Löb's Theorem in Coq

First Bachelor Seminar Talk

Janis Bailitis

Universität des Saarlandes
Advisors: Dr. Yannick Forster, Dr. Dominik Kirst
Supervisor: Prof. Dr. Gert Smolka
Programming Systems Lab

April 30, 2024

What is Löb's theorem?

Problem (Henkin, 1952)

- Assume sufficiently strong formal system, e.g. Peano arithmetic
- There is a sentence S expressing own provability

Question: S independent or provable?

- Kreisel 1953: It depends on provability predicate
  - ➤ Only inspected restricted set of provability predicates
- Löb 1955: S is provable
  - ➤ But for a strong notion of provability predicate

Provability

Intuitively, $\mathrm{Prov}(x)$ expresses provability in T if for all φ, $T \vdash \varphi \leftrightarrow T \vdash \mathrm{Prov}(\overline{\varphi})$[1,2]

Theorem (Löb's theorem, 1955)

Let $\mathrm{Prov}(x)$ *express provability in* T. *For all sentences* φ, *we have*

$$(T \vdash \mathrm{Prov}(\overline{\varphi}) \dot{\to} \varphi) \to (T \vdash \varphi).$$

- Generalises Gödel's second incompleteness theorem
- Relevant beyond pure logic
  - ➤ Program verification: Assume property holding **later**, i.e. at a lower step index
- Agda-mechanisation by Gross et al. 2016
  - ➤ Using Curry-Howard and quines

[1] This is the notion Kreisel used in 1953.
[2] We assume that $\overline{\varphi}$ is some encoding of φ as term.

Provability

Intuitively, $\mathrm{Prov}(x)$ expresses provability in T if for all φ, $T \vdash \varphi \leftrightarrow T \vdash \mathrm{Prov}(\overline{\varphi})$[1][2]

Theorem (Löb's theorem, 1955)

Let $\mathrm{Prov}(x)$ *express provability in* T*. For all sentences* φ*, we have*

$$(T \vdash \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \varphi) \rightarrow (T \vdash \varphi).$$

- Intuitive notion: Too weak
- Sufficiently strong formulae $\rightarrow$ high technical overhead
- Can we do this more abstractly?
  - Löb isolated abstract axioms
  - Assuming them, proof is mechanical and short; also in Coq

[1] This is the notion Kreisel used in 1953.

[2] We assume that $\overline{\varphi}$ is some encoding of φ as term.

Löb's axioms

We use Peano arithmetic (PA).

Löb's axioms (cf. [BBJ07], slightly more general)

Suppose $\mathrm{Prov}(x) : \mathbb{F}$, and φ, ψ any sentence. $\mathrm{Prov}(x)$ is a modality satisfying

- **necessitation** if $\mathrm{PA} \vdash \varphi$ implies $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi})$
- **the modal fixed point theorem** if for any $F(x) : \mathbb{F}$ we find $\tau : \mathbb{F}$ such that $\mathrm{PA} \vdash \tau \dot{\leftrightarrow} F(\mathrm{Prov}(\overline{\tau}))$
- **internal necessitation** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\mathrm{Prov}(\overline{\varphi})})$
- **the distributivity law** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi \dot{\rightarrow} \psi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\psi})$

- Goal: Use Church's Thesis for arithmetic ($\mathrm{CT}_{\mathrm{PA}}$)[1]
  - ➤ Gives more abstract formula
  - ➤ Investigate which axioms hold

[1] Formalised for first-order arithmetic by Hermes and Kirst 2022, proven consistent by Kirst and Peters 2023.

Defining a provability candidate

Axiom ($\mathsf{CT}_{\mathsf{PA}}$)

Let $f : \mathbb{N} \to \mathbb{N}$ be a function.
There is a formula $\varphi(x, y)$ such that for all $n : \mathbb{N}$ we have

$$\mathsf{PA} \vdash \dot{\forall} x.\, \varphi(\overline{n}, x) \mathrel{\dot{\leftrightarrow}} x \equiv \overline{f\ n}.$$

Lemma (Weak representability, cf. [HK23])

Suppose $P : \mathbb{N} \to \mathbb{P}$ is enumerable.
There is a formula $\varphi(x)$ such that for all $n : \mathbb{N}$ we have

$$P\ n \leftrightarrow \mathsf{PA} \vdash \varphi(\overline{n}).$$

Now inspect $\lambda\varphi.\, \mathsf{PA} \vdash \varphi$, enumerable by [FKS19]

Corollary

We find a formula $\mathsf{Prov}(x)$ such that for all $\varphi : \mathbb{F}$ we have $\mathsf{PA} \vdash \varphi \leftrightarrow \mathsf{PA} \vdash \mathsf{Prov}(\overline{\varphi})$.

Löb's axioms revisited

Löb's axioms

Suppose $\mathrm{Prov}(x) : \mathbb{F}$, and φ, ψ any sentence. $\mathrm{Prov}(x)$ is a modality satisfying

- **necessitation** if $\mathrm{PA} \vdash \varphi$ implies $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi})$
- **the modal fixed point theorem** if for any $F(x) : \mathbb{F}$ we find $\tau : \mathbb{F}$ such that $\mathrm{PA} \vdash \tau \dot{\leftrightarrow} F(\mathrm{Prov}(\overline{\tau}))$
- **internal necessitation** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\mathrm{Prov}(\overline{\varphi})})$
- **the distributivity law** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi \dot{\rightarrow} \psi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\psi})$

- Next obligation: Modal fixed points
  - Problem: Construct formula 'out of nowhere' $\rightarrow$ auxiliary result needed

Gödel's diagonal lemma

Lemma (Diagonal lemma, cf. [Nor18])

Suppose $\varphi(x) : \mathbb{F}$*. Then, there is a sentence* G *satisfying*

$$\mathsf{PA} \vdash G \dot{\leftrightarrow} \varphi(\overline{G}).$$

- Key result behind Gödel's first incompleteness theorem (among others)

Lemma (Recursion theorem, Kleene (1938))

Suppose $f : \mathbb{N} \to \mathbb{N}$ *is computable (and total).*
Then, there is $g : \mathbb{N}$ *such that* $M_g = M_{f(g)}$*.*

- Proofs closely related

Modal fixed points

With the **diagonal lemma**, we can find modal fixed points.

Lemma (Modal fixed points)

Let $F(x)$ be a formula. There is a sentence ψ such that

$$\mathrm{PA} \vdash \psi \dot{\leftrightarrow} F(\mathrm{Prov}(\overline{\psi})).$$

Proof.

- Use diagonal lemma on instance $F(\mathrm{Prov}(x))$
- We obtain $\mathrm{PA} \vdash \psi \dot{\leftrightarrow} F(\mathrm{Prov}(\overline{\psi}))$

□

Löb's axioms revisited

Löb's axioms

Suppose $\mathrm{Prov}(x) : \mathbb{F}$, and φ, ψ any sentence. $\mathrm{Prov}(x)$ is a modality satisfying

- **necessitation** if $\mathrm{PA} \vdash \varphi$ implies $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi})$
- **the modal fixed point theorem** if for any $F(x) : \mathbb{F}$ we find $\tau : \mathbb{F}$ such that $\mathrm{PA} \vdash \tau \dot{\leftrightarrow} F(\mathrm{Prov}(\overline{\tau}))$
- **internal necessitation** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\mathrm{Prov}(\overline{\varphi})})$
- **the distributivity law** if $\mathrm{PA} \vdash \mathrm{Prov}(\overline{\varphi \dot{\rightarrow} \psi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\varphi}) \dot{\rightarrow} \mathrm{Prov}(\overline{\psi})$

- Other axioms to be investigated
  - We have Kreisel's notion of provability
  - Too weak to show all axioms

Corollaries of the diagonal lemma

Theorem (Gödel's first incompleteness theorem)

There exists a sentence G with $\mathrm{PA} \nvdash G$ and $\mathrm{PA} \nvdash \dot{\neg} G$.

Transfers along all consistent and enumerable extensions of PA.

- Similar statement shown by Kirst and Peters 2023
  - Different approach: Computational argument
- Our approach: Use diagonal lemma on instance $\dot{\neg}\mathrm{Prov}(x)$

Corollaries of the diagonal lemma (continued)

Theorem (Tarski's theorem, cf. [BBJ07])

There is no formula True(x) *such that for all formulae* φ

$$(\mathbb{N} \vDash \varphi \to \mathbb{N} \vDash \mathrm{True}(\overline{\varphi})) \text{ and } (\mathbb{N} \nvDash \varphi \to \mathbb{N} \vDash \dot{\neg}\mathrm{True}(\overline{\varphi})).$$

Theorem (Essential undecidability[1])

Suppose $T \supseteq \mathrm{PA}$ *consistent.*
Then, $\lambda\varphi.\ T \vdash \varphi$ *is not decidable.*

[1] Also shown by Kirst and Hermes 2022 using different approach.

Further work

- Find provability formula strong enough to show all Löb axioms
  - Stay as abstract as possible
  - Exploit CT_{PA} as far as possible
- In Löb's original 1955 paper, axioms are different
  - Ours used in more recent literature
  - Investigate how they relate
- Derive Gödel's second incompleteness theorem from Löb's theorem
- Diagonal lemma requires formula to have at most one free variable
  - What happens if we allow for more?

Thanks for your attention.

[BBJ07] George S. Boolos, John P. Burgess and Richard C. Jeffrey. **Computability and Logic**. 5. edition. Cambridge University Press, 2007.

[FKS19] Yannick Forster, Dominik Kirst and Gert Smolka. 'On synthetic undecidability in coq, with an application to the entscheidungsproblem'. In: **Proceedings of the 8th ACM SIGPLAN International Conference on Certified Programs and Proofs, CPP 2019, Cascais, Portugal, January 14-15, 2019**. Ed. by Assia Mahboubi and Magnus O. Myreen. ACM, 2019, pp. 38–51. DOI: 10.1145/3293880.3294091. URL: https://doi.org/10.1145/3293880.3294091.

[Hen54] Leon Henkin. 'G. Kreisel. On a problem of Henkin's. Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings, series A, vol. 56 (1953), pp. 405–406; also Indagationes mathematicae, vol. 15 (1953), pp. 405–406.'. In: **Journal of Symbolic Logic** 19.3 (1954), 219–220. DOI: 10.2307/2268627.

[HK23] Marc Hermes and Dominik Kirst. 'An Analysis of Tennenbaum's Theorem in Constructive Type Theory'. In: **CoRR** abs/2302.14699 (2023). DOI: 10.48550/ARXIV.2302.14699. arXiv: 2302.14699. URL: https://doi.org/10.48550/arXiv.2302.14699.

[KH23] Dominik Kirst and Marc Hermes. ‘Synthetic Undecidability and Incompleteness of First-Order Axiom Systems in Coq’. In: **J. Autom. Reason.** 67.1 (2023), p. 13. DOI: 10.1007/S10817-022-09647-X. URL: https://doi.org/10.1007/s10817-022-09647-x.

[Kle38] S. C. Kleene. ‘On Notation for Ordinal Numbers’. In: **The Journal of Symbolic Logic** 3.4 (1938), pp. 150–155. ISSN: 00224812. URL: http://www.jstor.org/stable/2267778 (visited on 28/04/2024).

[KP23] Dominik Kirst and Benjamin Peters. 'Gödel's Theorem Without Tears - Essential Incompleteness in Synthetic Computability'. In: **31st EACSL Annual Conference on Computer Science Logic (CSL 2023)**. Ed. by Bartek Klin and Elaine Pimentel. Vol. 252. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 30:1–30:18. ISBN: 978-3-95977-264-8. DOI: 10.4230/LIPIcs.CSL.2023.30. URL: https://drops-dev.dagstuhl.de/entities/document/10.4230/LIPIcs.CSL.2023.30.

[Lob55] M. H. Lob. 'Solution of a Problem of Leon Henkin'. In: **The Journal of Symbolic Logic** 20.2 (1955), pp. 115–118. ISSN: 00224812. URL: http://www.jstor.org/stable/2266895 (visited on 28/04/2024).

[Nor18] Michael Norrish. **LSS 2018: Computability and Incompleteness. Gödel's First Incompleteness Theorem, Tarski's Indefinability of Truth**. 2018. URL: https://comp.anu.edu.au/lss/lectures/2023/lss-computability-4.pdf.

[Smo24] Gert Smolka. **Modeling and Proving in Computational Type Theory Using the Coq Proof Assistant**. Textbook under construction. 2024. URL: https://www.ps.uni-saarland.de/~smolka/drafts/mpctt.pdf (visited on 28/04/2024).

Technical background: Gödel numberings

Problem

Let $\varphi(x), \psi : \mathbb{F}$.
We used $\varphi(\overline{\psi})$ for 'substituting some encoding of ψ for x in φ'.
ψ is not a **number**, but a **formula** $\rightarrow \overline{\psi}$ not a numeral.

Typical issue. Gödel faced it himself.

Remark (Gödelisation)

There are functions $\text{göd} : \mathbb{F} \rightarrow \mathbb{N}$, $\text{göd}^{-1} : \mathbb{N} \rightarrow \mathbb{F}$ *inverting each other.*

$\varphi(\overline{\psi}) \rightsquigarrow \varphi(\overline{\text{göd}(\psi)})$

Technical background: Diagonal lemma

- Functions $\text{diag} := \lambda\varphi.\ \varphi(\overline{\varphi})$, and $\text{diag}_{\mathbb{N}} := \lambda n.\ \text{göd}(\text{diag}(\text{göd}^{-1}(n)))$

Proof.

- Suppose $\varphi(x)$. To find: G such that $\text{PA} \vdash G \dot{\leftrightarrow} \varphi(\overline{G})$
- Plug $\text{diag}_{\mathbb{N}}$ into CT_{PA}, get $\text{dg}(x, y)$ with $\forall n : \mathbb{N}.\ \text{PA} \vdash \dot{\forall} x.\ \text{dg}(\overline{n}, x) \dot{\leftrightarrow} x \equiv \overline{\text{diag}_{\mathbb{N}}\ n}$
- Define $G' := \dot{\exists} y.\ \text{dg}(x, y) \dot{\wedge} \varphi(y)$ and $G := G'(\overline{G'})$
- Argue inside PA that

$$\begin{aligned} G = G'(\overline{G'}) &= \dot{\exists} y.\ \text{dg}(\overline{G'}, y) \dot{\wedge} \varphi(y) \\ &\leftrightarrow \dot{\exists} y.\ y \equiv \overline{\text{diag}_{\mathbb{N}}(\text{göd}(G'))} \dot{\wedge} \varphi(y) \\ &\leftrightarrow \dot{\exists} y.\ y \equiv \overline{\text{göd}(G)} \dot{\wedge} \varphi(y) \\ &\leftrightarrow \varphi(\overline{G}) \end{aligned}$$

□

Technical background: Tarski's theorem

Theorem (Tarski's theorem)

There is no formula True(x) *such that for all formulae* φ

$$(\mathbb{N} \vDash \varphi \rightarrow \mathbb{N} \vDash \mathrm{True}(\overline{\varphi})) \textit{ and } (\mathbb{N} \nvDash \varphi \rightarrow \mathbb{N} \vDash \dot{\neg}\mathrm{True}(\overline{\varphi})).$$

Proof.

- Suppose True(x) has this property
- By diagonal lemma and soundness, find G such that $\mathbb{N} \vDash G \dot{\leftrightarrow} \dot{\neg}\mathrm{True}(\overline{G})$
- Case distinction
  - If $\mathbb{N} \vDash G$, then $\mathbb{N} \vDash \mathrm{True}(\overline{G})$
    Further, $\mathbb{N} \vDash \dot{\neg}\mathrm{True}(\overline{G})$ from $\mathbb{N} \vDash G \dot{\leftrightarrow} \dot{\neg}\mathrm{True}(\overline{G})$, i.e. $\mathbb{N}$ is inconsistent
  - If $\mathbb{N} \nvDash G$, have $\mathbb{N} \vDash \dot{\neg}\mathrm{True}(\overline{G})$
    Show $\mathbb{N} \vDash G$. Easy from $\mathbb{N} \vDash G \dot{\leftrightarrow} \dot{\neg}\mathrm{True}(\overline{G})$ □

Proof still constructive by [Smo24] (classical reasoning for stable claims).

Technical background: Gödel's first incompleteness theorem

Theorem (Strong separability, cf. [HK23])

Suppose $P, Q : \mathbb{N} \to \mathbb{P}$ *are*

- *both semi-decidable and*
- *disjoint (i.e. for all* $n : \mathbb{N}$*, we have* $P\ n \to Q\ n \to \bot$*).*

Then, there is a formula $\varphi(x)$ *such that for all* $n : \mathbb{N}$ *we have*

$$(P\ n \to \mathrm{PA} \vdash \varphi(\overline{n})) \text{ and } (Q\ n \to \mathrm{PA} \vdash \dot{\neg}\varphi(\overline{n})).$$

Corollary

We find a formula $\mathrm{SProv}(x)$ *such that for all formulae* φ

$$(\mathrm{PA} \vdash \varphi \to \mathrm{PA} \vdash \mathrm{SProv}(\overline{\varphi})) \land (\mathrm{PA} \vdash \dot{\neg}\varphi \to \mathrm{PA} \vdash \dot{\neg}\mathrm{SProv}(\overline{\varphi}))$$

Technical background: Gödel's first incompleteness theorem (continued)

We have $\mathrm{SProv}(x)$ such that for all formulae φ

$$(\mathrm{PA} \vdash \varphi \rightarrow \mathrm{PA} \vdash \mathrm{SProv}(\overline{\varphi})) \land (\mathrm{PA} \vdash \dot{\neg}\varphi \rightarrow \mathrm{PA} \vdash \dot{\neg}\mathrm{SProv}(\overline{\varphi}))$$

Proof (of Gödel's first incompleteness theorem).

- Need to find: Sentence G with $\mathrm{PA} \nvdash G$ and $\mathrm{PA} \nvdash \dot{\neg} G$
- Plug $\dot{\neg}\mathrm{SProv}(x)$ into diagonal lemma, obtain $\mathrm{PA} \vdash G \dot{\leftrightarrow} \dot{\neg}\mathrm{SProv}(\overline{G})$
- If $\mathrm{PA} \vdash G$
  - Obtain $\mathrm{PA} \vdash \mathrm{SProv}(\overline{G})$ by property of $\mathrm{SProv}(x)$
  - Observe that $\mathrm{PA} \vdash \dot{\neg}\mathrm{SProv}(\overline{G})$ from diagonal lemma, contradiction
- If $\mathrm{PA} \vdash \dot{\neg} G$
  - Obtain $\mathrm{PA} \vdash \dot{\neg}\mathrm{SProv}(\overline{G})$ by property of $\mathrm{SProv}(x)$
  - Observe that $\mathrm{PA} \vdash G$ from diagonal lemma, contradiction

$\square$

Technical background: Essential undecidability

Theorem (Essential undecidability)

Suppose $T \supseteq \mathsf{PA}$ consistent.
Then, $\lambda\varphi.\ T \vdash \varphi$ is not decidable.

Lemma

Suppose $P : \mathbb{F} \to \mathbb{P}$ is decidable.
We can find formula $\varphi(x)$ such that for any formula ψ

$$P\,\psi \to \mathsf{PA} \vdash \varphi(\overline{\psi}) \text{ and } \neg P\,\psi \to \mathsf{PA} \vdash \dot\neg\varphi(\overline{\psi}).$$

Proof (of essential undecidability).

- Suppose $\lambda\varphi.\ T \vdash \varphi$ was decidable
- Invoke lemma to obtain $\varphi(x)$
- By weakening, $\forall\psi.\ T \vdash \psi \to T \vdash \varphi(\overline{\psi}) \wedge T \nvdash \psi \to T \vdash \dot\neg\varphi(\overline{\psi})$, contradiction

□