

Topics in Applied Math: Logic and Foundations of Mathematics

Part 4. First order arithmetic

Kazuyuki Tanaka

BIMSA

October 31, 2025



清华大学求真书院
Qiu Zhen College, Tsinghua University

Logic and Foundations

- **Part 1. Equational theory**
- **Part 2. First order theory**
- **Part 3. Basic Model theory**
- **Part 4. First order arithmetic and incompleteness theorems**
- **Part 5. Models of first-order arithmetic**
- **Part 6. Second order arithmetic and reverse mathematics**

Part 4. Schedule

- Oct. 24, (1) Peano arithmetic and representation theorems
- Oct. 29, (2) The first incompleteness theorem
- Oct. 31, (3) **The second incompleteness theorem**

- **Peano arithmetic** PA is a first-order theory of natural numbers in the language $\mathcal{L}_{\text{OR}} = \{+, \cdot, 0, 1, <\}$, consisting of basic axioms for arithmetical operations and **Induction**: $\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x\varphi(x)$.
- The formulas in \mathcal{L}_{OR} are classified as Σ_i and Π_i ($i \in \mathbb{N}$). In particular, a Σ_0 ($=\Pi_0$) formula is a bounded formula (only with bounded quantifiers $\forall x < t$ and $\exists x < t$). If φ is bounded, $\exists x_1 \cdots \exists x_k \varphi$ is Σ_1 , and $\forall x_1 \cdots \forall x_k \varphi$ is Π_1 .
- For a class Γ of formulas, $\text{I}\Gamma$ denotes a subsystem of PA obtained by restricting ($\varphi(x)$ of) induction to Γ . E.g., $\text{PA}^- \subset \text{IOpen} \subset \text{I}\Sigma_0 \subset \text{I}\Sigma_1 \subset \text{PA}$.
- R_0 consists of the following axioms.

$$\overline{m} \neq \overline{n} \quad (\text{when } m \neq n), \quad \overline{m} + \overline{n} = \overline{m+n}, \quad \overline{m} \cdot \overline{n} = \overline{m \cdot n},$$

$$\neg(x < \overline{0}), \quad x < \overline{n+1} \leftrightarrow x = \overline{0} \vee \cdots \vee x = \overline{n}.$$
 R_0 is Σ_1 -complete, i.e., it proves all true Σ_1 sentences.
- R is obtained from R_0 by adding **R4**: $x < \overline{n} \vee x = \overline{n} \vee \overline{n} < x$.
In the following, all the systems we will consider contain R .

Theorem 2.7 ((Strong) Representation Theorem for Δ_1 Sets)

Let T be a theory containing R . For any Δ_1 set C , there exists a Σ_1 formula $\varphi(x)$ which strongly represents C , i.e., $n \in C \Rightarrow T \vdash \varphi(\bar{n})$, $n \notin C \Rightarrow T \vdash \neg\varphi(\bar{n})$.

Proof.

- For a Δ_1 set C , there exist Σ_0 formulas $\theta_1(x, y), \theta_2(x, y)$ such that

$$n \in C \Leftrightarrow \mathfrak{N} \models \exists y \theta_1(\bar{n}, y), \quad n \notin C \Leftrightarrow \mathfrak{N} \models \exists y \theta_2(\bar{n}, y).$$

Now, let $\varphi(x)$ be a Σ_1 formula $\exists y(\theta_1(\bar{n}, y) \wedge \forall z \leq y \neg\theta_2(\bar{n}, z))$. By the Σ_1 -completeness of T , $n \in C \Rightarrow T \vdash \varphi(\bar{n})$.

- To show $n \notin C \Rightarrow T \vdash \neg\varphi(\bar{n})$, take $n \notin C$. Since $\mathfrak{N} \models \exists y \theta_2(\bar{n}, y)$, there exists an $m \in \mathbb{N}$ such that $\mathfrak{N} \models \theta_2(\bar{n}, \bar{m})$, and then by the Σ_1 completeness of T , $T \vdash \theta_2(\bar{n}, \bar{m})$. Also, since $\mathfrak{N} \not\models \exists y \theta_1(\bar{n}, y)$, for all $l \in \mathbb{N}$, $\mathfrak{N} \models \neg\theta_1(\bar{n}, \bar{l})$, i.e., $T \vdash \neg\theta_1(\bar{n}, \bar{l})$. Therefore, if $\theta_1(\bar{n}, a)$ in some model of T , then a is a non-standard, i.e., (by axiom R4) $a > l$ for all $l \in \mathbb{N}$. Thus, $T \vdash \forall y(\theta_1(\bar{n}, y) \rightarrow \exists z \leq y \theta_2(\bar{n}, z))$, that is, $T \vdash \neg\varphi(\bar{n})$. \square

Theorem 2.8 (Representation Theorem for Δ_1 Function)

Suppose $T \supseteq R$. For any Δ_1 function $f(\vec{x})$, there exists a Σ_1 formula $\varphi(\vec{x}, y)$ which functionally represents $f(\vec{x}) = y$.

§3. $I\Sigma_1$ and related systems

Definition 3.1

For a formula $\varphi(x, y_1, \dots, y_k)$ in \mathcal{L}_{OR} , the following formula is called the **collection principle** or **bounding principle** of φ , denoted $(B\varphi)$:

$$\forall x < u \exists y_1 \cdots \exists y_k \varphi(x, y_1, \dots, y_k) \rightarrow \exists v \forall x < u \exists y_1 < v \cdots \exists y_k < v \varphi(x, y_1, \dots, y_k).$$

For a class Γ of formulas, we let $B\Gamma = I\Sigma_0 \cup \{(B\varphi) : \varphi \in \Gamma\}$.

Lemma 3.2

In $B\Sigma_n$ ($n \geq 1$), the class of Σ_n (Π_n) formulas are closed under adding bounded quantifiers $\forall x < t, \exists x < t$ in front of them.

Lemma 3.3

For any $n \geq 1$, $B\Sigma_n$ is a subsystem of $I\Sigma_n$.

Lemma 3.4

For any n , $I\Sigma_n$ and $I\Pi_n$ are equivalent.

Homework # 3-3

(1) The following formula is called the **least number principle** for a formula φ and denoted as $(L\varphi)$,

$$\exists x\varphi(x) \rightarrow \exists x(\varphi(x) \wedge \forall y < x \neg \varphi(y)).$$

$L\Sigma_n$ stands for $PA^- \cup \{(L\varphi) : \varphi \text{ is } \Sigma_n\}$. Then, show that $I\Sigma_n$ is equivalent to $L\Sigma_n$.

(2) For any n , show $B\Sigma_{n+1} \supset I\Sigma_n$.

It is also known that the relation $I\Sigma_{n+1} \supset B\Sigma_{n+1} \supset I\Sigma_n$ is strict.^{1 2}

¹Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Springer, 1993

²Kaye R. *Models of Peano arithmetic*, Oxford University Press, 1991.

Lemma 3.5

In $\text{I}\Sigma_1$, for a Σ_1 formula $\varphi(x)$ and a Π_1 formula $\psi(x)$, we can prove

$$\forall x(\varphi(x) \leftrightarrow \psi(x)) \rightarrow \forall u \exists m, n > 0 \forall x < u (\varphi(x) \leftrightarrow m(x+1) + 1 \text{ is a divisor of } n).$$

- In the above lemma, the triple (u, m, n) satisfying

$$\forall x < u \left(\varphi(x) \leftrightarrow m(x+1) + 1 \text{ is a divisor of } n \right)$$

is called a **u -piece code** of a Δ_1 set $\{x \mid \varphi(x)\}$.

- Similarly, if $\varphi(x, y)$ is a Δ_1 formula, then for any $u = \langle u_1, u_2 \rangle$, there exist m, n s.t.

$$\forall x < u_1 \forall y < u_2 (\varphi(x, y) \leftrightarrow m(\langle x, y \rangle + 1) + 1 \text{ is a divisor of } n),$$

where $\langle x, y \rangle = \frac{(x+y)(x+y+1)}{2} + x$.

Theorem 3.6 (Definability of primitive recursive functions)

In $\text{I}\Sigma_1$, the graph of a primitive recursive function $f(\vec{x}, y) = z$ can be represented by a Δ_1 formula $\varphi(\vec{x}, y, z)$, and the following is provable

$$\forall \vec{x} \forall y \exists! z \varphi(\vec{x}, y, z).$$

- We prove this theorem by induction on the construction of prim. rec. functions. The kernel of the proof is the construction of a predicate $\delta(u, m, n)$ such that

$$\delta(\langle u_1, u_2 \rangle, m, n) \Leftrightarrow \forall y < u_1 \exists z < u_2 f(y) = z,$$

by using a Σ_0 formula $\gamma(x, m, n)$ expressing “ $m(x + 1) + 1$ is a divisor of n .”

- The above theorem shows that adding a symbol for a primitive recursive function and its definition to $\text{I}\Sigma_1$, we obtain a conservative extension.

Lemma 3.7

The CE sets are exactly the same as the Σ_1 sets. Hence, the computable (recursive) sets are exactly the same as the Δ_1 sets.

Proof.

- Any CE relation $R(\vec{x})$ can be expressed by $\exists y S(\vec{x}, y)$ for some primitive recursive relation S . By lemma 3.6, any primitive recursive relation S can be expressed by a Σ_1 formula, and so $R(\vec{x})$ is Σ_1 .
- Conversely, a Σ_1 formula is expressed in the form $\exists y \theta(\vec{x}, y)$ with $\theta(\vec{x}, y) \in \Sigma_0$. Since a Σ_0 formula is a primitive recursive, the set defined by $\exists y \theta(\vec{x}, y)$ is CE.

§4. The first incompleteness theorem

We assume that all theories are given in the language \mathcal{L}_{OR} and contain R , so Σ_1 -complete.

Definition 4.1

A theory T is **ω -consistent** \Leftrightarrow for any formula $\varphi(x)$, if $\varphi(\bar{n})$ is provable in T for all $n \in \mathbb{N}$, then $\exists x \neg \varphi(x)$ is not provable in T . A theory is **1-consistent** \Leftrightarrow it is ω -consistent with $\varphi(x)$ restricted to Σ_0 (or Δ_1). A theory is **Σ_n -sound** if all the Σ_n theorems of T are true.

Homework # 3-2

- (1) Show that a Σ_1 -complete theory T is 1-consistent iff it is Σ_1 -sound.
- (2) Show that any ω -consistent Σ_1 -complete theory is Π_3 -sound, but may not be Σ_3 -sound.

Lemma 4.2 ((Weak) Representation Theorem for CE sets)

Suppose that a theory T contains R and 1-consistent. Then, for any CE set C , there exists a Σ_1 formula $\varphi(x)$ such that for any n , $n \in C \Leftrightarrow T \vdash \varphi(\bar{n})$.

Theorem 4.3 (Gödel's first incompleteness theorem, a naïve version)

Let T be a Σ_1 theory containing R and 1-consistent. Then T is incomplete, that is, there is a sentence σ which T cannot prove or disprove.

Proof.

- We know K is CE but not co-CE. By the weak representation theorem for CE sets, there exists a formula $\varphi(x)$ such that

$$n \in K \Leftrightarrow T \vdash \varphi(\bar{n}).$$

- On the other hand, since $\mathbb{N} - K$ is not a CE, there exists some d such that

$$d \in \mathbb{N} - K \not\vdash T \vdash \neg\varphi(\bar{d}).$$

Thus, $(d \in K \text{ and } T \vdash \neg\varphi(\bar{d}))$ or $(d \notin K \text{ and } T \not\vdash \neg\varphi(\bar{d}))$.

- In the former case, since $d \in K$ implies $T \vdash \varphi(\bar{d})$, T is inconsistent, contradicting with the 1-consistency assumption.
- In the latter case, T is incomplete because $\varphi(\bar{d})$ cannot be proved or disproved.

We prepare some useful prim. rec. functions for coding things.

First, recall that if A, B are prim. rec. n -ary relations, also are $\neg A$, $A \wedge B$, $A \vee B$.

Lemma 4.4

For a primitive recursive function $h(\vec{x})$, $\mu y < h(\vec{x}) A(\vec{x}, y)$ is primitive recursive.

Example

Let $p(x) = "(x + 1)\text{-th prime number}"$, that is ,

$$p(0) = 2, p(1) = 3, p(2) = 5, \dots$$

Then, $p(x)$ is a primitive recursive function since it is defined as follows:

$$p(0) = 2, \quad p(x + 1) = \mu y < p(x)! + 2 (p(x) < y \wedge \text{prime}(y)).$$

- A finite sequence of natural numbers (x_0, \dots, x_{n-1}) can be represented by a single natural number x (called a **sequence number**) as follows,

$$x = p(0)^{x_0+1} \cdot p(1)^{x_1+1} \cdot \dots \cdot p(n-1)^{x_{n-1}+1}.$$

- Fixing n , such a mapping from \mathbb{N}^n to \mathbb{N} is a primitive recursive function.
- Conversely, the function $c(x, i)$ taking the i -th element x_i from a sequence number x is defined as

$$x_i = c(x, i) = \mu y < x (\neg \exists z < x (p(i)^{y+2} \cdot z = x)).$$

- The length of the sequence represented by x is

$$\text{leng}(x) = \mu i < x (\neg \exists z < x (p(i) \cdot z = x)).$$

- Furthermore, we define a primitive recursive relation $\text{Seq}(x)$ to denote that a natural number x is a sequence number as follows:

$$\text{Seq}(x) \Leftrightarrow \forall i < x \forall z < x (p(i) \cdot z = x \rightarrow i \leq \text{leng}(x)).$$

Definition 4.5

Let Ω be a finite (or countably infinite) set of symbols, and an injection $\phi : \Omega \rightarrow \mathbb{N}$. For a string $s = a_0 \cdots a_{n-1}$, the following natural number $\psi(s)$ is called the **Gödel number** of s , denoted by $\ulcorner s \urcorner$.

$$\psi(s) = p(0)^{\phi(a_0)+1} \cdot p(1)^{\phi(a_1)+1} \cdot \dots \cdot p(n-1)^{\phi(a_{n-1})+1}.$$

The mapping $\ulcorner \ \urcorner$ is an injection from the set of all symbols Ω^* to \mathbb{N} .

Example

Let $\Omega = \{0, 1, +, (,)\}$, $\phi(0) = 0$, $\phi(1) = 1$, $\phi(+)$ = 3, $\phi(($) = 5 and $\phi($) = 6.

Then,

$$\ulcorner (1 + 0) + 1 \urcorner = 2^6 \cdot 3^2 \cdot 5^4 \cdot 7^1 \cdot 11^7 \cdot 13^4 \cdot 17^2$$

Problem 5

Show that $\text{Term}(x)$ expressing “ x is the Gödel number of a term” is primitive recursive.

Definition 4.6

A theory T is Σ_i (Π_i/Δ_i /**primitive recursive**, etc.) if the set of Gödel numbers of its axioms $\{\ulcorner \sigma \urcorner : \sigma \in T\}$ is Σ_i (Π_i/Δ_i /primitive recursive, etc.).

- Ordinary theories in mathematics are finite or at most primitive recursive.
- The theories of arithmetic introduced so far (PA, $\text{I}\Sigma_1$, etc.) are all primitive recursive.
- To derive the incompleteness theorem, we need to assume that a theory is CE.
- Without this condition, for example, if we take all true arithmetic formulas as axioms, we would have a complete theory, but it would not be a formal system.
- By the following theorem, a Σ_1 set of axioms can be always be replaced by a primitive recursive set.

Theorem 4.7 (Craig's lemma)

For any CE theory T , there exists a primitive recursive theory T' that proves the same theorems.

Proof. Let T be a theory defined by a Σ_1 formula $\varphi(x) \equiv \exists y\theta(x, y)$ (θ is Σ_0). That is, $\sigma \in T \Leftrightarrow \mathfrak{N} \models \varphi(\ulcorner\sigma\urcorner)$, where $\ulcorner\sigma\urcorner$ is the Gödel number of a sentence σ . Then, we define a primitive recursive theory T' as follows:

$$T' = \{\overbrace{\sigma \wedge \sigma \wedge \cdots \wedge \sigma}^{n+1 \text{ copies}} : \theta(\ulcorner\sigma\urcorner, \bar{n})\}.$$

Then, T and T' are equivalent, since $\vdash \sigma \leftrightarrow \sigma \wedge \sigma \wedge \cdots \wedge \sigma$. Thus T' is primitive recursive. □

Because Gödel numbers and their decodings are heavily used in T' , T' cannot be easily expressed in Σ_0 .

Based on Craig's lemma, a Σ_1 theory is primitive recursively axiomatizable. Then, "a finite sequence (or finite tree) P of formulas is a proof in T " can be defined in a primitive recursive way (with T as a parameter).

Definition 4.8

- Let T be a Σ_1 theory and T' its p.r. counterpart. A **proof** in T' is a finite sequence of formulas where each formula is either a logical axiom, an equality axiom, or an axiom of T' , or obtained by applying MP or quantification rules from formulas appearing before. The formula that appears at the end of a proof is a **theorem** of T .
- Now, we define the primitive recursive predicate Proof_T as follows.

$$\text{Proof}_T(\ulcorner P \urcorner, \ulcorner \sigma \urcorner) \Leftrightarrow P \text{ is a proof of formula } \sigma \text{ in } T'.$$

- By Proof_T , we also denote a Δ_1 formula expressing the above Proof_T in $\text{I}\Sigma_1$. A Σ_1 formula Bew_T is defined as

$$\text{Bew}_T(x) \equiv \exists y \text{Proof}_T(y, x).$$

The formula $\text{Bew}_T(x)$ expresses that " x is the Gödel number of a theorem of T ". "Bew" stands for the German beweisbar (provable).

Lemma 4.9 (Diagonalization lemma)

Let T be a theory containing R. For any formula $\psi(x)$ in which x is the unique free variable, there exists a sentence σ such that $T \vdash \text{“}\sigma \leftrightarrow \psi(\overline{\overline{\sigma}})\text{”}$.

Proof.

- A formula with only x as a free variable is computably enumerated as $\varphi_0(x), \varphi_1(x), \dots$, and then $f(n) = \overline{\overline{\varphi_n(\overline{n})}}$ is also a computable function. By the functional representation theorem, there exists a Σ_1 formula χ such that

$$f(m) = n \Rightarrow T \vdash \chi(\overline{m}, \overline{n}) \wedge \forall y \neq \overline{n} \neg \chi(\overline{m}, y).$$

- The formula $\exists y(\chi(x, y) \wedge \psi(y))$ must be listed as $\varphi_k(x)$ for some k . Now, let σ be $\varphi_k(\overline{k})$. Since $f(k) = \overline{\overline{\sigma}}$, $T \vdash \chi(\overline{k}, \overline{\overline{\sigma}})$. Thus, in T , $\psi(\overline{\overline{\sigma}}) \rightarrow \exists y(\chi(\overline{k}, y) \wedge \psi(y))$ ($\equiv \varphi_k(\overline{k}) \equiv \sigma$).
- On the other hand, since $T \vdash \forall y \neq \overline{\overline{\sigma}} \neg \chi(\overline{k}, y)$, in T ,

$$\neg \psi(\overline{\overline{\sigma}}) \rightarrow \forall y(\chi(\overline{k}, y) \rightarrow \neg \psi(y)) \rightarrow \neg \exists y(\chi(\overline{k}, y) \wedge \psi(y)) \quad (\equiv \neg \sigma).$$

- Therefore, $T \vdash \sigma \leftrightarrow \psi(\overline{\overline{\sigma}})$, that is, σ is a fixed point of ψ . □

Theorem 4.10 (Gödel's first incompleteness theorem, a formal version)

Let T be a 1-consistent Σ_1 theory containing R . Then, T is incomplete, that is, there is a sentence σ such that $T \not\vdash \sigma$ and $T \not\vdash \neg\sigma$.

Proof.

- By the diagonalization lemma, $\neg\text{Bew}_T(x)$ has a fixed point, that is, there exists σ such that $T \vdash \sigma \leftrightarrow \neg\text{Bew}_T(\overline{\overline{\sigma}})$.
- We will show this σ is neither provable nor disprovable in T as follows.
- Let $T \vdash \sigma$. Then $\text{Bew}_T(\overline{\overline{\sigma}})$ is true. Hence $T \vdash \text{Bew}_T(\overline{\overline{\sigma}})$ from Σ_1 completeness. Since σ is the fixed point of $\neg\text{Bew}_T(x)$, we have $T \vdash \neg\sigma$, which means that T is inconsistent.
- On the other hand, if $T \vdash \neg\sigma$, $T \vdash \text{Bew}_T(\overline{\overline{\sigma}})$ because σ is a fixed point. Here, using the 1-consistency of T , $\text{Bew}_T(\overline{\overline{\sigma}})$ is true, and so $T \vdash \sigma$, which is a contradiction. \square

The sentence σ in the above proof “asserts its own unprovability” because “ $\sigma \leftrightarrow T \not\vdash \sigma$ ” holds. This σ is called the **Gödel sentence** of T . Since $T \not\vdash \sigma$, $\mathfrak{N} \models \neg\text{Bew}_T(\overline{\overline{\sigma}})$ is true. So, the Gödel sentence is a “true Π_1 sentence.”

To weaken the assumption of incompleteness theorem, Rosser modified $\text{Bew}_T(x)$ as follows

$$\text{Bew}_T^*(x) \equiv \exists y(\text{Proof}_T(y, x) \wedge \forall z < y \neg \text{Proof}_T(z, \neg x)),$$

where $\neg x$ means the code of $\neg\varphi$ when x is the code of a formula φ .

Lemma 4.11

Let T be a Σ_1 theory containing R. Then, for any sentence σ ,

- (1) $T \vdash \sigma \Rightarrow T \vdash \text{Bew}_T^*(\overline{\neg\sigma})$,
- (2) $T \vdash \neg\sigma \Rightarrow T \vdash \neg\text{Bew}_T^*(\overline{\neg\sigma})$.

Proof. If T is inconsistent, the lemma holds trivially, so we assume T is consistent. If $T \vdash \sigma$, it is easy to see that $\text{Bew}_T^*(\overline{\neg\sigma})$ is true. Then (1) follows from Σ_1 completeness. To show (2), assume $T \vdash \neg\sigma$. There exists $n \in \mathbb{N}$ such that the following holds in \mathfrak{N}

$$\text{Proof}_T(\overline{n}, \overline{\neg\sigma}) \wedge \forall z \leq \overline{n} \neg \text{Proof}_T(z, \overline{\neg\sigma}).$$

By Σ_1 completeness, the above formula is provable in T . So, in T , $\text{Proof}_T(y, \overline{\neg\sigma}) \rightarrow y > \overline{n}$, and thus

$$\forall y(\text{Proof}_T(y, \overline{\neg\sigma}) \rightarrow \exists z < y \text{Proof}_T(z, \overline{\neg\sigma}))$$

is provable. Therefore, $T \vdash \neg\text{Bew}_T^*(\overline{\neg\sigma})$.

The fixed point σ of $\neg\text{Bew}_T^*(x)$, i.e., $T \vdash \sigma \leftrightarrow \neg\text{Bew}_T^*(\overline{\neg\sigma})$ is called a **Rosser sentence**.

Theorem 4.12 (Gödel-Rosser)

If T is a consistent Σ_1 theory containing R , then there exists a sentence σ such that $T \not\vdash \sigma$ and $T \not\vdash \neg\sigma$.

Proof.

- If $T \vdash \sigma$, then by the last lemma $T \vdash \text{Bew}_T^*(\overline{\neg\sigma})$, and so by the definition of the fixed point σ , $T \vdash \neg\sigma$, which implies that T is inconsistent.
- If $T \vdash \neg\sigma$, then by the last lemma, $T \vdash \neg\text{Bew}_T^*(\overline{\neg\sigma})$. By definition of the fixed point σ , we have $T \vdash \sigma$, which implies that T is inconsistent. \square

Two applications of the first incompleteness thm

The next theorem is a very important application of the argument of the first incompleteness theorem.

Lemma 4.13

In a consistent theory T containing R , there exists no formula $\psi(x)$ such that for any sentence σ , $T \vdash \sigma \leftrightarrow \psi(\overline{\Gamma\sigma\overline{\Gamma}})$.

Proof. If there were such a $\psi(x)$, then a fixed point σ of $\neg\psi(x)$ clearly does not satisfy the condition. \square

In the above lemma, letting T be $\text{Th}(\mathfrak{N})$, we obtain the following theorem.

Theorem 4.14 (Tarski's undefinability of truth)

There is no formula $\psi(x)$ such that $\mathfrak{N} \models \sigma \leftrightarrow \psi(\overline{\Gamma\sigma\overline{\Gamma}})$ for all sentence σ .

Lemma 4.15

For a consistent Σ_1 -complete theory T , there is no formula $\psi(x)$ s.t. for any sentence σ ,

$$\begin{aligned} (1) \quad T \vdash \sigma &\Rightarrow T \vdash \psi(\overline{\ulcorner \sigma \urcorner}), \\ (2) \quad T \not\vdash \sigma &\Rightarrow T \vdash \neg\psi(\overline{\ulcorner \sigma \urcorner}). \end{aligned}$$

Proof. Suppose there were such a $\psi(x)$, and let σ be a fixed point of $\neg\psi(x)$. Then, if $T \vdash \sigma$ then $T \vdash \neg\psi(\overline{\ulcorner \sigma \urcorner})$, which means (1) does not hold. If $T \not\vdash \sigma$ then $T \not\vdash \neg\psi(\overline{\ulcorner \sigma \urcorner})$, which means (2) does not hold. \square

Lemma 4.16

For a consistent Σ_1 -complete theory T , the set $\{\ulcorner \sigma \urcorner : T \vdash \sigma, \sigma \text{ is a sentence}\}$ is not computable.

Proof. If the set of theorems of T is computable, by the strong representation theorem, there would be such a $\psi(x)$ that satisfies the above lemma. \square

The following theorem was due to Church. Turing also obtained a similar result by expressing the halting problem as a satisfaction problem of first-order logic.

Theorem 4.17 (Undecidability of first-order logic)

The set $\{\ulcorner \sigma \urcorner : \sigma \text{ is a valid sentence in the language } \mathcal{L}_{\text{OR}}\}$ is not computable. Therefore, the satisfiability of first order logic is not decidable.

Proof.

- First note that $I\Sigma_1$ is finitely axiomatizable, because the Σ_1 -induction schema can be expressed as a single induction axiom for a universal Σ_1 -formula (a universal CE set). Or, instead of $I\Sigma_1$, you may take $Q_{<}$ or any other finitely axiomatized theory for which the first incompleteness theorem can be shown.
- Let ξ be a sentence obtained by connecting all the axioms of $I\Sigma_1$ by \wedge .
- Then, from the deduction theorem, $I\Sigma_1 \vdash \sigma \Leftrightarrow \vdash \xi \rightarrow \sigma$. If $\{\ulcorner \sigma \urcorner : \vdash \sigma\}$ is computable, $\{\ulcorner \sigma \urcorner : \vdash \xi \rightarrow \sigma\} = \{\ulcorner \sigma \urcorner : I\Sigma_1 \vdash \sigma\}$ is also computable. By the representation theorem, there exists a which contradicts with the above lemma.
- Finally, note that the satisfiability of first order logic can be expressed as $\{\ulcorner \sigma \urcorner : \not\models \neg \sigma\}$ and that if it were computable then $\{\ulcorner \sigma \urcorner : \vdash \neg \sigma\}$ would be computable.

§5. Introducing the second incompleteness theorem

 $\text{I}\Sigma_1$ and
primitive
recursive
functionsThe first
incompleteness
theoremIntroducing the
second theorem

Commentaries

- The first incompleteness theorem says that a consistent CE theory T including R is neither prove nor disprove the Gödel sentence.
- The second incompleteness theorem says that a consistent CE theory T including $\text{I}\Sigma_1$ does not prove its consistency.
- To obtain the second theorem, it is sufficient to show that the consistency implies the Gödel sentence, or equivalently the consistency implies the unprovability of the Gödel sentence.
- Thus, the main part of the proof of the second theorem is to formalize the proof of the first theorem in the system T .
- Although this requires extremely elaborate arguments, the main points are summarized as the three properties of the derivability predicate $\text{Bew}_T(x)$ as shown in the next slide.

Lemma 5.1 (Hilbert-Bernays-Löb's derivability condition)

Let T be a consistent CE theory containing $\text{I}\Sigma_1$. For any φ, ψ ,

D1. $T \vdash \varphi \Rightarrow T \vdash \text{Bew}_T(\overline{\Gamma\varphi\overline{}})$.

D2. $T \vdash \text{Bew}_T(\overline{\Gamma\varphi\overline{}}) \wedge \text{Bew}_T(\overline{\Gamma\varphi \rightarrow \psi\overline{}}) \rightarrow \text{Bew}_T(\overline{\Gamma\psi\overline{}})$.

D3. $T \vdash \text{Bew}_T(\overline{\Gamma\varphi\overline{}}) \rightarrow \text{Bew}_T(\overline{\Gamma\text{Bew}_T(\overline{\Gamma\varphi\overline{}})\overline{}})$.

Proof.

- D1 is obtained from the Σ_1 completeness of T , since $\text{Bew}_T(\overline{\Gamma\varphi\overline{}})$ is a Σ_1 formula.
- For D2, it is clear that the proof of ψ is obtained by applying MP to the proof of φ and the proof of $\varphi \rightarrow \psi$.
- D3 formalizes D1 in T . This is the most difficult, since we can not find a simple machinery to transform a proof of φ in T to a proof of $\text{Bew}_T(\overline{\Gamma\varphi\overline{}})$. There are several known ways to deal with this problem, but below we will briefly explain how to deal with the representability of primitive recursive functions within the system.

- Since the function from a number n to the Gödel number of its numeral $\ulcorner \bar{n} \urcorner$ is primitive recursive, we denote the function by \dot{x} .
- For an expression $\varphi(x)$, $\varphi(\dot{y})$ denotes the expression obtained by substituting the term with the Gödel number \dot{y} to every free occurrence of the variable x . If the value of y is a standard natural number n , this is nothing but a substitution of the numeral \bar{n} , but $\varphi(\dot{y})$ is just an expression with the variable y , which can be formulated within Bew_T .
- With this notation, our goal is to prove

$$T \vdash \text{Proof}_T(x, y) \rightarrow \text{Bew}_T(\overline{\ulcorner \text{Proof}_T(\dot{x}, \dot{y}) \urcorner}). \quad (1)$$

- In general, we prove that for any primitive recursive function f ,

$$T \vdash f(x_1, \dots, x_k) = y \rightarrow \text{Bew}_T(\overline{\ulcorner f(\dot{x}_1, \dots, \dot{x}_k) = \dot{y} \urcorner}). \quad (2)$$

- The above formula can be proved by meta-induction on the construction of the primitive recursive function f .
- As a example, we will prove for addition $x + y = z$, the above formula (2) holds.

- By Σ_1 induction on variable y (assuming other variables are constants), we prove that

$$x + y = z \rightarrow \text{Bew}_T(\overline{\ulcorner \dot{x} + \dot{y} = \dot{z} \urcorner}). \quad (3)$$

- First, if $y = 0$, then $x + 0 = z$ and so $x = z$. By A3 of PA, $\text{Bew}_T(\overline{\ulcorner \dot{x} + 0 = \dot{x} \urcorner})$. Thus

$$x + 0 = z \rightarrow \text{Bew}_T(\overline{\ulcorner \dot{x} + 0 = \dot{z} \urcorner}).$$

- Next assuming $x + y = w \rightarrow \text{Bew}_T(\overline{\ulcorner \dot{x} + \dot{y} = \dot{w} \urcorner})$, we want to show

$$x + (y + 1) = z \rightarrow \text{Bew}_T(\overline{\ulcorner \dot{x} + (\dot{y} + 1) = \dot{z} \urcorner}).$$

- Suppose $x + (y + 1) = z$. Let $w = x + y$. Then, we have $z = w + 1$, since $x + (y + 1) = (x + y) + 1$ by A4. Hence, By the definition of \dot{x} , $\text{Bew}_T(\overline{\ulcorner \dot{z} = \dot{w} + 1 \urcorner})$.
- From $\text{Bew}_T(\overline{\ulcorner \dot{x} + \dot{y} = \dot{w} \urcorner})$, by using A4 in Bew_T , $\text{Bew}_T(\overline{\ulcorner \dot{x} + (\dot{y} + 1) = \dot{w} + 1 \urcorner})$. Then from $\text{Bew}_T(\overline{\ulcorner \dot{z} = \dot{w} + 1 \urcorner})$, we obtain $\text{Bew}_T(\overline{\ulcorner \dot{x} + (\dot{y} + 1) = \dot{z} \urcorner})$.
- Thus, we have shown (3) by $\text{I}\Sigma_1$.
- As for other p.r. functions, their defining formulas are given as axioms in the theory T, so (2) can be proved using a similar argument.

- To prove D3, assume $\text{Bew}_T(\overline{\ulcorner\varphi\urcorner})$ in addition to T. Then, there is a numeral c that satisfies $\text{Proof}_T(c, \overline{\ulcorner\varphi\urcorner})$.
- Now by (1), we have $\text{Bew}_T(\overline{\ulcorner\text{Proof}_T(\dot{c}, \ulcorner\dot{\varphi}\urcorner)\urcorner})$. Here, $\ulcorner\dot{\varphi}\urcorner$ is a standard natural number, so it is nothing but $\overline{\ulcorner\varphi\urcorner}$.
- Since $T \vdash \text{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}) \rightarrow \exists x \text{Proof}_T(x, \overline{\ulcorner\varphi\urcorner})$ can be deduced from a quantification axiom of first-order logic, we have

$$T \vdash \text{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}) \rightarrow \text{Bew}_T(\overline{\ulcorner\varphi\urcorner}).$$

- Then, by D1,

$$T \vdash \text{Bew}_T(\overline{\ulcorner\text{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}) \rightarrow \text{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner}).$$

By D2,

$$T \vdash \text{Bew}_T(\overline{\ulcorner\text{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner})\urcorner}) \rightarrow \text{Bew}_T(\overline{\ulcorner\text{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner}).$$

- Finally, $\text{Bew}_T(\overline{\ulcorner\text{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner})$ is obtained with the first assumption by MP. Thus, D3 is proven.

□

In the following, let π_G denote a Gödel sentence in the proof of the first incompleteness theorem. That is,

$$T \vdash \pi_G \leftrightarrow \neg \text{Bew}_T(\overline{\ulcorner \pi_G \urcorner}).$$

By $\text{Con}(T)$, we denote the sentence meaning “ T is consistent”, formally defined as

$$\text{Con}(T) \equiv \neg \text{Bew}_T(\overline{\ulcorner 0 = 1 \urcorner}).$$

Then we have the following.

Lemma 5.2

$$T \vdash \text{Con}(T) \leftrightarrow \pi_G.$$

Proof. • To show $\pi_G \rightarrow \text{Con}(T)$.

Obviously, $T \vdash 0 = 1 \rightarrow \pi_G$. So, by D1 and D2,

$$T \vdash \text{Bew}_T(\overline{\ulcorner 0 = 1 \urcorner}) \rightarrow \text{Bew}_T(\overline{\ulcorner \pi_G \urcorner}).$$

Taking the contraposition, we have $T \vdash \pi_G \rightarrow \text{Con}(T)$.

- To show $\text{Con}(T) \rightarrow \pi_G$.

First, from $T \vdash \pi_G \leftrightarrow \neg \text{Bew}_T(\overline{\neg \pi_G})$ and D1,

$$T \vdash \text{Bew}_T(\overline{\text{Bew}_T(\overline{\neg \pi_G})} \rightarrow \neg \pi_G).$$

Using D2, we have

$$T \vdash \text{Bew}_T(\overline{\text{Bew}_T(\overline{\text{Bew}_T(\overline{\neg \pi_G})})} \rightarrow \text{Bew}_T(\overline{\neg \pi_G})).$$

Combining this with D3: $T \vdash \text{Bew}_T(\overline{\neg \pi_G}) \rightarrow \text{Bew}_T(\overline{\text{Bew}_T(\overline{\neg \pi_G})})$, we obtain

$$T \vdash \text{Bew}_T(\overline{\neg \pi_G}) \rightarrow \text{Bew}_T(\overline{\neg \pi_G}).$$

Then, by using $T \vdash \pi_G \rightarrow (\neg \pi_G \rightarrow 0 = 1)$ and D2, we get

$$T \vdash \text{Bew}_T(\overline{\neg \pi_G}) \rightarrow \text{Bew}_T(\overline{0 = 1}).$$

Taking the contraposition,

$$T \vdash \neg \text{Bew}_T(\overline{0 = 1}) \rightarrow \neg \text{Bew}_T(\overline{\neg \pi_G}).$$

That is, $T \vdash \text{Con}(T) \rightarrow \pi_G$.

Theorem 5.3 (Gödel's second incompleteness theorem)

Let T be a consistent CE theory, which contains $\text{I}\Sigma_1$. Then $T \not\vdash \text{Con}(T)$.
In other words, T cannot prove its own consistency $\text{Con}(T)$.

Proof

By the proof of the first incompleteness theorem, $T \not\vdash \pi_G$.

By the above lemma, $T \vdash \text{Con}(T) \leftrightarrow \pi_G$. So, $T \not\vdash \text{Con}(T)$. □

Remark

- The first incompleteness theorem is a negative result in the sense that it shows the limit of provability, whereas the second incompleteness theorem shows that the concrete proposition $\text{Con}(T)$ is not provable in T , which provides a positive tool from an application perspective.
- In mathematical logic, the second incompleteness theorem is often used to separate two axiomatic theories by showing the consistency of one over the other. E.g. $\text{I}\Sigma_1$ is a proper subsystem of PA, since the consistency of the former can be proved in the latter.

Homework # 3-4

- (1) Show that there is a consistent theory T that proves its own contradiction $\neg\text{Con}(T)$.
- (2) Let $\text{Bew}_T^\#(x) \equiv (\text{Bew}_T(x) \wedge x \neq \overline{\overline{0=1}})$. For any true proposition σ ,

$$\text{Bew}_T^\#(\overline{\overline{\sigma}}) \leftrightarrow \text{Bew}_T(\overline{\overline{\sigma}})$$

and

$$T \vdash \neg\text{Bew}_T^\#(\overline{\overline{0=1}}).$$

Does it contradict with the second incompleteness theorem?

- As a variant of the Gödel sentence, a sentence meaning “this sentence is provable” is known as a **Henkin sentence**. That is, H is a Henkin sentence if

$$H \leftrightarrow \text{Bew}_T(\overline{\overline{H}}).$$

If H is provable and true, then both sides are true and there is no problem.

On the other hand, if H is false and unprovable, both sides are also equivalent. So, there does not seem to be any clue to determine whether or not H is provable or true. Yet, we can show it is actually provable.

- To this end, first let C denote the sentence “this sentence is consistent with T ”, i.e., $C \leftrightarrow \neg \text{Bew}_T(\overline{\overline{\neg C}})$.
- Since the theory $T+C$ proves its own consistency, it is inconsistent by the second incompleteness theorem. Thus, T proves $\neg C$.
- On the other hand, since $\neg C \leftrightarrow \text{Bew}_T(\overline{\overline{\neg C}})$, $\neg C$ is the same as H , and therefore H is provable.

The above fact can be also stated as follows.

Theorem 5.4 (Löb's theorem)

Let T be a consistent Σ_1 theory containing $\text{I}\Sigma_1$. If T proves “if T proves σ , then σ ”, then T proves σ .

Proof.

Suppose that T proves that “if T proves σ , then σ ”, which means that “If $\neg\sigma$, then T does not prove σ , that is, $T + \neg\sigma$ is consistent.” That is, since $T + \neg\sigma$ proves the consistency of $T + \neg\sigma$, by the second incompleteness theorem, $T + \neg\sigma$ is inconsistent. Therefore, T proves σ . □

The Henkin sentence H satisfies that T proves “if T proves H , then H ”. So by the theorem, T proves H .

A paradoxical fact derived from this theorem is that any proposition σ can be proven by assuming that there is a proof of σ .

- For simplicity, let T be PA. We also identify a formula $\varphi(x)$ with the set $\{n : \varphi(n)\}$.
- In T , we can prove a countable version of the completeness theorem of first-order logic. A countable model M can be treated as its coded diagram, i.e., the set of the Gödel numbers of \mathcal{L}_M -sentences true in M . The arithmetized completeness theorem says that if T' is consistent then there exists (a formula expressing the diagram of) a model of T' .
- Now, we going to prove $\text{Con}(T) \rightarrow \pi_G$ in T . By the completeness theorem, it is sufficient to show that any model M of $T + \text{Con}(T)$ satisfies π_G . First, note that π_G is equivalent to $\neg \text{Bew}_T(\ulcorner \pi_G \urcorner)$, which is also equivalent to $\text{Con}(T + \neg \pi_G)$. Since M satisfies $\text{Con}(T)$, we can make a model M_1 of T over M . So, if M_1 satisfies $\neg \pi_G$, then M shows $\text{Con}(T + \neg \pi_G)$. If M_1 satisfies π_G , M also satisfies π_G since π_G is Π_1 and M is a submodel of M_1 . (This proof is due to Kikuchi-Tanaka.)

- D. Hilbert and P. Bernays, *Grundlagen der Mathematik I-II*, Springer-Verlag, 1934-1939, 1968-1970 (2nd ed.). This gives the first complete proof of the second incompleteness theorem by analyzing the provability predicate.
- R.M. Smullyan, *Theory of Formal Systems*, revised edition, Princeton Univ. 1961. A classic masterpiece introducing recursive inseparability, etc.
- *Handbook of Mathematical Logic* (1977), edited by J. Barwise
Smoryński's chapter on incompleteness theorems includes various unpublished results (particularly by Kreisel) and a wide range of mathematical viewpoints.
- P. Lindström, *Aspects of Incompleteness*, *Lecture Notes in Logic* 10, Second edition, Assoc. for Symbolic Logic, A K Peters, 2003.
A technically advanced book. It has detailed information on Pour-El and Kripke's theorem (1967) that between any two recursive theories (including PA) there exists a recursive isomorphism that preserves propositional connectives and provability.

- R.M. Solovay (1976) studied modal propositional logic GL with $\text{Bew}_T(x)$ as modality \Box , which is described by
 - (1) $\vdash A \Rightarrow \vdash \Box A$,
 - (2) $(\Box A \wedge \Box(A \rightarrow B)) \rightarrow \Box B$,
 - (3) $\Box A \rightarrow \Box \Box A$,
 - (4) $\Box(\Box A \rightarrow A) \rightarrow \Box A$
- The following two books are good on this topic.

Smoryński, Self-Reference and Modal Logic, Springer 1977.

G. Boolos, The Logic of Provability, Cambridge 1993.

The following are excellent introductory books.

- T. Franzen, Gödel's Theorem: An Incomplete Guide to Its Use and Abuse(2005).
On the use and misuse of the incompleteness theorem as a broader understanding of Godel's theorem. A Japanese translation (with added explanations) by Tanaka (2011).
- P. Smith, Gödel's Without (Too Many) Tears, Second Edition 2022.
<https://www.logicmatters.net/resources/pdfs/GWT2edn.pdf>
Easy to read. The best reference to this lecture.

Thank you for your attention!