

Diagonalization

We bring together all the tools developed so far to prove the crucial step toward incompleteness: the diagonal lemma.

1. Previously on...

The final goal is now close at hand: proving the Incompleteness Theorem. This is another technical chapter, centered on a crucial lemma that requires all the preceding chapters. Before diving into the details, let us briefly review all the concepts we will need.

1.1. Peano arithmetic

- Peano arithmetic is based on the language $\mathcal{L}_0 = (0, S, +, \times)$, in which we construct a theory $\mathcal{PA}$ (the Peano axioms).
- The natural numbers $\mathbb{N}$ form a model of this theory: the Peano axioms are satisfied in $\mathbb{N}$.
- For $n \in \mathbb{N}$, its numeral $\bar{n} = SS \dots S0$ is its representation in $\mathcal{L}_0$ (there are exactly n symbols S).

1.2. Gödel coding

- Every formula of $\mathcal{L}_0$ is assigned a natural number: its Gödel number. We also write it as $n = \ulcorner \mathcal{P} \urcorner$.
- In $\mathcal{L}_0$, its numeral is therefore $\bar{n} = \overline{\ulcorner \mathcal{P} \urcorner}$, which we will also often denote by $\ulcorner \mathcal{P} \urcorner$ (this is a slight abuse of notation, but it cannot cause confusion within formulas).
- We already gave an overview of the diagonalization of a formula during the informal proof of incompleteness. The idea is to consider $\mathcal{G} = \mathcal{P}(\ulcorner \mathcal{G} \urcorner)$.
- The goal of this chapter is to formalize the formula $\mathcal{G}$ rigorously.

1.3. Primitive recursive functions

- Recall that primitive recursive functions are readily computable functions $f : \mathbb{N}^p \rightarrow \mathbb{N}$.
- The important point for this chapter is that the function $\text{substitution}(n, v, k) : \mathbb{N}^3 \rightarrow \mathbb{N}$ is primitive recursive: n denotes the Gödel number of a formula $\mathcal{P}$, v the code of a variable x , and k the value to be substituted for x . The function therefore returns $\ulcorner \mathcal{P}(k) \urcorner \in \mathbb{N}$.

1.4. Representable functions

- Every primitive recursive function is representable in $\mathcal{PA}$. We can therefore translate all our mathematical concepts into $\mathcal{L}_0$.
- For a primitive recursive function $f : \mathbb{N} \rightarrow \mathbb{N}$, this means that there exists a formula $\mathcal{P}(x, y)$ of $\mathcal{L}_0$ such that, for every $n \in \mathbb{N}$, there is a proof in $\mathcal{PA}$ of the equivalence:

$$\forall y \quad \mathcal{P}(\bar{n}, y) \leftrightarrow y = \overline{f(n)}$$

- The formula $\mathcal{P}$ plays the role of the graph of f : it uniquely characterizes the value $y = f(n)$ in $\mathcal{L}_0$.

2. Diagonal lemma

Let us tackle the cornerstone of the proof of the Incompleteness Theorem.

2.1. Statement

Lemma 1 (Diagonal lemma).

For every formula $\mathcal{P}(x)$ (where x is the only free variable), there exists a closed formula $\mathcal{G}$ such that:

$$\boxed{\mathcal{PA} \vdash (\mathcal{G} \leftrightarrow \mathcal{P}(\ulcorner \mathcal{G} \urcorner))}$$

- Thus $\mathcal{G}$ appears twice in the statement: once as a logical formula and once through its Gödel number. The formula $\mathcal{G}$ therefore satisfies a property about itself.
- The diagonal lemma can be viewed as a fixed-point theorem: instead of obtaining $f(x) = x$, we have $\mathcal{P}(\ulcorner \mathcal{G} \urcorner) \leftrightarrow \mathcal{G}$.
- The equivalence is the syntactically correct way to say that $\mathcal{G}$ and $\mathcal{P}(\ulcorner \mathcal{G} \urcorner)$ are equivalent statements (but they are not equal formulas in $\mathcal{L}_0$). On the other hand, the proof below will show that $\mathcal{G}$ is defined exactly by $\mathcal{G} = \mathcal{H}(\ulcorner \mathcal{H} \urcorner)$.

2.2. Proof

Diagonalization on natural numbers

We will construct a primitive recursive function $\text{diag} : \mathbb{N} \rightarrow \mathbb{N}$.

- Given $\mathcal{Q}(x)$, we know how to assign it its Gödel number $n = \ulcorner \mathcal{Q} \urcorner \in \mathbb{N}$. We will consider the reverse operation.
- For $n \in \mathbb{N}$, we can decide whether this natural number corresponds to a well-formed formula $\mathcal{Q}(x)$ of $\mathcal{L}_0$ whose only free variable is x . If it does not, set $\text{diag}(n) = 0$.
- If n represents the formula $\mathcal{Q}(x)$, then we can evaluate $\mathcal{Q}$ at n , that is, form the substitution $\mathcal{Q}(\bar{n})$. This is a formula of $\mathcal{L}_0$ whose Gödel number is $\ulcorner \mathcal{Q}(\bar{n}) \urcorner$. We then define the natural number:

$$\text{diag}(n) = \ulcorner \mathcal{Q}(\bar{n}) \urcorner$$

That is, $\text{diag}(n) = \ulcorner \mathcal{Q}(\ulcorner \mathcal{Q} \urcorner) \urcorner$.

- Let us summarize the construction with a diagram:

$$n \in \mathbb{N} \xrightarrow{\text{Gödel decoding}} Q \in \mathcal{L}_0 \xrightarrow{\text{substitution}} Q(\bar{n}) \in \mathcal{L}_0 \xrightarrow{\text{Gödel number}} \text{diag}(n) = \ulcorner Q(\bar{n}) \urcorner \in \mathbb{N}$$

- For a formula $Q(x)$ with Gödel number $n = \ulcorner Q \urcorner$, the computation of $Q(k)$ is performed by the function $\text{substitution}(n, \ulcorner x \urcorner, k)$. For the formula Q with Gödel number n , this function replaces the occurrences of the variable x by $\bar{k}$ and then returns the Gödel number of the resulting formula, namely $\ulcorner Q(\bar{k}) \urcorner$.

Remark: throughout the rest of the chapter, the variable being replaced will always be x .

- Deciding whether a natural number n denotes a well-formed formula $Q(x)$ whose only free variable is x is a primitive recursive operation. Substitution is primitive recursive as well (see the chapters “Gödel coding” and “Primitive recursive functions”). Thus the function $\text{diag} : \mathbb{N} \rightarrow \mathbb{N}$ is also primitive recursive.

Diagonalization in the language

- Every primitive recursive function is representable (see the chapter “Representable Functions”). Therefore, for our function $\text{diag} : \mathbb{N} \rightarrow \mathbb{N}$, there exists a formula $\text{Diag}(x, y)$ of $\mathcal{L}_0$ such that, for every $n \in \mathbb{N}$:

$$\mathcal{PA} \vdash \left(\forall y \text{ Diag}(\bar{n}, y) \leftrightarrow y = \overline{\text{diag}(n)} \right)$$

The formula $\mathcal{G}$

- Fix the formula $\mathcal{P}(x)$ from the statement and define a new formula $\mathcal{H}$ by:

$$\mathcal{H}(x) : \exists y \text{ Diag}(x, y) \wedge \mathcal{P}(y)$$

Let p denote the Gödel number of this last formula: $p = \ulcorner \mathcal{H} \urcorner$.

- Define the Gödel formula by:

$$\mathcal{G} = \mathcal{H}(\bar{p})$$

Thus, by definition, $\mathcal{G}$ denotes “ $\mathcal{H}(\bar{p})$ ” and therefore also “ $\exists y \text{ Diag}(\bar{p}, y) \wedge \mathcal{P}(y)$ ”.

- Moreover, $\text{diag}(p) = \ulcorner \mathcal{G} \urcorner$. Indeed, apply the diagram above to the natural number p :

$$p \in \mathbb{N} \xrightarrow{\text{Gödel decoding}} \mathcal{H}(x) \in \mathcal{L}_0 \xrightarrow{\text{substitution}} \mathcal{H}(\bar{p}) \in \mathcal{L}_0 \xrightarrow{\text{Gödel number}} \text{diag}(p) = \ulcorner \mathcal{H}(\bar{p}) \urcorner = \ulcorner \mathcal{G} \urcorner \in \mathbb{N}$$

End of the proof

Recall that the goal is to show:

$$\mathcal{PA} \vdash \left(\mathcal{G} \leftrightarrow \mathcal{P}(\ulcorner \mathcal{G} \urcorner) \right)$$

We therefore work in the theory $\mathcal{PA}$. Recall that $p = \ulcorner \mathcal{H} \urcorner$, $\mathcal{G} = \mathcal{H}(\bar{p})$, and $\text{diag}(p) = \ulcorner \mathcal{G} \urcorner$.

- **Forward direction** $\rightarrow$

Suppose $\mathcal{G}$

therefore $\mathcal{H}(\bar{p})$ by the definition of $\mathcal{G}$

therefore $\exists y \text{ Diag}(\bar{p}, y) \wedge \mathcal{P}(y)$ by the definition of $\mathcal{H}$

therefore $\exists y (y = \overline{\text{diag}(p)}) \wedge (\mathcal{P}(y))$ since diag is represented by Diag in $\mathcal{PA}$

therefore $\mathcal{P}(\overline{\text{diag}(p)})$ because the equality determines the value of y

therefore $\mathcal{P}(\ulcorner \mathcal{G} \urcorner)$

Thus $\mathcal{G} \rightarrow \mathcal{P}(\ulcorner \mathcal{G} \urcorner)$.

- **Reverse direction** $\leftarrow$

Suppose $\mathcal{P}(\ulcorner \mathcal{G} \urcorner)$. We must deduce $\mathcal{G} = \mathcal{H}(\overline{p})$, that is:

$$\exists y \quad \text{Diag}(\overline{p}, y) \wedge \mathcal{P}(y)$$

We will show that $y = \ulcorner \mathcal{G} \urcorner$ works.

- On the one hand, $y = \ulcorner \mathcal{G} \urcorner = \overline{\text{diag}(\overline{p})}$, so $\text{Diag}(\overline{p}, y)$ holds by representability.
- Moreover, $\mathcal{P}(y) = \mathcal{P}(\ulcorner \mathcal{G} \urcorner)$ is our assumption.
- Therefore, for this y , we do have $\text{Diag}(\overline{p}, y) \wedge \mathcal{P}(y)$.

By $\exists$ introduction, we have:

$$\mathcal{P}(\ulcorner \mathcal{G} \urcorner) \rightarrow [\exists y \quad \text{Diag}(\overline{p}, y) \wedge \mathcal{P}(y)]$$

which is exactly $\mathcal{P}(\ulcorner \mathcal{G} \urcorner) \rightarrow \mathcal{G}$.

This completes the proof of the diagonal lemma. We have shown:

$$\mathcal{P}\mathcal{A} \vdash (\mathcal{G} \leftrightarrow \mathcal{P}(\ulcorner \mathcal{G} \urcorner))$$

3. Tarski's theorem

3.1. Motivation

The rest of this chapter is not needed for the proof of the Incompleteness Theorem. However, it is an excellent application of the diagonal lemma and another opportunity to emphasize the difference between truth and provability. Tarski's theorem is a powerful version of the liar paradox. I make the following statement:

“I am lying right now.”

- If the statement is true, then I am lying, so the statement is false. Contradiction.
- If the statement is false, then I am not lying, so the statement is true. Another contradiction.

A variant is:

“This statement is false.”

3.2. Truth is not definable

Informally, Tarski's theorem (1933) says that there is no formula $\mathcal{V}(x)$ that recognizes true formulas; that is, there is no formula $\mathcal{V}(x)$ such that $\mathcal{V}(\overline{n})$ is true exactly when the closed formula $\mathcal{P}$ with Gödel number n is true.

In our context, a closed formula $\mathcal{P}$ is true if $\models_{\mathbb{N}} \mathcal{P}$, that is, if it is satisfied in the standard model of the natural numbers $\mathbb{N}$.

Here is the precise statement.

Theorem 1 (Tarski).

There is no formula $\mathcal{V}(x)$ of $\mathcal{L}_0$ such that, for every closed formula $\mathcal{P}$ of $\mathcal{L}_0$:

$$\models_{\mathbb{N}} \mathcal{P} \quad \text{if and only if} \quad \models_{\mathbb{N}} \mathcal{V}(\ulcorner \mathcal{P} \urcorner)$$

An equivalent formulation is “truth is not definable”. Indeed, let A be the set of natural numbers $n = \ulcorner \mathcal{P} \urcorner$ such that $\models_{\mathbb{N}} \mathcal{P}$ (that is, the Gödel numbers of true formulas). Then there is no formula $\mathcal{V}(x)$ such that $\models_{\mathbb{N}} \mathcal{V}(\overline{n})$ holds for every $n \in A$ and only for those n . For this notion of “definable”, see the subsection “Definable sets” in the chapter “Peano arithmetic”. On the other hand, the next chapter will show that “being provable” is definable.

3.3. Proof

Suppose for a contradiction that such a formula $\mathcal{V}(x)$ exists. Apply the diagonal lemma to $\neg\mathcal{V}(x)$ (in place of $\mathcal{P}(x)$ in the statement of the lemma). Thus there exists a closed formula $\mathcal{G}$ such that:

$$\mathcal{PA} \vdash (\mathcal{G} \leftrightarrow \neg\mathcal{V}(\overline{\mathcal{G}}))$$

This sentence $\mathcal{G}$ is our version of the liar talking about itself. We work in the structure $\mathbb{N}$ for the theory $\mathcal{PA}$ over $\mathcal{L}_0$. By the Soundness Theorem, if $\mathcal{PA} \vdash \mathcal{Q}$, then $\models_{\mathbb{N}} \mathcal{Q}$: everything provable is true. Thus the equivalence above is true:

$$\mathcal{G} \leftrightarrow \neg\mathcal{V}(\overline{\mathcal{G}})$$

Let us verify that the sentence $\mathcal{G}$ contradicts the equivalence required in Theorem 1 (where $\mathcal{G}$ plays the role of the sentence $\mathcal{P}$).

- If $\mathcal{G}$ is true, that is, $\models_{\mathbb{N}} \mathcal{G}$, then by the equivalence from the diagonal lemma we also have $\models_{\mathbb{N}} \neg\mathcal{V}(\overline{\mathcal{G}})$. In other words, $\mathcal{V}(\overline{\mathcal{G}})$ is false, so we do not have $\models_{\mathbb{N}} \mathcal{V}(\overline{\mathcal{G}})$. Thus, in this case, $\mathcal{V}$ fails to recognize that $\mathcal{G}$ is true.
- If $\mathcal{G}$ is false (that is, we do not have $\models_{\mathbb{N}} \mathcal{G}$), then by the equivalence from the diagonal lemma, $\neg\mathcal{V}(\overline{\mathcal{G}})$ is false, so $\mathcal{V}(\overline{\mathcal{G}})$ is true. Thus we have $\models_{\mathbb{N}} \mathcal{V}(\overline{\mathcal{G}})$. Consequently, in this case, $\mathcal{V}$ fails to recognize that $\mathcal{G}$ is false.
- Thus, without knowing whether $\mathcal{G}$ is true or false, $\mathcal{G}$ is nevertheless a counterexample to the desired equivalence.