Proof interpretations: a modern perspective

Lecture 4 - Classical logic and the negative translation

Anupam Das & Thomas Powell

University of Copenhagen & Technische Universität Darmstadt

North American Summer School on Logic, Language, and Information

Carnegie Mellon University

28 June 2018

These slides are available at `http://www.anupamdas.com/nasslli18`.

Outline

In Lecture 3, we presented Gödel's functional interpretation of intuitionistic arithmetic:

$$\mathrm{HA}^{\omega} \vdash A \Rightarrow \text{System T} \vdash A_D(t, y)$$

The soundness theorem states: If HA^{ω} proves an existential statement, we can find an term t which *computes* that object.

In Lecture 3, we presented Gödel's functional interpretation of intuitionistic arithmetic:

$$\mathrm{HA}^\omega \vdash A \Rightarrow \text{System T} \vdash A_D(t, y)$$

The soundness theorem states: If HA^ω proves an existential statement, we can find an term t which *computes* that object.

For example:

$$\forall n (X(n) \geq n \wedge X(n) \text{ prime})$$

where

$$X(n) = \text{least } p \leq 1 + n! \text{ such that } p \text{ prime}$$

In Lecture 3, we presented Gödel's functional interpretation of intuitionistic arithmetic:

$$\mathrm{HA}^{\omega} \vdash A \Rightarrow \text{System T} \vdash A_D(t, y)$$

The soundness theorem states: If HA^{ω} proves an existential statement, we can find an term t which *computes* that object.

For example:

$$\forall n (X(n) \geq n \wedge X(n) \text{ prime})$$

where

$$X(n) = \text{least } p \leq 1 + n! \text{ such that } p \text{ prime}$$

But in ordinary mathematics, we frequently make use of the law of excluded-middle, namely:

$$A \vee \neg A$$

which is explicitly banned in HA^{ω}.

Plan of the lecture

1. We first establish that classical logic poses a genuine problem for program extraction.

2. But things are not quite as bad as they may seem! We can often extract 'indirect' information.

3. In fact, under certain conditions we can always do this, and there is a logical technique for making this formal: The negative translation

4. Moreover, for $\forall\exists$ theorems, classical logic can be circumvented entirely! We can still extract programs from nonconstructive proofs of purely existential statements.

As before, we give lots of examples (even more than last lecture).

References

We primarily follow Chapters 2 and 10 of

- Kohlenbach, U. (2008). *Applied Proof Theory - Proof Interpretations and their Use in Mathematics.*
  Springer Monographs in Mathematics. Springer

Outline

Back to rational powers

Theorem

There exists a pair of irrational numbers x, y such that x^y is rational.

Back to rational powers

Theorem

There exists a pair of irrational numbers x, y such that x^y is rational.

Proof.

Suppose that $\sqrt{2}^{\sqrt{2}}$ is rational. Then we can just set $x = y = \sqrt{2}$.

Otherwise, $\sqrt{2}^{\sqrt{2}}$ must be irrational, and we can set $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, since

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^2 = 2.$$

Done. □

Theorem

There exists a pair of irrational numbers x, y such that x^y is rational.

Proof.

Suppose that $\sqrt{2}^{\sqrt{2}}$ is rational. Then we can just set $x = y = \sqrt{2}$.

Otherwise, $\sqrt{2}^{\sqrt{2}}$ must be irrational, and we can set $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, since

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^2 = 2.$$

Done. $\square$

However, while the above proof gives us two *candidates* for x and y, namely

$$(x, y) = (\sqrt{2}, \sqrt{2}) \text{ or } (\sqrt{2}^{\sqrt{2}}, \sqrt{2})$$

we don't know which one works, since we have no prodedure for deciding whether or not $\sqrt{2}^{\sqrt{2}}$ is irrational.

Theorem

There exists a pair of irrational numbers x, y such that x^y is rational.

Proof.

Suppose that $\sqrt{2}^{\sqrt{2}}$ is rational. Then we can just set $x = y = \sqrt{2}$.

Otherwise, $\sqrt{2}^{\sqrt{2}}$ must be irrational, and we can set $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, since

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{2} = 2.$$

Done. □

However, while the above proof gives us two *candidates* for x and y, namely

$$(x, y) = (\sqrt{2}, \sqrt{2}) \text{ or } (\sqrt{2}^{\sqrt{2}}, \sqrt{2})$$

we don't know which one works, since we have no prodedure for deciding whether or not $\sqrt{2}^{\sqrt{2}}$ is irrational.

Remark. Actually, it is known that $\sqrt{2}^{\sqrt{2}}$ is irrational, but this is a deep result in its own right.

The drinkers paradox

There is someone in the pub such that, if they are drinking, then everyone in the pub is drinking.

The drinkers paradox

> *There is someone in the pub such that, if they are drinking, then everyone in the pub is drinking.*

For pubs with infinitely many drinkers, this is non-computable!

> *There is someone in the pub such that, if they are drinking, then everyone in the pub is drinking.*

For pubs with infinitely many drinkers, this is non-computable!

Theorem (Formal drinkers paradox)

Let P be some predicate on the natural numbers. Then

$$\exists n (P(n) \rightarrow \forall m P(m)).$$

The drinkers paradox

> *There is someone in the pub such that, if they are drinking, then everyone in the pub is drinking.*

For pubs with infinitely many drinkers, this is non-computable!

Theorem (Formal drinkers paradox)

Let P be some predicate on the natural numbers. Then

$$\exists n (P(n) \to \forall m P(m)).$$

Proof.

Suppose that $P(k)$ is true for all k. Set $n := 0$.

Otherwise, $P(k)$ fails for some k. Set $n := k$. □

Again, we have two candidates, but no way to pick them, since we cannot decide in finite time whether or not $P(k)$ holds for some k.

The minimum principle

Theorem

Let $f : \mathbb{N} \to \mathbb{N}$ *be a function. There exists some* $n \in \mathbb{N}$ *such that* $\forall m (f(n) \leq f(m))$.

The minimum principle

Theorem
Let $f : \mathbb{N} \to \mathbb{N}$ *be a function. There exists some* $n \in \mathbb{N}$ *such that* $\forall m (f(n) \leq f(m))$*.*

Proof.
Suppose that this were not the case. Then for any n there would exists some m with $f(n) > f(m)$.

Define the sequence (x_i) by

$$x_0 := n \text{ and } x_{i+1} \text{ satisfies } f(x_i) > f(x_{i+1})$$

Then we have an infinite decreasing sequence

$$f(x_0) > f(x_1) > f(x_2) > \ldots$$

which contradicts the wellfoundedness of $\mathbb{N}$. □ □

The minimum principle

Theorem

Let $f : \mathbb{N} \to \mathbb{N}$ be a function. There exists some $n \in \mathbb{N}$ such that $\forall m(f(n) \leq f(m))$.

Proof.

Suppose that this were not the case. Then for any n there would exists some m with $f(n) > f(m)$.

Define the sequence (x_i) by

$$x_0 := n \text{ and } x_{i+1} \text{ satisfies } f(x_i) > f(x_{i+1})$$

Then we have an infinite decreasing sequence

$$f(x_0) > f(x_1) > f(x_2) > \dots$$

which contradicts the wellfoundedness of $\mathbb{N}$. $\square$ $\square$

As before, the proof tells us that some n exists, but doesn't tell us how to find it!

But is this a problem merely with the *proof*, or is it a fundamental property of the *theorem* itself?

Some theorems are just noncomputable

Theorem

There is no computable functional $\Phi : (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ *which satisfies*

$$(*) \quad \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

Some theorems are just noncomputable

Theorem

There is no computable functional $\Phi : (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ *which satisfies*

$$(*) \quad \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

Proof.

Suppose that one did exist, and define $f = \mathbf{1}$ i.e. f is the constant 1-function. Since Φ is computable, it only looks at a finite amount of its input i.e. there exists some N such that

$$(\dagger) \quad \forall g : \mathbb{N} \to \mathbb{N} (\forall i \leq N (g(i) = 1) \to \Phi(g) = \Phi(\mathbf{1}))$$

Some theorems are just noncomputable

Theorem

There is no computable functional $\Phi : (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ *which satisfies*

$$(*) \quad \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

Proof.

Suppose that one did exist, and define $f = \mathbf{1}$ i.e. f is the constant 1-function. Since Φ is computable, it only looks at a finite amount of its input i.e. there exists some N such that

$$(\dagger) \quad \forall g : \mathbb{N} \to \mathbb{N} (\forall i \leq N (g(i) = 1) \to \Phi(g) = \Phi(\mathbf{1}))$$

Now define

$$h(n) := \begin{cases} 1 & \text{if } n \leq \max\{N, \Phi(\mathbf{1})\} \\ 0 & \text{otherwise} \end{cases}$$

Some theorems are just noncomputable

Theorem

There is no computable functional $\Phi : (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ which satisfies

$$(*) \quad \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

Proof.

Suppose that one did exist, and define $f = \mathbf{1}$ i.e. f is the constant 1-function. Since Φ is computable, it only looks at a finite amount of its input i.e. there exists some N such that

$$(\dagger) \quad \forall g : \mathbb{N} \to \mathbb{N} (\forall i \leq N (g(i) = 1) \to \Phi(g) = \Phi(\mathbf{1}))$$

Now define

$$h(n) := \begin{cases} 1 & \text{if } n \leq \max\{N, \Phi(\mathbf{1})\} \\ 0 & \text{otherwise} \end{cases}$$

- Then $\forall i \leq N(h(i) = 1)$ and so $\Phi(h) = \Phi(\mathbf{1})$ by $(\dagger)$.
- But by $(*)$ we have $\exists n \leq \Phi(\mathbf{1})(g(n) \leq 0)$

Some theorems are just noncomputable

Theorem

There is no computable functional $\Phi : (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ *which satisfies*

$$(*) \quad \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

Proof.

Suppose that one did exist, and define $f = \mathbf{1}$ i.e. f is the constant 1-function. Since Φ is computable, it only looks at a finite amount of its input i.e. there exists some N such that

$$(\dagger) \quad \forall g : \mathbb{N} \to \mathbb{N} (\forall i \leq N (g(i) = 1) \to \Phi(g) = \Phi(\mathbf{1}))$$

Now define

$$h(n) := \begin{cases} 1 & \text{if } n \leq \max\{N, \Phi(\mathbf{1})\} \\ 0 & \text{otherwise} \end{cases}$$

- Then $\forall i \leq N (h(i) = 1)$ and so $\Phi(h) = \Phi(\mathbf{1})$ by $(\dagger)$.
- But by $(*)$ we have $\exists n \leq \Phi(\mathbf{1})(g(n) \leq 0)$

But $g(n) = 1$ for all $n \leq \Phi(\mathbf{1})$, a contradiction. □

There is no classical functional interpretation

It is impossible to extend the functional interpretation to classical logic.

There is no classical functional interpretation

It is impossible to extend the functional interpretation to classical logic.

If it were, then since PA^ω proves

$$\forall f \exists n \forall m (f(n) \leq f(m))$$

we would expect to extract a term t of System T satisfying

$$\forall f, m (f(t(f)) \leq f(m))$$

Therefore, in particular, there would be a computable functional $\Phi(f) := t(f)$ satisfying

$$\exists n \leq \Phi(f) \forall m (f(n) \leq f(m))$$

which we just demonstrated was not possible.

What about everyday mathematics?

Theorem

Let (x_n) be a nondecreasing sequence of rational numbers in the unit interval $[0,1]$. Then (x_n) converges.

What about everyday mathematics?

Theorem

Let (x_n) be a nondecreasing sequence of rational numbers in the unit interval $[0, 1]$. Then (x_n) converges.

Theorem

Formal version. Let (x_n) be a nondecreasing sequence of rational numbers in $[0, 1]$. Then

$$\forall k \exists n \forall m (|x_{n+m} - x_n| \leq 2^{-k}).$$

What about everyday mathematics?

Theorem

Let (x_n) be a nondecreasing sequence of rational numbers in the unit interval $[0,1]$. Then (x_n) converges.

Theorem

Formal version. Let (x_n) be a nondecreasing sequence of rational numbers in $[0,1]$. Then

$$\forall k \exists n \forall m (|x_{n+m} - x_n| \leq 2^{-k}).$$

Theorem (Functional interpretation)

Let (x_n) be a nondecreasing sequence of rational numbers in $[0,1]$. Then there exists a function $N : \mathbb{N} \to \mathbb{N}$ such that

$$\forall k, m (|x_{N(k)+m} - x_{N(k)}| \leq 2^{-k}).$$

Remark. *The function N is a so-called modulus of convergence for (x_n).*

Even basic things like convergence are fundamentally non-computable

Theorem (E. Specker, 1949)

There exist computable, monotonically increasing, bounded sequences of rational numbers which do not have a computable modulus of convergence.

Note. Just sequences are known as Specker sequences.

Even basic things like convergence are fundamentally non-computable

Theorem (E. Specker, 1949)

There exist computable, monotonically increasing, bounded sequences of rational numbers which do not have a computable modulus of convergence.

Note. Just sequences are known as Specker sequences.

Conclusion.

- There are simple, everyday mathematical facts which are fundamentally non-computable.
- Direct program extraction only works for proofs which don't use any law of excluded-middle.
- The vast majority of normal mathematical proofs are beyond program extraction...

Even basic things like convergence are fundamentally non-computable

Theorem (E. Specker, 1949)

There exist computable, monotonically increasing, bounded sequences of rational numbers which do not have a computable modulus of convergence.

Note. Just sequences are known as Specker sequences.

Conclusion.

- There are simple, everyday mathematical facts which are fundamentally non-computable.
- Direct program extraction only works for proofs which don't use any law of excluded-middle.
- The vast majority of normal mathematical proofs are beyond program extraction...

But it's not quite as bad as it looks!

Outline

The drinkers paradox revisited

Theorem

$\exists n \forall m (P(n) \rightarrow P(m))$.

The drinkers paradox revisited

Theorem

$\exists n \forall m (P(n) \to P(m))$.

Proof (computational version).

Suppose that the statement is false, in other words

$$\forall n \exists m (\neg P(n) \wedge P(m)).$$

Then there must exist a function $g : \mathbb{N} \to \mathbb{N}$ such that

$$\forall n (\neg P(n) \wedge P(g(n))).$$

Theorem
$\exists n \forall m (P(n) \to P(m))$.

Proof (computational version).
Suppose that the statement is false, in other words

$$\forall n \exists m (\neg P(n) \wedge P(m)).$$

Then there must exist a function $g : \mathbb{N} \to \mathbb{N}$ such that

$$\forall n (\neg P(n) \wedge P(g(n))).$$

But this is impossible: Either $P(g(0))$ is false, in which case the following fails:

$$\neg P(0) \wedge P(g(0))$$

or $P(g(0))$ is true, in which case the following fails

$$\neg P(g(0)) \wedge P(g(g(0)))$$

Therefore the statement must be true. □

A computational drinkers paradox

Let's rephrase this argument in a slightly more formal way. Over classical logic we have the following set of equivalences:

$$
\begin{aligned}
&\exists n \forall m (P(n) \to P(m)) \\
&\Leftrightarrow \neg \forall n \exists m (\neg P(n) \wedge P(m)) \\
&\Leftrightarrow \neg \exists g \forall n (\neg P(n) \wedge P(g(n))) \\
&\Leftrightarrow \forall g \exists n (P(n) \to P(g(n))).
\end{aligned}
$$

A computational drinkers paradox

Let's rephrase this argument in a slightly more formal way. Over classical logic we have the following set of equivalences:

$$\begin{aligned} &\exists n \forall m (P(n) \to P(m)) \\ \Leftrightarrow\ &\neg \forall n \exists m (\neg P(n) \wedge P(m)) \\ \Leftrightarrow\ &\neg \exists g \forall n (\neg P(n) \wedge P(g(n))) \\ \Leftrightarrow\ &\forall g \exists n (P(n) \to P(g(n))). \end{aligned}$$

There is no computable way to find an n satisfying the first formula. But we can find a functional Φ realizing the second formula:

$$\Phi(g) := \begin{cases} 0 & \text{if } P(g(0)) \\ g(0) & \text{if } \neg P(g(0)) \end{cases}$$

What does the functional Φ actually do?

The original drinkers paradox $\exists n \forall m (P(n) \to P(m))$ asserts:

There exists an 'ideal' drinker n, such that if they drink then $\boxed{\textit{everyone drinks}}$

What does the functional Φ actually do?

The original drinkers paradox $\exists n \forall m (P(n) \to P(m))$ asserts:

There exists an 'ideal' drinker n, such that if they drink then $\boxed{\textit{everyone drinks}}$

The reformulated drinks paradox $\forall g \exists n (P(n) \to P(g(n)))$ asserts:

$\boxed{\textit{For any function } g}$ *there exists an approximation n to an ideal drinker, such that if they drink then* $g(n)$ *drinks.*

What does the functional Φ actually do?

The original drinkers paradox $\exists n \forall m (P(n) \to P(m))$ asserts:

> *There exists an 'ideal' drinker n, such that if they drink then* *everyone drinks*

The reformulated drinks paradox $\forall g \exists n (P(n) \to P(g(n)))$ asserts:

> *For any function g* *there exists an approximation n to an ideal drinker, such that if they drink then $g(n)$ drinks.*

Key idea.

- We may not be able to compute ideal objects whose existence relies on classical logic, but we can compute approximations to those ideal objects.
- Functionals which compute these approximations can be formally extracted from the classical proof.

The minimum principle revisited

Theorem

$\forall f \exists n \forall m (f(n) \leq f(m))$.

The minimum principle revisited

Theorem

$\forall f \exists n \forall m (f(n) \leq f(m))$.

Proof (computational version).

Suppose that the statement is false, in other words

$$\exists f \forall n \exists m (f(n) > f(m)).$$

The minimum principle revisited

Theorem

$\forall f \exists n \forall m (f(n) \leq f(m))$.

Proof (computational version).

Suppose that the statement is false, in other words

$$\exists f \forall n \exists m (f(n) > f(m)).$$

Then in particular there must exists a function $g : \mathbb{N} \to \mathbb{N}$ such that

$$\forall n (f(n) > f(g(n))).$$

The minimum principle revisited

Theorem
$\forall f \exists n \forall m (f(n) \leq f(m))$.

Proof (computational version).
Suppose that the statement is false, in other words

$$\exists f \forall n \exists m (f(n) > f(m)).$$

Then in particular there must exists a function $g : \mathbb{N} \to \mathbb{N}$ such that

$$\forall n (f(n) > f(g(n))).$$

But this means that

$$f(0) > f(g(0)) > f(g^{(2)}(0)) > \ldots > f(g^{(k)}(0))$$

which is a contradiction for any $k > f(0)$. Therefore we have

$$f(n) \leq f(g(n))$$

where n is one of $g(0), g^{(2)}(0), \ldots, g^{(f(0))}(0)$, and so the original statement must be true. $\square$

A computational minimum principle

In general there is no computable functional Φ such that

$$\forall f \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

However, we *can* find a functional Φ such that

$$\forall f, g \exists n \leq \Phi(f, g)(f(n) \leq f(g(n))).$$

namely:

$$\Phi(f, g) := \max\{g(0), g^{(2)}(0), \ldots, g^{(f(0))}(0)\}$$

A computational minimum principle

In general there is no computable functional Φ such that

$$\forall f \exists n \leq \Phi(f) \forall m (f(n) \leq f(m)).$$

However, we *can* find a functional Φ such that

$$\forall f, g \exists n \leq \Phi(f, g)(f(n) \leq f(g(n))).$$

namely:

$$\Phi(f, g) := \max\{g(0), g^{(2)}(0), \ldots, g^{(f(0))}(0)\}$$

Alternatively put

- There is no computable bound for a minimal n;
- There is a computable bound for an approximation to a minimal n.

What is really going on here?

We are seeing the following phenomenon.

- We cannot compute direct witnesses for existential statements proven using classical logic.
- We can compute witnesses for the 'not not' version of these statements.
- The latter can be viewed as approximatations to the former.

What is going on in general?

Exercises

1. What is the functional interpretation of the law of excluded-middle for $\exists$-formulas i.e.
$$\exists x P(x) \vee \forall y \neg P(y).$$
What is the corresponding 'indirect' interpretation.

2. Repeat question 1, but this time for the general minimum principle
$$\exists x P(x) \rightarrow \exists y (P(y) \wedge \forall z < y \neg P(z)).$$

3. There is a non-computable f which solves the Halting problem:
$$\exists f \forall e, x (f(e, x) = 1 \leftrightarrow \{e\} \text{ terminates on input } x)$$
What would an approximation of f be? Can we compute it?

Outline

The Gödel-Gentzen negative translation

Let A be a formula in predicate logic. We define the negative translation of A by

$$A^N :\equiv \neg\neg A^*$$

where A^* is defined inductively as

$$\begin{aligned} A^* &:\equiv A \text{ if } A \text{ is a prime formula} \\ (A \Box B)^* &:\equiv A^* \Box B^* \text{ if } \Box \in \{\wedge, \vee, \rightarrow\} \\ (\exists x A)^* &:\equiv \exists x A^* \\ (\forall x A)^* &\equiv \forall x \neg\neg A^* \end{aligned}$$

Soundness of the negative translation

The negative translation obeys the following general pattern: Suppose that

$$\mathcal{P}_{\text{class}} \vdash A$$

for some classical theory $\mathcal{P}_{\text{class}}$. Then

$$\mathcal{P} \vdash A^N$$

where $\mathcal{P}$ is the intuitionistic version of that theory.

Soundness of the negative translation

The negative translation obeys the following general pattern: Suppose that

$$\mathcal{P}_{\text{class}} \vdash A$$

for some classical theory $\mathcal{P}_{\text{class}}$. Then

$$\mathcal{P} \vdash A^N$$

where $\mathcal{P}$ is the intuitionistic version of that theory.

In particular, this is true for Peano/Heyting arithmetic.

Theorem

If $\mathrm{PA}^\omega \vdash A$ *then* $\mathrm{HA}^\omega \vdash A^N$.

Proof.

Induction over the structure of derivations in PA^ω. □

The negative translation of $\forall\exists\forall$ formulas

Suppose that $A :\equiv \forall k \exists n \forall m P(k, n, m)$ for $P(k, n, m)$ quantifier-free. Then

$$\begin{aligned} A^N &\equiv \neg\neg A^* \\ &\equiv \neg\neg(\forall k \exists n \forall m P(k, n, m))^* \\ &\equiv \neg\neg\forall k \neg\neg(\exists n \forall m P(k, n, m))^* \\ &\equiv \neg\neg\forall k \neg\neg\exists n \forall m \neg\neg P(k, n, m). \end{aligned}$$

The negative translation of $\forall\exists\forall$ formulas

Suppose that $A :\equiv \forall k \exists n \forall m P(k, n, m)$ for $P(k, n, m)$ quantifier-free. Then

$$\begin{aligned} A^N &\equiv \neg\neg A^* \\ &\equiv \neg\neg(\forall k \exists n \forall m P(k, n, m))^* \\ &\equiv \neg\neg\forall k \neg\neg(\exists n \forall m P(k, n, m))^* \\ &\equiv \neg\neg\forall k \neg\neg\exists n \forall m \neg\neg P(k, n, m). \end{aligned}$$

This looks complicated, but in arithmetic we have

$$\neg\neg Q \leftrightarrow Q$$

for all quantifier-free formulas, and

$$\neg\neg\forall k \neg\neg B \leftrightarrow \forall k \neg\neg B$$

is provable intuitionistically. Therefore

$$A^N \leftrightarrow \forall k \neg\neg\exists n \forall m P(k, n, m).$$

and so

$$\mathrm{PA} \vdash \forall k \exists n \forall m P(k, n, m) \Rightarrow \mathrm{HA} \vdash \forall k \neg\neg\exists n \forall m P(k, n, m).$$

We cannot give a direct computational interpretation to classical arithmetic i.e. it is *not* the case that

$$\text{if } \mathrm{PA}^\omega \vdash A \text{ then } \mathrm{HA}^\omega \vdash \forall y A_D(t, y)$$

for some $t \in \mathrm{T}$. However, what we do have is:

1. A computational interpretation of Heyting arithmetic
2. An embedding of Peano arithmetic into Heyting arithmetic

We cannot give a direct computational interpretation to classical arithmetic i.e. it is *not* the case that

$$\text{if } \mathrm{PA}^\omega \vdash A \text{ then } \mathrm{HA}^\omega \vdash \forall y A_D(t, y)$$

for some $t \in \mathrm{T}$. However, what we do have is:

1. A computational interpretation of Heyting arithmetic
2. An embedding of Peano arithmetic into Heyting arithmetic

So why not combine them? I.e.

$$\mathrm{PA}^\omega \mapsto \mathrm{HA}^\omega \mapsto \text{System } \mathsf{T}$$

Gödel's main theorem (second part)

Gödel's soundness theorem for *classical logic* says that we can translate a proof of A to a program witnessing $\exists x \forall y (A^N)_D(x, y)$.

Gödel's main theorem (second part)

Gödel's soundness theorem for *classical logic* says that we can translate a proof of A to a program witnessing $\exists x \forall y (A^N)_D(x, y)$.

Theorem (K. Gödel, 1958)

Suppose that

$$\mathrm{PA}^\omega \vdash A$$

Then there exists a term t of System T *such that*

$$\mathrm{HA}^\omega \vdash \forall y (A^N)_D(t, y)$$

and moreover, we can formally extract t from the proof of A.

Gödel's main theorem (second part)

Gödel's soundness theorem for *classical logic* says that we can translate a proof of A to a program witnessing $\exists x \forall y (A^N)_D(x, y)$.

Theorem (K. Gödel, 1958)

Suppose that

$$\mathrm{PA}^\omega \vdash A$$

Then there exists a term t of System T *such that*

$$\mathrm{HA}^\omega \vdash \forall y (A^N)_D(t, y)$$

and moreover, we can formally extract t from the proof of A.

Proof.

Combine the soundness theorem for intuitionistic logic with the negative translation. □

The classical functional interpretation of ∀∃∀ theorems

What is the functional interpretation of $B :\equiv \forall k \neg\neg\exists n \forall m P(k, n, m)$?

The classical functional interpretation of ∀∃∀ theorems

What is the functional interpretation of $B :\equiv \forall k \neg\neg \exists n \forall m P(k,n,m)$?

Going back to the last lecture we have

$$\begin{aligned}
\forall k \neg\neg \exists n \forall m P(k,n,m) &\mapsto \forall k \neg(\exists n \forall m P(k,n,m) \to \bot) \\
&\mapsto \forall k \neg \exists g \forall n \neg P(k,n,g(n)) \\
&\mapsto \forall k (\exists g \forall n \neg P(k,n,g(n)) \to \bot) \\
&\mapsto \exists \Phi \forall k, g P(k, \Phi(k,g), g(\Phi(k,g)))
\end{aligned}$$

The classical functional interpretation of $\forall\exists\forall$ theorems

What is the functional interpretation of $B :\equiv \forall k \neg\neg \exists n \forall m P(k,n,m)$?

Going back to the last lecture we have

$$\begin{aligned}
\forall k \neg\neg \exists n \forall m P(k,n,m) &\mapsto \forall k \neg(\exists n \forall m P(k,n,m) \to \bot) \\
&\mapsto \forall k \neg \exists g \forall n \neg P(k,n,g(n)) \\
&\mapsto \forall k (\exists g \forall n \neg P(k,n,g(n)) \to \bot) \\
&\mapsto \exists \Phi \forall k, g P(k, \Phi(k,g), g(\Phi(k,g)))
\end{aligned}$$

Therefore in the special case of theorems of this form, we have

$$\text{if } \mathrm{PA}^\omega \vdash \forall k \exists n \forall m P(k,n,m) \text{ then System } \mathrm{HA}^\omega \vdash \forall k, g P(k, t(k,g), g(t(k,g)))$$

for some term t if System T.

We can equivalently view this as a bound i.e.

$$\mathrm{T} \vdash \forall k, g, \exists n \leq t(g,k) P(k,n,g(n))$$

We now see what was going on with our earlier examples.

We now see what was going on with our earlier examples.

Drinkers paradox

$$\begin{aligned}\exists n \forall m (P(n) \to P(m)) &\rightsquigarrow \neg\neg\exists n \forall m (P(n) \to P(m)) \\ &\rightsquigarrow \forall g \exists n \leq \Phi(g)(P(n) \to P(g(n)))\end{aligned}$$

and a witness for n is given by

$$\Phi(g) = g(0)$$

We now see what was going on with our earlier examples.

Drinkers paradox

$$\begin{aligned}\exists n \forall m (P(n) \to P(m)) &\leadsto \neg\neg\exists n \forall m (P(n) \to P(m)) \\ &\leadsto \forall g \exists n \leq \Phi(g)(P(n) \to P(g(n)))\end{aligned}$$

and a witness for n is given by

$$\Phi(g) = g(0)$$

Least element principle

$$\begin{aligned}\forall f \exists n \forall m (f(n) \leq f(m)) &\leadsto \forall f \neg\neg\exists n \forall m (f(n) \leq f(m)) \\ &\leadsto \forall f, g \exists n (f(n) \leq f(g(n)))\end{aligned}$$

and a *bound* for n is given by

$$\Phi(f, g) = \max\{g(0), g^{(2)}(0), \ldots, g^{(f(0))}(0)\}$$

Outline

The classical functional interpretation of $\forall\exists$ statements

Suppose that $\mathrm{PA}^\omega \vdash B$ where $B :\equiv \forall u \exists v Q(u, v)$. What does the classical functional interpretation do in this case?

The classical functional interpretation of $\forall\exists$ statements

Suppose that $\mathrm{PA}^\omega \vdash B$ where $B :\equiv \forall u \exists v Q(u, v)$. What does the classical functional interpretation do in this case?

Let's first look at the negative translation. We have

$$B^N \equiv \neg\neg(\forall u \exists v Q(u, v)) \equiv \neg\neg\forall u \neg\neg\exists v \neg\neg Q(u, v) \leftrightarrow \forall u \neg\neg\exists v Q(u, v)$$

where the equivalence $\leftrightarrow$ is possible in Heyting arithmetic. Therefore

$$\mathrm{HA} \vdash \forall u \neg\neg\exists v Q(u, v)$$

The classical functional interpretation of $\forall\exists$ statements

Suppose that $\mathrm{PA}^\omega \vdash B$ where $B :\equiv \forall u \exists v Q(u,v)$. What does the classical functional interpretation do in this case?

Let's first look at the negative translation. We have

$$B^N \equiv \neg\neg(\forall u \exists v Q(u,v)) \equiv \neg\neg\forall u \neg\neg\exists v \neg\neg Q(u,v) \leftrightarrow \forall u \neg\neg\exists v Q(u,v)$$

where the equivalence $\leftrightarrow$ is possible in Heyting arithmetic. Therefore

$$\mathrm{HA} \vdash \forall u \neg\neg\exists v Q(u,v)$$

But what is the functional interpretation of this? We have

$$\begin{aligned} \forall u \neg\neg\exists v Q(u,v) &\mapsto \forall u \neg\exists v \neg Q(u,v) \\ &\mapsto \forall u \exists v \neg\neg Q(u,v) \\ &\mapsto \exists f \forall u Q(u, f(u)). \end{aligned}$$

But this is the same as the direct, intuitionistic functional interpretation!

The classical functional interpretation of ∀∃ statements

Suppose that $\mathrm{PA}^\omega \vdash B$ where $B :\equiv \forall u \exists v Q(u, v)$. What does the classical functional interpretation do in this case?

Let's first look at the negative translation. We have

$$B^N \equiv \neg\neg(\forall u \exists v Q(u, v)) \equiv \neg\neg\forall u \neg\neg\exists v \neg\neg Q(u, v) \leftrightarrow \forall u \neg\neg\exists v Q(u, v)$$

where the equivalence $\leftrightarrow$ is possible in Heyting arithmetic. Therefore

$$\mathrm{HA} \vdash \forall u \neg\neg\exists v Q(u, v)$$

But what is the functional interpretation of this? We have

$$\begin{aligned} \forall u \neg\neg\exists v Q(u, v) &\mapsto \forall u \neg\exists v \neg Q(u, v) \\ &\mapsto \forall u \exists v \neg\neg Q(u, v) \\ &\mapsto \exists f \forall u Q(u, f(u)). \end{aligned}$$

But this is the same as the direct, intuitionistic functional interpretation!

Remark. What really going on here is that the functional interpretation admits Markov's principle $\neg\neg\exists x A_0(x) \to \exists x A_0(x)$ for any quantifier-free formula $A_0(x)$.

Program extraction theorem

Theorem

Suppose that

$$\mathrm{PA}^{\omega} \vdash \forall u \exists v Q(u, v).$$

Then there exists a term t of System T *such that*

$$\mathrm{HA}^{\omega} \vdash \forall u Q(u, tu)$$

and moreover, we can formally extract t from the proof of A.

In other words, for the special case of $\forall\exists$ theorems, we can extract a **direct** witness from their proof, even if their proof uses non-constructive reasoning and therefore doesn't seem to have any computational meaning.

Program extraction theorem

Theorem
Suppose that

$$\mathrm{PA}^\omega \vdash \forall u \exists v Q(u, v).$$

Then there exists a term t of System T *such that*

$$\mathrm{HA}^\omega \vdash \forall u Q(u, tu)$$

and moreover, we can formally extract t from the proof of A.

In other words, for the special case of $\forall\exists$ theorems, we can extract a **direct** witness from their proof, even if their proof uses non-constructive reasoning and therefore doesn't seem to have any computational meaning.

How is this possible?

Program extraction theorem

Theorem

Suppose that

$$\mathrm{PA}^{\omega} \vdash \forall u \exists v Q(u, v).$$

Then there exists a term t of System T *such that*

$$\mathrm{HA}^{\omega} \vdash \forall u Q(u, tu)$$

and moreover, we can formally extract t from the proof of A.

In other words, for the special case of $\forall\exists$ theorems, we can extract a **direct** witness from their proof, even if their proof uses non-constructive reasoning and therefore doesn't seem to have any computational meaning.

How is this possible?

When the existence of ideal objects are used in the proof of purely existential statements, we only need approximations to those ideal objects to extract a witness for the statement.

Why it works

Suppose that a theorem $B :\equiv \forall u \exists v B(u, v)$ is proven using some nonconstructive lemma $A :\equiv \exists x \forall y A(x, y)$.

Naive idea. In order to find a function f satisfying $\forall u B(u, fu)$ we need to find some x satisfying $\forall y A(x, y)$. We cannot compute this x, therefore no computable f exists.

Why it works

Suppose that a theorem $B :\equiv \forall u \exists v B(u, v)$ is proven using some nonconstructive lemma $A :\equiv \exists x \forall y A(x, y)$.

Naive idea. In order to find a function f satisfying $\forall u B(u, fu)$ we need to find some x satisfying $\forall y A(x, y)$. We cannot compute this x, therefore no computable f exists.

Recall the functional interpretation of implication:

$$(\exists x \forall y A(x, y) \to \forall u \exists v B(u, v)) \mapsto \exists V, Y \forall x, u (A(x, Yxu) \to B(u, Vxu))$$

Why it works

Suppose that a theorem $B :\equiv \forall u \exists v B(u,v)$ is proven using some nonconstructive lemma $A :\equiv \exists x \forall y A(x,y)$.

Naive idea. In order to find a function f satisfying $\forall u B(u, fu)$ we need to find some x satisfying $\forall y A(x,y)$. We cannot compute this x, therefore no computable f exists.

Recall the functional interpretation of implication:

$$(\exists x \forall y A(x,y) \to \forall u \exists v B(u,v)) \mapsto \exists V, Y \forall x, u(A(x, Yxu) \to B(u, Vxu))$$

Suppose we have functionals V, Y satisfying the interpretation of implication together with an indirect interpretation of $\exists x \forall y A(x,y)$ i.e. a functional Φ such that

$$(*) \quad \forall g A(\Phi g, g(\Phi g)).$$

Suppose that a theorem $B :\equiv \forall u \exists v B(u,v)$ is proven using some nonconstructive lemma $A :\equiv \exists x \forall y A(x,y)$.

Naive idea. In order to find a function f satisfying $\forall u B(u, fu)$ we need to find some x satisfying $\forall y A(x,y)$. We cannot compute this x, therefore no computable f exists.

Recall the functional interpretation of implication:

$$(\exists x \forall y A(x,y) \to \forall u \exists v B(u,v)) \mapsto \exists V, Y \forall x, u(A(x, Yxu) \to B(u, Vxu))$$

Suppose we have functionals V, Y satisfying the interpretation of implication together with an indirect interpretation of $\exists x \forall y A(x,y)$ i.e. a functional Φ such that

$$(*) \quad \forall g A(\Phi g, g(\Phi g)).$$

For each u define the function $g_u : \mathbb{N} \to \mathbb{N}$ by $g_u(x) := Yxu$, and define

$$f(u) := V(\Phi g_u)u.$$

Why it works

Suppose that a theorem $B :\equiv \forall u \exists v B(u,v)$ is proven using some nonconstructive lemma $A :\equiv \exists x \forall y A(x,y)$.

Naive idea. In order to find a function f satisfying $\forall u B(u, fu)$ we need to find some x satisfying $\forall y A(x,y)$. We cannot compute this x, therefore no computable f exists.

Recall the functional interpretation of implication:

$$(\exists x \forall y A(x,y) \to \forall u \exists v B(u,v)) \mapsto \exists V, Y \forall x, u (A(x, Yxu) \to B(u, Vxu))$$

Suppose we have functionals V, Y satisfying the interpretation of implication together with an indirect interpretation of $\exists x \forall y A(x,y)$ i.e. a functional Φ such that

$$(*) \quad \forall g A(\Phi g, g(\Phi g)).$$

For each u define the function $g_u : \mathbb{N} \to \mathbb{N}$ by $g_u(x) := Yxu$, and define

$$f(u) := V(\Phi g_u)u.$$

Then for any input u, by $(*)$ we have $A(\Phi g_u, g_u(\Phi g_u)) \equiv A(\Phi g_u, Y(\Phi g_u)u)$. Therefore $B(u, V(\Phi g_u)u) \equiv B(u, f(u))$ holds.

The drinkers paradox as a lemma

Theorem

Let P be some predicate on the natural numbers. Then

$$\forall u \exists v (P(v) \rightarrow P(u^{v+7}))$$

The drinkers paradox as a lemma

Theorem

Let P be some predicate on the natural numbers. Then

$$\forall u \exists v (P(v) \to P(u^{v+7}))$$

Proof.

Fix some u. By the drinkers paradox there exists some x satisfying

$$P(x) \to \forall y P(y).$$

Set $v := x$. Then

$$P(v) \to \forall y P(y) \to P(u^{v+7}).$$

□

Can we find a function f satisfying

$$\forall u (P(f(u)) \to P(u^{f(u)+7}))?$$

An approximation to the drinkers paradox as a lemma

Our classical proof uses the implication

$$\exists x \forall y (P(x) \to P(y)) \to \forall u \exists v (P(v) \to P(u^{v+7}))$$

which has functional interpretation

$$\exists V, Y \forall x, u \left((P(x) \to P(\underbrace{Yxu}_{u^{x+7}})) \to (P(\underbrace{Vxu}_{x}) \to P(\underbrace{u^{Vxu+7}}_{u^{x+7}})) \right)$$

An approximation to the drinkers paradox as a lemma

Our classical proof uses the implication

$$\exists x \forall y (P(x) \to P(y)) \to \forall u \exists v (P(v) \to P(u^{v+7}))$$

which has functional interpretation

$$\exists V, Y \forall x, u \left((P(x) \to P(\underbrace{Yxu}_{u^{x+7}})) \to (P(\underbrace{Vxu}_{x}) \to P(\underbrace{u^{Vxu+7}}_{u^{x+7}}) \right)$$

This is solved by $Vxu := x$ and $Yxu := u^{x+7}$. But the indirect interpretation of the drinkers paradox:

$$\forall g (P(\Phi g) \to P(g(\Phi g)))$$

is solved by

$$\Phi g := \begin{cases} 0 & \text{if } P(g0) \\ g0 & \text{if } \neg P(g0) \end{cases}$$

An approximation to the drinkers paradox as a lemma

Our classical proof uses the implication

$$\exists x \forall y (P(x) \to P(y)) \to \forall u \exists v (P(v) \to P(u^{v+7}))$$

which has functional interpretation

$$\exists V, Y \forall x, u \left((P(x) \to P(\underbrace{Yxu}_{u^{x+7}})) \to (P(\underbrace{Vxu}_{x}) \to P(\underbrace{u^{Vxu+7}}_{u^{x+7}}) \right)$$

This is solved by $Vxu := x$ and $Yxu := u^{x+7}$. But the indirect interpretation of the drinkers paradox:

$$\forall g (P(\Phi g) \to P(g(\Phi g)))$$

is solved by

$$\Phi g := \begin{cases} 0 & \text{if } P(g0) \\ g0 & \text{if } \neg P(g0) \end{cases}$$

So putting these together, we have $g_u(x) := Yxu = u^{x+7}$ and therefore

$$f(u) := V(\Phi g_u)u = \Phi g_u = \begin{cases} 0 & \text{if } P(u^7) \\ u^7 & \text{if } \neg P(u^7) \end{cases}$$

Looking ahead

- Are there any non-trivial mathematical theorems, whose proofs can be analysed using the functional interpretation to obtain genuinely new numerical information?
- Do the indirect reformulations for $\forall\exists\forall$ statements have a mathematical meaning? Do they play a role in 'normal' mathematics?
- Can we extend the functional interpretation of classical arithmetic to stronger theories involving e.g. the axiom of choice, or Zorn's lemma?

Outline

References I

Kohlenbach, U. (2008).
Applied Proof Theory - Proof Interpretations and their Use in Mathematics.
Springer Monographs in Mathematics. Springer.