On limits of applicability of Gödel's second incompleteness theorem

F.N. Pakhomov
Steklov Mathematical Institute, Moscow
pakhf@mi-ras.ru

PDMI Logic Seminar,
March 06, 2019

Peano arithmetic

Robinson's arithmetic Q:

1. $S(x) \neq 0$;
2. $S(x) = S(y) \rightarrow x = y$;
3. $x \leq 0 \leftrightarrow x = 0$;
4. $x \leq S(y) \leftrightarrow x \leq y \vee x = S(y)$;
5. $x + 0 = x$;
6. $x + S(y) = S(x + y)$;
7. $x0 = 0$;
8. $x(Sy) = xy + x$.

PA $=$ Q $+$ the following scheme:

$$\varphi(0) \wedge \forall x\, (\varphi(x) \rightarrow \varphi(Sx)) \rightarrow \forall x \varphi(x).$$

First incompleteness theorem

Theorem (Gödel'1931)

Suppose c.e. theory T contains PA *and is arithmetically sound (e.g. it doesn't prove false sentences of first-order arithmetic). Then there is a sentence φ such that $T \nvdash \varphi$ and $T \nvdash \neg\varphi$.*

Note: Actually Gödel worked over much stronger formal theory P that was a variant of Principia Mathematica system. It contained higher types, but it wasn't important for Gödel's argument. Also Gödel used the notion ω-consistency instead of soundedness.

Theorem (Rosser'36; Tarski, Mostowski, Robinson'53)

Suppose $T \supseteq \mathrm{Q}$ and T is consistent. Then there is a sentence φ such that $T \nvdash \varphi$ and $T \nvdash \neg\varphi$.

Formalization of provability

We encode formulas by numbers:

string in finite alphabet $\varphi \longmapsto$ binary string α encoding $\varphi \longmapsto$ number n which binary expansion is 1α.

For a formula φ, the expression $\ulcorner\varphi\urcorner$ is the term $S^n(0)$, where n is the number corresponding to φ.

Recall that Hilbert-style proof is a list of formulas, where each formula is either an axiom or is a result of application of an inference rule to some preceding formulas.

For a given c.e. theory T we have predicate $\mathrm{Prf}_T(x, y)$:

"number x encodes some proof in the theory T and the last formula in it is y."

$\mathrm{Prv}_T(x)$ is the formula $\exists y\, \mathrm{Prf}_T(y, x)$.

Second incompleteness theorem

The consistency assertion $\mathrm{Con}(T)$ is $\neg\mathrm{Prv}_T(\ulcorner 0 = S0 \urcorner)$.

Theorem (Gödel'31)

Suppose c.e. theory $T \supseteq \mathrm{PA}$ *and* T *is consistent. Then* $T \nvdash \mathrm{Con}(T)$.

Note: In this case Göodel also considered extensions of system P. Instead of c.e. extensions he considered extensions by primitive recursive sets of axioms.

Hilbert-Bernays-Löb derivability conditions

Abbreviations:

- $\Box_T \varphi$ is an abbreviation for $\mathrm{Prv}_T(\ulcorner \varphi \urcorner)$;
- $\Diamond_T \varphi$ is an abbreviation for $\neg \mathrm{Prv}_T(\ulcorner \neg \varphi \urcorner)$;
- $\bot$ is an abbreviation for $0 = S(0)$;
- $\top$ is an abbreviation for $0 = 0$;

Note that $\mathrm{Con}(T)$ is $\Diamond \top$.

Hilbert-Bernays-Löb derivability conditions:

HBL-1 $T \vdash \varphi \quad \Rightarrow \quad T \vdash \Box_T \varphi$;

HBL-2 $T \vdash \Box_T(\varphi \to \psi) \to (\Box_T \varphi \to \Box_T \psi)$;

HBL-3 $T \vdash \Box_T \varphi \to \Box_T \Box_T \varphi$.

Theorem (Löb'55)

Suppose c.e. theory $T \supseteq \mathrm{Q}$, T is consistent and the predicate Prv_T satisfies HBL conditions. Then $T \nvdash \mathrm{Con}(T)$.

Fixed-point lemma

Lemma (Gödel'31)

For any formula $\varphi(x)$ *there is a sentence* ψ *such that*

$$\mathsf{Q} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

Proof:

$$\mathsf{subst}_x : \langle \ulcorner \varphi(x) \urcorner, \ulcorner \psi \urcorner \rangle \longmapsto \ulcorner \varphi(\ulcorner \psi \urcorner) \urcorner.$$

$$\text{For all } \varphi, \psi\text{: } \mathsf{Q} \vdash \mathsf{subst}_x(\ulcorner \varphi(x) \urcorner, \ulcorner \psi \urcorner) = \ulcorner \varphi(\ulcorner \psi \urcorner) \urcorner.$$

Let $\chi(x)$ be $\varphi(\mathsf{subst}_x(x, x))$. We put ψ to be $\chi(\ulcorner \chi(x) \urcorner)$.
Observe that

$$\begin{aligned} \mathsf{Q} \vdash \psi &\leftrightarrow \chi(\ulcorner \chi(x) \urcorner) \\ &\leftrightarrow \varphi(\mathsf{subst}_x(\ulcorner \chi(x) \urcorner, \ulcorner \chi(x) \urcorner)) \\ &\leftrightarrow \varphi(\ulcorner \chi(\ulcorner \chi(x) \urcorner) \urcorner) \\ &\leftrightarrow \varphi(\ulcorner \psi \urcorner). \end{aligned}$$

Proof of second incompleteness theorem

Let ψ be such that $\mathsf{Q} \vdash \psi \leftrightarrow \neg\Box_T\psi$.

We reason in T:

1. $\bot \to \varphi$;
2. $\Box_T(\bot \to \varphi)$ (HBL-1);
3. $\Box_T\bot \to \Box_T\varphi)$ (HBL-2);
4. $\Box_T\varphi \to \Box_T\Box_T\varphi$ (HBL-3);
5. $\Box_T\varphi \to \Box_T\neg\Box_T\varphi$ (fixed-point property of φ);
6. $\Box_T\varphi \to \Box_T\bot$ (4., 5., and HBL-1+HBL-2);
7. $\Box_T\varphi \leftrightarrow \Box_T\bot$;
8. $\neg\Box_T\varphi \leftrightarrow \neg\Box_T\bot$;
9. $\varphi \leftrightarrow \Diamond_T\top$.
10. $\Diamond_T\top \leftrightarrow \neg\Box_T\Diamond_T\top$.

If $T \vdash \Diamond_T\top$ then $T \vdash \neg\Box_T\Diamond_T\top$ (by 10.) and $T \vdash \Box_T\Diamond_T\top$ (by HBL-1), hence T is inconsistent.

Proving HBL conditions

Δ_0 formulas are formulas built of propositional connectives and bounded quantifiers $\forall x \leq t$ and $\exists x \leq t$ (here $x \notin \mathrm{FV}(t)$).
Σ_1 formulas are $\exists \vec{x}\, \varphi$, where φ is Δ_0.

Note that $\Box_T \varphi$ is a Σ_1 sentence.
HBL-1: $T \vdash \varphi \Rightarrow T \vdash \Box_T \varphi$.

Lemma

If φ is a true Σ_1 sentence then $\mathrm{Q} \vdash \varphi$.

HBL-2: $T \vdash \Box_T(\varphi \to \psi) \to (\Box_T \varphi \to \Box_T \psi)$.
To prove this T should be able to concatenate proofs of $\varphi \to \psi$ and φ and add formula ψ at the end.

HBL-3: $T \vdash \Box_T \varphi \to \Box_T \Box_T \varphi$.
This requires formalization of HBL-1 in T. To prove the lemma inside T we need to transform a proof p of φ into a proof q of the fact that p is a proof of φ. Note that $|q|$ is polynomial in $|p|$.

Theory $I\Delta_0 + \Omega_1$

$I\Delta_0 = Q +$ the following scheme:

$$\varphi(0) \wedge \forall x\,(\varphi(x) \to \varphi(Sx)) \to \forall x \varphi(x), \text{ where } \varphi \text{ is } \Delta_0.$$

The length $|x| = \lceil \log_2(x) \rceil = \min\{y \mid \exp(y) \geq x\}$.

Smash function: $x \# y = 2^{|x||y|}$.

Axiom Ω_1 is $\forall x, y \exists z\,(x \# y = z)$.

Proposition

If $T \supseteq I\Delta_0 + \Omega_1$ is $\mathtt{NP}$-axiomatizable theory. Then HBL conditions hold for T with the natural provability predicate for it.

Corollary

If $T \supseteq I\Delta_0 + \Omega_1$ is $\mathtt{NP}$-axiomatizable consistent theory. Then $T \nvdash \mathrm{Con}(T)$.

Pudlak's version of second incompleteness theorem

Theorem (Pudlak'85)

If $T \supseteq \mathrm{Q}$ is c.e. consistent theory. Then $T \nvdash \mathrm{Con}(T)$.

Idea of proof (part 1):

A T-cut $J(x)$ is a formula such that

$$T \vdash J(0) \land \forall x \, (J(x) \to (\forall y \leq S(x))J(y)).$$

A T-cut $J(x)$ is called closed under the function $f(x_1, \dots, x_k)$ if

$$T \vdash \forall x_1, \dots, x_k \, (J(x_1) \land \dots \land J(x_k) \to J(f(x_1, \dots, x_k)).$$

For a fornmula φ we denote by φ^J the result of replacement of each quantifier $\forall x \, \varphi$ with the quantifier $\forall x \, (J(x) \to \varphi)$ and each quantifier $\exists x \, \varphi$ with the quantifier $\exists x \, (J(x) \land \varphi)$.
For T-cuts $J(x)$ that are closed under $+$ and $\cdot$ we have absoluteness for Δ_0 formulas:

$$T \vdash \forall \vec{x}(\varphi(\vec{x}) \leftrightarrow (\varphi(\vec{x}))^J), \text{ for } \Delta_0 \text{ formulas } \varphi.$$

Pudlak's version of second incompleteness theorem

Theorem

If $T \supseteq \mathrm{Q}$ is c.e. consistent theory. Then $T \nvdash \mathrm{Con}(T)$.

Idea of proof (part 2):

Lemma

In Q there is a cut $I(x)$ that is closed under $+, \cdot,$ and $\#$ and

$$\mathrm{Q} \vdash \varphi^I, \textit{ for any axiom } \varphi \textit{ of } \mathrm{I}\Delta_0 + \Omega_1.$$

Assume for a contradiction that $T \vdash \mathrm{Con}(T)$. By Δ_0 absoluteness, $T \vdash (\mathrm{Con}(T))^I$. Let U be theory with $\mathtt{NP}$ axiomatization

$$\{\underbrace{\varphi \wedge \ldots \wedge \varphi}_{|p| \text{ times}} \mid p : T \vdash \varphi^I\}.$$

It is easy to see that $\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Con}(T) \rightarrow \mathrm{Con}(U)$. Thus $U \vdash \mathrm{Con}(U)$, since $U \supseteq \mathrm{I}\Delta_0 + \Omega_1$ we get to a contradiction.

Weak set theory H.

Let us consider theory H in the language of set theory with additional unary function $\overline{\mathrm{V}}$:

1. $\forall z\,(z \in x \leftrightarrow z \in y) \rightarrow x = y$ (Extensionality);
2. $\exists y \forall z\,(z \in y \leftrightarrow z \in x \wedge \varphi(z))$ (Separation);
3. $y \in \overline{\mathrm{V}}(x) \leftrightarrow \exists z \in x\,(y \subseteq \overline{\mathrm{V}}(z))$.

Note that the last axiom essentially states

$$\overline{\mathrm{V}}(x) = \bigcup_{z \in x} \mathcal{P}(\overline{\mathrm{V}}(z)).$$

In ZFC cummulative hierarchy V_α, for $\alpha \in \mathrm{On}$:

- $V_0 = \emptyset$;
- $V_{\alpha+1} = \mathcal{P}(V_\alpha)$;
- $V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$, for $\lambda \in \mathrm{Lim}$.

It is easy to see that

$$\overline{\mathrm{V}} \colon x \longmapsto V_\alpha, \text{ where } \alpha \text{ is least such that } x \subseteq V_\alpha.$$

It is easy to prove that the models of second-order version of H up to isomorphism are $(V_\alpha, \in, \overline{\mathrm{V}})$.

Embedding of arithmetic in H

We make some standard definitions in H:

1. $x \in \mathsf{Trans} \overset{\text{def}}{\iff} \forall y \in x\, (y \subseteq x)$;
2. $x \in \mathsf{On} \overset{\text{def}}{\iff} x \in \mathsf{Trans} \wedge \forall y \in x\, (y \in \mathsf{Trans})$;
3. $x \leq y \overset{\text{def}}{\iff} x \in \mathsf{On} \wedge y \in \mathsf{On} \wedge (x \in y \vee x = y)$;
4. $\alpha = S(\beta) \overset{\text{def}}{\iff} \alpha \in \mathsf{On} \wedge \beta \in \mathsf{On} \wedge (\forall \gamma \in \mathsf{On})(\gamma \in \beta \leftrightarrow \gamma \in \alpha \vee \gamma = \alpha)$;
5. $\alpha \in \mathsf{Nat} \overset{\text{def}}{\iff} \alpha \in \mathsf{On} \wedge (\forall \beta \leq \alpha)(\beta = \emptyset \vee \exists \gamma\, (\beta = S(\gamma)))$.

Note that however we couldn't prove totality of successor function in H.

We define partial functions $+ \colon \mathsf{On} \times \mathsf{On} \to \mathsf{On}$ and $\times \colon \mathsf{On} \times \mathsf{On} \to \mathsf{On}$ such that

- $\alpha + \beta = \bigcup \{ S(\alpha + \gamma) \mid \gamma < \beta \}$;
- $\alpha\beta = \bigcup \{ \alpha\gamma + \alpha \mid \gamma < \beta \}$.

In the equalities above the left part should be defined whenever the right part is defined.

H and $\mathsf{H}_{<\omega}$ are non-Gödelian

Theory $\mathsf{H}_{<\omega}$ is an extension of H by the infinite series of axioms $\exists x\, \mathsf{Nmb}_n(x)$ stating that all individual natural numbers n exist

$$\mathsf{Nmb}_0(x) \stackrel{\text{def}}{\Longleftrightarrow} (\forall y \in x) y \neq y,$$

$$\mathsf{Nmb}_{n+1}(x) \stackrel{\text{def}}{\Longleftrightarrow} \exists y\, (\mathsf{Nmb}_n(y) \wedge \forall z\, (z \in x \leftrightarrow z \in y \vee z = y).$$

Note that the theory $\mathsf{H}_{<\omega}$ could prove existence of all the individual hereditary finite sets.
Since our interpretation of arithmetical functions isn't total, we naturally switch to the predicate only arithmetical signature:

$$x = y, \quad x \leq y, \quad x = S(y), \quad x = y + z, \quad x = yz.$$

We could naturally express $\mathsf{Prf}_{\mathsf{H}_{<\omega}}(x, y)$ by a predicate-only Σ_1 formula. And $\mathsf{Con}(\mathsf{H}_{<\omega})$ by a Π_1 predicate-only formula.

Theorem

Theory H *proves* $\mathsf{Con}(\mathsf{H}_{<\omega})$.

Idea of proof of non-Gödelian property for $\mathrm{H}_{<\omega}$

Argument outside of specific formal theory:

To prove consistency of $\mathrm{H}_{<\omega}$ one could assume for a contradiction that there is a $\mathrm{H}_{<\omega}$ proof p of $\exists x\ x \neq x$. We consider number n_p that is the maximum of all n s.t. the axiom $\exists x\ \mathrm{Nmb}_n(x)$ appear in p. Next we show that $(V_{n_p+1}, \in, \overline{\mathrm{V}})$ is a model of all the axioms that appear in p and hence p couldn't exist.

Idea of proof of non-Gödelian property for $\mathrm{H}_{<\omega}$

Intuition of why $\mathrm{H} \vdash \mathrm{Con}(\mathrm{H}_{<\omega})$:

The number $n_p \leq \lfloor p/2 \rfloor$ (moreover $n_p \leq \lfloor \log_2(p) \rfloor$).
Hence for large enough p, from mere presence of a proof p we could conclude that there is model $(V_{n_p+1}, \in, \overline{\mathrm{V}})$ with a given iteration of powerset on top of it. It is enough to formalize the argument that there p isn't a proof of inconsistency.

Conservation result between EA and $\mathrm{H}_{>\omega}$

EA is Kalmar elementary functions arithmetic. It is the variant of $\mathrm{I}\Delta_0$ in the language with binary exponentiation function $\exp(x)$.

Lemma

Let $S(x)$ be superexponential cut in EA, e.g.

$$S(x) \stackrel{\text{def}}{\iff} \text{“}\underbrace{2^{\cdot^{\cdot^{\cdot^{2}}}}}_{n\ \textit{times}} \textit{ is defined.}$$

Let Nat^{-n} be the class in H that consists of all x s.t. $S^n(x)$ is defined. For each predicate-only Π_1 sentence φ of the form $\forall\vec{x}\,\psi(\vec{x})$, where ψ is Δ_0:

$$\mathrm{EA} \vdash \varphi^S \iff \mathrm{H} \vdash \forall\vec{x}(\vec{x} \in \mathrm{Nat}^{-n} \to \psi(\vec{x})), \textit{ for some } n.$$

Thank you!