

Gödel's Incompleteness Theorem

Deepak D'Souza

Department of Computer Science and Automation
Indian Institute of Science, Bangalore.

9 April 2024

Outline

- 1 Theory of Arithmetic
- 2 Peano's Proof System for Arithmetic
- 3 Proof of Gödel's theorem

Gödel's Incompleteness Theorem

Theorem (Gödel (1931))

*There **cannot** exist a sound and complete proof system for arithmetic (i.e. First-Order Logic of natural numbers with addition and multiplication $(\mathbb{N}, +, \cdot)$).*

Arithmetic

First-order logic of $(\mathbb{N}, +, \cdot)$:

- Domain is $\mathbb{N} = \{0, 1, 2, \dots\}$.
- Terms: 0 , 1 , $0 + 1$, $1 \cdot x$, $x + y$, $x \cdot y$, etc.
- Atomic formulas: $t = t$
- Note that relations like “ $<$ ” are definable in the logic: $t < t'$ is definable as $\exists x(x \neq 0 \wedge t + x = t')$.
- Formulas:
 - Atomic formulas
 - Quantification: $\forall x\varphi$, $\exists x\varphi$
 - Boolean combinations: $\neg\varphi$, $\varphi \vee \psi$, $\varphi \wedge \psi$.

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Integer division of x by y gives quotient q and leaves remainder r ”

$$\text{intdiv}(x, y, q, r) \stackrel{\text{def}}{=} x = (q \cdot y) + r \wedge r < y.$$

- “ y divides x ”

$$\text{divides}(y, x) \stackrel{\text{def}}{=} \exists q(x = q \cdot y).$$

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Integer division of x by y gives quotient q and leaves remainder r ”

$$\text{intdiv}(x, y, q, r) \stackrel{\text{def}}{=} x = (q \cdot y) + r \wedge r < y.$$

- “ y divides x ”

$$\text{divides}(y, x) \stackrel{\text{def}}{=} \exists q(x = q \cdot y).$$

- “ x is prime”

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Integer division of x by y gives quotient q and leaves remainder r ”

$$\text{intdiv}(x, y, q, r) \stackrel{\text{def}}{=} x = (q \cdot y) + r \wedge r < y.$$

- “ y divides x ”

$$\text{divides}(y, x) \stackrel{\text{def}}{=} \exists q (x = q \cdot y).$$

- “ x is prime”

$$\text{prime}(x) \stackrel{\text{def}}{=} x \geq 2 \wedge \forall y (\text{divides}(y, x) \implies (y = 1 \vee y = x)).$$

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Integer division of x by y gives quotient q and leaves remainder r ”

$$\text{intdiv}(x, y, q, r) \stackrel{\text{def}}{=} x = (q \cdot y) + r \wedge r < y.$$

- “ y divides x ”

$$\text{divides}(y, x) \stackrel{\text{def}}{=} \exists q (x = q \cdot y).$$

- “ x is prime”

$$\text{prime}(x) \stackrel{\text{def}}{=} x \geq 2 \wedge \forall y (\text{divides}(y, x) \implies (y = 1 \vee y = x)).$$

- “ x is a power of 2”

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Integer division of x by y gives quotient q and leaves remainder r ”

$$\text{intdiv}(x, y, q, r) \stackrel{\text{def}}{=} x = (q \cdot y) + r \wedge r < y.$$

- “ y divides x ”

$$\text{divides}(y, x) \stackrel{\text{def}}{=} \exists q (x = q \cdot y).$$

- “ x is prime”

$$\text{prime}(x) \stackrel{\text{def}}{=} x \geq 2 \wedge \forall y (\text{divides}(y, x) \implies (y = 1 \vee y = x)).$$

- “ x is a power of 2”

$$\text{power}_2(x) \stackrel{\text{def}}{=} \forall p ((\text{prime}(p) \wedge \text{divides}(p, x)) \implies p = 2).$$

What we can say in $\text{FO}(\mathbb{N}, +, \cdot)$

- “Every number has a successor”

$$\forall n \exists m (m = n + 1).$$

- “Every number has a predecessor”

$$\forall n \exists m (n = m + 1).$$

- “There are only finitely many primes”

$$\exists n \forall p (\text{prime}(p) \implies p < n).$$

- “There are infinitely many primes”

$$\forall n \exists p (\text{prime}(p) \wedge p > n).$$

Theory of $\text{FO}(\mathbb{N}, +, \cdot)$

$Th(\mathbb{N}, +, \cdot)$ is the set of sentences of $\text{FO}(\mathbb{N}, +, \cdot)$ that are **true**. For example:

- “Every number has a successor”

$$\forall n \exists m (m = n + 1).$$

belongs to $Th(\mathbb{N}, +, \cdot)$, while

- “There are only finitely many primes”

$$\exists n \forall p (\text{prime}(p) \implies p < n).$$

does not.

Note that there is a mathematical definition of truth based on the mathematical definition of the semantics of the logic.

Peano's Proof System for Arithmetic

- Axioms:

$$\begin{aligned}& \forall x \neg(0 = x + 1) \\& \forall x \forall y (x + 1 = y + 1 \implies x = y) \\& \forall x (x + 0 = x) \\& \forall x \forall y \forall z (x + (y + z) = (x + y) + z) \\& \forall x (x \cdot 0 = 0) \\& \forall x \forall y \forall z (x \cdot (y + z) = ((x \cdot y) + (x \cdot z))) \\& (\varphi(0) \wedge \forall x (\varphi(x) \implies \varphi(x + 1))) \implies \forall x \varphi(x).\end{aligned}$$

- Other axioms like $(\varphi \wedge \psi) \implies \varphi$, $\forall x(\varphi) \implies \varphi(17)$.
- Inference rules like 'Modus Ponens'

Given φ and $\varphi \implies \psi$, infer ψ .

Proof

A proof of φ in a proof system is a finite sequence of sentences

$$\varphi_0, \varphi_1, \dots, \varphi_n$$

such that each φ_i is either an axiom or follows from two previous ones by an inference rule, and $\varphi_n = \varphi$.

A proof system is “**sound**” if whatever it proves is indeed true (i.e. in $Th(\mathbb{N})$).

A proof system is “**complete**” if it can prove whatever is true (i.e. in $Th(\mathbb{N})$).

Gödel's Incompleteness Theorem

Theorem (Gödel (1931))

*There **cannot** exist a sound and complete proof system for arithmetic (i.e. First-Order Logic of natural numbers with addition and multiplication $(\mathbb{N}, +, \cdot)$).*

Proof of Gödel's theorem

- Gödel's original proof was an intricate construction of an $FO(\mathbb{N}, +, \cdot)$ sentence φ which (for a given proof system like Peano's) asserts that
"I am not provable in the given proof system"
- The sentence φ cannot be *false*. If it were, then φ would be provable, which would mean the proof system is **unsound**. So φ must be *true*, which means that there is a true sentence (name φ itself) which is true but has no proof in the system.
- Here we will follow a subsequent proof given by Turing which shows

$$\neg \text{HP} \leq \text{Th}(\mathbb{N}).$$

- Hence $\text{Th}(\mathbb{N})$ is not even r.e. and hence there cannot be a proof system that is sound and complete (why?).

Encoding computations of M on x

Let $M = (Q, A, \Gamma, s, \delta, \vdash, \flat, t, r)$ be a given TM and let $x = a_1 a_2 \cdots a_n$ be an input to it.

We can represent a configuration of M as follows:

$$\begin{array}{ccccccc} \vdash & b_1 & b_2 & b_3 & \cdots & b_m \\ - & - & q & - & & - \end{array}$$

Thus a configuration is encoded over the alphabet $\Gamma \times (Q \cup \{-\})$.

Encoding computations of M on x

A computation of M on x is a string of the form

$$c_0 \# c_1 \# \cdots \# c_N \#$$

such that

- ① Each c_i is the encoding of a configuration of M .
- ② c_0 is (encoding of) the start configuration of M on x .

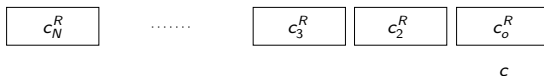
$$\begin{array}{ccccccc} \vdash & a_1 & a_2 & a_3 & \cdots & a_n \\ s & - & - & - & & - \end{array}$$

- ③ All c_i 's are of the **same** length.
- ④ Each $c_i \xRightarrow{1} c_{i+1}$, and
- ⑤ c_N is a halting configuration (i.e. state component is t or r).



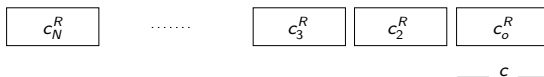
Basic idea

View a computation of M on x as a number whose representation in base $p \geq |\Delta|$ looks like:



Now construct a sentence $\varphi_{M,x}$ which asserts that “there is a number n whose base- p representation encodes a valid halting computation of M on x .”

The sentence $\varphi_{M,x}$



- Define $valcomp_{M,x}(v)$ to be

$$\begin{aligned} \exists c \exists d (& power_p(c) \wedge power_p(d) \\ & \wedge length(v, d) \wedge start(v, c) \\ & \wedge move(v, c, d) \wedge halt(v, d)). \end{aligned}$$

- Define $\varphi_{M,x}$ to be

$$\exists v \, valcomp_{M,x}(v).$$

Expressing the components of $\varphi_{M,x}$

The key predicate we need is “ $digit_p(v, d, a)$ ”: which says that d is a power of p (say $d = p^k$), and in the base- p representation, the k -th digit of v (from the least significant end) is a .

$$digit_p(v, d, a) \stackrel{\text{def}}{=} \exists u \exists r (v = u \cdot p \cdot d + a \cdot d + r \wedge r < d).$$