# Complexity of the interpretability logic IL

Luka Mikec[1] Fedor Pakhomov[2] Mladen Vuković[1]

[1]Department of Mathematics, Faculty of Science, University of Zagreb, Croatia

[2]Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia;
Research University Higher School of Economics, Moscow, Russia

Dubrovnik, 2018

# Interpretability

- Let $T_1$ and $T_2$ be some first order theories given through their sets of axioms (*not* sets of theorems).
- Roughly, translation of $T_2$ in $T_1$ is a pair $(f, U)$ where:
  - $f(R(\vec{x})) = f(R)(\vec{x})$;
  - $f(A \to B) = f(A) \to f(B)$ etc.;
  - $f(\forall x F) = \forall x(U(x) \to f(F))$ etc.;
- Thus, translation mostly preserves structure.
- $T_1$ interprets $T_2$ ($T_1 \rhd T_2$) if

$$T_2 \vdash F \;\Rightarrow\; T_1 \vdash f(F),$$

for all sentences $F \in \mathcal{L}(T_2)$.

# Interpretability

- In particular, we can study interpretability between finite extensions of a given theory.
- For example, since $PA \nvdash \Diamond\top$, we have:

$$PA + \Diamond_{PA}\top \;\rhd\; PA,$$

  where $\Diamond_{PA}\top$ formalizes consistency of $PA$ within $PA$.
- Furthemore, we can ask which *interpretabilities* can be proven within the base theory.

$$T + A \rhd T + B \quad \Rightarrow \quad T \vdash Int(\ulcorner A \urcorner, \ulcorner B \urcorner)$$

# Interpretability logics

- The language of interpretability logics is given by

$$A ::= p \,|\, \bot \,|\, A \to A \,|\, \Box A \,| A \rhd A,$$

  where $p$ is a propositional variable.
- Let $T$ be a formal theory, and $Int(\ulcorner A \urcorner, \ulcorner B \urcorner)$ a sentence formalizing $T + A \ \rhd \ T + B$.
- Arithmetical interpretation $*$ assigns sentences to modal formulas, such that:
  - $(A \to B)^* = A^* \to B^*$ etc.;
  - $p^*$ is a sentence (fixed by $*$);
  - $(\Box A)^* = Pr_T(A^*)$;
  - $(A \rhd B)^* = Int_T(A^*, B^*)$.

# Interpretability logics

- Given a formal theory $T$,

$$A \in IL(T) \;:\Leftrightarrow\; T \vdash A^*.$$

- Research focuses on sufficiently strong theories, able to deal with syntax ("sequential theories").
- Interpretability logics of sequential theories contain the basic interpretability logic IL.

# Basic interpretability logic **IL**

- Basic interpretability logic **IL**:

  propositionally valid formulas (in the new language);

  K $\Box(A \to B) \to (\Box A \to \Box B)$;

  Löb $\Box(\Box A \to A) \to \Box A$;

  J1 $\Box(A \to B) \to A \rhd B$;

  J2 $(A \rhd B) \ \wedge \ (B \rhd C) \to A \rhd C$;

  J3 $(A \rhd C) \ \wedge \ (B \rhd C) \to A \vee B \rhd C$;

  J4 $A \rhd B \to (\Diamond A \to \Diamond B)$;

  J5 $\Diamond A \rhd A$.

  - rules: modus ponens and necessitation $A/\Box A$.

  (parentheses priority: $\neg, \Box, \Diamond; \ \wedge, \vee; \ \rhd; \ \to, \leftrightarrow$)

- $\Box A$ is provably equivalent with $\neg A \rhd \bot$, and $\Diamond A$ is defined as $\neg\Box\neg A$.
- There is no $T$ such that $IL(T) = IL$. In fact, we always have $IL \subset ILW \subseteq IL(T)$. But, $IL$ has nice semantics.

# Models

- Semantics: extend the usual relational (Kripke) model.
- **IL**-frame (Veltman frame): $\mathcal{F} = \langle W, R, \underline{\{S_w \; : \; w \in W\}} \rangle$, where:
  1. $W \neq \emptyset$;
  2. $R^{-1}$ is well-founded (no $x_0 R x_1 R x_2 R \ldots$ chains);
  3. $R$ is transitive;
  4. $S_w \subseteq R(w)^2$ is reflexive, transitive, contains $R \cap R(w)^2$ ($wRuRv$ implies $uS_w v$);
- **IL**-model (Veltman model): $\mathcal{M} = \langle W, R, \{S_w \; : \; w \in W\}, V \rangle$, where:
  1. $\langle W, R, \{S_w \; : \; w \in W\} \rangle$ is a **IL**-frame;
  2. $V \subseteq W \times Prop$ (or $V : Prop \to \mathcal{P}(W)$).

# Models

- Veltman model: $\mathcal{M} = \langle W, R, \underline{\{S_w \ : \ w \in W\}}, V \rangle$.
- $w \Vdash p$ if and only if $wVp$, for $p \in Prop$.
- Logical connectives have classical semantics.
- Truth of a formula $F \rhd G$ ("$F$ interprets $G$") in a world $w \in \mathcal{M}$:

$$w \Vdash F \rhd G \quad :\Leftrightarrow \quad \forall x \in R(w) : \quad x \Vdash F \Rightarrow \exists y \in S_w(x) : \ y \Vdash G.$$

- Modal soundness and completeness:

$$\mathbf{IL} \vdash F \quad \Leftrightarrow \quad \forall \mathcal{F} : \mathcal{F} \vDash F.$$

# Extensions and frame conditions

- Some extensions of **IL**:

| | |
|---|---|
| $\mathbf{IL}P$ | $\mathbf{IL} + A \rhd B \to \Box(A \rhd B)$ |
| $\mathbf{IL}M$ | $\mathbf{IL} + A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C$ |
| $\mathbf{IL}M_0$ | $\mathbf{IL} + A \rhd B \to \Diamond A \wedge \Box C \rhd B \wedge \Box C$ |
| $\mathbf{IL}W$ | $\mathbf{IL} + A \rhd B \to A \rhd B \wedge \Box\neg A$ |
| $\mathbf{IL}W^*$ | $\mathbf{IL} + A \rhd B \to B \wedge \Box C \rhd B \wedge \Box C \wedge \Box\neg A$ |

- These logics are complete w.r.t. certain classes of frames:

| | |
|---|---|
| $(P)$ | $wRuS_x v \Rightarrow wRv$; |
| $(M)$ | $wRuS_w v \Rightarrow R(v) \subseteq R(u)$; |
| $(M_0)$ | $wRuRxS_w v \Rightarrow R(v) \subseteq R(u)$; |
| $(W)$ | $S_w \circ R$ is reverse well-founded for each $w$; |
| $(W^*)$ | $(M_0)$ and $(W)$. |

- **IL**W-frame is **IL**-frame that satisfies $(W)$ etc.
- Current best guess for $IL(All)$ is (a possibly modally incomplete logic) $\mathbf{IL}W$ + $(R_n)_n$ + $(R^n)_n$. (Joost Joosten)

# Complexity

- **IL** conservatively extends **GL** ("provability logic"); **GL** is in $\mathtt{PSPACE}$.
- Closed fragment of **IL** is $\mathtt{PSPACE}$-hard (Bou, Joosten).
- FMP for **IL**: if $x \Vdash F$, then there is finite $\mathcal{M}$ and $x' \in \mathcal{M}$ s.t. $x' \Vdash F$.
- Standard approach: to check if $\vdash F$, we can (soundness, completeness, FMP) check if there is a finite model of $\neg F$.
- So, to prove $\mathbf{IL} \in \mathtt{PSPACE}$, it suffices to construct a $\mathtt{PSPACE}$ algorithm for checking satisfiability.

# Complexity (satisfiability)

- Let $\Gamma$ be an adequate set for $A \in \mathcal{L}$: set of subformulas, closed under certain operations (in fact, we use four different adequate set).
- $|\Gamma|$ is polynomial in $|A|$.
- Our algorithm builds models world-by-world (nondetereministically or with backtracking).
- There are functions named (1), (2) and (3).
- (1) only calls (2), which only calls (3), which only calls (1).

# Function (1)

- (1) takes $\Delta \subseteq \Gamma$ and checks whether there is a rooted Veltman model of $\Delta$ ($W = \{w\} \cup R(w)$, $w \Vdash \Delta$)
- The starting call will be with $\Delta = \{A\}$.
- (1) looks at all the maximal Boolean consistent $\Delta' \supseteq \Delta$, and returns a positive result if at least one extension is satsfiable.
- Lemma: (1) returns a positive result if and only if $\Delta$ is satisfiable.

# Function (2)

- (2) takes a maximal Boolean consistent $\Delta \subseteq \Gamma$ and checks whether there is a rooted Veltman model of $\Delta$.

$$\Delta^+ := \{E \rhd G \in \Gamma : E \rhd G \in \Delta\}$$
$$\Delta^- := \{E \rhd G \in \Gamma : \neg(E \rhd G) \in \Delta\}$$

- (2) returns a positive answer if the sets $\{\neg(C \rhd D)\} \cup \Delta^+$ are satisfiable for all $\neg(C \rhd D) \in \Delta^-$.
- Lemma: (2) returns a positive result if and only if $\Delta$ is satisfiable. (Proof by merging roots)

# Function (3)

- (3) takes a Boolean consistent $\Delta \subseteq \Gamma$ consisting of one negated $\rhd$-formula $\neg(C \rhd D)$ and a set of positive $\rhd$-formulas $\Delta^+$, and checks whether there is a rooted Veltman model of $\Delta$.
- We say that $(N, P)$ is a $(\neg(C \rhd D), \Delta)$-pair if:
  1. $N, P \subseteq \Gamma$;
  2. $D \in N$;
  3. $\bot \notin P$;
  4. $E \rhd G \in \Delta^+ \;\Rightarrow\; E \in N$ or $G \in P$.
- (3) returns a positive answer if there is a $(\neg(C \rhd D), \Delta)$-pair $(N, P)$ such that the following holds:
  1. $\{\neg B, B \rhd \bot \mid B \in N\} \cup \{C, C \rhd \bot\}$ is satisfiable;
  2. $\{\neg B, B \rhd \bot \mid B \in N\} \cup \{G, \theta \rhd \bot\}$ is satisfiable for all $G$ in $P$.
- Lemma: (3) returns a positive result if and only if $\Delta$ is satisfiable. (Proof by joining the models, adding a new root $w$, and adding the $S_w$ where needed – or even make it total).

# Wrapping up

- Note that (1) can be calculated in terms of (2) etc.
- Each (1)-(2)-(3) chain adds a new $\Box\neg B$ formula for some $B \in \Gamma$: so the procedure terminates.
- Algorithm works locally correct: each function does what it is supposed to do assuming the next one does. Correctness follows by induction (starting with leaf nodes in the execution tree).
- **IL** was known to be $\mathtt{PSPACE}$-hard (conservatively extends **GL**; also $\mathbf{IL}_0$). Thus, **IL** is $\mathtt{PSPACE}$-complete.

# Currnt/future work

- I believe to have shown that $\mathbf{IL}W$, $\mathbf{IL}M$ and $\mathbf{IL}\mathrm{P}$ are also PSPACE-complete.
- After checking/finalizing those proofs, the next natural step would be to prove complexity results regarding $\mathbf{IL}M_0$ and $\mathbf{IL}W^*$. These logics were only recently proven to have FMP (Luka Mikec, Tin Perkov, Mladen Vuković), but with respect to a much more complex semantics.

# Papers


- L. Mikec, F. Pakhomov, M. Vuković. Complexity of interpretability logics **IL**. Logic Journal of the IGPL, 2018.
- L. Mikec, T. Perkov, M. Vuković. Decidability of interpretability logics **IL**$M_0$ and **IL**$W^*$. Logic Journal of the IGPL, Volume 25, Issue 5, 1 October 2017, Pages 758–772,