Gödel's First Incompleteness Theorem

Alex Edmonds

MAT 477

January 2014

Incompleteness of Peano Arithmetic (Main Theorem):

If the axioms of PA are true, then there is an $\mathcal{L}_A$-sentence G called the Gödel sentence of $\mathcal{L}_A$ such that G is true and unprovable. In particular, PA $\nvdash$ G and PA $\nvdash \neg$G.

Idea: Our proof constructs G explicitly. Loosely speaking, we want G to say, **"This sentence is unprovable."** However, $\mathcal{L}_A$-formulas can't talk about $\mathcal{L}_A$-formulas. This motivates us to 'code' $\mathcal{L}_A$-expressions into $\mathbb{N}$ using uniqe factorization into primes:

Gödel Numbers: We assign each $\mathcal{L}_A$-expression E a unique natural number $\ulcorner \mathrm{E} \urcorner$ called the *Gödel number* (shortly g.n.) of E. To do this, assign each non-variable $\mathcal{L}_A$-symbol an odd code number; the code number of the variable symbol $\mathrm{v_k}$ is 2k. For an $\mathcal{L}_A$-expression E given by the sequence of symbols $\{s_1, \ldots, s_k\}$ with corresponding code numbers $\{c_1, \ldots, c_k\}$, the Gödel number of E is defined as $\ulcorner \mathrm{E} \urcorner = \pi_1^{c_1} \pi_2^{c_2} \ldots \pi_k^{c_k}$ where π_i denotes the i^{th} prime.

e.g. Say code number of S is 7. Then, $\ulcorner \mathrm{v_3} \urcorner = 2^{2\cdot 3}$ and $\ulcorner \mathrm{Sv_3} \urcorner = 2^7 3^{2\cdot 3}$.

Super Gödel Numbers: We may also code sequences of $\mathcal{L}_A$-expressions with $\mathbb{N}$. Let $\mathsf{P} = \{\mathsf{E}_1, \ldots, \mathsf{E}_k\}$ be a sequence of $\mathcal{L}_A$-expressions with corresponding Gödel numbers $\{g_1, \ldots, g_k\}$. The *super Gödel number* (shortly sup. g.n.) of P is defined by $[\mathsf{P}] = \pi_1^{g_1}\pi_2^{g_2}\ldots\pi_k^{g_k}$.

Diagonalization: Gödel numbers allow formulas to be self-referential. In particular, for a formula φ of one free variable, define its diagonalization φ_d as $\varphi(\ulcorner\varphi\urcorner)$. However, we want more: formulas need to be able to express properties of formulas via their Gödel numbers.

Expressability: A relation R on $\mathbb{N}^k$ is called *expressible in $\mathcal{L}_A$* if there exists an $\mathcal{L}_A$-formula $\varphi(\mathsf{v}_1, \ldots, \mathsf{v}_\mathsf{k})$ such that, for all $n_1, \ldots, n_k \in \mathbb{N}$,

- if $R(n_1, \ldots, n_k)$, then $\varphi(\bar{\mathsf{n}}_1, \ldots, \bar{\mathsf{n}}_\mathsf{k})$ is true;
- if not $R(n_1, \ldots, n_k)$, then $\neg\varphi(\bar{\mathsf{n}}_1, \ldots, \bar{\mathsf{n}}_\mathsf{k})$ is true.

Say function $f : \mathbb{N}^k \to \mathbb{N}$ is expressible if the relation $f(n_1, \ldots, n_k) = n_{k+1}$ is expressible.

e.g. The function $x - y$ is expressed by $\mathsf{x} = \mathsf{y} + \mathsf{z}$.

Main Claim: The following functions and relations are expressible:

- function $diag(n)$: satisfies, for every $\mathcal{L}_A$-formula φ, $diag(\ulcorner\varphi\urcorner) = \ulcorner\varphi_d\urcorner$
- rel. $Prf(m, n)$: holds when $m =$ [P], $n = \ulcorner\varphi\urcorner$ for some proof P of φ
- rel. $Gdl(m, n)$: holds when $m =$ [P], $n = \ulcorner\varphi\urcorner$ for some proof P of φ_d

We'll return to a proof of this later on. First, the heart of it all:

Main Claim $\Rightarrow$ Main Theorem

Proof: By Claim, there exists an $\mathcal{L}_A$-formula Gdl(x, y) which expresses

Gdl. Let T(y) be the formula $\forall x \neg$Gdl(x, y). Let G be the diagonalization of T(y), namely $\forall x \neg \mathrm{Gdl}(x, \ulcorner T \urcorner)$. Then,

G false $\Leftrightarrow$ There exists $m \in \mathbb{N}$ such that $Gdl(m, \ulcorner T \urcorner)$

$\Leftrightarrow$ There exists $m \in \mathbb{N}$ such that, for some finite sequence P of $\mathcal{L}_A$-formulas, $m = \ulcorner P \urcorner$ and P is a proof of diag. of T, namely G

$\Leftrightarrow$ G is provable

(Compare to "This sentence is unprovable.") Now, by soundness, G provable $\Rightarrow$ G is true. Therefore, G is true and unprovable. By soundness again, $\neg$G is also unprovable. $\square$

Primitive recursive functions

Def: f *is defined from* g *and* h *by primitive recursion* if:

- $f(\vec{x}, 0) = g(\vec{x})$
- $f(\vec{x}, Sy) = h(\vec{x}, y, f(\vec{x}, y))$

Def: The *primivitive recursive* (shortly p.r.) functions are:

- The **initial functions**, namely the successor function S, the zero functions Z^k, and the co-ordinate functions I_i^k;
- Any composition of p.r. functions;
- Any function defined by primitive recursion from p.r. functions.

Proposition: $\{$p.r. functions$\} \subset \{$expressible functions$\}$

Proof: It suffices to show

(1) $\mathcal{L}_A$ expresses the initial functions;

(2) $\mathcal{L}_A$ expresses g and $h \Rightarrow \mathcal{L}_A$ expresses any composition of g and h;

(3) $\mathcal{L}_A$ expresses g and $h \Rightarrow \mathcal{L}_A$ expresses any function defined by

primitive recursion from g and h.

Proving **(1)** is trivial: S is expressed by $\mathsf{Sx} = \mathsf{y}$; Z^k by $\mathsf{y} = 0$; I_i^k by $\mathsf{v_i} = \mathsf{y}$.

(2): If functions $g(x)$ and $h(x)$ are expressed by G(x,y) and H(x,y), then their composition $f(x) = h(g(x))$ is expressed by $\exists \mathsf{z}(\mathsf{G}(\mathsf{x}, \mathsf{z}) \wedge \mathsf{H}(\mathsf{z}, \mathsf{y}))$.

(3) is tricky and relies on a sort of Gödel numbering again. Suppose $\mathsf{H}(\mathsf{x}, \mathsf{y})$ expresses $h(x) = y$. Let $f(x)$ be defined primitive recursively by $f(0) = a$ and $f(Sx) = h(f(x))$. Note that $f(x) = y$ iff

> **(A)** There is a sequence of numbers $k_0, k_1, \ldots, k_x$ such that:
>
> $k_0 = a$; for $u < x$, $k_{Su} = h(k_u)$; and $k_x = y$.

For any such sequence $k_0, k_1, \ldots, k_n$, we wish to encode it by a pair of numbers c and d. We need a function $\beta(c, d, i) = i^{th}$ element of the sequence coded by c and d. A result from number theory (see Appendix) tells us that, for every sequence $k_0, \ldots, k_n$, there exists c, d s.th., for all $i \leq n$, k_i is the remainder of c divided by $d(i+1)+1$. Hence, taking $\beta(c, d, i)$ to be said remainder, **(A)** may be reformulated as

> **(B)** There exists c, d such that: $\beta(c, d, 0) = a$; For $u < x$, $\beta(c, d, Su) = h(\beta(c, d, u))$; And $\beta(c, d, x) = y$

Easy fact: The remainder function is expressible.

It follows β is expressible. Say B(c,d,i,k) expresses $\beta(c, d, i) = k$. Then,

(B) may be translated into $\mathcal{L}_A$ as

(C) $\exists c \exists d\{B(c,d,0,\bar{a}) \wedge (\forall u \leq x)$

$[u \neq x \rightarrow \exists v \exists w\{B(c,d,u,v) \wedge B(c,d,Su,w) \wedge H(v,w)\}] \wedge B(c,d,x,y)\}$

We may conclude that f is expressible in $\mathcal{L}_A$. This argument generalizes easily to the multivariable case.$\square$

Constructing our p.r. fucntions – a sketch

It remains to show that the functions and relations we need are p.r. This is a large but mostly mechanical task aided by the following.

Four useful facts: (1) $f(\vec{x})$ p.r. function $\Rightarrow f(\vec{x}) = y$ is a p.r. relation

(2) 'Truth-functional combinations' (conjunction, implication etc.) of p.r. relations are p.r.

(3) A relation defined from a p.r. relation by bounded quantification is p.r.

(4) Functions defined by p.r. cases from p.r. functions are p.r.

Claim: Relation $Term(n)$, which holds when $n = \ulcorner \tau \urcorner$ for a term τ, is p.r.

Our proof uses the following useful notion:

Def: A *term-sequence* is a finite sequence of expressions such that each is either:

a. 0;

b. a variable symbol;

c. Sτ where τ is an expression which appears earlier in the sequence;

d. $(\tau_1 + \tau_2)$ where τ_1 and τ_2 are earlier expressions in the sequence; or

e. $(\tau_1 \times \tau_2)$ where τ_1 and τ_2 are earlier expressions in the sequence.

Note that an expression is a term iff it is the last expression in some term sequence. Before proceeding with Claim, we need:

Lemma 1: relation $Var(n)$, which holds iff n is g.n. of a variable, is p.r.

Lemma 2: relation $Termseq(m, n)$, which holds iff $m =$ [T] and $n = \ulcorner\tau\urcorner$ for a term-sequence T ending in τ, is p.r.

Basic Fact: The following functions are primitive recursive:

- $prime(i)$, returns the i^{th} prime
- $len(n)$, returns the number of distinct prime factors of n. In particular, $len(\ulcorner \mathsf{E} \urcorner)$ is length of $\mathcal{L}_A$-expression E; $len([\mathsf{T}])$ length of term-sequence T
- $exp(n, i)$, returns degree of i^{th} prime in factorization of n. **Importantly**, for sequence of expressions E, $exp([\mathsf{E}], i)$ returns g.n. of i^{th} expression in E
- $m * n$, which returns the g.n. of the concactenation of the expression with g.n. m and the expression with g.n. n.

Proof of Lemma 1 (i.e. that $Var(n)$ is p.r.)

The g.n. of the variable $\mathsf{v_k}$ is 2^{2k}. Hence, $Var(n) \Leftrightarrow \exists k(n = 2^{2k})$.

However, unbounded quantification is not necessarily p.r.-preserving. This is dealt with by noting that $n = 2^{2k} \Rightarrow k < n$. Hence, $Var(n) \Leftrightarrow (\exists k < n)(n = 2^{2x})$ which is constructed in p.r. preserving ways. Therefore, $Var(n)$ is p.r. $\square$

Proof of Lemma 2 (i.e. that $Termseq(m, n)$ is p.r. – easy but technical)

$Termseq(m, n)$ is equivalent to the statement

(1) $exp(m, len(m)) = n$; and

(2) for $1 \leq k \leq len(m)$:

a' $exp\{m, k\} = \ulcorner 0 \urcorner$; or

b' $Var(exp(m, k))$; or

c' $(\exists j < k)(exp(m, k) = \ulcorner \mathsf{S} \urcorner * exp(m, j))$; or

d' $(\exists i < k)(\exists j < k)(exp(m, k) = \ulcorner (\urcorner * exp(m, j) * \ulcorner + \urcorner * exp(m, j) * \ulcorner) \urcorner)$

or

e' $(\exists i < k)(\exists j < k)(exp(m,k) = \ulcorner(\urcorner * exp(m,j) * \ulcorner\times\urcorner * exp(m,j) * \ulcorner)\urcorner)$

Indeed, **(1)** guarantees that the sequence with super g.n. m ends with the expression with g.n. n. Also, **(2)** guarrantees that m is actually the g.n. of a term sequence. In particular, **a'**, **b'**, **c'**, **d'**, **e'** correspond to **a**, **b**, **c**, **d**, **e**, resp. Since the relation above is constructed in p.r.-preserving ways from p.r. functions and relations, conclude $Term(m,n)$ is p.r.

Proof of Claim: $Term(n) \Leftrightarrow \exists m\ Termseq(m, n)$, but again we need quantification to be bounded. Suppose $[\mathsf{T}] = \pi_1^{d_1} \dots \pi_k^{d_k}$ for a term-sequence T of τ. Length of T is bounded by length of τ, i.e. $k \leq len(n)$. Also, $[\mathsf{T}] \leq \pi_k^{k \max_i d_i}$ and $\max_i d_i \leq n$. Therefore, $Term(n) \Leftrightarrow (\exists m \leq prime(len(n))^{n(len(n))})\ Termseq(m, n)$. Since the relation is constructed from p.r. relations in p.r. preserving ways, it is p.r.

$\square$

Using construction histories: Note that our definitions of *term-sequence* and *term* aren't very different from our definitions of *proof* an *provability*. Not suprisingly then, the proof that $Prf(m, n)$ is p.r. is very close to the proof for $Term(m, n)$. Actually, most of the important proofs of primivitive recursiveness that we want (i.e. for formulas, sentences, axioms) follow the same structure.

Sketch that Prf(m,n) is p.r.

Fact: We'll omit a proof that following relations are p.r.:

- $Sent(n)$, holds when $n = \ulcorner\varphi\urcorner$ for some sentence φ;
- $Axiom_{PA}(n)$, holds when $n = \ulcorner\varphi\urcorner$ for some axiom φ of PA;
- $Ded(l, m, n)$, holds when $l = \ulcorner\varphi\urcorner$, $m = \ulcorner\psi\urcorner$, $n = \ulcorner\gamma\urcorner$, where sentence γ

follows by rule of deduction from sentences φ and ψ.

Then, $Prf(m,n)$ is equivalent to the statement:

$exp(m, len(m)) = n$; and $Sent(n)$; and $\forall k \leq len(m)\{$

$Axiom_{PA}(exp(m,k))$, or

$(\exists i \leq k)(\exists j < k)[Ded(exp(m,i), exp(m,j), exp(m,k)]\ \}$

which is constructed from p.r. functions in p.r.-preserving ways. This result completes part of the proof of Main Claim. We'll omit the rest.

Important Note: The only facts about PA we use in the proof of Main Theorem are to show that $Axiom_{PA}(n)$ is primitive recursive!!

Generalizing the incompleteness argument

We have shown that PA is incomplete. The obvious question is whether PA is 'completable'. That is to say, can we add axioms to PA so that, for every sentence φ, either φ or $\neg\varphi$ is provable? **No!** Since the only fact used about PA is that $Axiom_{PA}(n)$ is p.r., our argument holds for any theory T where $Axiom_T(n)$ is p.r. We call such a theory *p.r. axiomatized*.

Gödel's First Incompleteness Theorem (Semantic Version)

Let T be a theory whose language includes $\mathcal{L}_A$. Suppose T is p.r.

axiomatized and all $\mathcal{L}_A$ axioms of T are true. Then, there exists an

$\mathcal{L}_A$-sentence φ such that $\mathrm{T} \nvdash \varphi$ and $\mathrm{T} \nvdash \neg\varphi$.

Appendix: Beta-function

Def: Let $rem(c, d)$ denote the remainder when c is divided by d.

Def: For $\mathbf{d} = \langle d_0, \ldots, d_n \rangle \in \mathbb{N}^n$, $c \in \mathbb{N}$, let $Rm(c, \mathbf{d}) = \langle k_0, \ldots, k_n \rangle$ where

$k_i = rem(c, d_i)$.

β-**Theorem:** For every sequence $\mathbf{k} \in \mathbb{N}^n$, there exists $c, d \in \mathbb{N}$ such that

$Rm(c, \mathbf{d}) = \mathbf{k}$, where $d_i = d(i+1) + 1$. In particular, $\forall i \leq n$

$\beta(c, d, i) := rem(c, d(i+1)+1) = k_i$.

To prove this, we need:

Chinese Remainder Theorem: Let $\mathbf{d} = \langle d_0, \ldots, d_n \rangle$ and suppose all d_i are relatively prime. Then, for distinct $c_1, c_2 < |\mathbf{d}| := d_0 \times \ldots \times d_n$, $Rm(c_1, \mathbf{d}) \neq Rm(c_2, \mathbf{d})$.

Proof: By way of contradiction, assume $Rm(c_1, \mathbf{d}) = Rm(c_2, \mathbf{d})$. Let $c = |c_1 - c_2|$, Then each d_i divides c. Since d_i are relatively prime, this implies $|\mathbf{d}| = d_0 \times \ldots \times d_n$ divides c. Hence, $|\mathbf{d}| \leq c \leq max\{c_1, c_2\}$. $\Box$

Proof of β-Theorem

Step 1: Let $s = \max\{n, k_1, \ldots, k_n\}$ and let $d = s!$. For $i \leq n$, the numbers $d_i := d(i+1)+1$ are relatively prime. Indeed, suppose otherwise. Then, there exists distinct $i, j \leq n$ s.th. both $d(i+1)+1$ and $d(j+1)+1$ are divisible by p. In particular, p divides $d|i-j|$. Also, since p divides $d(i+1)+1$, p does not divide d. Hence, p divides $(i-j) \leq n \leq s$, i.e. $p \leq s$. But, if p doesn't divide $d = s!$, then $p > s$. Contradiction. Conclude d_i's are relatively prime.

Step 2: Note that, for all c, $Rm(c, \mathbf{d}) \in S_1 \times \cdots \times S_n$ where

$S_i = \{0, 1, \ldots, d_i - 1\}$. Furthermore, $|S_1 \times \cdots \times S_k| = d_0 \ldots d_n = |d|$.

The Chinese Remainder Theorem says each $c < |d|$ is mapped to a

distinct element, i.e. $Rm(c, \mathbf{d})$ takes on $|d|$ many values for $c < |d|$.

Therefore, $Rm(c, \mathbf{d})$ takes on each $\mathbf{a} \in S_1 \times \cdots \times S_n$ for $c < |d|$. Since

$\mathbf{k} \in S_1 \times \cdots \times S_n$, this completes the proof. $\square$