Undecidability of First-Order Logic

Guram Bezhanishvili* Lawrence S. Moss†

Notes to the instructor

The material here will take around four weeks to cover in a classroom setting. The exact amount depends on the background of the students, of course. For students with no background in logic, one would have to do more in the way of examples of sentences. For those with no background on Turing machines, it would be useful to look up Turing machine simulators on the internet. Very simple programs that the class writes could then serve as examples throughout the rest of the module, illustrating the model construction and the long sentences as well. For an on-line educational presentation of register machines (a variant of Turing machines), one might see the courseware material by the second author [6]. That presentation also allows one to present results such as the undecidability of the halting problem *without any coding whatsoever*. However, it is not based on historical sources.

We have tried hard to emphasize the ideas and the significance of the results. Surely a classroom presentation should allow for much discussion in addition to a lecture.

Our presentation avoids Gödel numbering. So students who have never seen this will have questions about what "computability" means for a function from Turing machines to sentences. We did not want to present any details, partly because different instructors have different ways to do this, and partly because these details could easily obscure the main ideas.

There are many ways to prove the undecidability of first-order satisfiability. Büchi [2] provided the first proof along the lines that we presented it. Our approach also owes a lot to work that uses the technique of *tiling* that was pioneered by Hao Wang [4, 9] and later studied by many people, including Robert Berger [1] and Raphael Robinson [8]. Indeed, one alternative approach to our work would be to first show that the tiling problem is undecidable, and then to use this as a stepping-stone to the undecidability of first-order satisfiability. See [6] for one exposition of this.

Another way would be to prove the undecidability of the *word problem* for Thue processes, or the undecidability of some other class of word problems. This is the approach of Davis et al [3] (which goes back to the work of Post [7][1] and Andrey Andreyevich Markov (1903–1979) [5]).

Yet another way to do all of our work would be to use an equational presentation of Turing machines. Here is a sketch of this approach. Given a Turing machine M with alphabet symbols $s_1, \ldots, s_j$ and states, $q_1, \ldots, q_k$, we construct a finite, purely equational set of sentences $E(M)$. The signature of these equations contains the states $q_1, \ldots, q_k$ as constants, another constant ε (representing the empty word), and finally a constant h (representing halting). Further, the alphabet symbols become *unary function symbols* $\mathsf{g}_1, \ldots, \mathsf{g}_j$ (corresponding to $s_1, \ldots, s_j$, respectively). There is one additional unary function symbol m, and one 5-place symbol f. The basic idea is that

*Department of Mathematical Sciences; New Mexico State University; Las Cruces, NM 88003-8001; gbezhani@nmsu.edu.

†Department of Mathematics; Indiana University; Bloomington, IN 47405-7106; lsm@cs.indiana.edu.

[1]It is interesting that it was Church who suggested the problem to Post!

terms correspond to configurations of M with an extra "time stamp." For example, if at step 4 of a run, a machine M is in state q_a and the head is reading symbol s_b, and the tape is

$$s_c \qquad s_d \qquad s_d \qquad \underline{s_b} \qquad s_d \qquad s_c$$

(we have underlined the head position), then we would encode this by the term

$$\mathsf{f}(q_a, s_b, \mathsf{g}_d(\mathsf{g}_d(\mathsf{g}_c(\varepsilon))), \mathsf{g}_d(\mathsf{g}_c(\varepsilon)), \mathsf{m}(\mathsf{m}(\mathsf{m}(\mathsf{m}(\varepsilon)))))$$

Notice that the portion of the tape to the left of the reading head is encoded from *right-to-left*, the opposite of the standard encoding. The action of the machine is expressed in equations in this signature. For example, suppose that M has $(q_a, s_b) \Rightarrow (q_c, L, s_a)$. We then expect the tape to read

$$s_c \qquad s_d \qquad \underline{s_d} \qquad s_a \qquad s_d \qquad s_c$$

and we take the equations

$$\mathsf{f}(q_a, s_b, \mathsf{g}_d(x), y, z) \quad = \quad \mathsf{f}(q_c, s_d, x, \mathsf{g}_a(y), \mathsf{m}(z))$$

Note that x, y, and z here are *variables*; as usual, we understand equations as being universally quantified. There would be several equations corresponding to each $\Rightarrow$ assertion about M. One final equation would be $\mathsf{f}(q_0, s_0, x, y, z) = \mathsf{h}$; here q_0 is the halting state, and we assume that whenever M is in q_0, the symbol under the reading head is s_0. This defines a finite set $E(M)$ of equational axioms, and $M \mapsto E(M)$ is computable. The main point is to check that

$$E(M) \vdash \mathsf{f}(q_1, s_0, \varepsilon, \varepsilon, \varepsilon) = \mathsf{h}$$

if M halts when started in its starting state q_1, s_0 is under the reading head, the rest of the tape is blank, and the time counter is at 0. To prove this result takes some work.

In the (more significant) left-to-right direction, it is useful to use semantic arguments based on an "intended model." It is also useful to know that first-order provability from equational axioms coincides with provability in the smaller and more manageable *equational logic*. This use of equational logic would be a distraction for students who have not seen it, and probably the treatment of Church's Theorem in the main body of this module would be more accessible.

References

[1] Robert Berger, *The undecidability of the domino problem*, Mem. Amer. Math. Soc. No. **66** (1966), 72.

[2] J. Richard Büchi, *Turing-machines and the Entscheidungsproblem*, Math. Ann. **148** (1962), 201–213.

[3] Martin D. Davis, Ron Sigal, and Elaine J. Weyuker, *Computability, complexity, and languages: Fundamentals of theoretical computer science*, second ed., Computer Science and Scientific Computing, Academic Press, Boston, MA, 1994.

[4] A. S. Kahr, Edward F. Moore, and Hao Wang, *Entscheidungsproblem reduced to the ∀∃∀ case*, Proc. Nat. Acad. Sci. U.S.A. **48** (1962), 365–377.

[5] A. Markov, *On the impossibility of certain algorithms in the theory of associative systems*, Doklady Akad. Nauk SSSR (N.S.) **55** (1947), 583–586.

[6] Lawrence S. Moss, `1#`*: A text register machine introduction to computability theory*, courseware avaliable from `http://www.indiana.edu/~iulg/trm/`, 2007.

[7] Emil L. Post, *Recursive unsolvability of a problem of Thue*, J. Symbolic Logic **12** (1947), 1–11.

[8] Raphael M. Robinson, *Undecidability and nonperiodicity for tilings of the plane*, Invent. Math. **12** (1971), 177–209.

[9] Hao Wang, *Dominoes and the AEA case of the decision problem*, Proc. Sympos. Math. Theory of Automata (New York, 1962), Polytechnic Press of Polytechnic Inst. of Brooklyn, Brooklyn, N.Y., 1963, pp. 23–55.