A Simple Proof of Gödel's Incompleteness Theorems

Arindama Singh, Department of Mathematics, IIT Madras, Chennai-600036
Email: asingh@iitm.ac.in

1 Introduction

Gödel's incompleteness theorems are considered as achievements of twentieth century mathematics. The theorems say that the natural number system, or *arithmetic*, has a true sentence which cannot be proved and the consistency of arithmetic cannot be proved by using its own proof system; see [1]. Though the ideas involved in their proofs are very complex, they can be presented in a simple and comprehensible way.

2 Background

We assume a theory of arithmetic, say $\mathcal{N} = (\mathbb{N}, +, \times)$ to be consistent. Write $\vdash X$ for "X is a theorem in $\mathcal{N}$." The usual theorems or laws of logic hold true in this theory. We will be using explicitly the laws of Double Negation, Contradiction, Distribution of implication, Contraposition, Modus Ponens and Hypothetical Syllogism, as spelled out below.

$$\vdash \neg\neg X \leftrightarrow X. \tag{1}$$

$$\vdash X \to (\neg X \to Y). \tag{2}$$

$$\vdash (X \to (Y \to Z)) \to ((X \to Y) \to (X \to Z)). \tag{3}$$

$$\text{If } \vdash X \to \neg Y, \text{ then } \vdash Y \to \neg X. \tag{4}$$

$$\text{If } \vdash X \text{ and } \vdash X \to Y, \text{ then } \vdash Y. \tag{5}$$

$$\text{If } \vdash X \to Y \text{ and } \vdash Y \to Z, \text{ then } \vdash X \to Z. \tag{6}$$

Besides the logical laws, there are some more theorems specific to arithmetic, which are obtained by encoding formulas as natural numbers. The encoding is the so called Gödel numbers. We define Gödel number $g(\cdot)$ of symbols, formulas (in general, strings), and proofs as follows.

Enumerate the symbols such as connectives, quantifiers, punctuation marks, predicates, function symbols, and variables as:

$$\top,\ \bot,\ \neg,\ \wedge,\ \vee,\ \ \rightarrow,\ \ \leftrightarrow,\ (,\ ),\ P_1,\ f_1,\ x_1,\ P_2,\ f_2,\ x_2, \ldots$$

Define $g(\sigma) = n$, where the symbol σ comes as the n-th in the above list. Extend g to strings of these symbols by

$$g(\sigma_1\sigma_2 \cdots \sigma_m) = 2^{g(\sigma_1)} \times 3^{g(\sigma_2)} \times \cdots \times p_m^{g(\sigma_m)}$$

where p_i is the i-th prime number. This defines g of terms and formulas. Next, extend g to proofs of formulas by

$$g(X_1\ X_2\ \ldots\ X_m) = 2^{g(X_1)} \times 3^{g(X_2)} \times \cdots \times p_m^{g(X_m)}.$$

where again p_i is the i-th prime number.

Due to prime factorization theorem in $\mathcal{N}$, the function g has the following properties:

(a) g is a computable function.

(b) $g(uv)$ can be computed from those of $g(u)$ and $g(v)$.

(c) Given $n \in \mathbb{N}$, if $n = g(X)$ and X is known to be a symbol, or a formula, or a proof, then X can be computed from n.

Let $Proof(x, y)$ be a binary predicate that translates "x is the Gödel number of a proof of a formula whose Gödel number is y. Write

$$Pr(y) = \exists x Proof(x, y).$$

That is, we interpret Pr as the subset (unary relation) of $\mathbb{N}$ which is the set of Gödel numbers of all provable (in $\mathcal{N}$) formulas. $Pr(g(X))$ means that there is a natural number which is the Gödel number of a proof of a formula whose Gödel number is $g(X)$. Which, in turn, means that there is a natural number which is the Gödel number of a proof of X. We thus use a further abbreviation such as $P(X) = Pr(y) = Pr(g(X))$. The predicate Pr is the provability predicate. We may also say that P is a predicate whose arguments are formulas, and that $P(X)$ means that X is provable in $\mathcal{N}$. $P(X)$ is a formula in the theory $\mathcal{N}$. We thus loosely call P as the provability predicate.

3 Provability Predicate

The provability predicate P has the following properties:

$$\text{If } \vdash X, \text{ then } \vdash P(X). \tag{7}$$

$$\vdash P(X \to Y) \to (P(X) \to P(Y)). \tag{8}$$

$$\vdash P(X) \to P(P(X)). \tag{9}$$

Since $0 \neq 1$ in $\mathcal{N}$, $P(0 = 1)$ expresses inconsistency of $\mathcal{N}$. Therefore, consistency of $\mathcal{N}$ may be formulated by asserting that the sentence $P(0 = 1)$ is not a theorem of $\mathcal{N}$. Our assumption of consistency of $\mathcal{N}$ thus gives

$$\nvdash P(0 = 1). \tag{10}$$

Let $B_1(n)$, $B_2(n)$, $\ldots$ be an enumeration of all formulas in $\mathcal{N}$ having exactly one free variable. Consider the formula $\neg P(B_n(n))$. This is one in the above list, say $B_k(n)$. Since $\vdash p \leftrightarrow p$, we have $\vdash B_k(n) \leftrightarrow \neg P(B_n(n))$. Then. $\vdash \forall n(B_k(n) \leftrightarrow \neg P(B_n(n)))$, by universal generalization. And

$$\vdash B_k(k) \leftrightarrow \neg P(B_k(k)),$$

by universal specification with $[n/k]$. Abbreviating the sentence $B_k(k)$ to A, we obtain:

$$\vdash A \leftrightarrow \neg P(A). \tag{11}$$

The statement (11) says that the sentence "This sentence is not provable" is expressible and is a theorem in $\mathcal{N}$; ingenuity of Gödel.

For a formal proof of (11), start with the formula $B(x)$, having exactly one free variable. Let the *diagonalization* of $B(x)$ be the expression

$$\exists x(B(x) \wedge (x = g(B(x)))).$$

Since g a computable function, the relation

> $diag(m, n)$: n is the Gödel number of the diagonalization of the formula having exactly one free variable with Gödel number m

is recursive and hence representable in $\mathcal{N}$ by some binary predicate, say, $C(x, y)$. Next, define

the formula $F(x)$ as $\exists y(C(x, y) \wedge B(y))$
the sentence G as $\exists x(\exists y(C(x, y) \wedge B(y)) \wedge (x = g(F(x))))$.

Finally, show that $\vdash G \leftrightarrow B(g(G))$.

You have thus proved the Diagonalization Lemma:

> *for each formula $B(y)$ with exactly one free variable there exists a sentence G such that $\vdash G \leftrightarrow B(g(G))$.*

Next, take $B(y)$ as $\neg Pr(y)$ to obtain (11); see for example, [2].

Here are some more properties of this special sentence A.

$$\vdash P(A) \to P(\neg A) \tag{12}$$

Proof:

$\vdash A \leftrightarrow \neg P(A)$	(11)
$\vdash P(A) \to \neg A$	(4)
$\vdash P(P(A) \to \neg A)$	(7)
$\vdash P(P(A) \to \neg A) \to (P(P(A)) \to P(\neg A))$	(8)
$\vdash P(P(A)) \to P(\neg A)$	(5)
$\vdash P(A) \to P(P(A))$	(9)
$\vdash P(A) \to P(\neg A)$	(6)

The sentence A is connected to inconsistency as the following property shows:

$$\vdash P(A) \to P(0 = 1). \tag{13}$$

Proof:

$$
\begin{array}{ll}
\vdash A \to (\neg A \to (0=1)) & (2)\\
\vdash P(A \to (\neg A \to (0=1))) & (7)\\
\vdash P(A \to (\neg A \to (0=1))) & \\
\quad \to (P(A) \to P(\neg A \to (0=1))) & (8)\\
\vdash P(A) \to P(\neg A \to (0=1)) & (5)\\
\vdash P(\neg A \to (0=1)) \to (P(\neg A) \to P(0=1)) & (8)\\
\vdash P(A) \to (P(\neg A) \to P(0=1)) & (6)\\
\vdash (P(A) \to (P(\neg A) \to P(0=1))) & \\
\quad \to ((P(A) \to P(\neg A)) \to (P(A) \to P(0=1))) & (3)\\
\vdash (P(A) \to P(\neg A)) \to (P(A) \to P(0=1)) & (5)\\
\vdash P(A) \to P(\neg A) & (12)\\
\vdash P(A) \to P(0=1) & (5)
\end{array}
$$

4 Incompleteness Theorems

Theorem 1 *There exists a sentence C in $\mathcal{N}$ such that $\nvdash C$ and $\nvdash \neg C$.*

Proof: Take $C = A$, the sentence used in (11). If $\vdash A$, then by (7), $\vdash P(A)$. But by (11), $\vdash A$ implies that $\vdash \neg P(A)$. This is a contradiction. On the other hand, if $\vdash \neg A$, then by (11), $\vdash P(A)$. By (13) and (5), $\vdash P(0=1)$. This contradicts (10). Therefore, neither $\vdash A$ nor $\vdash \neg A$. □

Theorem 2 *There exists a true sentence in $\mathcal{N}$ which is not provable in $\mathcal{N}$.*

Proof: Consider the sentence A used in (11). Either A is true or $\neg A$ is true. But neither is provable by Theorem 1. Whichever of A or $\neg A$ is true serves as the sentence asked in the theorem. □

Theorem 3 $\nvdash \neg P(0=1)$.

Proof: Suppose $\vdash \neg P(0=1)$. Due to (1), (4) and (13), $\vdash \neg P(A)$. By (11), $\vdash A$. By (7), $\vdash P(A)$. By (13), $\vdash P(0=1)$. This contradicts (10). □

Notice that the sentence $\neg P(0=1)$ can also serve the purpose of the sentence asked in Theorems 1-2.

Theorems 1-2 are called as Gödel's First Incompleteness theorem; they are, in fact one theorem. Theorem 1 shows that Arithmetic is negation incomplete. Its other form, Theorem 2 shows that no axiomatic system for Arithmetic can be complete. Since axiomatization of Arithmetic is truly done in second order logic, it shows also that any axiomatic system such as Hilbert's calculus for second order logic will remain incomplete.

In (9-12), we could have used any other inconsistent sentence instead of $(0 = 1)$. Since $\neg P(0 = 1)$ expresses consistency of Arithmetic, its unprovability in Theorem 3 proves that consistency of Arithmetic cannot be proved using the proof mechanism of Arithmetic. It shattered Hilbert's program for proving the consistency of Arithmetic. Herman Wyel thus said:

> God exists since mathematics is consistent,
> and the Devil exists since its consistency cannot be proved.

References

[1] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, *Monatshefte für Mathematik und Physik* 38: 173-98, 1931.Trans: On formally undecidable propositions of Principia Mathematica and related systems I, In: Solomon Feferman, ed., 1986. Kurt Gdel Collected works, Vol. I. Oxford University Press: 144-195.

[2] A. Singh and C. Goswami, *Fundamentals of Logic,* Indian Council of Philosophical Research, New Delhi, 1998.