

The Undecidability of First Order Logic

A first order logic is given by a set of function symbols and a set of predicate symbols. Each function or predicate symbol comes with an arity, which is natural number. Function symbols of arity 0 are known as constant symbols. Now *terms* are recursively defined by

- variables are terms, and
- if t_i are terms for $i=1, \dots, n$ and f is an n -ary function symbol, then $f(t_1, \dots, t_n)$ is a term.

Formulas are recursively defined by

- if t_i are terms for $i=1, \dots, n$ and p is an n -ary predicate symbol, then $p(t_1, \dots, t_n)$ is a formula.
- if P and Q are formulas, then $\neg P$ is a formula, $P \wedge Q$ is a formula, $P \vee Q$ is a formula, $P \rightarrow Q$ is a formula, and $P \Leftrightarrow Q$ is a formula.
- if P is a formula and x a variable, then $\forall x: P$ and $\exists x: P$ are formulas.

The completeness theorem for first order logic says that a formula is provable from the laws of first order logic (not given here) if and only if it is true in under all possible interpretations, i.e. regardless of the meaning of the function and predicate symbols.

Theorem: It is undecidable whether a first order logic formula is provable (or true under all possible interpretations).

Proof: Suppose there is an algorithm B that, given a first order logic and a formula in that logic, decides whether that formula is valid (holds under all possible interpretations). I will use that to give a decision algorithm for the language $\{(M, w) \mid M \text{ is the description of a Turing machine that accepts the string } w\}$. As the latter problem is undecidable this will show that B cannot exist.

Given M and w , create a first order logic by declaring a constant λ , a unary function symbol a for every letter a in the alphabet, and a binary predicate f_q for every state q of M .

Consider the following interpretation of this logic: Variables x range over strings over the given alphabet, λ denotes the empty string, $a(w)$ denotes the string aw , and $f_q(x, y)$ indicates that M , when given input w , can reach a

configuration with state q , in which xy is on the tape, with x in reverse order, and the head of M points at the first position of y . Under this interpretation $f_{q_0}(\lambda, w)$ is certainly a true formula, as the initial configuration is surely reachable. Here q_0 is the initial state, and w is a representation of w made from the constant and function symbols of the logic. Furthermore the formula $\exists x \exists y: f_{q\text{-acc}}(x, y)$ with $q\text{-acc}$ the acceptance state, holds iff M accepts w .

Whenever M has a transition from state q to state r , reading a , writing b , and moving right, the formula

$$\forall x \forall y: f_q(x, ay) \Rightarrow f_r(bx, y)$$

holds. Here x and y are variables. Likewise, if M has a transition from state q to state r , reading a , writing b , and moving left, the formulas

$$\forall x \forall y: f_q(cx, ay) \Rightarrow f_r(x, cby)$$

hold for every choice of a letter c . In addition we have

$$\forall x \forall y: f_q(\lambda, ay) \Rightarrow f_r(\lambda, by),$$

covering the case that M cannot move left, because its head is already in the left-most position. Finally, there are variants of the formulas above for the case that a is the blank symbol and that square of the tape is visited for the first time:

$$\begin{aligned} \forall x \forall y: f_q(x, \lambda) &\Rightarrow f_r(bx, \lambda) \\ \forall x \forall y: f_q(cx, \lambda) &\Rightarrow f_r(x, cb\lambda) \\ \forall x \forall y: f_q(\lambda, \lambda) &\Rightarrow f_r(\lambda, b\lambda). \end{aligned}$$

Let T be the conjunction of all implication formulas mentioned above. As M has finitely many transitions and the alphabet is finite, this conjunction is finite as well, and thus a formula of first order logic. Now consider the formula

$$f_{q_0}(\lambda, w) \ \& \ T \Rightarrow \exists x \exists y: f_{q\text{-acc}}(x, y).$$

In case M accepts w , there is a valid computation leading to an accept state. Each step therein corresponds with a substitution instance of one of the conjuncts in T , and using the laws of first order logic it is easy to check that the formula above is provable and thus true under all interpretations. If, on the other hand, the formula above is true under all interpretations, it is surely true in the given interpretation, which implies that M has an accepting computation starting on w .

Thus, in order to decide whether or not M accepts w , it suffices to check whether or not the formula above is a theorem of first order logic.

[Rob van Glabbeek](#)

rvg@cs.stanford.edu