Interpreting classical theories in constructive ones

Jeremy Avigad

Department of Philosophy

Carnegie Mellon University

avigad+@cmu.edu

http://macduff.andrew.cmu.edu

A brief history of proof theory

Before the 19th century: There is no sharp distinction between constructive and nonconstructive reasoning in mathematics.

19th century: Foundational interest in the "concrete" content of abstract reasoning. Dedekind, Cantor, etc. introduce radically nonconstructive methods to mathematics. Kronecker objects.

Early 20th century: Hilbert tries to reconcile constructive and classical reasoning by justifying the latter on finitistic grounds.

1931: Gödel shows this to be infeasible.

Modified Hilbert's program: justify classical theories on constructive grounds; more generally, elucidate the relationships between them.

Classical theories vs. constructive theories

S^1_2	IS^1_2
$I\Sigma_1$	$I\Sigma^i_1$, PRA
PA, ACA_0 $\Sigma^1_1{-}AC_0$, KP	HA, T ML, IKP
$\Sigma^1_1{-}AC$, $\widehat{ID}_1$	$\Sigma^1_1{-}AC^i$, $ML+U$
ATR_0, $\widehat{ID}_{<\omega}$, KPl_0	$\widehat{ID}_{<\omega}$, $ML+U_{<\omega}$
$KP\omega$, ID_1 $\Pi^1_1{-}CA^-_0$	$IKP\omega$, $ID^{i,acc}_1$ CZF, $ML+U^e$
$\Delta^1_2{-}CA_0$, KPi	T_0, $IKPi$
Z_2	Z^i_2
ZFC	IZF

Bridging the gap

- The Gödel-Gentzen double-negation interpretation reduces PA to HA, Z_2 to Z_2^i, ZF to IZF.
- The Friedman-Dragalin translation recovers Π^0_2 theorems.

But these methods do not work for S^1_2, $I\Sigma_1$, $\Sigma^1_1{-}AC$, KP. For these purposes, we can turn to

- Ordinal analysis
- Functional interpretation

These methods provide additional information, but from the reductive point of view, they are indirect.

What goes wrong? Some examples:

- The double-negation interpretation of Σ_1 induction involves induction on predicates of the form $\neg\neg\exists x\, A(x,y)$ (or equivalently, $\neg\forall x\,\neg A(x,y)$).
- The double negation translation of the Σ^1_1 axiom of choice is of the form

  $$\forall x\, \neg\neg\exists Y\, \varphi(x,Y) \rightarrow \neg\neg\exists Y\, \forall x\, \varphi(x,Y_x)$$

  where φ is arithmetic.

Repairing the double-negation translation

We can supplement the double-negation translation with a generalization of the Friedman-Dragalin translation, and reduce

- S_2^1 to IS_2^1
- S_2 to IS_2
- $I\Sigma_1$ to $I\Sigma_1^i$
- PA to HA
- $\Sigma_1^1{-}AC$ to $\Sigma_1^1{-}AC^i$
- KP to IKP
  - with or without infinity
  - with or without $\mathbb{N}$ as urelements
  - with foundation for all or just Σ_1 formulae
  - without extensionality in IKP

Credits

Buchholz '81: Reduces theories of iterated inductive definitions ID_α to intuitionistic theories of strictly positive inductive definitions (and even accessibility ones).

Coquand '98: Inspired by the Buchholz translation (with $\alpha = 1$), finds a remarkably simple reduction for $I\Sigma_1$.

Avigad '98: Recasts the Coquand interpretation slightly, and extends it to the other theories mentioned.

(Coquand and Hofmann independently obtained a different reduction for S^1_2.)

The idea

Intuitionistic logic has a well-known constructive interpretation. Unfortunately, the negation of a formula, $\varphi \to \perp$, carries no useful constructive information.

- The Friedman-Dragalin solution: replace $\perp$ with a formula $\exists x\, A(x)$.
- The Buchholz-Coquand solution: replace $\perp$ dynamically; reinterpret implication as well.

A simple translation

Start with an intuitionistic language L, conditions $p, q, \ldots,$ an order relation $\prec$, and a forcing notion $p \Vdash A$ for atomic formulae A.

Assume $p \Vdash A$ is monotone, and $p \Vdash \perp$ implies $p \Vdash A$.

Define:

$$\begin{array}{rcl} p \Vdash (\varphi \wedge \psi) & \equiv & p \Vdash \varphi \wedge p \Vdash \psi \\ p \Vdash (\varphi \vee \psi) & \equiv & p \Vdash \varphi \vee p \Vdash \psi \\ p \Vdash (\varphi \rightarrow \psi) & \equiv & \forall q \preceq p \; (q \Vdash \varphi \rightarrow q \Vdash \psi) \\ p \Vdash \forall x \; \varphi & \equiv & \forall x \; p \Vdash \varphi \\ p \Vdash \exists x \; \varphi & \equiv & \exists x \; p \Vdash \varphi \end{array}$$

Write $\Vdash \varphi$ if every condition forces φ.

Notes:

1. Treat $\perp$ as an atomic formula
2. Monotonicity holds
3. If one has a "meet" operation, we have

$$p \Vdash (\varphi \rightarrow \psi) \equiv \forall q \; (q \Vdash \varphi \rightarrow p \wedge q \Vdash \psi)$$

The main theorem

Theorem. Suppose Γ proves φ intuitionistically. Then $\Vdash \Gamma$ proves $\Vdash \varphi$.

Corollary. Suppose in an intuitionistic theory T' we can define such a forcing relation and prove that every axiom of another theory T is forced. Then whenever T proves φ, T' proves $\Vdash \varphi$

The trick is to pick useful forcing conditions.

Interpreting $I\Sigma_1$ in $I\Sigma_1^i + (MP_{pr})$

Under the double-negation interpretation, induction on $\exists x\ B(x,y)$ translates to induction on $\neg\forall x\ \neg B(x,y)$. We would be happier if the latter formula were again Σ_1.

For primitive recursive matrices, Markov's principle takes the form

$$\neg\forall x\ A(x) \to \exists x\ \neg A(x) \qquad (MP_{pr})$$

In $I\Sigma_1^i$, (MP_{pr}) implies that the double-negation interpretation of any Σ_1 formula is again Σ_1, so $I\Sigma_1$ is interpretable in $I\Sigma_1^i + (MP_{pr})$.

Interpreting Markov's principle

To interpret (M_{pr}), use the forcing framework. Conditions p are finite sets of Π_1 sentences,

$$\{\forall x\ A_1(x), \forall x\ A_2(x), \ldots, \forall x\ A_k(x)\}.$$

Define $p \preceq q$ to be $p \supseteq q$.

Write $p \vdash \varphi$ for

$$\exists y\ (A_1(y) \wedge \ldots \wedge A_k(y) \rightarrow \varphi).$$

For θ atomic, define $p \Vdash \theta$ to be $p \vdash \theta$.

Note that we have

$$p \Vdash (\varphi \rightarrow \psi) \equiv \forall q\ (q \Vdash \varphi \rightarrow p \cup q \Vdash \psi).$$

Some details

Lemma. The following are provable in $I\Sigma^i_1$:

1. $\forall x\ A(x) \Vdash \forall x\ A(x)$

2. If $p \Vdash \neg\forall x\ A(x)$, then $p \Vdash \exists x\ \neg A(x)$.

Proof. For 1, we have

$$\begin{aligned} \forall x\ A(x) \Vdash \forall x\ A(x) &\equiv \forall z\ (\forall x\ A(x) \Vdash A(z)) \\ &\equiv \forall z\ (\forall x\ A(x) \vdash A(z)) \\ &\equiv \forall z\ \exists y\ (A(y) \to A(z)). \end{aligned}$$

For 2, let p be the set $\{\forall x\ B_1(x), \ldots, \forall x\ B_k(x)\}$, and suppose $p \Vdash \neg\forall x\ A(x)$. Then whenever $q \Vdash \forall x\ A(x)$, we have $p, q \Vdash \bot$.

By 1, we have $p, \forall x\ A(x) \Vdash \bot$. In other words,

$$\exists y\ (B_1(y) \wedge \ldots \wedge B_k(y) \wedge A(y) \to \bot)$$

which implies

$$\exists x, y\ (B_1(y) \wedge \ldots \wedge B_k(y) \to \neg A(x)),$$

which is to say

$$\exists x\ (p \vdash A(x)).$$

But this is just $p \Vdash \exists x\ A(x)$.

Conclusion

Theorem. If $I\Sigma_1^i + (MP_{pr})$ proves φ then then $I\Sigma_1^i$ proves $\Vdash \varphi$.

Proof. The preceeding lemma handles (MP_{pr}), induction on $\exists x\ B(x, y)$ translates to induction on $p \Vdash \exists x\ B(x, y)$, and the quantifier-free axioms are easy.

Corollary. $I\Sigma_1^i + (MP_{pr})$, and hence $I\Sigma_1$, are conservative over $I\Sigma_1^i$ for Π_2^0 sentences.

Proof. $\Vdash \forall x\ \exists y\ A(x, y)$ is equivalent to $\forall x\ \exists y\ A(x, y)$.

Admissible set theory

In the language of set theory, take equality to be *defined* by

$$x = y \equiv \forall z\ (z \in x \leftrightarrow z \in y).$$

The axioms of Kripke-Platek set theory (KP) are as follows:

1. Extensionality: $x = y \rightarrow (x \in w \rightarrow y \in w)$
2. Pair: $\exists x\ (y \in x \wedge z \in x)$
3. Union: $\exists x\ \forall z \in y\ \forall w \in z\ (w \in x)$
4. Δ_0 separation: $\exists x\ \forall z\ (z \in x \leftrightarrow z \in y \wedge \varphi(z))$ where φ is Δ_0 and x does not occur in φ
5. Δ_0 collection: $\forall x \in z\ \exists y\ \varphi(x, y) \rightarrow \exists w\ \forall x \in z\ \exists y \in w\ \varphi(x, y)$, where φ is Δ_0
6. Foundation: $\forall x\ (\forall y \in x\ \psi(y) \rightarrow \psi(x)) \rightarrow \forall x\ \psi(x)$, for arbitrary ψ

Note that the double-negation interpretation of collection is equivalent to

$$\forall x \in z\ \neg\forall y\ \neg\varphi^N(x, y) \rightarrow \neg\forall w\ \neg\forall x \in z\ \neg\forall y \in w\ \neg\varphi^N(x, y).$$

A three-step reduction

1. Remove extensionality:

   interpret KP in KP^{int}

2. Apply a double-negation translation:

   interpret KP^{int} in $IKP^{int,\#} + (MP_{res})$

3. Use a forcing relation:

   interpret $IKP^{int,\#} + (MP_{res})$ in IKP^{int}

Eliminating extensionality

Life in an intensional universe can be strange. For example, there may be many "empty sets". That is: we can have simultaneously,

$$\forall z\ (z \notin x), \forall z\ (z \notin y), x \in w, y \notin w.$$

Friedman: to interpret extensionality, say "x is isomorphic to y," $x \sim y$, if

$$\forall u \in x\ \exists v \in y\ (u \sim v) \wedge \forall u \in y\ \exists v \in x\ (u \sim v).$$

Then replace "element of" by "isomorphic to an element of"; i.e. define

$$x \in^* y \equiv \exists u \in x\ (y \sim u).$$

To make this work in the context of KP, one needs to show that isomorphism is Δ definable.

Theorem. KP is interpretable in KP^{int}.

The intermediate theory

Define an intermediate theory, $IKP^{int,\#}$, with axioms:

1. Pair and union: as before
2. Δ_0 separation: for negative formulae only
3. Δ_0 collection$^{\#}$:
   $$\forall x \in z \, \exists y \, \varphi(x,y) \rightarrow \exists w \, \forall x \in z \, \neg \forall y \in w \, \neg \varphi(x,y)$$
   where φ is Δ_0 and negative.
4. Foundation: for negative formulae only

Define an axiom schema, (MP_{res}):

$$\neg \forall x \, \varphi \rightarrow \exists w \, \neg \forall x \in w \, \neg \varphi$$

for Δ_0 formulae φ.

Theorem. KP^{int} is interpretable in $IKP^{int,\#} + (MP_{res})$.

The forcing relation

Take conditions p to be finite sets of Π_1 setences,

$$\{\forall x\ \varphi_1(x), \forall x\ \varphi_2(x), \ldots, \forall x\ \varphi_k(x)\},$$

where each φ_i is Δ_0.

Write $p \vdash \psi$ for

$$\exists y\ (\forall x \in y\ \varphi_1(x) \wedge \ldots \wedge \forall x \in y\ \varphi_k(x) \rightarrow \psi).$$

For θ atomic, define $p \Vdash \theta$ to be $p \vdash \theta$.

Some details

Lemma. If φ is negative and Δ_0, then IKP^{int} proves the all the following:

1. $p \Vdash \varphi$ is equivalent to $p \vdash \varphi$.
2. If $p \Vdash \neg\forall x\ \varphi$ then $p \Vdash \exists w\ \neg\forall x \in w\ \varphi$
3. If $p \Vdash \forall x \in y\ \exists z\ \varphi$ then $p \Vdash \exists w\ \forall x \in y\ \neg\forall z \in w\ \neg\varphi$

Theorem. $IKP^{int,\#} + (MP_{res})$ is interpretable in IKP^{int}.

Corollary. If KP^{int} proves $\forall x\ \exists y\ \varphi$, where φ is Δ_0, then IKP^{int} proves $\forall x\ \exists w\ \neg\forall y \in w\ \neg\varphi$.

Interpreting $\Sigma^1_1{-}AC$

$\Sigma^1_1{-}AC$ is a theory in the language of second-order arithmetic with axioms

1. the quantifier-free axioms of PA

2. induction

3. arithmetic comprehension

4. arithmetic choice:

$$\forall x\ \exists Y\ \varphi(x, Y) \to \exists Y\ \forall x\ \varphi(x, Y_x)$$

   where φ is arithmetic and the second "Y" codes a sequence of sets.

To interpret $\Sigma^1_1{-}AC$, replace arithmetic choice by

$$\forall x\ \exists Y\ \varphi(x, Y) \to \exists W\ \forall x\ \exists Y \in W\ \varphi(x, Y),$$

where W codes a countable collection of sets. Then "proceed as before," using a version of (MP) for arithmetic formulae.

Final questions

1. We now have yet another way of showing that PA is Π_2 conservative over HA. How does this relate to other methods?

2. Can this be extended to other theories, like ATR_0, KPl, or KPi?