Contents

CHAPTER 1

Introduction

1.1 Preliminaries

The language of a first-order formula A is denoted by $\mathrm{L}(A)$ and contains all predicate, constant, function and free variable symbols that occur in A. These are also referred to as the *non-logical symbols* of A.

An occurrence of Φ-term is called *maximal* if it does not occur as subterm of another Φ-term.

$\bar{x}$ denotes $x_1, \ldots, x_n$.

1.2 Craig Interpolation

Theorem 1.1 (Interpolation). *Let Γ and Δ be sets of first-order formulas such that $\Gamma \cup \Delta$ is unsatisfiable. Then there exists a first-order formula I, called interpolant, such that*

1. $\Gamma \models I$
2. $\Delta \models \neg I$
3. $\mathrm{L}(I) \subseteq \mathrm{L}(\Gamma) \cap \mathrm{L}(\Delta)$. $\square$

In the context of interpolation, every non-logical symbol is assigned a color which indicates the its origin(s). A non-logical symbol is said to be Γ *(Δ)-colored* if it only occurs in Γ (Δ) and *grey* in case it occurs in both Γ and Δ.

CHAPTER 2

The Resolution Calculus

2.1 Resolution

Resolution calculus, in the formulation as given here, is a sound and complete calculus for first-order logic with equality. Due to the simplicity of its rules, it is widely used in the area of automated deduction.

Definition 2.1. A *clause* is a finite set of literals. The empty clause will be denoted by $\Box$. A *resolution refutation* of a set of clauses Γ is a derivation of $\Box$ consisting of applications of resolution rules (cf. figure 2.1) starting from clauses in Γ. $\triangle$

Theorem 2.2. *A clause set Γ is unsatisfiable if and only if there is resolution refutation of Γ.*

Proof. See [Rob65]. $\Box$

Clauses will usually be denoted by C or D, literals by l.

$$\textit{Resolution:} \quad \frac{C \vee l \qquad D \vee \neg l'}{(C \vee D)\sigma} \quad \sigma = \mathrm{mgu}(l, l')$$

$$\textit{Factorisation:} \quad \frac{C \vee l \vee l'}{(C \vee l)\sigma} \quad \sigma = \mathrm{mgu}(l, l')$$

$$\textit{Paramodulation:} \quad \frac{C \vee s = t \qquad D[r]}{(C \vee D[t])\sigma} \quad \sigma = \mathrm{mgu}(s, r)$$

Figure 2.1: The rules of resolution calculus

2.2 Resolution and Interpolation

In order to apply resolution to arbitrary first-order formulas, they have to be converted to clauses first. This usually makes use of intermediate normal forms which are defined as follows:

Definition 2.3. A formula is in *Negation Normal Form (NNF)* if negations only occur directly befor of atoms. A formula is in *Conjunctive Normal Form (CNF)* if it is a conjunction of disjunctions of literals. △

In this context, the conjuncts of a CNF-formula are interpreted as clauses. A well-established procedure for the translation to CNF is comprised of the following steps:

1. NNF-Transformation
2. Skolemisation
3. CNF-Transformation

Step 1 can be achieved by solely pushing the negation inwards. As this transformation yields an equivalent formula, it clearly has no effect on the interpolants. Step 2 and 3 on the other hand do not produce equivalent formulas since they introduce new symbols. In this section, we will show that they nonetheless do preserve the set of interpolants. This fact is vital for the use of resolution-based methods for interpolant computation of arbitrary formulas.

2.2.1 Interpolation and Skolemisation

Skolemisation is a procedure for replacing existential quantifiers with Skolem terms:

Definition 2.4. Let $V_{\exists x}$ be the set of universally bound variables in the scope of the occurrence of $\exists x$ in a formula. The skolemisation of a formula A in NNF, denoted by $\mathrm{sk}(A)$, is the result of replacing every occurrence of an existential quantifier $\exists x$ in A by a term $f(y_1, \dots, y_n)$ where f is a new Skolem function symbol and $V_{\exists x} = \{y_1, \dots, y_n\}$. In case $V_{\exists x}$ is empty, the occurrence of $\exists x$ is replaced by a new Skolem constant symbol c.

The skolemisation of a set of formulas Φ is defined to be $\mathrm{sk}(\Phi) = \{\mathrm{sk}(A) \mid A \in \Phi\}$. △

Proposition 2.5. *Let $\Gamma \cup \Delta$ be unsatisfiable. Then I is an interpolant for $\Gamma \cup \Delta$ if and only if it is an interpolant for $\mathrm{sk}(\Gamma) \cup \mathrm{sk}(\Delta)$.*

Proof. Since $\mathrm{sk}(\cdot)$ adds fresh symbols to both Γ and Δ individually, none of them are containd in $\mathrm{L}(\mathrm{sk}(\Gamma)) \cap \mathrm{L}(\mathrm{sk}(\Delta))$. Therefore condition 3 of theorem 1.1 is satisfied in both directions.

As for any set of formulas Φ, each model of Φ can be extended to a model of $\mathrm{sk}(\Phi)$ and every model of $\mathrm{sk}(\Phi)$ is a witness for the satisfiability of Φ, $\Phi \models I$ iff $\mathrm{sk}(\Phi) \models I$. Hence conditions 1 and 2 of theorem 1.1 remain satisfied for I as well. □

2.2.2 Interpolation and structure-preserving Normal Form Transformation

A common method for transforming a skolemised formula A into CNF while preserving their structure is defined as follows:

Definition 2.6. For every occurrence of a subformula B of A, introduce a new atom L_B which acts as a label for the subformula. For each of them, create a defining clause D_B:

does it suffice to not treat universal quantifiers specifically here? (subterms have free variables; possibly need to mention to just pull universal quantifiers outwards to get prenex form and drop quantifiers)

If B is atomic:

$$D_B \equiv (\neg B \vee L_B) \wedge (B \vee \neg L_B)$$

If B is $\neg G$:

$$D_B \equiv (L_B \vee L_G) \wedge (\neg L_B \vee \neg L_G)$$

If B is $G \wedge H$:

$$D_B \equiv (\neg L_B \vee L_G) \wedge (\neg L_B \vee L_H) \wedge (L_B \vee \neg L_G \vee \neg L_H)$$

If B is $G \vee H$:

$$D_B \equiv (L_B \vee \neg L_G) \wedge (L_B \vee \neg L_H) \wedge (\neg L_B \vee L_G \vee L_H)$$

If B is $G \supset H$:

$$D_B \equiv (L_B \vee L_G) \wedge (L_B \vee \neg L_H) \wedge (\neg L_B \vee \neg L_G \vee L_H)$$

If B is $\forall x G$:

$$D_B \equiv \forall x(\neg L_B \vee L_G) \wedge \forall x(L_B \vee \neg L_G)$$

Let $\delta(A)$ be defined as $\bigwedge_{B \in \Sigma(A)} D_B \wedge L_A$, where $\Sigma(A)$ denotes the set of occurrences of subformulas of A. $\triangle$

Proposition 2.7. *Let A be a formula. Then $\mathrm{sk}(A)$ is unsatisfiable if and only if $\delta(\mathrm{sk}(A))$ is unsatisfiable.*

Proposition 2.8. *Let $\mathrm{sk}(\Gamma) \cup \mathrm{sk}(\Delta)$ be unsatisfiable. Then I is an interpolant for $\mathrm{sk}(\Gamma) \cup \mathrm{sk}(\Delta)$ if and only if I is an interpolant for $\delta(\mathrm{sk}(\Gamma)) \cup \delta(\mathrm{sk}(\Delta))$.*

Proof. As δ introduces fresh symbols for each $\mathrm{sk}(\Gamma)$ and $\mathrm{sk}(\Delta)$, they must not occur in any interpolant of $\mathrm{sk}(\Gamma)$ and $\mathrm{sk}(\Delta)$. This establishes condition 3 of theorem 1.1 in both directions.

Using proposition 2.7, condition 1 and 2 of theorem 1.1 are immediate. $\square$

CHAPTER 3

Proof by Reduction

A common theme of proofs in theoretical computer science is to instead of proving the result from first principles to reduce the problem to another one, which then is easier to solve. In this instance, we are able to give a reduction for finding interpolants for first-order logic *with* equality to first-order logic *without* equality, where it is simpler to give an appropriate algorithm.

The general layout of this approach is the following: From two sets Γ and Δ, where $\Gamma \cup \Delta$ is unsatisfiable, we compute Γ' and Δ' which do not make use of equality but simulate equality it via axioms. In the process of this transformation, also function symbols are replaced by predicate symbols with appropriate axioms to make sure that their behaviour is compatible to the one of functions. Now an interpolant of Γ' and Δ' can be derived using an algorithm that is only capable of handling predicate symbols, as all other non-logical symbols have been removed. Since the additional axioms ensure that the newly added predicate symbols mimic equality and functions respectively, we will see that the occurrences of these predicates in the interpolant can be translated back to occurrences of equality and function symbols in first-order logic with equality in the language of Γ and Δ, thereby yielding the originally desired interpolant.

$\Gamma \models I$
$\Delta \models \neg I$
$L(I) \subseteq L(\Gamma) \cap L(\Delta)$
to show:
- can find I' such that I' is interpolant between Γ' and Δ'.
- $\Phi' \models I'$ implies $\Phi \models I$ ($\Phi', \neg I' \models \bot$ implies $\Phi, \neg I \models \bot$)

3.1 Reduction to first-order logic without equality

As we shall see in this section, first-order formulas with equality can be transformed into first-order formulas without equality in a way that is satisfiability-preserving, which is sufficient for our purposes.

First, we define the axioms which allow for simulation of equality and functions in first order logic without equality and function symbols:

Definition 3.1. For first-order formulas A and a fresh predicate symbol E, we define:

$$\mathrm{FAx}(A) \stackrel{\text{def}}{=} \bigwedge_{f \in \mathrm{FS}(A)} \forall \bar{x} \exists y (F_f(\bar{x}, y) \wedge (\forall z (F_f(\bar{x}, z) \supset E(z, y))))$$

$$\begin{aligned}\mathrm{EAx}(A) \stackrel{\text{def}}{=} {} & \forall x E(x, x) \wedge \\ & \bigwedge_{\substack{P \in \mathrm{PS}(A) \cup \{E\} \cup \\ \{F_f \mid f \in \mathrm{FS}(A)\}}} \forall x_1 \ldots \forall x_{\mathrm{ar}(P)} \forall y_1 \ldots \forall y_{\mathrm{ar}(P)} \\ & \qquad ((E(x_1, y_1) \wedge \ldots \wedge E(x_{\mathrm{ar}(P)}, y_{\mathrm{ar}(P)})) \supset \\ & \qquad (P(x_1, \ldots, x_{\mathrm{ar}(P)}) \Leftrightarrow P(y_1, \ldots, y_{\mathrm{ar}(P)})))\end{aligned}$$

For sets of first-order formulas Φ and $h \in \{\mathrm{FAx}, \mathrm{EAx}\}$, $h(\Phi) \stackrel{\text{def}}{=} \bigcup_{A \in \Phi} h(A)$. △

Wie EAx schöner formulieren? Auch: besserer Name?

Definition 3.2. Let A be a first-order formula. Then $\mathrm{T}(A)$ is the result of applying the following algorithm to A:

1. Replace every occurrence of $s = t$ in A by $E(s, t)$
2. As long as there is an occurrence of a function symbol f in A:
   Let B be the atom in which f occurs.
   Then B is of the form $P(s_1, \ldots, s_{j-1}, f(\bar{t}), s_{j+1}, \ldots s_m)$.
   Replace B in A by $\exists y (F_f(\bar{t}, y) \wedge P(s_1, \ldots, s_{j-1}, y, s_{j+1}, \ldots s_m))$ for a variable y which does not occur free in B.

For sets of first-order formulas Φ, $\mathrm{T}(\Phi) \stackrel{\text{def}}{=} \bigcup_{A \in \Phi} \mathrm{T}(A)$. △

replace by $P(s_1, \ldots, s_m)$ where $s_j = f(\bar{t})$ for some j? (also in proof below

Definition 3.3. For a first-order formula A, let $\mathrm{T}_{\mathrm{Ax}}(A) = \mathrm{FAx}(A) \wedge \mathrm{EAx}(A) \wedge \mathrm{T}(A)$ and for a set of first-order formulas Φ, let $\mathrm{T}_{\mathrm{Ax}}(\Phi) = \mathrm{FAx}(\Phi) \cup \mathrm{EAx}(\Phi) \cup \mathrm{T}(\Phi)$. △

Note that $\mathrm{FAx}(A)$, $\mathrm{EAx}(A)$ and $\mathrm{T}(A)$ contain neither the equality predicate nor function symbols. Hence they translate formulas A in the language $\mathrm{L}(A)$ to formulas in the language $\mathrm{L}(A) \setminus (\{=\} \cup \mathrm{FS}(A))$.

Proposition 3.4. *A first-order formula A is satisfiable if and only if $\mathrm{T}_{\mathrm{Ax}}(A)$ is satisfiable.*

// TODO: go through proof another time

Proof. Suppose A is satisfiable. Let M be a model of A. We show that $\mathrm{T}_{\mathrm{Ax}}(A)$ is satisfiable by extending M to satisfy this formula.

First, let $M \models E(s, t)$ if and only if $M \models s = t$. By reflexivity of equality, it follows that $M \models \forall x E(x, x)$ and since equality of all arguments implies the same truth value for predicates, we get that M is a model of $\mathrm{EAx}(A)$.

Second, let $M \models F_f(\bar{x}, y)$ if and only if $M \models f(\bar{x}) = y$ for all $f \in \mathrm{FS}(A)$. Since M is a model of A, it maps f to a function, which returns a unique result for every combination of parameters. Hence M is also a model of $\mathrm{FAx}(A)$.

By the above definition of E in M, step 1 of the algorithm in definition 3.2 yields a formula that is satisfied by M. For step 2, suppose $P(s_1, \ldots, s_{j-1}, f(\bar{t}), s_{j+1}, \ldots s_m)$

does (not) hold under M. Let y such that $M \models f(\bar{t}) = y$. By our definition of F under M, $M \models F(\bar{t}, y)$ with this unique y. Hence $\exists y(F(\bar{t}, y) \wedge P(s_1, \dots, s_{j-1}, y, s_{j+1}, \dots s_m))$ does (not) hold under M.

For the other direction, suppose $\mathrm{T}_{\mathrm{Ax}}(A)$ is satisfiable. We again extend a model M of this formula to a model of A.

First, let $M \models s = t$ if and only if $M \models E(s,t)$. As M is a model of $\mathrm{EAx}(A)$, E and consequently $=$ are reflexive, symmetric and transitive.

i do have to show this, right? then show in more detail

Second, let $M \models f(\bar{x}) = y$ if and only if $M \models F(\bar{x}, y)$. As by assumption M is a model of $\mathrm{FAx}(A)$, we know that for every $\bar{x}$, some y exists and is uniquely defined. Hence f in M refers to a well-defined function.

To show that $M \models A$, consider that the predicates E and $=$ coincide in M. Furthermore, let B be an occurrence of $\exists y(F_f(\bar{t}, y) \wedge P(s_1, \dots, s_{j-1}, y, s_{j+1}, \dots s_m))$ in $\mathrm{T}(A)$ which was introduced by T. First suppose that B holds in M. Then there is a y such that $F_f(\bar{t}, y)$ and $P(s_1, \dots, s_{j-1}, y, s_{j+1}, \dots s_m)$ hold in M. By our definition of f in M, $M \models f(\bar{t}) = y$, hence also $P(s_1, \dots, s_{j-1}, f(\bar{t}), s_{j+1}, \dots s_m)$. On the other hand, suppose that B does not hold in M. Then no y exists such that $F_f(\bar{t}, y)$ and $P(s_1, \dots, s_{j-1}, y, s_{j+1}, \dots s_m)$. Hence by our definition of f, $P(s_1, \dots, s_{j-1}, f(\bar{t}), s_{j+1}, \dots s_m)$ does not hold as well. $\square$

Corollary 3.5. *A set of first-order formulas* Φ *is satisfiable if and only if* $\mathrm{T}_{\mathrm{Ax}}(\Phi)$ *is satisfiable.*

Proof. Suppose Φ is satisfiable. Then there is a model M which satisfies every formula A in Φ. Hence $M \models \bigwedge_{A \in \Phi} A$, and by proposition 3.4, $M \models \mathrm{T}_{\mathrm{Ax}}(\bigwedge_{A \in \Phi} A)$

TODO: this does not work like that, possibly show for EAX, FAX and Trans extra and combine then $\square$

3.2 Compuation of interpolants in first-order logic without equality and function symbols

3.3 Hence

Proof of Theorem 1.1 (Interpolation). By proposition 3.4, as $\Gamma \cup \Delta$ is unsatisfiable, so is $\mathrm{T}_{\mathrm{Ax}}(\Gamma) \cup \mathrm{T}_{\mathrm{Ax}}(\Delta)$ $\square$

CHAPTER 4

Proofs

4.1 WT: Interpolation extraction in one pass

easy for constants, just as in huang but in one pass

terms can grow unpredictably, order cannot be determined during pass

4.2 WT: Interpolation extraction in two passes

4.2.1 huang proof revisited

propositional part

Let $\Gamma \cup \Delta$ be unsatisfiable. Let π be a proof of $\Box$ from $\Gamma \cup \Delta$. Then PI is a function that returns a interpolant w.r.t. the current clause.

Definition 4.1. θ is a *propositional interpolant* with respect to a clause C in a resolution refutation π of $\Gamma \cup \Delta$ if

1. $\Gamma \models \theta \vee C$
2. $\Delta \models \neg\theta \vee C$
3. $\mathrm{PS}(\theta) \subseteq (\mathrm{PS}(\Gamma) \cap \mathrm{PS}(\Delta)) \cup \{\top, \bot\}$. △

The third condition will sometimes be referred to as *language restriction*. It is easy to see that a propositional interpolant with respect to $\Box$ is a propositional interpolant, i.e. it is an interpolant without the language restriction on constant, variable and function symbols.

add this to the definition, i.e. possible define rel prop interpol from prop interpol

We proceed by defining a procedure PI which extracts interpolants from a resolution refutation.

Definition 4.2. PI is defined as follows:

Base case. If $C \in \Gamma$, $\mathrm{PI}(C) = \bot$. If otherwise $C \in \Delta$, $\Delta(C) = \top$.

Resolution. Suppose the clause C is the result of a resolution step. Then it has the following form:

If the clause C is the result of a resolution step of $C_1 : D \vee l$ and $C_2 : E \vee \neg l'$ using a unifier σ such that $l\sigma = l'\sigma$, then $\mathrm{PI}(C)$ is defined as follows:

1. If $\mathrm{PS}(l) \in \mathrm{L}(\Gamma) \setminus \mathrm{L}(\Delta)$:$\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$
2. If $\mathrm{PS}(l) \in \mathrm{L}(\Delta) \setminus \mathrm{L}(\Gamma)$: $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$
3. If $\mathrm{PS}(l) \in \mathrm{L}(\Gamma) \cap \mathrm{L}(\Delta)$: $\mathrm{PI}(C) = [(l \wedge \mathrm{PI}(C_2)) \vee (l' \wedge \mathrm{PI}(C_1))]\sigma$

change to "is Γ-colored?"

Factorisation. If the clause C is the result of a factorisation of $C_1 : l \vee l' \vee D$ using a unifier σ such that $l\sigma = l'\sigma$, then $\mathrm{PI}(C) = \mathrm{PI}(C_1)\sigma$.

Paramodulation. If the clause C is the result of a paramodulation of $C_1 : s = t \vee C$ and $C_2 : D[r]$ using a unifier σ such that $r\sigma = s\sigma$, then $\mathrm{PI}(C)$ is defined according to the following case distinction:

1. If r occurs in a maximal Δ-term $h(r)$ in $D[r]$ and $h(r)$ occurs more than once in $D[r] \vee \mathrm{PI}(D[r])$:
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma \vee (s = t \wedge h(s) \neq h(t))$
2. If r occurs in a maximal Γ-term $h(r)$ in $D[r]$ and $h(r)$ occurs more than once in $D[r] \vee \mathrm{PI}(D[r])$:
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma \wedge (s \neq t \vee h(s) = h(t))$
3. Otherwise:
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma$ △

Proposition 4.3. *Let C be a clause of a resolution refutation. Then $\mathrm{PI}(C)$ is a propositional interpolant with respect to C.*

Proof. Proof by induction on the number of rule applications including the following strenghtenings: $\Gamma \models \mathrm{PI}(C) \vee C_\Gamma$ and $\Delta \models \neg\mathrm{PI}(C) \vee C_\Delta$, where D_Φ denotes the clause D with only the literals which are contained in $\mathrm{L}(\Phi)$. They clearly imply conditions 1 and 2 of definition 4.1.

Base case. Suppose no rules were applied. We distinguish two possible cases:

1. $C \in \Gamma$. Then $\mathrm{PI}(C) = \bot$. Clearly $\Gamma \models \bot \vee C_\Gamma$ as $C_\Gamma = C \in \Gamma$, $\Delta \models \neg\bot \vee C_\Delta$ and $\bot$ satisfies the restriction on the language.
2. $C \in \Delta$. Then $\mathrm{PI}(C) = \top$. Clearly $\Gamma \models \top \vee C_\Gamma$, $\Delta \models \neg\top \vee C_\Delta$ as $C_\Delta = C \in \Delta$ and $\top$ satisfies the restriction on the language.

Suppose the property holds for n rule applications. We show that it holds for $n+1$ applications by considering the last one:

Resolution. Suppose the last rule application is an instance of resolution. Then it is of the form:

$$\frac{C_1 : D \vee l \qquad C_2 : E \vee \neg l'}{C : (D \vee E)\sigma} \quad l\sigma = l'\sigma$$

By the induction hypothesis, we can assume that:

$\Gamma \models \mathrm{PI}(C_1) \vee (D \vee l)_\Gamma$

$\Delta \models \neg\mathrm{PI}(C_1) \vee (D \vee l)_\Delta$

$\Gamma \models \mathrm{PI}(C_2) \vee (E \vee \neg l')_\Gamma$

$\Delta \models \neg\mathrm{PI}(C_2) \vee (E \vee \neg l')_\Delta$

We consider the respective cases from definition 4.2:

1. $\mathrm{PS}(l) \in \mathrm{L}(\Gamma) \setminus \mathrm{L}(\Delta)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$.

   As $\mathrm{PS}(l) \in \mathrm{L}(\Gamma)$, $\Gamma \models (\mathrm{PI}(C_1) \vee D_\Gamma \vee l)\sigma$ as well as $\Gamma \models (\mathrm{PI}(C_2) \vee E_\Gamma \vee \neg l')\sigma$. By a resolution step, we get $\Gamma \models (\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2))\sigma \vee ((D \vee E)\sigma)_\Gamma$.

   Furthermore, as $\mathrm{PS}(l) \notin \mathrm{L}(\mathrm{PI})$, $\Delta \models (\neg\mathrm{PI}(C_1) \vee D_\Delta)\sigma$ as well as $\Delta \models (\neg\mathrm{PI}(C_2) \vee E_\Delta)\sigma$. Hence it certainly holds that $\Delta \models (\neg\mathrm{PI}(C_1) \vee \neg\mathrm{PI}(C_2))\sigma \vee (D \vee E)\sigma_\Delta$.

   The language restriction clearly remains satisfied as no non-logical symbols are added.

2. $\mathrm{PS}(l) \in \mathrm{L}(\Delta) \setminus \mathrm{L}(\Gamma)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$.

   As $\mathrm{PS}(l) \notin \mathrm{L}(\Gamma)$, $\Gamma \models (\mathrm{PI}(C_1) \vee D_\Gamma)\sigma$ as well as $\Gamma \models (\mathrm{PI}(C_2) \vee E_\Gamma)\sigma$. Suppose that in a model M of Γ, $M \not\models D_\Gamma$ and $M \not\models E_\Gamma$. Then $M \models \mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)$. Hence $\Gamma \models (\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2))\sigma \vee ((D \vee E)\sigma)_\Gamma$.

   Furthermore due to $\mathrm{PS}(l) \in \mathrm{L}(\Delta)$, $\Delta \models (\neg\mathrm{PI}(C_1) \vee D_\Delta \vee l)\sigma$ as well as $\Delta \models (\neg\mathrm{PI}(C_2) \vee E_\Delta \vee \neg l')\sigma$. By a resolution step, we get $\Delta \models (\neg\mathrm{PI}(C_1) \vee \neg\mathrm{PI}(C_2))\sigma \vee (D_\Delta \vee E_\Delta)\sigma$ and hence $\Delta \models \neg(\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2))\sigma \vee (D_\Delta \vee E_\Delta)\sigma$.

   The language restriction again remains intact.

3. $\mathrm{PS}(l) \in \mathrm{L}(\Delta) \cap \mathrm{L}(\Gamma)$: Then $\mathrm{PI}(C) = [(l \wedge \mathrm{PI}(C_2)) \vee (\neg l' \wedge \mathrm{PI}(C_1))]\sigma$

   First, we have to show that $\Gamma \models [(l \wedge \mathrm{PI}(C_2)) \vee (l' \wedge \mathrm{PI}(C_1))]\sigma \vee ((D \vee E)\sigma)_\Gamma$. Suppose that in a model M of Γ, $M \not\models D_\Gamma$ and $\Gamma \not\models E$. Otherwise we are done. The induction assumtion hence simplifies to $M \models \mathrm{PI}(C_1) \vee l$ and $M \models \mathrm{PI}(C_2) \vee \neg l'$ respectively. As $l\sigma = l'\sigma$, by a case distinction argument on the truth value of $l\sigma$, we get that either $M \models (l \wedge \mathrm{PI}(C_2))\sigma$ or $M \models (\neg l' \wedge \mathrm{PI}(C_1))\sigma$.

   Second, we show that $\Delta \models ((l \vee \neg\mathrm{PI}(C_1)) \wedge (\neg l' \vee \neg\mathrm{PI}(C_2)))\sigma \vee ((D \vee E)\sigma)_\Delta$. Suppose again that in a model M of Δ, $M \not\models D_\Delta$ and $\Gamma \not\models E_\Delta$. Then the required statement follows from the induction hypothesis.

   The language condition remains satisfied as only the common literal l is added to the interpolant.

Factorisation. Suppose the last rule application is an instance of factorisation. Then it is of the form:

$$\frac{C_1 : l \vee l' \vee D}{C_1 : (l \vee D)\sigma} \quad \sigma = \mathrm{mgu}(l, l')$$

Then the propositional interpolant $\mathrm{PI}(C)$ is defined as $\mathrm{PI}(C_1)$. By the induction hypothesis, we have:

$\Gamma \models \mathrm{PI}(C_1) \vee (l \vee l' \vee D)_\Gamma$

$\Delta \models \mathrm{PI}(C_1) \vee (l \vee l' \vee D)_\Delta$

It is easy to see that then also:

$\Gamma \models (\mathrm{PI}(C_1) \vee (l \vee D)_\Gamma)\sigma$

$\Delta \models (\mathrm{PI}(C_1)\sigma \vee (l \vee D)_\Delta)\sigma$

The restriction on the language trivially remains intract.

Paramodulation. Suppose the last rule application is an instance of paramodulation. Then it is of the form:

$$\frac{C_1 : D \vee s = t \qquad C_2 : E[r]}{C : (D \vee E[t])\sigma} \quad \sigma = \mathrm{mgu}(s, r)$$

By the induction hypothesis, we have:

$\Gamma \models \mathrm{PI}(C_1) \vee (D \vee s = t)_\Gamma$

$\Delta \models \neg\mathrm{PI}(C_1) \vee (D \vee s = t)_\Delta$

$\Gamma \models \mathrm{PI}(C_2) \vee (E[r])_\Gamma$

$\Delta \models \neg\mathrm{PI}(C_2) \vee (E[r])_\Delta$

First, we show that $\mathrm{PI}(C)$ as constructed in case 3 of the definition is a propositional interpolant in any of these cases:

$\mathrm{PI}(C) = (s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))$

Suppose that in a model M of Γ, $M \not\models D\sigma$ and $M \not\models E[t]\sigma$. Otherwise we are done. Furthermore, assume that $M \models (s = t)\sigma$. Then $M \not\models E[r]\sigma$, but then necessarily $M \models \mathrm{PI}(C_2)\sigma$.
On the other hand, suppose $M \models (s \neq t)\sigma$. As also $M \not\models D\sigma$, $M \models \mathrm{PI}(C_1)\sigma$. Consequently, $M \models [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma \vee [(D \vee E)_\Gamma]\sigma$

By an analogous argument, we get $\Delta \models [(s = t \wedge \neg\mathrm{PI}(C_2)) \vee (s \neq t \wedge \neg\mathrm{PI}(C_1))]\sigma \vee [(D \vee E)_\Delta]\sigma$, which implies $\Delta \models [(s \neq t \vee \neg\mathrm{PI}(C_2)) \wedge (s = t \vee \neg\mathrm{PI}(C_1))]\sigma \vee ((D \vee E)_\Delta)\sigma$

The language restriction again remains satisfied as the only predicate, that is added to the interpolant, is $=$.

This concludes the argumentation for case 3.

The interpolant of case 1 differs only by an additional formula added via a disjunction and hence condition 1 of definition 4.1 holds by the above reasoning. As the adjoined formula is a contradiction, its negation is valid which in combination with the above reasoning establishes condition 2. Since no new predicated are added, the language condition remains intact.

The situation in case 2 is somewhat symmetric: As a tautology is added to the interpolant with respect to case 1, condition 1 is satisfied by the above reasoning. For condition 2, consider that the negated interpolant of case 1 implies the negated interpolant of this case. The language condition again remains intact. □

proof that we are allowed to overbind
TODO: define procedure
TODO: proof

overbinding

Algorithm (input: propositional interpolant θ):

1. Let $t_1, \dots, t_n$ be the maximal occurrences of noncommon terms in θ. Order t_i ascendingly by term size.

2. Let θ^* be θ with maximal occurrences of Δ-terms $r_1, \dots, r_k$ replaced by fresh variables $x_1, \dots, x_k$ and maximal occurrences of Γ-terms $s_1, \dots, s_{n-k}$ by fresh variables $x_{k+1}, \dots, x_n$

3. Return $Q_1 x_1, \dots Q_n x_n \theta^*$, where Q_i is $\forall$ if t_i is a Δ-term and $\exists$ otherwise.

Language condition easily established. To prove:
$\Gamma \models Q_1 x_1, \dots Q_n x_n \theta^*$
$\Delta \models \neg Q_1 x_1, \dots Q_n x_n \theta^*$
We know that θ works, just the terms are missing.

4.2.2 final step of huang's proof

Theorem 4.4. *$Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\Box)^*(z_1, \dots, z_n)$ is a craig interpolant (order as in huang).*

Proof. By lemma **??**, $\Gamma \models \forall x_1 \dots \forall x_n \overline{\mathrm{PI}(\Box)}(x_1, \dots, x_n)$.

The terms in $\overline{PI(\Box)}$ are either among the x_i, $1 \leq i \leq n$ or grey terms or Γ-terms. Let t be a maximal Γ-term in $\overline{\mathrm{PI}(\Box)}$. Then it is of the form $f(x_{i_1}, \dots, x_{i_{n_x}}, u_1, \dots, u_{n_u}, v_1, \dots, v_{n_v})$, where f is Γ-colored, the x_j are as before, the u_j are grey terms and the v_j are Γ-terms. Note that the Δ-terms, which are replaced by the $x_{i_1}, \dots, x_{i_{n_x}}$ are of strictly smaller size than t as they are "strict" subterms of t.

basically only need the x_j

In $\mathrm{PI}(\Box)^*$, t will be replaced by some z_j, which is existentially quantified. For this z_j, t is a witness as due to the quantifier ordering, all the $x_{i_1}, \dots, x_{i_{n_x}}$ will be quantified before the existential quantification of z_j. Therefore $\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\Box)^*(z_1, \dots, z_n)$

□

Bibliography

[BJ13] Maria Paola Bonacina and Moa Johansson. On interpolation in automated theorem proving. Technical Report 86/2012, Dipartimento di Informatica, Università degli Studi di Verona, 2013. Submitted to journal August 2013.

[BL11] Matthias Baaz and Alexander Leitsch. *Methods of Cut-Elimination*. Trends in Logic. Springer, 2011.

[CK90] C.C. Chang and H.J. Keisler. *Model Theory*. Studies in Logic and the Foundations of Mathematics. Elsevier Science, 1990.

[Cra57a] William Craig. Linear Reasoning. A New Form of the Herbrand-Gentzen Theorem. *The Journal of Symbolic Logic*, 22(3):250–268, September 1957.

[Cra57b] William Craig. Three uses of the herbrand-gentzen theorem in relating model theory and proof theory. *The Journal of Symbolic Logic*, 22(3):269–285, September 1957.

[Hua95] Guoxiang Huang. Constructing Craig Interpolation Formulas. In *Proceedings of the First Annual International Conference on Computing and Combinatorics*, COCOON '95, pages 181–190, London, UK, UK, 1995. Springer-Verlag.

[Kle67] Stephen Cole Kleene. *Mathematical logic*. Wiley, New York, NY, 1967.

[Kra97] Jan Krajíček. Interpolation Theorems, Lower Bounds for Proof Systems, and Independence Results for Bounded Arithmetic. *Journal of Symbolic Logic*, pages 457–486, 1997.

[Lyn59] Roger C. Lyndon. An interpolation theorem in the predicate calculus. *Pacific Journal of Mathematics*, 9(1):129–142, 1959.

[McM03] Kenneth L. McMillan. Interpolation and SAT-Based Model Checking. In Jr. Hunt, Warren A. and Fabio Somenzi, editors, *Computer Aided Verification*, volume 2725 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin Heidelberg, 2003.

[Pud97] Pavel Pudlák. Lower Bounds for Resolution and Cutting Plane Proofs and Monotone Computations. *J. Symb. Log.*, 62(3):981–998, 1997.

[Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *J. ACM*, 12(1):23–41, January 1965.

[Sho67] Joseph R. Shoenfield. *Mathematical logic*. Addison-Wesley series in logic. Addison-Wesley Pub. Co., 1967.

[Sla70] James R. Slagle. Interpolation theorems for resolution in lower predicate calculus. *J. ACM*, 17(3):535–542, July 1970.

[Tak87] Gaisi Takeuti. *Proof Theory*. Studies in logic and the foundations of mathematics. North-Holland, 1987.

[Wei10] Georg Weissenbacher. *Program Analysis with Interpolants*. PhD thesis, 2010.