Contents

CHAPTER 1

Introduction

1.1 Preliminaries

The language of a first-order formula A is denoted by $L(A)$ 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 term is called *maximal* if it does not occur as subterm of another term.

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. $L(I) \subseteq L(\Gamma) \cap L(\Delta)$. $\Box$

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. A *resolution refutation* of a set of clauses Γ is a number of resolution rule applications (cf. figure 3.1) starting from clauses in Γ which results in the empty clause. △

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

Proof. See [Rob65]. □

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

2.1.1 Interpolation and Skolemisation

In order to apply resolution to arbitrary first-order formulas, they have to be converted to clauses first. This process is composed of a CNF-transformation as well as a skolemisation to remove existential quantifiers. The CNF-transformation clearly has no influence on the interpolant as no symbols are added or removed and the resulting formula is logically equivalent. Skolemisation on the other hand does introduce new symbols and is only satisfiablity-preserving. As we will now see, this does not affect the interpolants.

not the case for tseitin-style

Definition 2.3. Let $V_{\exists x}$ be the set of universally bound variables in the scope of the occurrence of $\exists x$. The skolemisation of a formula A, denoted by $\mathrm{sk}(A)$, is the result of replacing every occurrence of an existential quantifier $\exists x$ in A by $f(y_1, \ldots, y_n)$ where f is a new Skolem function symbol and $V_{\exists x} = \{y_1, \ldots, y_n\}$. In case $V_{\exists x}$ is empty, $\exists x$ 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 2.4. *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 new symbols to both Γ and Δ, I does not contain any of them as they are not contained in $L(\mathrm{sk}(\Gamma)) \cap L(\mathrm{sk}(\Delta))$. Therefore condition 3 of theorem 1.1 is satisfied in both directions.

Since for a 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. □

CHAPTER 3

Constructive Proofs

3.1 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 3.1. A *clause* is a finite set of literals. A *resolution refutation* of a set of clauses Γ is a number of resolution rule applications (cf. figure 3.1) starting from clauses in Γ which results in the empty clause. $\triangle$

Theorem 3.2. *A clause set Γ is unsatisfiable if and only if there is resolution refutation of Γ.*

Proof. See [Rob65]. $\square$

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 3.1: The rules of resolution calculus

3.1.1 Interpolation and Skolemisation

In order to apply resolution to arbitrary first-order formulas, they have to be converted to clauses first. This process is composed of a CNF-transformation as well as a skolemisation to remove existential quantifiers. The CNF-transformation clearly has no influence on the interpolant as no symbols are added or removed and the resulting formula is logically equivalent. Skolemisation on the other hand does introduce new symbols and is only satisfiablity-preserving. As we will now see, this does not affect the interpolants.

not the case for tseitin-style

Definition 3.3. Let $V_{\exists x}$ be the set of universally bound variables in the scope of the occurrence of $\exists x$. The skolemisation of a formula A, denoted by $\mathrm{sk}(A)$, is the result of replacing every occurrence of an existential quantifier $\exists x$ in A by $f(y_1, \ldots, y_n)$ where f is a new Skolem function symbol and $V_{\exists x} = \{y_1, \ldots, y_n\}$. In case $V_{\exists x}$ is empty, $\exists x$ 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 3.4. *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 new symbols to both Γ and Δ, I does not contain any of them as they are not contained in $L(\mathrm{sk}(\Gamma)) \cap L(\mathrm{sk}(\Delta))$. Therefore condition 3 of theorem 1.1 is satisfied in both directions.

Since for a 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. □

3.2 Reduction to first order logic without equality

Let A be a first order formula.

Let $U(E)$ be the conjunction of all $\forall \bar{x} \exists y F_i(\bar{x}, y) \wedge (\forall z F_i(\bar{x}, z) \supset z = y)$ for $f_i \in \mathrm{FS}(E)$.

Let E' be inductively defined as follows: If E does not contains an occurrence of a function symbol, let $E' = E$. Otherwise let f_i be a maximal occurrence of a function symbol and A be the atom in which it occurs. Then A is of the form $P(s_1, \ldots, s_{j-1}, f_i(\bar{t}), s_{j+1}, \ldots s_n)$. Let E_F be E where A is replaced by $\exists y F_i(\bar{t}, y) \wedge P(s_1, \ldots, s_{j-1}, y, s_{j+1}, \ldots s_n)])$ and $E' = E'_F$.

Clearly $E \models_= A$ iff $U(E) \wedge E' \models_= A$.

Let $I(E)$ denote a conjunction between $\forall x\, x = x$ and for all $P \in \mathrm{PS}(E)$, $\forall \bar{x}, \bar{y}\; x_1 = y_1 \supset \ldots \supset x_n = y_n \supset P(\bar{x}) \supset P(\bar{y})$, where n is the arity of P. If $U(E) \wedge E' \models_= A$, also $I(E) \wedge U(E) \wedge E' \models A$.

As $E \models_= A$ iff $I(E) \wedge U(E) \wedge (E) \models A$, E is unsatisfiable iff $I(E) \wedge U(E) \wedge E'$ is. Note that this does not rely on equality and contains no function symbols. Hence by the interpolation theorem for first order logic without equality, there is an interpolant for $\left(\bigcup_{A \in \Gamma} I(A) \wedge U(A) \wedge A\right) \cup \left(\bigcup_{A \in \Delta} I(A) \wedge U(A) \wedge A\right)$ for unsatisfiable $\Gamma \cup \Delta$. Since the equality axioms added via I ensure a valid interpretation of the equality symbol and the F_i can be translated back to f_i in a natural way (as guaranteed by the U), the interpolant

how to state?

more verbose and precise

we receive is also an interpolant for $\Gamma \cup \Delta$. Note that by adding the axiom of reflexivity to both Γ and Δ, it is contained in the intersection of the languages and hence is allowed to appear in the interpolant, which is required.

3.3 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

3.4 WT: Interpolation extraction in two passes

3.4.1 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 relative interpolant w.r.t. the current clause.

Definition 3.5. θ is a *relative 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. $\mathrm{PS}(\theta) \subseteq (\mathrm{PS}(\Gamma) \cap \mathrm{PS}(\Delta)) \cup \{\top, \bot\}$. $\triangle$

The third condition will sometimes be referred to as *language restriction*. It is easy to see that a relative 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 relative interpolants from a resolution refutation.

Definition 3.6. 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 $\mathrm{PS}(l) \in L(\Gamma) \setminus L(\Delta)$:$\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$

   change to "is Γ-colored?"
2. If $\mathrm{PS}(l) \in L(\Delta) \setminus L(\Gamma)$: $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$
3. If $\mathrm{PS}(l) \in L(\Gamma) \cap L(\Delta)$: $\mathrm{PI}(C) = [(l \wedge \mathrm{PI}(C_2)) \vee (l' \wedge \mathrm{PI}(C_1))]\sigma$

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:
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma$ $\triangle$

Proposition 3.7. *Let C be a clause of a resolution refutation. Then* $\mathrm{PI}(C)$ *is a relative 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 $L(\Phi)$. They clearly imply conditions 1 and 2 of definition 3.5.

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 3.6:

1. $\mathrm{PS}(l) \in L(\Gamma) \setminus L(\Delta)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$.
   As $\mathrm{PS}(l) \in L(\Gamma)$, $\Gamma \models (\mathrm{PI}(C_1) \vee D_\Gamma \vee l)\sigma$ 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 $\mathrm{PS}(l) \notin L(\mathrm{PI})$, $\Delta \models (\neg\mathrm{PI}(C_1) \vee D_\Delta)\sigma$ 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 nonlogical symbols are added.
2. $\mathrm{PS}(l) \in L(\Delta) \setminus L(\Gamma)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$.
   As $\mathrm{PS}(l) \notin L(\Gamma)$, $\Gamma \models (\mathrm{PI}(C_1) \vee D_\Gamma)\sigma$ 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 $\mathrm{PS}(l) \in L(\Delta)$, $\Delta \models (\neg\mathrm{PI}(C_1) \vee D_\Delta \vee l)\sigma$ 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. $\mathrm{PS}(l) \in L(\Delta) \cap L(\Gamma)$: 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 relative 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 relative 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 3.5 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 $t_1, \ldots, t_n$ 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 $r_1, \ldots, r_k$ replaced by fresh variables $x_1, \ldots, x_k$ and maximal occurrences of Γ-terms $s_1, \ldots, s_{n-k}$ by fresh variables $x_{k+1}, \ldots, x_n$
3. Return $Q_1 x_1, \ldots Q_n x_n \theta^*$, where Q_i is $\forall$ if t_i is a Δ-term and $\exists$ otherwise.

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

Attempt without P_P:

Definition 3.8. Overline as in paper, replace Δ-terms $t_1, \ldots, t_k$ by respective fresh variables in parenthesis $\triangle$

Lemma 3.9. *$(\overline{C\sigma}(x_1, \ldots, x_n))$ reduces to $(\overline{C}(x_1, \ldots, x_n))\sigma'$, where $\sigma' = \sigma[t_1/x_1] \ldots [t_n/x_n]$.*

$(\overline{C}(x_1, \ldots, x_n))\sigma$ reduces to $(\overline{C\sigma'}(x_1, \ldots, x_n))$ if σ does not change any of $x_1, \ldots, x_n$ or any of $t_1, \ldots, t_n$.

it would work to fix substitutions of x_i by substituting t_i for that instead, as long as the result isn't another t_i, but this isn't actually relevant here.

Proposition 3.10. *$\Gamma = \overline{\Gamma}(x_1, \ldots, x_n)$.*

Proof. By definition, Δ-terms only appear in Δ and not in Γ. $\square$

Lemma 3.11. *$\Gamma \models \overline{(\mathrm{PI}(C) \vee C)}(x_1, \ldots, x_n)$.*

Proof. By induction on the resultion refutation.

Base case: Either $C \in \Gamma$, then it does not contain Δ-terms. Otherwise $C \in \Delta$ and $\mathrm{PI}(C) = \top$.

Induction step:

Resolution.

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

By the induction hypothesis, we can assume that:

$\Gamma \models \overline{\mathrm{PI}(C_1) \vee (D \vee l)}(x_1, \ldots, x_n)$

$\Gamma \models \overline{\mathrm{PI}(C_2) \vee (E \vee \neg l')}(x_1, \ldots, x_n)$

1. $\mathrm{PS}(l) \in L(\Gamma) \setminus L(\Delta)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$.
   We show that $\Gamma \models \overline{(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D \vee E)\sigma}(x_1, \ldots, x_n)$. This is by lemma 3.9 with σ' as in the lemma equivalent to $\Gamma \models \overline{(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D \vee E)}(x_1, \ldots, x_n)\sigma'$.
   By Lemma 11 (Huang) and the induction hypothesis,
   $\Gamma \models \overline{\mathrm{PI}(C_1)} \vee \overline{D} \vee \overline{l}$
   $\Gamma \models \overline{\mathrm{PI}(C_2)} \vee \overline{E} \vee \overline{\neg l'}$
   As $l\sigma = l'\sigma$, $\overline{l\sigma} = \overline{l'\sigma}$.
   Hence $\Gamma \models \overline{\mathrm{PI}(C_1)} \vee \overline{D} \vee \overline{\mathrm{PI}(C_2)} \vee \overline{E}$ and again by Lemma 11 (Huang), $\Gamma \models \overline{\mathrm{PI}(C_1) \vee D \vee \mathrm{PI}(C_2) \vee E}$.
   Also $\Gamma \models \overline{\mathrm{PI}(C_1) \vee D \vee \mathrm{PI}(C_2) \vee E\sigma}$. As $t_1, \ldots, t_n$ do not appear in $\overline{\mathrm{PI}(C_1) \vee D \vee \mathrm{PI}(C_2) \vee E}$ and these are the only variables where σ and σ' differs, we get that $\Gamma \models \overline{\mathrm{PI}(C_1) \vee D \vee \mathrm{PI}(C_2) \vee E}\sigma'$.

2. $\text{PS}(l) \in L(\Delta) \setminus L(\Gamma)$: Then $\text{PI}(C) = [\text{PI}(C_1) \wedge \text{PI}(C_2)]\sigma$.
   We show that $\Gamma \models \overline{((\text{PI}(C_1) \wedge \text{PI}(C_2)) \vee D \vee E)\sigma}(x_1, \ldots, x_n)$. By lemma 3.9 with σ' as in the lemma, $\Gamma \models \overline{((\text{PI}(C_1) \wedge \text{PI}(C_2)) \vee D \vee E)}(x_1, \ldots, x_n)\sigma'$.
   TODO

Paramodulation.

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

By the induction hypothesis, we have:

$\Gamma \models \overline{\text{PI}(C_1) \vee (D \vee s = t)}$

$\Gamma \models \overline{\text{PI}(C_2) \vee (E[r])}$

easy case: $\text{PI}(C) = [(s = t \wedge \text{PI}(C_2)) \vee (s \neq t \wedge \text{PI}(C_1))]\sigma$

to show: $\Gamma \models \overline{[((s = t \wedge \text{PI}(C_2)) \vee (s \neq t \wedge \text{PI}(C_1))) \vee (D \vee E[t])]\sigma}$

proof idea: either $s = t$, then also $\text{PI}(C_2)$, or else $s \neq t$, but then also $\text{PI}(C_1)$

by lemma 3.9 for σ' as in lemma, $\Gamma \models \overline{((s = t \wedge \text{PI}(C_2)) \vee (s \neq t \wedge \text{PI}(C_1))) \vee (D \vee E[t])}\sigma'$

by lemma 11 (huang) $\Gamma \models [((\overline{s} = \overline{t} \wedge \overline{\text{PI}(C_2)}) \vee (\overline{s \neq t} \wedge \overline{\text{PI}(C_1)})) \vee (\overline{D} \vee \overline{E[t]})]\sigma'$

reformulate: $\Gamma \models ((\overline{s}\sigma' = \overline{t}\sigma' \wedge \overline{\text{PI}(C_2)}\sigma') \vee (\overline{s}\sigma' \neq \overline{t}\sigma' \wedge \overline{\text{PI}(C_1)}\sigma')) \vee (\overline{D}\sigma' \vee \overline{E[t]}\sigma')$

By the rule: $s\sigma = r\sigma$, hence also $\overline{s\sigma} = \overline{r\sigma}$ and $\overline{s}\sigma' = \overline{r}\sigma'$ REALLY TRUE? – think so...

Suppose $M \models \Gamma$ and $M \not\models (\overline{D}\sigma' \vee \overline{E[t]}\sigma')$.

Suppose $M \models \overline{s}\sigma' = \overline{t}\sigma'$.

By induction hypothesis (and lemma 11 (huang) and adding the substitution σ'), $\Gamma \models \overline{\text{PI}(C_2)}\sigma' \vee \overline{(E[r])}\sigma'$.

However by assumption $\Gamma \not\models \overline{E[t]}\sigma'$.

Hence $\Gamma \not\models \overline{E[s]}\sigma'$, and $\Gamma \not\models \overline{E[r]}\sigma'$. Therefore $\Gamma \models \overline{\text{PI}(C_2)}\sigma'$.

Suppose on the other hand $M \models \overline{s}\sigma' \neq \overline{t}\sigma'$.

By the induction hypothesis, $M \models \overline{\text{PI}(C_1)}\sigma' \vee (\overline{D}\sigma' \vee (\overline{s} = \overline{t})\sigma')$, hence then $M \models \overline{\text{PI}(C_1)}\sigma'$.

Consequently, $M \models (\overline{s}\sigma' \neq \overline{t}\sigma' \wedge \overline{\text{PI}(C_1)}\sigma') \vee (\overline{s}\sigma' = \overline{t}\sigma' \wedge \overline{\text{PI}(C_2)}\sigma')$.

By lemma 11 (huang), $M \models \overline{(s \neq t \wedge \text{PI}(C_1) \vee (s = t \wedge \text{PI}(C_2))}\sigma'$.

Hence $\Gamma \models \overline{(s \neq t \wedge \text{PI}(C_1) \vee (s = t \wedge \text{PI}(C_2))}\sigma' \vee (\overline{D} \vee \overline{E[t]})\sigma')$.

IS THIS REALLY WHAT I NEED TO SHOW?

□

3.4.2 final step of huang's proof

Theorem 3.12. *$Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\square)^*(z_1, \dots, z_n)$ is a craig interpolant (order as in huang).*

Proof. By lemma 3.11, $\Gamma \models \forall x_1 \dots \forall x_n \overline{\mathrm{PI}(\square)}(x_1, \dots, x_n)$.

The terms in $\overline{PI(\square)}$ are either among the x_i, $1 \le i \le n$ or grey terms or Γ-terms. Let t be a maximal Γ-term in $\overline{\mathrm{PI}(\square)}$. 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 $\mathrm{PI}(\square)^*$, 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 $\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\square)^*(z_1, \dots, z_n)$ □

Conjecture 3.13. *Suppose every variable occurs only once in $\Gamma \cup \Delta$. Then the order of the quantifiers for $\mathrm{PI}(\square)^*$ does not matter.*

The subterm-relation is reflexive.

Definition 3.14. Let C be a clause in $\Gamma \cup \Delta$.

A maximal term s of C is said to be *smaller* than a maximal term t of C if a subterm of s occurs in t and s is of smaller length than t. △

3.4.3 Half-baked approaches

Definition 3.15. Direct interpolation extraction.

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

$Q_1 z_1 \dots Q_{n_1} z_{n_1} \mathrm{PI}(C_1)^*(z_1, \dots, z_{n_1})$
$Q_1 z_1 \dots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \dots, z_{n_2})$

△

Conjecture 3.16. *$Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\square)^*(z_1, \dots, z_n)$, with the z_i ordered by the terms they replace with ordering defined as in 3.14, is a craig interpolant.*

Proof. By lemma 3.11, $\Gamma \models \forall x_1 \dots \forall x_n \overline{\mathrm{PI}(\square)}(x_1, \dots, x_n)$.

The terms in $\overline{PI(\square)}$ are either among the x_i, $1 \le i \le n$ or grey terms or Γ-terms.

Let t be a maximal Γ-term in $\overline{\mathrm{PI}(\square)}$. 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. □

Proposition 3.17. *$\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(C)^*(z_1, \dots, z_n) \vee C$, quantifiers ordered as in 3.14, is a craig interpolant.*

Proof. proof, only for $\Box$:

We know that $\Gamma \models \overline{\mathrm{PI}(\Box)}(x_1, \dots, x_n)$. It still contains the respective maximal Γ-terms.

OTHER ATTEMPT:
Induction.
Base case: simple.
Suppose Resolution.

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

$\Gamma \models Q_1 z_1 \dots Q_{n_1} z_{n_1} \mathrm{PI}(C_1)^*(z_1, \dots, z_{n_1}) \vee D \vee l$
$\Gamma \models Q_1 z_1 \dots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \dots, z_{n_2}) \vee E \vee \neg l'$
to show: $\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(C)^*(z_1, \dots, z_n) \vee (D \vee E)\sigma$

Note that every term that is colored and in C has been introduced by the substitution σ, as otherwise it would occur in both Γ and Δ.

not true if C_1 and C_2 both from same set.

Note that it occurs in either l or l', but not in both.

Let t be a colored term in $\mathrm{PI}(C)$, which has just been added W.l.o.g. let it occur in l, i.e. in C_1.

Case distinction:

t is Γ-colored. Then σ replaces variables $y_1, \dots, y_k$ in $E \vee \neg l'$ with terms that contain t.

By the induction hypothesis, $\Gamma \models Q_1 z_1 \dots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \dots, z_{n_2}) \vee E \vee \neg l'$.

Hence $\Gamma \models (Q_1 z_1 \dots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \dots, z_{n_2}) \vee E \vee \neg l')\sigma$.

Also $\Gamma \models Q_1 z_1 \dots Q_{n_2} z_{n_2} (\mathrm{PI}(C_2)^*(z_1, \dots, z_{n_2})\sigma) \vee E\sigma \vee \neg l'\sigma$.

Similarly, $\Gamma \models Q_1 z_1 \dots Q_{n_1} z_{n_1} (\mathrm{PI}(C_1)^*(z_1, \dots, z_{n_1})\sigma) \vee D\sigma \vee l\sigma$

$\Gamma \models Q_1 z_1 \dots Q_n z_n ((\neg l \wedge \mathrm{PI}(C_2)) \vee (l \wedge \mathrm{PI}(C_1)))^*(z_1, \dots, z_n)\sigma) \vee D\sigma \vee l\sigma$

l basically is the only new thing ($l\sigma = l'\sigma$).

Either l does not contain any subterms of other terms, then it does not depend on anything and l serves as witness for itself.

Otherwise it does depend on other terms and we have to make sure that that term is available. Depending on another term means that it uses information that is only available from another term, i.e. it contains a subterm of another term. but then that subterm is quantified over before the variable that replaces t is, so it works out.

t is Δ-colored. Then it is replaced by a universally quantified variable. But it "was already universally quantified" in the induction hypothesis. There, it was some free variable, because that's the only thing that can be substituted, but even with this free var, it worked out.

$\Box$

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.