MA5219 LOGIC AND FOUNDATION OF MATHEMATICS I

Lecture 9: Provability

Tin Lok Wong

10 September, 2018

The aim of this lecture is to investigate what incompleteness tells us about provability, and vice versa.

Rosser's Incompleteness Theorem states that no recursive consistent theory $T \vdash \mathrm{R}(\exp)$ can decide all Π_1 sentences. If a theory T does not decide a sentence σ, then whether one should add σ or $\neg\sigma$ to T as an extra axiom is generally a philosophical question. However, in the simplest case when the undecided sentence σ is Π_1, there is an obvious choice.

Note 9.1. Let σ be a Π_1 sentence and T be an $\mathcal{L}_\mathrm{A}(\exp)$ theory such that $T \vdash \mathrm{R}(\exp)$. If $T \nvdash \neg\sigma$, then $\mathbb{N} \models \sigma$.

Proof. Suppose $\mathbb{N} \not\models \sigma$. Then $\mathbb{N} \models \neg\sigma$ by the truth definition. From Lemma 5.2(1), we see that $\neg\sigma$ is equivalent to a Σ_1 sentence. So the Σ_1 completeness of $\mathrm{R}(\exp)$ implies $T \vdash \mathrm{R}(\exp) \vdash \neg\sigma$. □

When we proved the Henkinization Lemma in Lecture 7, we saw a procedure to extend any consistent theory to one that decides all sentences. This procedure cannot be entirely algorithmic because otherwise it would contradict the First Incompleteness Theorem. The only non-algorithmic part in this procedure is the case splitting, which involves a question on provability. Therefore, provability cannot be recursive, although it is r.e. by Assignment 8.2. In other words, proofs are algorithmically checkable, but provability is not: there is no pattern in proofs that one can exploit to show unprovability in a uniform way.

Theorem 9.2 (Raphael Robinson)**.** The theory $\mathrm{R}(\exp)$ is *essentially undecidable*, i.e., for every recursive consistent $\mathcal{L}_\mathrm{A}(\exp)$ theory $T \vdash \mathrm{R}(\exp)$,

$$\{\ulcorner\sigma\urcorner : \sigma \text{ is an } \mathcal{L}_\mathrm{A}(\exp) \text{ sentence and } T \vdash \sigma\}$$

is not recursive.

Proof. For each $j \in \mathbb{N}$, denote by σ_j the $\mathcal{L}_\mathrm{A}(\exp)$ sentence whose Gödel number is the $(j+1)$th smallest element in

$$\{\ulcorner\sigma\urcorner : \sigma \text{ is an } \mathcal{L}_\mathrm{A}(\exp) \text{ sentence}\}.$$

Notice $(\sigma_j)_{j\in\mathbb{N}}$ is recursive by the Church–Turing Thesis. We will construct another sequence of $\mathcal{L}_\mathrm{A}(\exp)$ sentences $(\tau_j)_{j\in\mathbb{N}}$ by recursion, with the inductive assumption that at each stage $j \in \mathbb{N}$,

$$T_j := T \cup \{\tau_i : i < j\} \nvdash \bot.$$

At the end, we will set $T^* = T \cup \{\tau_i : i \in \mathbb{N}\}$. The inductive assumption ensures that every finite subset of T^* is consistent. Thus $T^* \nvdash \bot$ by the Compactness Lemma.

Given $(\tau_i)_{i<j}$, define

$$\tau_j = \begin{cases} \neg\sigma_j, & \text{if } T_j + \sigma_j \vdash \bot; \\ \sigma_j, & \text{otherwise.} \end{cases}$$

Then (cut) ensures that the inductive assumption is maintained.

For every $\mathcal{L}_\mathrm{A}(\exp)$ sentence σ, either $T^* \vdash \sigma$ or $T^* \vdash \neg\sigma$ by (asn). Notice $T^* \supseteq T \vdash \mathrm{R}(\exp)$. Given an $\mathcal{L}_\mathrm{A}(\exp)$ sentence σ, we can 'decide' whether or not $\sigma \in T^*$ as follows:

(1) find $j \in \mathbb{N}$ such that $\sigma = \sigma_j$;

(2) run the construction above to get τ_j;

(3) if $\tau_j = \sigma_j$, then return `true`, else return `false`.

Clearly, if this procedure returns `true`, then $\sigma = \sigma_j = \tau_j \in T^*$. If it returns `false`, then $\tau_j = \neg\sigma_j$ by construction, and so $\sigma = \sigma_j \notin T^*$ by ($\bot$) and (asn), as $T^* \not\vdash \bot$. Hence the answers given by this 'decision procedure' are always correct.

By the First Incompleteness Theorem, this procedure cannot always give an answer. As a result, there is no algorithm which, when given $(\tau_i)_{i<j}$ and σ_j as inputs, decides whether or not

$$T_j + \sigma_j = T \cup \{\tau_i : i < j\} \cup \{\sigma_j\} \vdash \bot,$$

or equivalently, whether or not $T \vdash \bigwedge_{i<j} \tau_i \wedge \sigma_j \to \bot$. In particular, the set in question is not recursive by the Church–Turing Thesis. $\square$

Theorem 9.2 gives a natural example of a subset of $\mathbb{N}$ that is r.e. but not recursive. Hence, without **par-while**, strictly fewer sets can be computed.

After seeing what the First Incompleteness Theorem tells us about provability, we look at what certain properties of provability can tell us about incompleteness. More specifically, we would like to know what the undecided sentence given by the Incompleteness Theorem means. Simply knowing that a sentence asserts its own unprovability may not be informative because some formulas have non-equivalent fixed point.

Assignment 9.3. Prove that there are an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula $\varphi(y)$ and $\mathcal{L}_{\mathrm{A}}(\exp)$ sentences σ, τ such that $\mathrm{R}(\exp)$ proves

$$\sigma \leftrightarrow \varphi(\underline{\sigma}) \quad \text{and} \quad \tau \leftrightarrow \varphi(\underline{\tau}) \quad \text{and} \quad \neg(\sigma \leftrightarrow \tau).$$

(Hint: use a $\Delta_0(\exp)$ formula $\varphi(y)$ which expresses

'the number of symbols in y is odd under our Gödel numbering';

for example, let $\varphi(y)$ be

$$\exists z {\leqslant} y \ \left(2^{4(2z)} \leqslant y \wedge y < 2^{4(2z+1)}\right).$$

Fixed points of such $\varphi(y)$ can be found without using the Diagonal Lemma.) [7 points]

It turns out that, under suitable conditions on the provability predicate, the Gödel sentence asserting its own unprovability is unique up to logical equivalence. These conditions on the provability predicate are typically formulated using notation from modal logic, in which $\square \ldots$ is intended to mean 'it is necessary that ...'. For us $\square \underline{\sigma}$ is intended to mean $T \vdash \sigma$. When there is no risk of ambiguity, we often omit the parentheses and write $\square(\ldots)$ as $\square \ldots$. Read $\square \ldots$ as 'box ...'.

Definition (Löb)**.** An $\mathcal{L}_{\mathrm{A}}(\exp)$ formula $\square(y)$ is said to satisfy the *derivability conditions* over an $\mathcal{L}_{\mathrm{A}}(\exp)$ theory T if the following hold for all $\mathcal{L}_{\mathrm{A}}(\exp)$ sentences σ, τ.

(N) If $T \vdash \sigma$, then $T \vdash \square \underline{\sigma}$.

(IN) $T \vdash \square \underline{\sigma} \to \square \underline{\square \underline{\sigma}}$.

($\square$D) $T \vdash \square \underline{(\sigma \to \tau)} \to (\square \underline{\sigma} \to \square \underline{\tau})$.

The names N, IN, and $\square$D stand for necessitation, internal necessitation, and $\square$-distributivity respectively. As Assignment 8.2 shows, provability is r.e. and so can be expressed by a Σ_1 formula. Thus the Σ_1 completeness of $\mathrm{R}(\exp)$ suggests that provability implies provable provability, i.e., necessitation is true. Internal necessitation asserts the provability of necessitation. $\square$-distributivity says that the provable sentences are provably closed under *modus ponens*.

Incompleteness Theorems (Gödel)**.** Let T be any $\mathcal{L}_{\mathrm{A}}(\exp)$ theory and $\square(y)$ be an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula satisfying Löb's derivability conditions over T. If σ is an $\mathcal{L}_{\mathrm{A}}(\exp)$ sentence such that $T \vdash \sigma \leftrightarrow \neg \square \underline{\sigma}$, then

(1) $T \nvdash \bot$ if and only if $T \nvdash \sigma$; and

(2) $T \vdash \neg\Box\underline{\bot} \leftrightarrow \sigma$.

Proof. (1) If $T \vdash \sigma$, then

$$\begin{array}{lll} & T \vdash \Box\underline{\sigma} & \text{by (N);} \\ \therefore & T \vdash \neg\sigma & \text{by the choice of } \sigma; \\ \therefore & T \vdash \bot & \text{by } (\bot). \end{array}$$

Conversely, if $T \vdash \bot$, then

$$\begin{array}{lll} & T + \neg\sigma \vdash \bot & \text{by the definition of } \vdash; \\ \therefore & T \vdash \sigma & \text{by (RAA).} \end{array}$$

(2) By the Completeness Theorem, it suffices to show semantic entailment. Fix $M \models T$. If $M \models \neg\sigma$, then

$$\begin{array}{lll} & M \models \Box\underline{\sigma} & \text{by the choice of } \sigma; \\ \therefore & M \models \Box\underline{\Box\underline{\sigma}} & \text{by (IN);} \\ \therefore & M \models \Box\underline{\neg\sigma} & \text{by (N) and } (\Box\text{D}), \text{ as the choice of } \sigma \text{ implies } T \vdash \Box\underline{\sigma} \to \neg\sigma; \\ \therefore & M \models \Box\underline{\bot} & \text{by (N) and } (\Box\text{D}), \text{ as } T \vdash \sigma \to (\neg\sigma \to \bot). \end{array}$$

Conversely, if $M \models \Box\underline{\bot}$, then

$$\begin{array}{lll} & M \models \Box\underline{\sigma} & \text{by (N) and } (\Box\text{D}), \text{ as } T \vdash \bot \to \sigma; \\ \therefore & M \models \neg\sigma & \text{by the choice of } \sigma. \end{array}$$

□

Notice (the proof of) part (2) is essentially a T-provable version of (the proof of) part (1). As Gödel's Incompleteness Theorems show, provided Löb's derivability conditions hold over a theory T, all sentences that assert their own unprovability in T are equivalent to the consistency of T. As a consequence, under these conditions, no consistent theory can prove its own consistency.

This consequence, which is often referred to as Gödel's Second Incompleteness Theorem, is very useful in separating theories. To show that an extension T^* of a consistent theory T is strictly stronger, by definition one has to produce a model of T which does not satisfy T^*. Alternatively, one can proceed by proving the consistency of T in T^* because T cannot prove its own consistency by the Second Incompleteness Theorem. This alternative method is so useful in set theory that set-theoretic axioms are often classified in terms of their consistency strengths, i.e., the theories whose consistencies can be proved under the axioms.

Gödel's Incompleteness Theorems tells us that sentences asserting their own unprovability are all equivalent. The same is true of sentences asserting their own provability: they are actually all equivalent to $\top$. The proof is not long, but we omit it.

Löb's Theorem. Let T be an $\mathcal{L}_{\mathrm{A}}(\exp)$ theory and $\Box(y)$ be an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula satisfying Löb's derivability conditions over T. Then for all $\mathcal{L}_{\mathrm{A}}(\exp)$ sentences σ,

$$T \vdash \sigma \leftrightarrow \Box\underline{\sigma} \quad \Leftrightarrow \quad T \vdash \sigma.$$

The special case when $\sigma = \bot$ in Löb's Theorem tells us

$$T \vdash \bot \quad \Leftrightarrow \quad T \vdash \bot \leftrightarrow \Box\underline{\bot} \quad \Leftrightarrow \quad T \vdash \neg\Box\underline{\bot}.$$

Taking negations everywhere, we deduce that T is consistent if and only if it does not prove its own consistency. This is Gödel's Second Incompleteness Theorem. So Löb's Theorem is a generalization of Gödel's Second Incompleteness Theorem. It also says that the only cases in which the truth of a sentence is provably equivalent to its provability are the trivial ones, i.e., when the sentence (and hence by (N) its provability) is outright provable. In particular, truth and provability are really different, not only intentionally, but also extensionally.

Notice that in Gödel's Incompleteness Theorems, we could have taken $\Box(y)$ to be $\top(y)$, in which case the conclusion is trivial: no consistent theory can prove $\neg\Box\underline{\bot} = \neg\top = \bot$. To have a non-trivial conclusion, one needs to construct a formula $\Box(y)$ satisfying the derivability conditions which reasonably expresses provability. This task will occupy us for the next three lectures.