Small initial segments and consistency

Zofia Adamowicz

Institute of Mathematics, Polish Academy of Science

Model Theory and Proof Theory of Arithmetic
A Memorial Conference in Honor of Henryk Kotlarski and Zygmunt Ratajczyk

Introduction and motivation

The main association a mathematician has with the word consistency is probably the Gödel Second Incompleteness Theorem.

$$T \nvdash Cons(T).$$

However, this does not always hold. There are some restrictions on T (Willard), but also there are some requirements of the predicate *Cons* we use (Pudlak). For instance

$$I\Delta_0 \vdash HCons^J(I\Delta_0),$$

where J is some definable initial segment (e.g. $\log\log\log$). Our focuss will be on $Cons^J(\cdot)$ (consistency relativized to J), for some definable initial segment J.

Introduction and motivation

Assume T is recursive, consistent and contains $I\Delta_0 + B\Sigma_1$.

Small initial segments

We consider initial segments $J = J_T$ depending on T. The definition of T is built into the definition of J_T.

We assume that J is a Σ_1 or Π_1 formula. We shall identify J with the set definable by the formula J.
We are interested in the following properties of J:

Small initial segments

When an initial segment is small? Key properties

- J is an initial segment provably in T,
- $\mathbb{N} \subseteq J$ provably in T
- J is $\mathbb{N}$ in some non standard models of T.

Is there a non trivial Σ_1 or Π_1 definable J with the above properties?

What J can be

The amount of consistency of T

Let $Cons(\cdot)$ denote the Hilbert or the Herbrand consistency predicate. Let $Cons(\cdot)^x$ express the meaning that there is no inconsistency proof which is $\leq x$.

Consider the following definable initial segment J: let $x \in J$ iff $Cons^x(T)$.

Note that the definition of J depends on the formula defining T. Thus, we should write J_T.

We shall call J_T *the amount of consistency of* T.

The amount of consistency of T

- $x \in J_T$ iff $Cons^x(T)$

Evidently, J has the *non Gödel* property:

$$T \vdash Cons^{J_T}(T).$$

Note that J_T has the following properties:

- J_T is an initial segment provably in T,
- $\mathbb{N} \subseteq J_T$ provably in T
  However, J_T is not $\mathbb{N}$ in non standard models of T (is not small).

The amount of consistency of T. Questions

Let $Pr_T(\cdot)$ be defined as $\neg Cons(T + \neg\cdot)$.

We have:

- If $Pr_T^x(\phi)$, then $Cons^{x+y}(T)$ implies $Cons^y(T + \phi)$.

Question
For what ϕ, y, $Cons^{x+y}(T)$ implies $Cons^x(T + \phi)$?
Candidate $\neg Cons(T)$.

The amount of consistency of T. Questions

In the classical case we have:

$$Cons(T) \Rightarrow Cons(T + \neg Cons(T))$$

Question

$$Cons^x(T) \Rightarrow Cons^x(T + \neg Cons(T))?$$

Question
For what ϕ

$$Pr_T^x(\phi)) \Rightarrow Pr_T^x(0 = 1)?$$

Candidate: $Cons(T)$.

What J_T can be

The amount of consistency of the Π_1 or Σ_1 truth

By $Cons^x(T + \Sigma_1)$ we shall mean the sentence stating the following: for every Σ_1 sentence η if $Sat_{\Sigma_1}(\eta)$ holds and $\eta \leq x$, then $Cons^x(T + \eta)$ holds.

By $Cons^x(T + \Pi_1)$ we shall mean the sentence stating the following: for every Π_1 sentence η if $Sat_{\Pi_1}(\eta)$ holds and $\eta \leq x$, then $Cons^x(T + \eta)$ holds.

What J_T can be

The amount of consistency of the Π_1 or Σ_1 truth
Assume $T \supseteq I\Delta_0 + exp$ and $\mathbb{N} \models T$.

Consider the following definable initial segment J_T: let $x \in J_T$ iff $Cons^x(T + \Pi_1)$. This initial segment is Σ_1 definable.
We shall call J_T *the amount of consistency of* Π_1*-truth*.

Dual
Consider the following definable initial segment J_T: let $x \in J_T$ iff $Cons^x(T + \Sigma_1)$. This initial segment is Π_1 definable.
We shall call J_T *the amount of consistency of* Σ_1*-truth*.

The amount of consistency of the Π_1 or Σ_1 truth

Is there a non standard model of T in which $J_T = \mathbb{N}$, i.e. is J_T small?

Can J_T be a closed under successor (be a cut)?

What J_T can be

Consider the following formula $\mathbb{N}_{T,\Pi_1}(x)$ expressing the meaning that there is a set (i.e. a characteristic function of a set) of size x consisting of Π_1 sentences containing all true Π_1 sentences and x-consistent with T:

$$\exists t \in \{0,1\}^x \Big(\forall \varphi < x \big(Sat_{\Pi_1}(\varphi) \Rightarrow t(\varphi) = 1 \big)$$

$$\&\text{the theory } \{\varphi < x : \ t(\varphi) = 1\} \text{ is } x\text{-consistent with } T \Big)$$

We may call $\mathbb{N}_{T,\Pi_1}(x)$, *the amount of codability of the* Π_1 *truth*. This is Σ_1.

Dual

Consider the following formula $\mathbb{N}_{T,\Sigma_1}(x)$ expressing the meaning that there is a set (i.e. a characteristic function of a set) of size x consisting of Σ_1 sentences containing all true Σ_1 sentences and x-consistent with T:

$$\exists t \in \{0,1\}^x \Big(\forall \varphi < x \big(Sat_{\Sigma_1}(\varphi) \Rightarrow t(\varphi) = 1 \big)$$

$$\&\text{the theory } \{\varphi < x : \ t(\varphi) = 1\} \text{ is } x\text{-consistent with } T \Big)$$

that is

$$\forall y \exists t \in \{0,1\}^x \Big(\forall \varphi < x \big(Sat_{\Sigma_1}(\varphi^y) \Rightarrow t(\varphi) = 1 \big)$$

$$\&\text{the theory } \{\varphi < x : \ t(\varphi) = 1\} \text{ is } x\text{-consistent with } T \Big)$$

We may call $\mathbb{N}_{T,\Sigma_1}(x)$, *the amount of the codability of the* Σ_1 *truth*. This is Π_1.

Without exponentiation

- $\mathbb{N}_{T,\forall\Sigma_m^b}$ = the amount of codability of $\forall\Sigma_m^b$ truth
- $\mathbb{N}_{T,\exists\Pi_m^b}$ = the amount of codability of $\exists\Pi_m^b$ truth
- the amount of consistency of $\exists\Pi_m^b$ truth
- the amount of consistency of $\forall\Sigma_m^b$ truth

The amount of codability of the Π_1 or Σ_1 truth. Questions

For what T,

- $\mathbb{N}_{T,\Pi_1}$, $\mathbb{N}_{T,\Sigma_1}$ are small, i.e. $= \mathbb{N}$ in some non standard model of T?
- $\mathbb{N}_{T,\Pi_1} = \mathbb{N}_{T,\Sigma_1}$ in some non standard model of T?
- $\mathbb{N}_{T,\Sigma_1} < \mathbb{N}_{T,\Pi_1}$ in some model of T?
- $\mathbb{N}_{T,\Pi_1} < \mathbb{N}_{T,\Sigma_1}$ in some model of T?

The amount of codability of the Π_1 or Σ_1 truth. Questions

The most interesting is $\mathbb{N}_{T,\Pi_1}$.
Can $\mathbb{N}_{T,\Pi_1}$ be non standard and closed under successor (be a cut)?
Can $\mathbb{N}_{T,\Pi_1}$ be a non standard model of $I\Delta_0$?
Can $\mathbb{N}_{T,\Pi_1}$ be a non standard model of $I\Delta_0 + exp$?
We may also consider $\mathbb{N}_{T,\Pi_1}$ in a model of a theory which is weaker than T, e.g. $\mathbb{N}_{I\Delta_0,\Pi_1}$ in a model of Q or $\mathbb{N}_{I\Delta_0+exp,\Pi_1}$ in a model of $I\Delta_0$.

Definition

Recall:

$$\omega_0(x) = x^2,$$

$$\omega_1(x) = 2^{(\log x)^2},$$

$$\omega_2(x) = 2^{2^{(\log\log x)^2}},$$

$$\omega_{i+1}(x) = 2^{\omega_i(\log x)}.$$

$$\Omega_i : \ \forall x \exists y \ y = \omega_i(x).$$

$\log_0(x) = x, \ \log_{i+1}(x) = \log\log_i(x).$

$\log^{i+2}$

We have

$$I\Delta_0 + \Omega_i \nvdash \forall x \exists \omega_i^{\log_{i+2} x}(x)$$

$$I\Delta_0 + \Omega_{i+1} \vdash \forall x \exists \omega_1^{\log_{i+2} x}(x).$$

For, the following formula is easy to be checked by induction:

$$\omega_1^n(x) = 2^{(\log x)^{2^n}},$$

for $n \geq 1$.
Hence, for instance,

$$\omega_1^{\log^3 x}(x) = 2^{(\log x)^{2^{\log^3 x}}} == 2^{(\log x)^{\log \log x}} = 2^{2^{(\log \log x)^2}} = \omega_2(x).$$

Examples

Assume $T = I\Delta_0 + \Omega_i$, $i \geq 0$. Then

$$T \vdash HCons^{\log_{i+3}}(T)$$

and if $I\Delta_0$ is finitely axiomatizable,

$$T \vdash HCons^{\log_{i+2}}(T).$$

Hence, if our *Cons* is *Hcons*, then

- the amount of consistency of $T \supseteq \log_{i+3}$.

and if $I\Delta_0$ is finitely axiomatizable,

- the amount of consistency of $T \supseteq \log_{i+2}$.

Examples

Assume that $I\Delta_0$ is finitely axiomatizable. Assume $T = I\Delta_0 + \Omega_i$, $i \geq 0$. Then
Let $T^{\#} \subseteq \Sigma_1$ be maximal consistent with T. Assume that elements Σ_1 definable are downward cofinal in M above $\mathbb{N}$.
Suppose that in M the amount of consistency of Σ_1 truth is $> \mathbb{N}$. Then $M \models Hcons^I(T + \Sigma_1)$ for an Σ_1 definable nonstandard I. But then, by maximality of $T^{\#}$, $Hcons^I(T + \phi)$ gives a Π_1 truth definition for Σ_1 sentences in M, contradiction.
Thus, in M we have:

- the amount of consistency of Σ_1 truth $= \mathbb{N}$.

Thus, J_T=the amount of consistency of Σ_1 truth, is small.
What about the amount of consistency of Π_1 truth?

Examples

Since

- the amount of consistency of Σ_1 truth $\geq$ the amount of codabiltiy of Σ_1 truth.

We have in M:

- the amount of consistency of Σ_1 truth $=$ the amount of codabiltiy of Σ_1 truth$= \mathbb{N}$.

The same can be shown without the assumption that $I\Delta_0$ is finitely axiomatizable with $\exists\Pi^b_m$ in place of Σ_1.

What about the amount of codability of Π_1 truth?

Examples

Assume that $I\Delta_0$ is finitely axiomatizable. Assume $T = I\Delta_0 + \Omega_i$, $i \geq 0$. Let $M \models I\Delta_0 + \Omega_{i+1}$. Then $M \models Hcons^{\log_{i+2}}(T + \Sigma_1)$ (by the fact that $\omega_{i+1}(x) = \omega_i^{\log_{i+2}(x)}(x)$). Hence

- the amount of consistency of Σ_1 truth $\supseteq \log_{i+2}$.

Examples

Assume $T = I\Delta_0 + \Omega_i$, $i \geq 0$.
Let $T^{\#} \subseteq \exists\Pi^b_m$ be maximal consistent with $I\Delta_0 + \Omega_{i+1}$. Let $M \models I\Delta_0 + \Omega_{i+1} + T^{\#}$. Suppose $M \models Hcons^{\log_{i+2}}(T + \exists\Pi^b_m)$. Since $Hcons^{\log_{i+2}}(T + \phi)$ is not a truth definition for $\exists\Pi^b_m$ sentences in M, for some $\phi \in \exists\Pi^b_m$, $\phi \notin T^{\#}$, $M \models Hcons^{\log_{i+2}}(T + \phi)$. But then ϕ is consistent with T and inconsistent with $I\Delta_0 + \Omega_{i+1}$, by maximality of $T^{\#}$. Thus, if in M,

- the amount of consistency of $\exists\Pi^b_m$ truth $\supseteq \log_{i+2}$,

then $I\Delta_0 + \Omega_{i+1}$ is not Π_1 conservative over $I\Delta_0 + \Omega_i$.

Consistency

Recall By $Cons^J(T + \Sigma_1)$ we shall mean the sentence stating

the following: for every Σ_1 sentence η if $Sat_{\Sigma_1}(\eta)$ holds and $\eta \in J$, then $Cons^J(T + \eta)$ holds.

By $Cons^J(T + \Pi_1)$ we shall mean the sentence stating the following: for every Π_1 sentence η if $Sat_{\Pi_1}(\eta)$ holds and $\eta \in J$, then $Cons^J(T + \eta)$ holds.
For $J_T = \mathbb{N}_{T,\Pi_1}$ or $J_T = \mathbb{N}_{T,\Sigma_1}$, by definition we have the non Gödel property:

$$T \vdash Cons^{\mathbb{N}_{T,\Pi_1}}(T + \Pi_1),$$

$$T \vdash Cons^{\mathbb{N}_{T,\Sigma_1}}(T + \Sigma_1).$$

Questions

$$T + \Pi_1\text{-truth} \vdash Cons^{\mathbb{N}_{T,\Sigma_1}}(T + \Pi_1)?$$

$$T + \Sigma_1\text{-truth} \vdash Cons^{\mathbb{N}_{T,\Pi_1}}(T + \Sigma_1)?$$

The answer is NO, even without exponentiation (for $\exists\Pi^b_m$ and $\forall\Sigma^b_m$ instead of Σ_1, Π_1).
Thus, we have the Gödel property.

Consistency

Usually a predicate $Cons(\cdot)$ is considered as expressing consistency if

T is consistent iff $\mathbb{N} \models Cons(T)$.
Some other properties are usually expected, e.g. the Hilbert Bernays derivability conditions:

- $T \vdash \phi$ implies $T \vdash Pr_T(\phi)$
- $T \vdash \big(Pr_T(\phi) \Rightarrow Pr_T(Pr_T(\phi))\big)$
- $T \vdash \Big(\big(Pr_T(\phi) \& Pr_T(\phi \Rightarrow \psi)\big) \Rightarrow Pr_T(\psi)\Big)$

Consistency

Note two other useful properties:
$Cons(T) \& Pr_T(\phi)$ implies $Cons(T + \phi)$.
If $Cons^J(\cdot)$ denotes *Cons* relativized to a definable initial segment J, then
$Cons^{2J}(T) \& Pr_T^J(\phi)$ implies $Cons^J(T + \phi)$.
We shall call the above properties **basic**.

Later we shall consider some unusual consistency predicates $Cons^J(\cdot)$, for some initial segments J, having the basic properties.

Usual properties of consistency

1. $Cons(\cdot)$ is Π_1

2. Σ_1 **completeness:**

- (a) $T \vdash \big(\eta \Rightarrow Pr_T(\eta)\big)$ for $\eta \in \Sigma_1$
- (b) $T + Cons(T) \vdash \big(Pr_T(\eta) \Rightarrow \eta\big)$ for $\eta \in \Pi_1$
- (c) $Cons(T) \Leftrightarrow Cons(T + \Sigma_1)$

Proofs

Remarks:
(a) implies the first Hilbert Bernays condition. From (a) $Cons(T)$ is provably equivalent to $Cons(T + \Sigma_1)$;
Proof of (b):
Suppose $T + Cons(T) + Pr_T(\eta) + \neg\eta$. Then, by (a), $Pr_T(\neg\eta)$, whence $Pr_T(\neg\eta) \& Pr_T(\neg\eta)$, contradicting $Cons(T)$.

Usual properties of consistency

If we consider $Cons(T + \Sigma_1 + \cdot)$, then (a):

$$T \vdash \big(\eta \Rightarrow Pr_{T+\Sigma_1}(\eta)\big)$$

for $\eta \in \Sigma_1$, is for free.

$Cons^{\mathbb{N}_{T,\Pi_1}}(T+\Sigma_1)$

Consider $Cons^{J_T}(T+\Sigma_1+\cdot)$, where $J_T = \mathbb{N}_{T,\Pi_1}$.

1. $Cons^{J_T}(T+\Sigma_1+\cdot)$ is Π_1 (remark: $Cons^{\mathbb{N}_{T,\Sigma_1}}(T+\Pi_1+\cdot)$ is Σ_1)

2. Σ_1 completeness for free

- $T \vdash \left(\eta \Rightarrow Pr^{J_T}_{T+\Sigma_1}(\eta)\right)$ for $\eta \in \Sigma_1$
- $T + Cons^{2J_T}(T) \vdash \left(Pr^{J_T}_{T+\Sigma_1}(\eta) \Rightarrow \eta\right)$ for $\eta \in \Pi_1$
- * $Cons^{J_T}(T) \not\Leftrightarrow Cons^{J_T}(T+\Sigma_1)$

Usual properties

3. Π_1 **conservativeness**

- (a) $T + \neg Cons(T)$ is Π_1 conservative over T
- (b) $T + \eta \nvdash Cons(T)$, for $\eta \in \Sigma_1$
- (c) If $T^{\#} \subseteq \Sigma_1$ is maximal consistent with T, then "$\neg Cons(T)$" $\in T^{\#}$

Proofs

Proof of (a):
Let $\eta \in \Sigma_1$ and assume that $T + \eta$ is consistent. Suppose $T + \eta \vdash Cons(T)$. Then $T + \eta \vdash Cons(T) \& Pr_T(\eta)$, whence $T + \eta \vdash Cons(T + \eta)$, which contradicts the Gödel theorem for $T + \eta$.
Proof of (b): similar.
Proof of (c):
Let $\eta \in \Sigma_1$ and assume that $T + \eta$ is consistent. if $T + \eta \vdash Cons(T)$, then $T + \eta \vdash Cons(T + \eta)$, which contradicts the Gödel theorem for $T + \eta$. Thus $T + \neg Cons(T)$ is consistent.

$Cons^{\mathbb{N}_{T,\Pi_1}}(T+\Sigma_1)$

If $J_T = \mathbb{N}_{T,\Pi_1}$,

3. Π_1 conservativeness

- $T + \neg Cons^{J_T}(T+\Sigma_1)$ is Π_1 conservative over T
- $T + \eta \nvdash Cons^{J_T}(T+\Sigma_1)$, for $\eta \in \Sigma_1$
- If $T^\# \subseteq \Sigma_1$ is maximal consistent with T, then "$\neg Cons^{J_T}(T+\Sigma_1)$" $\in T^\#$

Usual properties

4. Gödel:

- (a) $T \nvdash Cons(T)$;
- (b) If T is true then $T \nvdash \neg Cons(T)$ (note that $T + \neg Cons(T) \vdash \neg Cons(T + \neg Cons(T))$)
- (c) If $\theta \Leftrightarrow Cons(T + \neg\theta)$ provably in T, then $\theta \Leftrightarrow Cons(T)$ provably in T
- (d) $T + Cons(T) \vdash Cons(T + \neg Cons(T))$

Proofs

Proof of (c): Work in T. Assume θ. Then $Cons(T + \neg\theta)$, whence, in particular, $Cons(T)$. Assume $Cons(T)$. Suppose $\neg\theta$. Since $\neg\theta$ is Σ_1 we infer $Cons(T + \neg\theta)$, whence θ.
Proof of (d): Let θ be as in (c). Then, by (c),
$T + Cons(T) \vdash \theta$, whence, by (c),
$T + Cons(T) \vdash Cons(T + \neg\theta)$.

$Cons^{\mathbb{N}_{T,\Pi_1}}(T+\Sigma_1)$

4.Gödel:

If $J_T = \mathbb{N}_{T,\Pi_1}$,

- $T \nvdash Cons^{J_T}(T+\Sigma_1)$
- * $T \nvdash \neg Cons^{J_T}(T+\Sigma_1)$ (this means that $J_{T+\neg Cons^{J_T}(T+\Sigma_1)} \neq J_T$)
- If $\theta \Leftrightarrow Cons^{J_T}(T+\Sigma_1+\neg\theta)$ provably in T, then $\theta \Leftrightarrow Cons^{J_T}(T+\Sigma_1)$ provably in T
- $T + Cons^{2J_T}(T+\Sigma_1) \vdash Cons^{J_T}(T+\neg Cons^{J_T}(T+\Sigma_1))$

$Cons^{\mathbb{N}_{T,\Pi_1}}(T + \Sigma_1)$

Special

5. If $J_T = \mathbb{N}_{T,\Pi_1}$ and the set of true Π_1 sentences is maximal consistent with T and is not coded, then $Cons^{J_T}(T + \Sigma_1)$ (consistency holds in short models)

6. $T \vdash Cons^{J_T}(T + \Pi_1)$

7. $T + Cons^{2J_T}(T + \Sigma_1)$ is Σ_1 conservative over T

For what T, $\mathbb{N}_{T,\Pi_1}$, $\mathbb{N}_{T,\Sigma_1}$ have the key properties?

Let T denote a Π_2 axiomatizable consistent recursive theory. E.g. $I\Delta_0 + exp$, $I\Delta_0 + \Omega_1$.

- T has pointwise Σ_1 definable models. Every model of T has a Σ_1 elementary submodel pointwise Σ_1 definable models.
- T has models in which the set $\Sigma_1(M)$ of true Σ_1 sentences is not coded.

The key properties of $\mathbb{N}_{T,\Pi_1}$, $\mathbb{N}_{T,\Sigma_1}$

Lemma 2.1. *For every $n \in \mathbb{N}$ and every model M of T, $M \models \mathbb{N}_{T,\Pi_1}(n)$, $M \models \mathbb{N}_{T,\Sigma_1}(n)$.*

Lemma 2.2. *For every theory $T^\# \subseteq \Pi_1$ which is is maximal consistent w.r.t. T and every model M of $T + T^\#$ having the property that $T^\#$ is not coded in M, $\mathbb{N}_{T,\Pi_1}$ defines $\mathbb{N}$ in M. For every theory $T^\# \subseteq \Sigma_1$ which is is maximal consistent w.r.t. T and every model M of $T + T^\#$ having the property that $T^\#$ is not coded in M, $\mathbb{N}_{T,\Sigma_1}$ defines $\mathbb{N}$ in M.*

The key properties of $\mathbb{N}_{T,\Pi_1}$, $\mathbb{N}_{T,\Sigma_1}$

Proof. Let M satisfy the requirements of the lemma.
We shall show that $\mathbb{N}_{T,\Sigma_1}$ defines $\mathbb{N}$ in M.
For, assume $x \in \mathbb{N}$. Let $t \in \{0,1\}^x$ be such that

$$t(\varphi) = 1 \text{ iff } M \models Sat_{\Sigma_1}(\varphi).$$

Then t is as required in $\mathbb{N}_{T,\Sigma_1}$.
Assume now $\mathbb{N}_{T,\Sigma_1}(x)$ and suppose $x > \mathbb{N}$. Take the $t \in M$ existing by $\mathbb{N}_{T,\Sigma_1}$. Then the theory

$$\{\varphi : \ M \models t(\varphi) = 1\}$$

is consistent with T since

$$M \models \text{the theory } \{\varphi < x : \ t(\varphi) = 1\} \text{ is } x\text{-consistent with } T.$$

The key properties of $\mathbb{N}_{T,\Pi_1}$, $\mathbb{N}_{T,\Sigma_1}$

On the other hand this theory contains $T^{\#}$, since whenever φ is true i.e. $M \models Sat_{\Sigma_1}(\varphi)$, then $t(\varphi) = 1$.
So, by the maximality of $T^{\#}$, the theory

$$\{\varphi : \ M \models t(\varphi) = 1\}$$

equals $T^{\#}$. But so, t is its code on M. Contradiction.

Existence of models whose Σ_1 truth is not coded

Theorem 2.3. *If $M \models T$ is pointwise Σ_1 definable then $\Sigma_1(M)$ is not coded in M.*

Proof. Suppose the converse. Let $x \in M$ be a code for $\Sigma_1(M)$. Let η be the Σ_1 definition of X. Then we have for ϕ running over Σ_1 sentences:

$$\phi \text{ iff } \forall x \big(\eta(x) \Rightarrow \phi \in x\big).$$

This gives a Π_1 definition of the Σ_1 truth. Contradiction with the Tarski theorem.

Existence of models whose Σ_1 or $\exists\Pi^b_m$ truth is not coded

Theorem 2.4. *Every model for $I\Delta_0$ has a Σ_1 elementary submodel satisfying $I\Delta_0 + B\Sigma_1$ whose Σ_1 truth is not coded.*

(A. J. Wilkie and J. B. Paris, On the existence of end extensions of models of bounded induction, in: Proceedings of the International Congress of Logic, Philosophy and Methodology of Sciences, Moscow 1987.)

Lemmas

Let J_T be Π_1 definable.

Lemma 3.1. *Let θ be the diagonal sentence such that*

$$T \vdash \big(\theta \Leftrightarrow Cons^{J_T}(T + \Pi_1 + \neg\theta)\big).$$

Call θ the Gödel sentence.
Then

$$T \vdash \big(\theta \Leftrightarrow Cons^{J_T}(T + \Pi_1)\big).$$

Proof. Work in T. Assume θ. Then, in particular, $Cons^{J_T}(T + \Pi_1)$. Assume now $Cons^{J_T}(T + \Pi_1)$. Suppose $\neg\theta$. Since $\neg\theta$ is Π_1 we infer $Cons^{J_T}(T + \Pi_1 + \neg\theta)$. Hence θ.

Lemma 3.2. *Let θ be the diagonal sentence such that*

$$T \vdash \big(\theta \Leftrightarrow Cons(T + \Sigma_1 + \neg\theta)\big).$$

Call θ the Gödel sentence.
Then

$$T \vdash \big(\theta \Leftrightarrow Cons(T + \Sigma_1)\big).$$

Proof.
Work in T. Assume θ. Then $Cons(T + \Sigma_1 + \neg\theta)$, whence, in particular, $Cons(T + \Sigma_1)$. Assume $Cons(T + \Sigma_1)$. Suppose $\neg\theta$. Since $\neg\theta$ is Σ_1 we infer $Cons(T + \Sigma_1 + \neg\theta)$, whence θ.

Corollary
$T \nvdash Cons(T + \Sigma_1)$.

Lemmas

Lemma 3.3. *The sentence $Cons^{J_T}(T + \Pi_1)$ is independent from T.*

Proof. To see that the theory $T + Cons^{J_T}(T + \Pi_1)$ is consistent it suffices to observe that is is true in every model M of T in which ${J_T}^M = \mathbb{N}$.
We shall prove that $T \nvdash Cons^{J_T}(T + \Pi_1)$. Suppose the converse. Let θ Gödel sentence. Then, by Lemma 3.1, $T \vdash \theta$. Let M be a model of T. Then $M \models \theta$. Thus, $M \models Cons^{J_T}(T + \Pi_1 + \neg\theta)$. Since ${J_T}^M \supseteq \mathbb{N}$, the theory $T + \neg\theta$ is consistent. But on the other hand $T \vdash \theta$. Contradiction.

In weak arithmetic

Here we replace Σ_1 by $\exists\Pi^b_m$ and Π by $\forall\Sigma^b_m$. In particular we consider the $\forall\Sigma^b_m$ formula $\mathbb{N}_{T,\exists\Pi^b_m}$. It has the key properties. We let J_T be $\mathbb{N}_{T,\exists\Pi^b_m}$.

1. $Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(\cdot + \forall\Sigma^b_m)$ is $\exists\Pi^b_m$

2. $\forall\Sigma^b_m$ completeness

- $T \vdash \left(\eta \Rightarrow Pr^{\mathbb{N}_{T,\exists\Pi^b_m}}_{T+\forall\Sigma^b_m}(\eta)\right)$ for $\eta \in \forall\Sigma^b_m$
- $T + Cons^{2\forall\Sigma^b_m}(T) \vdash \left(Pr^{\mathbb{N}_{T,\exists\Pi^b_m}}_{T+\forall\Sigma^b_m}(\eta) \Rightarrow \eta\right)$ for $\eta \in \exists\Pi^b_m$
- * $Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T) \not\Leftrightarrow Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$

In weak arithmetic

3. $\exists\Pi^b_m$ **conservativeness**

- $T + \neg Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$ is $\exists\Pi^b_m$ conservative over T
- $T + \eta \not\vdash Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$, for $\eta \in \forall\Sigma^b_m$
- If $T^{\#} \subseteq \forall\Sigma^b_m$ is maximal consistent with T, then "$\neg Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$" $\in T^{\#}$

In weak arithmetic

4. Gödel:

- $T \nvdash Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$
- * $T \nvdash \neg Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$
  (this means that $\mathbb{N}_{T+\neg Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T+\forall\Sigma^b_m),\exists\Pi^b_m} \neq \mathbb{N}_{T,\exists\Pi^b_m}$)
- If $\theta \Leftrightarrow Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m + \neg\theta)$ provably in T, then $\theta \Leftrightarrow Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m)$ provably in T
- $T + Cons^{2\forall\Sigma^b_m}(T + \forall\Sigma^b_m) \vdash$
  $Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \neg Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T + \forall\Sigma^b_m))$

In weak arithmetic

Special

5. If the set of true $\exists\Pi^b_m$ sentences is maximal consistent with T and is not coded, then $Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T+\forall\Sigma^b_m)$(consistency holds in tall models)

6. $T \vdash Cons^{\mathbb{N}_{T,\exists\Pi^b_m}}(T+\exists\Pi^b_m)$

7. $T + Cons^{2^{\mathbb{N}_{T,\exists\Pi^b_m}}}(T+\forall\Sigma^b_m)$ is $\forall\Sigma^b_m$ conservative over T