LIVING WITHOUT BETH AND CRAIG

Based on joint work with Alessandro Artale, Jędrzej Kołodziejski, Andrea Mazzullo, Ana Ozaki, Frank Wolter

22/04/2024

Jean Christoph Jung
TU Dortmund University

AMSTERDAM, CIBD'24

MOTIVATION

- **Craig interpolants** and **explicit definitions** have many applications in knowledge representation and reasoning
- many logics have **Craig Interpolation Property / Projective Beth Definability Property**
  - propositional logic, basic modal logic, first-order logic …
  - **consequence:** there „always“ is an interpolant if you need one
  - **similarly:** there „always“ is an explicit definition if possible
- many logics **do not** enjoy Beth's / Craig's properties
  - FO^2, guarded-fragment, LTL, extensions of ML and description logics …

What to do? How to live without Beth and Craig?

CRAIG INTERPOLATION

θ is an **interpolant** for φ, φ' if $\varphi \models \theta \models \varphi'$ and $\mathrm{sig}(\theta) \subseteq \mathrm{sig}(\varphi) \cap \mathrm{sig}(\varphi')$.

Logic $\mathscr{L}$ has **Craig interpolation property** if for every $\varphi, \varphi' \in \mathscr{L}$:

$$\varphi \models \varphi' \quad \text{iff} \quad \text{there is an } \mathscr{L}\text{-interpolant for } \varphi, \varphi'.$$

Example propositional logic has CIP

Let $\varphi \models \varphi'$ and let $p_1, \ldots, p_n$ be the propositions in φ, but not in φ'

Then $\exists p_1 \ldots \exists p_n \,.\, \varphi$ is an interpolant for φ, φ'

where $\exists x \,.\, \psi := \psi[x/0] \vee \psi[x/1]$

- interpolant is **uniform** (does not depend on φ')
- construction is **exponential** (not known whether this is necessary)
- alternative: extract from **resolution proof** for inconsistency of $\varphi, \neg\varphi'$

FIRST ORDER LOGIC

FO has CIP

Craig constructed an interpolant for $\varphi \models \varphi'$ from a proof for $\varphi \to \varphi'$.

Uniform interpolants do **not** always exist:

$$\varphi = \forall x \Big(\quad A(x) \to B(x) \quad \wedge \quad (B(x) \to \exists y (R(x,y) \wedge B(y)) \Big)$$

Then: $A(x)$ implies infinite R-path, not expressible in FO (over A, R)

CIP **not** preserved for sublogics or for subclasses of structures:

- Guarded and Two-Variable Fragments, some modal logics
- FO over words (=LTL)

} lack CIP

BETH DEFINABILITY

θ is **explicit Σ-definition** for R under φ if $\varphi \models \forall \mathbf{x} R(\mathbf{x}) \leftrightarrow \theta(\mathbf{x})$ and $\mathrm{sig}(\theta) \subseteq \Sigma$.

Logic $\mathscr{L}$ has **projective Beth definability property** if for every $\varphi \in \mathscr{L}$, R, Σ:

R "determined" by φ and Σ iff there is explicit $\mathscr{L}(\Sigma)$-definition for R under φ

Well-known:

- Explicit Definability reduces to Interpolant Existence
- CIP implies PBDP
- Explicit Definability reduces to validity in logics with PBDP
- FO, ML, PL have PBDP, **but** LTL, FO^2, GF not

APPLICATION 1 — CONCEPT LEARNING

Knowledge base is pair $(\mathscr{O}, \mathscr{D})$ consisting of database $\mathscr{D}$ and ontology $\mathscr{O}$.

Let $P, N \subseteq \mathrm{dom}(\mathscr{D})$ be sets of positive and negative **examples**

We say that $\varphi(x)$ **fits** P, N **over** $(\mathscr{O}, \mathscr{D})$ if:

- $\mathscr{O} \cup \mathscr{D} \models \varphi(a)$ for all $a \in P$ and
- $\mathscr{O} \cup \mathscr{D} \models \neg\varphi(a)$ for all $a \in N$.

Example

$$\mathscr{D} = \{\text{teaches}(alice, logic), \text{student}(bob)\}$$

$$\mathscr{O} = \{\forall x \,.\, \text{student}(x) \rightarrow \neg \exists y \,.\, \text{teaches}(x, y)\}$$

Then: $\varphi(x) = \text{student}(x)$ **fits** $P = \{bob\}$ and $N = \{alice\}$ over $(\mathscr{O}, \mathscr{D})$

Fitting formula can be thought of as a **classifier** $\Rightarrow$ machine learning

APPLICATION 1 — CONCEPT LEARNING

Knowledge base is pair $(\mathcal{O}, \mathcal{D})$ consisting of database $\mathcal{D}$ and ontology $\mathcal{O}$.

Let $P, N \subseteq \mathrm{dom}(\mathcal{D})$ be sets of positive and negative **examples**
We say that $\varphi(x)$ **fits** P, N **over** $(\mathcal{O}, \mathcal{D})$ if:

- $\mathcal{O} \cup \mathcal{D} \models \varphi(a)$ for all $a \in P$ and
- $\mathcal{O} \cup \mathcal{D} \models \neg\varphi(a)$ for all $a \in N$.

Fitting existence asks for the existence of a fitting formula given $(\mathcal{O}, \mathcal{D}), P, N$

Lots of interest in description logic knowledge bases:

- theory [Baader, Funk, Hitzler, J, Lehmann, Lutz, Wolter, ...]
- several systems
  - DL-Learner [Hitzler & Lehmann MLJ 2010]
  - DL-Foil [Fanizzi, d'Amato, Esposito ILP 2008]
  - SPELL [ten Cate, Funk, J, Lutz IJCAI 2023]

APPLICATION 1 — CONCEPT LEARNING

In important cases: [Artale, J, Mazzullo, Ozaki, Wolter ToCL 2023]

- fitting existence can be reduced to interpolant existence
  in a way that fitting formulas correspond directly to interpolants

- interpolant existence can be reduced to fitting existence in the same way

$\Rightarrow$ interpolant existence and fitting existence are the same problem

Reduction needs **constants** in the language

Example Modal Logic with constants does **not** enjoy CIP

Consider: $\varphi = a \wedge \Diamond a$ and $\varphi' = b \rightarrow \Diamond b$ for constants a, b

Clearly: $\varphi \models \varphi'$, but in ML with only $\Diamond$ we cannot express a self-loop

APPLICATION 2 — DEFINITIONS

Ontology design

Ontology $\mathcal{O}$ describes domain knowledge as a set of logical formulas

Interesting from user perspective:

> *Does $\mathcal{O}$ define the meaning of A? If so, can we extract a definition?*

Referring Expressions

In many applications in data management constants are not meaningful to the user, e.g., identifier 0x1234

Interesting from user perspective:

> *Is there a definition for the constant 0x1234 in the given knowledge base? If so, can you provide it?*

Note: requires constants in the language

OVERVIEW

Interpolants and explicit definitions important in KR

If considered logics have CIP / PBDP 🙂 But if not ☹

Ways out

allow interpolants/definitions in larger logic

- works e.g., GF $\rightarrow$ GNFO
- not for FO^2: every extension with CIP is undecidable [Comer & ten Cate, 2024]
- same for hybrid logic [ten Cate, 2005]

Investigate existence problem

Given φ, φ', is there an interpolant?

$\Rightarrow$ **non-uniform** approach

1. **Decidable fragments of FO**
   modal and description logics
   guarded and two-variable fragment
2. **LTL**

STARTING POINT

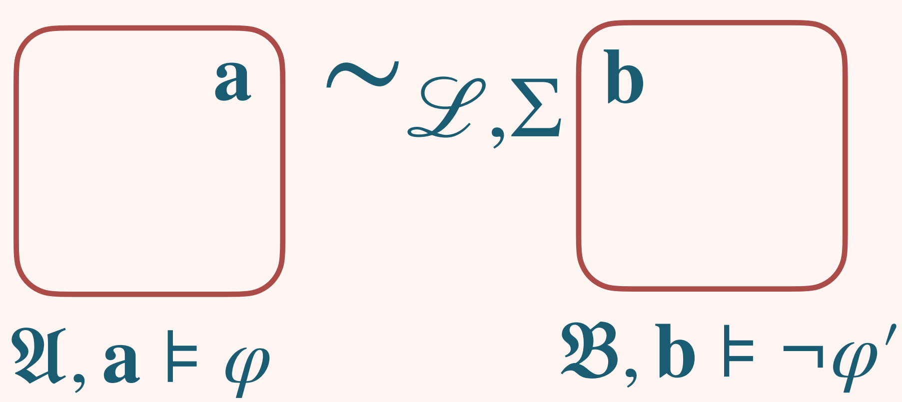
General Characterization a la Robinson [1956]

Let $\mathscr{L} \subseteq$ FO and $\varphi(\mathbf{x}), \varphi'(\mathbf{x}) \in \mathscr{L}$ with $\Sigma = \mathrm{sig}(\varphi) \cap \mathrm{sig}(\varphi')$. TFAE:

1. There is no interpolant for $\varphi(\mathbf{x}), \varphi'(\mathbf{x})$
2. There are models $\mathfrak{A}, \mathbf{a} \models \varphi$, $\mathfrak{B}, \mathbf{b} \models \neg\varphi'$ such that for all $\mathscr{L}(\Sigma)$-formulae ψ:

$$\mathfrak{A}, \mathbf{a} \models \psi \quad \text{iff} \quad \mathfrak{B}, \mathbf{b} \models \psi$$

3. $\varphi, \neg\varphi'$ are **jointly $\mathscr{L}, \Sigma$-consistent**



where $\sim_{\mathscr{L},\Sigma}$ is indistinguishability in the **infinite game** for $\mathscr{L}$, Σ, e.g.:

- bisimulation for modal logic
- guarded bisimulation for GF
- 2-pebble games for FO^2

BISIMULATION

2-Player game on structures: Spoiler vs Duplicator

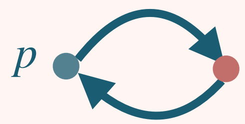
- starting position (u_0, v_0)
- in round i and position (u_i, v_i):

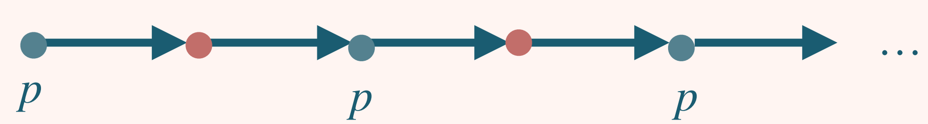
  (back) Spoiler chooses $u_i \to u_{i+1}$ and Duplicator replies with $v_i \to v_{i+1}$, or

  (forth) Spoiler chooses $v_i \to v_{i+1}$ and Duplicator replies with $u_i \to u_{i+1}$
- Spoiler wins if not:

  (atom) u_i and v_i agree on all Σ-propositions
- Duplicator has **winning strategy** from (u_0, v_0) if they can force infinite game

$$\Rightarrow \qquad u_0 \sim_{ML,\Sigma} v_0$$





Lemma $u \sim_{ML,\Sigma} v$ implies $\mathfrak{A}, u \models \psi$ iff $\mathfrak{A}, v \models \psi$, for all modal Σ-formulae

Converse direction holds over ω-saturated structures.

EXAMPLE MODAL LOGIC

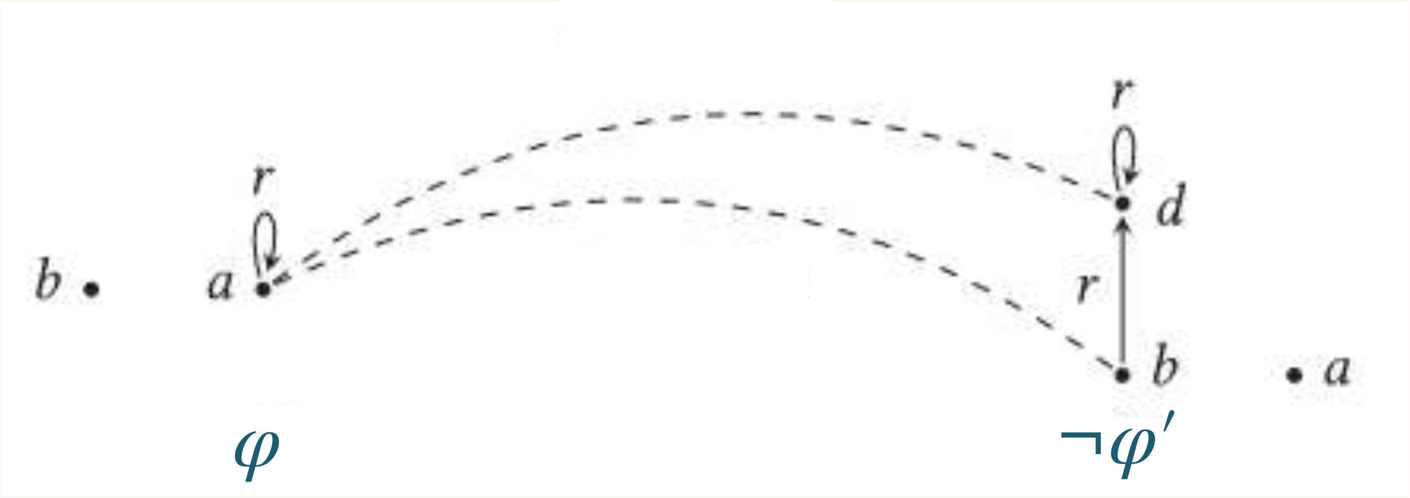
General Characterization

Let $\mathscr{L} \subseteq$ FO and $\varphi(\mathbf{x}), \varphi'(\mathbf{x}) \in \mathscr{L}$ with $\Sigma = \text{sig}(\varphi) \cap \text{sig}(\varphi')$. TFAE:

1. There is no interpolant for $\varphi(\mathbf{x}), \varphi'(\mathbf{x})$
2. $\varphi, \neg\varphi'$ are jointly $\mathscr{L}, \Sigma$-consistent

Example

$\varphi = a \wedge \Diamond a$ and $\varphi' = b \rightarrow \Diamond b$ where $\varphi \models \varphi'$ and a, b are constants



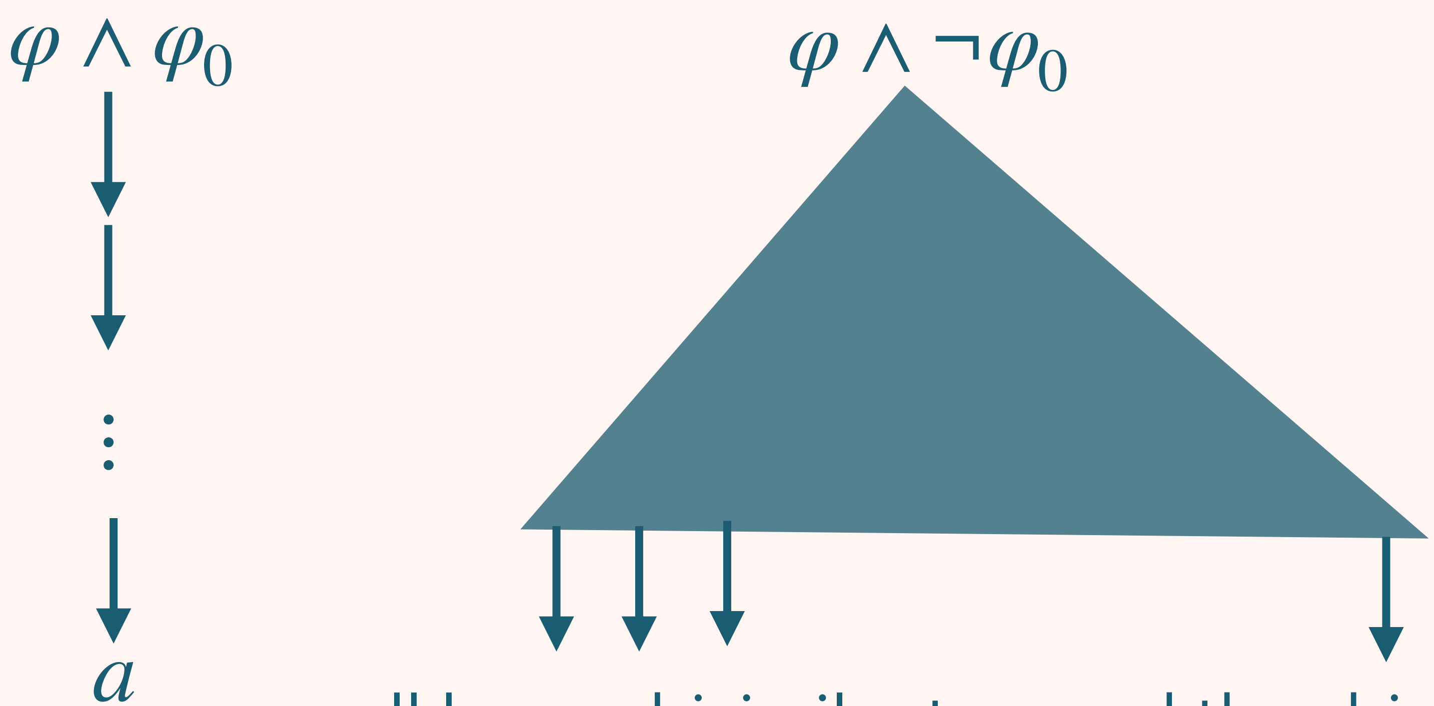
$\Rightarrow$ **no interpolant**

JOINT CONSISTENCY IS HARD

no interpolant for $\varphi \wedge \varphi_0 \models \varphi \rightarrow \varphi_0$ iff $\varphi \wedge \varphi_0, \varphi \wedge \neg\varphi_0$ jointly consistent

Consider $\varphi_0 = \Diamond^n a \wedge \Box^n a$

$\varphi = \big((\neg\varphi_0 \wedge \Diamond^n.\,\mathrm{true}) \rightarrow p\big) \wedge$ „p enforces binary tree of depth n“



all leaves bisimilar to a and thus bisimilar!

this idea ⇒

interpolant existence is coNExpTime-hard

DECIDING JOINT CONSISTENCY

Observation 1 witness for joint consistency can assumed to be tree-like

Observation 2 witness for joint consistency has bounded depth

Observation 3 witness for joint consistency has bounded outdegree

„Guess and check“ algorithm

- Guess witness for consistency of exponential size
- Check that the required bisimulation exists

Consequence Interpolant existence in ML + constants coNExpTime-complete (same for explicit definition existence)

[Artale, J, Mazzullo, Ozaki, Wolter ToCL 2023]

EXTENSIONS

coNExpTime-completeness extends to:

- converse modality
- multimodal logic $\Diamond_i$
- inclusion constraints between accessibility relations $R_i \subseteq R_j$
  (hardness already holds only under inclusion constraints w/o constants)

Important from KR / description logic perspective:

interpolants under ontologies as background knowledge

θ is interpolant of φ, φ' **under** φ_0 if

θ is interpolant of φ, φ' over structures that **globally** satisfy φ_0

$\Rightarrow \mathfrak{A}, \mathfrak{B}$ in joint consistency have to globally satsify φ_0

JOINT CONSISTENCY UNDER ONTOLOGIES

Standard procedures for **global satisfiability** in ML ist **type elimination**

type = syntactic description of single element

⇒ **fail** to capture any bisimilarities

Alternative: **mosaic** = (T_1, T_2) for sets T_1, T_2 of types,
„are realizable in joint consistent models of $\varphi, \neg\varphi'$ under φ_0"

Mosaic Elimination Procedure

- Start with the set of all mosaics
- remove mosaics not satisfying (atom)
- remove mosaic (T_1, T_2) if (back) or (forth) not satisfiable in current mosaics

2Exp many mosaics ⇒ 2ExpTime upper bound (for all extensions + tight)

GF AND FO^2

$$\mathrm{ML} \subseteq \mathrm{GF}^2 \subseteq \mathrm{GF} \subseteq \mathrm{FO}$$
$$\mathrm{GF}^2 \subseteq \mathrm{FO}^2 \subseteq \mathrm{FO}$$

ML, GF^2 enjoys CIP/PBDP,

but GF, FO^2 do not

uniform interpolants [J, Martel, Lutz, Schneider, Wolter ICALP 2017]

- GF^2 does not have uniform interpolants,
  but recognition is decidable (2ExpTime-c)
- GF, FO^2: even recognition is undecidable

interpolant existence [J, Wolter 2021]

- **decidable** in GF: 2/3 ExpTime-complete
- **decidable** in FO^2: 2ExpTime...2NExpTime
- **undecidable** in FO^2 with two equivalence relations
  [Wolter & Zakharyaschev 2024]

non-trivial extensions

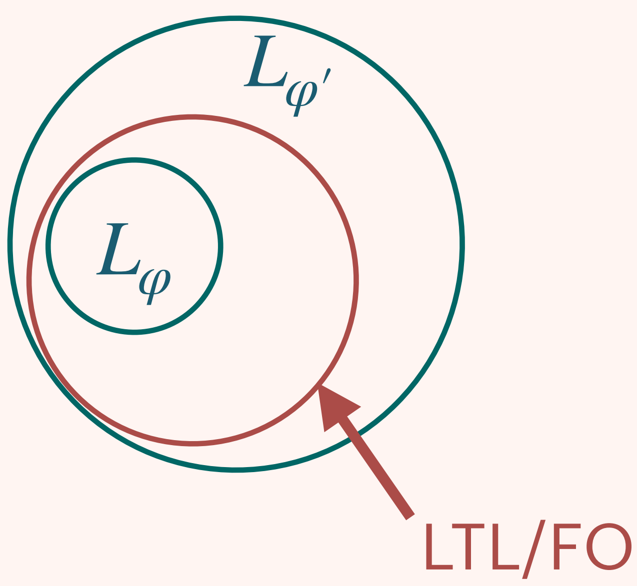
LTL

Lack of CIP/PBDP. Intuitively due to the fact that EVEN is implicitly definable

Interpolant existence Given $\varphi \models \varphi'$, is there an interpolant?

- $\exists p_1 \dots \exists p_n . \varphi$ is uniform interpolant for φ
- LTL not closed under projection $\Rightarrow$ $\exists p_1 \dots \exists p_n . \varphi$ **regular**

$\Rightarrow$ interpolant existence reduces to **separability** question



Separability of regular languages by FO language:

- decidable in ExpTime [Place & Zeitoun 2016]
- separator/interpolant computable
- notoriously open problem over trees

MODAL SEPARABILITY OF μ-FORMULAE

restrict logic instead of signature!

$\mathscr{L}, \mathscr{L}'$**-separability**:

Given $\varphi, \varphi' \in \mathscr{L}$, decide whether there is $\theta \in \mathscr{L}'$ with $\varphi \models \theta \models \varphi'$.

$\Rightarrow$ generalizes $\mathscr{L}'$**-definability** of $\mathscr{L}$-formulae:

φ is $\mathscr{L}'$-definable iff φ, φ are $\mathscr{L}'$-separable

μML, ML-separability [J, Kołodziejski, 2024]

- ML-definability of μML formula is ExpTime-complete [Otto, STACS 1999]
- ML-separability is ExpTime-complete, also over finite/infinite trees
- ML-separability **over words** is PSpace-complete

COMPUTING SEPARATORS

One contribution is computation of separating formula

Strategy

1. φ, φ' ML-separable $\Rightarrow$ separable by formula of depth $\ell \in O(2^{|\varphi|+|\varphi'|})$
2. compute ℓ**-universal consequence** θ of φ:

   uniform interpolant for ML formulas up to depth ℓ

   essentially by reading it off from $\mathscr{A}_\varphi$:

   θ describes behaviour of $\mathscr{A}_\varphi$ up to depth ℓ

Consequence If any, there is a separator of size 2^{2^n}, and this is optimal.

COMPUTING INTERPOLANTS

Might be most interesting from practical perspective, but:

Bad news our computation algorithm inspired by one for ML does not work (and currently we don't know how to fix it :-/)

Conjecture price of elegant characterization via joint realizability is loss of constructability (**compactness!**)

Way forward show bound on the quantifier depth ℓ of the potential interpolant

- bound is trivial for ML with constants / inclusion constraints, but not under ontologies and not for GF/FO^2!
- **brutal way** disjunction over all types of depth ℓ $\Rightarrow$ non-elementary
- **alternative** automata-based construction?

CONCLUSION

Take-home message

- **non-uniform** approach to interpolation / explicit definability
  $\Rightarrow$ decide in each case when an interpolant / explicit definition exists
- Interpolant / definition existence **usually harder than satisfiability** but often decidable

Challenges

- computation problem
- investigate other logics without CIP, e.g. for FO^2 + counting
- other areas?
- investigate separability / definability

THANK YOU VERY MUCH!

QUESTIONS?