notations. The prefix-vocabulary class and standard class are also defined. Decidable and undecidable standard classes under both semantics are examined in the third and fourth sections, respectively. A set of maximally decidable standard classes are presented in the fifth section.

Preliminaries

In this paper, *vocabularies* are finite sets of predicate constants and function constants. Every constant (predicate or function) is equipped with a natural number, its *arity*. The logical symbols are defined as usual, including a countable set of predicate variables and a countable set of individual variables. Terms, formulas, and sentences of a vocabulary σ (or shortly, σ-terms, σ-formulas, and σ-sentences) are built from σ and variables in a standard way. The only thing that may be special is that we treat $\neg \varphi$ as a shorthand of $\varphi \rightarrow \bot$.

Each *structure* $\mathfrak{A}$ of vocabulary σ (or shortly, σ-*structure* $\mathfrak{A}$) consists of a nonempty set A, the *universe* of $\mathfrak{A}$, of an n-ary relation P^A on A for each n-ary predicate constant P in σ, of an n-ary function f^A on A for each n-ary function constant f in σ. For the sake of simplicity, we will omit superscripts A in the relations and functions if no confusion occurs. Let $\mathfrak{A}$ be a σ-structure. An *assignment* in $\mathfrak{A}$ is a function α that maps each individual variable to an element in A and that maps each n-ary predicate variable to an n-ary relation on A. Given a σ-formula φ and an assignment α in $\mathfrak{A}$, we write $\mathfrak{A} \models \varphi[\alpha]$ if α *satisfies* φ in $\mathfrak{A}$ in the standard way. Specially, if φ is a sentence, $[\alpha]$ may be dropped, and $\mathfrak{A}$ is said to be a *model* of φ, or in other words, $\mathfrak{A}$ *satisfies* φ.

Circumscription

Let φ be a sentence of first-order logic and let $\bar{P}$ be a tuple $(P_1, \ldots, P_k)$ of predicate constants. For each n-ary P_i we introduce a new predicate variable X_{P_i} of arity n, and let $\bar{X}$ be short for $(X_{P_1}, \ldots, X_{P_k})$. Moreover, we write $\bar{P} = \bar{X}$ for the conjunction of formulas $\forall \bar{x}(P_i(\bar{x}) \leftrightarrow X_{P_i}(\bar{x}))$ such that $i \leq k$, and write $\bar{P} \leq \bar{X}$ for the conjunction of formulas $\forall \bar{x}(P_i(\bar{x}) \rightarrow X_{P_i}(\bar{x}))$ such that $i \leq k$. The *circumscription* $\mathrm{CIRC}(\varphi; \bar{P})$ for φ is defined to be the second-order sentence $\varphi \wedge \forall \bar{X}(\bar{X} < \bar{P} \rightarrow \neg \varphi(\bar{X}))$, where $\varphi(\bar{X})$ is the formula obtained from φ by substituting the variables

$\bar{X}$ for the constants $\bar{P}$, and $\bar{X} < \bar{P}$ short for the formula $(\bar{X} \leq \bar{P}) \wedge \neg(\bar{X} = \bar{P})$. If $\bar{P}$ is just the tuple of all the predicate constants occurring in φ, we will simply write $\mathrm{CIRC}(\varphi)$. Finally, a structure $\mathfrak{A}$ is called a *minimal model* of φ if it is a model of $\mathrm{CIRC}(\varphi)$.

Stable Models

In a similar way, the stable model semantics is defined by a syntax translation SM. Given a sentence φ, let $\mathrm{SM}(\varphi)$ stand for the second-order sentence $\varphi \wedge \forall \bar{X}(\bar{X} < \bar{P} \rightarrow \neg \varphi^*(\bar{X}))$, where $\bar{P}$ lists all the predicate constants occurring in φ, notations $\bar{X}$ and $\bar{X} < \bar{P}$ are the same as given above, and $\varphi^*(\bar{X})$ is defined recursively as follows:

- $P(\bar{t})^* = X_P(\bar{t})$ if P is a predicate constant.
- $\psi^* = \psi$ if ψ is $\bot$ or an equality.
- $(\psi_1 \circ \psi_2)^* = (\psi_1^* \circ \psi_2^*)$ if $\circ \in \{\wedge, \vee\}$.
- $(\psi_1 \rightarrow \psi_2)^* = (\psi_1 \rightarrow \psi_2) \wedge (\psi_1^* \rightarrow \psi_2^*)$.
- $(Qx\psi)^* = Qx\psi^*$ if $Q \in \{\forall, \exists\}$.

A structure $\mathfrak{A}$ is called a *stable model* of φ if it is a model of $\mathrm{SM}(\varphi)$. A sentence ψ is said to be a *consequence* of the sentence φ under the stable model semantics (written $\varphi \vDash_{\mathrm{SM}} \psi$) if every stable model $\mathfrak{A}$ of φ satisfies ψ, i.e. $\mathrm{SM}(\varphi) \vDash \psi$. Herein, $\vDash$ denotes the classical inference relation. Notice that $\mathrm{SM}(\varphi) \vDash \psi$ iff $\mathrm{SM}(\varphi) \vDash \neg\neg\psi$. By Proposition 2 of (Ferraris, Lee, and Lifschitz 2010), we have following fact.

Fact 1. $\varphi \vDash_{\mathrm{SM}} \psi$ *iff* $\varphi \wedge \neg\psi$ *has no stable model.*

Interestingly, this fact provide us an effective procedure to reduce the validity of an inference to the satisfiability of a formula under the stable models semantics.

Standard Classes

A *stand prefix-vocabulary class* (simply a *standard class*) in this paper is slightly different from that in (Börger, Grädel, and Gurevich 1997). Instead of omitting nullary function constants and nullary predicate constants, we will consider these symbols in our definition. Strings over the alphabet $\{\forall, \exists\}$ are called *prefixes*. A prefix set is said to be *standard* if either it is the set (abbreviated "*all*") of all prefixes, or else it can be expressed by a regular expression in $\{(\forall|\epsilon), (\exists|\epsilon), \forall^*, \exists^*\}^*$, where ϵ stands for the empty prefix. To simplify notations, we write $\forall$ for $(\forall|\epsilon)$, and $\exists$ for $(\exists|\epsilon)$. Let Π and Π' be two standard prefix sets, we say that Π dominates Π', symbolically $\Pi \geq \Pi'$, if S' is a subset of S, where S and S' are the prefix sets given by Π and Π' respectively.

An *arity sequence* is a function that maps each natural number to either a natural number or the first infinite ordinal ω. For convenience, we write an arity sequence p as a sequence $(p_0, p_1, \ldots)$ where $p_i = p(i)$. A tail of zeros may be omitted. An empty sequence will be denoted (0) rather than $()$. And the sequence $(\omega, \omega, \ldots)$ will be denoted "*all*". Let p and q be two arity sequences, we say p dominates q, symbolically $p \geq q$, if $\sum_{i \leq j} p(j) \geq \sum_{i \leq j} q(j)$ for all i.

Let Π be a standard prefix set, and let p and f be two arity sequences. The *standard class* $[\Pi, p, f]_= $ (or $[\Pi, p, f]$) is defined to be the class of prenex formulas φ of first-order

logic with (or without) equality and such that (i) the quantifier prefix of φ belongs to Π; (ii) for all $n \geq 0$, φ has at most $p(n)$ predicate symbols of arity n and at most $f(n)$ function constants. Given two standard classes K and K′, we write K $\geq$ K′ if (i) K allows equality iff so does K′, the prefix set of K dominates that of K, and both arity sequences of K dominate the corresponding arity sequence of K′ respectively; or (ii) K coincides with K′ on the prefix set and both arity sequences, and K′ allows equality whenever so does K. Furthermore, we write K > K′ if K $\geq$ K′ and K $\neq$ K′ hold.

Let K be a standard class. We say that K has the *finite model property* if every satisfiable sentence $\varphi \in$ K has a finite model, and say K is *decidable* under the semantics of stable models (or circumscription) if there is an algorithm to decide whether or not there is a stable model (or minimal model respectively) for a given sentence in K.

Decidability

In this section, we investigate the decidable fragments under both the stable model and the circumscriptive semantics.

Given a fragment which has the finite model property, we have that every sentence in this fragment has a minimal model iff it is satisfiable. So, the finite model property of every first-order fragment immediately implies its decidability under the circumscriptive semantics. The following standard classes are decidable due to their finite model property in classical logic (Börger, Grädel, and Gurevich 1997).

Fact 2. *The following standard classes are decidable under the circumscriptive semantics:*

(1) $[\exists^*\forall^*, all, (\omega)]_=$ (4) $[\exists^*\forall\exists^*, all, all]$
(2) $[\exists^*\forall^2\exists^*, all, (\omega)]$ (5) $[\exists^*, all, all]_=$
(3) $[all, (\omega, \omega), (\omega, \omega)]$

In classical first-order logic, there are two standard classes that have not the finite model property: $[all, (\omega, \omega), (\omega, 1)]$ and $[\exists^*\forall\exists^*, all, (\omega, 1)]_=$. It was proved by Routenberg and Vinner that the monadic theory of the language with unary predicates, equality and one unary function is decidable (Shelah 1975). By definitions of both semantics of stable models and circumscription, we have the following fact.

Fact 3. *The class $[all, (\omega, \omega), (\omega, 1)]_=$ is decidable under both the stable model semantics and circumscription.*

Actually, under the stable model semantics, decidability is not a consequence of the finite model property. In the next section, we will see some fragments with finite model property which are undecidable under the stable model semantics. Below, we consider the decidability of the existential class under the stable model semantics. As we have mentioned above, this class has the finite model property.

Theorem 1. *Let $\exists\bar{x}\varphi \in [\exists^*, all, all]_=$ be a sentence that has a stable model where $\bar{x} = (x_1, \ldots, x_p)$ and φ is quantifier free. Then there must exist a first-order sentence in $[\exists^*\forall^*, all, all]_=$ that is equivalent to $\mathrm{SM}(\exists\bar{x}\varphi)$.*

Proof. (Sketch) We only consider the special case in which ψ is built from one n-ary predicate P. A similar argument with minor modifications can be applied to the general case.

Notice that, according to Lin's transformation, we have that $\mathrm{SM}(\exists\bar{x}\varphi(P))$ is satisfiable iff $\mathrm{CIRC}(\exists\bar{x}\varphi^*(Q); Q) \wedge (X = Q)$ is satisfiable (Ferraris, Lee, and Lifschitz 2007). So, it suffices to show that, for every existential sentence $\exists\bar{x}\gamma(\bar{x}, P)$, there is a first-order sentence in $[\exists^*\forall^*, all, all]_=$ that is equivalent to $\mathrm{CIRC}(\exists\bar{x}\gamma(\bar{x}, P); P)$. This strengthens Theorem 1 in (Kolaitis and Papadimitriou 1990).

Similar to the proof of the original theorem, we write γ as a disjunction of formulas ψ_i $(1 \leq i \leq m)$ for some m. Then, $\mathrm{CIRC}(\exists\bar{x}\gamma(\bar{x}, P); P)$ is equivalent to a disjunction of the following sentences

$$\exists\bar{x}\psi_i \wedge \forall X (X < P \rightarrow \forall\bar{x}\neg\gamma(\bar{x}, X)) \qquad (1)$$

where $1 \leq i \leq m$. Let T_i be the set of term tuples $\bar{t}$ such that $P(\bar{t})$ is a conjunct in ψ_i. Finally, we complete the proof by show that (1) is equivalent to the following formula:

$$\exists\bar{x}\left(\psi_i \wedge \forall\bar{y}\left(P(\bar{y}) \leftrightarrow \bigvee_{\bar{s} \in T_i} (\bar{y} = \bar{s})\right) \wedge \bigwedge_{\varnothing \subsetneq S \subseteq T_i} \forall\bar{z}\vartheta_S\right) \tag{2}$$

where $\bar{y} = (y_1, \ldots, y_n)$ and $\bar{z} = (z_1, \ldots, z_p)$ are two sequences of pairwise distinct individual variables that have no occurrence in φ, the notation $\bar{y} = \bar{s}$ stands for the conjunction of all equalities $y_i = s_i$ for $1 \leq i \leq n$, and ϑ_S is a formula obtained from $\neg\gamma(\bar{z}, X)$ by substituting the formula $P(\bar{t}) \wedge \bigwedge_{\bar{s} \in S} \neg(\bar{t} = \bar{s})$ for each atomic formula $X(\bar{t})$. $\square$

A class of sentences is said to be *semi-decidable* under the stable model semantics if there is a deterministic algorithm such that, for each formula in this class, the algorithm halts with the result "yes" if it has no stable model, otherwise the algorithm does not halt or halts with "no".

The semi-decidability of the existential class immediately follows from the semi-decidability of first-order logic.

Corollary 2. *The class $[\exists^*, all, all]_=$ is semi-decidable under the stable model semantics.*

In the proof of Theorem 1, for every sentence without occurrences of function constants with arity ≥ 1, we can find an equivalent sentence of first-order logic in the class $[\exists^*\forall^*, all, (0)]_=$. Ramsey showed that the classical satisfiability problem for this class is decidable (Börger, Grädel, and Gurevich 1997). Hence, we have the following result.

Corollary 3. *The class $[\exists^*, all, (\omega)]_=$ is decidable under the stable model semantics.*

Undecidability

In this section, we address ourselves to undecidable fragments under the stable model semantics and circumscription.

Let φ be a first-order formula, let $P_1, \ldots, P_n$ be the set of all predicates occurring in φ where P_i is of k_i-ary. We define $\tilde{\varphi}$ to be the conjunction of φ and the following formula

$$\bigwedge_{1 \leq i \leq n} \forall x_1 \cdots \forall x_{k_i}(P_i(x_1, \ldots, x_{k_i}) \vee \tilde{P}_i(x_1, \ldots, x_{k_i})),$$

where $\tilde{P}_1, \ldots, \tilde{P}_n$ are pairwise distinct and each $\tilde{P}_i$ is a k_i-ary predicate constant that does not occur in φ. The following proposition can be easily obtained.

Lemma 4. *A first-order sentence φ is satisfiable iff $\tilde{\varphi}$ has a minimal model iff $\tilde{\varphi}$ has a stable model.*

A first-order formula is said to be in *universal implicational normal form*, if it is a prenex formula involving only universal first-order quantifiers, and the quantifier-free part of it is a conjunction of rules in the following form

$$\gamma_1 \wedge \cdots \wedge \gamma_m \rightarrow \gamma_{m+1} \vee \cdots \vee \gamma_n$$

where γ_i $(1 \le i \le n)$ are atomic formulas and $1 \le m \le n$.

Lemma 5. *For every first-order sentence φ in universal implicational normal form,* $\mathrm{SM}(\varphi)$ *is equivalent to* $\mathrm{CIRC}(\varphi)$.

Proof. By the definition. $\qquad\square$

Let n be a positive integer to denote the number of tile types and let H, V be two binary relations on $\mathbb{Z}_n$. A *domino system* is a triple $\mathcal{D} = (n, H, V)$. We say $\mathcal{D}$ *tiles* the grid $\mathbb{N} \times \mathbb{N}$ if there is a tiling $\tau : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Z}_n$ such that for all $i, j \in \mathbb{N}$: (i) if $\tau(i, j) = k$ and $\tau(i + 1, j) = l$ then $(k, l) \in H$; and (ii) if $\tau(i, j) = k$ and $\tau(i, j + 1) = l$ then $(k, l) \in V$. The domino problem is defined as follows: Given a domino system $\mathcal{D}$, does $\mathcal{D}$ tile $\mathbb{N} \times \mathbb{N}$? It was proved in (Berger 1966) that the domino problem is undecidable.

Theorem 6. *The class $[\forall^2, (0, 0, 1), (0, 1)]$ is undecidable under both the stable model and circumscriptive semantics.*

Proof. (Sketch) Let P be a binary predicate and s a unary function symbol. Given a domino system $\mathcal{D} = (n, H, V)$, let us first construct a class of formulas from $\mathcal{D}$ as follows:

$$\bigvee_{0 \le k \le n} P(x^k, x^k) \wedge \bigwedge_{0 \le k < l \le n} \neg(P(x^k, x^k) \wedge P(x^l, x^l)) \quad (3)$$

$$\bigvee_{1 \le k \le n} P(x, y^k) \wedge \bigwedge_{1 \le k < l \le n} \neg(P(x, y^k) \wedge P(x, y^l)) \quad (4)$$

$$\bigwedge_{(k,l) \notin H} \neg(P(x, y^{k+1}) \wedge P(x^{n+1}, y^{l+1})) \quad (5)$$

$$\bigwedge_{(k,l) \notin V} \neg(P(x, y^{k+1}) \wedge P(x, y^{l+n+2})) \quad (6)$$

where t^k stands for the term obtained from t by applying the function symbol s exactly k times if t is a term. Then, we define $\phi = \forall x \forall y ((3) \wedge (P(x, x) \wedge P(y, y) \rightarrow (4) \wedge (5) \wedge (6)))$. Clearly, ϕ is a sentence in the class $[\forall^2, (0, 0, 1), (0, 1)]$. Now we claim that $\mathcal{D}$ tiles $\mathbb{N} \times \mathbb{N}$ iff ϕ has a minimal model.

For the direction of "if", we assume that $\mathfrak{A}$ is a minimal model of ϕ. Consequently, $\mathfrak{A}$ is a model of (3), which implies: (i) there is an element $a_0 \in A$ such that $a_0 \in P$; (ii) for $k > 0$ and $1 \le i \le n$, $a_{k(n+1)} \in P$ and $a_{k(n+1)+i} \notin P$, where, for each $m \in \mathbb{N}$, a_m is defined to be the element obtained from a_0 by applying the function s^A exactly m times. Now we define a function $\tau : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Z}_n$ as follows: $\tau(i, j) = k$ iff $(a_i, a_{j+k+1}) \in P$. By property (ii) mentioned above and (4), τ is well-defined. And by (5) and (6), we can conclude that τ is a tiling of $\mathcal{D}$ that tiles $\mathbb{N} \times \mathbb{N}$.

For the converse direction, suppose τ is a tiling of $\mathcal{D}$ that tiles $\mathbb{N} \times \mathbb{N}$. Let σ be the vocabulary of ϕ, and let $\mathfrak{A}$ be a σ-structure with universe $\mathbb{N}$ defined as follows: (i) for all $i \in$ $\mathbb{N}$, $s(i) = i+1$; (ii) P consists of pair $(i \cdot (n+1), i \cdot (n+1))$ for each $i \in \mathbb{N}$ and of ordered pair $(i \cdot (n+1), j \cdot (n+1) + k + 1)$ for each ordered pair $(i, j) \in \mathbb{N}^2$ such that $\tau(i, j) = k$. It is easy to check that $\mathfrak{A}$ is a minimal model of ϕ. So the class $[\forall^2, (0, 0, 1), (0, 1)]$ is undecidable under circumscription.

Obviously, ϕ can be effectively translated to an equivalent sentence in prenex implicational normal form. By Lemma 5, undecidability for the stable model case is obtained. $\quad\square$

Theorem 7. *The following classes are undecidable under both the stable model and circumscriptive semantics.*

(1)	$[\forall\exists\forall, (0, \omega, 2), (0)]$	(10)	$[\forall, (0), (0, 2)]_=$
(2)	$[\forall^3\exists, (0, \omega, 2), (0)]$	(11)	$[\forall, (0), (0, 0, 1)]_=$
(3)	$[\forall^*\exists, (0, 0, 2), (0)]$	(12)	$[\forall^2, (0, 2), (0, 0, 1)]$
(4)	$[\forall\exists\forall^*, (0, 0, 2), (0)]$	(13)	$[\forall^2\exists, (0, \omega, 2), (0)]_=$
(5)	$[\forall\exists\forall\exists^*, (0, 0, 2), (0)]$	(14)	$[\exists^*\forall^2\exists, (0, 0, 2), (0)]_=$
(6)	$[\forall^3\exists^*, (0, 0, 2), (0)]$	(15)	$[\forall^2\exists^*, (0, 0, 2), (0)]_=$
(7)	$[\forall\exists\forall, (0, 0, 2), (0)]$	(16)	$[\forall\exists\forall, (0, 0, 2), (\omega)]$
(8)	$[\exists^*\forall\exists\forall, (0, 0, 2), (0)]$	(17)	$[\forall^3\exists, (0, 0, 2), (\omega)]$
(9)	$[\exists^*\forall^3\exists, (0, 0, 2), (0)]$	(18)	$[\forall^2\exists, (0, 0, 2), (\omega)]_=$

Proof. Notice that the following classes are undecidable in first-order logic (Börger, Grädel, and Gurevich 1997):

$\langle 1 \rangle$	$[\forall\exists\forall, (0, \omega, 1), (0)]$	$\langle 10 \rangle$	$[\forall, (0), (0, 2)]_=$
$\langle 2 \rangle$	$[\forall^3\exists, (0, \omega, 1), (0)]$	$\langle 11 \rangle$	$[\forall, (0), (0, 0, 1)]_=$
$\langle 3 \rangle$	$[\forall^*\exists, (0, 0, 1), (0)]$	$\langle 12 \rangle$	$[\forall^2, (0, 1), (0, 0, 1)]$
$\langle 4 \rangle$	$[\forall\exists\forall^*, (0, 0, 1), (0)]$	$\langle 13 \rangle$	$[\forall^2\exists, (0, \omega, 1), (0)]_=$
$\langle 5 \rangle$	$[\forall\exists\forall\exists^*, (0, 0, 1), (0)]$	$\langle 14 \rangle$	$[\exists^*\forall^2\exists, (0, 0, 1), (0)]_=$
$\langle 6 \rangle$	$[\forall^3\exists^*, (0, 0, 1), (0)]$	$\langle 15 \rangle$	$[\forall^2\exists^*, (0, 0, 1), (0)]_=$
$\langle 7 \rangle$	$[\forall\exists\forall, (0, 0, 1), (0)]$	$\langle 16 \rangle$	$[\forall\exists\forall, (0, 0, 1), (\omega)]$
$\langle 8 \rangle$	$[\exists^*\forall\exists\forall, (0, 0, 1), (0)]$	$\langle 17 \rangle$	$[\forall^3\exists, (0, 0, 1), (\omega)]$
$\langle 9 \rangle$	$[\exists^*\forall^3\exists, (0, 0, 1), (0)]$	$\langle 18 \rangle$	$[\forall^2\exists, (0, 0, 1), (\omega)]_=$

For each $i \le 18$, by Lemma 4, it is not difficult to see that the classical satisfiability of formulas in class $\langle i \rangle$ can be reduced to the minimal-model existence of some formulas in class (i). This implies the undecidability in circumscriptive semantics. For the case of stable models, notice that there is an effective algorithm to translate every first-order sentence φ to a sentence ψ in the prenex implicational normal form such that $\mathrm{CIRC}(\varphi)$ is equivalent to $\mathrm{CIRC}(\psi)$. By a similar argument, we can also conclude that class (i) is undecidable under the stable model semantics. $\quad\square$

Now we give a method to eliminate equalities under the stable model semantics. First, we define a syntax translation. Given an arbitrary formula φ, let $\widehat{\varphi} = \forall x E(x, x) \wedge \lambda_\varphi$, where λ_φ is the formula obtained from φ by substituting $\neg\neg E(t, t')$ for each equality $t = t'$, and E is a special binary predicate constant that does not occur in φ.

Lemma 8. *For every sentence φ of first-order logic, φ has a stable model iff so does $\widehat{\varphi}$.*

Proof. (Sketch) For the direction of "if", assume $\mathfrak{A}$ is a stable model of $\widehat{\varphi}$. First we show the claim that E must be interpreted as the identity relation on A. Then, let $\mathfrak{B}$ be a structure obtained from $\mathfrak{A}$ by omitting the relation E^A, and show $\mathfrak{B}$ is a stable model of φ. For the converse direction,

let $\mathfrak{B}$ be a stable model of φ, let E^A be the identity relation on B, and let $\mathfrak{A}$ be the $\sigma \cup \{E\}$-structure $(\mathfrak{B}, E^A)$. It is not difficult to verify that $\mathfrak{A}$ is a stable model of $\widehat{\varphi}$. $\qquad\square$

Theorem 9. *The following standard classes are undecidable under the stable model semantics.*

(1) $[\forall, (0,0,1),(0,2)]$
(2) $[\forall, (0,0,1),(0,0,1)]$
(3) $[\forall^2 \exists, (0,\omega,3),(0)]$
(4) $[\exists^* \forall^2 \exists, (0,0,3),(0)]$
(5) $[\forall^2 \exists, (0,0,3),(\omega)]$
(6) $[\forall^2 \exists^*, (0,0,3),(0)]$

Proof. Immediate by Lemma 8 and Theorem 7. $\qquad\square$

Theorem 10. *The class $[\forall^3, (0,\omega,4),(0)]$ is undecidable under the stable model semantics.*

Proof. We prove the theorem by constructing formulas in $[\forall^3, (0,\omega,4),(0)]$ to express the domino problem. Given an arbitrary domino system $\mathcal{D} = (n,H,V)$, let $\phi = \forall x \forall y \forall z \psi$ where ψ is the conjunction of the following formulas:

$$\neg\mathrm{succ}_H(x,y) \vee \mathrm{succ}_H(x,y) \qquad (7)$$

$$\neg\mathrm{succ}_V(x,y) \vee \mathrm{succ}_V(x,y) \qquad (8)$$

$$(\mathrm{succ}_H(x,y) \to \mathrm{ok}_H(x)) \wedge (\neg\mathrm{ok}_H(x) \to \mathrm{ok}_H(x)) \qquad (9)$$

$$(\mathrm{succ}_V(x,y) \to \mathrm{ok}_V(x)) \wedge (\neg\mathrm{ok}_V(x) \to \mathrm{ok}_V(x)) \qquad (10)$$

$$\mathrm{succ}_H(x,y) \wedge \mathrm{succ}_V(x,z) \to D(y,z) \qquad (11)$$

$$\mathrm{succ}_V(y,x) \wedge \mathrm{succ}_H(z,x) \wedge D(y,z) \to \mathrm{ok}_D(y,z) \qquad (12)$$

$$\neg\mathrm{ok}_D(y,z) \wedge D(y,z) \to \mathrm{ok}_D(y,z) \qquad (13)$$

$$\bigvee_{0 \le k < n} P_k(x) \wedge \bigwedge_{0 \le k < l < n} \neg(P_k(x) \wedge P_l(x)) \qquad (14)$$

$$\bigwedge_{(k,l) \notin H} \neg(\mathrm{succ}_H(x,y) \wedge P_k(x) \wedge P_l(y)) \qquad (15)$$

$$\bigwedge_{(k,l) \notin V} \neg(\mathrm{succ}_V(x,y) \wedge P_k(x) \wedge P_l(y)) \qquad (16)$$

Herein, for $0 \le k < n$, P_k are unary predicates; succ_H, succ_V, D and ok_D are binary predicates. It is clear that ϕ is just a sentence in $[\forall^3, (0,\omega,4),(0)]$. Now our task is to show that $\mathcal{D}$ tiles $\mathbb{N} \times \mathbb{N}$ iff ϕ has a stable model.

For the direction of "if", we first assume $\mathfrak{A}$ is a stable model of ϕ, and then show the following statements are true:

1. For all $a \in A$, there is $b \in A$ such that $(a,b) \in \mathrm{succ}_H$.
2. For all $a \in A$, there is $b \in A$ such that $(a,b) \in \mathrm{succ}_V$.
3. For all $a,b,c \in A$, if $(a,b) \in \mathrm{succ}_H$ and $(a,c) \in \mathrm{succ}_V$, then there exists $d \in A$ such that $(b,d) \in \mathrm{succ}_V$ and that $(c,d) \in \mathrm{succ}_H$.

Let R be the set of order pairs (b,c) such that $(a,b) \in \mathrm{succ}_H$ and $(a,c) \in \mathrm{succ}_V$ for some element $a \in A$. Let α be any assignment in $\mathfrak{A}$ which assigns R to X_D and which assigns Q^A to X_Q for any other predicate Q in ϕ. It is obvious that α satisfies $\phi^*(\bar{X})$ in $\mathfrak{A}$. Since $\mathfrak{A}$ is a stable model of ϕ, we can conclude that $D^A = R$. Assume that there is an ordered pair (b,c) such that, for any $d \in A$, order pairs (b,d) and (c,d) don't belong to succ_V and succ_H respectively. Let β be an assignment that assigns $D \setminus \{(b,c)\}$ to X_{ok_H} and assigns Q^A to the other predicate variables X_Q. It is easy to see that β satisfies both $\bar{X} < \bar{P}$ and $\phi^*(\bar{X})$ in $\mathfrak{A}$. But this

is impossible since $\mathfrak{A}$ is a stable model of ϕ. So statement 3 must be true. Similarly, we can prove statements 1 and 2.

By statements 1–3, there are elements $a_{i,j} \in A$ $(i,j \in \mathbb{N})$ such that $(a_{i,j}, a_{i+1,j}) \in \mathrm{succ}_H$ and $(a_{i,j}, a_{i,j+1}) \in \mathrm{succ}_V$ hold for all $i,j \in \mathbb{N}$. Define $\tau : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}_n$ as follows: $\tau(i,j) = k$ iff $a_{i,j} \in P_k$ for all $i,j \in \mathbb{N}$. Notice that $\mathfrak{A}$ is a model of (14). So, for each pair (i,j) there exists exactly one $k \in \mathbb{Z}_n$ such that $a_{i,j} \in P_k$, which implies τ is both well-defined and total. Moreover, since both (15) and (16) are satisfied by $\mathfrak{A}$, we have that τ is a tiling for $\mathbb{N} \times \mathbb{N}$.

For the converse direction, we assume that τ is a tiling of $\mathcal{D}$ for $\mathbb{N} \times \mathbb{N}$ firstly. Let σ be the vocabulary of ϕ, and let $\mathfrak{A}$ be a σ-structure with the universe $\mathbb{N} \times \mathbb{N}$ such that

- $\mathrm{succ}_H = \{((i,j),(i+1,j)) : i,j \in \mathbb{N}\}$;
- $\mathrm{succ}_V = \{((i,j),(i,j+1)) : i,j \in \mathbb{N}\}$;
- $\mathrm{ok}_H = \mathrm{ok}_V = \mathbb{N} \times \mathbb{N}$;
- $\mathrm{ok}_D = D = \{((i,j+1),(i+1,j)) : i,j \in \mathbb{N}\}$;
- for all $k \in \mathbb{Z}_n$, $P_k = \{(i,j) : i,j \in \mathbb{N} \text{ and } \tau(i,j) = k\}$.

It is easy to check that $\mathfrak{A}$ is a model of ϕ. Let α be an assignment in $\mathfrak{A}$. To obtain a contradiction, we assume that both formulas $\phi^*(\bar{X})$ and $\bar{X} < \bar{P}$ are satisfied by α in $\mathfrak{A}$. Since

$$\forall x \forall y (\neg\mathrm{succ}_H(x,y) \vee X_{\mathrm{succ}_H}(x,y))$$

can be inferred from $\phi^*(\bar{X})$, any proper subset of succ_H cannot be assigned to X_{succ_H} by α. As a consequence, it must be held that $\alpha(X_{\mathrm{ok}_H}) = \mathbb{N} = \mathrm{ok}_H$. By similar arguments, we can also show that relations succ_V, ok_V, D and ok_D are assigned to variables X_{succ_V}, X_{ok_V}, X_D and X_{ok_D} respectively. So, there is some $k \in \mathbb{Z}_n$ such that $\alpha(X_{P_k})$ is a proper subset of P_k. Consequently, there are $i,j \in \mathbb{N}$ such that (i,j) does not belong to $\alpha(X_{P_l})$ for any $l \in \mathbb{Z}_n$. But this is impossible since $\forall x \bigvee_{0 \le l < n} X_{P_l}(x)$ is a consequence of $\phi^*(\bar{X})$. Therefore, α does not satisfy both $\phi^*(\bar{X})$ and $\bar{X} < \bar{P}$ in $\mathfrak{A}$. Because the assignment α is arbitrary, $\mathfrak{A}$ must be a stable model of ϕ. And this completes the proof. $\qquad\square$

Remark. Theorems 9 and 10 show that the standard classes $[\exists^* \forall \exists^*, all, all]$, $[\exists^* \forall^2 \exists^*, all, (\omega)]$ and $[\exists^* \forall^*, all, (\omega)]_=$, which are decidable under circumscription, are undecidable under the semantics of stable models.

Maximally Decidable Standard Classes

If K is decidable under the stable model semantics (or circumscription), but every standard class $\mathsf{K}' > \mathsf{K}$ is not, then we say K is *maximally decidable* for the stable model semantics (or circumscription, respectively).

To obtain the exact boundary between decidability and undecidability over standard classes, we have to identified maximally decidable classes. It seems very difficult to find all such classes, but by the results proved in previous sections, we have the following:

Corollary 11. *The following classes are maximally decidable standard classes under the circumscriptive semantics:*

(1) $[\exists^* \forall^*, all, (\omega)]_=$
(2) $[\exists^* \forall^2 \exists^*, all, (\omega)]$
(3) $[all, (\omega,\omega),(\omega,\omega)]$
(4) $[\exists^* \forall \exists^*, all, all]$
(5) $[\exists^*, all, all]_=$
(6) $[all, (\omega,\omega),(\omega,1)]_=$

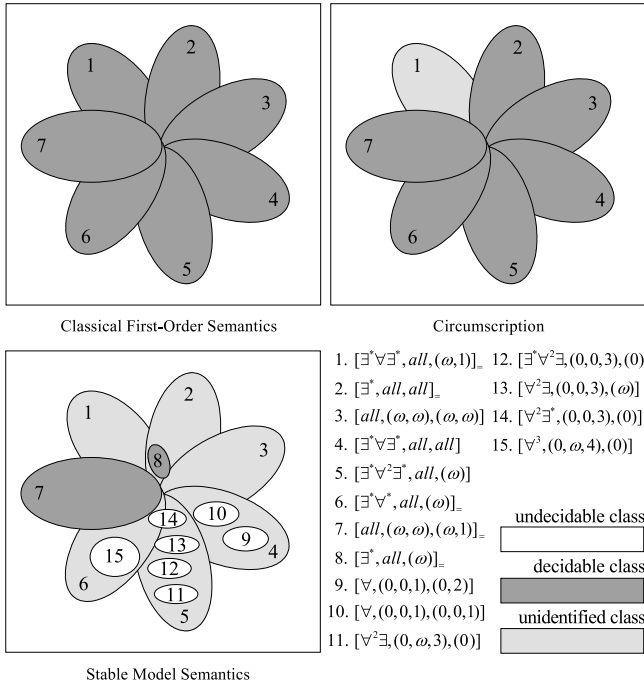|	
1. $[\exists^*\forall\exists^*, all, (\omega,1)]_=$	12. $[\exists^*\forall^2\exists, (0,0,3), (0)]$
2. $[\exists^*, all, all]_=$	13. $[\forall^2\exists, (0,0,3), (\omega)]$
3. $[all, (\omega,\omega), (\omega,\omega)]$	14. $[\forall^2\exists^*, (0,0,3), (0)]$
4. $[\exists^*\forall\exists^*, all, all]$	15. $[\forall^3, (0,\omega,4), (0)]$
5. $[\exists^*\forall^2\exists^*, all, (\omega)]$	
6. $[\exists^*\forall^*, all, (\omega)]_=$	
7. $[all, (\omega,\omega), (\omega,1)]_=$	undecidable class
8. $[\exists^*, all, (\omega)]_=$	decidable class
9. $[\forall, (0,0,1), (0,2)]$	
10. $[\forall, (0,0,1), (0,0,1)]$	unidentified class
11. $[\forall^2\exists, (0,\omega,3), (0)]$	

Figure 1: The Classification of Decidable Standard Classes

Proof. Immediate by Fact 2, 3 and Theorems 6 and 7. □

Corollary 12. *The class* $[all, (\omega,\omega), (\omega,1)]_=$ *is a maximally decidable standard class under the stable model semantics.*

Proof. Immediate by Fact 3, Theorems 6, and the undecidability of class (10) in Theorem 7. □

Conclusion

This paper characterizes some decidable and undecidable fragments of first-order language under the semantics of stable models and circumscription. A simple illustration of these results is given in Figure 1. From it we can see that the boundaries between decidability and undecidability under the two semantics are very different. This is rather counterintuitive due to the similarity of definitions for two semantics. According the viewpoint of (Pearce 2008), stable model = stable part + minimal model, that is, the stable model semantics is obtained by introducing stabilization into the circumscriptive semantics. The difference between decidability under these two semantics shows that the stabilization indeed enhance the expressive power of language.

Another interesting point arises from a comparison between decidability under circumscription and decidability of classical first-order logic. In spite of its second-order definition and high undecidability of the full first-order language, decidability of all the fragments considered in the paper under the semantics of circumscription coincides with that in classical first-order logic. We guess that such a coincidence comes from the finite model property of these fragments or the manageable behavior of monadic second-order logic.

References

Berger, R. 1966. The undecidability of the domino problem. *Memoirs of the AMS* (66):72.

Börger, E.; Grädel, E.; and Gurevich, Y. 1997. *The classical decision problem.* Berlin: Springer.

Ferraris, P.; Lee, J.; and Lifschitz, V. 2007. A new perspective on stable models. In *Proceedings of IJCAI'07*, 372–379.

Ferraris, P.; Lee, J.; and Lifschitz, V. 2010. Stable models and circumscription. *Artificial Intelligence* to appear.

Gelfond, M., and Lifschitz, V. 1988. The stable model semantics for logic programming. In *Proceedings of ICLP'88*, 1070–1080.

Gelfond, M. 2008. Answer sets. In van Harmelen, F.; Lifschitz, V.; and Porter, B., eds., *Handbook of Knowledge Representation*, 285–316. Elsevier.

Kolaitis, P. G., and Papadimitriou, C. H. 1990. Some computational aspects of circumscription. *Journal of the ACM* 37:1–14.

Lifschitz, V. 1985. Computing circumscription. In *Proceedings of IJCAI'85*, 121–127.

Lifschitz, V. 1994. Circumscription. In Gabbay, D.; Hogger, C.; and Robinson, J., eds., *The Handbook of Logic in AI and Logic Programming, volume 3*, 298–352. Oxford University Press.

Lin, F., and Zhou, Y. 2007. From answer set logic programming to circumscription via logic of GK. In *Proceedings of IJCAI'07*, 441–446.

McCarthy, J. 1980. Circumscription – a form of nonmonotonic reasoning. *Artificial Intelligence* 13(1-2):27–39.

McCarthy, J. 1986. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence* 28(1):89–116.

Pearce, D. 2008. Sixty years of stable models. In *Proceedings of ICLP'08*, 52.

Rabinov, A. 1989. A generalization of collapsible cases of circumscription. *Artificial Intelligence* 38:111–117.

Schlipf, J. S. 1986. How uncomputable is general circumscription? In *Proceedings of LICS'86*, 92–95.

Schlipf, J. S. 1987. Decidability and definability with circumscription. *Annals of Pure and Applied Logic* 35:173–191.

Shelah, S. 1975. The monadic theory of order. *Annals of Mathematics* 102(3):379–419.