Logic and Foundation I
Part 2. First-order logic

Kazuyuki Tanaka

BIMSA

December 23, 2023

北京雁栖湖应用数学研究院
Beijing Institute of Mathematical Sciences and Applications

Logic and Foundations I

- **Part 1. Equational theory**
- **Part 2. First order theory**
- **Part 3. Model theory**
- **Part 4. First order arithmetic and incompleteness theorems**

Part 2. Schedule

- Dec. 07, (1) Peano arithmetic and representation theorems
- Dec. 14, (2) The first incompleteness theorem
- Dec. 21, (3) The second incompleteness theorem
- Dec. 28, (4) Presburger arithmetic

Introduction

- Gödel's first incompleteness theorem shows the existence of statements that cannot be proven or disproved for an axiomatic system T such as Peano arithmetic.
- The second incompleteness theorem asserts that a statement with the specific meaning "T is consistent" cannot be proven with T.
- The second incompleteness theorem is obtained by formalizing the proof of the first incompleteness theorem within its own system T.
- For the first theorem, we arithmetized several metamathemacal concepts such as proofs and theorems by using Gödel numbers. For the second theorem, we further need to analyze more general concepts such as primitive recursiveness and Σ_1-completeness, which are used in the proof of the first theorem.
- In the last lectures, we studied two proofs of the first theorem. The second one is more robust, or suitable for elevating it to the second theorem.
- In this lecture, we assume $\mathrm{I}\Sigma_1$ from the beginning.
- We prove the second incompleteness theorem by using the derivability conditions.

Recap

- **Peano arithmetic** PA is a first-order theory of natural numbers in the language $\mathcal{L}_{\mathrm{OR}} = \{+, \cdot, 0, 1, <\}$, consisting of axioms for arithmetical operations and Induction: $\varphi(0) \land \forall x(\varphi(x) \to \varphi(x+1)) \to \forall x \varphi(x)$.
- The formulas in $\mathcal{L}_{\mathrm{OR}}$ 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, $\mathsf{I}\Gamma$ denotes a subsystem of PA obtained by restricting ($\varphi(x)$ of) induction to Γ.
- R, $\mathsf{Q}_<$ and PA^- are very weak subsystems with no induction. We have

  $$\mathsf{R} \subset \mathsf{Q}_< \subset \mathsf{PA}^- \subset \mathsf{IOpen} \subset \mathsf{I}\Sigma_0 \subset \mathsf{I}\Sigma_1 \subset \mathsf{PA}.$$
- The following is the **collection principle** or **bounding principle** of φ, denoted (Bφ):

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

$$\mathbf{B\Gamma} := \mathsf{I}\Sigma_0 \cup \{(\mathsf{B}\varphi) : \varphi \in \Gamma\}. \quad \mathsf{I}\Sigma_{n+1} \supset \mathsf{B}\Sigma_{n+1} \supset \mathsf{I}\Sigma_n.$$

- In $\mathsf{B}\Sigma_n (n \geq 1)$, Σ_n and Π_n are closed under bounded quantifiers.

Theorem (Σ_1-completeness of R)

R proves all true Σ_1 sentences. Therefore, $\mathsf{Q}_<$, PA^-, IOpen, etc. are all **Σ_1-complete**.

Definition

Theory T is **1-consistent** if, for any Σ_1 sentence σ, $T \vdash \sigma \Rightarrow \mathfrak{N} \models \sigma$.

- If a theory holds in the standard model $\mathfrak{N}$, it is 1-consistent, and indeed ω-consistent (i.e., for any formula $\varphi(x)$, if $T \vdash \varphi(\bar{n})$ for all $n \in \mathbb{N}$ then $T \nvdash \exists x \neg\varphi(x)$).

Theorem ((Weak) Representation Theorem for CE sets, repeated)

Let T be Σ_1-complete and 1-consistent. For a CE set C, there exists a Σ_1 formula $\varphi(x)$ s.t.

$$n \in C \quad \Leftrightarrow \quad T \vdash \varphi(\overline{n}).$$

Theorem ((Strong) Representation Theorem for Computable Sets, repeated)

Let T be Σ_1-complete. For a computable set C, there exists a Σ_1 formula $\varphi(x)$ such that

$$n \in C \Rightarrow T \vdash \varphi(\overline{n}), \quad n \notin C \Rightarrow T \vdash \neg\varphi(\overline{n}).$$

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

Let T be a Σ_1-complete and 1-consistent Σ_1 theory. 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

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

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

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

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

□

Formalizing metamathematics

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

Lemma

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

- $p(x) =$ "$(x+1)$-th prime number " is a primitive recursive function defined as follows.

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

- A finite sequence $(x_0, \ldots, x_{n-1})$ can be represented by a single number x as follows,

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

- For a natural number x, the function $c(x, i)$ takes the ith element x_i from x,

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

- The length of the sequence represented by x is

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

- Finally, $\mathrm{Seq}(x)$ denotes that x codes a sequence as follows:

$$\mathrm{Seq}(x) \Leftrightarrow \forall i < x \forall z < x \ (p(i) \cdot z = x \to i \le \mathrm{leng}(x)).$$

Gödel numbers

Definition

Let Ω be a finite (or countably infinite) set of symbols, and an injection $\phi : \Omega \to \mathbb{N}$. 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 $\ulcorner s \urcorner$.

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

The mapping $\ulcorner \quad \urcorner$ 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 $\mathrm{Term}(x)$ expressing "x is the Gödel number of a term" is primitive recursive.

Definition

A theory T is CE ($= \Sigma_1$) or primitive recursive, if the set of Gödel numbers of its axioms $\{\ulcorner\sigma\urcorner : \sigma \in T\}$ is Σ_1 or primitive recursive, respectively.

- Ordinary theories of arithmetic (PA, IΣ_1, etc.) are all primitive recursive.
- From the following theorem, a Σ_1 set of axioms can be always be replaced by a primitive recursive set.

Theorem (Craig's lemma)

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

Proof. Let T be a theory defined by Σ_1 formula $\varphi(x) \equiv \exists y\theta(x, y)$ (θ is Σ_0). Then, we define a primitive recursive theory T' as follows:

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

□

- Because Gödel numbers are heavily used in T', T' cannot be easily expressed in Σ_0.
- From now on, a Σ_1 theory T is automatically transformed and identified with its p.r. counterpart T'.

Definition

- Let T be a Σ_1 (= prim. rec.) theory.
- A proof in T is a finite sequence of formulas where each formula is either a logical axiom or an axiom of T, or obtained by applying MP or quantification rules from formulas appearing before in the sequence.
- The formula that appears at the end of the proof is the theorem of T.
- We can define a primitive recursive predicate Proof_T such as

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

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

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

  $\mathrm{Bew}_T(x)$ expresses that "x is the Gödel number of a theorem of T".

"Bew" stands for the German beweisbar (provable).

Theorem (Representation Theorem for Computable Function)

Let T be Σ_1-complete. For any computable function $f(\vec{x})$, there exists a Σ_1 formula $\varphi(\vec{x}, y)$ which represents $f(\vec{x}) = y$ and satisfies, for all natural numbers $m_1, \ldots, m_l$,

$$T \vdash \forall y \forall y' (\varphi(\overline{m_1}, \ldots, \overline{m_l}, y) \land \varphi(\overline{m_1}, \ldots, \overline{m_l}, y') \to y = y').$$

Proof. For simplicity, we assume that $l = 1$. Suppose $f(x) = y$ is represented by a Σ_1 formula $\varphi(x, y) \equiv \exists z \theta(x, y, z)$ with $\theta(x, y, z) \in \Sigma_0$. We define a Σ_0 formula $\psi(x, y, z)$ as

$$\theta(x, y, z) \land \forall y', z' \leq y + z (\theta(x, y', z') \to y + z \leq y' + z').$$

Then, $\exists z \psi(x, y, z)$ also represents $f(x) = y$. To show, the functional property of this representation. Take any m and let $n = f(m)$. Then the minimal k such that $\theta(\overline{m}, \overline{n}, \overline{k})$ satifies $\psi(\overline{m}, \overline{n}, \overline{k})$. By the definition, no other y, z satisfy ψ. So, we are done. □

Corollary (Strong Representation for primitive recursive functions)

For any primitive recursive function f, there is a Δ_1 formula $\chi(x, y)$ such that

$$f(m) = n \Rightarrow \mathsf{I}\Sigma_1 \vdash \chi(\overline{m}, \overline{n}) \quad \text{and} \quad \mathsf{I}\Sigma_1 \vdash \forall x \exists! y \chi(x, y).$$

Lemma (Diagonalization lemma)

Let T be Σ_1-complete. For any formula $\psi(x)$ in which x is the unique free variable, there exists a sentence σ such that $T \vdash$ "$\sigma \leftrightarrow \psi(\overline{\ulcorner \sigma \urcorner})$" .

Proof.

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

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

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

$$\neg \psi(\overline{\ulcorner \sigma \urcorner}) \to \forall y(\chi(\overline{k}, y) \to \neg \psi(y)) \to \neg \exists y(\chi(\overline{k}, y) \land \psi(y)) \ (\equiv \neg \sigma).$$

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

The First Incompleteness Theorem

Theorem (Gödel's first incompleteness theorem)

Any 1-consistent CE theory T including $\mathrm{I}\Sigma_1$ is incomplete.

Proof.

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

The sentence σ in the above proof "asserts its own unprovability" because "$\sigma \Leftrightarrow T \nvdash \sigma$" holds. This σ is called the **Gödel sentence** of T. Since $T \nvdash \sigma$, $\mathfrak{N} \models \neg\mathrm{Bew}_T(\overline{\ulcorner\sigma\urcorner})$ is true. So, the Gödel sentence is a "true Π_1 sentence."

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

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

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

Lemma

Let T be a Σ_1-complete Σ_1 theory. Then, for any sentence σ,
(1) $T \vdash \sigma \Rightarrow T \vdash \mathrm{Bew}^*_T(\overline{\ulcorner\sigma\urcorner})$,
(2) $T \vdash \neg\sigma \Rightarrow T \vdash \neg\mathrm{Bew}^*_T(\overline{\ulcorner\sigma\urcorner})$.

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 $\mathrm{Bew}^*_T(\overline{\ulcorner\sigma\urcorner})$ 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 $\mathfrak{N}$

$$\mathrm{Proof}_T(\overline{n}, \overline{\ulcorner\neg\sigma\urcorner}) \land \forall z \le \overline{n} \neg\mathrm{Proof}_T(z, \overline{\ulcorner\sigma\urcorner}).$$

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

$$\forall y(\mathrm{Proof}_T(y, \overline{\ulcorner\sigma\urcorner}) \to \exists z < y \mathrm{Proof}_T(z, \overline{\ulcorner\neg\sigma\urcorner}))$$

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

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

Theorem (Gödel-Rosser)

If T is a consistent Σ_1-complete Σ_1 theory, then there exists a sentence σ such that $T \nvdash \sigma$ and $T \nvdash \neg\sigma$.

Proof.

- If $T \vdash \sigma$, then by the last lemma $T \vdash \mathrm{Bew}^*_T(\overline{\ulcorner\sigma\urcorner})$, 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, $T \vdash \neg\mathrm{Bew}^*_T(\overline{\ulcorner\sigma\urcorner})$. By definition of the fixed point σ, we have $T \vdash \sigma$, which implies that T is inconsitent. □

Two applications of the first incomp. theorem

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

Lemma

In a consistent Σ_1-complete theory T, there exists no formula $\psi(x)$ such that for any sentence σ, $T \vdash \sigma \leftrightarrow \psi(\overline{\ulcorner \sigma \urcorner})$.

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

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

Theorem (Tarski's undefinability of truth)

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

Lemma

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

$$
\begin{aligned}
(1)\ T \vdash \sigma \quad &\Rightarrow \quad T \vdash \psi(\overline{\ulcorner \sigma \urcorner}),\\
(2)\ T \nvdash \sigma \quad &\Rightarrow \quad T \vdash \neg\psi(\overline{\ulcorner \sigma \urcorner}).
\end{aligned}
$$

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 $T \nvdash \sigma$ then $T \nvdash \neg\psi(\overline{\ulcorner \sigma \urcorner})$, which means (2) does not hold. □

Lemma

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. □

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 (Undecidability of first-order logic)

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

Proof.

- First note that $\mathrm{I}\Sigma_1$ 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 $\mathrm{I}\Sigma_1$, you may take $\mathrm{Q}_<$ 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 $\mathrm{I}\Sigma_1$ by $\wedge$.
- Then, from the deduction theorem, $\mathrm{I}\Sigma_1 \vdash \sigma \Leftrightarrow \vdash \xi \to \sigma$. If $\{\ulcorner\sigma\urcorner : \vdash \sigma\}$ is computable, $\{\ulcorner\sigma\urcorner : \vdash \xi \to \sigma\} = \{\ulcorner\sigma\urcorner : \mathrm{I}\Sigma_1 \vdash \sigma\}$ 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.

□

Intoducing the second incompleteness theorem

- The first incompleteness theorem says that a consitent 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 $\mathsf{I}\Sigma_1$ 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 $\mathrm{Bew}_T(x)$ as shown in the next slide.

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

Let T be a consistent CE theory containing IΣ_1. For any φ, ψ,
D1. $T \vdash \varphi \Rightarrow T \vdash \mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner})$.
D2. $T \vdash \mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner}) \wedge \mathrm{Bew}_T(\overline{\ulcorner \varphi \to \psi \urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner \psi \urcorner})$.
D3. $T \vdash \mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner \mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner}) \urcorner})$.

Proof.

- D1 is obtained from the Σ_1 completeness of T, since $\mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner})$ 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 $\varphi \to \psi$.
- 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 $\mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner})$. 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 $\ulcorner \overline{n} \urcorner$ 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 $\overline{n}$, 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

$$T \vdash \mathrm{Proof}_T(x, y) \to \mathrm{Bew}_T(\overline{\ulcorner \mathrm{Proof}_T(\dot{x}, \dot{y}) \urcorner}). \tag{1}$$

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

$$T \vdash f(x_1, \ldots, x_k) = y \to \mathrm{Bew}_T(\overline{\ulcorner f(\dot{x_1}, \ldots, \dot{x_k}) = \dot{y}) \urcorner}). \tag{2}$$

- The above formula can be proved by meta-induction on the construction of the primitive recursive function f.
- As a exmaple, 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

$$x + y = z \to \mathrm{Bew}_T(\ulcorner \overline{\dot{x} + \dot{y} = \dot{z}} \urcorner). \tag{3}$$

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

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

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

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

- 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 $\dot{x}$, $\mathrm{Bew}_T(\ulcorner \overline{\dot{z} = \dot{w} + 1} \urcorner)$.
- From $\mathrm{Bew}_T(\ulcorner \overline{\dot{x} + \dot{y} = \dot{w}} \urcorner)$, by using A4 in Bew_T, $\mathrm{Bew}_T(\ulcorner \overline{\dot{x} + (\dot{y} + 1) = \dot{w} + 1)} \urcorner)$. Then from $\mathrm{Bew}_T(\ulcorner \overline{\dot{z} = \dot{w} + 1} \urcorner)$, we obtain $\mathrm{Bew}_T(\ulcorner \overline{\dot{x} + (\dot{y} + 1) = \dot{z})} \urcorner)$.
- Thus, we have shown (3) by IΣ_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 $\mathrm{Bew}_T(\overline{\ulcorner\varphi\urcorner})$ in addition to T. Then, there is a numeral c that satisfies $\mathrm{Proof}_T(c, \overline{\ulcorner\varphi\urcorner})$.
- Now by (1), we have $\mathrm{Bew}_T(\overline{\ulcorner\mathrm{Proof}_T(\dot{c}, \ulcorner\dot{\varphi}\urcorner)\urcorner})$. Here, $\ulcorner\dot{\varphi}\urcorner$ is a standard natural number, so it is nothing but $\overline{\ulcorner\varphi\urcorner}$.
- Since $T \vdash \mathrm{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}) \to \exists x \mathrm{Proof}_T(x, \overline{\ulcorner\varphi\urcorner})$ can be deduced from a quantification axiom of first-order logic, we have

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

- Then, by D1,

$$T \vdash \mathrm{Bew}_T\left(\overline{\ulcorner\mathrm{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner}\right).$$

  By D2,

$$T \vdash \mathrm{Bew}_T\left(\overline{\ulcorner\mathrm{Proof}_T(\dot{c}, \overline{\ulcorner\varphi\urcorner}))\urcorner}\right) \to \mathrm{Bew}_T\left(\overline{\ulcorner\mathrm{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner}\right).$$

- Finally, $\mathrm{Bew}_T(\overline{\ulcorner\mathrm{Bew}_T(\overline{\ulcorner\varphi\urcorner})\urcorner})$ 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,

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

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

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

Then we have the following.

Lemma

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

Proof. • To show $\pi_G \to \mathrm{Con}(T)$.

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

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

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

- To show $\mathrm{Con}(T) \to \pi_G$.

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

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

Using D2, we have

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

Combining this with D3: $T \vdash \mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner \mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}) \urcorner})$, we obtain

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

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

$$T \vdash \mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner 0 = 1 \urcorner})$$

Taking the contraposition,

$$T \vdash \neg\mathrm{Bew}_T(\overline{\ulcorner 0 = 1 \urcorner}) \to \neg\mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}).$$

That is, $T \vdash \mathrm{Con}(T) \to \pi_G$. □

Theorem (Gödel's second incompleteness theorem)

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

Proof
By the proof of the first incompleteness theorem, $T \nvdash \pi_G$.
By the above lemma, $T \vdash \mathrm{Con}(T) \leftrightarrow \pi_G$. So, $T \nvdash \mathrm{Con}(T)$. □

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 $\mathrm{Con}(T)$ 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. $\mathrm{I}\Sigma_1$ is a proper subsystem of PA, since the consistency of the former can be proved in the latter.

Homework

(1) Show that there is a consistent theory T that proves its own contradiction $\neg\mathrm{Con}(T)$.

(2) Let $\mathrm{Bew}^{\#}_T(x) \equiv (\mathrm{Bew}_T(x) \wedge x \neq \overline{\ulcorner 0 = 1 \urcorner})$. For any true proposition σ,

$$\mathrm{Bew}^{\#}_T(\overline{\ulcorner \sigma \urcorner}) \leftrightarrow \mathrm{Bew}_T(\overline{\ulcorner \sigma \urcorner})$$

and

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

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

$$\mathrm{H} \leftrightarrow \mathrm{Bew}_T(\overline{\ulcorner \mathrm{H} \urcorner}).$$

  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., $\mathrm{C} \leftrightarrow \neg\mathrm{Bew}_T(\overline{\ulcorner \neg\mathrm{C} \urcorner})$.
- Since the theory T+C proves its own consistency, it is inconsistent by the second incompleteness theorem. Thus, T proves $\neg\mathrm{C}$.
- On the other hand, since $\neg\mathrm{C} \leftrightarrow \mathrm{Bew}_T(\overline{\ulcorner \neg\mathrm{C} \urcorner})$, $\neg\mathrm{C}$ is the same as H, and therefore H is provable.

The above fact can be also stated as follows.

Theorem (Löb's theorem)

Let T be a consistent Σ_1 theory containing $\mathrm{I}\Sigma_1$. 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.

Alternative proof of D3

- 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 $\mathrm{Con}(T) \to \pi_G$ in T. By the completeness theorem, it is sufficient to show that any model M of $T + \mathrm{Con}(T)$ satisfies π_G. First, note that π_G is equivalet to $\neg\mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner})$, which is also equivalet to $\mathrm{Con}(T + \neg\pi_G)$. Since M satisfies $\mathrm{Con}(T)$, we can make a model M_1 of T over M. So, if M_1 satisfies $\neg\pi_G$, then M shows $\mathrm{Con}(T + \neg\pi_G)$. 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.)

Some commentaries on Gödel's theorem

- 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 $\mathrm{Bew}_T(x)$ as modality $\Box$, which is described by

  (1) $\vdash A \Rightarrow \vdash \Box A$,

  (2) $(\Box A \wedge \Box(A \to B)) \to \Box B$,

  (3) $\Box A \to \Box\Box A$,

  (4) $\Box(\Box A \to A) \to \Box A$

- 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.

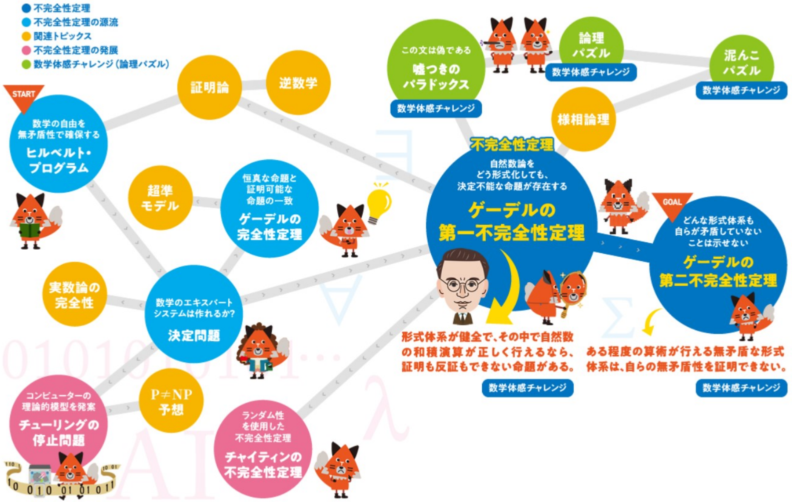
- T. Franzen, Gödel's Theorem: An Incomplete Guide to Its Use and Abuse(2005).

  On the use and misuse of the incompleteness theorem as a broader understanding of Godel's theorem. A Janpanse 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.

Math classroom: https://www.asahi.com/ads/math2022/



Summary

Theorem (Gödel's first incompleteness theorem)

Any Σ_1-complete and 1-consistent CE theory is incomplete, that is, there is a sentence that cannot be proved or disproved.

Theorem (Gödel-Rosser incompleteness theorem)

Any Σ_1-complete and consistent CE theory is incomplete.

Theorem (Gödel's second incompleteness theorem)

Let T be a consistent CE theory, which contains $\mathrm{I}\Sigma_1$. Then $\mathrm{Con}(T)$ cannot be proved in T.

Appendix

- Since Gödel, many researchers were looking for a proposition that has a natural mathematical meaning and is independent of Peano arithmetic, etc.

- Paris and Harrington found the first example in 1977. This is a slight modification of Ramsey's theorem in finite form.

Jeff Paris

Leo Harrington

- Following their findings, Kirby and Paris (1982) showed that the propositions on the Goodstein sequence and the Hydra game are independent of PA.

- H. Friedman showed that Kruskal's theorem (1982) and the Robertson-Seimor theorem in graph theory (1987) are independent of a stronger subsystem of second-order arithmetic, and also discovered various independent propositions for set theory.

Thank you for your attention!