# Axiomatizing provable 1-provability

Evgeny Kolmakov

kolmakov-ea@yandex.ru

Lomonosov Moscow State University

Joint with Lev D. Beklemishev

Workshop on Proof Theory, Modal Logic and Reflection Principles
Moscow, Russia

October 19, 2017

# Preliminaries

We deal with r.e. consistent extensions of EA (or $I\Delta_0(\exp)$).

$$T \subseteq U \Leftrightarrow \text{every theorem of } T \text{ is a theorem of } U$$
$$T \equiv U \Leftrightarrow T \subseteq U \text{ and } U \subseteq T$$
$$T \equiv_\Gamma U \Leftrightarrow T \text{ and } U \text{ have the same } \Gamma\text{-consequences}$$

$\varphi$ is 1-provable in $T$, if $\varphi$ is provable in $T +$ all true $\Pi_1$-sentences.

$$[1]_T\varphi := \exists z\, (\mathrm{True}_{\Pi_1}(z) \wedge \Box_T(\mathrm{True}_{\Pi_1}(\dot{z}) \to \varphi))\,.$$

$[1]_T$ satisfies derivability conditions and is $\Sigma_2$-complete provably in EA:

1. If $T \vdash \varphi$, then $\mathrm{EA} \vdash [1]_T\varphi$.
2. $\mathrm{EA} \vdash [1]_T(\varphi \to \psi) \to ([1]_T\varphi \to [1]_T\psi)$.
3. $\mathrm{EA} \vdash [1]_T\varphi \to [1]_T[1]_T\varphi$.
4. $\mathrm{EA} \vdash \sigma(\vec{x}) \to [1]_T\sigma(\dot{\vec{x}})$, whenever $\sigma(\vec{x})$ is a $\Sigma_2$-formula.

# Preliminaries

The local reflection principle

$$\mathrm{Rfn}(T)\colon \Box_T\varphi \to \varphi,\ \ \varphi \text{ a sentence.}$$

If we restrict $\varphi \in \Gamma$ we get the partial reflection principle $\mathrm{Rfn}_\Gamma(T)$.

Given some elementary well-ordering $(D, \prec)$, the transfinite iteration of the local reflection schemata along $(D, \prec)$ is defined by formalizing the following condition

$$\mathrm{Rfn}(T)_\alpha \equiv T + \{\mathrm{Rfn}(\mathrm{Rfn}(T)_\beta) \mid \beta \prec \alpha\}.$$

# Provable 1-provability

A sentence $\varphi$ is provably 1-provable in PA, if $\mathrm{PA} \vdash [1]_{\mathrm{PA}}\varphi$.

Question. Can we axiomatize the set of all provably 1-provable sentences by the means of the standard provability predicate $\Box_{\mathrm{PA}}$ (without mentioning the notion of 1-provability)?

More generally, given a pair $T$ and $S$ of r.e. consistent extensions of EA we consider the set of sentences $T$-provably 1-provable in $S$

$$C_S(T) = \{\varphi \mid T \vdash [1]_S\varphi\}.$$

Derivability conditions for $[1]_S \Longrightarrow C_S(T)$ is a theory extending $S$.

$C_S(T)$ can be viewed as a theory with the provability predicate $\Box_T[1]_S$.

# Connection to local reflection

Fact. $\mathrm{Rfn}(S) \subseteq C_S(T)$.

Proof. Consider arbitrary instance $\Box_S \varphi \to \varphi$ of $\mathrm{Rfn}(S)$. We derive

$$\begin{aligned} T \vdash \Box_S \varphi &\to [1]_S \varphi \\ &\to [1]_S(\Box_S \varphi \to \varphi) \end{aligned}$$

The sentence $\neg \Box_S \varphi$ is $\Pi_1$, so by provable $\Sigma_2$-completeness of $[1]_S$

$$\begin{aligned} T \vdash \neg \Box_S \varphi &\to [1]_S \, (\neg \Box_S \varphi) \\ &\to [1]_S(\Box_S \varphi \to \varphi), \end{aligned}$$

This shows $T \vdash [1]_S(\Box_S \varphi \to \varphi)$, and hence $\mathrm{Rfn}(S) \subseteq C_S(T)$.

# Main results

In the same fashion using transfinite induction in PA one can show

$$\mathrm{Rfn}(S)_{\varepsilon_0} \subseteq C_S(\mathrm{PA}).$$

The natural hypothesis is then $C_S(\mathrm{PA}) \equiv \mathrm{Rfn}(S)_{\varepsilon_0}$, and the main difficulty is to prove the reverse inclusion

$$C_S(\mathrm{PA}) \subseteq \mathrm{Rfn}(S)_{\varepsilon_0}.$$

We prove the following results:

- For any $n \geqslant 1$ we have $C_S(I\Sigma_n) \equiv \mathrm{Rfn}(S)_{\omega_n}$.
- It follows that $C_S(\mathrm{PA}) \equiv \mathrm{Rfn}(S)_{\varepsilon_0}$.
- In general, if $\alpha$ is the $\Sigma^0_2$-ordinal of $T$ measured w.r.t. transfinite iterations of local $\Sigma_2$-reflection schema over EA, that is,

  $$T \equiv_{\Sigma_2} \mathrm{Rfn}_{\Sigma_2}(\mathrm{EA})_\alpha,$$

  then we have

  $$C_S(T) \equiv \mathrm{Rfn}(S)_{1+\alpha}.$$

# $\Sigma_2$-conservation results

$[1]_S\varphi$ is a $\Sigma_2$-sentence $\Rightarrow$ $C_S(T)$ depends on $\Sigma_2$-consequences of $T$.

We want to show $C_S(I\Sigma_n) \subseteq \mathrm{Rfn}(S)_{\omega_n} \Longrightarrow$ it would be more convenient for us to work with some form of iterated reflection under $C_S(\cdot)$ equivalent to the induction $I\Sigma_n$.

Theorem (Beklemishev, Visser; 2005). $\Sigma_2$-consequences of $I\Sigma_n$ are axiomatized by $\mathrm{Rfn}_{\Sigma_2}(\mathrm{EA})_{\omega_n}$ for $n \geqslant 1$.

Theorem (Beklemishev; 2003). $\mathrm{EA} \vdash \forall\alpha\, \mathrm{Rfn}_{\Sigma_2}(T)_\alpha \equiv_{\Sigma_2} \mathrm{Rfn}(T)_\alpha$.

Lemma. $C_S(I\Sigma_n) = C_S(\mathrm{Rfn}_{\Sigma_2}(\mathrm{EA})_{\omega_n}) = C_S(\mathrm{Rfn}(\mathrm{EA})_{\omega_n})$.

# Commutative property

Lemma. $\mathsf{EA} \vdash \forall\alpha\, C_S(\mathrm{Rfn}(T)_\alpha) \equiv \mathrm{Rfn}(C_S(T))_\alpha$.

Proof (sketch). Argue informally in EA by reflexive induction on $\alpha$:

$$\frac{\forall\alpha(\Box_{\mathsf{EA}}(\forall\beta \prec \alpha\, A(\beta)) \to A(\alpha))}{\forall\alpha\, A(\alpha)}$$

$\forall\alpha\, C_S(\mathrm{Rfn}(T)_\alpha) \supseteq \mathrm{Rfn}(C_S(T))_\alpha$ can be proved along the same lines as $S + \mathrm{Rfn}(S) \subseteq C_S(T)$.

We concentrate on $\forall\alpha\, C_S(\mathrm{Rfn}(T)_\alpha) \subseteq \mathrm{Rfn}(C_S(T))_\alpha$.

Denote $T^\alpha := \mathrm{Rfn}(T)_\alpha$ and $U^\alpha := \mathrm{Rfn}(C_S(T))_\alpha$.

We need to show $C_S(T^\alpha) \subseteq U^\alpha$ given $\Box_{\mathsf{EA}}(\forall\beta \prec \alpha\, C_S(T^\beta) \subseteq U^\beta)$.

Assume $T^\alpha \vdash [1]_S\varphi$. By the definition of $T^\alpha$ and the formalized deduction theorem there exist sentences $\psi_1, \ldots, \psi_n$ and $\beta \prec \alpha$ with

$$T^\beta \vdash \bigwedge_{i=1}^{n}(\Box_{T^\beta}\psi_i \to \psi_i) \to [1]_S\varphi.$$

We show that for any subset $I \subseteq \{1, \ldots, n\}$ it holds that

$$U^{\beta+1} \vdash \left(\bigwedge_{i\in I}\neg\Box_{T^\beta}\psi_i \wedge \bigwedge_{i\notin I}\Box_{T^\beta}\psi_i\right) \to \varphi,$$

whence $U^{\beta+1} \vdash \varphi$ follows, and hence $U^\alpha \vdash \varphi$, as required.

Main fact: provably in EA for any $I \subseteq \{1, \ldots, n\}$

$$T^\beta \vdash \bigwedge_{i\in I}(\Box_{T^\beta}\psi_i \to \psi_i) \to [1]_S\varphi \Longrightarrow T^\beta \vdash [1]_S\left(\left[\bigwedge_{i\in I}\neg\Box_{T^\beta}\psi_i\right] \to \varphi\right).$$

Now, fix some $I \subseteq \{1, \ldots, n\}$. By propositional logic we have

$$T^\beta \vdash \bigwedge_{i \notin I} \psi_i \to \bigwedge_{i \notin I} (\Box_{T^\beta} \psi_i \to \psi_i),$$

and hence

$$T^\beta \vdash \bigwedge_{i \notin I} \psi_i \to \left( \bigwedge_{i \in I} (\Box_{T^\beta} \psi_i \to \psi_i) \to [1]_S \varphi \right).$$

By $\Sigma_1$-completeness this implies

$$\mathsf{EA} \vdash \Box_{T^\beta} \bigwedge_{i \notin I} \psi_i \to \Box_{T^\beta} \left( \bigwedge_{i \in I} (\Box_{T^\beta} \psi_i \to \psi_i) \to [1]_S \varphi \right).$$

Using main fact we obtain

$$\mathsf{EA} \vdash \Box_{T^\beta} \bigwedge_{i \notin I} \psi_i \to \Box_{T^\beta} [1]_S \left( \left[ \bigwedge_{i \in I} \neg \Box_{T^\beta} \psi_i \right] \to \varphi \right).$$

By the reflexive induction hypothesis for $\beta$ we have

$$\mathsf{EA} \vdash \Box_{T^\beta}[1]_S \left( \left[ \bigwedge_{i \in I} \neg \Box_{T^\beta} \psi_i \right] \to \varphi \right) \to \Box_{U^\beta} \left( \left[ \bigwedge_{i \in I} \neg \Box_{T^\beta} \psi_i \right] \to \varphi \right),$$

and hence

$$\mathsf{EA} \vdash \Box_{T^\beta} \bigwedge_{i \notin I} \psi_i \to \Box_{U^\beta} \left( \left[ \bigwedge_{i \in I} \neg \Box_{T^\beta} \psi_i \right] \to \varphi \right).$$

Finally, this yields

$$U + \mathsf{Rfn}(U^\beta) \vdash \left( \bigwedge_{i \notin I} \Box_{T^\beta} \psi_i \right) \to \left( \left[ \bigwedge_{i \in I} \neg \Box_{T^\beta} \psi_i \right] \to \varphi \right).$$

# Computing $C_S(\mathrm{EA})$

Using the commutative property we get

$$C_S(I\Sigma_n) = C_S(\mathrm{Rfn}(\mathrm{EA})_{\omega_n}) \equiv \mathrm{Rfn}(C_S(\mathrm{EA}))_{\omega_n},$$

and hence it is left to find $C_S(\mathrm{EA})$.

Lemma. $\mathrm{EA}^+ \vdash C_S(\mathrm{EA}) \equiv S + \mathrm{Rfn}(S)$.

Proof (sketch). $\supseteq$ was shown above.

For $\subseteq$ we apply a version of the Herbrand theorem for $\Sigma_2$-formulas. Assume $\varphi \in C_S(\mathrm{EA})$, that is, $\mathrm{EA} \vdash [1]_S\varphi$, then

$$\mathrm{EA} \vdash \exists \pi \, (\mathrm{True}_{\Pi_1}(\pi) \wedge \Box_S(\mathrm{True}_{\Pi_1}(\pi) \to \varphi)) \, ,$$

and by the definition of $\mathrm{True}_{\Pi_1}$ in terms of $\mathrm{Sat}_{\Delta_0}$ this can be rewritten as the following $\Sigma_2$-sentence

$$\mathrm{EA} \vdash \exists p \, \exists \delta \, \forall x \, (\mathrm{Sat}_{\Delta_0}(\delta, x) \wedge \mathrm{Prf}_S(p, \ulcorner \mathrm{True}_{\Pi_1}(\forall t \delta) \to \varphi \urcorner)) \, .$$

Applying Herbrand theorem we get a sequence of terms

$$p_0, \delta_0, p_1(x_1), \delta_1(x_1), \ldots, p_k(x_1, \ldots, x_k), \delta_k(x_1, \ldots, x_k),$$

such that the following disjunction is provable in EA:

$$(\mathrm{Sat}_{\Delta_0}(\delta_0, x_1) \wedge \mathrm{Prf}_S(p_0, \ulcorner \mathrm{True}_{\Pi_1}(\forall t\, \delta_0) \to \varphi \urcorner)) \vee$$
$$(\mathrm{Sat}_{\Delta_0}(\delta_1(x_1), x_2) \wedge \mathrm{Prf}_S(p_1(x_1), \ulcorner \mathrm{True}_{\Pi_1}(\forall t\, \delta_1(x_1)) \to \varphi \urcorner)) \vee$$
$$(\mathrm{Sat}_{\Delta_0}(\delta_2(x_1, x_2), x_3) \wedge \mathrm{Prf}_S(p_2(x_1, x_2), \ulcorner \mathrm{True}_{\Pi_1}(\forall t\, \delta_2(x_1, x_2)) \to \varphi \urcorner)) \vee$$
$$\ldots$$
$$(\mathrm{Sat}_{\Delta_0}(\delta_k(\vec{x}^k), x_{k+1}) \wedge \mathrm{Prf}_S(p_k(\vec{x}^k), \ulcorner \mathrm{True}_{\Pi_1}(\forall t\, \delta_k(\vec{x}^k)) \to \varphi \urcorner)),$$

with $x_1, x_2, \ldots, x_{k+1}$ as free variables. We denote the $i$th member of the disjunction by $C_i(x_1, \ldots, x_i)$ for $i \in \{1, \ldots, k+1\}$. This formulas are $\Delta_1$, however, they can be replaced with $\Delta_0$-formulas (that leads to more complicated notations).

Our aim is to show $\mathrm{EA} + \mathrm{Rfn}(S) \vdash \varphi$. We successively consider each line of the disjunction.

For $i = 1$ we consider the cases $\forall x_1\, C_1(x_1)$ and $\exists x_1 \neg C_1(x_1)$:

- $\forall x_1\, C_1(x_1)$ implies $\mathrm{True}_{\Pi_1}(\forall t \delta_0) \wedge \Box_S(\mathrm{True}_{\Pi_1}(\forall t \delta_0) \to \varphi)$, and hence using $\mathrm{Rfn}(S)$ we derive $\varphi$.
- $\exists x_1 \neg C_1(x_1)$ implies the existence of $c_1 = \mu x. \neg C_1(x)$. Key properties: $c_1$ is $\Delta_0$-definable and $\neg C_1(c_1)$.

We fix this $c_1$ and continue in the same fashion for $i = 2$:

- $\forall x_2\, C_2(c_1, x_2)$ implies

$$\mathrm{True}_{\Pi_1}(\forall t \delta_1(c_1)) \wedge \Box_S(\mathrm{True}_{\Pi_1}(\forall t \delta_1(c_1)) \to \varphi).$$

We can't use the local reflection $\mathrm{Rfn}(S)$ right away, because we have the parameter $c_1$ under $\Box_S$.

However, using $\Delta_0$-definability and provable uniqueness of $c_1$ it can be shown that under the assumptions above we get

$$\forall y[(y = \mu x.\neg C_1(x)) \to \mathrm{True}_{\Pi_1}(\forall t \delta_1(y))]$$

and

$$\Box_S(\forall y[(y = \mu x.\neg C_1(x)) \to (\mathrm{True}_{\Pi_1}(\forall t \delta_1(y)) \to \varphi)]).$$

In this way we get rid of $c_1$ under $\Box_S$. Now we are able to use $\mathrm{Rfn}(S)$, and in this case we also derive $\varphi$ as for $i = 1$.

- $\exists x_2 \neg C_2(c_1, x_2)$ implies the existence of $c_2 = \mu x.\neg C_2(c_1, x)$. As above we have $\neg C_2(c_1, c_2)$.

We proceed in a similar way for $i = 3, 4, \ldots,$ and as above we either derive $\varphi$ or falsify the corresponding line of the disjunction. It can't be that all the lines are falsified, so we must obtain $\varphi$ at some step.

Hence we showed $\mathrm{EA} \vdash [1]_S \varphi \Longrightarrow \mathrm{EA} + \mathrm{Rfn}(S) \vdash \varphi$, as required.

Combining all the Lemmas we get the following chain of equivalences

| | |
|---|---|
| | $C_S(I\Sigma_n)$ |
| $\Sigma_2$-consequences of $I\Sigma_n$ | $\equiv$ |
| | $C_S(\mathrm{Rfn}_{\Sigma_2}(\mathrm{EA})_{\omega_n})$ |
| $\Sigma_2$-conservativity of Rfn over $\mathrm{Rfn}_{\Sigma_2}$ | $\equiv$ |
| | $C_S(\mathrm{Rfn}(\mathrm{EA})_{\omega_n})$ |
| Commutative property | $\equiv$ |
| | $\mathrm{Rfn}(C_S(\mathrm{EA}))_{\omega_n}$ |
| $C_S(\mathrm{EA}) \equiv S + \mathrm{Rfn}(S)$ | $\equiv$ |
| | $\mathrm{Rfn}(S + \mathrm{Rfn}(S))_{\omega_n}$ |
| $1 + \omega_n = \omega_n$ for $n \geqslant 1$ | $\equiv$ |
| | $\mathrm{Rfn}(S)_{\omega_n}$ |

**Thank you for your attention!**