0.1 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

0.2 WT: Interpolation extraction in two passes

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

Definition 0.1. 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 0.2. 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 $\mathrm{PS}(l) \in L(\Gamma) \setminus L(\Delta)$: $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$ 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 0.3. *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 0.1.

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 0.2:

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:

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

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 0.1 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 $t_1, \ldots, t_n$ 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 0.4. Overline as in paper, replace Δ-terms $t_1, \dots, t_k$ by respective fresh variables in parenthesis △

Lemma 0.5. *$(\overline{C\sigma}(x_1, \dots, x_n))$ reduces to $(\overline{C}(x_1, \dots, x_n))\sigma'$, where $\sigma' = \sigma[t_1/x_1] \dots [t_n/x_n]$.*

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

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 0.6. *$\Gamma = \overline{\Gamma}(x_1, \dots, x_n)$.*

Proof. By definition, Δ-terms only appear in Δ and not in Γ. □

Lemma 0.7. *$\Gamma \models \overline{(\mathrm{PI}(C) \vee C)}(x_1, \dots, x_n)$.*

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:

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

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

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 $\Gamma \models \overline{(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D \vee E)\sigma}(x_1, \dots, x_n)$. This is by lemma 0.5 with σ' as in the lemma equivalent to $\Gamma \models \overline{(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D \vee E)}(x_1, \dots, x_n)\sigma'$.
   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 $t_1, \dots, t_n$ 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. $\mathrm{PS}(l) \in L(\Delta) \setminus L(\Gamma)$: Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$.
   We show that $\Gamma \models \overline{((\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)) \vee D \vee E)\sigma}(x_1, \ldots, x_n)$. By lemma 0.5 with σ' as in the lemma, $\Gamma \models \overline{((\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)) \vee D \vee E)}(x_1, \ldots, x_n)\sigma'$.
   TODO

Paramodulation.

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

By the induction hypothesis, we have:

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

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

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

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

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

by lemma 0.5 for σ' as in lemma, $\Gamma \models \overline{((s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))) \vee (D \vee E[t])}\sigma'$

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

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

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'), $\Gamma \models \overline{\mathrm{PI}(C_2)}\sigma' \vee \overline{(E[r])}\sigma'$.

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 $\Gamma \models \overline{\mathrm{PI}(C_2)}\sigma'$.

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

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

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

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

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

IS THIS REALLY WHAT I NEED TO SHOW?

$\square$

0.2.2 final step of huang's proof

Theorem 0.8. *$Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\Box)^*(z_1, \dots, z_n)$ is a craig interpolant (order as in huang).*

Proof. By lemma 0.7, $\Gamma \models \forall x_1 \dots \forall x_n \overline{\mathrm{PI}(\Box)}(x_1, \dots, x_n)$.

The terms in $\overline{PI(\Box)}$ are either among the x_i, $1 \le i \le n$ 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.

basically only need the x_j

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.

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 $\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\Box)^*(z_1, \dots, z_n)$

$\Box$

Conjecture 0.9. *Suppose every variable occurs only once in $\Gamma \cup \Delta$. Then the order of the quantifiers for $\mathrm{PI}(\Box)^*$ does not matter.*

The subterm-relation is reflexive.

Definition 0.10. (OLD) Let s be a term that is in $\mathrm{PI}(C)$ but not in any predecessor $\mathrm{PI}(C_i)$, $i \in \{1, 2\}$. s is smaller than a term t in $\mathrm{PI}(C)$ if s is of strictly smaller length than t and there is a subterm in s which also occurs in t. $\triangle$

Definition 0.11. (NEW) A term s is smaller than a term t if in some parent of a rule application, a subterm of s, which is a variable, appears in t and s is smaller in length than t.

$\triangle$

0.2.3 Half-baked approaches

Definition 0.12. Direct interpolation extraction.

This version of overline and star does NOT overbind variables! If they happen to be in the final interpolant, just overbind them somehow, but not earlier. This is ok as the interpolant only contains variables if both corresponding atoms in Γ and Δ do. Variables are the only terms in the interpolant that can "change their color", so we don't know a priori if there are constraints on the quantifier to overbind them with.

Convention w.r.t. a clause C which has been derived from C_1 and C_2: $\bar{Q}_n = Q_1 z_1 \dots Q_n z_n$, such that the z_i correspond to the maximal terms t_i in $\mathrm{PI}(C)$. Same terms must be overbound by same variable, see 101a for counterexample to per-occurrence-overbinding. The z_i are ordered such that

1. the orderings in the Q_{n_1} and Q_{n_2} are respected (no circlular relations can occur in combination with merging as a term is only smaller than another term if it is smaller in length as well, which excludes cycles)
2. as well as ordering constraints of terms newly introduced in $\mathrm{PI}(C)$ (i.e. those that were not present in $\mathrm{PI}(C_1)$ and $\mathrm{PI}(C_2)$).

Basically, track dependencies and define actual order later.

Resolution.

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

$\bar{Q}_{n_1}\mathrm{PI}(C_1)^*$

$\bar{Q}_{n_2}\mathrm{PI}(C_2)^*$

1. l and l' Γ-colored:
   $\mathrm{PI}(C) \equiv (\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2))\sigma$
   $\mathrm{PI}(C)^* \equiv (\mathrm{PI}(C_1)^* \vee \mathrm{PI}(C_2)^*)\sigma$ (just replace maximal terms)
   intended meaning of σ: to change the free variables still in the $\mathrm{PI}(C_i)$
   Let $t_1, \dots, t_{n_1}$ be terms overbound in $\mathrm{PI}(C_1)$ and $s_1, \dots, s_{n_2}$ terms overbound in $\mathrm{PI}(C_2)$.
   $\{z_1, \dots, z_n\} = \{t_1, \dots, t_{n_1}\}\sigma \cup \{s_1, \dots, s_{n_2}\}\sigma$ // common terms are merged
   order relations as in C_1, C_2
   $\bar{Q}_n\mathrm{PI}(C)^* \equiv \bar{Q}_n(\mathrm{PI}(C_1)^* \vee \mathrm{PI}(C_2)^*)$
2. l and l' Δ-colored:
   similar to first case
3. l and l' grey:
   $\mathrm{PI}(C) \equiv [(\neg l' \wedge \mathrm{PI}(C_1)) \vee (l \wedge \mathrm{PI}(C_2))]\sigma$
   $\mathrm{PI}(C)^* \equiv [(\neg l'^* \wedge \mathrm{PI}(C_1)^*) \vee (l^* \wedge \mathrm{PI}(C_2)^*)]\sigma$
   Let $t_1, \dots, t_{n_1}$ be terms overbound in $\mathrm{PI}(C_1)$, $s_1, \dots, s_{n_2}$ terms overbound in $\mathrm{PI}(C_2)$ and $r_1, \dots, r_{n_3}$ be the maximal colored terms of $l\sigma$ and $l'\sigma$ (need to apply σ here because we there might be grey variables replaced by colored terms)
   $\{z_1, \dots, z_n\} = \{t_1, \dots, t_{n_1}\}\sigma \cup \{s_1, \dots, s_{n_2}\}\sigma \cup \{r_1, \dots, r_{n_3}\}$
   order relations as in C_1, C_2 plus:
   + If r_i is smaller in length than t_j (s_j) and a subterm of r_i occurs in t_j (s_j), then r_i is smaller than t_j (s_j).
   + If r_i is larger in length than t_j (s_j) and a subterm of t_j (s_j) occurs in r_i, then r_i is larger than t_j (s_j).
   + If $z_i\sigma \neq z_i$, we have to potentially add new dependencies (cf. 102b). TODO: check only t_j if change in s_j and similar?

   $\bar{Q}_n\mathrm{PI}(C)^* \equiv \bar{Q}_n[(\neg l'^* \wedge \mathrm{PI}(C_1)^*) \vee (l^* \wedge \mathrm{PI}(C_2)^*)]\sigma$

$\triangle$

Conjecture 0.13. *$Q_1 z_1 \dots Q_n z_n \mathrm{PI}(\Box)^*(z_1, \dots, z_n)$, with the z_i ordered by the terms they replace with ordering defined as in 0.11, is a craig interpolant.*

Proof. By lemma 0.7, $\Gamma \models \forall x_1 \dots \forall x_n \overline{\mathrm{PI}(\Box)}(x_1, \dots, x_n)$.

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. $\Box$

Proposition 0.14. *$\Gamma \models Q_1 z_1 \dots Q_n z_n \mathrm{PI}(C)^*(z_1, \dots, z_n) \vee C$, quantifiers ordered as in 0.11, is a craig interpolant.*

Proof. Induction.

Base case: simple.

Suppose Resolution.

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

$\Gamma \models \bar{Q}_{n_1} \mathrm{PI}(C_1)^* \vee D \vee l$
$\Gamma \models \bar{Q}_{n_2} \mathrm{PI}(C_2)^* \vee E \vee \neg l'$
to show: $\Gamma \models \bar{Q}_n \mathrm{PI}(C)^* \sigma \vee (D \vee E)\sigma$

Note that a term newly introduced in $\mathrm{PI}(C)$ 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:

1. Suppose l, l' are from Γ alone:

   By induction hypothesis:

   $\Gamma \models (\bar{Q}_{n_1} \mathrm{PI}(C_1)^* \vee D \vee l)\sigma$

   $\Gamma \models (\bar{Q}_{n_2} \mathrm{PI}(C_2)^* \vee E \vee \neg l')\sigma$

   By resolution:

   $\Gamma \models (\bar{Q}_{n_1} \mathrm{PI}(C_1)^* \vee \bar{Q}_{n_2} \mathrm{PI}(C_2)^*)\sigma \vee (D \vee E)\sigma$

   Suppose t is Γ-colored.

   Then it will be replaced by x_i and existentially quantified. It appears in either $\mathrm{PI}(C_1)$ or $\mathrm{PI}(C_2)$.

   t is a witness for x_i because it contains subterms $t_1, \dots, t_n$. If they are overbound as well, they are so before t and are available here.

   TODO: derive properties using examples 103 or so

Then σ replaces variables $y_1, \ldots, y_k$ in $E \vee \neg l'$ with terms that contain t.
By the induction hypothesis, $\Gamma \models Q_1 z_1 \ldots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \ldots, z_{n_2}) \vee E \vee \neg l'$.
Hence $\Gamma \models (Q_1 z_1 \ldots Q_{n_2} z_{n_2} \mathrm{PI}(C_2)^*(z_1, \ldots, z_{n_2}) \vee E \vee \neg l')\sigma$.
Also $\Gamma \models Q_1 z_1 \ldots Q_{n_2} z_{n_2} (\mathrm{PI}(C_2)^*(z_1, \ldots, z_{n_2})\sigma) \vee E\sigma \vee \neg l'\sigma$.
Similarily, $\Gamma \models Q_1 z_1 \ldots Q_{n_1} z_{n_1} (\mathrm{PI}(C_1)^*(z_1, \ldots, z_{n_1})\sigma) \vee D\sigma \vee l\sigma$
$\Gamma \models Q_1 z_1 \ldots Q_n z_n ((\neg l \wedge \mathrm{PI}(C_2)) \vee (l \wedge \mathrm{PI}(C_1)))^*(z_1, \ldots, z_n)\sigma) \vee D\sigma \vee l\sigma$
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.

□

Proposition 0.15. *Let $A(x_1, \ldots, x_n)$ be an atom in a relative interpolant. A variable occurs in one of the x_i if and only if there are atoms $A(y_1, \ldots, y_n)$ and $A(z_1, \ldots, z_n)$ in Γ and Δ respectively, where x_i can be unified with z_i and y_i such that there is still a variable at that location.*

This means that either the term structure above the variable is the same in the original clauses or there are some variables. Intended meaning: the original clauses prove at least the x_i, i.e. are at least as or more general.

Special case for outermost variables:

Let $A(x_1, \ldots, x_n)$ be an atom in a relative interpolant. An x_i is a variable if and only if there are atoms $A(y_1, \ldots, y_n)$ and $A(z_1, \ldots, z_n)$ in Γ and Δ respectively, where y_i and z_i are variables.

need more narrow version: clauses do appear in parent clauses in derivation.

Proposition 0.16. *Suppose in a partial interpolant, there are two maximal terms t_1 and t_2 such that w.l.o.g. t_1 is smaller (as defined in 0.11) than t_2. Then it the final interpolant, an overbinding can be defined where the variable corresponding to t_1 is quantified over before the variable corresponding to t_2 is.*