MA5219 LOGIC AND FOUNDATION OF MATHEMATICS I

Lecture 25: Finite consistency

Tin Lok Wong

12 November, 2018

The aim of this lecture is to demonstrate the pervasiveness of the incompleteness phenomenon by establishing a version of the Incompleteness Theorem for finite consistency statements.

The unprovability of Con(PA) in PA may seem disappointing. The Σ_1 completeness of PA provides some consolation: every instance of Con(PA) is provable in PA, i.e., we know PA $\vdash \mathrm{Con}(\underline{n}, \mathrm{PA})$ for every $n \in \mathbb{N}$, where $\mathrm{Con}(w, \mathrm{PA})$ is a $\Delta_0(\exp)$ formula expressing 'there is no proof of PA $\vdash \bot$ of length less than w'. The complication is that this involves infinitely many proofs. Given $n \in \mathbb{N}$, is it feasible to find a proof π_n of PA $\vdash \mathrm{Con}(\underline{n}, \mathrm{PA})$? More precisely, can the length of this π_n be bounded above by a polynomial in the length of n?

To address this question, we need a proper definition of length.

Definition.
- The *length* $\mathrm{len}(s)$ of a term/formula/proof s is the number of symbols in s.
- The length of the binary representation of $n \in \mathbb{N}$ is denoted $|n|$.

Remark 25.1. By the *length* of a proof, some authors mean the number of sequents/line in the proof. In their terminology, our notion of length is often referred to as *size*.

According to our arithmetization (from Lecture 10, say), if s is a term/formula/proof in $\mathcal{L}_{\mathrm{A}}(\exp)$, then $4\,\mathrm{len}(s) - 4 < |\ulcorner s \urcorner| \leqslant 4\,\mathrm{len}(s)$. So which notion of length we use for a syntactical object matters very little. One important fact to note is that numbers are exponential in their lengths.

Note 25.2. If $n \in \mathbb{N}$, then $|n| = \lceil \log_2(n+1) \rceil$.

Example 25.3. $|10| = |1010_2| = 4$.

Example 25.4. If $n \in \mathbb{N}$, then

$$\mathrm{len}(\underline{n}) = \mathrm{len}\big((\cdots((0 + \underbrace{1) + 1) \cdots + 1}_{n \text{ -many 1's}})\big) = 4n + 1,$$

which is exponential in $|n|$.

Example 25.4 points out one problem in the formulation of our question: one obviously cannot have proofs of PA $\vdash \mathrm{Con}(\underline{n}, \mathrm{PA})$ of lengths bounded by a polynomial in $|n|$ because we cannot even write down $\underline{n}$ with this number of symbols. To avoid this triviality, we employ more efficient numerals. These efficient numerals are sometimes called *dyadic numerals* since they are based on the binary representation of numbers. Following usual practice, we use the same notation for efficient and inefficient numerals.

Redefinition (for this and the next lecture)**.** Let $\underline{0} = 0$ and $\underline{1} = 1$. For all $n \in \mathbb{N}$, define

$$\underline{2(n+1)} = (((1+1) \times \underline{n+1}) + 0) \quad \text{and} \quad \underline{2(n+1)+1} = (((1+1) \times \underline{n+1}) + 1).$$

Example 25.5. $(((1+1) \times (((1+1) \times \underbrace{\underbrace{(((1+1) \times 1) + 0)}_{\underline{10_2}}) + 1)) + 0) = \underline{10}$.

$\underline{101_2}$

$\underline{1010_2}$

As one can see from the example above, the dyadic numeral of a number is reminiscent of its binary representation. Therefore, the length of the dyadic numeral of a number is bounded above by a fixed polynomial in the length of the number.

Observation 25.6. $\mathrm{len}(\underline{1}) = 1$ and $\mathrm{len}(\underline{n}) = 9\,|n| + 1$ whenever $n \in \mathbb{N} \setminus \{1\}$. □

Our question now becomes non-trivial. To state the answer, it is convenient to be able to indicate in the turnstile notation the length of the proof involved. Our notation for negating decorated turnstiles to be introduced below (is non-standard and) originates from Frege's *Begriffsschrift*. Our meaning of this notation is slightly different from his: his stroke negates the formula on the right of the turnstile, while our stroke negates the turnstile itself.

Definition. Write $T \mathrel{|\!\overset{n}{\vdash}} \sigma$ for 'there is a proof of $T \vdash \sigma$ of length strictly less than n'. The negation of $T \mathrel{|\!\overset{n}{\vdash}} \sigma$ is denoted $T \mathrel{|\!\overset{n}{\nvdash}} \sigma$. Write $T \mathrel{|\!\overset{|n|}{\vdash}_{*}} S(n)$ to mean

there is $t(X) \in \mathbb{N}[X]$ such that for all $n \in \mathbb{N}$,

$$T \mathrel{|\!\overset{t(|n|)}{\vdash}} S(n).$$

The main theorem of this lecture resembles our Incompleteness Theorems from Lecture 9. Here we only have one direction because apparently the lengths do not match in the converse. The proof is also very similar. The additional part is in realizing that many simple operations on proofs like *modus ponens* incur only a polynomial increase in length. The formula $\Box^{\leqslant}(w, y)$ below is intended to express $T \mathrel{|\!\overset{w}{\vdash}} y$.

Theorem 25.7 (H. Friedman, Pudlák, independently)**.** Let T be an $\mathcal{L}_{\mathrm{A}}(\exp)$ theory and $\Box^{\leqslant}(w, y)$ be an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula. Suppose $p(X), q(X) \in \mathbb{N}[X]$ satisfying the following conditions for all $n \in \mathbb{N}$ and all $\mathcal{L}_{\mathrm{A}}(\exp)$ sentences σ, τ.

(M) $T \vdash \forall w, y\ \big(\Box^{\leqslant}(w, y) \to \forall w' \geqslant w\ \Box^{\leqslant}(w', y)\big)$.

(N) If $T \mathrel{|\!\overset{n}{\vdash}} \sigma$, then $T \mathrel{|\!\overset{p(n)}{\vdash}} \Box^{\leqslant}(\underline{n}, \underline{\sigma})$.

(IN) $T \mathrel{|\!\overset{q(|n|+\mathrm{len}(\sigma))}{\vdash}} \Box^{\leqslant}(\underline{n}, \underline{\sigma}) \to \Box^{\leqslant}(\underline{p(n)}, \underline{\Box^{\leqslant}(\underline{n}, \underline{\sigma})})$.

(□D) $T \mathrel{|\!\overset{q(|n|+\mathrm{len}(\sigma)+\mathrm{len}(\tau))}{\vdash}} \Box^{\leqslant}(\underline{n}, \underline{\sigma \to \tau}) \to \big(\Box^{\leqslant}(\underline{n}, \underline{\sigma}) \to \Box^{\leqslant}(\underline{p(n)}, \underline{\tau})\big)$.

Let $\sigma(w)$ be an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula such that for all $n \in \mathbb{N}$,

$$T \mathrel{|\!\overset{q(|n|)}{\vdash}} \sigma(\underline{n}) \leftrightarrow \neg\, \Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})}).$$

Then there are $r(X), s(X) \in \mathbb{N}[X]$ satisfying, for all $n \in \mathbb{N}$,

(1) if $T \mathrel{|\!\overset{r(n)}{\nvdash}} \bot$, then $T \mathrel{|\!\overset{n}{\nvdash}} \sigma(\underline{n})$;

(2) $T \mathrel{|\!\overset{s(|n|)}{\vdash}} \neg\, \Box^{\leqslant}(\underline{r(n)}, \underline{\bot}) \to \sigma(\underline{n})$.

Proof. (1) Let $n \in \mathbb{N}$ such that $T \mathrel{|\!\overset{n}{\vdash}} \sigma(\underline{n})$. Then

	$T \mathrel{|\!\overset{p(n)}{\vdash}} \Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})})$		by (N);
$\therefore$	$T \mathrel{|\!\overset{p'(n)}{\vdash}} \neg\sigma(\underline{n})$	for some $p'(X) \in \mathbb{N}[X]$,	by the choice of σ;
$\therefore$	$T \mathrel{|\!\overset{p''(n)}{\vdash}} \bot$	for some $p''(X) \in \mathbb{N}[X]$,	by ($\bot$).

(2) By the choice of σ and by usual logic,

(a) $T \mathrel{|\!\overset{|n|}{\vdash}_{*}} \Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})}) \to \neg\sigma(\underline{n})$; and

(b) $T \mathrel{|\!\overset{|n|}{\vdash}_{*}} \sigma(\underline{n}) \to (\neg\sigma(\underline{n}) \to \bot)$.

Now

$$\begin{array}{lll} & T \mathrel{\overset{|n|}{\underset{*}{\vdash}}} \neg\sigma(\underline{n}) \to \Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})}) & \text{by the choice of } \sigma; \\ \therefore & T \mathrel{\overset{|n|}{\underset{*}{\vdash}}} \neg\sigma(\underline{n}) \to \Box^{\leqslant}\big(\underline{p(n)}, \underline{\Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})})}\big) & \text{by (IN);} \\ \therefore & T \mathrel{\overset{|n|}{\underset{*}{\vdash}}} \neg\sigma(\underline{n}) \to \Box^{\leqslant}\big(\underline{p'(n)}, \underline{\neg\sigma(\underline{n})}\big) \quad \text{for some } p'(X) \in \mathbb{N}[X], & \text{by (a), (N) and } (\Box\text{D}); \\ \therefore & T \mathrel{\overset{|n|}{\underset{*}{\vdash}}} \neg\sigma(\underline{n}) \to \Box^{\leqslant}(\underline{p''(n)}, \underline{\bot}) \quad \text{for some } p''(X) \in \mathbb{N}[X], & \text{by (b), (N) and } (\Box\text{D}). \end{array}$$

□

In a way similar to how we constructed our $\Box(y)$ in Lecture 10, one can construct a $\Delta_0(\exp)$ formula $\Box^{\leqslant}(w, y)$ satisfying the derivability conditions in the statement of Theorem 25.7 for any recursive $\mathcal{L}_{\mathrm{A}}(\exp)$ theory $T \supseteq \mathrm{I}\Delta_0(\exp)$ such that

$$\{(n, m) \in \mathbb{N}^2 : \mathbb{N} \models \Box^{\leqslant}(n, m)\} = \{(n, \ulcorner\theta\urcorner) \in \mathbb{N}^2 : \theta \text{ is an } \mathcal{L}_{\mathrm{A}}(\exp) \text{ formula and } T \mathrel{\overset{n}{\vdash}} \theta\}.$$

In the case when some polynomial-time algorithm can decide whether a formula is in the theory T or not, Pudlák showed one can reduce $\mathrm{I}\Delta_0(\exp)$ here to Q using some results connecting $\mathrm{I}\Delta_0(\exp)$ and Q from Wilkie and Paris. Given a formula $\Box^{\leqslant}(w, y)$, it is not hard to modify our proof of the Diagonal Lemma to obtain a self-referential formula $\sigma(w)$ satisfying the conditions in Theorem 25.7 *except* the length bound: simply add a free variable w everywhere and consider a function

$$D\colon (n, \ulcorner\theta\urcorner) \mapsto \ulcorner\theta(\underline{n}, \underline{\theta})\urcorner$$

for all $n \in \mathbb{N}$ and all $\mathcal{L}_{\mathrm{A}}(\exp)$ formulas $\theta(w, x)$. To obtain the length bound, apparently one needs a rather involved argument. So we omit the proof here.

Diagonal Lemma (numeral version with polynomial length bounds)**.** For every $\mathcal{L}_{\mathrm{A}}(\exp)$ formula $\varphi(w, y)$, there exist an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula $\sigma(w)$ and a polynomial $q(X) \in \mathbb{N}[X]$ such that whenever $n \in \mathbb{N}$,

$$\mathrm{Q}(\exp) \mathrel{\overset{q(|n|)}{\vdash\!\!\!-\!\!\!-}} \sigma(\underline{n}) \leftrightarrow \varphi(\underline{n}, \underline{\sigma(\underline{n})}).$$ □

From Theorem 25.7, one can readily derive a negative answer to the question posed at the beginning of this lecture. Amongst other things, this negative answer tells us that, even when efficient numerals are used, the proofs of true $\Delta_0(\exp)$ formulas in $\mathrm{R}(\exp)$ must in general be at least exponential in the lengths of the formulas. In the case of finite consistency statements, for finite theories capable of coding sequences, Pudlák established also an exponential upper bound on the proof lengths; see Theorem 26.1 in the next lecture.

Finite Incompleteness Theorem (H. Friedman, Pudlák, independently)**.** Let T be a consistent $\mathcal{L}_{\mathrm{A}}(\exp)$ theory extending $\mathrm{Q}(\exp)$ and $\Box^{\leqslant}(w, y)$ be an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula with $p(X), q(X) \in \mathbb{N}[X]$ satisfying (M), (N), (IN), ($\Box$D) in the statement of Theorem 25.7 for all $n \in \mathbb{N}$ and all $\mathcal{L}_{\mathrm{A}}(\exp)$ sentences σ, τ. Then there is a rational $\varepsilon > 0$ such that for all $n \in \mathbb{N}$,

$$T \mathrel{\overset{2^{\varepsilon|n|}}{\vdash\!\!\!-\!\!\!-}} \neg\Box^{\leqslant}(\underline{n}, \underline{\bot}).$$

Proof. Apply the numeral version of Diagonal Lemma above to find an $\mathcal{L}_{\mathrm{A}}(\exp)$ formula $\sigma(w)$ and a polynomial $q'(X) \in \mathbb{N}[X]$ such that for all $n \in \mathbb{N}$,

$$T \mathrel{\overset{q'(|n|)}{\vdash\!\!\!-\!\!\!-}} \sigma(\underline{n}) \leftrightarrow \neg\Box^{\leqslant}(\underline{n}, \underline{\sigma(\underline{n})}).$$

Without loss of generality, assume $q'(X) = q(X)$. Let $r(X), s(X) \in \mathbb{N}[X]$ we get from applying Theorem 25.7 to the current situation. Theorem 25.7(1) gives, for each $n \in \mathbb{N}$, a proof π_n of $T \vdash \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot}) \to \sigma(\underline{n})$ of length less than $s(|n|)$. For every proof ν_n of $T \vdash \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot})$, where $n \in \mathbb{N}$, let

$$\mu(\pi_n, \nu_n) \quad = \quad \dfrac{\begin{array}{cc} \vdots\ \pi_n & \vdots\ \nu_n \\ T \vdash \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot}) \to \sigma(\underline{n}) & T \vdash \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot}) \end{array}}{T \vdash \sigma(\underline{n})}\ (\mathrm{MP})$$

By studying (MP) closely, one obtains $s'(X) \in \mathbb{N}[X]$ which satisfies

$$\mathrm{len}(\mu(\pi_n, \nu_n)) \leqslant \mathrm{len}(\pi_n) + \mathrm{len}(\nu_n) + s'(|n|) < s(|n|) + \mathrm{len}(\nu_n) + s'(|n|)$$

for all such ν_n's. Let $\delta \in \mathbb{Q}$ with $0 < \delta < 1$. Some calculations reveal that for all sufficiently large $n \in \mathbb{N}$ and all proofs ν_n of $T \vdash \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot})$,

$$\mathrm{len}(\nu_n) > \mathrm{len}(\mu(\pi_n, \nu_n)) - s(|n|) - s'(|n|) \geqslant 2^{|n|-1} - s(|n|) - s'(|n|) \geqslant 2^{\delta|n|}$$

by Theorem 25.7(1). Hence $T \vdash^{2^{\delta|n|}}_{\top} \neg\Box^{\leqslant}(\underline{r(n)}, \underline{\bot})$ for all sufficiently large $n \in \mathbb{N}$. Some further calculations involving (M) then gives a positive $\varepsilon < \delta$ such that, for all sufficiently large $n \in \mathbb{N}$,

$$T \vdash^{2^{\varepsilon|n|}}_{\top} \neg\Box^{\leqslant}(\underline{n}, \underline{\bot}).$$

By choosing a smaller positive ε, one can make this hold for all $n \in \mathbb{N}$. □

Very loosely, the transition from the Second Incompleteness Theorem to the Finite Incompleteness Theorem suggests a general principle: if we bound all the quantifiers in a certain family of statements about arithmetic, then unprovability becomes non-polynomial provability. However, complexity-theoretic issues have prevented one from proving any reasonable formulation of this general principle so far. So the success in the Finite Incompleteness Theorem seems to be largely isolated.

Suppose T and $\Box^{\leqslant}(w, y)$ are as in the Finite Incompleteness Theorem. We can express the consistency of T as $\forall w\, \neg\Box^{\leqslant}(w, \underline{\bot})$. Then (∀R) gives

$$T + \forall w\, \neg\Box^{\leqslant}(w, \underline{\bot}) \vdash^{|n|}_{*} \neg\Box^{\leqslant}(\underline{n}, \underline{\bot}).$$

Therefore, in view of the Finite Incompleteness Theorem, we must have $T \nvdash \forall w\, \neg\Box^{\leqslant}(w, \underline{\bot})$. This shows a variant of the Second Incompleteness Theorem. This also shows one way of utilizing consistency: it helps shorten proofs of finite consistency statements. We will pursue this line further in the next lecture.