1 Arrow-Algo

1. In the original clauses, find all occurrences of variables.

   Common case: If a variable appears as outermost symbol or only has grey ancestor-terms, add an arrow from it to all other occurrences.

   Uncommon case: if there is more than one occurrence of a variable under a Φ-colored term, add a *weak* dependency between them all (symmetric relation).

   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 $\text{PI}(C_i)^*$, $i \in \{1,2\}$, i.e. use ancestor PI without colored terms.

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

   c) Replace colored terms in $\text{PI}(C)$ (from new atom and unifier applied to $\text{PI}(C_i)^*$) 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 $\text{PI}(C_i)^*$, $i \in \{1,2\}$ and ones from atom added to $\text{PI}(C)$. Order such that arrows only point to variables to the right AND weakly connected variables appear in the same quantifier block.

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

1.1 algo more formally

Every literal in any initial clause set has a globally unique id/number

Ex: $P(y, a, f(z, g(y, b))) \vee Q(x)$

Term position:

0.2.1.0 means first literal, 3rd arg, 2nd arg, fst arg: y

0.1 is a

0.2.1 is $g(z, b)$

ρ calculates the position of a term or the term of a position, depending on the argument type.

for a position p_i, $\rho(p_i)$ denotes whatever p_i refers to in its respective clause.

for a term t, $\rho(t)$ denotes the position in t in its respective clause.

for a position p, $\rho_{\text{lit}}(p)$ denotes the position of the literal

for a position p, $\rho_{\text{term}}(p)$ denotes the position of the term in p_i

$\Rightarrow p = \rho_{\text{lit}}(p) . \rho_{\text{term}}(p)$

for a position p, $p \mod i$ denotes p with i least significant places cut off, $0.2.1.0 \mod 2 = 0.2$

1.2 Arrows:

$\mathcal{A}$ is a set of ordered pairs of term positions which point to positions in terms in literals

$\mathcal{W}$ is a set of unordered pairs of term positions which point to positions in terms in literals

w.r.t a refutation π of $\Gamma \cup \Delta$:

1. For each initial clause C in $\Gamma \cup \Delta$:

   Add to $\mathcal{A}$ all (p_1, p_2) in C such that p_1 contains only grey symbol and $\rho(p_1)$ is a variable and $\rho(p_1) = \rho(p_2)$ but $p_1 \neq p_2$.

   Add to $\mathcal{W}$ all $\{p_1, p_2\}$ such that there is a colored symbol in p_1 and a possible different one in p_2 and $\rho(p_1)$ is a var and $\rho(p_1) = \rho(p_2)$.

2. For each C resulting from a resolution step from $C_1 : D \vee l$ and $C_2 : E \vee \neg l$ to $C = D \vee E$ with prop interpolant PI($\cdot$):

   Note: literals are added to the interpolant if they occur in both ancestors.

   Merge the respective ids of l and $\neg l$, i.e. their arrows. Their term structure will be the same, so all arrows point to valid positions.

1.3 algo

1.3.1 $\text{AI}_{\text{matrix}}$ and $\text{AI}_{\text{clause}}$

1. For each initial clause C, $\text{AI}_{\text{matrix}}(C) = \text{PI}(C)$ and $\text{AI}_{\text{clause}}(C) = \ell_{\Gamma \cup \Delta, z}[C]$.

2. For each C resulting from a resolution step from $C_1 : D \vee l$ and $C_2 : E \vee \neg l$ to $C = (D \vee E)\sigma$ with $l\sigma = l'\sigma$ with prop interpolant PI($\cdot$):

- if l and l' don't have the same color:

  NOTE: same overbinding for l and $\neg l$

  $$\mathrm{AI}_{\mathrm{matrix}}(C)' = \Big((\neg l \wedge \mathrm{AI}_{\mathrm{matrix}}(C_1)) \vee (l \wedge \mathrm{AI}_{\mathrm{matrix}}(C_2))\Big)\sigma$$

  $$\mathrm{AI}_{\mathrm{clause}}(C)' = \Big((\mathrm{AI}_{\mathrm{clause}}(C_1) \setminus \{l\}) \vee (\mathrm{AI}_{\mathrm{clause}}(C_2) \setminus \{\neg l'\})\Big)\sigma \;//\; \text{setminus: remove clause with that ancestor}$$

  same term in matrix/clause $\Rightarrow$ overbound by same variable

  $$\ell_{\Gamma \cup \Delta, z}[\mathrm{AI}_{\mathrm{matrix}}(C)' \vee \mathrm{AI}_{\mathrm{clause}}(C)'] = \underbrace{\ell_{\Gamma \cup \Delta, z}[\mathrm{AI}_{\mathrm{matrix}}(C)']}_{\mathrm{AI}_{\mathrm{matrix}}(C)} \vee \underbrace{\ell_{\Gamma \cup \Delta, z}[\mathrm{AI}_{\mathrm{clause}}(C)']}_{\mathrm{AI}_{\mathrm{clause}}(C)}$$

- if l and l' have the same color:

  as above but with $\mathrm{AI}_{\mathrm{matrix}}(C)' = \Big(\mathrm{AI}_{\mathrm{matrix}}(C_1) \vee \mathrm{AI}_{\mathrm{matrix}}(C_2)\Big)\sigma$

1.3.2 AI

$\mathrm{AI}(C) = Q_1 u_1 \ldots Q_m u_m \, \mathrm{AI}_{\mathrm{matrix}}(C)$

$u_1, \ldots, u_m$ are comprised of all x_i and y_i PLUS all free variables in $\mathrm{AI}_{\mathrm{matrix}}(C)$.

Q_i is $\exists$ if $u_i = y_i$ for some i, $\forall$ if $u_i = x_i$ for some i. For free variables, Q_i as in C_1/C_2. if $u_i \sigma \neq u_i$, u_i becomes one of x_j or y_j or grey term.

$(p_1, p_2) \in \mathrm{TransitiveClosure}(\mathcal{A})$ implies that $u_i < u_j$ if u_i replaces t_i and u_j replaces t_j and p_1 points into t_i and p_2 points into t_j.

2 current proof attempts

Lemma 1. *Whenever the same variable appears multiple times in $\mathrm{PI}(C) \vee C$ for $C \in \pi$, there are arrows.*

- *if both variables are contained only in grey terms, there is a double arrow // they unify to exactly the same*

- *if only one variable is only contained in grey terms, there is an arrow from it to the other one // either unify the one in grey term, then other one must be overbound later. if otherwise var in the colored term is unified, we can still overbind the grey one first.*

- *otherwise there are weak arrows between them // have same quantifier, so order does not matter, but want to keep dependencies on both the same*

Proof. By induction. Note: As required by resolution, all initial clauses are variable disjoint.

Base case: In the initial clause sets, consider for a clause C two different positions p_1 and p_2 pointing to the same variable. Then either:

- p_1 and p_2 contain only grey symbols. Then $(p_1, p_2) \in \mathcal{A}$.

- Only p_i, $i \in \{1, 2\}$ contains only grey symbols. Then $(p_i, p_{(i \mod 2)+1}) \in \mathcal{A}$.

- There are not only grey symbols in both p_1 and p_2, i.e. both contain at least a colored symbol. Then $\{p_1, p_2\} \in \mathcal{W}$.

Induction step: Suppose a clause C is the result of a resolution of $C_1 : D \vee l$ and $C_2 : E \vee \neg l$ with $l\sigma = l'\sigma$. $\mathrm{PI}(C)$ is $[\mathrm{PI}(C_1) \circ \mathrm{PI}(C)]\sigma$ or $[(l \wedge \mathrm{PI}(C_2)) \vee (\neg l \wedge \mathrm{PI}(C_1))]\sigma$.

Assumption: C_1 and C_2 are variable disjoint, i.e. variables are renamed in case C_1 and C_2 are derived from some common original clause and share variables.

By the induction hypothesis, there are appropriate arrows in both $\mathrm{PI}(C_i) \vee C_i$, $i \in \{1, 2\}$.

If the variables were present in C_1 or C_2, the arrow is still there, either in $\mathrm{PI}(C)$ (in the case of l or l'), C (in case of D and E) or in currently not shown literal (in case l and l' have the same color). Otherwise, it was introduced by unification in $l\sigma$ or $\mathrm{PI}(C_i)\sigma$. In this case, there is some term position q in with $\rho(l).q$ a variable and $\rho(l').q$ a variable or a term containing variables (or other way around). Hence unification maps a variable to a variable or a term containing variables. The variable being unified is in $\mathrm{PI}(C_i) \vee C_i$ for some $i \in \{1, 2\}$. But by the induction hypothesis, all occurrences of each variable does already have appropriate arrows, which are still present. $\square$

Lemma 2. *In $\mathrm{PI}(C) \vee C$ for $C \in \pi$, if there is a Δ-colored term s in a Γ-term t, then there is an arrow from p_1 to p_2 such that $\rho(p_1) = s$ and $\rho(p_2) = s$ and for some i, $\rho(p_2 \mod i) = t$.*

Note: p_1 might be in some clause, the prop interpolant or none of both.

Proof. By induction.

Base case: There are no foreign terms in the initial clause sets, so no arrows necessary.

Induction step:

Resolution. Suppose a clause C is the result of a resolution of $C_1 : D \lor l$ and $C_2 : E \lor \neg l$ with $l\sigma = l'\sigma$.

1. Suppose l is colored. This case is similar to the grey one, with the exception that the cases applying to l in PI do not apply.

2. Suppose l is grey. Then $\mathrm{PI}(C) = [(l \land \mathrm{PI}(C_2)) \lor (\neg l \land \mathrm{PI}(C_1))]\sigma$

   By the induction hypothesis, there are appropriate arrows in $\mathrm{PI}(C_1) \lor C_1$ and $\mathrm{PI}(C_2) \lor C_2$.

   We show that for all maximal Γ-terms in $\mathrm{PI}(C) \lor C$ with Δ-terms in them which were not present in $\mathrm{PI}(C_i) \lor C_i$, $i \in \{1, 2\}$, there is an arrow.

   Γ-terms and Δ-terms are not unifiable. Hence all pairs of terms (t_1, t_2) in the same positions in l and l' (if both positions exist) either point to the same symbol or (w.l.o.g.) t_1 is a variable and t_2 is a term. If there are Δ-terms in Γ-terms in the prefix, they are present in both ancestors and handled by the induction hypothesis.

   The only way a Δ-colored term may enter a Γ-colored term is in the situation where t_1 is a variable and t_2 a colored term. But then $\mathrm{mgu}(t_1, t_2)$ applied to t_1 yields t_2, i.e. "the parts of σ concerned with unifying t_1 and t_2" do not introduce new Δ-terms in Γ-terms. In other words, all such situation have been present in $\mathrm{PI}(C_i) \lor C_i$ for $i \in \{1, 2\}$ and since the arrows for l and l' are merged, they are present for $l\sigma$ in $\mathrm{PI}(C)$.

   This handles the case where terms t_1 and t_2 are unified. But unification also affects all other occurrences of variables, this means "the parts of σ not concerned with unifying t_1 and t_2". The relevant case for this lemma is when a Γ-term contains a variable, that is substituted by a term containing Δ-terms. But in this case, by Lemma 1, there is an arrow from the other occurrence of the variable to the one in the Γ-term: either double arrow in $\mathcal{A}$ if both prefixes are grey, one in $\mathcal{A}$ if one of the prefixes is grey or one in $\mathcal{W}$ if both prefixes contain a colored symbol. $\qquad\square$

Lemma 3. $\mathrm{AI}_{\mathrm{matrix}}(C) \sim \ell_{\Gamma \cup \Delta, z}[PI(C)]$ for $C \in \pi$.

Proof. Clearly the propositional structure is the same.

The coloring of the terms is the same, but some variables are "split up" into multiple ones in $\mathrm{AI}_{\mathrm{matrix}}$ as not all terms of the same form are replaced by the same variable if the terms have entered in different stages.

-> these are only the variables which have no logical connection, TODO: formally define what this means

$\qquad\square$