Gödel's Incompleteness Theorems

In the early 1900's there was a drive to find adequate axiomatic foundations for mathematics. Russell's paradox (If S is the set of all sets that do not contain themselves, does S contain itself?) helped to point out how difficult it is to find a good axiom system for set theory. David Hilbert, the most prominent mathematician of the time, proposed a program of finding axiom systems, and proving them consistent by "finitary" means; that is finite combinatorial methods that do not involve questionable set-theoretic constructions. Gödel's 1931 paper effectively destroyed hopes for the success of this program. Gödel proved that **PA** cannot even prove its own consistency, let alone the consistency of a more powerful system such as set theory.

In his 1931 paper Gödel proved two results (his two "incompleteness theorems"). The second incompleteness theorem states that the consistency of **PA** cannot be proved in **PA**. Here we prove the first incompleteness theorem, and outline the proof of the second. (In fact, Gödel did not include a complete proof of his second theorem, but complete proofs now appear in text and reference books.)

Here we consider only the theory **PA**, although the first incompleteness theorem applies to any consistent extension of **RA**, and the second incompleteness theorem applies to "nice" theories of arithmetic, which in general must include some form of induction among their axioms.

The first theorem formulates a sentence G which asserts "I am not provable", and the theorem states that indeed G is not provable in **PA**, so G is true. By soundness of **PA**, $\neg G$ is also not provable in **PA**. The method of constructing G follows the method of constructing the sentence "I am false" in the proof of Tarski's Theorem. (Historically, Gödel's theorems came first.)

Let Γ be the set of axioms of **PA**. Thus Γ consists of P1,...,P6, together with the induction axioms. Let $Proof(x,y)$ be the recursive relation "y codes an $LK-\Gamma$ proof of the sentence coded by x". Thus $\exists y Proof(n,y)$ holds iff $n = \#A$, where A is a sentence provable in **PA**. Let $d(x)$ be the diagonal substitution function (defined on page 89). Recall that $d(n) = sub(n,n) = \#A(s_n)$ when $\#A(x) = n$. Then $d(x)$ is total and computable, so the relation $S(x)$ is r.e., where

$$S(x) = \exists y Proof(d(x), y)$$

Let $A(x)$ be an $\exists\Delta_0$ formula which represents $S(x)$ in **RA** (and hence in **PA**). Then for all $n \in \mathbb{N}$,

$$\exists y Proof(d(n), y) \quad \Leftrightarrow \quad \mathbf{PA} \vdash A(s_n) \tag{1}$$

Let $e = \#\neg A(x)$, so

$$d(e) = \#\neg A(s_e) \tag{2}$$

Let

$$G =_{syn} \neg A(s_e)$$

so $\#G = d(e)$. Since $A(x)$ represents the relation $\exists y Proof(d(x), y)$, it follows that the formula $\neg A(s_e)$ asserts that the formula whose number is $d(e)$ is not provable in **PA**. But that formula is $\neg A(s_e)$, so this formula, i.e. the formula G, asserts "I am not provable".

Gödel's First Incompleteness Theorem: If **PA** is consistent, then **PA** does not prove G.

Remark: Note that in this course we take for granted that **PA** is consistent. The reason that Gödel did not, is that there is no known "finitary" proof that **PA** is consistent. Our proof of consistency involves the assertion that **PA** is sound. That is, all of the axioms of **PA** are true in the standard model $\underline{\mathbb{N}}$, and hence all logical consequences of these axioms are true in $\underline{\mathbb{N}}$. But this proof is not finitary, because it involves an induction on a statement mentioning the infinite set $\mathbb{N}$.

Proof: We prove the contrapositive. Suppose that $\mathbf{PA} \vdash G$, i.e. $\mathbf{PA} \vdash \neg A(s_e)$. Then sentence number $d(e)$ is provable, so $\exists y Proof(d(e), y)$ holds. Hence $\mathbf{PA} \vdash A(s_e)$, by the left-to-right direction of (1). Thus **PA** proves both a formula and its negation, so it is inconsistent. □

The above proof is finitary, in that it involves only finite objects. Later we will argue, as Gödel did, that the proof can be formalized in **PA**. It is important that the proof only uses the left-to-right direction of (1), since this direction is finitary: From a proof of the sentence whose number is $d(n)$ one can construct a proof of the sentence $A(s_n)$. Our proof of the converse direction of (1) is not finitary, since it involves the soundness of **PA**. It is not clear that **PA** can prove this converse direction. However, using the right-to-left direction we can prove the following:

Proposition: If **PA** is sound, then **PA** does not prove $\neg G$.

Proof: Suppose **PA** proves $\neg G$; i.e. **PA** proves $A(s_e)$. By the right-to-left direction of (1), this implies $\exists y Proof(d(e), y)$; that is, **PA** proves sentence number $d(e)$, so **PA** proves $\neg A(s_e)$, so **PA** proves G. Thus **PA** is inconsistent, and hence unsound. □

Remark: We say that a theory Σ is ω-*consistent* provided that for each formula $C(x)$, if Σ proves $\neg C(s_n)$ for each $n \in \mathbb{N}$, then Σ does not prove $\exists x C(x)$. Every sound theory is ω-consistent, but not conversely. It is not hard to see the assumption that **PA** is ω-consistent is sufficient to prove the right-to-left direction in (1), and hence this assumption can replace the stronger assumption that **PA** is sound, in the above Proposition.

Exercise 1 *Show that there is a consistent extension of* **PA** *which is not ω-consistent.*

Formulating consistency in PA

Let $B(x, y)$ be an $\exists\Delta_0$ formula which represents $Proof(x, y)$ in **RA** (and hence in **PA**).

Thus for each sentence C,

$$\mathbf{PA} \vdash C \quad \Leftrightarrow \quad \mathbf{PA} \vdash \exists y B(\#C, y) \tag{3}$$

where here (and below) we write $B(\#C, y)$ for $B(s_{\#C}, y)$.

We require that the formula $B(x, y)$ represent the relation $Proof(x, y)$ in a straightforward way, so that Lemma 2 and Lemma 3 below both hold.

Recall that $A(x)$ represents the relation $\exists y Proof(d(x), y)$ in **PA**. By constructing the formula $A(x)$ from $B(x, y)$ in a straightforward manner, we can insure that for each $n \in \mathbb{N}$

$$\mathbf{PA} \vdash \quad A(s_n) \supset \exists y B(s_{d(n)}, y) \tag{4}$$

Note that **PA** is consistent iff **PA** does not prove $0 \neq 0$. Thus we make the definition

$$con(PA) =_{syn} \neg \exists y B(\#0 \neq 0, y)$$

Gödel's Second Incompleteness Theorem: If **PA** is consistent, then **PA** does not prove $con(PA)$.

This follows from the following Lemma:

Lemma 1: (Gödel) $\mathbf{PA} \vdash \ con(PA) \supset G$

The Second Incompleteness Theorem follows immediately from the Lemma and the First Incompleteness Theorem.

The Lemma is proved by formalizing in **PA** the proof of the First Incompleteness Theorem. To see that "$con(PA) \supset G$" is an accurate translation of the First Incompleteness Theorem, note that G is $\neg A(s_e)$, which asserts that formula number $d(e)$ is not provable in **PA**; i.e. G asserts that G is not provable in **PA**.

Now we formalize the proof of the First Incompleteness Theorem in **PA**. Thus we must show that **PA** proves the contrapositive of the formula in Lemma 1; that is we must show

$$\mathbf{PA} \vdash \quad A(s_e) \supset \exists y B(\#0 \neq 0, y) \tag{5}$$

We need to formalize the left-to-right direction of (1), which involves formalizing the proof of Corollary 2 to the MAIN LEMMA, page 84. This corollary states that every true $\exists\Delta_0$ sentence C is provable in **RA** (and hence in **PA**). Thus we must show

Lemma 2: For each $\exists\Delta_0$ sentence C,

$$\mathbf{PA} \vdash \quad C \supset \exists z B(\#C, z)$$

The proof of this Lemma is the main work in the proof of the Second Incompleteness Theorem, and will not be given here. However we note that Lemma 2 is immediate for the case in which C is true, since then by Corollary 2 (to the MAIN LEMMA) C has a proof π in **RA**, and hence

$$\mathbf{RA} \vdash B(\#C, \#\pi)$$

because $B(x, y)$ represents $Proof(x, y)$ in **RA**. Despite this easy argument, the proof of Lemma 2 for the case in which C is false requires formalizing the proof of Corollary 2 (and the MAIN LEMMA itself), as mentioned above. (Note that there are false $\exists\Delta_0$ formulas C such that $\neg C$ is not provable in **PA**.)

If we take $C =_{syn} A(s_e)$ in Lemma 2 we obtain

$$\mathbf{PA} \vdash \quad A(s_e) \supset \exists z B(\#A(s_e), z) \tag{6}$$

Now from (4) with $n = e$ and (2) we obtain

$$\mathbf{PA} \vdash \quad A(s_e) \supset \exists z B(\#\neg A(s_e), z) \tag{7}$$

Finally, (5) follows from (7), (6), and the following lemma:

Lemma 3: For any sentence C,

$$\mathbf{PA} \vdash \forall x \forall z[(B(\#C, x) \wedge B(\#\neg C, z)) \supset \exists y B(\#0 \neq 0, y)]$$