Logic and Computation II

Part 4. Formal arithmetic and Gödel's incompleteness theorems

Kazuyuki Tanaka

BIMSA

March 22, 2023

北京雁栖湖
应用数学研究院
YANQI LAKE BEIJING INSTITUTE OF
MATHEMATICAL SCIENCES AND APPLICATIONS

Logic and Computation II

- **Part 4. Formal arithmetic and Gödel's incompleteness theorems**
- **Part 5. Automata on infinite objects**
- **Part 6. Recursion-theoretic hierarchies**
- **Part 7. Admissible ordinals and second order arithmetic**

Part 4. Schedule

- Mar. 7, (1) First-order logic
- Mar. 9, (2) Arithmetical formulas
- Mar.14, (3) Gödel's first incompleteness theorem
- Mar.16, (4) Gödel's second incompleteness theorem
- Mar.21, (5) Second-order logic
- Mar.23, (6) Analytical formulas

Today's topics

Introduction

- 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.

Lemma (Strong Representation for primitive recursive functions)

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

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

Then, $\mathrm{I}\Sigma_1 + \forall x \varphi(x, f(x))$ is conservative over $\mathrm{I}\Sigma_1$.

Lemma (Diagonalization lemma)

For any formula $\psi(x)$ with a unique free variable x, there exists a sentence σ such that $\mathrm{I}\Sigma_1 \vdash \sigma \leftrightarrow \psi(\overline{\ulcorner \sigma \urcorner})$.

Definition (Provability predicate Bew)

Let T be a CE theory including $\mathrm{I}\Sigma_1$. Then, we define a prim. rec. relation $\mathrm{Proof}_T(\ulcorner P \urcorner, \ulcorner \sigma \urcorner)$ to express "P is a proof of formula σ in T".
By Proof_T, we also denote a Δ_1 formula expressing the relation Proof_T in $\mathrm{I}\Sigma_1$.
A Σ_1 formula Bew_T is defined as $\mathrm{Bew}_T(x) \equiv \exists y\, \mathrm{Proof}_T(y, x)$.

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

The First Incompleteness Theorem

Theorem (Gödel's first incompleteness theorem)

Any 1-consistent CE theory T including $\mathsf{I}\Sigma_1$ 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 thus constructed "asserts its own unprovability" because "$\sigma \Leftrightarrow T \nvdash \sigma$" holds. This σ is called the **Gödel sentence** of T.

Using Homework of lecture-04-03, the assumption of T can be weakened from 1-consistency to consistency (Gödel-Rosser's theorem).

Homework

Complete the proof of Gödel-Rosser's theorem.

- Let $A = \{\ulcorner\sigma\urcorner : T \vdash \sigma\}$, $B = \{\ulcorner\sigma\urcorner : T \vdash \neg\sigma\}$. If T is consitent CE theory, then A, B are disjoint CE sets.
- Similarly to the proof of the strong representation theorem for computable sets, costruct a formula $\psi(x)$ such that $A \subset \{n : T \vdash \psi(\overline{n})\}$ and $B \subset \{n : T \vdash \neg\psi(\overline{n})\}$.
- Considering the sentence σ such that $T \vdash (\sigma \leftrightarrow \neg\psi(\ulcorner\sigma\urcorner))$, prove that $\ulcorner\sigma\urcorner \notin A \cup B$.
- Also, notice that if A, B were computably separable, we could construct a formula $\psi(x)$ such that $\{n : T \vdash \psi(\overline{n})\} \cup \{n : T \vdash \neg\psi(\overline{n})\} = \mathbb{N}$

Two applications of the first incomp. theorem

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 $\mathsf{I}\Sigma_1$ 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 $\mathsf{I}\Sigma_1$, you may take $\mathsf{Q}_<$ 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 $\mathsf{I}\Sigma_1$ by $\wedge$.
- Then, from the deduction theorem, $\mathsf{I}\Sigma_1 \vdash \sigma \Leftrightarrow \vdash \xi \to \sigma$. If $\{\ulcorner\sigma\urcorner : \vdash \sigma\}$ is computable, $\{\ulcorner\sigma\urcorner : \vdash \xi \to \sigma\} = \{\ulcorner\sigma\urcorner : \mathsf{I}\Sigma_1 \vdash \sigma\}$ is also computable, which contradicts with the above homework argument.
- Finally, note that the satisfiability of first order logic can be expressed as $\{\ulcorner\sigma\urcorner : \not\vdash \neg\sigma\}$ and that if it were computable then $\{\ulcorner\sigma\urcorner : \vdash \neg\sigma\}$ would be computable.

The next theorem is also a very important corollary of the argument of the first incompleteness theorem. Note that T in the diagonalization lemma does not need be a CE theory. So, letting T be $\mathrm{Th}(\mathfrak{N})$, i.e., the set of all sentences true in $\mathfrak{N}$, we have

Theorem (Tarski's Truth Indefinability)

For any sentence σ, there is no formula $\psi(x)$ such that

$$\mathfrak{N} \models \sigma \leftrightarrow \psi(\overline{\ulcorner \sigma \urcorner}).$$

In other words, $\{\ulcorner \sigma \urcorner : \mathfrak{N} \models \sigma\}$ is not arithmetically definable.

Proof. Consider a fixed point σ for $\neg\psi(x)$.

Intoducing the second incompleteness theorem

- A version of the first incompleteness theorem says that a consitent CE theory T including $\mathsf{I}\Sigma_1$ (indeed $\mathsf{Q}_<$ is enough) neither prove (nor disprove) the Gödel sentence.
- A main part of the second incompleteness theorem says that a CE theory T including $\mathsf{I}\Sigma_1$ proves that the consistency of T implies the Gödel sentence (equivalently, its unprovability).
- Then, we obtain the second incompleteness theorem that a consistent T does not prove its consistency, since if it did then it would also prove the Gödel sentence, which contradicts with the first theorem.
- 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 $\mathsf{I}\Sigma_1$. 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 $\mathrm{Bew}_T(\overline{\ulcorner\varphi\urcorner})$. We will explain an idea of this machinery in the next slide.

- First, we prove that, for any primitive recursive function f,

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

  Here, the function $\dot{x}$ is a primitive recursive function from a number n to the Gödel number of its numeral $\ulcorner \bar{n} \urcorner$.
- The above formula can be proved by meta-induction on the construction of the primitive recursive function f.
- Now, assume $\mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner})$. Then, there is a numeral c that satisfies $\mathrm{Proof}_T(c, \overline{\ulcorner \varphi \urcorner})$. So, substituing (the numeral of the Gödel number of) this formula into $\mathrm{Bew}_T(x)$, we finally obtain $\mathrm{Bew}_T(\overline{\ulcorner \mathrm{Bew}_T(\overline{\ulcorner \varphi \urcorner}) \urcorner})$ by a simple computation.
- For more details, please refer to my book[1].

□

[1] https://www.shokabo.co.jp/mybooks/ISBN978-4-7853-1575-7.htm

In the following, let π_G denote a Gödel sentence such that

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

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 $\pi_G \to \mathrm{Con}(T)$. $T \vdash 0 = 1 \to \pi_G$, 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 get $T \vdash \pi_G \to \mathrm{Con}(T)$.

Proof. • 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,

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

By D3, $T \vdash \mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}) \to \mathrm{Bew}_T(\overline{\ulcorner \mathrm{Bew}_T(\overline{\ulcorner \pi_G \urcorner}) \urcorner})$, so

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

Using $T \vdash \pi_G \to (\neg\pi_G \to 0 = 1)$ and D2, from above

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

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 $\mathrm{Con}(T)$ cannot be proved in 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

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 $\mathrm{Bew}^{\#}_T(x) \equiv (\mathrm{Bew}_T(x) \land x \neq \overline{\ulcorner 0 = 1 \urcorner})$. 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?

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.

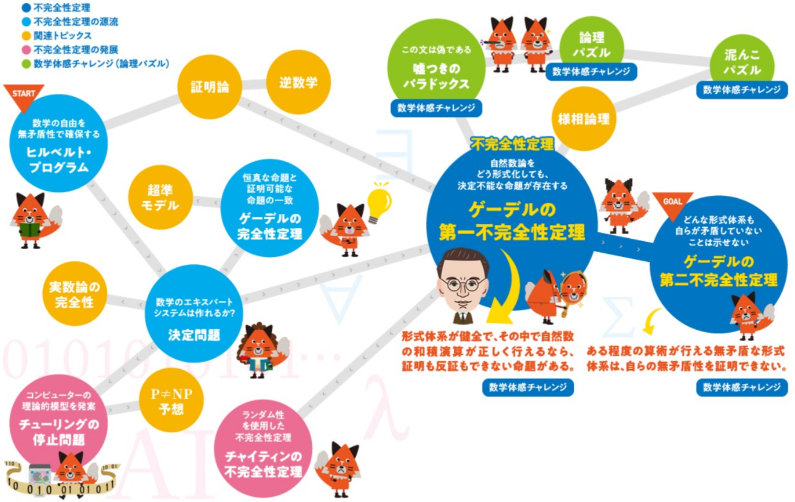
- T. Franzen, Gödel's Theorem: An Incomplete Guide to Its Use and Abuse(2005).

  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!