1 Proof of the correctness of Huang's algorithm without propositional refutations

TODO: Notation: $\ell^x_\Delta[t] = x_t$, then $x_t\sigma' = x_{t\sigma}$

1.1 Lemmas

Intuition of σ':
If we pull a substitution out of a lifting which replaces Δ-terms, we also have to replace the Δ-terms in the "codomain" of the substitution. This is the second case in the definition of σ' below.
There is just a problem in the following case: $\ell^x_\Delta[f(x)\sigma]$, where $x\sigma = a$ and f is a Δ-symbol. Then $\ell^x_\Delta[f(x)\sigma] = \ell^x_\Delta[f(a)] = x_i$, but $\ell^x_\Delta[f(x)]\sigma = x_j$ with $i \neq j$. The first case of the definition of x_j then fixes this by replacing x_j with x_i.

Lemma 1. *Let C be a clause and σ a substitution. Let $t_1, \dots, t_n$ be all maximal Δ-terms in this context, i.e. those that occur in C or $C\sigma$, and $x_1, \dots, x_n$ the corresponding fresh variables to replace the t_i (i.e. none of the x_i occur in C). Define σ' with $\mathrm{dom}(\sigma') = \mathrm{dom}(\sigma) \cup \{x_t \mid t\sigma \neq t\}$ such that for a variable z,*

$$z\sigma' = \begin{cases} x_{t\sigma} & \text{if } z = x_t \text{ and } t\sigma \neq t \\ \ell^x_\Delta[z\sigma] & \text{otherwise} \end{cases}$$

Then $\ell^x_\Delta[C\sigma] = \ell^x_\Delta[C]\sigma'$.

Proof. As substitutions and liftings only affect the terms of a clause, it suffices to show that $\ell^x_\Delta[t\sigma] = \ell^x_\Delta[t]\sigma'$ for for a term t in C. More precisely, only variables of $\mathrm{dom}(\sigma)$ and maximal Δ-terms are affected. We show that for terms t of either kind that $\ell^x_\Delta[t\sigma] = \ell^x_\Delta[t]\sigma'$ holds, which proves the lemma.
For the first kind of affected terms, suppose that v is a variable in $\mathrm{dom}(\sigma)$, which occurs in C. Then $\ell^x_\Delta[v\sigma] = x_{v\sigma}$. But $\ell^x_\Delta[v]\sigma' = v\sigma' = \ell^x_\Delta[v\sigma] = x_{v\sigma}$.
For the second kind of affected terms, let t be a maximal Δ-term in C. Then $\ell^x_\Delta[t\sigma] = x_{t\sigma}$. We show that $x_{t\sigma} = \ell^x_\Delta[t]\sigma'$.
Suppose that $t\sigma = t$. Then $\ell^x_\Delta[t]\sigma' = x_t\sigma' = x_t = x_{t\sigma}$. Note that x_r must not occur in t for some term r, as $x_r\sigma = x_r$, but potentially $x_r\sigma' \neq x_r$.
Otherwise it is the case that $t\sigma \neq t$. Then $\ell^x_\Delta[t]\sigma' = x_t\sigma'$, and by the definition of σ', $x_t\sigma' = x_{t\sigma}$. $\square$

Lemma 2 (corresponds to Lemma 4.8 in thesis and Lemma 11 in Huang)**.** *Let A and B be first-order formulas and s and t be terms. Then it holds that:*

1. $\ell^x_\Phi[\neg A] \Leftrightarrow \neg\ell^x_\Phi[A]$
2. $\ell^x_\Phi[A \circ B] \Leftrightarrow (\ell^x_\Phi[A] \circ \ell^x_\Phi[B])$ *for* $\circ \in \{\wedge, \vee\}$
3. $\ell^x_\Phi[s = t] \Leftrightarrow (\ell^x_\Phi[s] = \ell^x_\Phi[t])$

Lemma 3. *Let M be a model, E a formula and s and t terms such that $M \not\models \ell^x_\Delta[E[t]_p]$ and $M \models (\ell^x_\Delta[s]) = (\ell^x_\Delta[t])$.*
Let $h[t]$ be a maximal Δ-colored term containing t at p in $E[t]_p$, if such a term exists.

- *If $h[t]$ does not exists, then $M \not\models \ell^x_\Delta[E[s]_p]$.*
- *Otherwise $M \not\models \ell^x_\Delta[E[s]_p]$ or $M \models (\ell^x_\Delta[h[s]]) \neq (\ell^x_\Delta[h[t]])$ holds.*

Proof. Suppose that t at p in $E[t]_p$ is not contained in a Δ-colored term. Then $\ell^x_\Delta[E[t]_p]$ and $\ell^x_\Delta[E[s]_p]$ only differ at position p, where at the first, there is $\ell^x_\Delta[t]$, and at the latter, there is $\ell^x_\Delta[s]$. But in M, they are interpreted the same way, hence $M \models \ell^x_\Delta[E[s]_p] \Leftrightarrow M \models \ell^x_\Delta[E[t]_p]$, which implies the result.
Otherwise t at p in $E[t]_p$ is contained in the maximal Δ-colored term $h[t]$. Suppose that $M \models (\ell^x_\Delta[h[s]]) = (\ell^x_\Delta[h[t]])$ as otherwise we would be done. But then $M \models \ell^x_\Delta[E[s]_p] \Leftrightarrow \ell^x_\Delta[E[t]_p]$. $\square$

1.2 Definition of PI

We use basically the same definition of PI as Huang with minor adaptions for paramodulation (deviations are marked):

Definition 4 (Propositional interpolant extraction.)**.** Let π be a resolution refutation of $\Gamma \cup \Delta$. $\mathrm{PI}(\pi)$ is defined to be $\mathrm{PI}(\square)$, where $\square$ is the empty clause derived in π.
For a clause C in π, $\mathrm{PI}(C)$ is defined as follows:

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

Resolution. 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 l is Γ-colored: $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$
2. If l is Δ-colored: $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2)]\sigma$
3. If l is grey: $\mathrm{PI}(C) = [(l \wedge \mathrm{PI}(C_2)) \vee (\neg l' \wedge \mathrm{PI}(C_1))]\sigma$

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. Suppose 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 $r\sigma = s\sigma$. Let $h[r]$ be the maximal colored term in which r occurs in $D[r]$. Then $\mathrm{PI}(C)$ is defined according to the following case distinction:

1. If $h[r]$ is Δ-colored: // Huang has the additional clause (not applied here): $h[r]$ occurs more than once in $D[r] \vee \mathrm{PI}(D[r])$
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma \vee (s = t \wedge h[s] \neq h[t])\sigma$
2. If $h[r]$ is Γ-colored: // Huang has the additional clause (not applied here): $h[r]$ occurs more than once in $D[r] \vee \mathrm{PI}(D[r])$
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma \wedge (s \neq t \vee h[s] = h[t])\sigma$

3. If r does not occur in a colored term in $D[r]$:
   $\mathrm{PI}(C) = [(s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))]\sigma$ $\triangle$

1.3 Adaption of central lemma

Now we show the "main" lemma of Huang's proof without using a propositional deduction P_P. The remaining part of his proof after this lemma does not use the restriction to propositional deductions and hence goes through.

Lemma 5 (corresponds to Lemma 12 in Huang and Lemma 4.9 in the thesis)**.** *Let π be a resolution refutation of $\Gamma \cup \Delta$. Then for $C \in \pi$, $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C) \vee C]$.*

Proof. By induction on the resolution refutation of the strengthening: $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C) \vee C_\Gamma]$, i.e. we only consider literals of C which are contained in $\mathrm{L}(\Gamma)$.
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:

$\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1) \vee (D \vee l)_\Gamma]$ and $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_2) \vee (E \vee \neg l')_\Gamma]$

which by Lemma 2 implies that

$\Gamma \overset{(*)}{\models} \ell^x_\Delta[\mathrm{PI}(C_1)] \vee \ell^x_\Delta[D_\Gamma] \vee \ell^x_\Delta[l_\Gamma]$ and $\Gamma \overset{(\circ)}{\models} \ell^x_\Delta[\mathrm{PI}(C_2)] \vee \ell^x_\Delta[E_\Gamma] \vee \neg\ell^x_\Delta[l'_\Gamma]$

Let σ' be defined as in Lemma 1 with $t_1, \dots, t_n$ all Δ-terms in this context (we need that every maximal Δ-term has a distinct index, so take all occurring in C_1, C_2, $\mathrm{PI}(C_1)$, $\mathrm{PI}(C_2)$, with and without σ applied to them).

Case distinction:

1. l is Γ-colored. Then $\mathrm{PI}(C) = [\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2)]\sigma$.

   We show that $\Gamma \models \ell^x_\Delta[(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2))\sigma \vee (D \vee E)_\Gamma\sigma]$,
   i.e. $\Gamma \models \ell^x_\Delta[\Big(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D_\Gamma \vee E_\Gamma\Big)\sigma]$.

   Hence by Lemma 1, $\Gamma \models \ell^x_\Delta[(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D_\Gamma \vee E_\Gamma)]\sigma'$.

   Since $\sigma = \mathrm{mgu}(l, l')$, $l\sigma$ and $l'\sigma$ are syntactically equal and so $\ell^x_\Delta[l\sigma] = \ell^x_\Delta[l'\sigma]$.

   As by Lemma 1 $\ell^x_\Delta[l\sigma] = \ell^x_\Delta[l]\sigma'$ and $\ell^x_\Delta[l'\sigma] = \ell^x_\Delta[l']\sigma'$, we get $\ell^x_\Delta[l]\sigma' = \ell^x_\Delta[l']\sigma'$.

   So by applying σ' to $(*)$ and $(\circ)$ (note that $l_\Gamma = l$ and $l'_\Gamma = l'$ as they are Γ-colored), we can perform a resolution step on $\ell^x_\Delta[l]\sigma'$ and get

   $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma' \vee \ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma' \vee \ell^x_\Delta[E_\Gamma]\sigma'$.

   and consequently $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D_\Gamma \vee E_\Gamma]\sigma'$.

   So by Lemma 1,

   $\Gamma \models \ell^x_\Delta[\Big(\mathrm{PI}(C_1) \vee \mathrm{PI}(C_2) \vee D_\Gamma \vee E_\Gamma\Big)\sigma]$.

2. l is Δ-colored. Then $\mathrm{PI}(C) = (\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2))\sigma$.

   We show that $\Gamma \models \ell^x_\Delta[(\mathrm{PI}(C_1) \wedge \mathrm{PI}(C_2))\sigma \vee (D_\Gamma \vee E_\Gamma)\sigma]$

   which by Lemma 2 is equivalent to
   $\Gamma \models \Big(\ell^x_\Delta[\mathrm{PI}(C_1)\sigma] \wedge \ell^x_\Delta[\mathrm{PI}(C_2)\sigma]\Big) \vee \ell^x_\Delta[D_\Gamma\sigma] \vee \ell^x_\Delta[E_\Gamma\sigma]$

   and by Lemma 1 is equivalent to
   $\Gamma \overset{(\circledast)}{\models} \Big(\ell^x_\Delta[\mathrm{PI}(C_1)]\sigma' \wedge \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'\Big) \vee \ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[E_\Gamma]\sigma'$

   As l and l' are Δ-colored, we can simplify $(*)$ and $(\circ)$ as follows and apply σ':

   $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma' \vee \ell^x_\Delta[D_\Gamma]\sigma'$ and $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma' \vee \ell^x_\Delta[E_\Gamma]\sigma'$

   These clearly imply $(\circledast)$.

3. l is grey. Then $\mathrm{PI}(C) = [(l \wedge \mathrm{PI}(C_2)) \vee (\neg l' \wedge \mathrm{PI}(C_2))]\sigma$.

   We show that $\Gamma \models \ell^x_\Delta[\Big((l \wedge \mathrm{PI}(C_2)) \vee (\neg l' \wedge \mathrm{PI}(C_2)) \vee D_\Gamma \vee E_\Gamma\Big)\sigma]$, which by Lemma 2 and Lemma 1 is equivalent to

   $\Gamma \models \Big(\ell^x_\Delta[l]\sigma' \wedge \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'\Big) \vee \Big(\neg\ell^x_\Delta[l']\sigma' \wedge \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'\Big) \vee \ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[E_\Gamma]\sigma'$.

   Suppose for a model M of Γ that $M \not\models \ell^x_\Delta[D_\Gamma]\sigma'$ and $M \not\models \ell^x_\Delta[E_\Gamma]\sigma'$ as otherwise we would be done. But then by $(*)$ and $(\circ)$, $M \models \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma' \vee \ell^x_\Delta[l]\sigma'$ and $M \models \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma' \vee \neg\ell^x_\Delta[l']\sigma'$.

   As observed in case 1, $\ell^x_\Delta[l]\sigma' = \ell^x_\Delta[l']\sigma'$. By a case distinction on the truth value of $\ell^x_\Delta[l]\sigma'$, we obtain the result.

Factorisation.

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

Then $\mathrm{PI}(C) = \mathrm{PI}(C_1)\sigma$.

The induction hypothesis gives that $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1) \vee l \vee l' \vee D]$. Let σ' be as in Lemma 1.

Then $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1) \vee l \vee l' \vee D]\sigma'$ and by Lemma 1, $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)\sigma \vee l\sigma \vee l'\sigma \vee D\sigma]$.

By Lemma 2, $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)\sigma] \vee \ell^x_\Delta[l\sigma] \vee \ell^x_\Delta[l'\sigma] \vee \ell^x_\Delta[D\sigma]$.

As $\sigma = \mathrm{mgu}(l, l')$, $l\sigma$ and $l'\sigma$ are syntactically equal, hence $\ell^x_\Delta[l\sigma] = \ell^x_\Delta[l'\sigma]$.

But then we can apply a factorisation step and get $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)\sigma] \vee \ell^x_\Delta[l\sigma] \vee \ell^x_\Delta[D\sigma]$ and by Lemma 1 and Lemma 2, $\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1)\sigma \vee l\sigma \vee D\sigma]$.

Paramodulation.

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

By the induction hypothesis, we have:

$\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_1) \vee (D \vee s = t)_\Gamma]$

$\Gamma \models \ell^x_\Delta[\mathrm{PI}(C_2) \vee (E[r]_p)_\Gamma]$

By Lemma 1 and Lemma 2, we get that:

$\Gamma \overset{(\circ)}{\models} \ell^x_\Delta[\mathrm{PI}(C_1)] \vee \ell^x_\Delta[D_\Gamma] \vee \ell^x_\Delta[s] = \ell^x_\Delta[t]$

$\Gamma \overset{(*)}{\models} \ell^x_\Delta[\mathrm{PI}(C_2)] \vee \ell^x_\Delta[(E[r]_p)_\Gamma]$

We distinguish two cases:

1. Suppose s does not occur in a maximal Δ-term $h[s]$ in $E[s]_p$

   We show that $\Gamma \models \ell^x_\Delta[\Big((s = t \wedge \mathrm{PI}(C_2)) \vee (s \neq t \wedge \mathrm{PI}(C_1))\Big)\sigma \vee \Big((D \vee E[t]_p)_\Gamma\Big)\sigma]$, which subsumes the cases 2 and 3 of the definition of PI for paramodulation. By Lemma 2, we can pull the liftings inwards and by Lemma 1, we can commute substitution and lifting by employing σ' to arrive at

   $\Gamma \models \Big((\ell^x_\Delta[s]\sigma') = (\ell^x_\Delta[t]\sigma') \wedge \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'\Big) \vee \Big((\ell^x_\Delta[s]\sigma') \neq (\ell^x_\Delta[t]\sigma') \wedge \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma'\Big) \vee \Big(\ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[(E[t]_p)_\Gamma]\sigma'\Big)$

   Let M be a model of Γ. Let $M \not\models \ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[(E[t]_p)_\Gamma]\sigma'$ as otherwise we would be done. We show that depending on the truth value of $(\ell^x_\Delta[s]) = (\ell^x_\Delta[t])$ in M, either the first or second conjunct of the above formula holds.

   Suppose that $M \models (\ell^x_\Delta[s]) \neq (\ell^x_\Delta[t])$. Then by $(\circ)$, $M \models \ell^x_\Delta[\mathrm{PI}(C_1)]$ and hence $M \models \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma'$.

   On the other hand, suppose that $M \models (\ell^x_\Delta[s]) = (\ell^x_\Delta[t])$. Then by Lemma 3, as s at p in $E[s]_p$ does not occur in a maximal Δ-term, $M \not\models \ell^x_\Delta[E[s]_p]$. Hence also $M \not\models \ell^x_\Delta[E[s]_p]\sigma'$ and by Lemma 1, $M \not\models \ell^x_\Delta[(E[s]_p)\sigma]$.

   Due to $\sigma = \mathrm{mgu}(s, r)$, $s\sigma$ and $r\sigma$ are syntactically equal. Suppose they are both not Δ-colored. Then the lifting does not affect them and $\ell^x_\Delta[(E[s]_p)\sigma]$ and $\ell^x_\Delta[(E[r]_p)\sigma]$ are the same formula. Otherwise the lifting will replace them with the same variable and we as well get that $\ell^x_\Delta[(E[s]_p)\sigma]$ and $\ell^x_\Delta[(E[r]_p)\sigma]$ are the same formula.

   By Lemma 1, $\ell^x_\Delta[(E[s]_p)]\sigma' = \ell^x_\Delta[(E[r]_p)]\sigma'$, so from $M \not\models \ell^x_\Delta[E[s]_p]\sigma'$, it follows that $M \not\models \ell^x_\Delta[(E[r]_p)]\sigma'$

   Then by $(*)$, we arrive at $M \models \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'$

2. Otherwise s occurs in a maximal Δ-term $h[s]_q$ in $E[s]_p$.

   Then a similar line of argument as in case 1 can be employed, with the difference that the application of Lemma 3 yields the extra case that $M \models (\ell^x_\Delta[h[s]]) \neq (\ell^x_\Delta[h[t]])$. Hence the following holds:

   $\Gamma \models \Big((\ell^x_\Delta[s]\sigma') = (\ell^x_\Delta[t]\sigma') \wedge \ell^x_\Delta[\mathrm{PI}(C_2)]\sigma'\Big) \vee \Big((\ell^x_\Delta[s]\sigma') \neq (\ell^x_\Delta[t]\sigma') \wedge \ell^x_\Delta[\mathrm{PI}(C_1)]\sigma'\Big) \vee \Big((\ell^x_\Delta[s]\sigma') = (\ell^x_\Delta[t]\sigma') \wedge (\ell^x_\Delta[h[s]]\sigma') \neq (\ell^x_\Delta[h[t]]\sigma')\Big) \vee \Big(\ell^x_\Delta[D_\Gamma]\sigma' \vee \ell^x_\Delta[(E[t]_p)_\Gamma]\sigma'\Big)$ □

From this point on, the following from Huang/my thesis go through:
Lemma 4.10: swap Γ and Δ and obtain logical negation as interpolant
Corollary 4.11: $\Delta \models \ell^y_\Gamma[\neg\,\mathrm{PI}(C) \vee C]$

Lemma 4.12: not important if lifting delta or gamma terms first
Thm 4.13: ordering