1 Overbinding in one step

Conjecture 1. *Suppose every variable occurs only once in $\Gamma \cup \Delta$. Then the order of the quantifiers for* $\mathrm{PI}(\Box)^*$ *does not matter.*

Proposition 2. *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 3. *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 5) 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.*

The subterm-relation is reflexive.

Definition 4. (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 5. (NEW)
Let C be a clause.
A maximal term s of C is smaller than a maximal term t of C if s is a variable and occurs in t, but $s \neq t$.

$\triangle$

2 Half-baked approaches

Definition 6. 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: $\bar{Q}_n = Q_1 z_1 \ldots Q_n z_n$, 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.

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

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

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

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 $\mathrm{PI}(C_i)$

   TODO: basically do nothing here since no new atoms (revisit after mixed colored case has been dealt with)

   Let $t_1, \ldots, t_{n_1}$ be terms overbound in $\mathrm{PI}(C_1)$ and $s_1, \ldots, s_{n_2}$ terms overbound in $\mathrm{PI}(C_2)$.

   $\{z_1, \ldots, z_n\} = \{t_1, \ldots, t_{n_1}\}\sigma \cup \{s_1, \ldots, s_{n_2}\}\sigma$ // 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:

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

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

   // just replace any atoms, note that vars are exempt

   // need to star at the end again for terms introduced by sigma

   order relations as in C_1, C_2 plus:

   Let C' and C'' be the clauses in $\Gamma \cup \Delta$ where l and l' originate.

   If in C' (C'') a maximal term s of l (l'') is smaller than a maximal term t of the same clause, and x_i replaces s and x_j replaces t in $\mathrm{PI}(C)^*$, then $x_i < x_j$.

   If in $\mathrm{PI}(C)^*$, x_i and x_j, $i \neq j$ replace t_i and t_j respectively, and t_i and t_j have a common origin where they were the same variable, then merge these variables in $\mathrm{PI}(C)^*$.

   Let $t_1, \ldots, t_{n_1}$ be the maximal colored terms in $\mathrm{PI}(C_1)$ and $s_1, \ldots, s_{n_2}$ the maximal colored terms in $\mathrm{PI}(C_2)$, Let $r_1, \ldots, r_{n_3}$ be the maximal colored terms in $[(\neg l' \wedge \mathrm{PI}(C_1)^*) \vee (l \wedge \mathrm{PI}(C_2)^*)]\sigma$

// this way, we catch all colored terms in the new atoms + every term that becomes colored due to σ changing a var.

$\{z_1, \ldots, z_{n_1}\} = \{t_1, \ldots, t_{n_1}\}$

$\{z_{n_1}, \ldots, z_{n_1+n_2}\} = \{s_1, \ldots, s_{n_2}\}$

$\{z_{n_1+n_2}, \ldots, z_{n_1+n_2+n_3}\} = \{r_1, \ldots, r_{n_3}\}$

$\bar{Q}_n \sim z_i$ ordered according to constraints and with quantifier.

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

$\bar{Q}_n\overline{\mathrm{PI}(C)} \equiv \bar{Q}_n\overline{[(\neg l' \wedge \mathrm{PI}(C_1)^*) \vee (l \wedge \mathrm{PI}(C_2)^*)]\sigma}$ // somewhat imprecise on $\bar{Q}_n$, but that's just useless quantifiers

$\triangle$

Proposition 7. $\Gamma \models Q_1 z_1 \ldots Q_n z_n \overline{\mathrm{PI}(C) \vee C}(z_1, \ldots, z_n)$ *, quantifiers ordered as in 5, is a craig interpolant.*

Proof. Induction.
Suppose Resolution.

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

$\Gamma \models \bar{Q}_{n_1} \overline{\mathrm{PI}(C_1) \vee D \vee l}$
$\Gamma \models \bar{Q}_{n_2} \overline{\mathrm{PI}(C_2) \vee E \vee \neg l'}$
to show:
$\Gamma \models \bar{Q}_n \overline{\mathrm{PI}(C) \vee (D \vee E)\sigma}$ // somewhat imprecise on $\bar{Q}_n$, but that's just useless quantifiers
$\Gamma \models (\bar{Q}_{n_1} \overline{PI(C_1)} \vee D \vee l)\sigma$
$\Gamma \models (\bar{Q}_{n_2} \overline{PI(C_2)} \vee E \vee \neg l')\sigma$
By resolution:
$\Gamma \models (\bar{Q}_{n_1} \overline{\mathrm{PI}(C_1)} \vee \bar{Q}_{n_2} \overline{\mathrm{PI}(C_2)})\sigma \vee (D \vee E)\sigma$

1. Suppose l, l' are from Γ alone: TODO
2. Suppose l and l' are colored with different colors and w.l.o.g l is Γ-colored and l' is Δ-colored.

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

   $\equiv \bar{Q}_n (\overline{\neg l'\sigma} \wedge \overline{\overline{\mathrm{PI}(C_1)}\sigma}) \vee (\overline{l\sigma} \wedge \overline{\overline{\mathrm{PI}(C_2)}\sigma})$

   Adapt Huang proof to this, need to consider quantifiers:

   If $\Gamma \not\models D\sigma$ and $\Gamma \not\models E\sigma$ (else we are done), then

   $\Gamma \models [(\neg l' \wedge \bar{Q}_{n_1} \overline{PI(C_1)}) \vee (l \wedge \bar{Q}_{n_2} \overline{\mathrm{PI}(C_2)})]\sigma$

   As $\bar{Q}_{n_1}$ and $\bar{Q}_{n_2}$ disjoint and their variables do not appear in l or l',

   $\Gamma \models (\bar{Q}_{n_1} \bar{Q}_{n_2} [(\neg l' \wedge \overline{PI(C_1)}) \vee (l \wedge \overline{PI(C_2)})])\sigma$

   $\Gamma \models \bar{Q}_{n_1} \bar{Q}_{n_2} [(\neg l'\sigma \wedge \overline{\mathrm{PI}(C_1)}\sigma) \vee (l\sigma \wedge \overline{\mathrm{PI}(C_2)}\sigma)]$

   Consider the maximal terms of this expression which are Δ-colored.

   The $\mathrm{PI}(C_i)$, $i \in \{1, 2\}$ contain no colored terms. σ can introduce one by replacing a free variable x by a Δ-term t. But then overline replaces it with an universally quantified variable again, hence the formula is still entailed by Γ.

   $\Gamma \models \bar{Q}_{n_1} \bar{Q}_{n_2} [(\neg l'\sigma \wedge \overline{\overline{\mathrm{PI}(C_1)}\sigma}) \vee (l\sigma \wedge \overline{\overline{\mathrm{PI}(C_2)}\sigma})]$

   TODO: should work out similarily as huang if using P_P or it's the same as what i'm trying above.

□

Proposition 8. $\Gamma \models Q_1 z_1 \ldots Q_n z_n \, \mathrm{PI}(C)^*(z_1, \ldots, z_n) \vee C$*, quantifiers ordered as in 5, is a craig interpolant.*

Proof. Induction.
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: $\Gamma \models \bar{Q}_n\, \mathrm{PI}(C)^* \vee (D \vee E)\sigma$
$\Gamma \models (\bar{Q}_{n_1}\, \mathrm{PI}(C_1)^* \vee D \vee l)\sigma$
$\Gamma \models (\bar{Q}_{n_2}\, \mathrm{PI}(C_2)^* \vee E \vee \neg l')\sigma$
By resolution:
$\Gamma \models (\bar{Q}_{n_1}\, \mathrm{PI}(C_1)^* \vee \bar{Q}_{n_2}\, \mathrm{PI}(C_2)^*)\sigma \vee (D \vee E)\sigma$

1. Suppose l, l' are from Γ alone: TODO

2. Suppose l and l' are colored with different colors and w.l.o.g l is Γ-colored and l' is Δ-colored.

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

   Adapt Huang proof to this, need to consider quantifiers:

   If $\Gamma \not\models D\sigma$ and $\Gamma \not\models E\sigma$ (else we are done), then

   $\Gamma \models [(\neg l' \wedge \bar{Q}_{n_1}\, \mathrm{PI}(C_1)^*) \vee (l \wedge \bar{Q}_{n_2}\, \mathrm{PI}(C_2)^*)]\sigma$

   As $\bar{Q}_{n_1}$ and $\bar{Q}_{n_2}$ disjoint and their variables do not appear in l or l',

   $\Gamma \models (\bar{Q}_{n_1}\bar{Q}_{n_2}[(\neg l' \wedge \mathrm{PI}(C_1)^*) \vee (l \wedge \mathrm{PI}(C_2)^*)])\sigma$

   The $\mathrm{PI}(C_i)$, $i \in \{1, 2\}$ contain no colored terms. σ can introduce one by replacing a free variable x.

   Consider the maximal terms of this expression which are Γ-colored.

   Either they only have grey subterms, then if they are existentially quantified, we can just use it as witness as the terms aren't replaced.

   Otherwise they contain at least a Γ- or a Δ-colored subterm.

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:

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

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

   By resolution:

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

   Suppose t is Γ-colored.

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

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

   TODO: derive properties using examples 103 or so

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

□

3 Arrow-Algo

1. In the original clauses, find all occurrences of variables. Add an arrow from an occurrence to each other occurrence with depth as least as high, if the full prefix to the occurrence with lower depth is shared by both occurrences (cf. 5).

   NOTE: this creates double arrows for occurrences at same depth. This appears to be necessary for terms which are only variables, and doesn't hurt if the variable is contained in a term.

2. For each step in the derivation:

   a) Build propositional interpolant using $\mathrm{PI}(C_i)^*$, $i \in \{1,2\}$, i.e. use ancestor PI without colored terms.

   b) If ancestors of atom added to $\mathrm{PI}(C)$ had arrows, merge them to atom in $\mathrm{PI}(C)$ (i.e. arrows starting in and leading to this atom).

   c) Replace colored terms in $\mathrm{PI}(C)$ (from new atom and unifier applied to $\mathrm{PI}(C_i)^*$) with fresh variables, except if a term has a double ended arrow to another overbinding variable, then use that variable.

      An arrow starts (ends) in one of the new variables if it starts (ends) somewhere in the term it replaced.

   d) Collect quantifiers: from $\mathrm{PI}(C_i)^*$, $i \in \{1,2\}$ and ones from atom added to $\mathrm{PI}(C)$. Order such that arrows only point to variables to the right.

      $\bar{Q}_n = \mathrm{sort}(Q_{n_1} \cup Q_{n_2} \cup \text{colored-terms}(l))$

Conjecture 9. *$\Gamma \cup \Delta$ unsat, π resolution refutation. Then $\Gamma \models \bar{Q}_n \, \mathrm{PI}(C)^* \vee C$ and $\Delta \models \neg\bar{Q}_n \, \mathrm{PI}(C)^* \vee C$ for all C in π.*

Proof. Base case as in Huang.
Induction.
Suppose Resolution.

$$\frac{C_1 : D \vee l \qquad C_2 : E \vee \neg l}{C : D \vee E}$$

$\Gamma \models \bar{Q}_{n_1} \, \mathrm{PI}(C_1)^* \vee D \vee l$
$\Gamma \models \bar{Q}_{n_2} \, \mathrm{PI}(C_2)^* \vee E \vee \neg l$
to show: $\Gamma \models \bar{Q}_n \, \mathrm{PI}(C)^* \vee D \vee E$, i.e.
$\Gamma \models \mathrm{sort}(Q_{n_1} \cup Q_{n_2} \cup \text{colored-terms}(l))((\neg l^* \wedge PI(C_1)^*) \vee (l^* \wedge PI(C_2)^*)) \vee D \vee E$
If $\Gamma \not\models D$ and $\Gamma \not\models E$ (else we are done), then
$\Gamma \models (\neg l \wedge \bar{Q}_{n_1} PI(C_1)^*) \vee (l \wedge \bar{Q}_{n_2} \, \mathrm{PI}(C_2)^*)$
As $\bar{Q}_{n_1}$ and $\bar{Q}_{n_2}$ disjoint and their variables do not appear in l or l,
$\Gamma \models \bar{Q}_{n_1} \bar{Q}_{n_2} [(\neg l \wedge PI(C_1)^*) \vee (l \wedge PI(C_2)^*)]$
Since we've pushed the variables outside, no colored terms appear in $\mathrm{PI}(C_i)^*$.
Suppose l does not contain colored terms. Then $l = l^*$ and we are done.
Otherwise let t be a maximal colored term in l.

By lemma 10, l appears in Γ with a certain polarity, say in clause E. l is an instance of E.
In fact, l is contained in $C\sigma$ where σ is the composition of unifiers applied in the deriviation up to the current point.
Hence $\Gamma \models C\sigma$.

1. Suppose t is Γ-colored. $\Gamma \models l$ implies that $\Gamma \models \exists y\, l[t/y]$
2. Suppose t is Δ-colored. TODO

$\square$

Lemma 10. *If an atom A appears in the interpolant, it appeared in both original clause sets, once positively and once negatively.*
A is contained in some instance of the respective clauses in Γ and Δ.

Lemma 11. *Let $C \in \Phi$ for some initial clause set Φ.*

1. *Let x be an occurrence of a variable in C and x' another occurrence of the same variable in a different position but at the same term depth. Then $\Phi \models QyC[x/y][x'/y]$ for $Q \in \{\forall, \exists\}$.*
2. *Let x be an occurrence of a variable in C with the lowest depth and x' another occurrence of the same variable with a higher depth. Let t be the maximal colored term which contains x'. t is Φ-colored since it appears in Φ. Then $\Phi \models Qy\exists zC[x/y]\{t/z\}$ for $Q \in \{\forall, \exists\}$.*

Lemma 12. *Let t be a maximal colored term in C in Φ. It is Φ-colored. Let $x_1, \ldots, x_n$ be the variables which occur in t. Then $\Phi \models Q\bar{x}\exists yC\{t/y\}$ for $Q \in \{\forall, \exists\}$.*